SISTEMAS DE COMUNICACIONES ELECTRÓNICAS

CUARTA EDICIÓN

Prentice Hall ®

TOMASI

Imagen protegida por derechos de autor

Contenido

CAPÍTULO 1 INTRODUCCIÓN A LAS COMUNICACIONES ELECTRÓNICAS 1

INTRODUCCIÓN 1
SISTEMAS ELECTRÓNICOS DE COMUNICACIONES 2
MODULACIÓN Y DEMODULACIÓN 2
EL ESPECTRO ELECTROMAGNÉTICO 4
Frecuencias de transmisión 5
Clasificación de los transmisores 7
ANCHO DE BANDA Y CAPACIDAD DE INFORMACIÓN 8
MODOS DE TRANSMISIÓN 10
Símplex (SX) 10
Semidúplex (HDX, de half duplex) 10
Dúplex total (FDX, de full duplex) 10
Dúplex total/general (F/FDX, de full/full duplex) 10
CONFIGURACIONES DE LOS CIRCUITOS 10
Transmisión a dos hilos 11
Transmisión a cuatro hilos 12
Híbridos y supresores de eco 12
ANÁLISIS DE SEÑALES 14
Señales senoidales 14
Ondas periódicas no senoidales (ondas complejas) 16
Serie de Fourier para una forma de onda rectangular 22
Espectros de potencia y energía 26
Transformadas de Fourier discreta y rápida 26
Efectos de limitación de banda sobre las señales 27
MEZCLADO 27
Suma lineal 27
Mezclado no lineal 29

ANÁLISIS DE RUIDO 34
Ruido no correlacionado 34
Voltaje del ruido 37
Ruido correlacionado 38
Ruido impulsivo 40
Interferencia 40
Resumen del ruido 40
Relación de potencia de señal a ruido 40
Factor de ruido y cifra de ruido 42
Temperatura equivalente de ruido 45
PREGUNTAS 46
PROBLEMAS 47

CAPÍTULO 2 GENERACIÓN DE SEÑAL 51

INTRODUCCIÓN 51
OSCILADORES 51
Osciladores retroalimentados 52
Osciladores no sintonizados 54
Osciladores sintonizados 55
Estabilidad de la frecuencia 59
Osciladores de cristal 60
OSCILADORES EN GRAN ESCALA DE INTEGRACIÓN 67
Generación de forma de onda con circuito integrado 68
LAZOS DE FASE CERRADA 72
Intervalos de enganche y de captura 73
Oscilador controlado por voltaje 74
Comparador de fases 75
Funcionamiento del lazo 80
Circuito integrado para un lazo de fase cerrada de precisión 84
SINTETIZADORES DE FRECUENCIAS 89
Sintetizadores directos de frecuencias 89
Sintetizadores indirectos de frecuencias 92
PREGUNTAS 95
PROBLEMAS 96

CAPÍTULO 3 TRANSMISIÓN POR MODULACIÓN DE AMPLITUD 100

INTRODUCCIÓN 100
PRINCIPIOS DE MODULACIÓN DE AMPLITUD 100
La envolvente de AM 101
Espectro de frecuencias y ancho de banda de AM 101
Representación fasorial de una onda de amplitud modulada 103
Coeficiente de modulación y porcentaje de modulación 103
Distribución de voltaje de AM 107
Análisis de AM en el dominio del tiempo 110
Distribución de potencia en AM 111
Cálculos de corriente en AM 114
Modulación con una señal compleja de información 115
CIRCUITOS MODULADORES DE AM 116
Modulador de AM de bajo nivel 116
Modulador de AM de potencia intermedia 118

Modulación simultánea de base y colector 122
Moduladores de AM en circuito integrado lineal 122
TRANSMISORES DE AM 127
Transmisores de bajo nivel 127
Transmisores de alto nivel 127
Patrones trapezoidales 127
Desplazamiento de la portadora 131
Envolventes de AM producidas por señales complejas no senoidales 132
MODULACIÓN DE AMPLITUD EN CUADRATURA 135
PREGUNTAS 135
PROBLEMAS 136

CAPÍTULO 4 RECEPCIÓN DE AMPLITUD MODULADA 140

INTRODUCCIÓN 140
PARÁMETROS DEL RECEPTOR 140
Selectividad 141
Mejoramiento del ancho de banda 142
Sensibilidad 142
Margen dinámico 143
Fidelidad 143
Pérdida de inserción 144
Temperatura de ruido y temperatura equivalente de ruido 145
RECEPTORES DE AM 145
Receptor de radiofrecuencia sintonizada 145
Receptor superheterodino 147
CIRCUITOS RECEPTORES DE AM 158
Circuitos amplificadores de RF 158
Amplificadores de bajo ruido 161
Circuitos de mezclador/convertidor 161
Circuitos amplificadores de FI 165
Circuitos detectores de AM 172
Controles automáticos de ganancia 176
Circuitos de reducción de ruido 178
Limitadores y eliminadores de ruido 179
Medidas alternas de señal a ruido 181
Receptores de AM en circuito integrado lineal 182
RECEPTORES DE AM DE DOBLE CONVERSIÓN 183
GANANCIA NETA DEL RECEPTOR 183
PREGUNTAS 186
PROBLEMAS 187

CAPÍTULO 5 SISTEMAS DE COMUNICACIONES DE BANDA LATERAL ÚNICA 189

INTRODUCCIÓN 189
SISTEMAS DE BANDA LATERAL ÚNICA 189
AM de banda lateral única y portadora de máxima potencia 190
AM de banda lateral única y portadora suprimida 191
AM de banda lateral única y portadora reducida 191
AM de banda lateral independiente 192
AM de banda lateral residual 193
Comparación de la transmisión con banda lateral única y la AM convencional 193

ANÁLISIS MATEMÁTICO DE AM CON PORTADORA SUPRIMIDA 197
GENERACIÓN DE BANDA LATERAL ÚNICA 197
Modulador de anillo balanceado 197
Modulador balanceado en contrafase (push-pull) con FET 200
Modulador de puente balanceado 201
Moduladores balanceados en circuito integrado lineal 203
TRANSMISORES DE BANDA LATERAL ÚNICA 205
Transmisor de banda lateral única: método con filtro 205
Transmisor de banda lateral única: método de desplazamiento de fase 211
Transmisor de banda lateral única: tercer método 213
Transmisor de banda lateral independiente 214
RECEPTORES DE BANDA LATERAL ÚNICA 214
Receptor BFO de banda lateral única 214
Receptor BFO coherente de banda lateral única 217
Receptor de detección de envolvente con banda lateral única 218
Receptor de banda lateral única con portadora piloto en multicanales 218
COMPRESIÓN Y EXPANSIÓN DE AMPLITUD DE BANDA LATERAL ÚNICA 220
BANDA LATERAL ÚNICA CON PORTADORA SUPRIMIDA Y MULTIPLEXADO POR DIVISIÓN DE FRECUENCIAS 221
DOBLE BANDA LATERAL CON PORTADORA SUPRIMIDA Y MULTIPLEXADO EN CUADRATURA 221
MEDICIONES DE BANDA LATERAL ÚNICA 223
PREGUNTAS 225
PROBLEMAS 225

CAPÍTULO 6 TRANSMISIÓN POR MODULACIÓN ANGULAR 228

INTRODUCCIÓN 228
MODULACIÓN ANGULAR 228
Análisis matemático 231
Sensibilidad a la desviación 232
Formas de onda de FM y PM 233
Desviación de fase e índice de modulación 234
Desviación de frecuencia 235
Moduladores y demoduladores de fase y de frecuencia 237
Análisis en frecuencia de las ondas con modulación angular 238
Requisitos de ancho de banda para ondas con modulación angular 241
BANDA COMERCIAL DE EMISIÓN EN FM 245
REPRESENTACIÓN FASORIAL DE LA ONDA CON MODULACIÓN ANGULAR 245
POTENCIA PROMEDIO DE UNA ONDA DE MODULACIÓN ANGULAR 246
RUIDO Y MODULACIÓN ANGULAR 248
Modulación de fase debida a una señal de interferencia 248
Modulación de frecuencia debida a una señal de interferencia 249
PREÉNFASIS Y DEÉNFASIS 251
MODULADORES DE FRECUENCIA Y FASE 253
Moduladores directos de FM 253
Moduladores directos de PM 258
CONVERSIÓN ELEVADORA DE FRECUENCIA 259
Método de heterodinado para elevar la frecuencia 259
Método de multiplicación para aumento de frecuencia 261

TRANSMISORES DIRECTOS DE FM 262
Transmisor directo de FM de Crosby 262
Trasmisor directo de FM con lazo de fase cerrada 265
TRANSMISORES INDIRECTOS DE FM 266
Transmisor indirecto de FM de Armstrong 266
Comparación de FM y PM 270
MODULACIÓN ANGULAR EN COMPARACIÓN CON MODULACIÓN DE AMPLITUD 270
Ventajas de la modulación angular 270
Desventajas de la modulación angular 270
PREGUNTAS 271
PROBLEMAS 272

CAPÍTULO 7 RECEPTORES POR MODULACIÓN DE ÁNGULO, FM ESTÉREO Y RADIO DE FM EN DOS SENTIDOS 275

INTRODUCCIÓN 275
RECEPTORES DE FM 276
DEMODULADORES DE FM 276
Discriminadores de frecuencia con circuito sintonizado 278
DEMODULADOR DE FM CON LAZO DE FASE CERRADA 282
DEMODULADOR DE FM POR CUADRATURA 284
SUPRESIÓN DE RUIDO EN FM 285
Limitadores de amplitud y umbral de FM 285
Circuitos limitadores 287
Efecto de captura de FM 287
COMPARACIÓN DE MODULACIÓN DE FRECUENCIA Y MODULACIÓN DE FASE 289
RECEPTORES DE FM EN CIRCUITO INTEGRADO LINEAL 290
Sistema de FI para FM en circuito integrado y baja potencia 290
Sistema de FI para FM de alto rendimiento, bajo voltaje y con mezclador 293
Sistema de radio de FM con un solo CI 293
EMISIÓN ESTEREOFÓNICA EN FM 294
Transmisión estereofónica de FM 297
Recepción estereofónica de FM 299
RADIOCOMUNICACIONES EN FM DE DOS SENTIDOS 303
Perspectiva histórica 304
Radiotransmisor de FM en dos sentidos 304
Radiorreceptor de FM en dos sentidos 305
PREGUNTAS 308
PROBLEMAS 309

CAPÍTULO 8 LÍNEAS DE TRANSMISIÓN 310

INTRODUCCIÓN 310
ONDAS ELECTROMAGNÉTICAS TRANSVERSALES 310
Características de las ondas electromagnéticas 311
TIPOS DE LÍNEAS DE TRANSMISIÓN 313
Balunes 314
Líneas de transmisión de conductores paralelos 315
Líneas de transmisión concéntricas o coaxiales 316
CIRCUITO EQUIVALENTE DE UNA LÍNEA DE TRANSMISIÓN 317
Líneas uniformemente distribuidas 317
Características de transmisión 317

PROPAGACIÓN DE ONDAS EN LÍNEA DE TRANSMISIÓN 322
Factor de velocidad 322
Longitud eléctrica de una línea de transmisión 324
Líneas de retardo 324
PÉRDIDAS EN LÍNEAS DE TRANSMISIÓN 325
Pérdidas en el conductor 325
Pérdida por calentamiento del dieléctrico 326
Pérdida por radiación 326
Pérdida por acoplamiento 326
Efecto corona (o efecto de arco voltaico) 326
ONDAS INCIDENTES Y REFLEJADAS 327
Líneas de transmisión resonantes y no resonantes 327
Coeficiente de reflexión 327
ONDAS ESTACIONARIAS 328
Relación de onda estacionaria 328
Ondas estacionarias en una línea abierta 330
Ondas estacionarias en una línea en corto 333
IMPEDANCIA DE ENTRADA DE UNA LÍNEA DE TRANSMISIÓN 333
Análisis fasorial de la impedancia de entrada: línea abierta 334
Análisis fasorial de la impedancia de entrada: línea en cortocircuito 337
Acoplamiento de impedancias en líneas de transmisión 338
REFLECTOMETRÍA EN EL DOMINIO DEL TIEMPO 340
LÍNEAS DE TRANSMISIÓN DE MICROCINTA Y DE CINTA 342
Microcinta 342
Línea de cinta 343
PREGUNTAS 344
PROBLEMAS 345

CAPÍTULO 9 PROPAGACIÓN DE LAS ONDAS ELECTROMAGNÉTICAS 347

INTRODUCCIÓN 347
POLARIZACIÓN ELECTROMAGNÉTICA 348
RAYOS Y FRENTES DE ONDA 348
RADIACIÓN ELECTROMAGNÉTICA 348
Densidad de potencia e intensidad de campo 348
Impedancia característica del espacio libre 349
FRENTE DE ONDA ESFÉRICO Y LA LEY DEL CUADRADO INVERSO 350
Frente de onda esférico 350
Ley del cuadrado inverso 350
ATENUACIÓN Y ABSORCIÓN DE ONDAS 351
Atenuación 351
Absorción 352
PROPIEDADES ÓPTICAS DE LAS ONDAS DE RADIO 353
Refracción 353
Reflexión 354
Difracción 357
Interferencia 357
PROPAGACIÓN TERRESTRE DE LAS ONDAS ELECTROMAGNÉTICAS 359
Propagación de ondas terrestres 359
Propagación de las ondas espaciales 361
Propagación por ondas celestes 362

TÉRMINOS Y DEFINICIONES DE PROPAGACIÓN 364
Frecuencia crítica y ángulo crítico 364
Altura virtual 365
Máxima frecuencia útil 365
Distancia de salto 366
PÉRDIDAS EN TRAYECTORIA POR EL ESPACIO LIBRE 366
MARGEN DE DESVANECIMIENTO 367
PREGUNTAS 368
PROBLEMAS 369

CAPÍTULO 10 ANTENAS Y GUÍAS DE ONDA 371

INTRODUCCIÓN 371
FUNCIONAMIENTO BÁSICO DE LA ANTENA 371
Circuitos equivalentes a una antena 372
RECIPROCIDAD DE ANTENAS 372
TERMINOLOGÍA Y DEFINICIONES PARA ANTENAS 373
Sistema de coordenadas de la antena 373
Diagrama de radiación 373
Campos cercano y lejano 375
Resistencia de radiación 376
Ganancia directiva y ganancia de potencia 377
Potencia isotrópica efectiva irradiada 378
Densidad de potencia capturada 380
Área de captura y potencia capturada 380
Polarización de antena 382
Abertura del haz de la antena 382
Ancho de banda de antena 383
Impedancia de entrada a la antena 383
ANTENAS BÁSICAS 383
Doblete elemental 383
Dipolo de media onda 385
Antena conectada a tierra 388
CARGA DE LA ANTENA 389
Bobinas de carga 389
Carga por el extremo superior 390
CONJUNTOS DE ANTENAS 391
Red de radiación lateral 391
Red de radiación longitudinal 392
Conjunto no resonante: la antena rómbica 393
ANTENAS DE USO ESPECIAL 394
Dipolo doblado 394
Antena de torniquete 394
Antena log-periódica 396
Antena de cuadro 397
Antenas de conjunto enfasado 398
Antena helicoidal 398
ANTENAS DE UHF Y DE MICROONDAS 400
Antena de reflector parabólico 401
Antena de cuerno cónico 408
GUÍAS DE ONDA 409
Guía de onda rectangular 410

OTROS TIPOS DE GUÍAS DE ONDA 417
Guía de onda circular 417
Guía de onda rígida 418
Guía de onda flexible 418
PREGUNTAS 419
PROBLEMAS 420

CAPÍTULO 11 COMUNICACIONES CON FIBRA ÓPTICA 422

INTRODUCCIÓN 422
HISTORIA DE LAS FIBRAS ÓPTICAS 422
COMPARACIÓN DE INSTALACIONES CON FIBRAS ÓPTICAS Y CON CABLES METÁLICOS 423
Ventajas de los sistemas de fibra óptica 423
Desventajas de los sistemas de fibra óptica 424
EL ESPECTRO ELECTROMAGNÉTICO 425
DIAGRAMA DE BLOQUES DEL SISTEMA DE COMUNICACIONES CON FIBRA ÓPTICA 425
TIPOS DE FIBRA 427
Construcción del cable 428
PROPAGACIÓN DE LA LUZ 429
La física de la luz 429
Velocidad de propagación 430
Refracción 430
Índice de refracción 430
Ángulo crítico 433
PROPAGACIÓN DE LA LUZ A TRAVÉS DE UNA FIBRA ÓPTICA 434
Modo de propagación 434
Perfil índice 435
CONFIGURACIONES DE FIBRA ÓPTICA 435
Fibra unimodal de índice escalonado 435
Fibra multimodal de índice escalonado 437
Fibra multimodal de índice graduado 437
Comparación de los tres tipos de fibra óptica 438
ÁNGULO DE ACEPTACIÓN Y CONO DE ACEPTACIÓN 439
Abertura numérica 440
PÉRDIDAS EN LOS CABLES DE FIBRA ÓPTICA 442
Pérdidas por absorción 444
Pérdidas en material, o por dispersión de Rayleigh 444
Dispersión cromática o de longitudes de onda 445
Pérdidas por radiación 446
Dispersión modal 446
Pérdidas en acoplamiento 448
FUENTES LUMINOSAS 450
Potencia óptica 451
FUENTES ÓPTICAS 452
Diodos emisores de luz 452
Diodo de láser de inyección 457

DETECTORES DE LUZ 458
Diodos PIN 458
Fotodiodos de avalancha 460
Características de los detectores de luz 460
LÁSERES 460
Tipos de láser 461
Características de los láseres 461
Construcción de los láseres 461
PROVISIÓN PARA ENLACE EN EL SISTEMA DE FIBRA ÓPTICA 462
PREGUNTAS 464
PROBLEMAS 465

CAPÍTULO 12 COMUNICACIONES DIGITALES 467

INTRODUCCIÓN 467
COMUNICACIONES DIGITALES 468
LÍMITE DE SHANNON DE CAPACIDAD DE INFORMACIÓN 468
RADIO DIGITAL 470
MODULACIÓN DIGITAL DE AMPLITUD 470
MANIPULACIÓN POR DESPLAZAMIENTO DE FRECUENCIA 471
Rapidez de bits FSK y baudios 472
Transmisor FSK 473
Consideraciones de FSK respecto del ancho de banda 474
Receptor FSK 476
Manipulación por desplazamiento de frecuencia con fase continua 477
MANIPULACIÓN POR DESPLAZAMIENTO DE FASE 478
Manipulación por desplazamiento binario de fase 478
Codificación M-aria 482
Manipulación por desplazamiento cuaternario de fase 484
PSK de ocho fases 490
PSK de dieciséis fases 496
MODULACIÓN DE AMPLITUD EN CUADRATURA 496
Ocho QAM 496
Dieciséis QAM 500
EFICIENCIA DE ANCHO DE BANDA 504
RECUPERACIÓN DE PORTADORA 505
Lazo cuadrático 505
Lazo de costas 506
Remodulador 507
MANIPULACIÓN POR DESPLAZAMIENTO DIFERENCIAL DE FASE 507
BPSK diferencial 507
RECUPERACIÓN DE RELOJ 509
PROBABILIDAD DE ERROR Y TASA DE ERRORES DE BITS 509
Errores en PSK 512
Errores en QAM 515
Errores en FSK 516
CODIFICACIÓN TRELLIS 517
PREGUNTAS 521
PROBLEMAS 522

CAPÍTULO 13 COMUNICACIÓN DE DATOS 524

INTRODUCCIÓN 524
HISTORIA DE LA COMUNICACIÓN DE DATOS 524
ORGANIZACIONES NORMATIVAS PARA COMUNICACIÓN DE DATOS 525
CIRCUITOS DE COMUNICACIÓN DE DATOS 526
- Transmisión de datos en serie y en paralelo 526
- Configuraciones 528
- Topologías 528
- Modos de transmisión 528
- Comparación de la operación con dos y con cuatro hilos 529

CÓDIGOS DE COMUNICACIÓN DE DATOS 530
- Código de Baudot 530
- Código ASCII 531
- Código EBCDIC 533
- Códigos de barras 533

CONTROL DE ERRORES 535
- Detección de errores 535
- Corrección de errores 541

SINCRONIZACIÓN 543
- Sincronización de caracteres 543

COMPONENTES EN LA COMUNICACIÓN DE DATOS 545
- Unidad de control de línea (LCU) 546

INTERFACES EN MODO SERIE 552
INTERFAZ RS-232 552
- Normas de las interfaces RS-449 y RS-530 557
- CCITT X.21 558

INTERFACES PARALELAS 559
- Interfaz paralela Centronics 560
- Bus IEEE 488 563

LA RED TELEFÓNICA 566
- Red DDD 566
- Servicio de línea privada 571

EL CIRCUITO TELEFÓNICO 572
- El lazo local 572
- Parámetros de transmisión 574

MÓDEMS DE DATOS 588
- Módems asíncronos 589
- Módems síncronos 590
- Sincronización del módem 591
- Módems de baja velocidad 591
- Módems de velocidad media y alta 592
- Control del módem–conjunto de órdenes AT 595
- Recomendaciones de CCITT para los módems 596
- Recomendación CCITT V.29 596
- Recomendación CCITT V.32 para módems 598
- Recomendaciones CCITT V.32bis y V.32terbo para módems 598
- Recomendación CCITT V.33 para módems 599
- Recomendación CCITT V.42 y V.42bis para módems 599
- Recomendación CCITT V.32 (V.fast) para módems 600
- Recomendación CCITT V.34+ para módems 601

PREGUNTAS 601
PROBLEMAS 603

CAPÍTULO 14 PROTOCOLOS DE COMUNICACIÓN DE DATOS Y CONFIGURACIONES DE RED 605

INTRODUCCIÓN 605
INTERCONEXIÓN DE SISTEMAS ABIERTOS 606
La jerarquía de protocolos ISO 606
Modo de carácter 608
Modo de bloque 608
PROTOCOLOS ASÍNCRONOS 608
PROTOCOLOS SÍNCRONOS 610
Protocolo bisync de IBM 610
Comunicaciones de enlace de datos síncronos 614
Operación del lazo SDLC 621
Transparencia 623
Abortar mensaje 623
Codificación por inversión en cero 623
Control de alto nivel para enlace de datos 624
RED PÚBLICA DE DATOS 626
Red con valor agregado 626
Red de conmutación de paquetes 626
Clase de servicio CCITT X.1 para usuario internacional 627
PROTOCOLO DE USUARIO Y RED CCITT X.25 628
Circuito virtual permanente 629
Llamada virtual 629
Diagrama de datos 630
Formato de paquete X.25 630
Las normas recomendadas de la serie X 631
Modo de transferencia asíncrona 631
Red óptica síncrona 634
RED DIGITAL DE SERVICIOS INTEGRADOS (ISDN) 636
Introducción a la ISDN 636
Arquitectura de la ISDN 638
Conexiones y unidades de interfaz del sistema ISDN 639
Protocolos ISDN 640
ISDN de banda ancha 643
REDES DE ÁREA LOCAL 645
Consideraciones sobre sistemas de red de área local 647
ANILLO DE PASO DE TESTIGO 651
Anillo con muescas 654
Bus de testigo 654
ETHERNET 654
Ethernet 10BASE-5 655
Ethernet 10BASE-2 656
Ethernet 10BASE-T 656
Formato de datos en Ethernet 657
Operación del sistema 658
INTERFAZ DE DATOS DISTRIBUIDOS EN FIBRA 659
Nodos FDDI 660

Red FDDI con líneas de transmisión de par de conductores metálicos trenzados 662
Canal Fibre 662
Ethernet 100BASE-T 663
PREGUNTAS 663
PROBLEMAS 665

CAPÍTULO 15 TRANSMISIÓN DIGITAL 667

INTRODUCCIÓN 667
Ventajas de la transmisión digital 667
Desventajas de la transmisión digital 668
MODULACIÓN DE PULSO 668
MODULACIÓN POR CÓDIGO DE PULSO 668
Muestreo PCM 669
Frecuencia de muestreo 673
Código binario reflejado 675
Rango dinámico 677
Eficiencia de codificación 680
Relación de señal a ruido de cuantización 680
Comparación de códigos PCM lineales y no lineales 681
Ruido de canal inactivo 682
Métodos de codificación 682
Compresión-expansión 683
Compresión-expansión analógica 683
Compresión-expansión digital 686
Error porcentual 693
Codificadores de voz 693
Velocidad de PCM de la línea 694
PCM DE MODULACIÓN DELTA 695
Transmisor con modulación delta 695
Receptor con modulación delta 696
PCM DE MODULACIÓN DELTA ADAPTATIVA 697
MODULACIÓN POR CÓDIGO DE IMPULSOS DIFERENCIAL 698
TRANSMISIÓN DE PULSOS 698
Interferencia entre símbolos 700
Patrones de ojo 702
POTENCIA DE SEÑALES DIGITALES BINARIAS 703
PREGUNTAS 704
PROBLEMAS 705

CAPÍTULO 16 MULTIPLEXADO 708

INTRODUCCIÓN 708
MULTIPLEXADO POR DIVISIÓN DE TIEMPO 708
SISTEMA DE PORTADORA DIGITAL T1 709
Bancos de canales tipo D 712
Formato de supertrama 713
Formato de supertrama extendido 714
SISTEMA CCITT DE PORTADORA MULTIPLEXADA POR DIVISIÓN DE TIEMPO 717
CODECS 717

CHIPS COMBINADOS 718
Funcionamiento general 718
Modo de velocidad fija de datos 719
Modo de velocidad variable de datos 726
Señalización de supervisión 728
JERARQUÍA DIGITAL EN NORTEAMÉRICA 728
Terminales de grupo maestro y de televisión comercial 728
Terminal de teléfono visual 730
Terminal de datos 731
CODIFICACIÓN DE LÍNEA 732
Voltajes de transmisión y componente de CD 732
Ciclo de trabajo 732
Consideraciones de ancho de banda 733
Recuperación de reloj 734
Detección de errores 734
Facilidad de detección y decodificación 734
Bifásica digital, de Miller y dicódigos 734
PORTADORAS T 736
Sistemas de portadora T1 y T1C 737
Sistema de portadora T2 738
Sistema de portadora T3 739
Sistema de portadora T4M 739
Sistema de portadora T5 739
SINCRONIZACIÓN DE TRAMA 739
Sincronización por dígito agregado 739
Sincronía por reemplazo de dígitos 740
Sincronía por canal agregado 740
Sincronía estadística 740
Sincronización por código de una línea 741
INTERCALACIÓN DE BIT E INTERCALACIÓN DE PALABRA 741
MULTIPLEXADO ESTADÍSTICO POR DIVISIÓN DE TIEMPO 741
MULTIPLEXADO POR DIVISIÓN DE FRECUENCIA 743
JERARQUÍA FDM DE AT&T 744
Comunicaciones de larga distancia con FDM 744
Canal de mensaje 744
Grupo básico 745
Supergrupo básico 745
Grupo maestro básico 745
Agrupamientos mayores 746
SEÑAL COMPUESTA EN BANDA BASE 746
Formación de grupos y supergrupos 746
FORMACIÓN DE UN GRUPO MAESTRO 748
Grupo maestro U600 748
Grupo maestro L600 751
Formación de un canal de radio 752
MULTIPLEXADO POR DIVISIÓN DE LONGITUD DE ONDA 754
Comparación de multiplexado por división de longitud de onda y por división de frecuencia 754
D-WDM, longitudes de onda y canales de longitudes de onda 754
Ventajas y desventajas del WDM 757
Componentes de los circuitos WDM 757

Acopladores por división de longitud de onda 758
WDM y la red óptica síncrona 758
PREGUNTAS 758
PROBLEMAS 759

CAPÍTULO 17 RADIOCOMUNICACIONES POR MICROONDAS Y GANANCIA DEL SISTEMA 761

INTRODUCCIÓN 761
VENTAJAS DE LAS RADIOCOMUNICACIONES POR MICROONDAS 762
MICROONDAS ANALÓGICAS Y DIGITALES 763
MODULACIÓN DE FRECUENCIA Y DE AMPLITUD 763
SISTEMA DE RADIO DE MICROONDAS CON FRECUENCIA MODULADA 763
Radiotransmisor de microondas de FM 764
Radiorreceptor de microondas de FM 765
RADIO REPETIDORAS DE MICROONDAS DE FM 765
DIVERSIDAD 768
Diversidad de frecuencia 769
Diversidad espacial 769
Diversidad de polarización 771
Diversidad híbrida 771
Diversidad cuádruple 771
ARREGLOS DE CONMUTACIÓN DE PROTECCIÓN 771
Reserva continua 771
Diversidad 773
Confiabilidad 773
ESTACIONES DE RADIO DE MICROONDAS FM 773
Estación terminal 773
Estación repetidora 776
CARACTERÍSTICAS DE LA TRAYECTORIA 781
Desvanecimiento 782
GANANCIA DEL SISTEMA 782
Pérdidas en la trayectoria en espacio libre 783
Margen de desvanecimiento 784
Umbral del receptor 785
Relaciones de portadora a ruido y de señal a ruido 786
Factor de ruido e índice de ruido 786
PREGUNTAS 790
PROBLEMAS 790

CAPÍTULO 18 COMUNICACIONES SATELITALES 793

INTRODUCCIÓN 793
HISTORIA DE LOS SATÉLITES 794
LEYES DE KEPLER 795
ÓRBITAS DE SATÉLITES 797
Categorías de elevación de satélites 797
Trayectorias orbitales de los satélites 798
SATÉLITES GEOESTACIONARIOS 801
Velocidad orbital de un satélite geosíncrono 803
Tiempo de retardo por viaje redondo para los satélites geosíncronos 803

Órbita de Clarke 803
Ventajas y desventajas de los satélites geosíncronos 803
ÁNGULOS VISUALES DE UNA ANTENA 805
Ángulo de elevación 805
Ángulo de azimut 807
Límites de visibilidad 808
CLASIFICACIONES, ESPACIAMIENTO Y ASIGNACIÓN DE FRECUENCIAS DE SATÉLITES 809
PATRÓN DE RADIACIÓN DE ANTENAS SATELITALES: HUELLAS 812
Haces locales y zonales 813
Haces hemisféricos 814
Haces globales 814
Reuso 814
MODELOS DE ENLACE DE SISTEMAS SATELITALES 815
Modelo de enlace de subida 815
Transpondedor 815
Modelo de enlace de bajada 816
Enlaces cruzados 816
PARÁMETROS DEL SISTEMA DE SATÉLITES 817
Pérdida por reducción 817
Potencia de transmisión y energía de bit 817
Potencia efectiva irradiada isotrópicamente 818
Temperatura equivalente de ruido 819
Densidad de ruido 820
Relación de portadora a densidad de ruido 821
Relación de energía de bit a densidad de ruido 821
Relación de ganancia a temperatura equivalente de ruido 825
ECUACIONES DE ENLACE DEL SISTEMA DE SATÉLITES 826
ECUACIONES DE ENLACE 826
Ecuación de enlace de subida 826
Ecuación de enlace de bajada 827
GASTOS DE ENLACE 827
PREGUNTAS 832
PROBLEMAS 832

CAPÍTULO 19 ARREGLOS DE ACCESO MÚLTIPLE A SATÉLITES 835

INTRODUCCIÓN 835
SISTEMAS SATELITALES FDM/FM 835
Satélite de comunicaciones Anik-E 837
ACCESO MÚLTIPLE 837
Acceso múltiple por división de frecuencia 838
Acceso múltiple por división de tiempo 842
Acceso múltiple por división de código 845
CAPACIDAD DE CANAL 850
Interfaces digitales no interpoladas 850
Interfaz digital interpolada de voz 850
Interpolación de voz por asignación de tiempo 851
RADIONAVEGACIÓN POR SATÉLITE 851
Navegación Loran 852
GPS Navstar 852
Servicios de posicionamiento global (GPS) 853

Segmentos de Navstar 853
GPS diferencial 862
PREGUNTAS 862
PROBLEMAS 863

CAPÍTULO 20 SERVICIO TELEFÓNICO MÓVIL 864

INTRODUCCIÓN 864
EVOLUCIÓN DEL TELÉFONO MÓVIL 865
SERVICIOS DE COMUNICACIONES MÓVILES EN DOS SENTIDOS 866
TELÉFONO CELULAR 866
Conceptos básicos del teléfono celular 867
Reutilización de frecuencia 868
Interferencia 870
División de célula 872
Sectorización 873
Segmentación y dualización 873
Estructura del sistema de células 874
TELÉFONO CELULAR ANALÓGICO 878
Asignación de frecuencias AMPS 878
Clasificación AMPS de teléfonos celulares 880
Canal de control AMPS 880
Señalización de canal de voz 882
Tonos de audio de supervisión y señalización 882
Acceso múltiple por división de frecuencia N-AMPS 884
Teléfono celular ETACS 884
Diagrama de bloques del teléfono celular analógico 884
TELÉFONO CELULAR DIGITAL 885
Canales de control USDC 886
Canal de voz digital USDC 887
Codificación de voz 888
Esquema de modulación digital USDC 891
Canal de control digital USDC 891
SISTEMA TELEFÓNICO CELULAR DIGITAL DE ACCESO MÚLTIPLE POR DIVISIÓN DE CÓDIGO (IS-95) 894
Acceso múltiple por división de código 894
Canales de tráfico CDMA 897
SISTEMA GLOBAL PARA COMUNICACIONES MÓVILES 898
Servicios de GSM 899
Arquitectura del sistema GSM 899
Subsistema de radio GSM 900
SISTEMA SATELITAL DE COMUNICACIONES PERSONALES 900
Mercado proyectado del PCSS 901
Requisitos de la industria del PCSS 901
Ventajas y desventajas del PCSS 901
Proveedores de PCSS 901
Sistema de satélites Iridium 901
PREGUNTAS 905
PROBLEMAS 906

APÉNDICE A EL DIAGRAMA DE SMITH 907

INTRODUCCIÓN 907
DEDUCCIÓN DEL DIAGRAMA DE SMITH 909
Líneas de $R_e(z)$ constante 909
Líneas de $X(z)$ constante 910
Inversión de impedancia (admitancia) 910
Complejo conjugado 911
GRAFICACIÓN DE LA IMPEDANCIA, ADMITANCIA Y SWR EN EL DIAGRAMA DE SMITH 911
IMPEDANCIA DE ENTRADA Y LA GRÁFICA DE SMITH 913
Compensación con transformador de cuarto de onda en la gráfica de Smith 918
Línea de acoplamiento mediante stub, con el diagrama de Smith 922
PROBLEMAS 925

RESPUESTAS A PROBLEMAS SELECCIONADOS 926

ÍNDICE 935

C A P Í T

Imagen protegida por derechos de autor

Introducción a las comunicaciones electrónicas

INTRODUCCIÓN

El objetivo fundamental de un *sistema electrónico de comunicaciones*, es transferir *información* de un lugar a otro. Por consiguiente, se puede decir que las comunicaciones electrónicas son la *transmisión*, *recepción* y *procesamiento* de información entre dos o más lugares, mediante circuitos electrónicos. La fuente original de información puede estar en forma *analógica* (continua), como por ejemplo la voz humana o la música, o en forma *digital* (discreta), como por ejemplo los *números codificados binariamente* o los *códigos alfanuméricos*. Sin embargo, todas las formas de información se deben convertir a *energía electromagnética* antes de ser propagadas a través de un sistema electrónico de comunicaciones.

Samuel Morse desarrolló en 1837 el primer sistema electrónico de comunicaciones. Usó la *inducción electromagnética* para transferir información en forma de puntos, rayas y espacios entre un transmisor y un receptor sencillos, usando una línea de transmisión que consistía en un tramo de conductor metálico. Llamó *telégrafo* a su invento. En 1876, Alexander Graham Bell y Thomas A. Watson fueron los primeros en transferir en forma exitosa la conversación humana a través de un sistema sencillo de comunicaciones con hilo metálico, al que llamaron *teléfono*.

Guglielmo Marconi transmitió por primera vez señales de radio, *sin hilos*, a través de la atmósfera terrestre, en 1894, y Lee DeForest inventó en 1908 el triodo, o válvula al vacío, que permitió contar con el primer método práctico para amplificar las señales eléctricas. La radio comercial comenzó en 1920, cuando las estaciones de radio comenzaron a emitir señales de *amplitud modulada* (AM), y en 1933 el mayor Edwin Howard Armstrong inventó la *modulación de frecuencia* (FM). La emisión comercial en FM comenzó en 1936.

Aunque los conceptos y principios fundamentales de las comunicaciones electrónicas han cambiado poco desde su introducción, los métodos y circuitos con que se realizan han sufrido grandes cambios. En los años recientes, los *transistores* y los *circuitos integrados lineales* han simplificado el diseño de los circuitos de comunicación electrónica, permitiendo así la miniaturización, mejor eficiencia y confiabilidad y costos generales menores. En los años recientes ha habido una necesidad abrumadora de comunicación entre cada vez más personas. Esta urgente

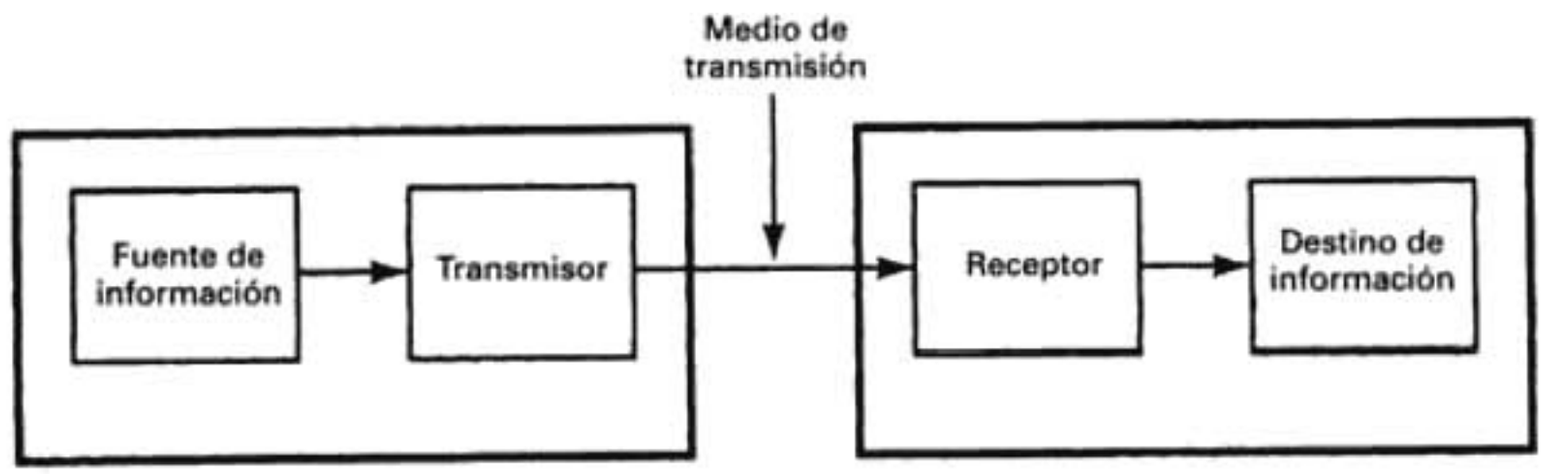


FIGURA 1-1 Diagrama simplificado de bloques de un sistema de comunicaciones electrónicas

necesidad ha estimulado un crecimiento gigantesco de la industria de comunicaciones electrónicas. Los sistemas electrónicos modernos de comunicación incluyen los de *cable metálico*, por *microondas* y los *satelitales*, así como los sistemas de *fibra óptica*.

SISTEMAS ELECTRÓNICOS DE COMUNICACIONES

La fig. 1-1 muestra un diagrama de bloques simplificado de un sistema electrónico de comunicaciones, que comprende un *transmisor*, un *medio de transmisión* y un *receptor*. Un transmisor es un conjunto de uno o más dispositivos o circuitos electrónicos que convierte la información de la fuente original en una señal que se presta más a su transmisión a través de determinado medio de transmisión. El medio de transmisión transporta las señales desde el transmisor hasta el receptor, y puede ser tan sencillo como un par de conductores de cobre que propaguen las señales en forma de flujo de corriente eléctrica. También se puede convertir la información a ondas electromagnéticas luminosas, propagarlas a través de cables de fibra óptica hechas de vidrio o de plástico, o bien se puede usar el espacio libre para transmitir ondas electromagnéticas de radio, a grandes distancias o sobre terreno donde sea difícil o costoso instalar un cable físico. Un receptor es un conjunto de dispositivos y circuitos electrónicos que acepta del medio de transmisión las señales transmitidas y las reconvierte a su forma original.

MODULACIÓN Y DEMODULACIÓN

Como a menudo no es práctico propagar señales de información a través de cables metálicos o de fibra óptica, o a través de la atmósfera terrestre, con frecuencia es necesario modular la información de la fuente, con una señal analógica de mayor frecuencia, llamada *portadora*. En esencia, la señal portadora transporta la información a través del sistema. La señal de información *modula* a la portadora, cambiando su amplitud, su frecuencia o su fase. *Modulación* no es más que el proceso de cambiar una o más propiedades de la portadora, en proporción con la señal de información.

Los dos tipos básicos de comunicaciones electrónicas son *analógico* y *digital*. Un *sistema analógico de comunicaciones* es aquel en el cual la energía se transmite y se recibe en forma analógica: una señal de variación continua, como por ejemplo una onda senoidal. En los sistemas analógicos de comunicaciones, tanto la información como la portadora son señales analógicas.

Sin embargo, el término *comunicaciones digitales* abarca una amplia variedad de técnicas de comunicación, que incluyen *transmisión digital* y *radio digital*. La transmisión digital es un sistema digital verdadero, donde los pulsos digitales (con valores discretos, como +5V y tierra) se transfieren entre dos o más puntos en un sistema de comunicaciones. Con la transmisión digital no hay portadora analógica, y la fuente original de información puede tener forma digital o analógica. Si está en forma analógica se debe convertir a pulsos digitales antes de transmitirla, y se debe reconvertir a la forma analógica en el extremo de recepción. Los sistemas de transmisión digital requieren una instalación física entre el transmisor y el receptor, como por ejemplo un conductor metálico o un cable de fibra óptica.

La *radio digital* es la transmisión de portadoras analógicas moduladas digitalmente, entre dos o más puntos en un sistema de comunicaciones. En la radio digital, la señal moduladora y la señal demodulada son pulsos digitales. Estos pulsos se pueden originar en un sistema digital de transmisión, en una fuente digital, como por ejemplo una computadora, o pueden ser una señal analógica codificada en binario. En los sistemas digitales de radio, el medio de transmisión puede ser una instalación física o el espacio libre (es decir, la atmósfera terrestre). Los sistemas analógicos de comunicaciones fueron los primeros en ser desarrollados; sin embargo, en tiempos recientes, se han popularizado más los sistemas digitales de comunicaciones.

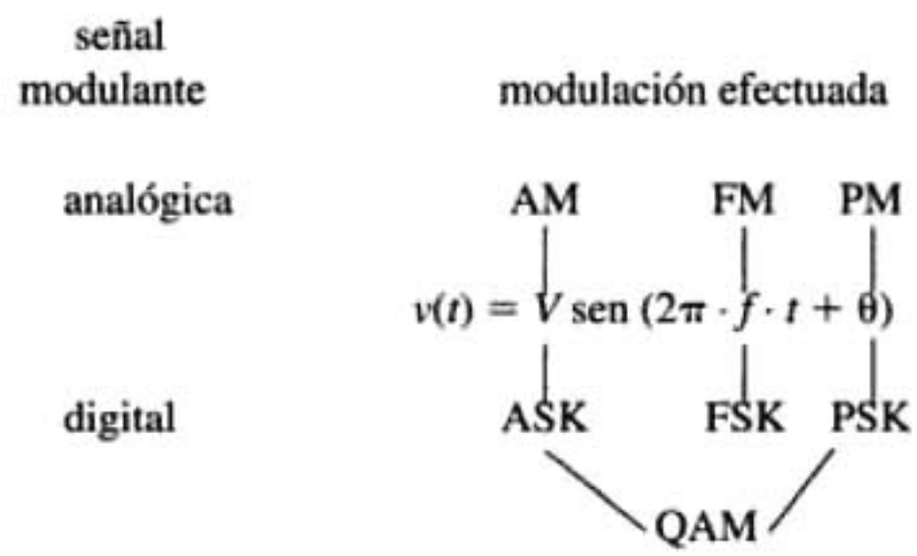
La ecuación 1-1 es la descripción general de una onda senoidal de voltaje, variable en el tiempo, como puede ser una señal portadora de alta frecuencia. Si la señal de información es analógica, y la amplitud (V) de la portadora es proporcional a ella, se produce la *modulación de amplitud* (AM, por *amplitude modulation*). Si se varía la frecuencia (f) en forma proporcional a la señal de información, se produce la *modulación de frecuencia* (FM, de *frequency modulation*); por último, si se varía la fase (θ) en proporción con la señal de información, se produce la *modulación de fase* (PM, de *phase modulation*).

Si la señal de información es digital, y la amplitud (V) de la portadora se varía proporcionalmente a la señal de información, se produce una señal modulada digitalmente, llamada *modulación por conmutación de amplitud* (ASK, de *amplitude shift keying*). Si la frecuencia (f) varía en forma proporcional a la señal de información se produce la *modulación por conmutación de frecuencia* (FSK, de *frequency shift keying*), y si la fase (θ) varía de manera proporcional a la señal de información, se produce la *modulación por conmutación de fase* (PSK, de *phase shift keying*). Si se varían al mismo tiempo la amplitud y la fase en proporción con la señal de información, resulta la *modulación de amplitud en cuadratura* (QAM, de *quadrature amplitude modulation*). Los sistemas ASK, FSK, PSK y QAM son formas de modulación digital, y se describirán con detalle en el capítulo 12.

$$v(t) = V \operatorname{sen}(2\pi ft + \theta) \tag{1-1}$$

donde $v(t)$ = voltaje variable senoidalmente en el tiempo
V = amplitud máxima (volts)
f = frecuencia (hertz)
θ = desplazamiento de fase (radianes)

A continuación se muestra un resumen de las diversas técnicas de modulación



La modulación se hace en un transmisor mediante un circuito llamado *modulador*. Una portadora sobre la que ha actuado una señal de información se llama *onda modulada* o *señal modulada*. La *demodulación* es el proceso inverso a la modulación, y reconvierte a la portadora modulada en la información original (es decir, quita la información de la portadora). La demodulación se hace en un receptor, con un circuito llamado *demodulador*.

Hay dos razones por las que la modulación es necesaria en las comunicaciones electrónicas: 1) Es en extremo difícil irradiar señales de baja frecuencia en forma de energía electromagnética, con una antena, y 2) ocasionalmente, las señales de la información ocupan la misma banda de frecuencias y si se transmiten al mismo tiempo las señales de dos o más fuentes, interferirán entre sí. Por ejemplo, todas las estaciones comerciales de FM emiten señales de voz y

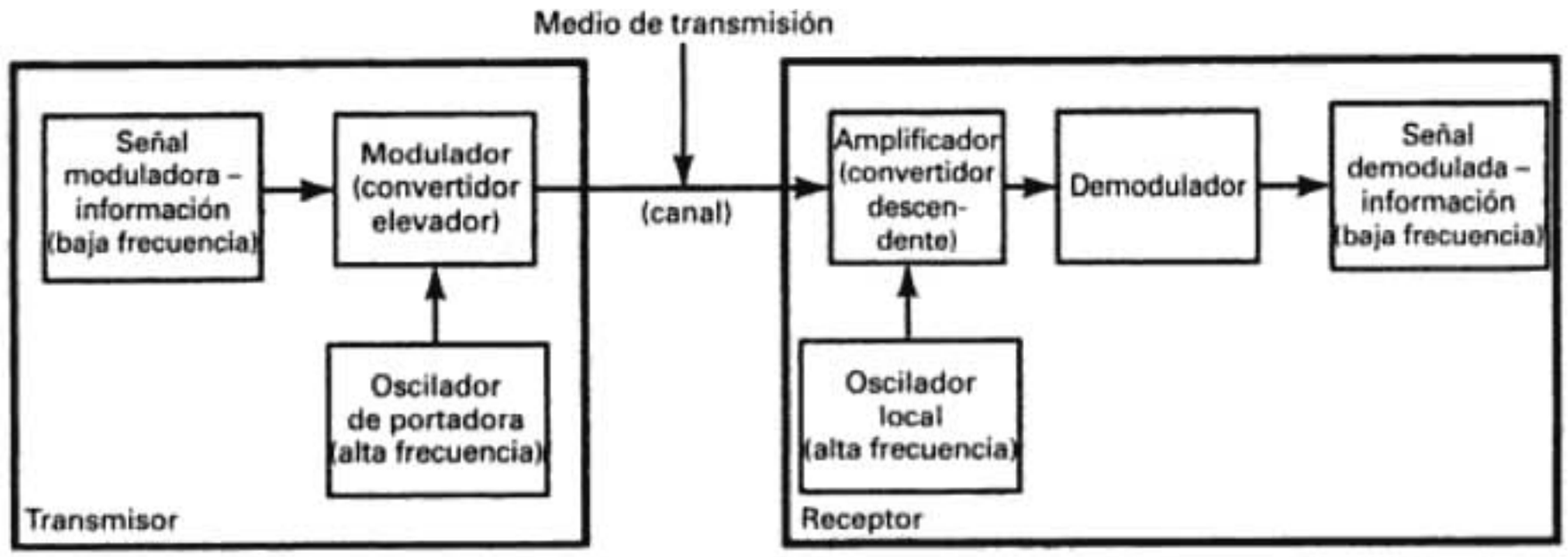


FIGURA 1-2 Diagrama de bloques de un sistema de comunicaciones

música que ocupan la banda de audiofrecuencias, desde unos 300 Hz hasta 15 kHz. Para evitar su interferencia mutua, cada estación convierte a su información a una banda o canal de frecuencia distinto. Se suele usar el término *canal* para indicar determinada banda de frecuencias asignada a determinado servicio. Un canal normal de banda de voz ocupa más o menos 3 kHz de ancho de banda, y se usa para transmitir señales como las de voz; los canales comerciales de emisión en AM ocupan una banda de frecuencias de 10 kHz, y en los canales de radio de microondas y vía satélite se requiere un ancho de banda de 30 MHz o más.

La fig. 1-2 es el diagrama simplificado de bloques de un sistema electrónico de comunicaciones, donde se ven las relaciones entre la señal moduladora, la portadora de alta frecuencia y la onda modulada. La señal de información (que a veces se llama en inglés señal de inteligencia) se combina con la portadora en el modulador, y se produce la onda modulada. La información puede estar en forma analógica o digital, y el modulador puede efectuar modulación analógica o digital. En el transmisor se hace una *conversión elevadora* de las señales de información, de bajas frecuencias a altas frecuencias, y se hace una *conversión descendente* en el receptor, de altas frecuencias a bajas frecuencias. El proceso de convertir una frecuencia, o banda de frecuencias, y pasarla a otro lugar en el espectro total de frecuencias, se llama *translación de frecuencia*. La translación de frecuencia es una parte intrincada de las comunicaciones electrónicas, porque se deben hacer muchas veces las conversiones elevadoras y descendentes cuando se transportan a través del sistema llamado canal. La señal modulada se transporta hasta el receptor a través de un sistema de transmisión. En el receptor se amplifica la señal modulada, se convierte en frecuencia menor y a continuación se demodula, para reproducir la información original de la fuente.

EL ESPECTRO ELECTROMAGNÉTICO

El objetivo de un sistema electrónico de comunicaciones es transferir información entre dos o más lugares, cuyo nombre común es *estaciones*. Esto se logra convirtiendo la información original a energía electromagnética, para transmitirla a continuación a una o más estaciones receptoras, donde se reconvierte a su forma original. La energía electromagnética se puede propagar en forma de voltaje o corriente, a través de un conductor o hilo metálico, o bien en forma de ondas de radio emitidas hacia el espacio libre, o como ondas luminosas a través de una fibra óptica. La energía electromagnética se distribuye en un intervalo casi infinito de frecuencias.

La *frecuencia* no es más que la cantidad de veces que sucede un movimiento periódico, como puede ser una onda senoidal de voltaje o de corriente, durante determinado periodo. Cada inversión completa de la onda se llama *ciclo*. La unidad básica de frecuencia es el hertz (Hz), y un hertz es igual a un ciclo por segundo (1 Hz = 1 cps). En electrónica se acostumbra usar prefijos métricos para representar las grandes frecuencias. Por ejemplo, se usa el kHz (kilohertz) para indicar miles de hertz, y el MHz (megahertz) para indicar millones de hertz.

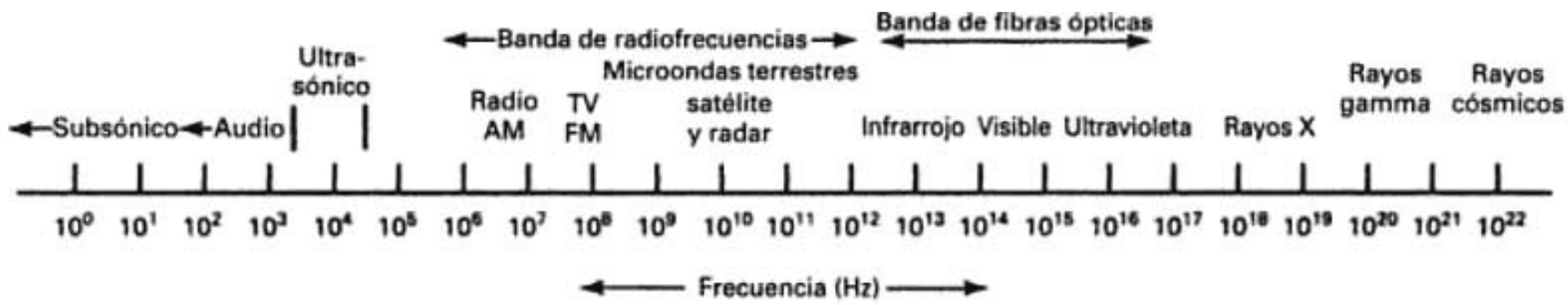


FIGURA 1-3 Espectro electromagnético de frecuencias

Frecuencias de transmisión

El *espectro electromagnético de frecuencias* total, donde se muestran los lugares aproximados de diversos servicios, se ve en la fig. 1-3. Este espectro de frecuencias va desde las *subsónicas* (unos pocos hertz) hasta los *rayos cósmicos* (10^{22} Hz).

El espectro de frecuencias se subdivide en *subsecciones o bandas*. Cada banda tiene un nombre y sus límites. En los Estados Unidos, las asignaciones de frecuencias para *radio propagación en el espacio libre* son realizadas por la *Comisión Federal de Comunicaciones* (FCC). Por ejemplo, la banda de emisión comercial en FM tiene asignadas las frecuencias de 88 MHz a 108 MHz. Las frecuencias exactas asignadas a transmisores específicos que funcionan en las diversas clases de servicio se actualizan y alteran en forma constante, para cumplir con las necesidades de comunicaciones en una nación.

El espectro total útil de *radiofrecuencias* (RF) se divide en bandas de frecuencia más angostas, a las que se dan nombres y números descriptivos, y algunas de ellas se subdividen a su vez en diversos tipos de servicios. Las designaciones de banda según el *Comité consultivo internacional de radio* (CCIR) se muestran en la tabla 1-1. Estas designaciones se resumen como sigue

Frecuencias extremadamente bajas (ELF, de *extremely low frequencies*). Son señales en el intervalo de 30 a 300 Hz, y comprenden las señales de distribución eléctrica (60 Hz) y las de telemetría de baja frecuencia.

Frecuencias de voz (VF, de *voice frequencies*). Son señales en el intervalo de 300 a 3000 Hz, e incluyen a las que generalmente se asocian a la voz humana. Los canales telefónicos normales tienen un ancho de banda de 300 a 3000 Hz, y con frecuencia se llaman *canales de frecuencia de voz*, o *canales de banda de voz*.

Frecuencias muy bajas (VLF, de *very low frequencies*). Son señales dentro de los límites de 3 a 30 kHz, que comprenden al extremo superior del intervalo audible humano. Las VLF se usan en algunos sistemas especiales, del gobierno y militares, como por ejemplo las comunicaciones con submarinos.

Frecuencias bajas (LF, de *low frequencies*). Son señales en el intervalo de 30 a 300 kHz, y se usan principalmente en la navegación marina y aeronáutica.

Frecuencias intermedias (MF, de *medium frequencies*). Son señales de 300 kHz a 3 MHz, y se usan principalmente para emisiones comerciales de radio AM (535 a 1605 kHz).

Frecuencias altas (HF, de *high frequencies*). Señales en el intervalo de 3 a 30 MHz, con frecuencia llamadas *ondas cortas*. La mayoría de las radiocomunicaciones en dos sentidos usa este intervalo, y la Voz de América y la Radio Europa Libre transmiten en él. También los radio aficionados y la banda civil (CB) usan señales de HF.

Muy altas frecuencias (VHF, por *very high frequencies*). Son señales de 30 a 300 MHz, y se usan en radios móviles, comunicaciones marinas y aeronáuticas, emisión comercial en FM (de 88 a 108 MHz) y en la emisión de televisión, en los canales 2 a 13 (54 a 216 MHz).

Frecuencias ultra altas (UHF, de *ultrahigh frequencies*). Son señales entre los límites de 300 MHz a 3 GHz, y las usa la emisión comercial de televisión, en los canales 14 a 83, en

TABLA 1-1 Designaciones de banda CCIR

Número de banda	Intervalo de frecuencias*	Designación
2	30 Hz–300 Hz	ELF (frecuencias extremadamente bajas)
3	0.3 kHz–3 kHz	VF (frecuencias de voz)
4	3 kHz–30 kHz	VLF (frecuencias muy bajas)
5	30 kHz–300 kHz	LF (bajas frecuencias)
6	0.3 MHz–3 MHz	MF (frecuencias intermedias)
7	3 MHz–30 MHz	HF (frecuencias altas)
8	30 MHz–300 MHz	VHF (frecuencias muy altas)
9	300 MHz–3 GHz	UHF (frecuencias ultra altas)
10	3 GHz–30 GHz	SHF (frecuencias super altas)
11	30 GHz–300 GHz	EHF (frecuencias extremadamente altas)
12	0.3 THz–3 THz	Luz infrarroja
13	3 THz–30 THz	Luz infrarroja
14	30 THz–300 THz	Luz infrarroja
15	0.3 PHz–3 PHz	Luz visible
16	3 PHz–30 PHz	Luz ultravioleta
17	30 PHz–300 PHz	Rayos X
18	0.3 EHz–3 EHz	Rayos gamma
19	3 EHz–30 EHz	Rayos cósmicos

*10^0, hertz (Hz); 10^3, kilohertz (kHz); 10^6, megahertz (MHz); 10^9 gigahertz (GHz); 10^{12}, terahertz (THz); 10^{15}, petahertz (PHz); 10^{18} exahertz (EHz)

los servicios móviles de comunicaciones terrestres, teléfonos celulares, algunos sistemas de radar y de navegación, y los sistemas de radio por microondas y por satélite. Hablando con generalidad, se considera que las frecuencias mayores que 1 GHz son de microondas, y eso incluye al extremo superior del intervalo de UHF.

Frecuencias super altas (SHF, por *superhigh frequencies*). Son señales de 3 a 30 GHz, donde está la mayoría de las frecuencias que se usan en sistemas de radiocomunicaciones por microondas y satelitales.

Frecuencias extremadamente altas (EHF, de *extremely high frequencies*). Son señales entre 30 y 300 GHz, y casi no se usan para radiocomunicaciones, a excepción de aplicaciones muy complicadas, costosas y especializadas.

Infrarrojo. Las frecuencias del infrarrojo son señales de 0.3 a 300 THz, y por lo general no se les considera como ondas de radio. Infrarrojo indica una radiación electromagnética que en general se asocia con el calor. Las señales infrarrojas se usan en sistemas de guía de proyectiles con blancos térmicos, o con la fotografía electrónica y la astronomía.

Luz visible. En la luz visible se incluyen las frecuencias electromagnéticas captadas por el ojo humano (0.3 a 3 PHz). Las comunicaciones con ondas luminosas se usan en los sistemas de fibra óptica, que en los últimos años han llegado a ser un medio principal de transmisión en los sistemas electrónicos de comunicaciones.

Rayos ultravioleta, rayos X, rayos gamma y *rayos cósmicos*: tienen poca aplicación en las comunicaciones electrónicas y en consecuencia no se describirán.

Cuando se manejan ondas de radio se acostumbra usar unidades de longitud de onda, y no de frecuencia. La longitud de onda es la distancia que ocupa en el espacio un ciclo de una onda electromagnética, es decir, la distancia entre los puntos correspondientes en una onda repetitiva. La longitud de onda es inversamente proporcional a la frecuencia de la onda, y directamente proporcional a su velocidad de propagación. Se supone que la velocidad

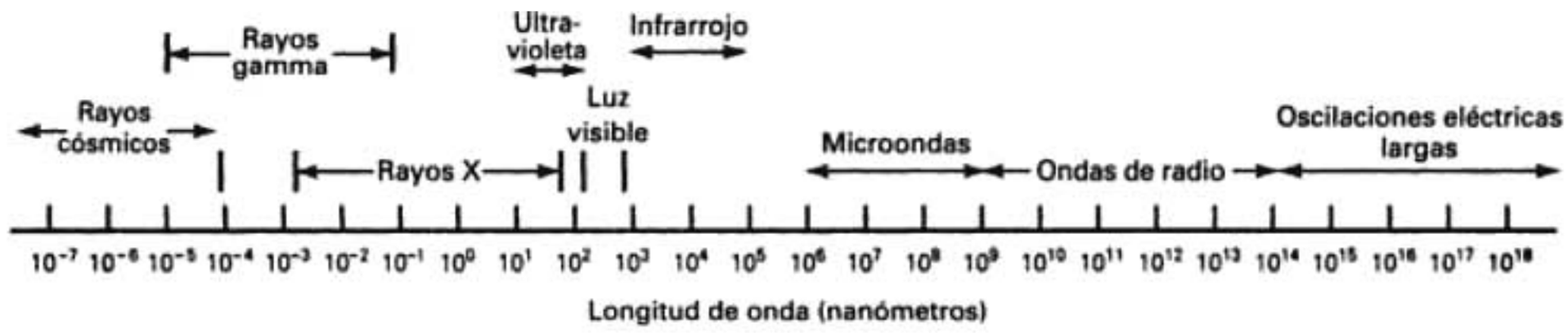


FIGURA 1-4 Espectro electromagnético de longitudes de onda

de propagación de la energía electromagnética en el espacio libre es 3×10^8 m/s. La relación entre frecuencia, velocidad y longitud de onda se expresa en forma matemática como sigue

$$\text{longitud de onda} = \frac{\text{velocidad}}{\text{frecuencia}}$$

$$\lambda = \frac{c}{f} \tag{1-2a}$$

donde λ = longitud de onda (metros por ciclo)
c = velocidad de la luz (300,000,000 metros por segundo)
f = frecuencia (hertz)

En la fig. 1-4 se ve el espectro de *longitudes de ondas electromagnéticas* donde se indican los diversos servicios con sus bandas.

Ejemplo 1-1

Calcular la longitud de onda, en metros, para las siguientes frecuencias: 1 kHz, 100 kHz y 10 MHz.

Solución Al sustituir en la ecuación 1-2a se obtiene

$$\lambda = \frac{300{,}000{,}000}{1000} = 300{,}000 \text{ m}$$

$$\lambda = \frac{300{,}000{,}000}{100{,}000} = 3000 \text{ m}$$

$$\lambda = \frac{300{,}000{,}000}{10{,}000{,}000} = 30 \text{ m}$$

La siguiente ecuación es para calcular la longitud de onda en pulgadas.

$$\lambda = \frac{c}{f} \tag{1-2b}$$

donde λ = longitud de onda (pulgadas por ciclo)
c = velocidad de la luz (11.8×10^9 pulgadas por segundo)
f = frecuencia (hertz)

Clasificación de los transmisores

Para fines de registro en Estados Unidos, los radiotransmisores se clasifican según su ancho de banda, esquema de modulación y tipo de información. Las *clasificaciones de emisión* se identifican con una clave de tres símbolos, que contiene una combinación de letras y números, como se ve en la tabla 1-2. El primer símbolo es una letra que indica el tipo de modulación de la portadora principal. El segundo símbolo es un número que identifica al tipo de emisión, y el tercer símbolo es otra letra que describe el tipo de información que se transmite. Por ejemplo, la designación A3E describe una señal por doble banda lateral, portadora completa, de amplitud modulada, que conduce información telefónica, de voz o de música.

TABLA 1-2 Clasificaciones FCC de emisiones

Símbolo	Letra	Tipo de modulación
Primero	No modulado	
	N	Portadora no modulada
	Modulación de amplitud	
	A	Doble banda lateral, portadora de máxima potencia (DSBFC)
	B	Banda lateral independiente, portadora de máxima potencia (ISBFC)
	C	Banda lateral residual, portadora de máxima potencia (VSB)
	H	Banda lateral única, portadora de máxima potencia (SSBFC)
	J	Banda lateral única, portadora suprimida (SSBSC)
	R	Banda lateral única, portadora reducida (SSBRC)
	Modulación de ángulo	
	F	Modulación de frecuencia (FM directa)
	G	Modulación de fase (FM indirecta)
	D	AM y FM simultáneas o secuenciadas
	Modulación de pulso	
	K	Modulación de amplitud de pulso (PAM)
	L	Modulación de ancho de pulso (PWM)
	M	Modulación de posición de pulso (PPM)
	P	Pulsos no modulados (datos binarios)
	Q	Ángulo modulado durante pulsos
	V	Cualquier combinación de categorías de modulación de pulso
	W	Cualquier combinación de dos o más de las formas anteriores de modulación
	X	Casos no descritos
Segundo	0	Sin señal moduladora
	1	Portadora manipulada digitalmente
	2	Tono manipulado digitalmente
	3	Analógica (sonido o video)
	7	Dos o más canales digitales
	8	Dos o más canales analógicos
	9	Analógica o digital
Tercero	A	Telegrafía manual
	B	Telegrafía automática (teletipo)
	C	Facsímil
	D	Datos, telemetría
	E	Telefonía (emisión de sonido)
	F	Televisión (emisión de video)
	N	Sin información transmitida
	W	Toda combinación de segundas letras

ANCHO DE BANDA Y CAPACIDAD DE INFORMACIÓN

Las dos limitaciones más importantes en el funcionamiento de un sistema de comunicaciones son el *ruido* y el *ancho de banda*. El ruido se describirá más adelante en este capítulo. El ancho de banda de una señal de información no es más que la diferencia entre las frecuencias máxima y mínima contenidas en la información, y el ancho de banda de un canal de comunicaciones es la diferencia entre las frecuencias máxima y mínima que pueden pasar por el canal (es decir, son su *banda de paso*). El ancho de banda de un canal de comunicaciones debe ser suficientemente grande (ancho) para pasar todas las frecuencias importantes de la información. En otras palabras, el ancho de banda del canal de comunicaciones debe ser igual o mayor que el ancho de banda de la información. Por ejemplo, las frecuencias de voz contienen señales de 300 a 3000 Hz. Por consiguiente, un canal para frecuencias de voz debe tener una amplitud igual o mayor que 2700 Hz (3000 Hz − 300 Hz). Si un sistema de transmisión de televisión por cable tiene una banda de paso de 500 a 5000 kHz, su amplitud de banda es 4500 kHz. Como regla general, un canal de comunicaciones no puede propagar una señal que contenga una frecuencia que cambie con mayor rapidez que la amplitud de banda del canal.

La *teoría de la información* es el estudio muy profundo del uso eficiente del ancho de banda para propagar información a través de sistemas electrónicos de comunicaciones. Esta teoría se puede usar para determinar la *capacidad de información* de un sistema de comunicaciones. La capacidad de información es una medida de cuánta información se puede transferir a través de un sistema de comunicaciones en determinado tiempo. La cantidad de información que se puede propagar en un sistema de transmisión es una función del ancho de banda y del tiempo de transmisión. R. Hartley, de los Bell Telephone Laboratories, desarrolló en 1920 la relación entre el ancho de banda, el tiempo de transmisión y la capacidad de información. La ley de Hartley sólo establece que mientras más amplio sea el ancho de banda y mayor sea el tiempo de transmisión, se podrá enviar más información a través del sistema. En forma matemática, la ley de Hartley es

$$I \propto B \times t \tag{1-3}$$

siendo I = capacidad de información
B = ancho de banda del sistema (hertz)
t = tiempo de transmisión (segundos)

La ecuación 1-3 indica que la capacidad de información es una función lineal, y es directamente proporcional tanto al ancho de banda del sistema como al tiempo de transmisión. Si sube al doble el ancho de banda en un sistema de comunicaciones, también se duplica la cantidad de información que puede transportar. Si el tiempo de transmisión aumenta o disminuye, hay un cambio proporcional en la cantidad de información que el sistema puede transferir.

En general, mientras más compleja sea la señal de información, se requiere más amplitud de banda para transportarla en determinado tiempo. Se requieren unos 3 kHz de amplitud de banda para transmitir las señales telefónicas con calidad de voz. En contraste, se asignan 200 kHz de ancho de banda a la transmisión comercial de FM para música, con alta fidelidad, y se requieren casi 6 MHz de ancho de banda para emitir señales de televisión de alta calidad.

C. E. Shannon (también de Bell Telephone Laboratories) publicó en 1948 un trabajo en el *Bell System Technical Journal*, donde relacionó la capacidad de información de un canal de comunicaciones, en bits por segundo (bps), con el ancho de banda y la relación de señal a ruido. La expresión matemática del *límite de Shannon de capacidad de información* es

$$I = B \log_2\left(1 + \frac{S}{N}\right) \tag{1-4a}$$

es decir,

$$I = 3.32\, B \log_{10}\left(1 + \frac{S}{N}\right) \tag{1-4b}$$

donde I = capacidad de información (bits por segundo)
B = ancho de banda (hertz)
$\frac{S}{N}$ = relación de potencia de señal a ruido (sin unidades)

Para un canal normal de comunicaciones en banda de voz, con una relación de potencias de señal a ruido de 1000 (30 dB) y un ancho de banda de 2.7 kHz, el límite de Shannon de capacidad de información es

$$\begin{aligned} I &= 2700 \log_2 (1 + 1000) \\ &= 26.9 \text{ kbps} \end{aligned}$$

Con frecuencia se entiende mal la fórmula de Shannon. Los resultados del ejemplo anterior indican que se pueden transferir 26.9 kbps a través de un canal de 2.7 kHz. Esto podría ser cierto, pero no se puede hacer en un sistema binario. Para alcanzar una rapidez de transmisión de información de 26.9 kbps a través de un canal de 2.7 kHz, cada símbolo que se transfiera debe contener más de un bit de información. Por consiguiente, para llegar al límite de Shannon de capacidad de información, se deben usar sistemas digitales de transmisión que tengan más de dos condiciones (símbolos) de salida.

La ecuación 1-4a se puede reordenar para usarla en la determinación de cuánto ancho de banda se requiere para propagar determinada cantidad de datos por un sistema.

$$B = \frac{I}{\log_2\left(1 + \frac{S}{N}\right)} \tag{1-5}$$

donde B = ancho de banda (hertz)
I = capacidad de información (bits por segundo)
$\frac{S}{N}$ = relación de potencia de señal a ruido (sin unidades)

MODOS DE TRANSMISIÓN

Los sistemas electrónicos de comunicaciones se pueden diseñar para manejar la transmisión sólo en una dirección, en ambas direcciones, sólo en una a la vez, o en ambas direcciones al mismo tiempo. A éstos se les llama *modos de transmisión*. Hay cuatro modos de transmisión posibles: *símplex*, *semidúplex*, *dúplex* y *dúplex/dúplex*.

Símplex (SX)

Con el funcionamiento símplex, las transmisiones sólo se hacen en una dirección. A veces, a los sistemas símplex se les llama *sólo en un sentido*, *sólo recibir* o *sólo transmitir*. Una estación puede ser un transmisor o un receptor, pero no ambos a la vez. Como ejemplo de transmisión símplex está la emisión comercial de radio o televisión: la estación de radio sólo transmite a uno, y uno siempre recibe.

Semidúplex (HDX, de half duplex)

En el funcionamiento semidúplex, las transmisiones se pueden hacer en ambas direcciones, pero no al mismo tiempo. A veces, a los sistemas semidúplex se les llama de *alternar en ambos sentidos*, *en uno de los sentidos*, o *de cambio y fuera*. Una estación puede ser transmisora y receptora, pero no al mismo tiempo. Los sistemas de radio en dos sentidos que usan botones para hablar (PTT, de *push-to-talk*) para conectar sus transmisores, como son los radios de banda civil y de policía, son ejemplos de transmisión en semidúplex.

Dúplex total (FDX, de full duplex)

Con el funcionamiento dúplex total, o simplemente dúplex, puede haber transmisiones en ambas direcciones al mismo tiempo. A veces, a los sistemas dúplex se les llama *simultáneos de dos direcciones*, *dúplex completos* o *líneas bilaterales* o *en ambos sentidos*. Una estación puede transmitir y recibir en forma simultánea; sin embargo, la estación a la que se transmite también debe ser de la que se recibe. Un sistema telefónico normal es un ejemplo de funcionamiento dúplex.

Dúplex total/general (F/FDX, de full/full duplex)

Con la operación en dúplex total/general es posible transmitir y recibir en forma simultánea, pero no necesariamente entre las mismas dos estaciones (es decir, una estación puede transmitir a una segunda estación, y recibir al mismo tiempo de una tercera estación). Las transmisiones dúplex total/general se usan casi exclusivamente en circuitos de comunicaciones de datos. El Servicio Postal en Estados Unidos es un ejemplo de funcionamiento en dúplex total/general.

CONFIGURACIONES DE LOS CIRCUITOS

Los circuitos de comunicaciones electrónicas se pueden configurar en varias formas distintas. A esas configuraciones se les llama *arreglos de circuito*, y pueden abarcar la transmisión a dos y a cuatro hilos.

Transmisión a dos hilos

Como su nombre implica, la *transmisión a dos hilos* usa dos conductores, uno para la señal y otro para una referencia, o tierra, o bien una configuración de circuito que es equivalente a sólo dos conductores. Los circuitos de dos hilos se adaptan en forma ideal a la transmisión símplex, aunque se pueden usar en transmisiones semidúplex y dúplex. La línea telefónica entre el hogar del lector y la central más cercana es un circuito a dos hilos.

La fig. 1-5 muestra los diagramas de bloques para dos configuraciones distintas de circuito a dos hilos. En la fig. 1-5a se ve la más sencilla, que es un circuito pasivo formado por dos conductores que conectan a una fuente de información, a través de un transmisor, con un destinatario en el receptor. Los conductores mismos son capaces de transmitir en dos sentidos, pero el transmisor y el receptor no lo son. Para intercambiar información en dirección opuesta habría que cambiar los lugares del transmisor y el receptor. Por consiguiente, esta configuración sólo es capaz de transmitir en un sentido, y no incorpora ganancia a la señal. Para lograr una transmisión semidúplex con un circuito de dos hilos habría que tener un transmisor y un receptor en cada estación, que deberían estar conectados al mismo par de conductores, en una forma tal que no interfieran entre sí.

La fig. 1-5b muestra un circuito activo a dos hilos (es decir, uno que incorpora ganancia). Con esta configuración, un amplificador se intercala en el circuito, entre el transmisor y el receptor. El amplificador es un dispositivo unidireccional y en consecuencia limita las transmisiones sólo a una dirección.

Para lograr las funciones de semidúplex o dúplex con un circuito de dos hilos, habría que alterar la información que viaja en direcciones opuestas de tal modo, o por algún método que convierta la fuente en destino y el destino en fuente. Se puede lograr la transmisión semidúplex con un circuito de dos hilos usando alguna forma de modulación, para *multiplexar* o combinar las dos señales de tal modo que no se interfieran entre sí, pero que todavía se puedan separar, o reconvertir a su forma original en el receptor. Después, en este libro, se describirán tanto la modulación como el multiplexado.

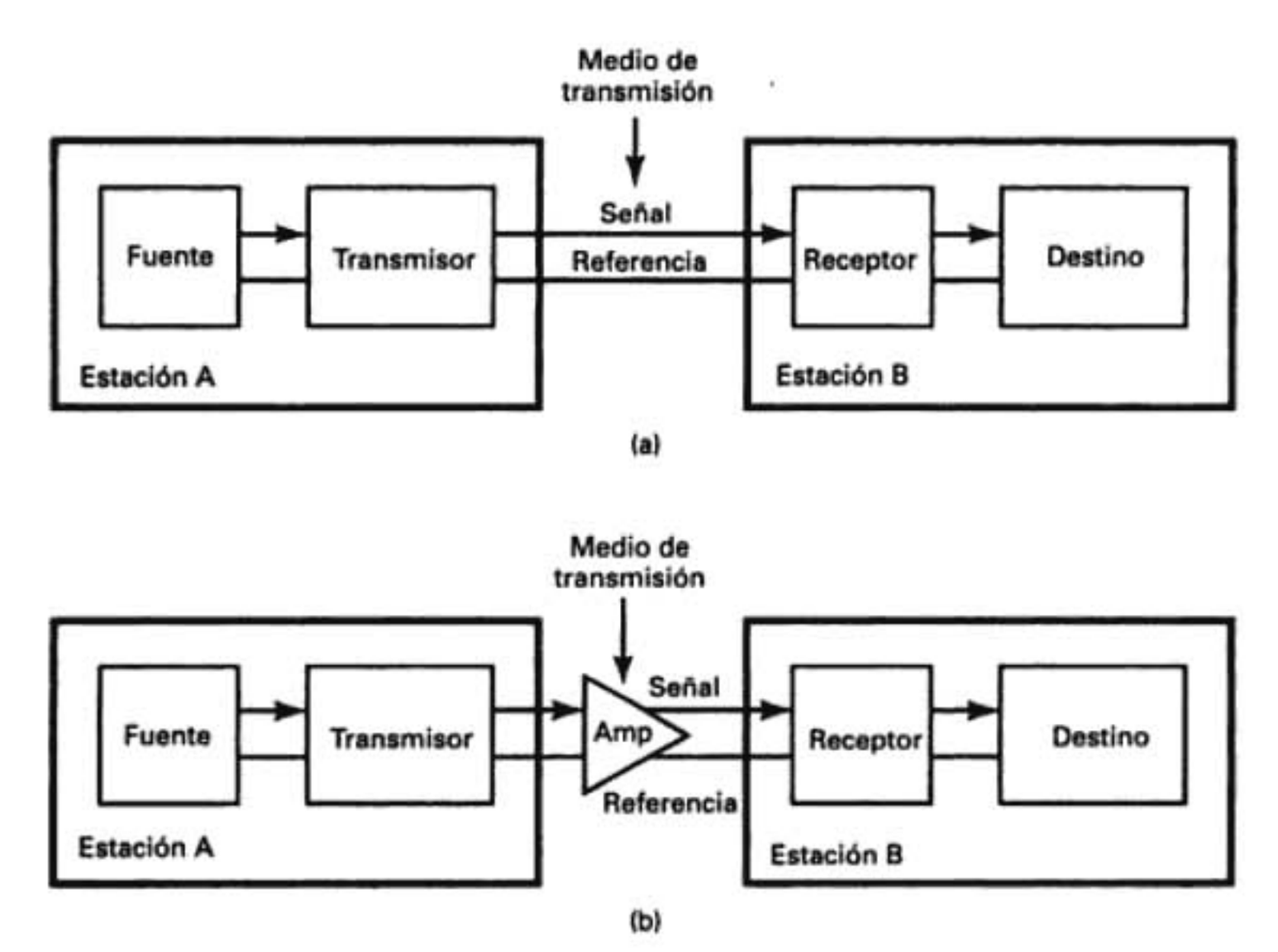


FIGURA 1-5 Configuraciones de circuito de dos hilos: (a) pasivo; (b) activo

Transmisión a cuatro hilos

En la *transmisión a cuatro hilos* se usan cuatro conductores, dos en cada dirección: de señal y de referencia o tierra, o bien una configuración de circuito que equivalga a cuatro conductores. Los circuitos de cuatro hilos se adaptan idealmente a la transmisión dúplex. La fig. 1-6 muestra el diagrama de bloques de un sistema activo de cuatro hilos. Como allí se ve, un circuito de cuatro hilos equivale a dos circuitos de dos hilos, uno para cada dirección de transmisión. Con el funcionamiento con cuatro hilos, el transmisor en un lugar se conecta, a través de un medio de transmisión, con el receptor en la otra estación, y viceversa. Sin embargo, en determinada estación los transmisores y receptores se pueden trabajar en forma totalmente independiente unos de otros.

Existen varias ventajas inherentes de los circuitos con cuatro hilos sobre los de dos hilos. Por ejemplo, los de cuatro hilos son bastante menos ruidosos y proporcionan más aislamiento entre las dos direcciones de transmisión cuando se opera en semidúplex o en dúplex. Sin embargo, los circuitos de dos hilos requieren menos conductor, menos circuitos y por consiguiente, menos inversión, que sus contrapartes de cuatro hilos. Las ventajas y desventajas de los circuitos con dos y cuatro hilos se apreciarán más a medida que el lector continúe su estudio de comunicaciones electrónicas.

Híbridos y supresores de eco

Cuando se conecta un circuito de dos hilos con uno de cuatro hilos, como por ejemplo en las llamadas telefónicas de larga distancia, se usa un circuito de interconexión, o de interfaz, llamado *híbrido* o *equipo terminador*, para efectuar la interfaz. El conjunto híbrido se usa para compensar impedancias y proporcionar aislamiento entre las dos direcciones de flujo de la señal.

La fig. 1-7 muestra el diagrama de bloques para conectar una red híbrida, de dos hilos con cuatro hilos. La bobina híbrida compensa las variaciones de impedancia en la parte del circuito que tiene dos hilos. Los amplificadores y los atenuadores ajustan los voltajes de señal a los valores requeridos, y los igualadores compensan defectos en la línea de transmisión que afectan la respuesta en frecuencia de la señal transmitida, como pueden ser la inductancia, la capacitancia y la resistencia de la línea. Las señales que van de oeste a este (W-E) entran al terminador desde la línea de dos hilos, donde se acoplan inductivamente en la sección transmisora de oeste a este, del circuito a cuatro hilos. Las señales recibidas de la línea se acoplan en la sección receptora de este a oeste del circuito a cuatro hilos, donde se aplican a las tomas centrales de las bobinas híbridas. Si las impedancias de la línea de dos hilos y de la red de balanceo se compensan en forma adecuada, todas las corrientes producidas por la señal E-W en la mitad superior del híbrido, serán de magnitud igual pero de polaridad opuesta. En consecuencia, los voltajes inducidos en los secundarios estarán desfasados 180° entre sí y, por lo tanto, se anularán. Esto evita que cualquier parte de la señal recibida se regrese al transmisor en forma de eco.

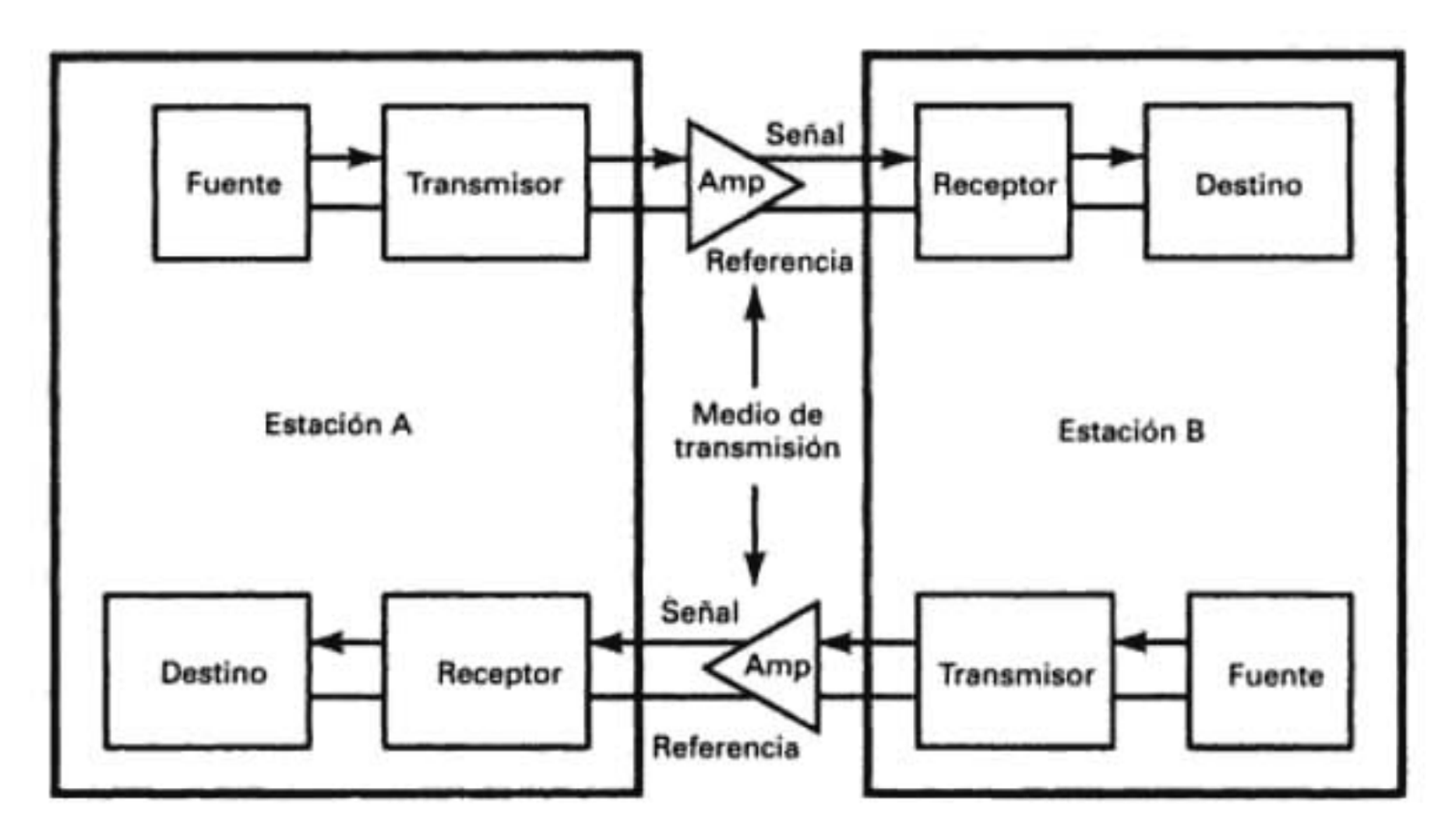


FIGURA 1-6 Circuito activo de cuatro hilos

Si no se compensa la impedancia de la línea de dos hilos con la de la red equilibradora, los voltajes inducidos en los secundarios de la bobina híbrida no se anularán por completo. Este desequilibrio hace que una parte de la señal recibida regrese al emisor, en la parte W-E del circuito de cuatro hilos. Quien habla oye como eco la parte regresada de la señal y, si la demora de esta señal en su ida y vuelta es mayor de unos 45 ms, el eco puede volverse bastante molesto. Para eliminarlo se insertan los llamados *supresores de eco* en un extremo del circuito de cuatro hilos. La fig. 1-8 muestra un diagrama de bloques simplificado de un supresor de eco. El detector de voz siente la presencia y la dirección de la señal. Entonces activa al

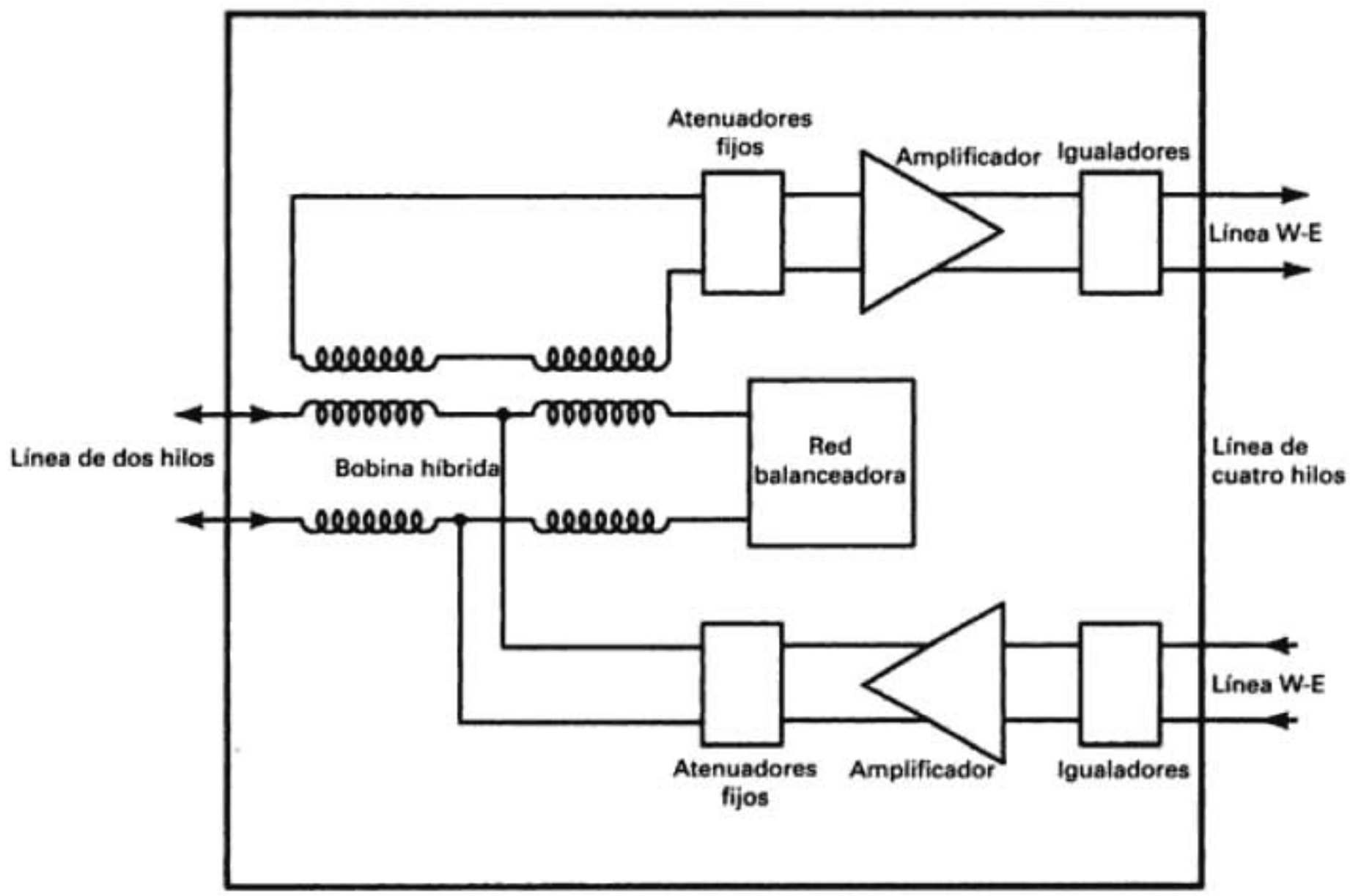


FIGURA 1-7 Terminador de dos a cuatro hilos (híbrido)

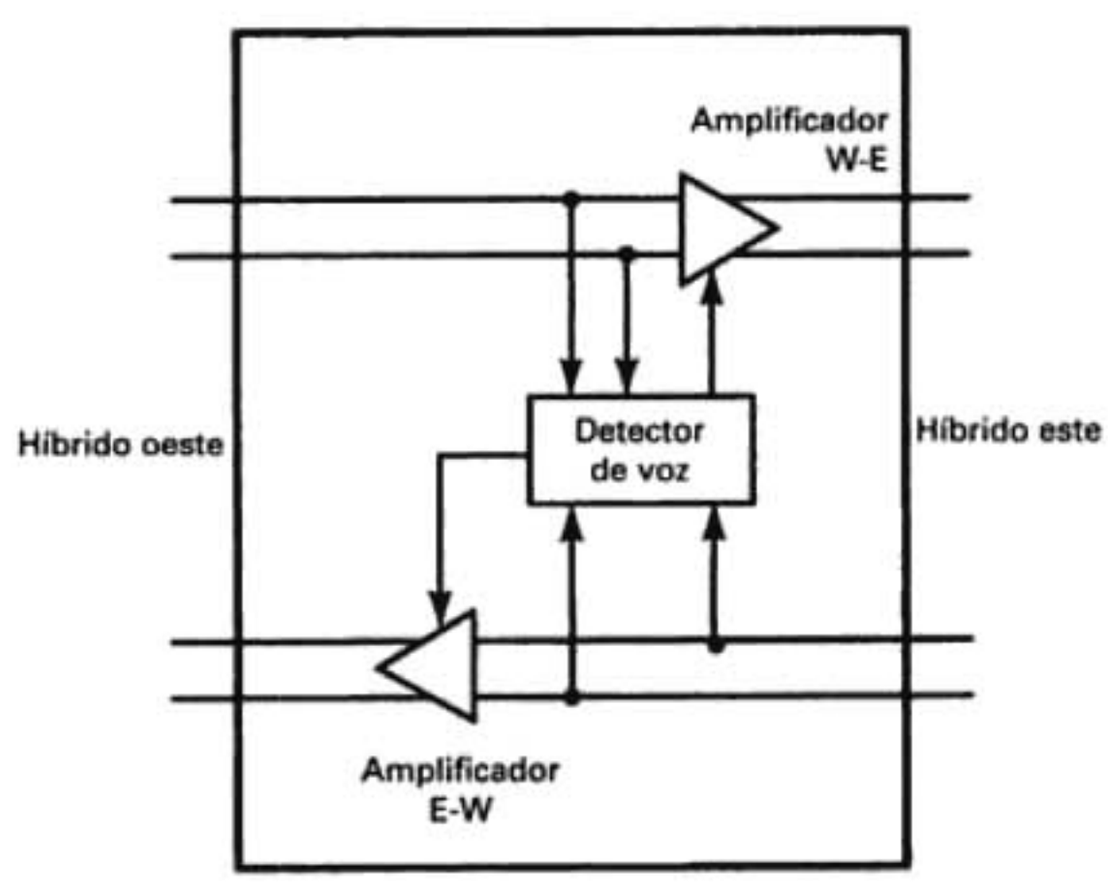


FIGURA 1-8 Supresor de ecos

amplificador en la dirección correcta y desactiva al amplificador en la dirección opuesta, evitando así que el eco regrese a quien habla. Si la conversación cambia rápidamente de dirección, la persona que escucha puede oír cuando el supresor de eco se enciende y se apaga, ya que cada vez que un supresor de eco detecta la voz y se activa, el primer instante de sonido se elimina del mensaje, y se obtiene una voz entrecortada. Con un supresor de eco en el circuito no puede haber transmisiones en ambas direcciones al mismo tiempo y, por consiguiente, la operación del circuito se limita a semidúplex. Las compañías comunes de larga distancia, como AT&T, por lo general ponen supresores de eco en circuitos de cuatro hilos de longitud mayor a 1500 millas eléctricas (ya que mientras más largo es el circuito, la demora por ida y vuelta es mayor).

ANÁLISIS DE SEÑALES

Cuando se diseñan los circuitos electrónicos de comunicaciones, con frecuencia se tiene que analizar y pronosticar el funcionamiento del circuito con base en la distribución de potencia y la composición de frecuencias de la señal de información. Esto se hace con el método matemático llamado *análisis de señales*. Aunque todas las señales en las comunicaciones electrónicas no son ondas senoidales o cosenoidales de una sola frecuencia, muchas de ellas sí lo son, y las que no lo son se pueden representar con una combinación de funciones de seno o de coseno.

Señales senoidales

En esencia, el análisis de señales implica la realización del análisis matemático de frecuencia, longitud de onda y valor de voltaje de una señal. Las señales eléctricas son variaciones de voltaje, o de corriente, respecto al tiempo, que se pueden representar por una serie de ondas seno o coseno. La descripción matemática de una onda de voltaje o de corriente con frecuencia única es

$$v(t) = V\,\text{sen}(2\pi ft + \theta) \quad \text{o} \quad v(t) = V\cos(2\pi ft + \theta)$$
$$i(t) = I\,\text{sen}(2\pi ft + \theta) \quad \text{o} \quad i(t) = I\cos(2\pi ft + \theta)$$

donde $v(t)$ = voltaje de la onda senoidal, variable respecto al tiempo t
$i(t)$ = corriente de la onda senoidal, variable respecto al tiempo t
V = voltaje máximo (volts)
f = frecuencia (hertz)
θ = desplazamiento de fase (radianes)
I = corriente máxima (amperes)
$2\pi f$ = ω velocidad angular (radianes por segundo)

El uso de una función seno o coseno para representar una señal es completamente arbitrario y depende de cuál se escoge como referencia. Sin embargo, se debe observar que sen θ = $\cos(\theta° - 90°)$. Por consiguiente, son válidas las siguientes ecuaciones

$$v(t) = V\,\text{sen}(2\pi ft + \theta) = V\cos(2\pi ft + \theta° - 90°)$$
$$v(t) = V\cos(2\pi ft + \theta) = V\,\text{sen}(2\pi ft + \theta° + 90°)$$

Las fórmulas anteriores son para una onda repetitiva, de una sola frecuencia. A esa forma de onda se le llama *onda periódica*, porque se repite con rapidez uniforme, es decir, cada ciclo sucesivo de la señal tarda exactamente el mismo tiempo y tiene exactamente las mismas variaciones de amplitud que en cualquier otro ciclo; cada ciclo tiene exactamente la misma forma. Una serie de ondas seno, coseno o cuadradas, son ejemplos de ondas periódicas. Las ondas periódicas se pueden analizar en el dominio del *tiempo* o en el *dominio de la frecuencia*. De hecho, con frecuencia se hace necesario pasar del dominio del tiempo al de la frecuencia y viceversa cuando se analiza el funcionamiento de un sistema.

Dominio del tiempo. Un osciloscopio normal es un instrumento de dominio del tiempo. La pantalla del tubo de rayos catódicos es una representación de la amplitud de la señal de entrada en función del tiempo, y se le suele llamar *forma de onda de la señal*. En esencia, una

forma de onda de la señal muestra la forma y la magnitud instantánea de la señal con respecto al tiempo, pero no necesariamente indica el valor de la frecuencia. Con un osciloscopio, la desviación vertical es proporcional a la amplitud de la señal total de entrada, y la deflexión horizontal es una función del tiempo (frecuencia de barrido). La fig. 1.9 muestra la forma de onda de una señal senoidal de frecuencia única de f hertz con amplitud máxima de V volts.

Dominio de la frecuencia. El analizador de espectro es un instrumento de dominio de la frecuencia. En esencia no se despliega ninguna forma de onda en la pantalla del tubo de rayos catódicos. En vez de lo anterior se muestra una gráfica de amplitud contra frecuencia (la cual se conoce como *espectro de frecuencia*). En un analizador de espectro, el eje horizontal representa la frecuencia y el eje vertical representa la amplitud. En consecuencia, existirá una deflexión vertical para cada frecuencia que está presente en la entrada. De hecho, la forma de onda de entrada se "barre" a una frecuencia variable, con la ayuda de un filtro de paso de banda con Q elevado (cuya frecuencia central está sincronizada con la velocidad de barrido horizontal del tubo de rayos catódicos. Cada frecuencia que está presente en la forma de onda entrada produce una línea vertical en la pantalla del tubo de rayos catódicos (éstas son las *componentes espectrales*). La deflexión vertical (altura) de cada línea es proporcional a la amplitud de la frecuencia que representa. Una representación en el dominio de la frecuencia de la onda muestra el contenido de la frecuencia, pero no indica necesariamente la forma de la onda o la amplitud combinada de todas las componentes de entrada en un instante específico de tiempo. La fig. 1-10 muestra el espectro de frecuencia de una señal senoidal de frecuencia única cuya amplitud pico es V volts y la frecuencia es f hertz.

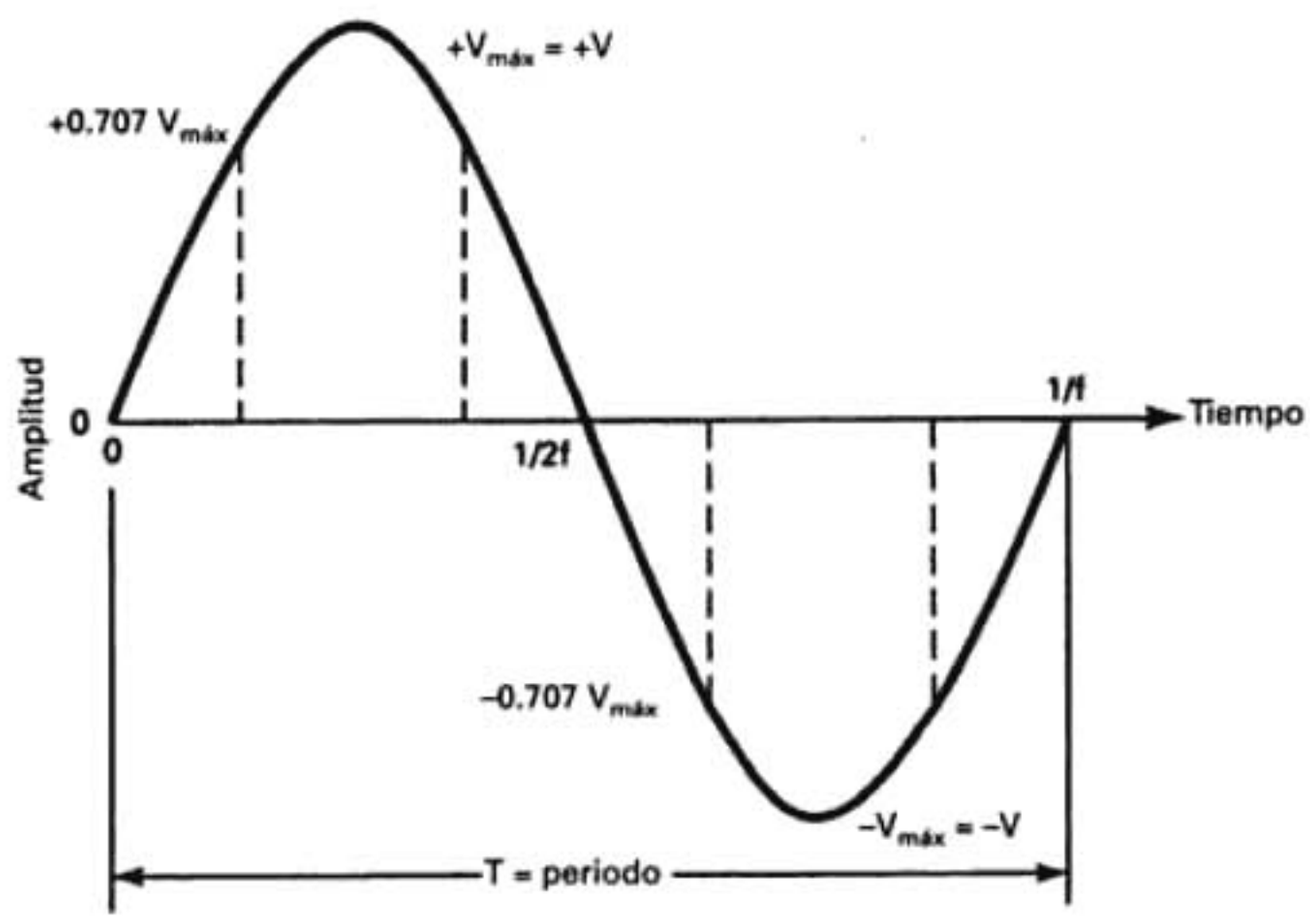


FIGURA 1-9 Representación en el dominio del tiempo (forma de onda de la señal) de una onda senoidal de frecuencia única

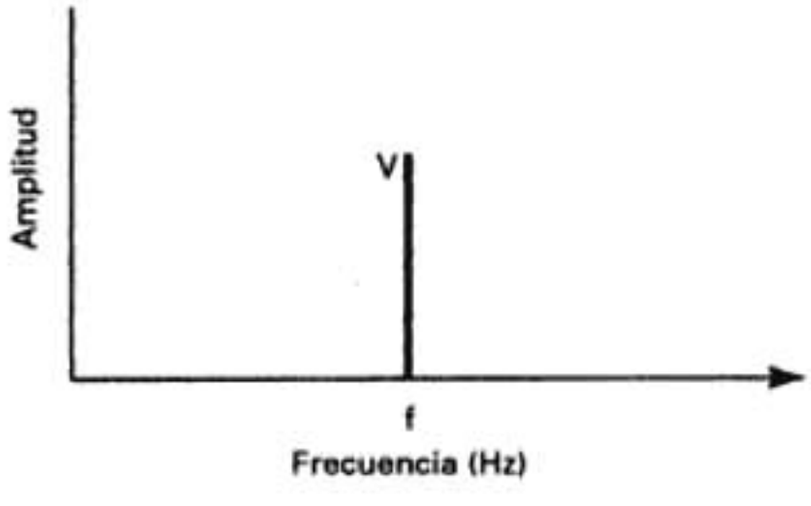


FIGURA 1-10 Representación en el dominio de la frecuencia (espectro) de una onda senoidal de frecuencia única

Ondas periódicas no senoidales (ondas complejas)

En esencia, toda onda repetitiva formada por más de una onda senoidal o cosenoidal relacionada armónicamente, es una *onda no senoidal* o una *onda periódica compleja*. Para analizar una onda periódica compleja es necesario usar una serie matemática inventada en 1826 por el físico y matemático francés, el barón Jean Fourier. Esta serie se llama *serie de Fourier*.

La serie de Fourier. Esta serie se usa en análisis de señales para representar las componentes senoidales de una onda periódica no senoidal, es decir, para cambiar una señal en el dominio del tiempo a una señal en el dominio de la frecuencia. En general, se puede obtener una serie de Fourier para cualquier función periódica, en forma de una serie de funciones trigonométricas con la siguiente forma matemática

$$\begin{aligned} f(t) = A_0 &+ A_1 \cos \alpha + A_2 \cos 2\alpha + A_3 \cos 3\alpha + \cdots + A_n \cos n\alpha \\ &+ B_1 \operatorname{sen} \beta + B_2 \operatorname{sen} 2\beta + B_3 \operatorname{sen} 3\beta + \cdots + B_n \operatorname{sen} n\beta \end{aligned} \tag{1-6}$$

donde $\alpha = \beta$

La ecuación 1-6 indica que la forma de onda $f(t)$ comprende un valor promedio (A_0) de cd, una serie de funciones cosenoidales en las que cada término sucesivo tiene una frecuencia que es múltiplo entero de la frecuencia del primer término cosenoidal de la serie, y una serie de funciones senoidales en la que cada término sucesivo tiene una frecuencia que es múltiplo entero de la del primer término senoidal de la serie. No hay restricciones para los valores o los valores relativos de las amplitudes de los términos seno y coseno. La ecuación 1-6 se enuncia como sigue en palabras: Cualquier *forma de onda periódica* está formada por un componente promedio y una serie de ondas senoidales y cosenoidales relacionadas armónicamente. Una *armónica* es un múltiplo entero de la frecuencia fundamental. La *frecuencia fundamental* es la *primera armónica*, y es igual a la frecuencia (*rapidez de repetición*) de la forma de onda. El segundo múltiplo de la fundamental se llama *segunda armónica*, el tercer múltiplo es la *tercera armónica*, y así sucesivamente. La frecuencia fundamental es la mínima necesaria para representar a una forma de onda. Por consiguiente, la ecuación 1-6 se puede escribir como sigue

$$f(t) = \text{dc} + \text{fundamental} + \text{2da. armónica} + \text{3ra. armónica} + \cdots + n\text{-ésima armónica}$$

Simetría de onda. Dicho en términos sencillos, la *simetría de la onda* describe la simetría de una forma de onda en el dominio del tiempo, esto es, su posición relativa con respecto a los ejes horizontal (tiempo) y vertical (amplitud).

Simetría par. Si una forma de onda periódica de voltaje es simétrica respecto al eje vertical (amplitud) se dice que tiene *simetría especular*, o de *ejes*, y se llama *función par*. Para todas las funciones pares, los coeficientes B de la ecuación 1-6 son cero. Por consiguiente, la señal sólo contiene un componente de cd y los términos cosenoidales (nótese que la misma onda cosenoide es una función par). La suma de una serie de funciones pares es una función par. Las funciones pares satisfacen la condición

$$f(t) = f(-t) \tag{1-7}$$

De acuerdo con la ecuación 1-7, la magnitud y la polaridad de la función en $+t$ es igual a la magnitud y la polaridad en $-t$. En la fig. 1-11a se ve una forma de onda que sólo contiene funciones pares.

Simetría impar. Si una forma periódica de onda de voltaje es simétrica respecto a una línea intermedia entre el eje vertical y el horizontal negativo (es decir, a los ejes en el segundo y cuarto cuadrantes) y pasa por el origen de las coordenadas, se dice que tiene una simetría *puntual* o que es *antisimétrica*, y se le llama *función impar*. Para todas las funciones impares, los coeficientes A de la ecuación 1-6 son cero. Por consiguiente, la señal tan sólo contiene un componente de cd y los términos senoidales (nótese que la misma onda seno es una función impar). La

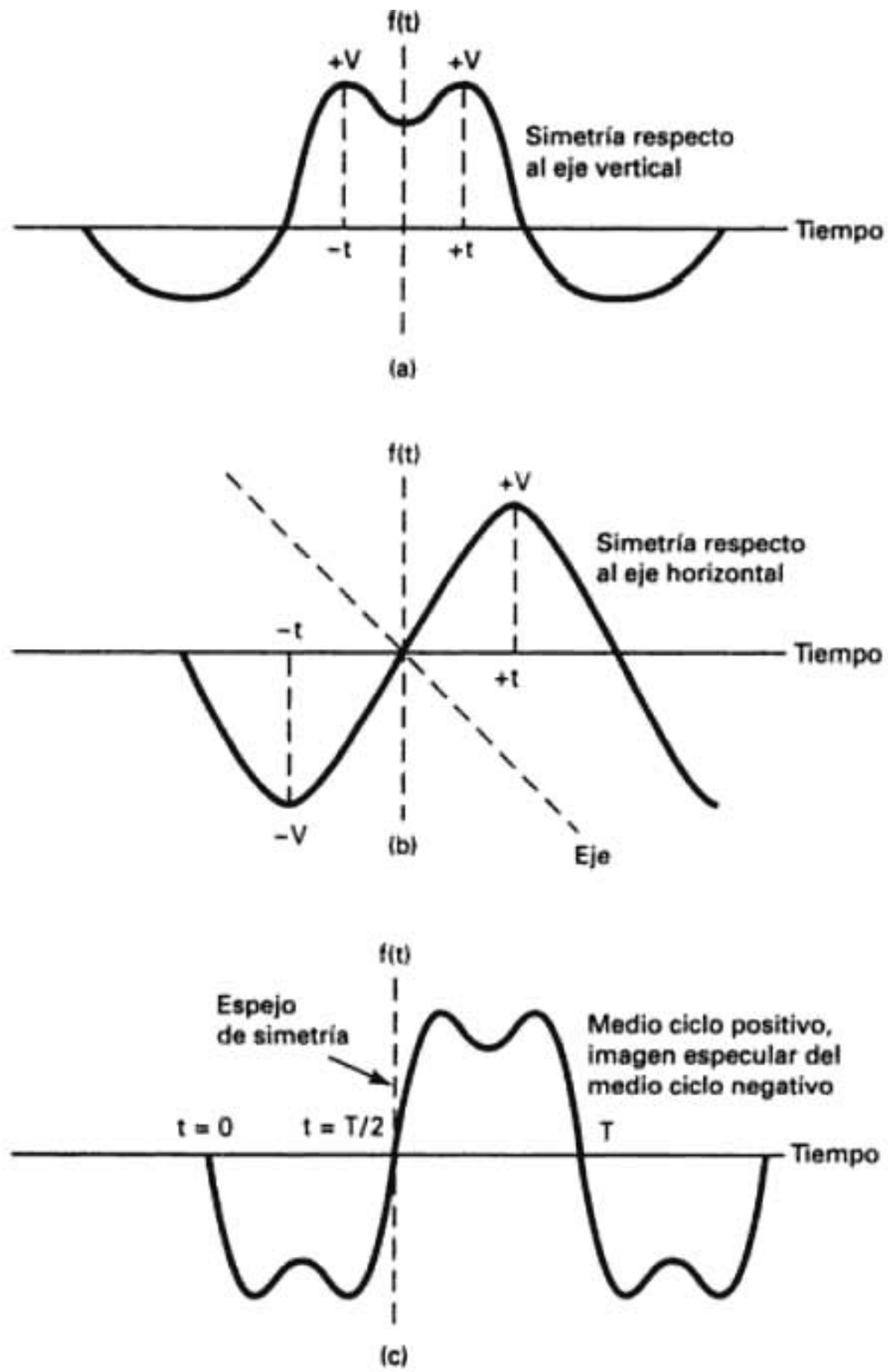


FIGURA 1-11 Simetrías de onda: (a) simetría par; (b) simetría impar; (c) simetría de media onda

suma de una serie de funciones impares es una función impar. A esta forma primero se le debe reflejar en el eje Y y después en el eje X para sobreponerla consigo misma. Así,

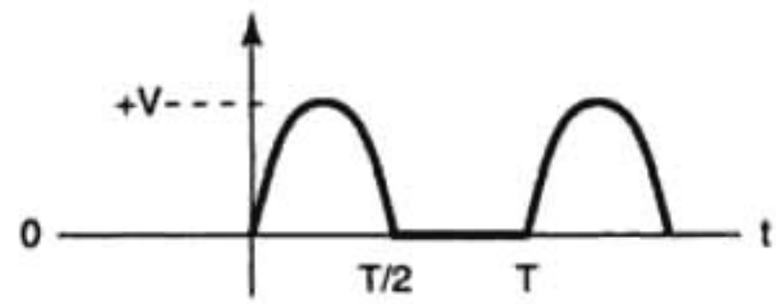
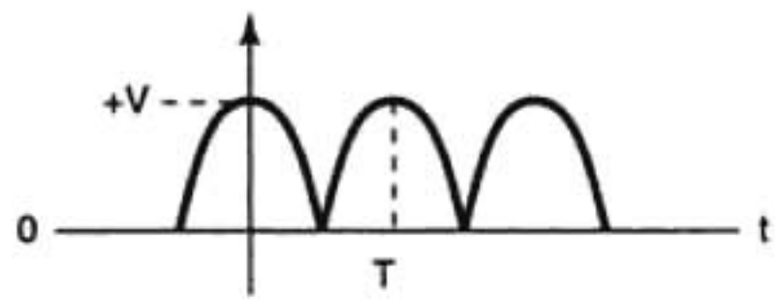
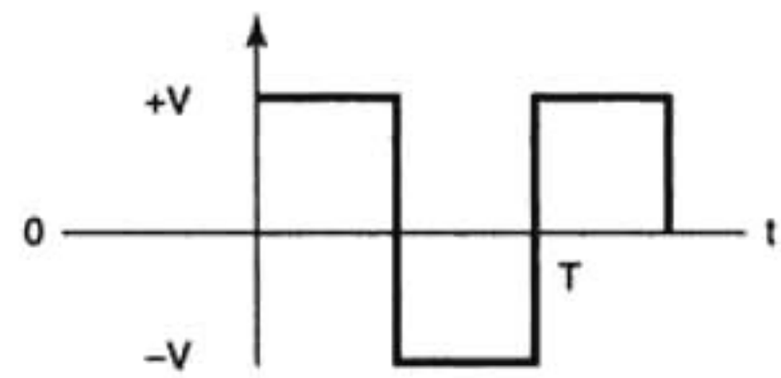
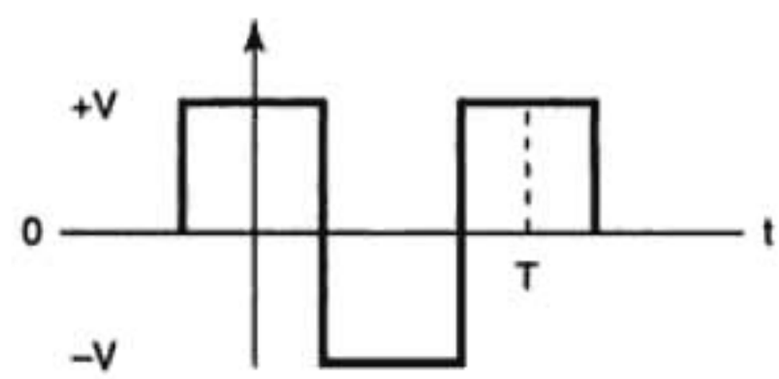
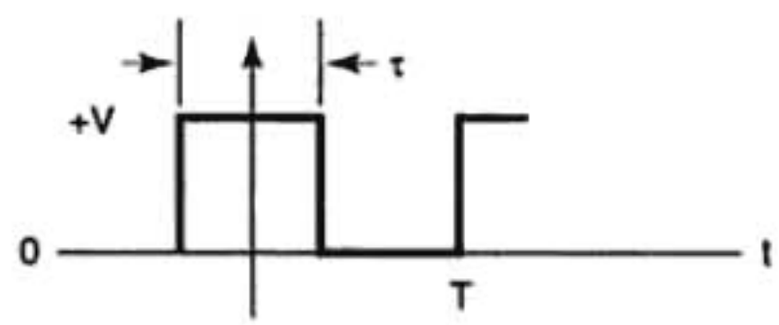
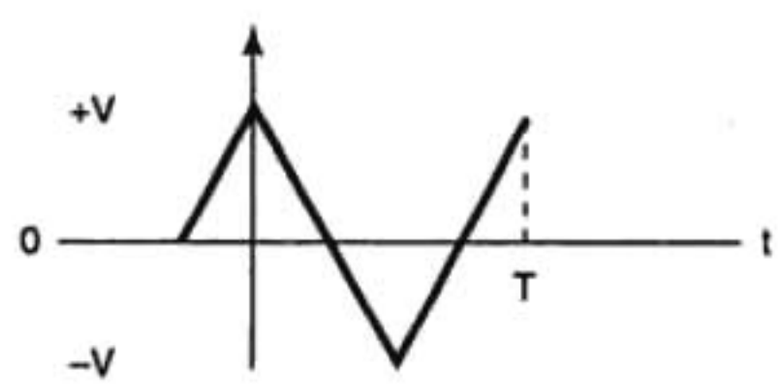
$$f(t) = -f(-t) \tag{1-8}$$

La ecuación 1-8 establece que la magnitud de la función en $+t$ es igual al negativo de la magnitud en $-t$, es decir, que las magnitudes en esos puntos son iguales, pero los signos son opuestos. En la fig. 1-11b se ve una forma de onda periódica que sólo contiene funciones impares.

Simetría de media onda. Si una forma de onda periódica de voltaje es tal que la onda del primer medio ciclo ($t = 0$ a $t = T/2$) se repite, pero con signo contrario, durante el segundo medio ciclo ($t = T/2$ a $t = T$), se dice que tiene *simetría de media onda*. Para todas las formas de onda con simetría de media onda, las armónicas pares de la serie, en los términos en seno y en coseno, son cero. Por consiguiente, las funciones de media onda cumplen con la condición

$$f(t) = -f\left(\frac{T}{2} + t\right) \tag{1-9}$$

TABLA 1-3 Resumen de series de Fourier

Forma de onda	Serie de Fourier
Impar 	$v(t) = \frac{V}{\pi} + \frac{V}{2}\text{sen}\,\omega t - \frac{2V}{3\pi}\cos 2\omega t - \frac{2V}{15\pi}\cos 4\omega t + \cdots$ $v(t) = \frac{V}{\pi} + \frac{V}{2}\text{sen}\,\omega t + \sum_{N=2}^{\infty} \frac{V[1+(-1)^N]}{\pi(1-N^2)}\cos N\omega t$
Par 	$v(t) = \frac{2V}{\pi} + \frac{4V}{3\pi}\cos \omega t - \frac{4V}{15\pi}\cos 2\omega t + \cdots$ $v(t) = \frac{2V}{\pi} + \sum_{N=1}^{\infty} \frac{4V(-1)^N}{\pi[1-(2N)^2]}\cos N\omega t$
Impar 	$v(t) = \frac{4V}{\pi}\text{sen}\,\omega t + \frac{4V}{3\pi}\text{sen}\,3\omega t + \cdots$ $v(t) = \sum_{N=\text{impar}}^{\infty} \frac{4V}{N\pi}\text{sen}\,N\omega t$
Par 	$v(t) = \frac{4V}{\pi}\cos \omega t - \frac{4V}{3\pi}\cos 3\omega t + \frac{4V}{5\pi}\cos 5\omega t + \cdots$ $v(t) = \sum_{N=\text{impar}}^{\infty} \frac{V\,\text{sen}\,N\pi/2}{N\pi/2}\cos N\omega t$
Par 	$v(t) = \frac{V\tau}{T} + \sum_{N=1}^{\infty}\left(\frac{2V\tau}{T}\frac{\text{sen}\,N\omega t/T}{N\pi t/T}\right)\cos N\pi t$
Par 	$v(t) = \frac{8V}{\pi^2}\cos \omega t + \frac{8V}{(3\pi)^2}\cos 3\omega t + \frac{8V}{(5\pi)^2}\cos 5\omega t + \cdots$ $v(t) = \sum_{N=\text{impar}}^{\infty} \frac{8V}{(N\pi)^2}\cos N\omega t$

En la fig. 1-11c se ve una forma de onda periódica con simetría de media onda. Se debe hacer notar que una forma de onda puede tener simetría de media onda y también puede ser impar o par, al mismo tiempo. Los coeficientes A_0, B_1 a B_n, y A_1 a A_n se pueden evaluar con las siguientes fórmulas integrales

$$A_0 = \frac{1}{T}\int_0^T f(t)\,dt \tag{1-10}$$

$$A_n = \frac{2}{T}\int_0^T f(t)\cos n\omega t\,dt \tag{1-11}$$

$$B_n = \frac{2}{T}\int_0^T f(t)\,\text{sen}\, n\omega t\,dt \tag{1-12}$$

Para resolver las ecuaciones 1-10, 1-11 y 1-12 se requiere aplicar el cálculo integral, que sale del alcance de este libro. Por lo anterior, en las descripciones que siguen se presentarán las soluciones correspondientes.

La tabla 1-3 es un resumen de las series de Fourier para algunas de las formas de onda periódicas no senoidales.

Ejemplo 1-2

Para el tren de ondas cuadradas de la fig. 1-12:

(a) Determinar las amplitudes máximas y las frecuencias de las primeras cinco armónicas impares.
(b) Trazar el espectro de frecuencias.
(c) Calcular el voltaje instantáneo total, para varios tiempos, y trazar la forma de onda en el dominio del tiempo.

Solución (a) Al inspeccionar la forma de onda de la fig. 1-12 se ve que el componente promedio de cd es 0 V, y que la forma de onda tiene al mismo tiempo simetría impar y de media onda. Si se evalúan las ecuaciones 1-10, 1-11 y 1-12 se obtiene la siguiente serie de Fourier para onda cuadrada con simetría impar.

$$v(t) = V_0 + \frac{4V}{\pi}\left[\text{sen}\,\omega t + \frac{1}{3}\,\text{sen}\,3\omega t + \frac{1}{5}\,\text{sen}\,5\omega t + \frac{1}{7}\,\text{sen}\,7\omega t + \frac{1}{9}\,\text{sen}\,9\omega t + \ldots\right] \tag{1-13a}$$

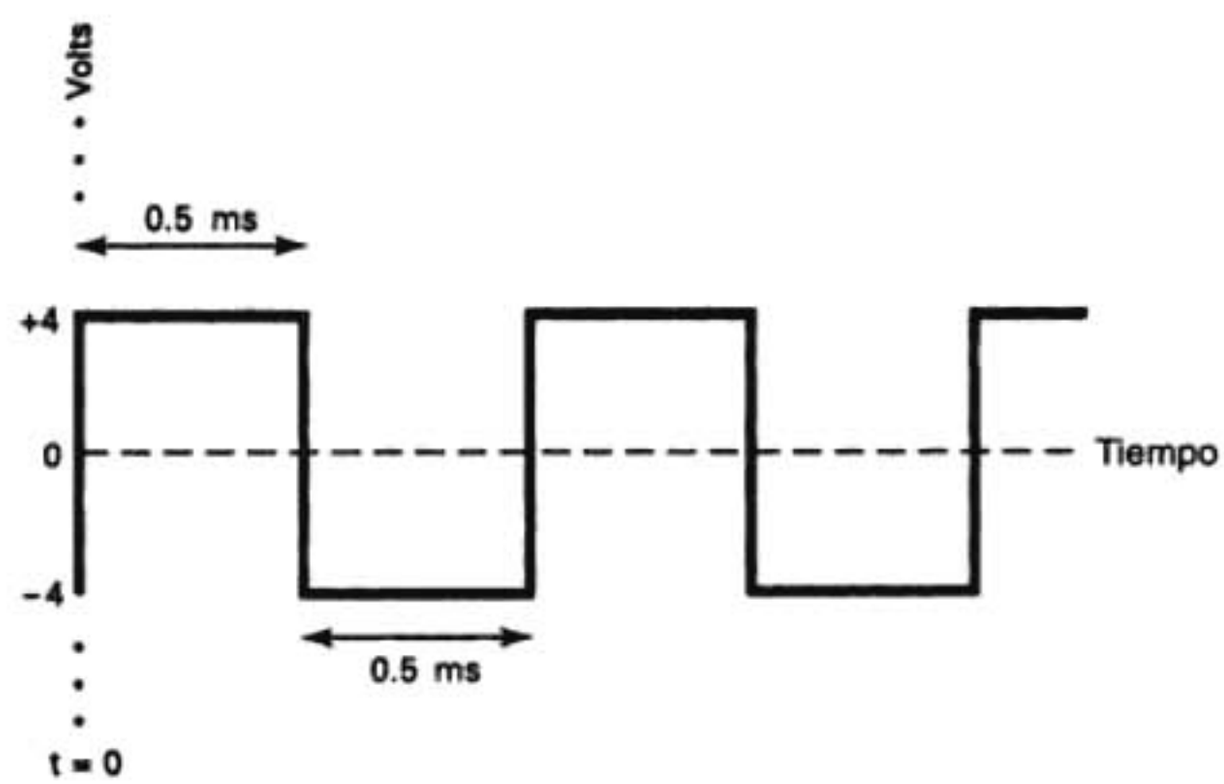


FIGURA 1-12 Forma de onda para el ejemplo 1-2

en donde $v(t)$ = voltaje variable en el tiempo
V_0 = voltaje promedio de cd (volts)
V = amplitud máxima de la onda cuadrada (volts)
ω = $2\pi f$ (radianes por segundo)
T = periodo de la onda cuadrada (segundos)
f = frecuencia fundamental de la onda cuadrada ($1/T$) (hertz)

La frecuencia fundamental de la onda cuadrada es

$$f = \frac{1}{T} = \frac{1}{1\text{ ms}} = 1\text{ kHz}$$

Se puede ver en la ecuación 1-13a que la frecuencia y la amplitud de la n-ésima armónica impar se determinan con las siguientes ecuaciones

$$f_n = n \times f \tag{1-13b}$$

$$V_n = \frac{4V}{n\pi} \qquad n = \text{entero positivo impar} \tag{1-13c}$$

siendo n = n-ésima armónica (sólo armónicas impares para una onda cuadrada)
f = frecuencia fundamental de la onda cuadrada (hertz)
V_n = amplitud máxima de la n-ésima armónica (volts)
f_n = frecuencia de la n-ésima armónica (hertz)
V = amplitud máxima de la onda cuadrada (volts)

Si se sustituye $n = 1$ en las ecuaciones 1-13b y 1-13c, el resultado es

$$V_1 = \frac{4(4)}{\pi} = 5.09\text{ V}_\text{p} \qquad f_1 = 1 \times 1000 = 1000\text{ Hz}$$

Al sustituir $n = 3, 5, 7$ y 9 en las ecuaciones 1-13b y 1-13c se obtiene

n	Armónica	Frecuencia (Hz)	Voltaje máximo (V_p)
1	Primera	1000	5.09
3	Tercera	3000	1.69
5	Quinta	5000	1.02
7	Séptima	7000	0.73
9	Novena	9000	0.57

(b) El espectro de frecuencias se ve en la fig. 1-13.
(c) Se sustituyen los resultados de los pasos anteriores en la ecuación 1-13a y se llega a

$$v(t) = 5.09\text{ sen}[2\pi 1000t] + 1.69\text{ sen}[2\pi 3000t] + 1.02\text{ sen}[2\pi 5000t] + 0.73\text{ sen}[2\pi 7000t] + 0.57\text{ sen}[2\pi 9000t]$$

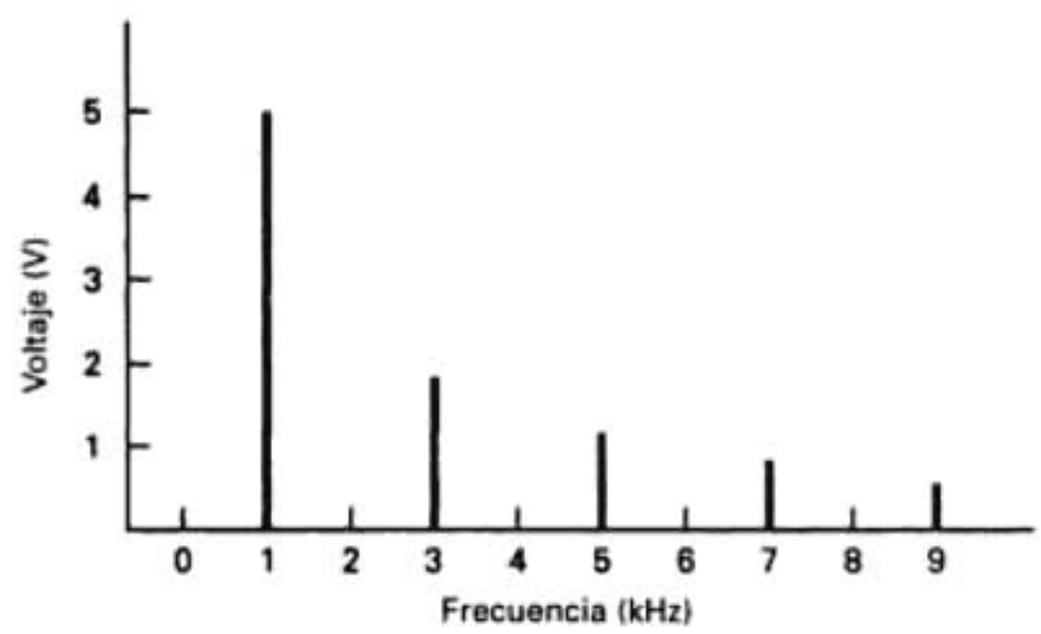


FIGURA 1-13 Espectro de frecuencias para el ejemplo 1-2

Se despeja $v(t)$ para $t = 62.5\ \mu s$ se obtiene

$$v(t) = 5.09\ \text{sen}[2\pi 1000(62.5\ \mu s)] + 1.69\ \text{sen}[2\pi 3000(62.5\ \mu s)] + 1.02\ \text{sen}[2\pi 5000(62.5\ \mu s)] + 0.73\ \text{sen}[2\pi 7000(62.5\ \mu s)] + 0.57\ \text{sen}[2\pi 9000(62.5\ \mu s)]$$

$$v(t) = 4.51\ \text{V}$$

De igual modo se despeja $v(t)$ para varios otros valores del tiempo, y se obtiene la siguiente tabla

Tiempo (μs)	$v(t)$ (volts máximo)
0	0
62.5	4.51
125	3.96
250	4.26
375	3.96
437.5	4.51
500	0
562.5	−4.51
625	−3.96
750	−4.26
875	−3.96
937.5	−4.51
1000	0

La señal en el dominio del tiempo se determina graficando los tiempos y los voltajes calculados arriba, en papel de gráficas. El resultado se ve en la fig. 1-14. Aunque la forma de onda que se ve no es exactamente una onda cuadrada, sí se parece mucho a ella. Para lograr una forma de onda más exacta en el dominio del tiempo sería necesario despejar $v(t)$ con más valores del tiempo que los que se indican en el diagrama.

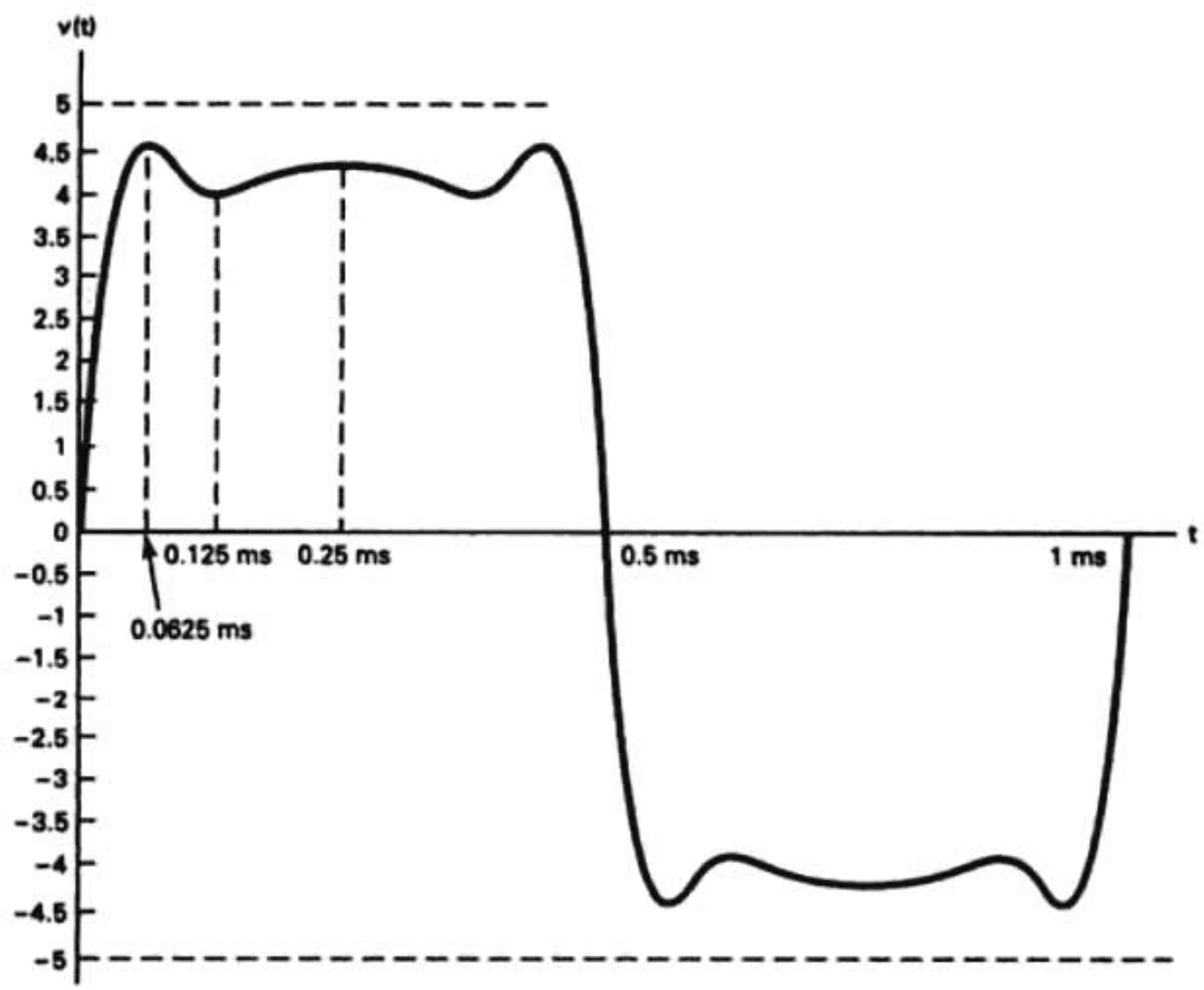


FIGURA 1-14 Señal del ejemplo 1-2 en el dominio del tiempo

Serie de Fourier para una forma de onda rectangular

Cuando se analizan los circuitos de comunicaciones electrónicas se hace necesario, con frecuencia, usar *pulsos rectangulares*. En la fig. 1-15 se ve una forma de onda que representa una serie de pulsos rectangulares. El *ciclo de trabajo* (DC, de *duty cycle*) en la onda es la relación del tiempo activo del pulso entre el periodo de la onda. En forma matemática, el ciclo de trabajo es

$$\mathrm{DC} = \frac{\tau}{T} \tag{1-14a}$$

$$\mathrm{DC}(\%) = \frac{\tau}{T} \times 100 \tag{1-14b}$$

en donde DC = ciclo de trabajo en decimales
DC(%) = ciclo de trabajo en porcentaje
τ = ancho del pulso de la onda rectangular (segundos)
T = periodo de la onda rectangular (segundos)

Independientemente de su ciclo de trabajo, una forma de onda rectangular está constituida por una serie de ondas senoidales relacionadas armónicamente. Sin embargo, la amplitud de los componentes espectrales depende del ciclo de trabajo. La serie de Fourier para una forma de onda rectangular de voltaje con simetría par es

$$v(t) = \frac{V\tau}{T} + \frac{2V\tau}{T}\left[\frac{\operatorname{sen} x}{x}(\cos \omega t) + \frac{\operatorname{sen} 2x}{2x}(\cos 2\omega t) + \cdots + \frac{\operatorname{sen} nx}{nx}(\cos n\omega t)\right] \tag{1-15}$$

donde $v(t)$ = voltaje de la onda, variable en el tiempo
τ = ancho del pulso de la onda rectangular (segundos)
T = periodo de la onda rectangular (segundos)
$x = \pi(\tau/T)$
n = n-ésima armónica, y puede ser cualquier valor entero positivo
V = amplitud máxima del pulso (volts)

Se ve en la ecuación 1-15 que una onda rectangular tiene un componente de 0 Hz (cd) igual a

$$V_0 = V \times \frac{\tau}{T} \quad \text{o} \quad V \times \mathrm{DC} \tag{1-16}$$

donde V_0 = voltaje de cd (volts)
DC = ciclo de trabajo en decimales
τ = ancho del pulso de la onda rectangular (segundos)
T = periodo de la onda rectangular (segundos)

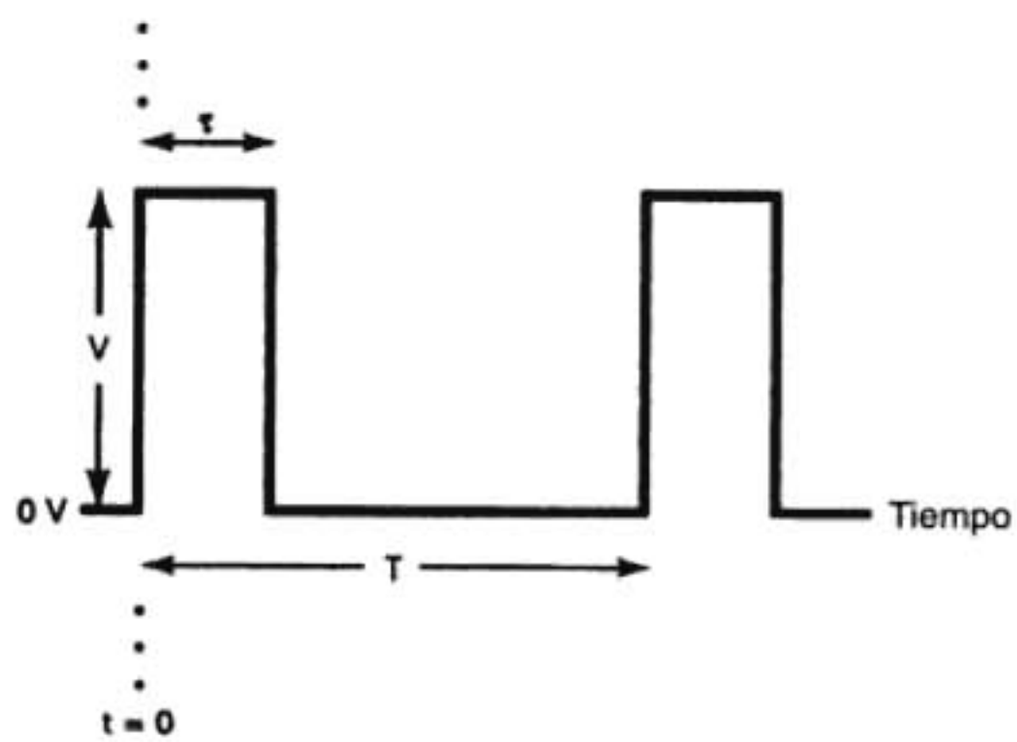


FIGURA 1-15 Forma de onda de un pulso rectangular

Mientras menor sea el ancho del pulso, la componente de cd será menor. También, según la ecuación 1-15, la amplitud de la n-ésima armónica es

$$V_n = \frac{2V\tau}{T} \times \frac{\text{sen } nx}{nx} \tag{1-17a}$$

o sea

$$V_n = \frac{2V\tau}{T} \times \frac{\text{sen}[(n\pi\tau)/T]}{(n\pi\tau)/T} \tag{1-17b}$$

donde V_n = amplitud máxima de la n-ésima armónica (volts)
n = n-ésima armónica (cualquier entero positivo)
π = 3.14159 radianes
V = amplitud máxima de la onda rectangular (volts)
τ = ancho de pulso de la onda rectangular (segundos)
T = periodo de la onda rectangular (segundos)

La función (sen x)/x se usa para describir formas de onda de pulsos repetitivos. El sen x es sólo una onda senoidal, cuya amplitud instantánea depende de x, y varía en sentido positivo entre sus amplitudes máximas, con una rapidez senoidal, cuando aumenta x. Si sólo hay x en el denominador, éste aumenta al aumentar x. Por consiguiente, una función (sen x)/x no es más que una onda senoidal amortiguada, en la que cada pico sucesivo es menor que el anterior. En la fig. 1-16 se ve una gráfica de la función (sen x)/x.

La fig. 1-17 muestra el espectro de frecuencias para un pulso rectangular, con relación de ancho de pulso a periodo igual a 0.1. Se puede ver que las amplitudes de las armónicas son como ondas senoidales amortiguadas. A la frecuencia cuyo periodo es igual a $1/\tau$ (es decir, a la frecuencia $10f$ hertz), hay una componente de 0 V. Un segundo cero está en $20f$ hertz (periodo = $2/\tau$), un tercero en $30f$ hertz (periodo = $3/\tau$), y así sucesivamente. Todas las componentes del espectro entre 0 Hz y la primera frecuencia cero se consideran en el primer lóbulo del espectro de frecuencias, y son positivas. Todas las componentes espectrales entre el primero y segundo cero de frecuencias están en el segundo lóbulo, y son negativas; las componentes entre el segundo y el tercer cero están todas en el tercer lóbulo y son positivas, etcétera.

Las siguientes características se cumplen en todas las formas de onda rectangulares repetitivas:

1. La componente de cd es igual a la amplitud del pulso por el ciclo de trabajo.
2. Hay componentes de 0 V en la frecuencia $1/\tau$ hertz y en todos los múltiplos enteros de esa frecuencia que cumplan con $T = n\tau$, siendo n = cualquier entero impar.
3. La envolvente de la frecuencia en función del tiempo, de las componentes del espectro, tiene la forma de una onda senoidal amortiguada en la que todas las componentes espectrales en lóbulos de número impar son positivas, y todas las componentes espectrales en los lóbulos pares son negativas.

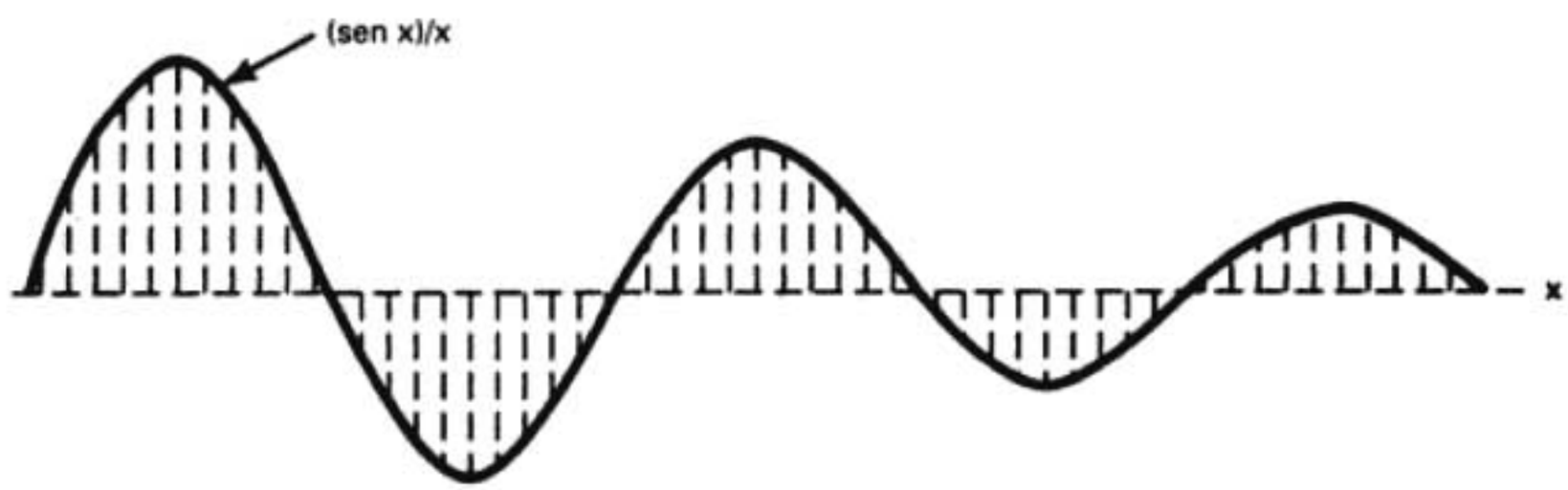


FIGURA 1-16 Función (sen x)/x

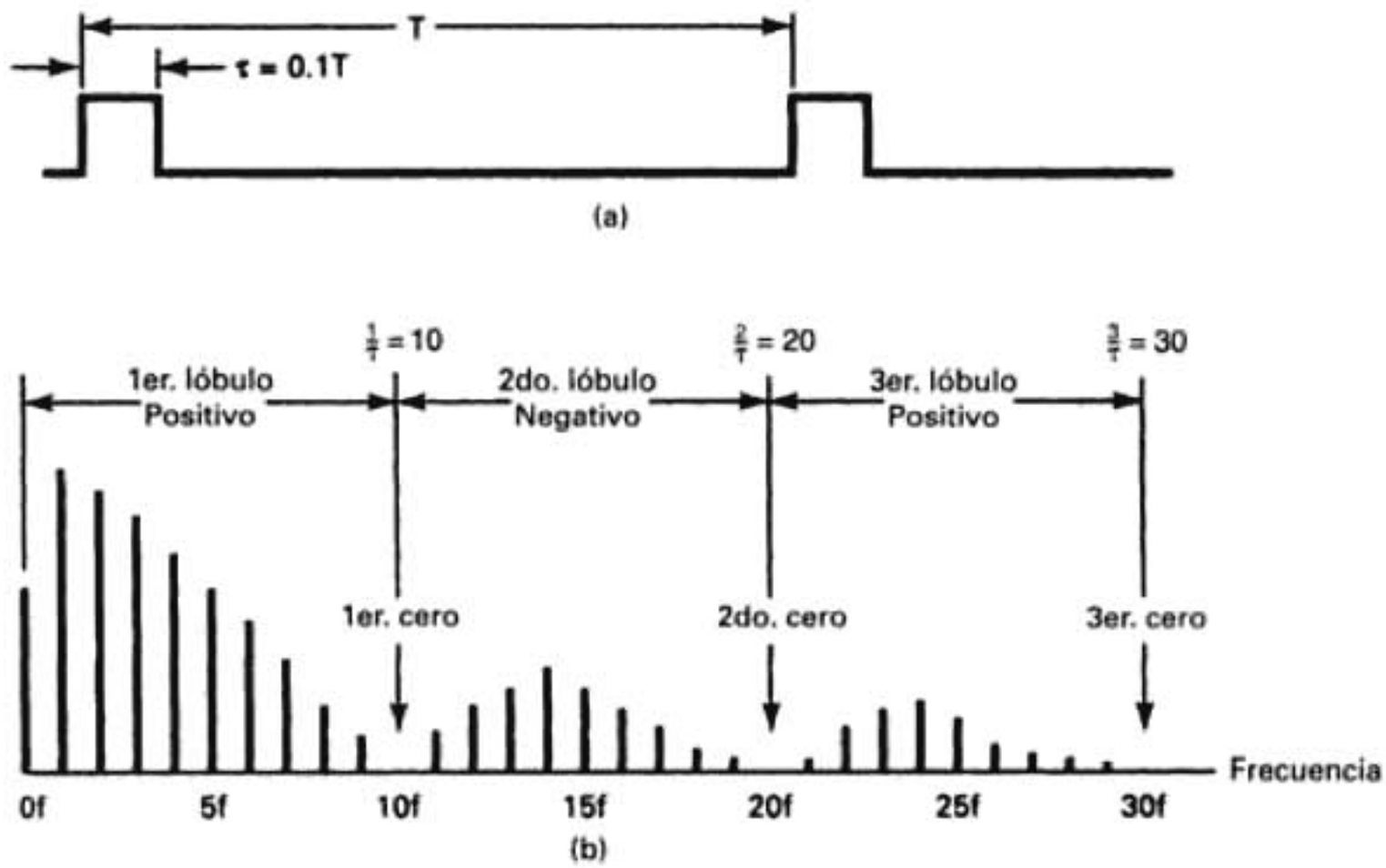


FIGURA 1-17 Función (sen x)/x: (a) forma de onda de un pulso rectangular; (b) espectro de frecuencias

Ejemplo 1-3

Para la forma de onda de pulsos que se ve en la fig. 1-18:

a) Determine la componente de cd.

b) Determine las amplitudes máximas de las 10 primeras armónicas.

c) Grafique la función (sen x)/x.

d) Trace el espectro de frecuencias.

Solución (a) De acuerdo con la ecuación 1-16, la componente de cd es

$$V_0 = \frac{1(0.4\text{ ms})}{2\text{ ms}} = 0.2\ V$$

(b) Las amplitudes máximas de las 10 primeras armónicas se calculan sustituyendo los valores de τ, T, V y n en la ecuación 1-17b, como sigue

$$V_n = 2(1)\left(\frac{0.4\text{ ms}}{2\text{ ms}}\right)\left\{\frac{\text{sen}[(n\pi)(0.4\text{ ms}/2\text{ ms})]}{(n\pi)(0.4\text{ ms}/2\text{ ms})}\right\}$$

n	Frecuencia (Hz)	Amplitud (volts)
0	0	0.2 V dc
1	500	0.374 V_p
2	1000	0.303 V_p
3	1500	0.202 V_p
4	2000	0.094 V_p
5	2500	0.0 V
6	3000	−0.063 V_p
7	3500	−0.087 V_p
8	4000	−0.076 V_p
9	4500	−0.042 V_p
10	5000	0.0 V

(c) En la fig. 1-19 se ve la gráfica de la función (sen x)/x.

(d) El espectro de frecuencias se ve en la fig. 1-20.

Aunque los componentes de frecuencia en los lóbulos pares son negativos, se acostumbra graficar todos los voltajes en la dirección positiva del espectro de frecuencias.

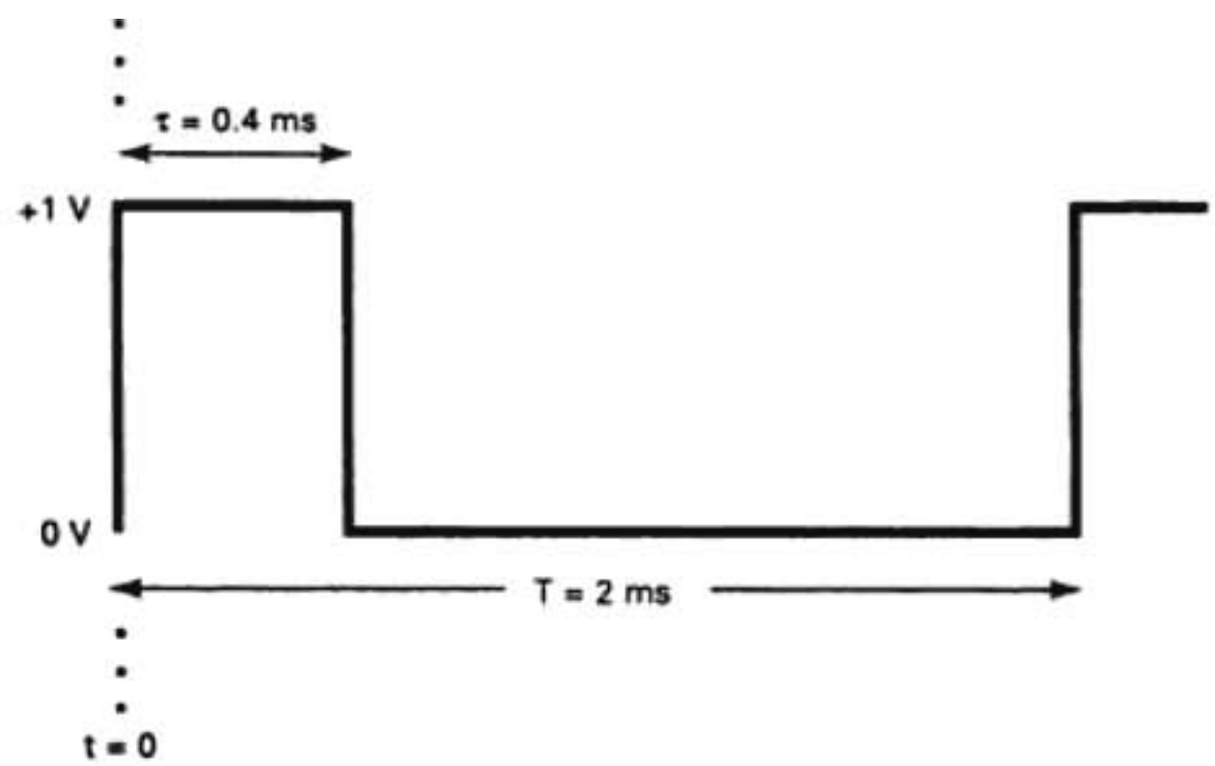


FIGURA 1-18 Forma de onda del pulso del ejemplo 1-3

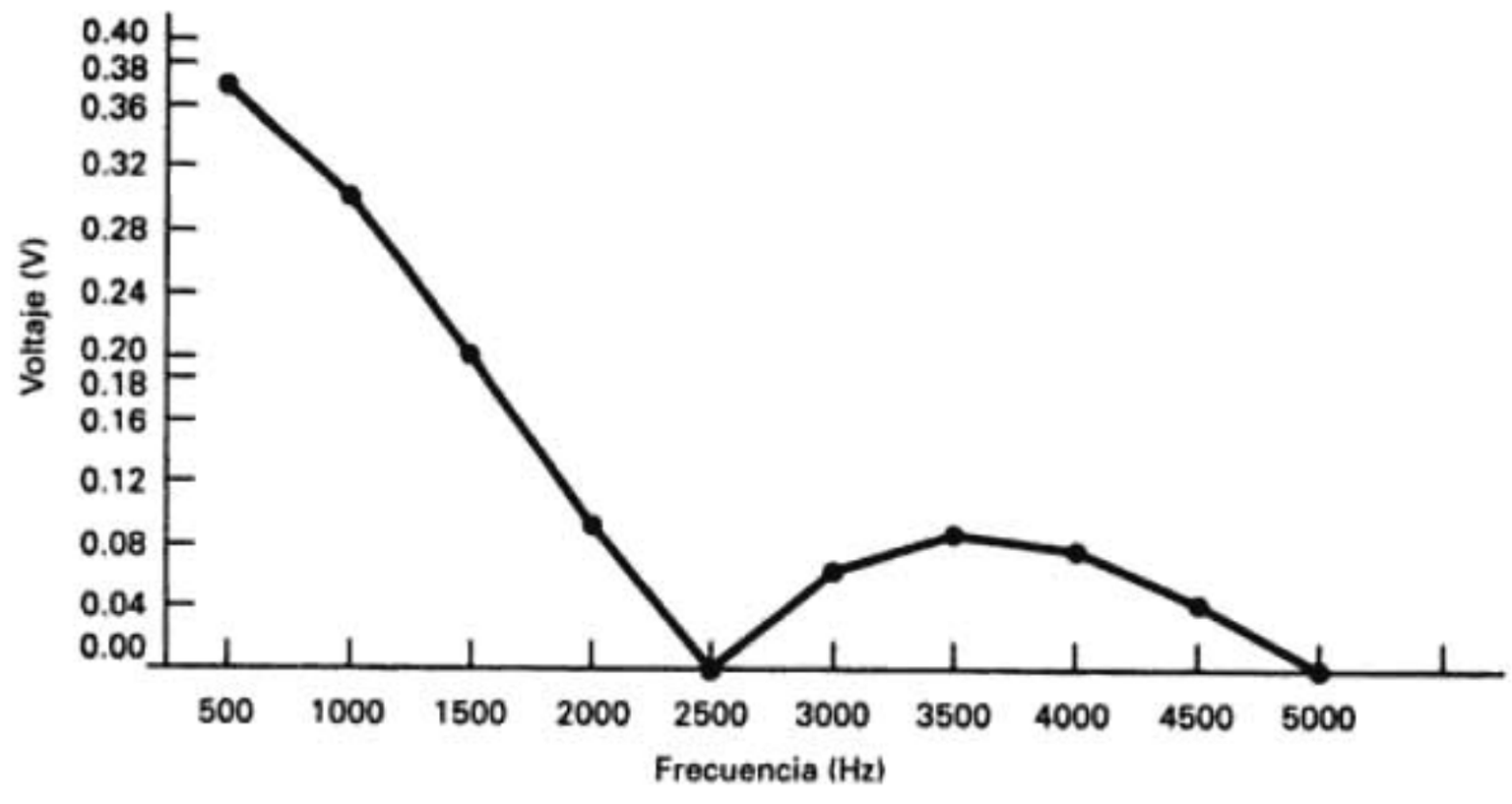


FIGURA 1-19 Función (sen x)/x para el ejemplo 1-3

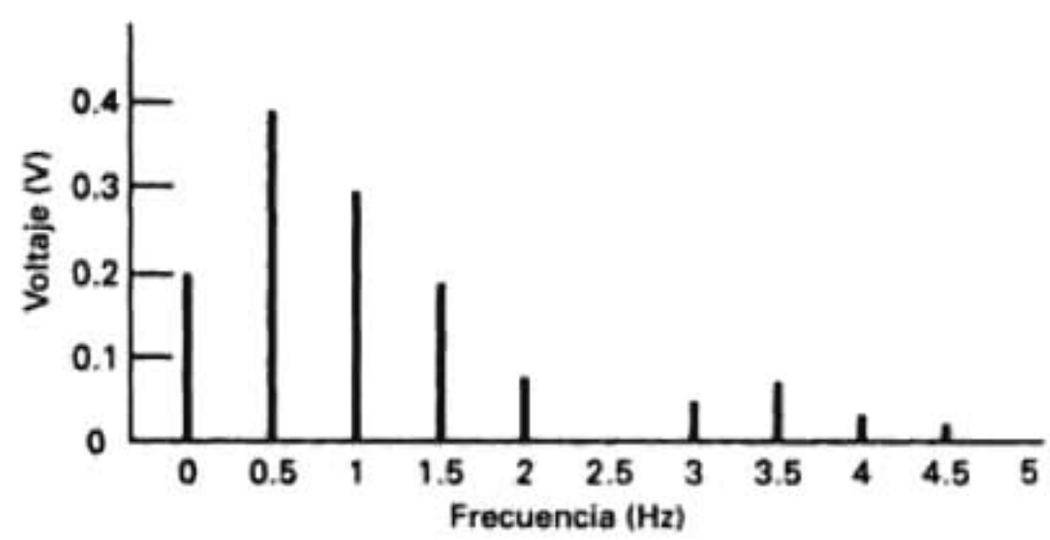


FIGURA 1-20 Espectro de frecuencias para el ejemplo 1-13

La fig. 1-21 muestra el efecto que tiene el reducir el ciclo de trabajo (es decir, reducir la relación τ/T) sobre el espectro de frecuencias, para una forma de onda no senoidal. Se puede apreciar que al angostar el ancho del pulso se produce un espectro de frecuencias con amplitud más uniforme. De hecho, para pulsos infinitamente angostos, el espectro de frecuencias abarca una cantidad infinita de frecuencias de igual amplitud, relacionadas armónicamente. Es imposible producir este espectro, y no digamos propagarlo. Eso explica por qué es difícil producir pulsos

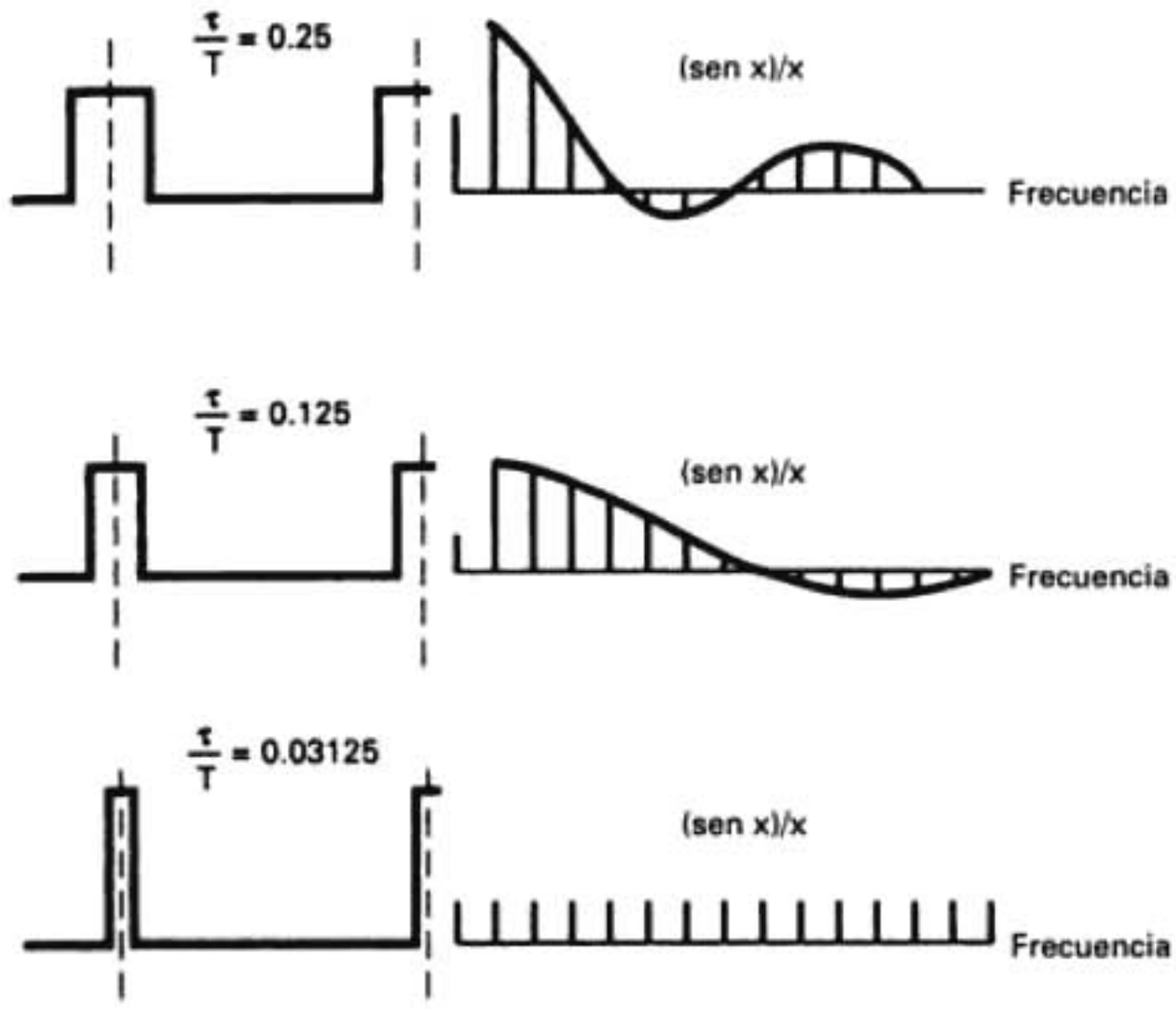


FIGURA 1-21 Efectos de reducir la relación τ/T (disminuyendo τ o aumentando T)

extremadamente angostos. Si se aumenta el periodo de una forma de onda rectangular, manteniendo constante el ancho del pulso, se obtiene el mismo efecto en el espectro de frecuencias.

Espectros de potencia y energía

En la sección anterior se usaron las series de Fourier para comprender mejor la representación de una señal compleja en el dominio de la frecuencia y del tiempo. Ambos dominios se pueden usar para ilustrar la relación de los voltajes (magnitudes) de señal con respecto a la frecuencia o al tiempo, para una señal variable en el tiempo.

Sin embargo, hay otra aplicación importante de la serie de Fourier. El objetivo de un canal de comunicaciones es transferir energía electromagnética de una fuente a un destino. Así, la relación entre la cantidad de energía transmitida y la recibida es una consideración importante. Entonces es importante examinar la relación entre la energía y la potencia en función de la frecuencia.

La *potencia eléctrica* es la rapidez con la cual se disipa, se entrega o se usa la energía, y es una función del cuadrado del voltaje o de la corriente ($P = E^2/R$ o $P = I^2 \times R$). Para las relaciones de potencia, la ecuación de Fourier, $f(t)$, se reemplaza por $[f(t)]^2$. La fig. 1-22 muestra el espectro de potencia para una forma de onda rectangular con 25% de ciclo de trabajo. Se parece al espectro de voltaje en función de frecuencia, excepto que tiene más lóbulos y un lóbulo primario mucho más grande. Nótese también que todos los lóbulos son positivos, porque no existe la potencia negativa.

En la fig. 1-22 se puede ver que la potencia de un pulso se dispersa dentro de un espectro de frecuencias relativamente ancho. Sin embargo, nótese que la mayor parte de la potencia está dentro del lóbulo primario. Por consiguiente, si el ancho de banda de un canal de comunicaciones tiene la suficiente anchura para pasar sólo las frecuencias del lóbulo primario, transferirá al receptor la mayor parte de la energía que contiene el pulso.

Transformadas de Fourier discreta y rápida

Muchas formas de onda que se manejan en los sistemas normales de comunicaciones no se pueden definir en forma satisfactoria con ecuaciones matemáticas; sin embargo, es de interés primordial su comportamiento en el dominio de la frecuencia. Con frecuencia hay necesidad de obtener este comportamiento de señales que se captan en el dominio del tiempo, es decir, en tiempo real.

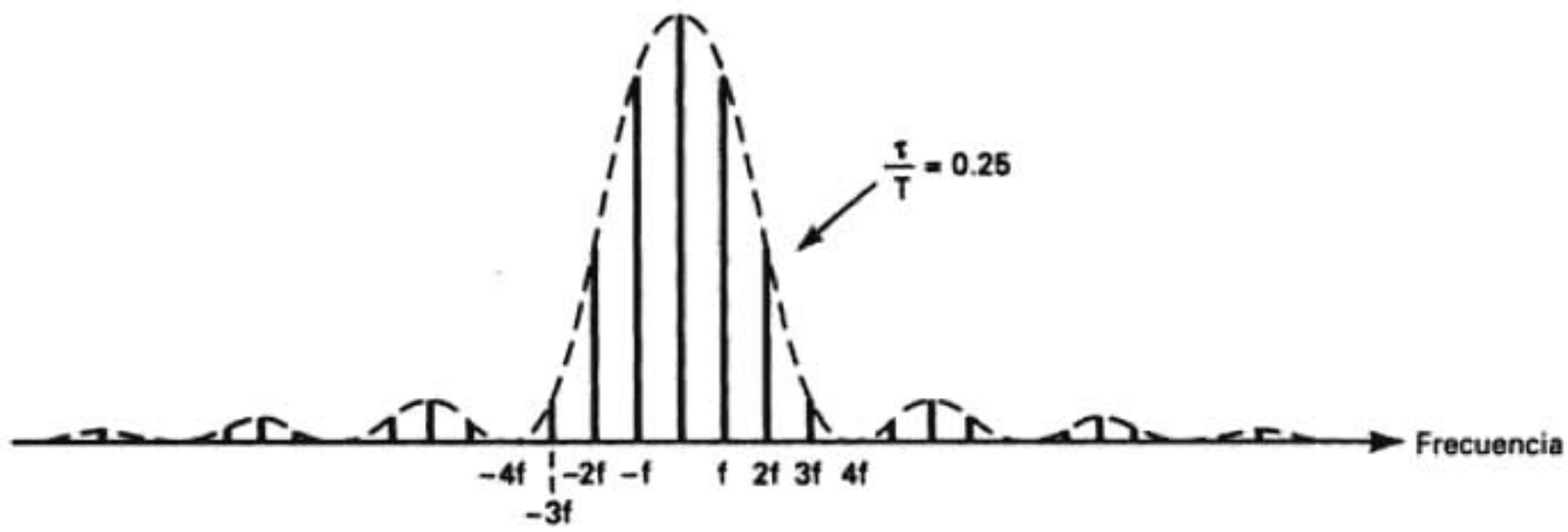


FIGURA 1-22 Espectro de potencias de un pulso rectangular con ciclo de trabajo de 25%

Ésta es la razón por la que se desarrolló la *transformada discreta de Fourier*. En esa transformación se muestrea una señal en el dominio del tiempo, en tiempos discretos. Las muestras se alimentan a una computadora donde un algoritmo calcula la transformación. Sin embargo, el tiempo de computación es proporcional a n^2, siendo n la cantidad de muestras. Para cualquier cantidad razonable de muestras, el tiempo de computación es excesivo. En consecuencia se desarrolló, en 1965, un nuevo algoritmo, llamado *transformada rápida de Fourier* (FFT, de *fast Fourier transform*), por Cooley y Tukey. Con la FFT, el tiempo de cómputo es proporcional a $n \log_2 n$, y no a n^2. Hoy se cuenta con la FFT en forma de subrutina, en muchas bibliotecas de subrutinas científicas en los grandes centros de cómputo.

Efectos de limitación de banda sobre las señales

Todos los canales de comunicaciones tienen un ancho de banda limitado y, en consecuencia, un efecto limitador sobre las señales que se propagan por ellos. Se puede considerar que un canal de comunicaciones es un *filtro ideal de fase lineal* con ancho de banda finito. Si una forma de onda repetitiva y no senoidal pasa por un filtro pasabajas ideal, se eliminan los componentes armónicos de frecuencia mayor que la frecuencia superior de corte del filtro. En consecuencia, cambia tanto el contenido de frecuencias como la forma de la onda. La fig. 1.23a muestra la forma de onda en el dominio del tiempo, para la onda cuadrada del ejemplo 1-2. Si esta forma de onda pasa por un filtro pasabajas, con frecuencia superior de corte de 8 kHz, se eliminan las frecuencias superiores a la octava armónica (9 kHz y mayores), y resulta la forma de onda de la fig. 1-23b. Las figs. 1-23c, d y e muestran las formas de onda que se producen cuando se usan filtros pasabajas con frecuencias superiores de corte de 6, 4 y 2 kHz, respectivamente.

Se puede ver en la fig. 1-23 que al *limitar la banda* de una señal cambia el contenido de frecuencias y por consiguiente la forma de onda. Si se impone una limitación suficiente de banda, al final la forma de onda sólo comprende a la frecuencia fundamental. En los sistemas de comunicaciones, la limitación de banda reduce la capacidad de información del sistema y si se impone demasiada limitación, se puede eliminar la señal de información de la forma de onda compuesta.

MEZCLADO

El *mezclado* es el proceso de combinar dos o más señales, y es un proceso esencial en comunicaciones electrónicas. En esencia hay dos formas en las que se pueden combinar o mezclar las señales: lineal y no lineal.

Suma lineal

La *suma lineal* se presenta cuando se combinan dos o más señales en un dispositivo lineal, como puede ser una red pasiva o un amplificador de señal pequeña. Las señales se combinan de tal manera que no se producen nuevas frecuencias, y la forma de onda combinada no es más que la suma lineal de las señales individuales. En la industria de grabación de audio, a veces se llama *mezclado lineal* a la suma lineal; sin embargo, en las radiocomunicaciones, el mezclado implica casi siempre un proceso no lineal.

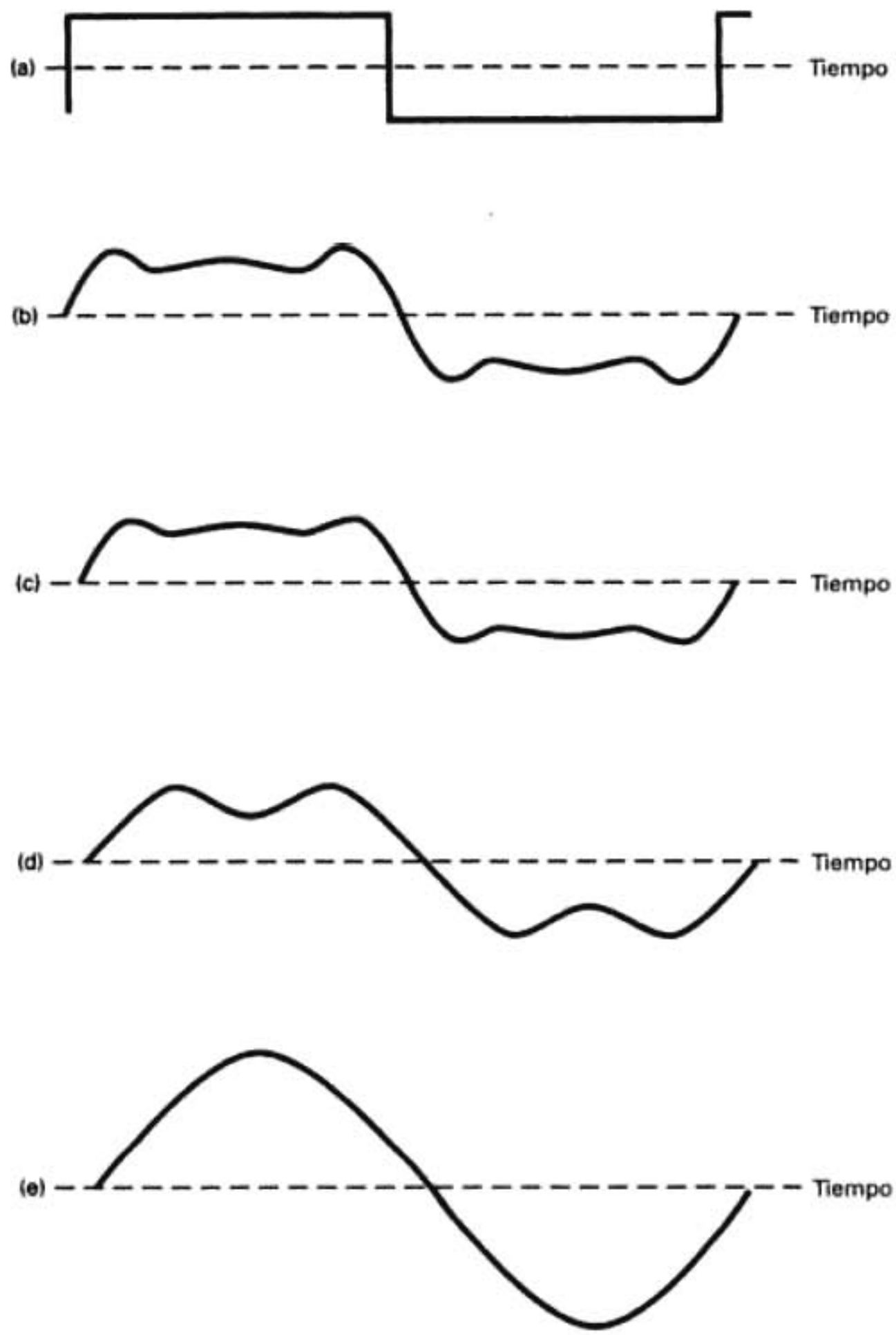


FIGURA 1-23 Señales de limitación de banda: (a) onda cuadrada de 1 kHz; (b) onda cuadrada de 1 kHz limitada por banda de 8 kHz; (c) onda cuadrada de 1 kHz limitada por banda de 6 kHz; (d) onda cuadrada de 1 kHz limitada por banda de 4 kHz; (e) onda cuadrada de 1 kHz limitada por banda de 2 kHz

Frecuencia única de entrada. La fig. 1-24a muestra la amplificación de una sola frecuencia de entrada en un amplificador lineal. La salida no es más que la señal original de entrada amplificada por la ganancia A del amplificador. La fig. 1-24b muestra la señal de salida en el dominio del tiempo, y la fig. 1-24c la indica en el dominio de la frecuencia. En forma matemática, la salida es

$$v_{\text{sal}} = Av_{\text{ent}} \tag{1-18}$$

o bien

$$v_{\text{ent}} = V_a \operatorname{sen} 2\pi f_a t$$

Por consiguiente,

$$v_{\text{sal}} = AV_a \operatorname{sen} 2\pi f_a t$$

Varias frecuencias de entrada. La fig. 1-25a muestra dos frecuencias de entrada que se combinan en un amplificador de señal pequeña. Cada frecuencia de entrada es amplificada con la ganancia A. Por consiguiente, la salida se expresa matemáticamente así

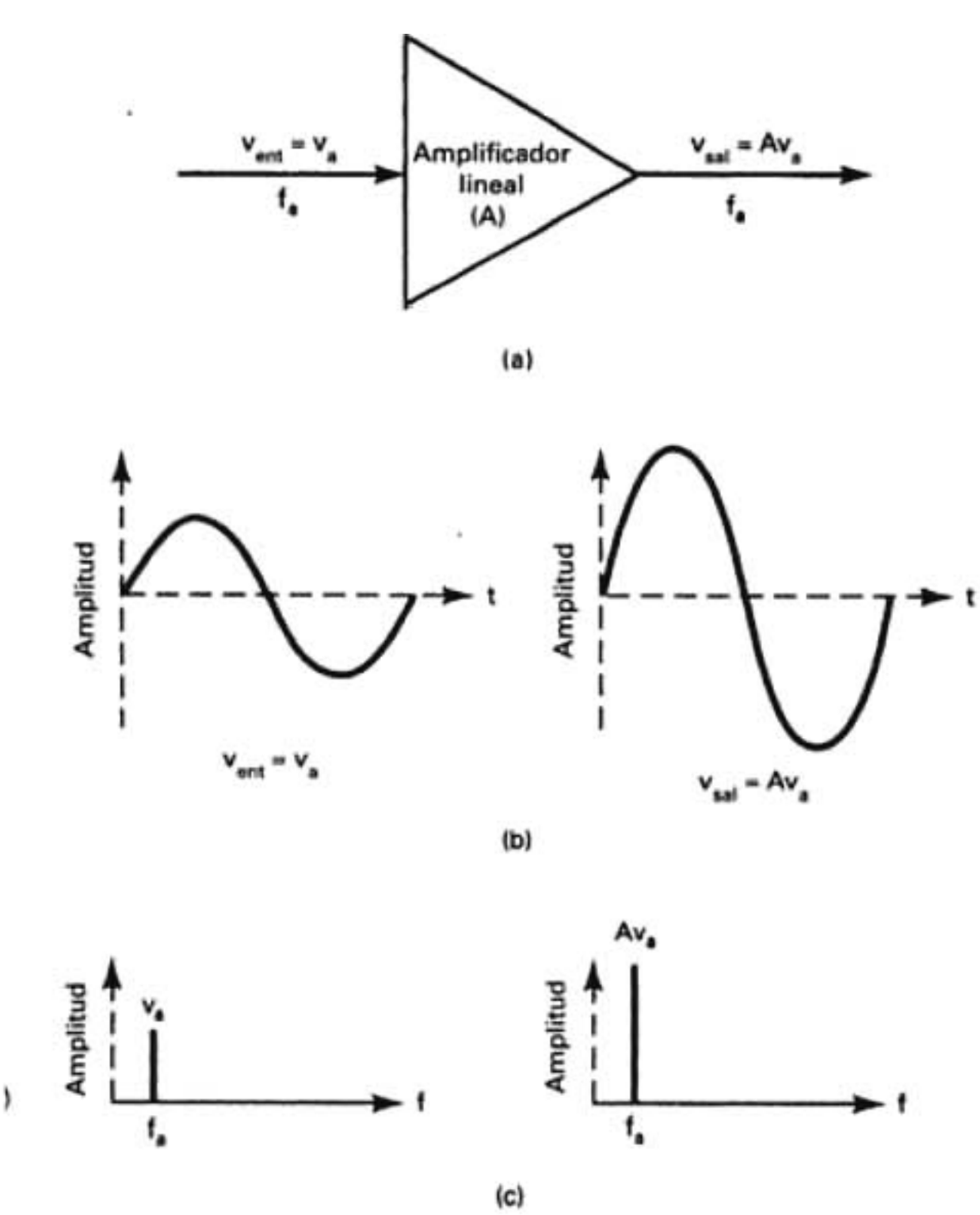


FIGURA 1-24 Amplificación lineal de una sola frecuencia de entrada: (a) amplificación lineal; (b) dominio del tiempo; (c) dominio de la frecuencia

$$v_{sal} = Av_{ent}$$

en donde $$v_{ent} = V_a \text{ sen } 2\pi f_a t + V_b \text{ sen } 2\pi f_b t$$

Por consiguiente, $$v_{sal} = A(V_a \text{ sen } 2\pi f_a t + V_b \text{ sen } 2\pi f_b t) \quad (1\text{-}19a)$$

o sea $$v_{sal} = AV_a \text{ sen } 2\pi f_a t + AV_b \text{ sen } 2\pi f_b t \quad (1\text{-}19b)$$

v_{sal} no es más que una forma de onda compleja que contiene las dos frecuencias de entrada, y es igual a la suma algebraica de v_a y v_b. La fig. 1-25b muestra la suma lineal de v_a y v_b en el dominio del tiempo, y la fig. 1-25c la muestra en el dominio de la frecuencia. Si se aplican más frecuencias de entrada al circuito, se suman linealmente con v_a y v_b. En los sistemas de audio de alta fidelidad es importante que el espectro de salida sólo contenga las frecuencias originales de entrada; en consecuencia, se prefiere la operación lineal. Sin embargo, en las radiocomunicaciones, donde es esencial la modulación, a veces es necesario el mezclado no lineal.

Mezclado no lineal

El *mezclado no lineal* sucede cuando se combinan dos o más señales en un dispositivo no lineal, como por ejemplo un diodo o un amplificador de señal grande. En el mezclado no lineal, las señales de entrada se combinan en forma no lineal y producen componentes adicionales de frecuencia.

Frecuencia única de entrada. La fig. 1-26a muestra la amplificación de una señal de entrada de frecuencia única mediante un amplificador no lineal. La salida de ese amplificador

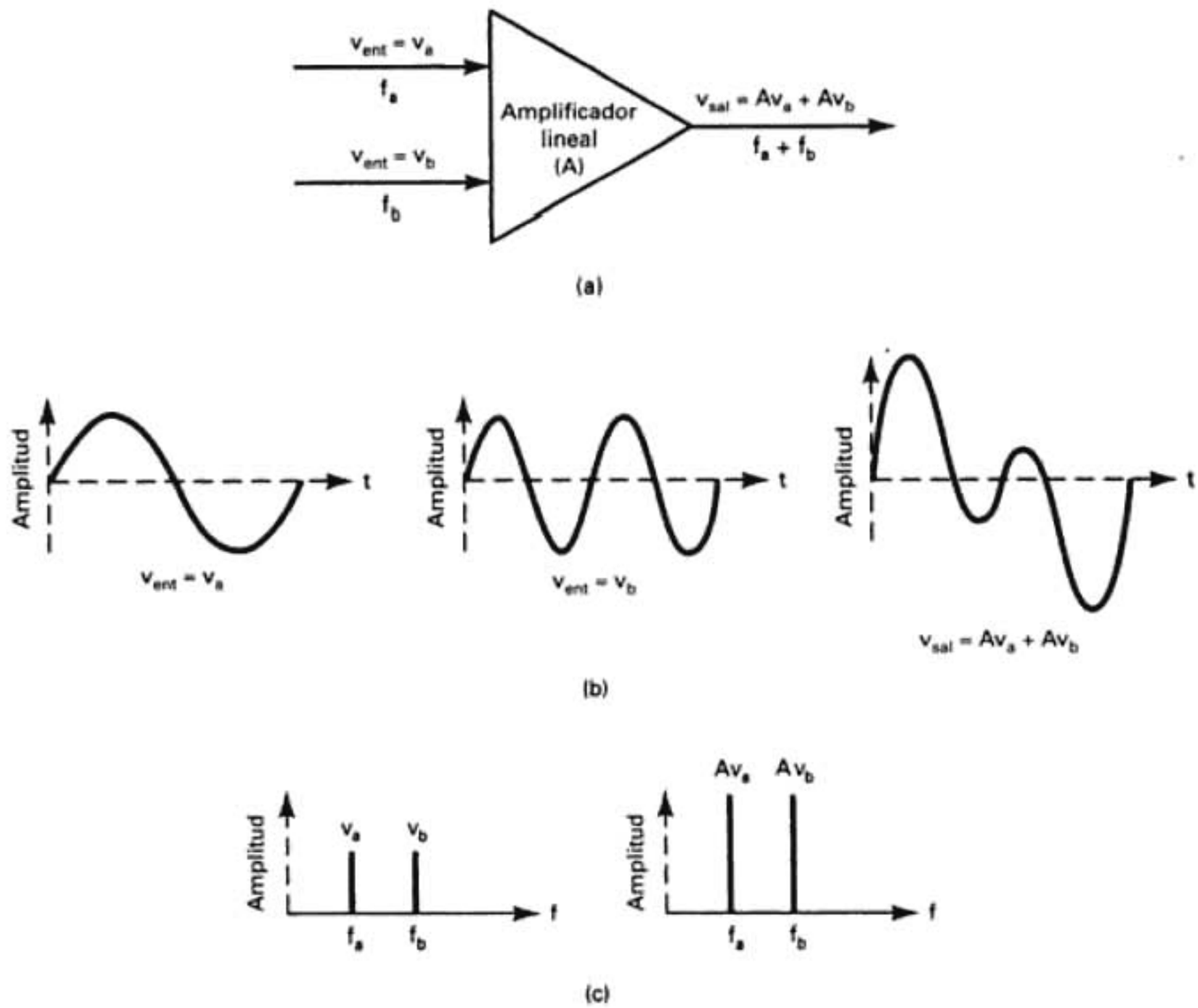


FIGURA 1-25 Mezclado lineal: (a) Amplificación lineal; (b) dominio del tiempo; (c) dominio de la frecuencia

no lineal en este caso no es una sola onda seno o coseno. Matemáticamente, la salida es la serie infinita de potencias

$$v_{sal} = Av_{ent} + Bv_{ent}^2 + Cv_{ent}^3 \tag{1-20a}$$

en donde $v_{ent} = V_a \text{ sen } 2\pi f_a t$

Por consiguiente, $v_{sal} = A(V_a \text{ sen } 2\pi f_a t) + B(V_a \text{ sen } 2\pi f_a t)^2 + C(V_a \text{ sen } 2\pi f_a t)^3$ (1-20b)

en donde Av_{ent} = término lineal, o simplemente la señal de entrada (f_a) amplificada por la ganancia A.

Bv_{ent}^2 = término cuadrático que genera la segunda frecuencia armónica ($2f_a$)

Cv_{ent}^3 = término cúbico que genera la tercera frecuencia armónica ($3f_a$)

v_{ent}^n produce una frecuencia igual a n por f. Por ejemplo, Bv_{ent}^2 genera una frecuencia igual a $2f_a$, Cv_{ent}^3 genera una frecuencia igual a $3f_a$, etcétera. A los múltiplos enteros de una frecuencia *base* se llaman *armónicas*. Como se dijo antes, la frecuencia de entrada original, f_a, es la primera armónica, o la frecuencia fundamental; $2f_a$ es la segunda armónica, $3f_a$ es la tercera, etcétera. La fig. 1-26b muestra la forma de onda de salida en el dominio del tiempo, para un amplificador no lineal con frecuencia única de entrada. Se ve que la forma de onda de salida no es más que la suma de la frecuencia de entrada con sus armónicas (múltiplos de la frecuencia fundamental). La fig. 1-26c muestra el espectro de salida en el dominio de la frecuencia. Nótese que las armónicas adyacentes están separadas entre sí por un valor igual a la frecuencia fundamental, f_a.

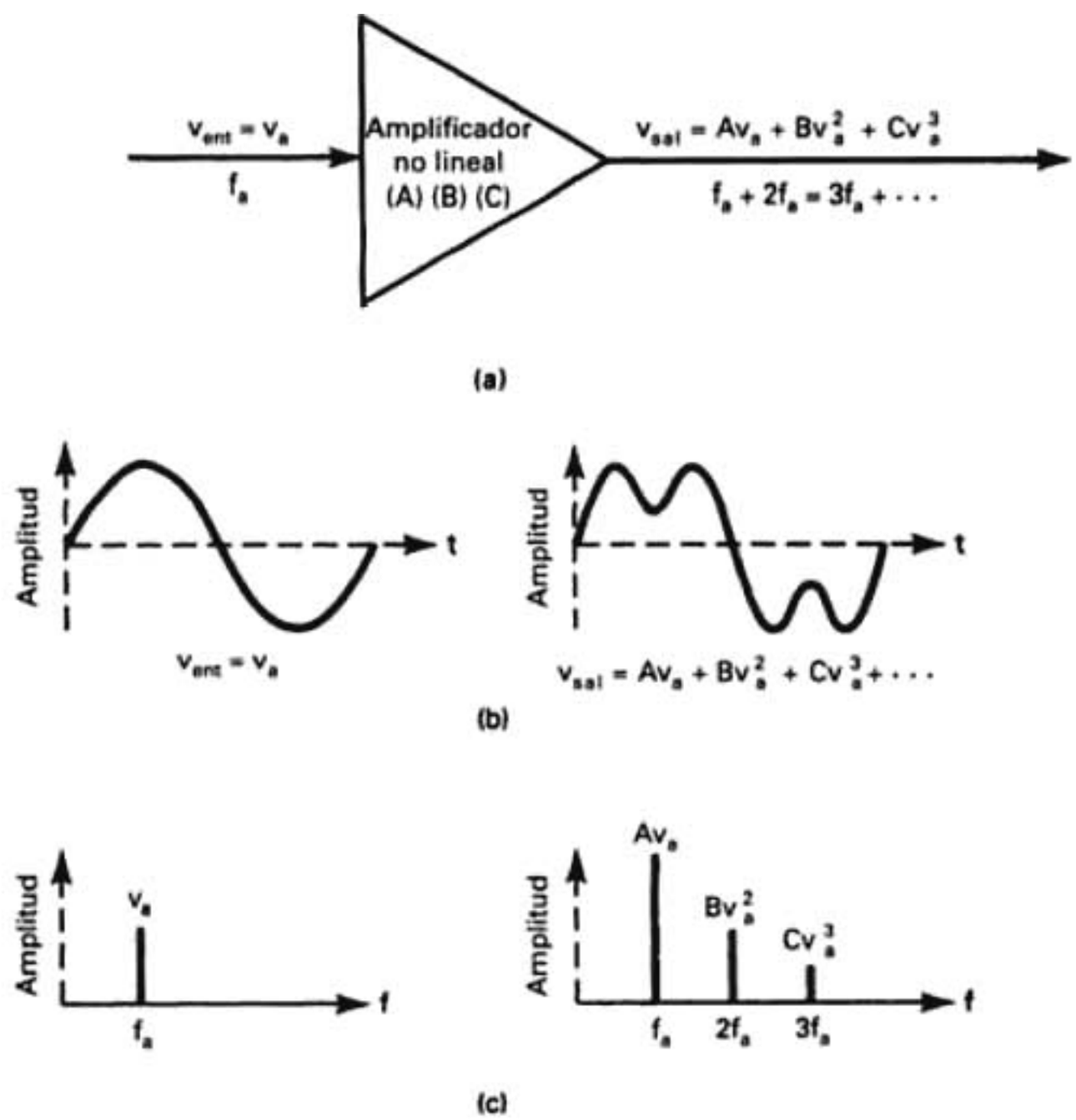


FIGURA 1-26 Amplificación no lineal de una sola frecuencia de entrada: (a) amplificación no lineal; (b) dominio del tiempo; (c) dominio de la frecuencia

La amplificación no lineal de una frecuencia única causa la generación de múltiplos, o armónicas, de esa frecuencia. Si las armónicas son perjudiciales, a esto se le llama *distorsión armónica*, o *distorsión por armónicas*. Si las armónicas son bienvenidas, se llama *multiplicación de frecuencia*.

Un JFET (transistor de efecto de campo de unión) es un caso especial de dispositivos no lineales, cuyas características se aproximan a las de uno de ley cuadrática. La salida de un dispositivo de ley cuadrática es

$$v_{sal} = Bv_{ent}^2 \tag{1-21}$$

La salida de un dispositivo de ley cuadrática con una sola frecuencia de entrada es cd y la segunda armónica. No se generan más armónicas que la segunda. Por consiguiente, se produce menos distorsión armónica con un JFET que con un BJT (transistor bipolar de unión).

Varias frecuencias de entrada. La fig. 1-27 muestra la amplificación no lineal de dos frecuencias de entrada por medio de un amplificador de señal grande (nc lineal). Matemáticamente, esa salida, con dos frecuencias de entrada, es

$$v_{sal} = Av_{ent} + Bv_{ent}^2 + Cv_{ent}^3$$

en donde

$$v_{ent} = V_a \operatorname{sen} 2\pi f_a t + V_b \operatorname{sen} 2\pi f_b t$$

Por consiguiente,

$$v_{sal} = A(V_a \operatorname{sen} 2\pi f_a t + V_b \operatorname{sen} 2\pi f_b t) + B(V_a \operatorname{sen} 2\pi f_a t + V_b \operatorname{sen} 2\pi f_b t)^2 + C(V_a \operatorname{sen} 2\pi f_a t + V_b \operatorname{sen} 2\pi f_b t)^3 + \cdots \tag{1-22a}$$

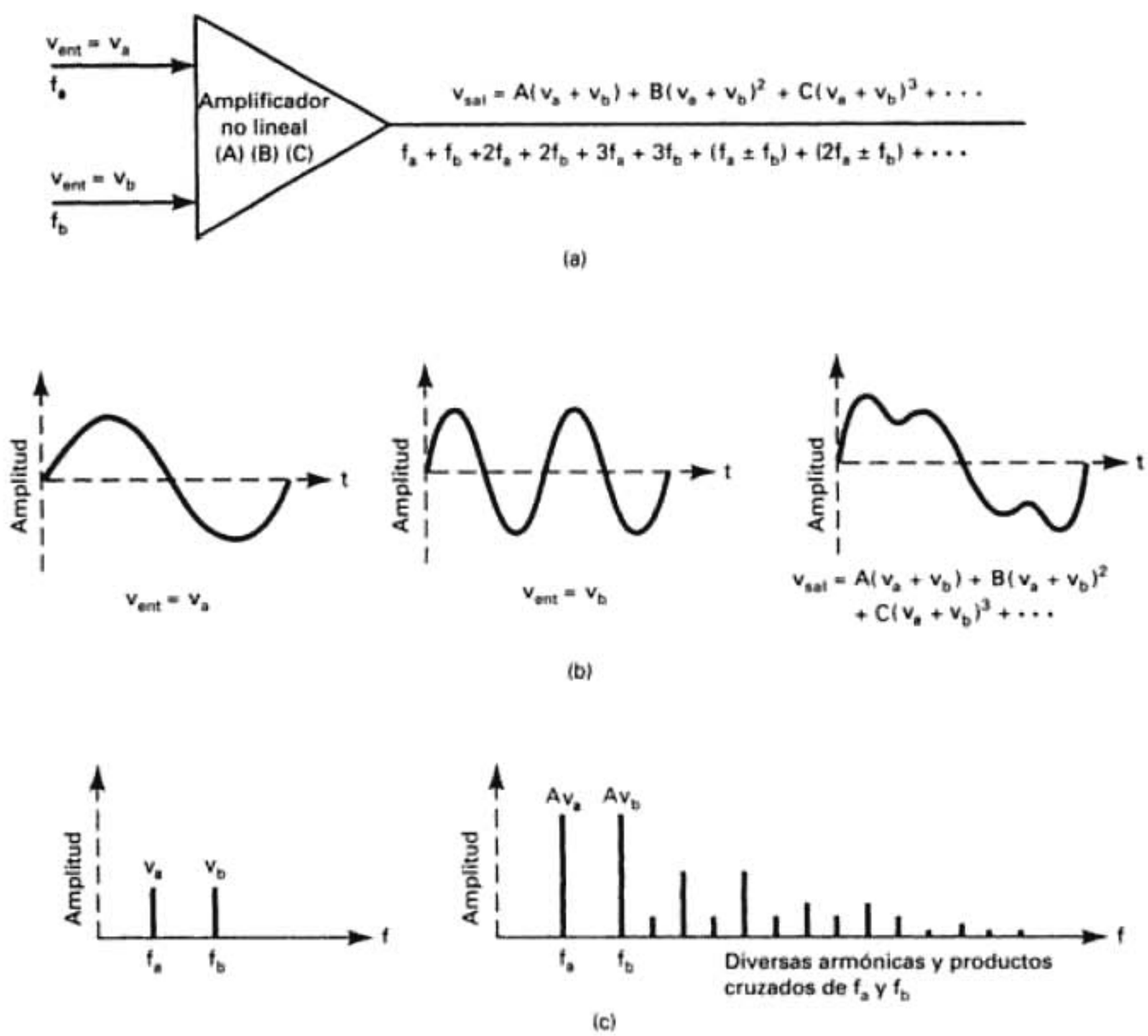


FIGURA 1-27 Amplificación no lineal de dos ondas senoides: (a) amplificación no lineal; (b) dominio del tiempo; (c) dominio de la frecuencia

La fórmula anterior es la de una serie infinita, y no hay límite de cantidad de términos que puede tener. Si se aplica el teorema del binomio a cada término de potencias mayores, la fórmula se puede reordenar para escribirla como sigue

$$v_{sal} = (Av_a' + Bv_a'^2 + Cv_a'^3 + \cdots) + (Av_b' + Bv_b'^2 + Cv_b'^3 + \cdots) + (2Bv_a'v_b' + 3Cv_a'^2v_b' + 3Cv_a'v_b'^2 + \cdots) \quad (1\text{-}22b)$$

en donde $v_a' = V_a \operatorname{sen} 2\pi f_a t$

$v_b' = V_b \operatorname{sen} 2\pi f_b t$

Los términos del primer conjunto entre paréntesis generan armónicas de f_a ($2f_a$, $3f_a$, etc.). Los términos en el segundo conjunto entre paréntesis generan armónicas de f_b ($2f_b$, $3f_b$, etc.). Los términos del tercer conjunto entre paréntesis generan los *productos cruzados* ($f_a + f_b$, $f_a - f_b$, $2f_a + f_b$, $2f_a - f_b$, etc.). Estos productos cruzados se producen en la *intermodulación* entre las dos frecuencias originales y sus armónicas. Los productos cruzados son las frecuencias de *suma* y de *diferencia*; son la suma y la diferencia de las dos frecuencias originales, las sumas y diferencias de sus armónicas, y las sumas y diferencias de las frecuencias originales y todas las armónicas. Se produce una cantidad infinita de frecuencias armónicas y de producto cruzado cuando se *mezclan* dos o más frecuencias en un dispositivo no lineal. Si no se desean los productos cruzados, se llama *distorsión por intermodulación*. Si se quieren tener los productos cruzados, se llama *modulación*. Matemáticamente, las frecuencias de suma y diferencia son

$$\text{productos cruzados} = mf_a \pm nf_b \quad (1\text{-}23)$$

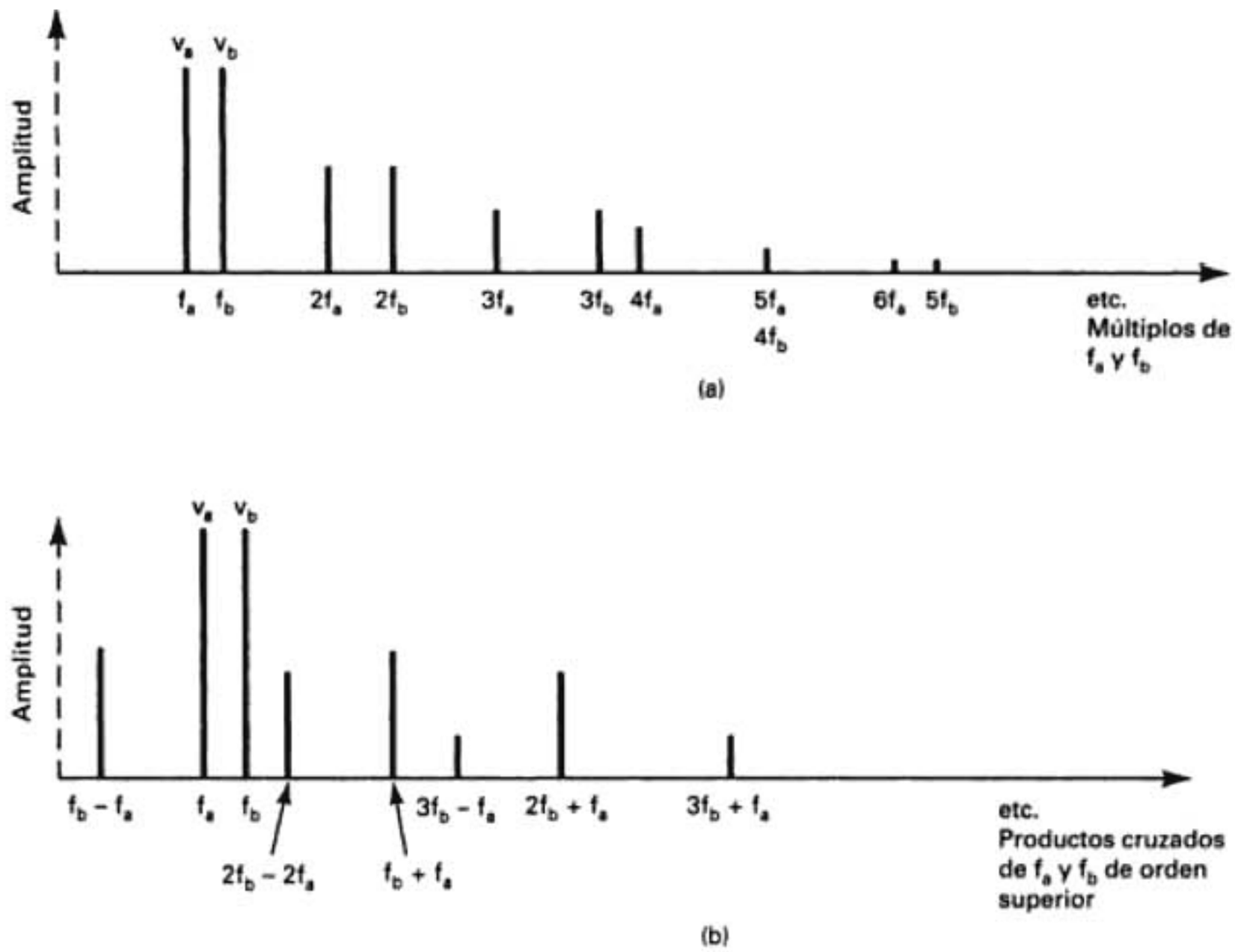


FIGURA 1-28 Espectro de salida de un amplificador no lineal con dos frecuencias de entrada: a) distorsión armónica; b) distorsión por intermodulación

siendo m y n enteros positivos, entre uno e infinito. La fig. 1-28 muestra el espectro de la salida de un amplificador no lineal con dos frecuencias de entrada.

La distorsión por intermodulación es la generación de cualquier frecuencia de producto cruzado no deseada, cuando se mezclan dos o más frecuencias en un dispositivo no lineal. En consecuencia, cuando se amplifican dos o más frecuencias en un dispositivo no lineal, en la salida habrá distorsiones tanto armónicas como de producto cruzado.

Ejemplo 1-4

Para un amplificador no lineal con 5 y 7 kHz de frecuencias de entrada:

(a) Determinar las tres primeras armónicas presentes en la salida, para cada frecuencia de entrada.

(b) Determinar los productos cruzados que se producen en la salida, para valores de m y n de 1 y 2.

(c) Trazar el espectro de frecuencias armónicas y de producto cruzado de salida, con las frecuencias determinadas en los pasos a) y b).

Solución (a) Las tres primeras armónicas comprenden las dos frecuencias originales de entrada, de 5 y 7 kHz; dos veces cada frecuencia original, 10 y 14 kHz, y tres veces cada frecuencia original, 15 y 21 kHz.

(b) Los productos cruzados con 1 y 2 como valores de m y n se determinan con la ecuación 1-23, y se resumen como sigue

m	n	Productos cruzados
1	1	7 kHz ± 5 kHz = 2 kHz y 12 kHz
1	2	7 kHz ± 10 kHz = 3 kHz y 17 kHz
2	1	14 kHz ± 5 kHz = 9 kHz y 19 kHz
2	2	14 kHz ± 10 kHz = 4 kHz y 24 kHz

(c) El espectro de frecuencias de salida se muestra en la fig. 1-29.

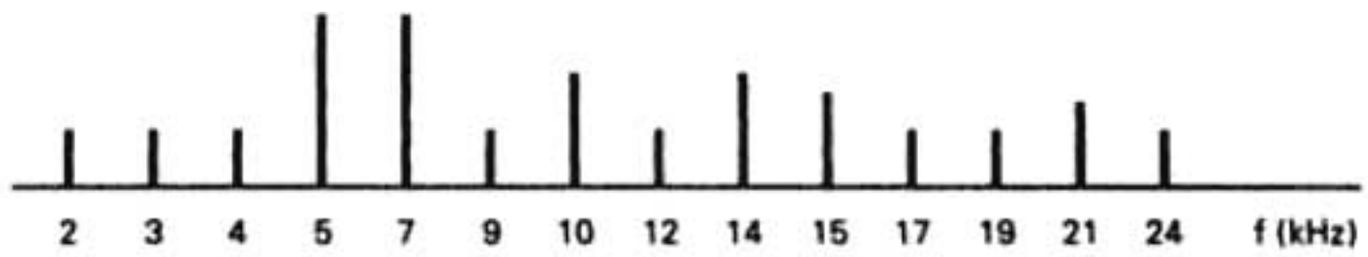


FIGURA 1-29 Espectro de salida para el ejemplo 1-4

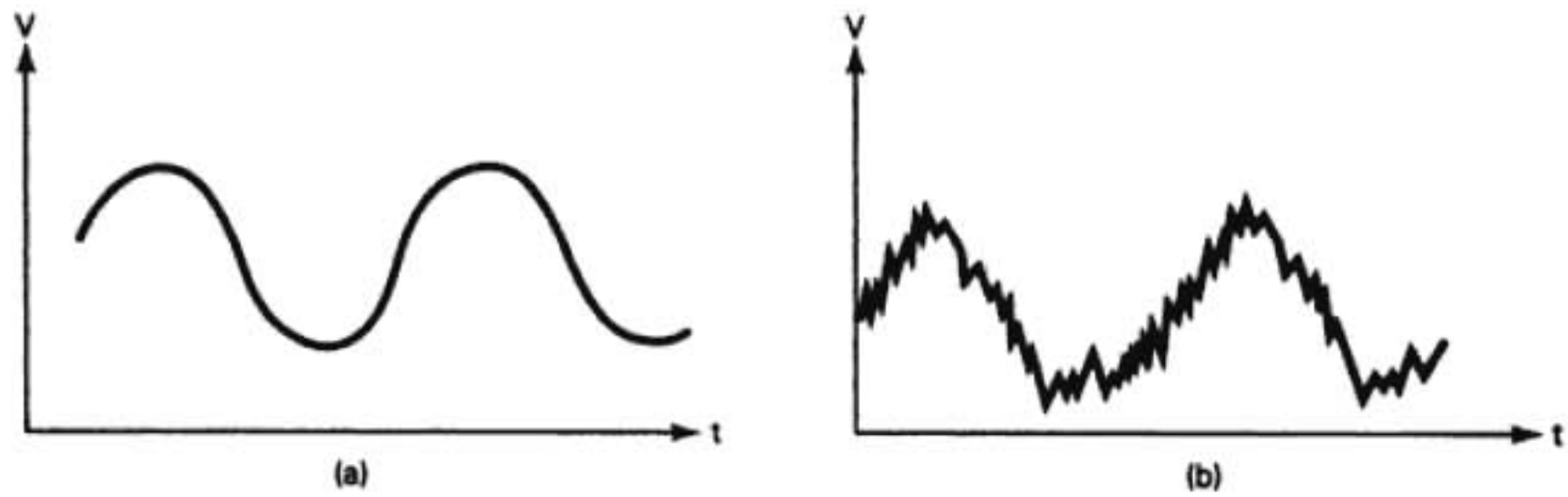


FIGURA 1-30 Efectos del ruido sobre una señal: (a) señal sin ruido; (b) señal con ruido

ANÁLISIS DE RUIDO

Se define al *ruido eléctrico* como cualquier energía eléctrica indeseable que queda entre la banda de paso de la señal. Por ejemplo, en la grabación de audio se consideran como ruido todas las señales eléctricas no deseadas que están dentro de la banda de frecuencias de audio, de 0 a 15 kHz, que interfieren con la música. La fig. 1-30 muestra el efecto que tiene el ruido sobre una señal eléctrica. En la fig. 1-30a se ve una senoide sin ruido, y en la fig. 1-30b se ve la misma señal, pero con la presencia de ruido.

Se puede clasificar al ruido en dos categorías: *correlacionado* y *no correlacionado*. La correlación implica una relación entre la señal y el ruido. Por consiguiente, el ruido correlacionado sólo existe cuando hay una señal. Por otra parte, el ruido no correlacionado está presente siempre, haya o no una señal.

Ruido no correlacionado

El ruido no correlacionado está presente independientemente de si haya una señal o no. Se puede seguir subdividiendo en dos categorías generales: externo e interno.

Ruido externo. El *ruido externo* es el que se genera fuera del dispositivo o circuito. Hay tres causas principales del ruido externo: atmosféricas, extraterrestres y generadas por el hombre.

Ruido atmosférico. El *ruido atmosférico* se origina en perturbaciones eléctricas naturales que se generan dentro de la atmósfera terrestre. Al ruido atmosférico se le suele llamar *electricidad estática*, parecido al de las frituras, que se oye con frecuencia en una bocina cuando no tiene señal. La fuente de la mayor parte de la electricidad estática se encuentra en las condiciones eléctricas naturales, como por ejemplo los rayos. A veces, la electricidad estática está en forma de pulsos que dispersan energía dentro de una amplia gama de frecuencias. Sin embargo, la magnitud de esta energía es inversamente proporcional a su frecuencia. Por consiguiente, el ruido atmosférico es relativamente insignificante a frecuencias mayores de más o menos 30 MHz.

Ruido extraterrestre. El *ruido extraterrestre* consiste en señales eléctricas que se originan fuera de la atmósfera de la Tierra y, en consecuencia, a veces se le llama *ruido de espacio profundo*. El ruido extraterrestre se origina en la Vía Láctea, en otras galaxias y en el Sol. También se subdivide en dos categorías: solar y cósmico.

El *ruido solar* lo genera en forma directa el calor solar. Hay dos partes de este ruido: la *producida* por una condición de calma, cuando existe una intensidad relativamente constante de radiación, y de *gran intensidad*, ocasionado por perturbaciones esporádicas debidas a *manchas solares* y a *protuberancias solares*. La magnitud del ruido esporádico causado por la actividad de manchas solares tiene una variación cíclica que se repite cada 11 años.

Las fuentes de *ruido cósmico* están distribuidas continuamente en las galaxias. Como las fuentes del ruido galáctico son mucho más lejanas que nuestro Sol, su intensidad de ruido es relativamente pequeña. Al ruido cósmico se le llama con frecuencia *ruido de cuerpo negro*, y se distribuye con bastante uniformidad por el cielo.

Ruido causado por el hombre. Las fuentes principales de este ruido son los mecanismos que producen chispas, como por ejemplo los conmutadores de los motores eléctricos, los sistemas de encendido automotriz, el equipo generador y conmutador de energía eléctrica y las lámparas fluorescentes. El ruido producido por el hombre tiene naturaleza de pulsos, y contiene una amplia gama de frecuencias, que se propagan por el espacio del mismo modo que las ondas de radio. Este ruido es más intenso en las metrópolis más densamente pobladas, y en las áreas industriales, por lo que a veces se le llama *ruido industrial*.

Ruido interno. El *ruido interno* es la interferencia eléctrica generada dentro de un dispositivo o circuito. Hay tres clases principales de ruido generado internamente: de disparo, de tiempo de tránsito y térmico.

Ruido de disparo. Este ruido se debe a la llegada aleatoria de portadoras (agujeros y electrones) al elemento de salida de un dispositivo electrónico, como por ejemplo un diodo, un transistor de efecto de campo o un transistor bipolar. El ruido de disparo se observó por primera vez en la corriente anódica de un amplificador de tubo al vacío, y W. Schottky lo describió matemáticamente en 1918. Los portadores de corriente (tanto para ca como cd) no se mueven en un flujo continuo y estable, porque la distancia que recorren varía debido a sus trayectorias aleatorias. El ruido de disparo varía en forma aleatoria, y se superpone a cualquier señal que haya. Cuando se amplifica, este ruido se oye como balines de metal que caen en un techo de lámina. A veces, al ruido de disparo se le llama *ruido de transistor*, y se suma al ruido térmico.

Ruido de tiempo de tránsito. Cualquier modificación a una corriente de portadores, cuando pasan de la entrada a la salida de un dispositivo (como por ejemplo, desde el emisor al colector de un transistor) produce una variación irregular y aleatoria, que se clasifica como *ruido de tiempo de tránsito*. Cuando el tiempo que tarda un portador en propagarse por un dispositivo es parte apreciable del tiempo de un ciclo de la señal, este ruido se hace perceptible. El ruido de tiempo de tránsito en los transistores se determina por la movilidad de los portadores, el voltaje de polarización y la clase de transistor. Los portadores que van del emisor al colector padecen demoras de tiempo en emisor, demoras de tiempo básico de tránsito y demoras de tiempo de recombinación y de propagación. Si las demoras de tránsito son excesivas en altas frecuencias, puede ser que el dispositivo agregue más ruido que amplificación de la señal.

Ruido térmico. Este ruido se asocia con el movimiento rápido y aleatorio de los electrones dentro de un conductor, producido por la *agitación térmica*. Este movimiento fue observado por primera vez por Robert Brown, botánico inglés. Observó primero pruebas de la naturaleza de la materia como partículas en movimiento, en los granos de polen. El movimiento aleatorio de los electrones fue reconocido por primera vez en 1927, por J. B. Johnson de los Bell Telephone Laboratories. Los electrones en el interior de un conductor portan una carga negativa unitaria, y la velocidad cuadrática media de uno de ellos es proporcional a su temperatura absoluta. En consecuencia, cada paso de un electrón entre choques con moléculas produce un corto pulso de corriente, que produce un voltaje pequeño a través del componente resistivo del conductor. Como este tipo de movimientos del electrón es totalmente aleatorio y es en todas direcciones, el voltaje promedio en la sustancia debido a esos movimientos es 0 V cd. Sin embargo, ese movimiento aleatorio sí produce una componente de ca.

La componente de ca debida a la agitación térmica tiene varios nombres, que incluyen al de *ruido térmico*, porque depende de la temperatura; también *movimiento browniano*, por su descubridor, *ruido de Johnson*, en honor de quien relacionó el movimiento browniano de las partículas con el movimiento de los electrones y, *ruido blanco*, porque el movimiento aleatorio se produce

en todas las frecuencias. Por consiguiente, el ruido térmico es el movimiento aleatorio de los electrones libres dentro de un conductor, causado por la agitación térmica.

Johnson demostró que la potencia del ruido térmico es proporcional al producto del ancho de banda por la temperatura. En forma matemática, la potencia del ruido es

$$N = KTB \tag{1-24}$$

donde N = potencia del ruido (watts)
B = ancho de banda (hertz)
K = constante de proporcionalidad de Boltzmann (1.38×10^{-23} joules por grado kelvin)
T = temperatura absoluta, en grados kelvin (la temperatura ambiente = 17° C o 290° K)

Para convertir de °C a grados kelvin sólo se suma 273°. Por consiguiente, T = °C + 273.

Ejemplo 1-5

Convertir las siguientes temperaturas a grados kelvin: 100° C, 0° C y −10° C

Solución Se aplica la fórmula T = °C + 273°, para convertir de °C a grados kelvin.

$$T = 100°\text{C} + 273° = 373°\ \text{K}$$
$$T = 0°\text{C} + 273° = 273°\ \text{K}$$
$$T = -10°\text{C} + 273° = 263°\ \text{K}$$

La potencia de ruido, expresada en dBm (decibelios referidos a 1 miliwatt), es una función logarítmica, igual a

$$N_{(\text{dBm})} = 10 \log \frac{KTB}{0.001} \tag{1-25}$$

Las ecuaciones 1-24 y 1-25 indican que en el cero absoluto (0° K o −273° C) no hay movimiento molecular aleatorio, y que el producto KTB es igual a cero.

Ejemplo 1-6

Convertir los siguientes valores de potencia absoluta en dBm: 0.002 W, 0.0001 W, 10 mW y 0.001 W.

Solución Los valores absolutos de potencia se convierten a unidades de dBm con la ecuación 1-25.

$$10 \log \frac{0.002}{0.001} = 3\ \text{dBm}$$
$$10 \log \frac{0.0001}{0.001} = -10\ \text{dBm}$$
$$10 \log \frac{10\ \text{mW}}{0.001} = 10\ \text{dBm}$$
$$10 \log \frac{0.001}{0.001} = 0\ \text{dBm}$$

Se puede ver, en el ejemplo 1-6, que las potencias mayores que 1 mW producen valores positivos de dBm, y las potencias menores que 1 mW producen valores negativos de dBm. Una potencia de 1 mW equivale a 0 dBm.

Al reordenar la ecuación 1-25 se obtiene

$$N_{(\text{dBm})} = 10 \log \frac{KT}{0.001} + 10 \log B \tag{1-26}$$

y para un ancho de banda de 1 Hz a temperatura ambiente,

$$N_{(\text{dBm})} = 10 \log \frac{(1.38 \times 10^{-23})(290)}{0.001} + 10 \log 1$$
$$= -174\ \text{dBm}$$

Así, la ecuación 1-25 se puede escribir, para cualquier ancho de banda a la temperatura ambiente, como sigue

$$N_{(\text{dBm})} = -174 \text{ dBm} + 10 \log B \tag{1-27}$$

El ruido aleatorio produce una densidad constante de potencia, en función de la frecuencia, y de acuerdo con la ecuación 1-24, la potencia disponible en una fuente de ruido térmico es proporcional al ancho de banda sobre cualquier intervalo de frecuencias. Esto se ha comprobado con frecuencias que van de 0 Hz hasta las máximas frecuencias de microondas que se usan en la actualidad. Así, si el ancho de banda es ilimitado, parece que la potencia disponible en una fuente de ruido térmico también es ilimitada. Eso, naturalmente, no es verdad, porque se puede demostrar que a frecuencias arbitrariamente altas la potencia de ruido térmico llega a bajar a cero. Como el ruido térmico se distribuye por igual en el espectro de frecuencias, a una fuente de ruido térmico se le llama a veces *fuente de ruido blanco*, por ser análoga a la luz blanca que contiene todas las frecuencias de la luz visible. Por lo anterior, la potencia rms (efectiva) medida a cualquier frecuencia en una fuente de ruido blanco es igual a la medida en cualquier otra frecuencia de la misma fuente de ruido. De igual manera, la potencia total rms de ruido medida en cualquier ancho de banda fijo es igual a la potencia total rms de ruido, medida en un ancho de banda igual en cualquier lugar del espectro total de ruido. En otras palabras, la potencia rms de ruido blanco presente en la banda de 1000 Hz a 2000 Hz es igual a la presente en la banda de 1,001,000 a 1,002,000 Hz.

El ruido térmico es aleatorio y continuo, y se produce en todas las frecuencias. También, es predecible, aditivo, y está presente en todos los dispositivos. Por lo anterior, el ruido térmico es el más importante de todos los ruidos.

Voltaje del ruido

La fig. 1-31 muestra el circuito equivalente de una fuente de ruido, donde su resistencia interna (R_I) está en serie con el voltaje rms de ruido (V_N). Para el peor de los casos, y para la transferencia máxima de la potencia del ruido, se iguala la resistencia de la carga (R) con la R_I. Así, la caída de voltaje de ruido a través de R es igual a la mitad del voltaje de la fuente ($V_R = V_N/2$), y según la ecuación 1-24, la potencia de ruido (N) desarrollada a través del resistor de carga es igual a KTB. La ecuación matemática de V_N se deduce como sigue

$$N = KTB = \frac{(V_N/2)^2}{R} = \frac{V_N^2}{4R}$$

Así,

$$V_N^2 = 4RKTB$$

y

$$V_N = \sqrt{4RKTB} \tag{1-28}$$

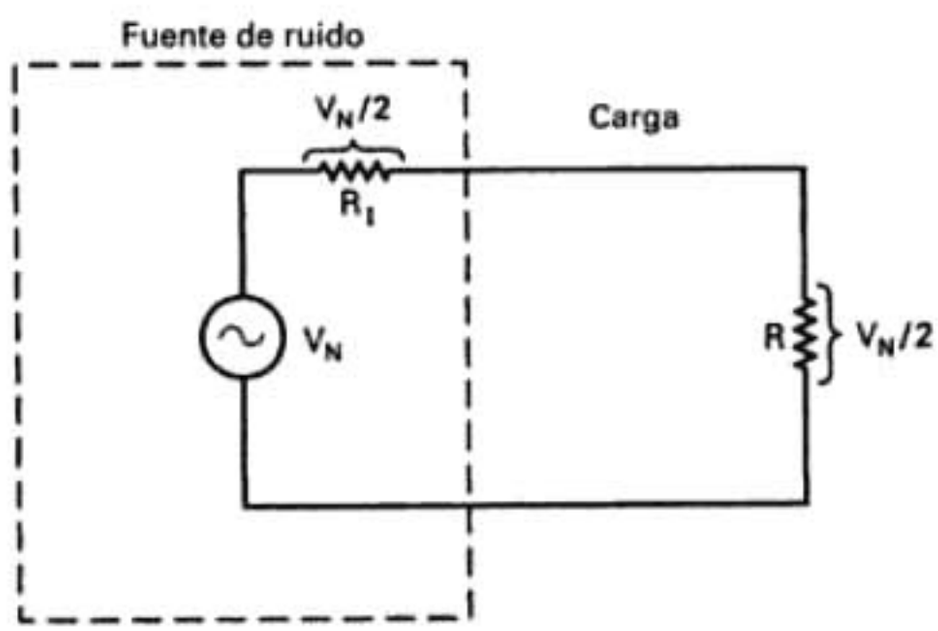


FIGURA 1-31 Circuito equivalente a una fuente de ruido

Ejemplo 1-7

Para un dispositivo electrónico que funciona a la temperatura de 17° C, con ancho de banda de 10 kHz, calcular:

(a) La potencia de ruido térmico, en watts y en dBm.

(b) El voltaje rms del ruido, para una resistencia interna de 100 Ω y una resistencia de carga de 100 Ω.

Solución (a) La potencia de ruido térmico se calcula sustituyendo en la ecuación 1-24.

$$N = KTB \qquad T\,(\text{kelvin}) = 17°\text{C} + 273° = 290°\ \text{K} \qquad B = 1 \times 10^4\ \text{Hz}$$
$$= (1.38 \times 10^{-23})(290)(1 \times 10^4) = 4 \times 10^{-17}\ \text{W}$$

Al sustituir en la ecuación 1-25 se obtiene la potencia del ruido en dBm.

$$N_{(\text{dBm})} = 10 \log \frac{[4 \times 10^{-17}]}{0.001} = -134\ \text{dBm}$$

o bien, se sustituye en la ecuación 1-27

$$N_{(\text{dBm})} = -174\ \text{dBm} + 10\ \text{Log}\ 10{,}000$$
$$= -174\ \text{dBm} + 40\ \text{dB}$$
$$= -134\ \text{dBm}$$

(b) El voltaje rms de ruido se calcula sustituyendo en la ecuación 1-28

$$V_N = \sqrt{4RKTB} \qquad \text{donde } KTB = 4 \times 10^{-17}$$
$$= \sqrt{(4)(100)(4 \times 10^{-17})} = 0.1265\ \mu\text{V}$$

Ejemplo 1-8

Convertir un valor de potencia de 13 dBm a watts.

Solución Los valores de potencia en dBm se pasan a watts reordenando la ecuación 1-25.

$$P_{(\text{dBm})} = 10 \log \frac{P_{(\text{watts})}}{0.001\ \text{W}}$$

$$13\ \text{dBm} = 10 \log \frac{P_{(\text{watts})}}{0.001\ \text{W}}$$

Se dividen entre 10 ambos lados de la ecuación, y se obtiene

$$\frac{13}{10} = \log \frac{P_{(\text{watts})}}{0.001\ \text{W}}$$

$$1.3 = \log \frac{P_{(\text{watts})}}{0.001\ \text{W}}$$

Se saca el antilogaritmo de ambos lados de la ecuación, para eliminar la función log del lado derecho.

$$10^{1.3} = \frac{P_{(\text{watts})}}{0.001\ \text{W}}$$

$$20 = \frac{P_{(\text{watts})}}{0.001\ \text{W}}$$

Al multiplicar por 0.001 ambos lados de la ecuación se llega al resultado

$$0.001(20) = P_{(\text{watts})}$$
$$P_{(\text{watts})} = 20\ \text{mW}$$

Ruido correlacionado

El *ruido correlacionado* es aquel que se relaciona mutuamente (se correlaciona) con la señal, y no puede estar en un circuito a menos que haya una señal de entrada. Dicho en términos sencillos: ¡no hay señal, no hay ruido! El ruido correlacionado se produce por amplificación no lineal, e incluye la distorsión armónica y de intermodulación, ya que las dos son formas de distorsión no lineal. Todos los amplificadores son no lineales en cierto grado. Por consiguiente, toda amplificación de señal produce distorsión no lineal. Esta distorsión también se produce cuando pasan las señales a través de dispositivos no lineales, como son los diodos. El ruido correlacionado es una forma de ruido interno.

Distorsión armónica. Hay *distorsión armónica* cuando se producen las armónicas no deseadas de una señal, debido a una amplificación no lineal (mezclado). Las armónicas son múltiplos enteros de la señal original de entrada. Esta señal original es la primera armónica, y se llama *frecuencia fundamental*. Dos por la frecuencia original de la señal es igual a la segunda armónica, tres origina la tercera, etcétera. Otro nombre de la distorsión armónica es *distorsión de amplitud*.

Hay varios grados de distorsión armónica. La *distorsión armónica de segundo orden* es la relación de la amplitud rms de la frecuencia de segunda armónica entre la amplitud rms de la frecuencia fundamental. La *distorsión armónica de tercer orden* es la relación de la amplitud rms de la tercera armónica entre la de la frecuencia fundamental, etcétera. La *distorsión armónica total* es la amplitud rms combinada de las armónicas superiores, dividida entre la amplitud rms de la frecuencia fundamental. La ecuación matemática de la distorsión armónica total (THD, de *total harmonic distortion*) es

$$\% \text{ THD} = \frac{v_{\text{superior}}}{v_{\text{fundamental}}} \times 100 \tag{1-29}$$

donde % THD = distorsión armónica porcentual total

v_{superior} = suma cuadrática de los voltajes rms de las armónicas superiores de la frecuencia fundamental, $\sqrt{v_2^2 + v_3^2 + v_n^2}$

$v_{\text{fundamental}}$ = voltaje rms de la frecuencia fundamental

Ejemplo 1-9

Determinar:

(a) La segunda, tercera y duodécima armónica de una onda repetitiva de 1 kHz.

(b) El porcentaje de distorsión armónica de segundo orden, tercer orden y total, para una frecuencia fundamental con amplitud de 8 *V*rms, una amplitud de segunda armónica de 0.2 *V*rms y de tercera armónica de 0.1 *V*rms.

Solución (a) las frecuencias armónicas no son más que múltiplos enteros de la frecuencia fundamental.

$$\text{2da armónica} = 2 \times \text{fundamental} = 2 \times 1 \text{ kHz} = 2 \text{ kHz}$$
$$\text{3ra armónica} = 3 \times \text{fundamental} = 3 \times 1 \text{ kHz} = 3 \text{ kHz}$$
$$\text{12a armónica} = 12 \times \text{fundamental} = 12 \times 1 \text{ kHz} = 12 \text{ kHz}$$

(b)
$$\% \text{ 2do orden} = \frac{V_2}{V_1} \times 100 = \frac{0.2}{8} \times 100 = 2.5\%$$
$$\% \text{ 3er orden} = \frac{V_3}{V_1} \times 100 = \frac{0.1}{8} \times 100 = 1.25\%$$
$$\% \text{ THD} = \frac{\sqrt{(0.2)^2 + (0.1)^2}}{8} = 2.795\%$$

Distorsión por intermodulación. Es la generación de *frecuencias indeseables de suma* y *diferencia*, cuando se amplifican dos o más señales en un dispositivo no lineal, que puede ser un amplificador de señal grande. Aquí la importancia la tiene la palabra *indeseable*, porque en los circuitos de comunicaciones con frecuencia se desea mezclar dos o más señales, y producir las frecuencias de suma y diferencia. Éstas son las llamadas frecuencias de *productos cruzados*. Los productos cruzados se producen cuando tanto las frecuencias armónicas como las fundamentales se mezclan en un dispositivo no lineal. Para que haya distorsión por intermodulación debe haber dos o más señales de entrada. La definición matemática de las frecuencias de suma y diferencia es

$$\text{productos cruzados} = mf_1 \pm nf_2 \tag{1-30}$$

donde f_1 y f_2 son frecuencias fundamentales, donde $f_1 > f_2$, y m y n son enteros positivos, entre uno e infinito.

Ejemplo 1-10

Para un amplificador no lineal con dos frecuencias de entrada, de 3 y 8 kHz, determinar:
(a) Las primeras tres armónicas presentes en la salida, para cada frecuencia de entrada.
(b) Las frecuencias de producto cruzado que se producen con los valores 1 y 2 de m y n.

Solución (a) En las tres primeras armónicas se incluyen las dos frecuencias originales, de 3 y 8 kHz; el doble de cada frecuencia original, 6 y 16 kHz, y tres veces cada frecuencia original, 9 y 24 kHz.
(b) Los productos cruzados para valores 1 y 2 de m y n en la ecuación 1-30 son los siguientes

m	n	Productos cruzados
1	1	8 kHz ± 3 kHz = 5 kHz y 11 kHz
1	2	8 kHz ± 6 kHz = 2 kHz y 14 kHz
2	1	16 kHz ± 3 kHz = 13 kHz y 19 kHz
2	2	16 kHz ± 6 kHz = 10 kHz y 22 kHz

Ruido impulsivo

El *ruido impulsivo* se caracteriza por tener picos de gran amplitud y corta duración dentro del espectro total del ruido. Como indica el nombre, el ruido impulsivo consiste en ráfagas repentinas de pulsos de forma irregular, que por lo general duran entre algunos microsegundos y una fracción de milisegundo, dependiendo de su amplitud y su origen. La importancia de los pulsos repentinos en las comunicaciones de voz suele ser más molesto que destructor, porque los pulsos producen un ruido corto, de explosión o de crepitación. Sin embargo, en los circuitos de datos este ruido impulsivo puede ser devastador.

Se produce más ruido impulsivo durante la transmisión por inducción mutua y por radiación electromagnética y, en consecuencia, se le suele considerar como una forma de ruido externo. Las fuentes normales del ruido impulsivo incluyen las transitorias producidas en los interruptores electromecánicos (como relevadores y solenoides), motores eléctricos, electrodomésticos y alumbrado (en especial, las lámparas fluorescentes); también, las líneas de transmisión eléctrica, los sistemas de encendido automotriz, las uniones soldadas de mala calidad y los rayos.

Interferencia

La *interferencia* es una forma de ruido externo y, como el nombre indica, significa "perturbar o estorbar". Se produce interferencia eléctrica cuando las señales de información de una fuente producen frecuencias que caen fuera de su ancho de banda asignado, e interfieren con otras señales de otra fuente. La mayor parte de la interferencia se produce cuando las armónicas o las frecuencias de producto cruzado de una fuente llegan a la banda de paso de un canal vecino. Por ejemplo, los radios CB transmiten señales en el intervalo de 27 a 28 MHz. Sus frecuencias de segunda armónica (54 a 56 MHz) caen dentro de la banda asignada a la televisión VHF (en especial el canal 3). Si una persona transmite en un radio CB y produce una componente de segunda armónica con gran amplitud, podría interferir con la recepción de TV de otras personas. La mayor parte de la interferencia se presenta en el espectro de radiofrecuencias, y se describirá con más detalle en capítulos posteriores de este libro.

Resumen del ruido

La tabla 1-4 es un resumen de las fuentes de ruido eléctrico que se describieron en este capítulo.

Relación de potencia de señal a ruido

La *relación de potencia de señal a ruido*, S/N (de signal-to-noise), es el cociente del valor de la potencia de la señal entre el valor de la potencia del ruido. Matemáticamente se expresa como sigue

$$\frac{S}{N} = \frac{P_s}{P_n} \tag{1-31}$$

en la cual P_s = potencia de la señal (watts)
P_n = potencia del ruido (watts)

TABLA 1-4 Fuentes de ruido eléctrico

- Ruido correlacionado (interno)
  - Distorsión no lineal
    - Distorsión armónica
    - Distorsión por intermodulación
- Ruido no correlacionado
  - Externo
    - Atmosférico
    - Extraterrestre
    - Solar
    - Cósmico
    - Causado por el hombre
    - Pulso
    - Interferencia
  - Interno
    - Térmico
    - Disparo
    - Tiempo de tránsito

Esta relación de potencia de señal a ruido se expresa con frecuencia en forma de función logarítmica, en unidades de decibeles.

$$\frac{S}{N}\,(\text{dB}) = 10 \log \frac{P_s}{P_n} \tag{1-32}$$

Ejemplo 1-11

Para un amplificador con potencia de señal de salida de 10 W y potencia de ruido de salida de 0.01 W, determinar la relación de potencia de señal a ruido.

Solución La relación de potencia de señal a ruido se calcula sustituyendo valores en la ecuación 1-31.

$$\frac{S}{N} = \frac{P_s}{P_n} = \frac{10}{0.01} = 1000$$

Para expresarla en dB, se sustituyen en la ecuación 1-32

$$\frac{S}{N}\,(\text{dB}) = 10 \log \frac{P_s}{P_n} = 10 \log \frac{10}{0.01} = 30 \text{ dB} \tag{1-33}$$

También se puede expresar la relación de potencia de señal a ruido en función de voltajes y resistencias, como se ve a continuación

$$\frac{S}{N}\,(\text{dB}) = 10 \log \frac{\dfrac{V_s^2}{R_{ent}}}{\dfrac{V_n^2}{R_{sal}}} \tag{1-34}$$

donde $\frac{S}{N}$ = relación de potencia de señal a ruido (decibeles)
R_{ent} = resistencia de entrada (ohms)
R_{sal} = resistencia de salida (ohms)
V_s = voltaje de la señal (volts)
V_n = voltaje del ruido (volts)

Si las resistencias de entrada y de salida del amplificador, el receptor o la red que se estén evaluando son iguales, la ecuación 1-34 se reduce a

$$\frac{S}{N}\,(\text{dB}) = 10 \log \left(\frac{V_s^2}{V_n^2}\right)$$

$$= 10 \log \left(\frac{V_s}{V_n}\right)^2$$

$$\frac{S}{N}(\text{dB}) = 20 \log \frac{V_s}{V_n} \tag{1-35}$$

Ejemplo 1-12

Para un amplificador con voltaje de señal de salida de 4 V, voltaje de ruido de salida de 0.005 V y resistencia de entrada y de salida de 50 Ω, calcular la relación de potencia de señal a ruido.

Solución Esta relación de potencias de señal a ruido se calcula sustituyendo en la ecuación 1-34.

$$\frac{S}{N}(\text{dB}) = 20 \log \frac{V_s}{V_n} = 20 \log \frac{4}{0.005} = 58.6 \text{ dB}$$

Factor de ruido y cifra de ruido

El *factor de ruido* (F) y la *cifra de ruido* (NF, de *noise figure*) son "cifras de mérito" para indicar cuánto se deteriora la relación de señal a ruido cuando una señal pasa por un circuito o una serie de circuitos. El factor de ruido no es más que un cociente de relaciones de potencia de señal a ruido en la entrada entre la relación de potencia de señal a ruido en la salida. La definición matemática del factor de ruido es

$$F = \frac{\text{relación de potencia de señal a ruido en entrada}}{\text{relación de potencia de señal a ruido en salida}} \text{ (relación adimensional)} \tag{1-36}$$

La cifra de ruido es sólo el factor de ruido expresado en dB, y es un parámetro de uso común para indicar la calidad de un receptor. La definición matemática de la cifra de ruido es

$$\text{NF(dB)} = 10 \log \frac{\text{relación de potencia de señal a ruido en entrada}}{\text{relación de potencia de señal a ruido en salida}} \tag{1-37}$$

o sea $\quad \text{NF(dB)} = 10 \log F$

En esencia, el coeficiente de ruido indica cuánto se deteriora la relación de señal a ruido cuando una forma de onda se propaga desde la entrada hasta la salida de un circuito. Por ejemplo, un amplificador con un coeficiente de ruido de 6 dB es aquel en el que la relación de señal a ruido en la salida es 6 dB menor que la que era en la entrada. Si un circuito no tiene ruido alguno en absoluto, y no agrega ruido adicional a la señal, la relación de señal a ruido en la salida será igual a la de la entrada. Para un circuito perfecto y sin ruido, el factor de ruido es 1 y la cifra de ruido es 0 dB.

Un circuito electrónico amplifica por igual señales y ruido, dentro de su banda de paso. Por consiguiente, si el amplificador es ideal y sin ruido, las señal de entrada y la de ruido se amplifican igual, y a la salida la relación de señal a ruido será igual a la de la entrada. Sin embargo, los amplificadores en realidad no son ideales. Por consiguiente, el amplificador agrega ruido generado internamente a la onda y reduce la relación general de señal a ruido. El ruido más predominante es el térmico, que se genera en todos los componentes eléctricos. Por lo anterior, todas las redes, amplificadores y sistemas agregan ruido a la señal y así reducen la relación general de señal a ruido, a medida que la señal avanza por ellos.

La fig. 1-32a muestra un amplificador ideal y sin ruido, con ganancia de potencia A_p, potencia de señal de entrada S_i y potencia de ruido en la entrada N_i. El valor de la señal de salida no es más que A_pS_i y el nivel de ruido en la salida es A_pN_i. Por lo anterior, las relaciones S/N en la entrada y la salida son iguales, y se expresan matemáticamente como sigue

$$\frac{S_{sal}}{N_{sal}} = \frac{A_pS_i}{A_pN_i} = \frac{S_i}{N_i}$$

siendo A_p la ganancia de potencia del amplificador.

La fig. 1-32b representa a un amplificador no ideal que genera un ruido interno N_d. Como en el caso del amplificador ideal sin ruido, tanto la señal de entrada como la de ruido son

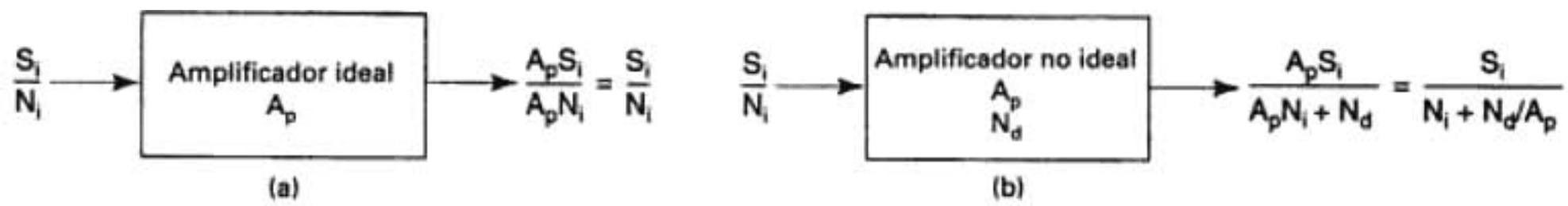


FIGURA 1-32 Cifra de ruido: (a) amplificador ideal sin ruido; (b) amplificador con ruido generado internamente

amplificadas en la ganancia del circuito. Sin embargo, el circuito agrega su ruido generado internamente a la onda. En consecuencia, la relación de señal a ruido en la salida es menor que a la entrada, en una cantidad proporcional a N_d.

$$\frac{S_{sal}}{N_{sal}} = \frac{A_pS_i}{A_pN_i + N_d} = \frac{S_i}{N_i + N_d/A_p}$$

donde A_p = ganancia de potencia del amplificador
N_d = potencia de ruido interno

Ejemplo 1-13

Para un amplificador no ideal con los siguientes parámetros, calcular:

(a) Relación S/N en la entrada (dB).
(b) Relación S/N en la salida (dB).
(c) Factor de ruido y la cifra de ruido.

Potencia de la señal de entrada = 2×10^{-10} W
Potencia de ruido en la entrada = 2×10^{-18} W
Ganancia de potencia = 1,000,000
Ruido interno (N_d) = 6×10^{-12} W

Solución (a) Para la señal de entrada y valores de potencia dados, y sustituyendo en la ecuación 1-33, la S/N en la entrada es

$$\frac{S}{N} = \frac{2 \times 10^{-10}\text{W}}{2 \times 10^{-18}\text{W}} = 100{,}000{,}000$$

$$10 \log (100{,}000{,}000) = 80 \text{ dB}$$

(b) La potencia de ruido en la salida es la suma del ruido interno y el ruido en la entrada amplificado

$$N_{sal} = 1{,}000{,}000(2 \times 10^{-18}) + 6 \times 10^{-12} = 8 \times 10^{-12} \text{ W}$$

La potencia de señal en la salida no es más que el producto de la potencia de entrada por la ganancia de potencia.

$$P_{sal} = 1{,}000{,}000(2 \times 10^{-10}) = 200 \ \mu\text{W}$$

Para la señal de salida y los valores de potencia calculados, y sustituyendo en la ecuación 1-33, la S/N a la salida es

$$\frac{S}{N} = \frac{200 \times 10^{-6}\text{ W}}{8 \times 10^{-12}\text{ W}} = 25{,}000{,}000$$

$$10 \log (25{,}000{,}000) = 74 \text{ dB}$$

(c) El factor de ruido se calcula sustituyendo los resultados de los pasos a) y b) en la ecuación 1-36

$$F = \frac{100{,}000{,}000}{25{,}000{,}000} = 4$$

y la cifra de ruido se calcula con la ecuación 1-37

$$NF = 10 \log 4 = 6 \text{ dB}$$

Cuando se conectan en cascada dos o más amplificadores, como se ve en la figura 1-33, el factor total de ruido es igual a la acumulación de los factores de ruido individuales. La *fórmula de Friiss* se usa para calcular el factor total de ruido de varios amplificadores en cascada. La fórmula de Friiss es la siguiente

$$F_T = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1A_2} + \frac{F_n - 1}{A_1A_2 \cdots A_n} \tag{1-38}$$

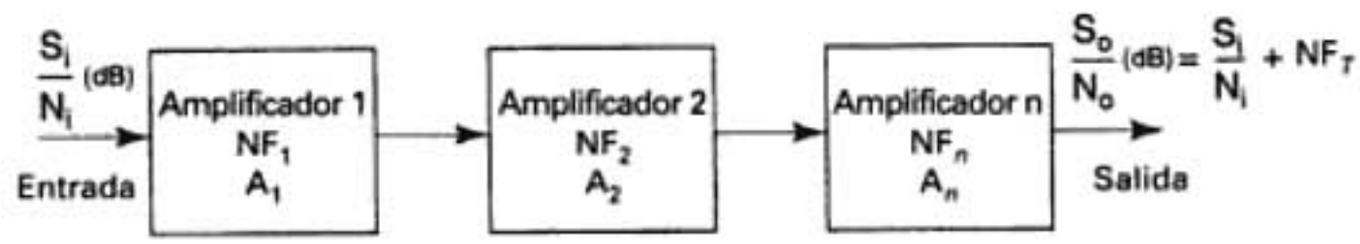


FIGURA 1-33 Cifra de ruido para amplificadores en cascada

en la que F_T = factor total de ruido para n amplificadores en cascada
F_1 = factor de ruido del amplificador 1
F_2 = factor de ruido del amplificador 2
F_3 = factor de ruido del amplificador 3
F_n = factor de ruido del amplificador n
A_1 = ganancia de potencia del amplificador 1
A_2 = ganancia de potencia del amplificador 2
A_3 = ganancia de potencia del amplificador n

Obsérvese que para usar la fórmula de Friiss, se deben convertir las cifras de ruido a factores de ruido. La cifra de ruido total es, simplemente,

$$NF_T(\text{dB}) = 10 \log F_T$$

Ejemplo 1-14

Calcular la cifra de ruido total para tres etapas de amplificación en cascada, cuyas cifras de ruido son 3 dB y sus ganancias de potencia son 10 dB.

Solución Hay que convertir las cifras de ruido en factores de ruido, y después sustituir en la ecuación 1-38. De este modo, el factor total de ruido es

$$F_T = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1 A_2} + \frac{F_n - 1}{A_1 A_2 \cdots A_n}$$
$$= 2 + \frac{2-1}{10} + \frac{2-1}{100} = 2.11$$

Así, la cifra de ruido total es

$$NF_T = 10 \log 2.11 = 3.24 \text{ dB}$$

Se deben hacer algunas observaciones importantes acerca del ejemplo 1-14. La primera es que la cifra de ruido total de 3.24 dB no fue mucho más grande que la cifra de ruido de la primera etapa (3 dB). Se puede ver en la ecuación 1-38 que la primera etapa en una serie de amplificadores como los de los amplificadores de audio y los receptores de radio, contribuye más la cifra de ruido total. Esto es cierto siempre y cuando la ganancia de la primera etapa sea suficiente para reducir los efectos de las etapas posteriores. Por ejemplo, si A_1 y A_2 en el ejemplo 1-14 sólo fueran 3 dB, la cifra de ruido total sería 4.4 dB, que indica un aumento importante. Es más, si la primera etapa fuera pasiva y tuviera una pérdida de 3 dB ($A = 0.5$), la cifra de ruido total aumentaría a 7.16 dB.

La figura 1-34 muestra cómo se puede reducir la relación de señal a ruido a medida que la señal pasa por un circuito amplificador de dos etapas. Como muestra la figura, tanto la señal de entrada como el ruido en la entrada se amplifican 10 dB en el amplificador 1. Sin embargo, este amplificador agrega 1.5 dB más de ruido, es decir, una cifra de ruido de 1.5 dB; entonces, reduce la relación de señal a ruido a la salida del amplificador 1 a 28.5 dB. De nuevo, la señal y el ruido se amplifican 10 dB en el amplificador 2. Sin embargo, el amplificador 2 agrega 2.5 dB de ruido, es decir, tiene un coeficiente de ruido de 2.5 dB; por consiguiente, reduce la relación de señal a ruido a la salida del amplificador 2 a 26 dB. La reducción general en la relación de señal a ruido, desde la entrada del amplificador 1 hasta la salida del amplificador 2 es 4 dB. Por lo anterior, la cifra de ruido total para los dos amplificadores es 4 dB.

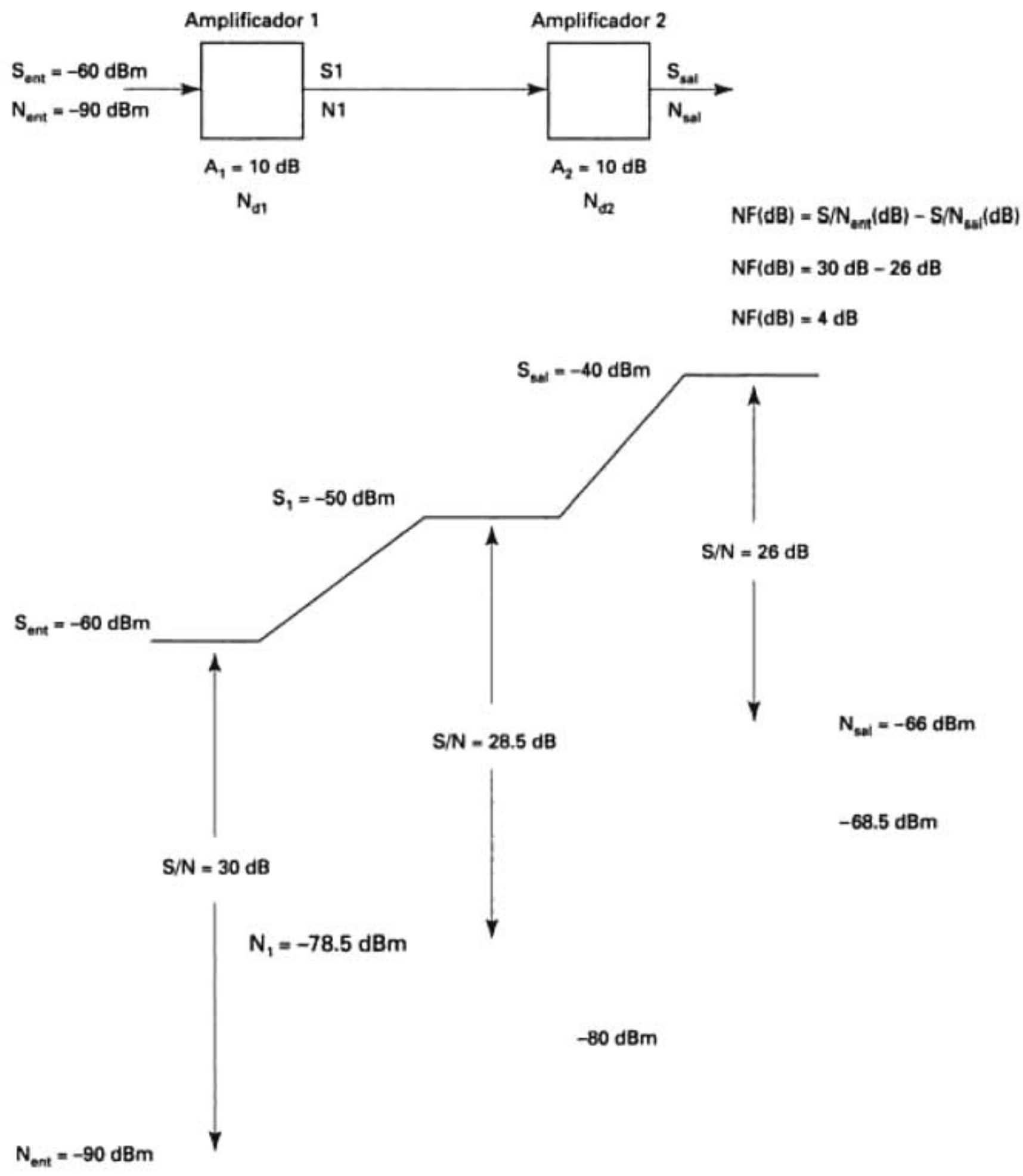


FIGURA 1-34 Degradación de la cifra de ruido en amplificadores conectados en cascada

Temperatura equivalente de ruido

Como el ruido producido por la agitación térmica es directamente proporcional a la temperatura, el ruido térmico se puede expresar en grados, y también en watts o en dBm. Si se reacomoda la ecuación 1-24 se obtiene

$$T = \frac{N}{KB} \tag{1-39}$$

en donde T = temperatura ambiente (en grados kelvin)
N = potencia del ruido (watts)
K = constante de Boltzmann (1.38×10^{-23} J/K)
B = ancho de banda (hertz)

La *temperatura equivalente de ruido*, T_e, es un valor hipotético que no se puede medir en forma directa. Es un parámetro conveniente que se usa con frecuencia en vez del coeficiente en los radiorreceptores complicados de bajo ruido, de VHF, UHF, microondas y satelitales. El uso de T_e como factor de ruido indica la reducción en la relación de señal a ruido que sufre una señal al propagarse a través de un receptor. Mientras menor sea la temperatura equivalente de ruido, la calidad del receptor es mejor. Un coeficiente de ruido de 1 dB corresponde a una temperatura equivalente de ruido de 75° K, y una de 6 dB corresponde a una temperatura equivalente de ruido de 870° K. Los valores normales de T_e van desde 20° K para receptores fríos hasta 1000° K para receptores ruidosos. La descripción matemática de T_e en la entrada de un receptor se expresa como sigue

$$T_e = T(\mathrm{F} - 1) \tag{1-40}$$

en donde T_e = temperatura equivalente de ruido (en grados kelvin)
T = temperatura del ambiente (valor de referencia: 290° K)
F = factor de ruido (adimensional)

Al revés, el factor de ruido se puede representar como una función de la temperatura equivalente de ruido, con la siguiente fórmula

$$\mathrm{F} = 1 + \frac{T_e}{T} \tag{1-41}$$

Ejemplo 1-15

Determinar:

(a) La cifra de ruido para una temperatura equivalente de ruido de 75° K (usar 290° K como temperatura de referencia).

(b) La temperatura equivalente de ruido para una cifra de ruido de 6 dB.

Solución (a) Se sustituye el dato en la ecuación 1-41, y se obtiene el factor de ruido de

$$\mathrm{F} = 1 + \frac{T_e}{T} = 1 + \frac{75}{290} = 1.258$$

y la cifra de ruido es tan sólo

$$\mathrm{NF} = 10 \log(1.258) = 1 \text{ dB}$$

(b) El factor de ruido se calcula reordenando la ecuación 1-37

$$\mathrm{F} = \text{antilog } (\mathrm{NF}/10) = \text{antilog } (6/10) = (10)^{0.6} = 4$$

Esto se sustituye en la ecuación 1-40 y se obtiene

$$T_e = T(\mathrm{F} - 1) = 290(4 - 1) = 870° \text{ K}$$

PREGUNTAS

1-1. Defina *comunicaciones electrónicas*.

1-2. ¿Cuándo se desarrolló el primer sistema electrónico de comunicaciones, quién lo desarrolló y qué clase de sistema era?

1-3. ¿Cuándo comenzaron las *radiocomunicaciones*?

1-4. ¿Cuáles son los tres componentes principales de un sistema de comunicaciones?

1-5. ¿Cuáles son los dos tipos básicos de *sistemas electrónicos de comunicaciones*?

1-6. ¿Qué organización asigna frecuencias para la radio propagación en el espacio libre, en Estados Unidos?

1-7. Describa lo siguiente: *señal portadora*, *señal moduladora* y *onda modulada*.

1-8. Describa los términos *modulación* y *demodulación*.

1-9. ¿Cuáles son las tres propiedades de una onda senoidal que se pueden variar, y qué tipo de modulación resulta en cada una de ellas?

1-10. Haga una lista y describa las razones por las que es necesaria la modulación en las comunicaciones electrónicas.

1-11. Describa la *conversión elevadora* de frecuencia, y dónde se hace.

1-12. Describa la *conversión reductora* de frecuencia, y dónde se hace.

1-13. Mencione y describa las dos limitaciones más importantes en el funcionamiento de un sistema de comunicaciones electrónicas.

1-14. ¿Qué es *capacidad de información* de un sistema de comunicaciones?

1-15. Describa en resumen el significado de la *ley de Hartley*.

1-16. Describa el *análisis de señales* en lo que concierne a las comunicaciones electrónicas.

1-17. ¿Qué quiere decir *simetría par*? ¿Cuál es un sinónimo de simetría par?

1-18. ¿Qué quiere decir *simetría impar*? ¿Cuál es un sinónimo de simetría impar?

1-19. ¿Qué quiere decir *simetría de media onda*?

1-20. Describa el significado del término *ciclo de trabajo*.

1-21. Describa una función (sen x)/x.

1-22. Defina la *suma lineal*.

1-23. Defina el *mezclado no lineal*.

1-24. Describa el *ruido eléctrico*.

1-25. ¿Cuáles son las dos categorías generales del ruido eléctrico?

1-26. La frase *no hay señal, no hay ruido* describe ¿a qué tipo de interferencia eléctrica?

1-27. Haga una lista de los tipos de ruido, y describa cuáles se consideran *ruido externo*.

1-28. ¿Cuál es el tipo predominante de *ruido interno*?

1-29. Describa la relación entre la *potencia de ruido térmico*, *ancho de banda* y *temperatura*.

1-30. Describa lo que es el *ruido blanco*.

1-31. Mencione y describa los dos tipos de *ruido correlacionado*.

1-32. Describa lo que es relación de *potencia de señal a ruido*.

1-33. ¿Qué quieren decir los términos *factor de ruido* y *cifra de ruido*?

1-34. Defina la *temperatura equivalente de ruido*.

1-35. Describa lo que es una *armónica* y una *frecuencia de producto cruzado*.

PROBLEMAS

1-1. ¿Cuál es la designación CCIR de los siguientes intervalos de frecuencia?
(a) 3 kHz a 30 kHz. **(b)** 0.3 MHz a 3 MHz. **(c)** 3 GHz a 30 GHz.

1-2. ¿Cuál es el intervalo de frecuencia para las siguientes designaciones CCIR?
(a) UHF. **(b)** ELF. **(c)** SHF.

1-3. ¿Cuál es el efecto, sobre la capacidad de información de un canal de comunicaciones, de ampliar al doble el ancho de banda asignado? ¿De triplicarlo?

1-4. ¿Cuál es el efecto, sobre la capacidad de información de un canal de comunicaciones, de reducir a la mitad el ancho de banda y subir al doble el tiempo de transmisión?

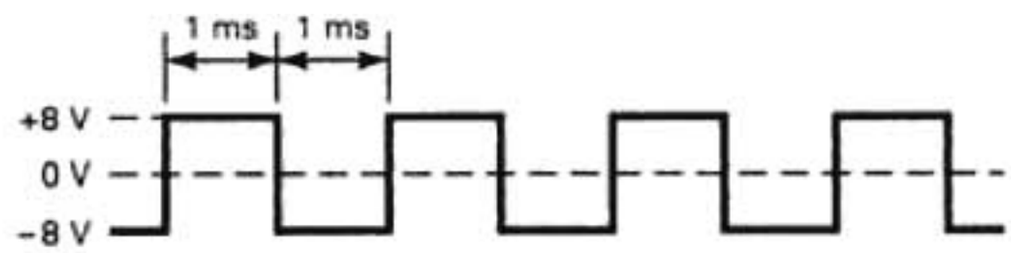
1-5. Convierta las siguientes temperaturas en grados kelvin:
(a) 17° C. **(b)** 27° C. **(c)** −17° C. **(d)** −50° C.

1-6. Convierta las siguientes potencias de ruido térmico en dBm:
(a) 0.001 μW. **(b)** 1 pW. **(c)** 2×10^{-15} W. **(d)** 1.4×10^{-16} W.

1-7. Convierta en watts las siguientes potencias de ruido térmico:
(a) −150 dBm. **(b)** −100 dBm. **(c)** −120 dBm. **(d)** −174 dBm.

1-8. Calcule la potencia de ruido térmico, en watts y en dBm, para los siguientes anchos de banda y temperaturas de un amplificador:
(a) $B = 100$ Hz, $T = 17°$ C.
(b) $B = 100$ kHz, $T = 100°$ C.
(c) $B = 1$ MHz, $T = 500°$ C.

1-9. Para el tren de ondas cuadradas de la figura siguiente:
(a) Determine las amplitudes de las primeras cinco armónicas.
(b) Trace el espectro de frecuencias.
(c) Trace el diagrama de la señal, en el dominio del tiempo, de las componentes de frecuencia hasta la quinta armónica.



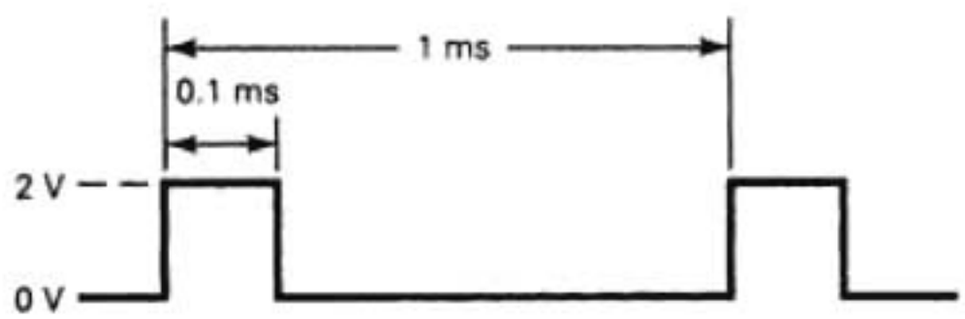
1-10. Para la forma de onda del pulso en la figura siguiente:

(a) Determine la componente de cd.

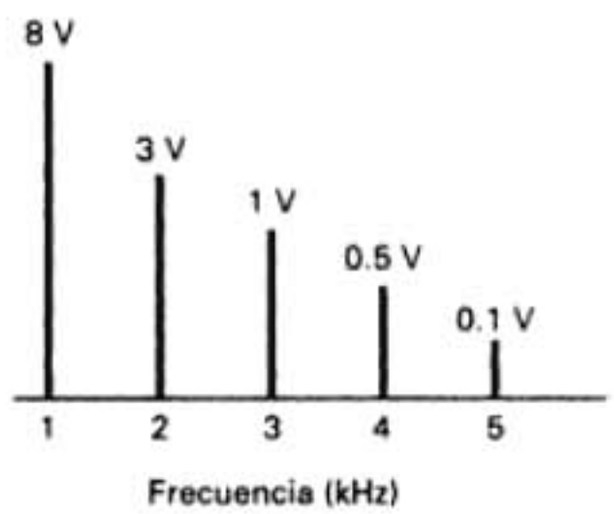
(b) Determine las amplitudes máximas de las cinco primeras armónicas.

(c) Trace la gráfica de la función (sen x)/x.

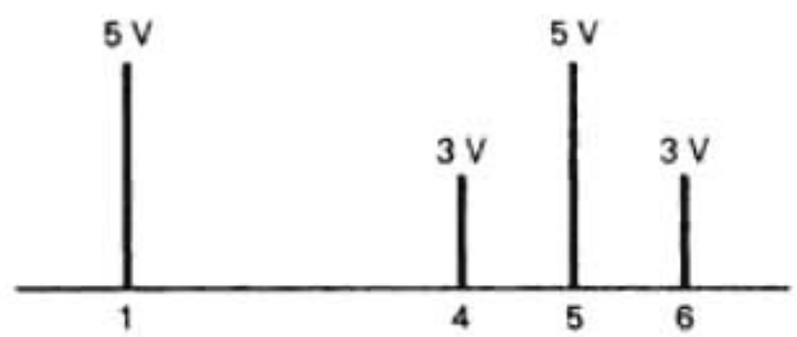
(d) Trace el espectro de frecuencias.



1-11. Describa el espectro que se ve a continuación. Determine la clase de amplificador (lineal o no lineal) y el contenido de frecuencias de la señal de entrada.



1-12. Repita el problema 1-11 con el siguiente espectro:



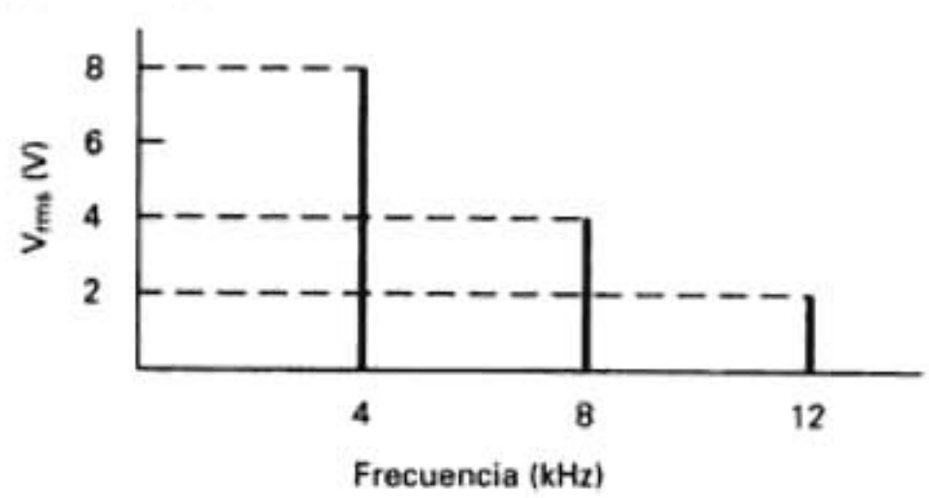
1-13. Un amplificador no lineal tiene dos frecuencias de entrada: 7 y 4 kHz.

(a) Determine las tres primeras armónicas presentes en la salida, para cada frecuencia.

(b) Determine las frecuencias de producto cruzado que se producen en la salida, para valores de m y n de 1 y 2.

(c) Trace el espectro de salida de las frecuencias armónicas y de productos cruzados determinadas en los pasos a) y b).

1-14. Determine la distorsión porcentual de segundo orden, tercer orden y armónica total, para el siguiente espectro de salida:



1-15. Determine el ancho de banda necesario para producir 8×10^{-17} watts de potencia de ruido térmico a la temperatura de 17° C.

1-16. Determine los voltajes de ruido térmico, para componentes que funcionen en las siguientes temperaturas, anchos de banda y resistencias equivalentes:
(a) $T = -50°$ C, $B = 50$ kHz y $R = 50\ \Omega$.
(b) $T = 100°$ C, $B = 10$ kHz y $R = 100\ \Omega$.
(c) $T = 50°$ C, $B = 500$ kHz y $R = 72\ \Omega$.

1-17. Determine la segunda, quinta y decimoquinta armónica para una onda repetitiva con frecuencia fundamental de 2.5 kHz.

1-18. Determine la distorsión de segunda y tercera armónica, y armónica total, para una banda repetitiva con amplitud de frecuencia fundamental de 10 V_{rms}, amplitud de segunda armónica de 0.2 V_{rms} y de tercera armónica de 0.1 V_{rms}.

1-19. Para un amplificador no lineal con frecuencias de ondas senoidales en la entrada de 3 y 5 kHz, determine las tres primeras armónicas presentes en la salida, para cada frecuencia de entrada, y las frecuencias de producto cruzado que se producen con valores de m y n igual a 1 y 2.

1-20. Determine las relaciones de potencia, en dB, con las siguientes potencias de entrada y salida:
(a) $P_{ent} = 0.001$ W, $P_{sal} = 0.01$ W.
(b) $P_{ent} = 0.25$ W, $P_{sal} = 0.5$ W.
(c) $P_{ent} = 1$ W, $P_{sal} = 0.5$ W.
(d) $P_{ent} = 0.001$ W, $P_{sal} = 0.001$ W.
(e) $P_{ent} = 0.04$ W, $P_{sal} = 0.16$ W.
(f) $P_{ent} = 0.002$ W, $P_{sal} = 0.0002$ W.
(g) $P_{ent} = 0.01$ W, $P_{sal} = 0.4$ W.

1-21. Determine las relaciones de voltaje, en dB, para los siguientes voltajes de entrada y de salida. Suponga valores iguales de resistencia de entrada y de salida.
(a) $v_{ent} = 0.001$ V, $v_{sal} = 0.01$ V.
(b) $v_{ent} = 0.1$ V, $v_{sal} = 2$ V.
(c) $v_{ent} = 0.5$ V, $v_{sal} = 0.25$ V.
(d) $v_{ent} = 1$ V, $v_{sal} = 4$ V.

1-22. Determine el factor de ruido general y la cifra de ruido general para tres amplificadores en cascada, con los siguientes parámetros:

$A_1 = 10$ dB
$A_2 = 10$ dB
$A_3 = 20$ dB
$NF_1 = 3$ dB
$NF_2 = 6$ dB
$NF_3 = 10$ dB

1-23. Determine el factor de ruido general y la cifra de ruido general para tres amplificadores en cascada, con los siguientes parámetros:

$A_1 = 3$ dB
$A_2 = 13$ dB
$A_3 = 10$ dB
$NF_1 = 10$ dB
$NF_2 = 6$ dB
$NF_3 = 10$ dB

1-24. Si el ancho de banda de un amplificador es $B = 20$ kHz, y su potencia total de ruido es $N = 2 \times 10^{-17}$ W, calcule la potencia total de ruido si el ancho de banda aumenta a 40 kHz. Calcúlela si el ancho de banda disminuye a 10 kHz.

1-25. Para un amplificador que funciona a una temperatura de 27° C, con ancho de banda de 20 kHz, determine:
(a) La potencia total de ruido, en watts y en dBm.
(b) El voltaje rms de ruido (V_N), con una resistencia interna de 50 Ω y un resistor de carga de 50 Ω.

1-26. **(a)** Determine la potencia de ruido, en watts y en dBm, de un amplificador que trabaja a una temperatura de 400° C con un ancho de banda de 1 MHz.

(b) Determine la disminución de potencia de ruido, en decibelios, si la temperatura bajara a 100° C

(c) Determine el aumento de potencia de ruido, en decibeles, si aumenta al doble el ancho de banda.

1-27. Determine la cifra de ruido para una temperatura equivalente de ruido de 1000° K; use 290° K como temperatura de referencia.

1-28. Determine la temperatura equivalente de ruido para una cifra de ruido de 10 dB.

1-29. Determine la cifra de ruido para un amplificador con relación de señal a ruido en la entrada igual a 100, y en la salida igual a 50.

1-30. Determine la cifra de ruido para un amplificador con relación de señal a ruido de 30 dB en la entrada y de 24 dB en la salida.

1-31. Calcule la relación de señal a ruido en la entrada para un amplificador con 16 dB de señal a ruido en la entrada y 5.4 dB de cifra de ruido.

1-32. Calcule la relación de señal a ruido en la salida de un amplificador con relación de señal a ruido de 23 dB en la entrada, y la cifra de ruido de 6.2 dB.

C A P Í

Imagen protegida por derechos de autor

Generación de señal

INTRODUCCIÓN

Los sistemas modernos de comunicaciones electrónicas tienen muchas aplicaciones que requieren formas de onda estables y repetitivas, tanto senoidales como no senoidales. En muchas de esas aplicaciones se requiere más de una frecuencia, y a menudo esas frecuencias se deben sincronizar entre sí. Por lo anterior, las partes esenciales de un sistema electrónico de comunicaciones son *generación de señal, sincronización de frecuencia* y *síntesis de frecuencia*. El objeto de este capítulo es presentar al lector el funcionamiento básico de los osciladores, los lazos de fase cerrada y los sintetizadores de frecuencia, e indicar cómo se usan para generar señales.

OSCILADORES

La definición de *oscilar* es fluctuar entre dos estados o condiciones. Por consiguiente, oscilar es vibrar o cambiar, y *oscilación* es el acto de fluctuar de uno a otro estado. Un *oscilador* es un dispositivo que produce oscilaciones, es decir, genera una forma de onda repetitiva. Hay muchas aplicaciones de los osciladores en las comunicaciones electrónicas, como las fuentes de portadora de alta frecuencia, fuentes piloto, relojes y circuitos de sincronización.

En las aplicaciones electrónicas, un oscilador es un dispositivo o circuito que produce oscilaciones eléctricas. Una oscilación eléctrica es un cambio repetitivo de voltaje o de corriente en una forma de onda. Si un oscilador es *autosostenido*, los cambios en la forma de onda son *continuos* y *repetitivos*; suceden con rapidez periódica. Un oscilador autosostenido también se llama oscilador autónomo o de funcionamiento libre. Los osciladores no autosostenidos requieren una señal externa de entrada, o *disparador*, para producir un cambio en la forma de onda de salida. Los osciladores que no son autosostenidos se llaman también osciladores *con disparo de inicio* u osciladores *monoestables*. El resto de este capítulo se dedica a explicar los osciladores autosostenidos, que no requieren más entrada externa que una fuente de voltaje de cd. En esencia, un oscilador convierte un voltaje de cd en la entrada a un voltaje de salida de ca. La forma de la onda de salida puede ser una senoidal, una onda cuadrada, una onda en diente de sierra o cualquier otra forma de onda, siempre y cuando se repita a intervalos periódicos.

Osciladores retroalimentados

Un *oscilador retroalimentado* es un amplificador con un *lazo de retroalimentación*, es decir, con un paso para que la energía se propague desde la salida y regrese a la entrada. Los osciladores autosostenidos (o monoestables) son osciladores retroalimentados. Una vez encendido, un oscilador retroalimentado genera una señal de salida de ca, de la cual se regresa una pequeña parte a la entrada, donde se amplifica. La señal amplificada de la entrada aparece en la salida, y el proceso se repite; se produce un proceso *regenerativo*, en el que la salida depende de la entrada y viceversa.

De acuerdo con el *criterio de Barkhausen*, para que un circuito retroalimentado sostenga oscilaciones, la ganancia neta de voltaje en torno al lazo de retroalimentación debe ser igual o mayor que la unidad, y el desplazamiento neto de fase en torno al lazo debe ser un múltiplo entero positivo de 360°.

Hay cuatro requisitos para que trabaje un oscilador retroalimentado: *amplificación, retroalimentación positiva, determinación de frecuencia* y una *fuente* de potencia eléctrica.

1. *Amplificación.* Un circuito oscilador debe tener cuando menos un dispositivo activo, y debe ser capaz de amplificar voltaje. De hecho, a veces se requiere que proporcione una ganancia infinita.
2. *Retroalimentación positiva.* Un circuito oscilador debe tener una trayectoria completa para que la señal de salida regrese a la entrada. La señal de retroalimentación debe ser *regenerativa*, y eso quiere decir que debe tener la fase correcta y la amplitud necesaria para sostener las oscilaciones. Si la fase es incorrecta, o si la amplitud es insuficiente, las oscilaciones cesan. Si la amplitud es excesiva, el amplificador se saturará. La *retroalimentación regenerativa* se llama también *retroalimentación positiva*, donde "positiva" simplemente indica que su fase ayuda en el proceso de oscilación, y no necesariamente indica una polaridad positiva (+) o negativa (−). La *retroalimentación degenerativa* se llama *retroalimentación negativa*, y proporciona una señal de retroalimentación que inhibe la producción de oscilaciones.
3. *Componentes que determinan la frecuencia.* Un oscilador debe tener componentes que determinen la frecuencia, como por ejemplo resistores, capacitores, inductores o cristales que permitan ajustar o cambiar la frecuencia de operación.
4. *Fuente de poder.* Un oscilador debe tener una fuente de energía eléctrica, que puede ser una fuente de cd.

La fig. 2-1 muestra un modelo eléctrico de un circuito *oscilador retroalimentado* (es decir, un amplificador de voltaje con retroalimentación regenerativa). Un oscilador retroalimentado es un circuito de *lazo cerrado* formado por un amplificador de voltaje con una *ganancia de voltaje de lazo abierto* (A_{ol}), una trayectoria regenerativa que determina la frecuencia con una *relación de retroalimentación* (β) y un circuito sumador o restador. La ganancia de voltaje de lazo abierto es la ganancia de voltaje del amplificador con la trayectoria de retroalimentación abierta. La *ganancia de voltaje de lazo cerrado* (A_{cl}) es la ganancia general de voltaje del circuito total, con el lazo de retroalimentación cerrado, y siempre es menor que la de lazo abierto. La relación de retroalimentación no es más que la función de transferencia de la red de retroalimentación, es decir, la relación de su voltaje de salida a su voltaje de entrada. Para una red de retroalimentación pasiva, la relación de retroalimentación siempre es menor que 1.

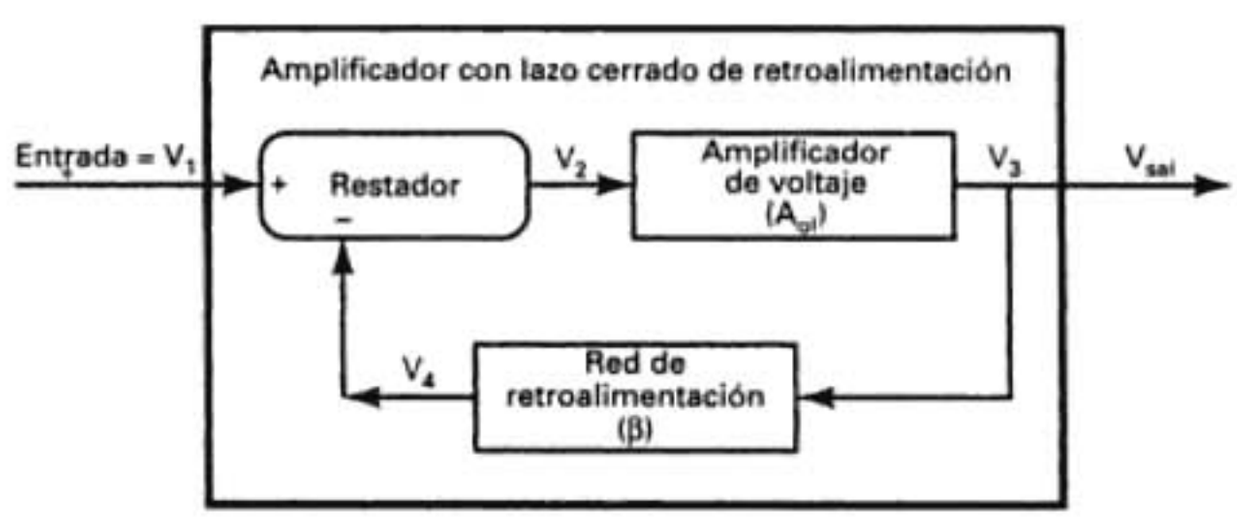


FIGURA 2-1 Modelo de un amplificador con retroalimentación

Las siguientes relaciones matemáticas se derivan de la fig. 2–1

$$\frac{V_{sal}}{V_{ent}} = \frac{V_3}{V_1}$$

$$V_2 = V_1 - V_4$$

$$V_3 = A_{ol}V_2$$

$$A_{ol} = \frac{V_3}{V_2}$$

$$V_4 = \beta V_3$$

$$\beta = \frac{V_4}{V_3}$$

en donde V_1 = voltaje de entrada externa
V_2 = voltaje de entrada al amplificador
V_3 = voltaje de salida
V_4 = voltaje de retroalimentación
A_{ol} = ganancia de voltaje de lazo abierto
β = relación de retroalimentación del lazo de retroalimentación

Se sustituye V_4 de la penúltima ecuación en la segunda $V_2 = V_1 - \beta V_3$

Así, $V_3 = (V_1 - \beta V_3)A_{ol}$

y $V_3 = V_1A_{ol} - V_3\beta A_{ol}$

Se reagrupa y factoriza esta ecuación, como sigue

$$V_3 + V_3\beta A_{ol} = V_1A_{ol}$$

Así, $V_3(1 + \beta A_{ol}) = V_1A_{ol}$

y

$$\frac{V_{sal}}{V_{ent}} = \frac{V_3}{V_1} = \frac{A_{ol}}{1 + \beta A_{ol}} = A_{cl} \qquad (2\text{-}1)$$

siendo A_{cl} la ganancia de voltaje de lazo cerrado.

La fórmula acostumbrada para la ganancia de voltaje de lazo cerrado de un amplificador retroalimentado es $A_{ol}/(1 + \beta A_{ol})$. Si en cualquier frecuencia βA_{ol} baja a -1, el denominador de la ecuación 2-1 baja a cero y V_{sal}/V_{ent} es infinito. Cuando esto sucede, el circuito oscilará y se podrá quitar la entrada externa.

Para que se produzcan las oscilaciones autosostenidas, un circuito debe cumplir con los cuatro requisitos básicos que se describieron anteriormente; también debe cumplir con el criterio de la ecuación 2-1 y debe apegarse al modelo básico de circuito retroalimentado que se muestra en la fig. 2-1. Aunque la acción de oscilador se puede lograr en muchas formas distintas, las configuraciones más comunes son las redes *RC* de desplazamiento de fase, los circuitos tanque *LC* (o circuitos *LC* resonantes), los cristales de cuarzo y los chips de circuito integrado. La clase de oscilación que se use en determinada aplicación depende de los siguientes criterios:

1. Frecuencia necesaria de operación.
2. Estabilidad requerida de frecuencia.
3. Operación con frecuencia variable o fija.
4. Requisitos o limitaciones de distorsión.
5. Potencia necesaria en la salida.
6. Tamaño físico.
7. Aplicación (por ejemplo, digital o analógica).
8. Costo.
9. Fiabilidad y durabilidad.
10. Exactitud necesaria.

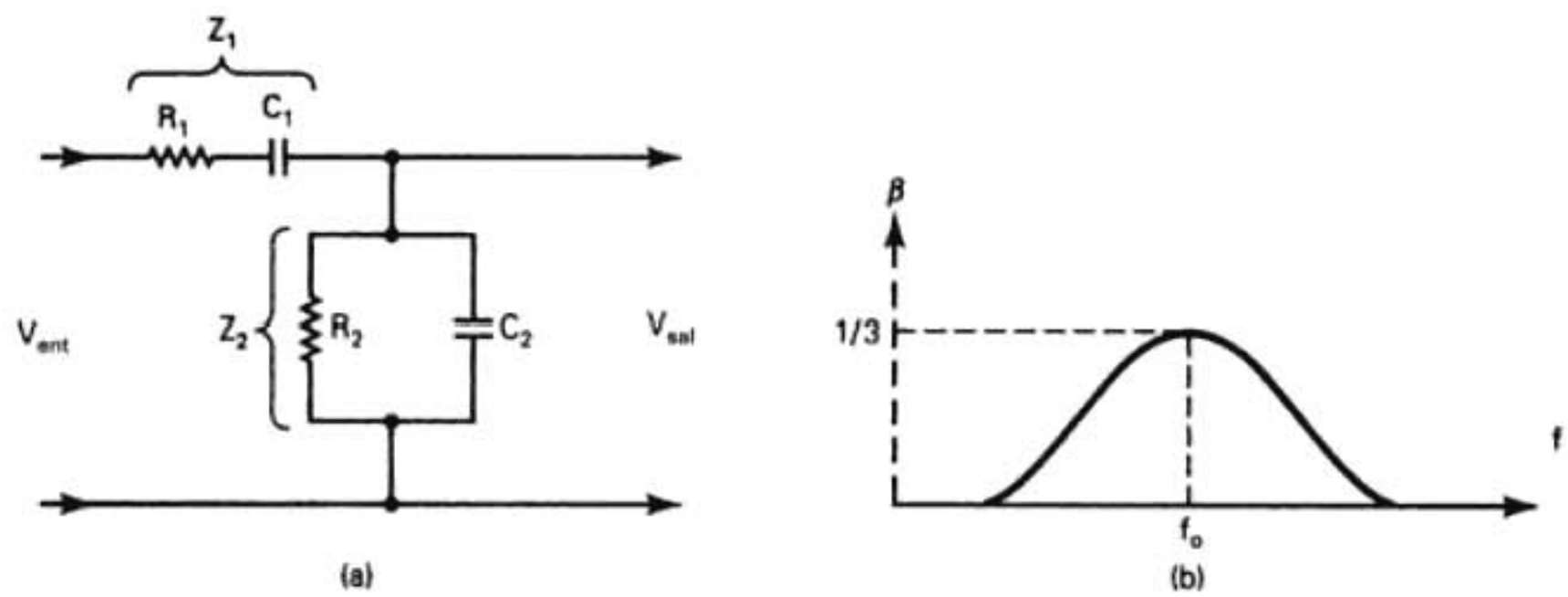


FIGURA 2-2 Red de adelanto-retraso: (a) configuración del circuito; (b) curva de transferencia de entrada a salida (β)

Osciladores no sintonizados

El oscilador de puente de Wien es un oscilador *RC* no sintonizado de corrimiento de fase, que usa retroalimentación tanto positiva como negativa. Es un circuito oscilador relativamente estable, de baja frecuencia, que se sintoniza con facilidad, y se suele usar en generadores de señal para producir frecuencias de 5 Hz a 1 MHz. Este circuito oscilador de puente de Wien fue el que utilizó la compañía Hewlett Packard en su diseño original de generador de señales.

La fig. 2-2a muestra una red sencilla de adelanto-retraso. A la frecuencia de oscilación (f_o), $R = X_C$ y la señal sufre un desplazamiento de fase de $-45°$ a través de Z_1 y otro de $+45°$ a través de Z_2. En consecuencia, en f_o, el desplazamiento total de fase a través de la red de adelanto-retraso es exactamente 0°. A frecuencias menores que la frecuencia de oscilación, el desplazamiento de fase a través de la red es de adelanto, y para frecuencias superiores el desplazamiento se retrasa. A frecuencias extremadamente bajas C_1 funciona como circuito abierto, y no hay salida. A frecuencias extremadamente altas, C_2 funciona como un cortocircuito y no hay salida.

Una red de adelanto-retraso es un divisor reactivo de voltaje, en el que el voltaje de entrada se divide entre Z_1 (la combinación de R_1 y C_1 en serie) y Z_2 (la combinación de R_2 y C_2 en paralelo). Por consiguiente, la red de adelanto-atraso es selectiva de frecuencia, y el voltaje de salida es máximo en f_o. La función de transferencia (β) para la red de retroalimentación es igual a $Z_2/(Z_1 + Z_2)$ y su máximo es igual a 1/3 en f_o. La fig. 2-2b muestra una gráfica de β en función de la frecuencia cuando $R_1 = R_2$ y $C_1 = C_2$. Así, f_o queda determinada por la siguiente ecuación

$$f_o = \frac{1}{2\pi RC}$$

en la que $R = R_1 = R_2$
$C = C_1 = C_2$

La fig. 2-3 muestra un oscilador con puente de Wien. La red de adelanto-atraso y el divisor resistivo de voltaje forman un puente de Wien (de donde procede el nombre de *oscilador con puente de Wien*). Cuando está balanceado el puente, la diferencia de voltaje es igual a cero. El divisor de voltaje proporciona una retroalimentación negativa o degenerativa, que compensa la retroalimentación positiva o regenerativa de la red de adelanto-atraso. La relación de los resistores en el divisor de voltaje es 2:1, y eso establece la ganancia de voltaje no inversor del amplificador A_1, que es igual a $R_f/R_i + 1 = 3$. Así, en f_o, la señal a la salida de A_1 se reduce en un factor de 3 al pasar por la red de adelanto-retraso ($\beta = 1/3$), y después se amplifica con factor de 3 en el amplificador A_1. Por lo anterior, en f_o, la ganancia de voltaje del lazo es igual a $A_{ol}\beta$ o sea $3 \times 1/3 = 1$.

Para compensar los desequilibrios del puente y las variaciones en los valores de los componentes debidos al calor, al circuito se le agrega un CAG o *control automático de ganancia* (AGC, de *automatic gain control*). Una forma sencilla de proporcionar control automático de ganancia es reemplazar R_i en la fig. 2-3 por un dispositivo con resistencia variable, como por ejemplo un FET. La resistencia del FET se hace inversamente proporcional a V_{sal}. El circuito se diseña de tal

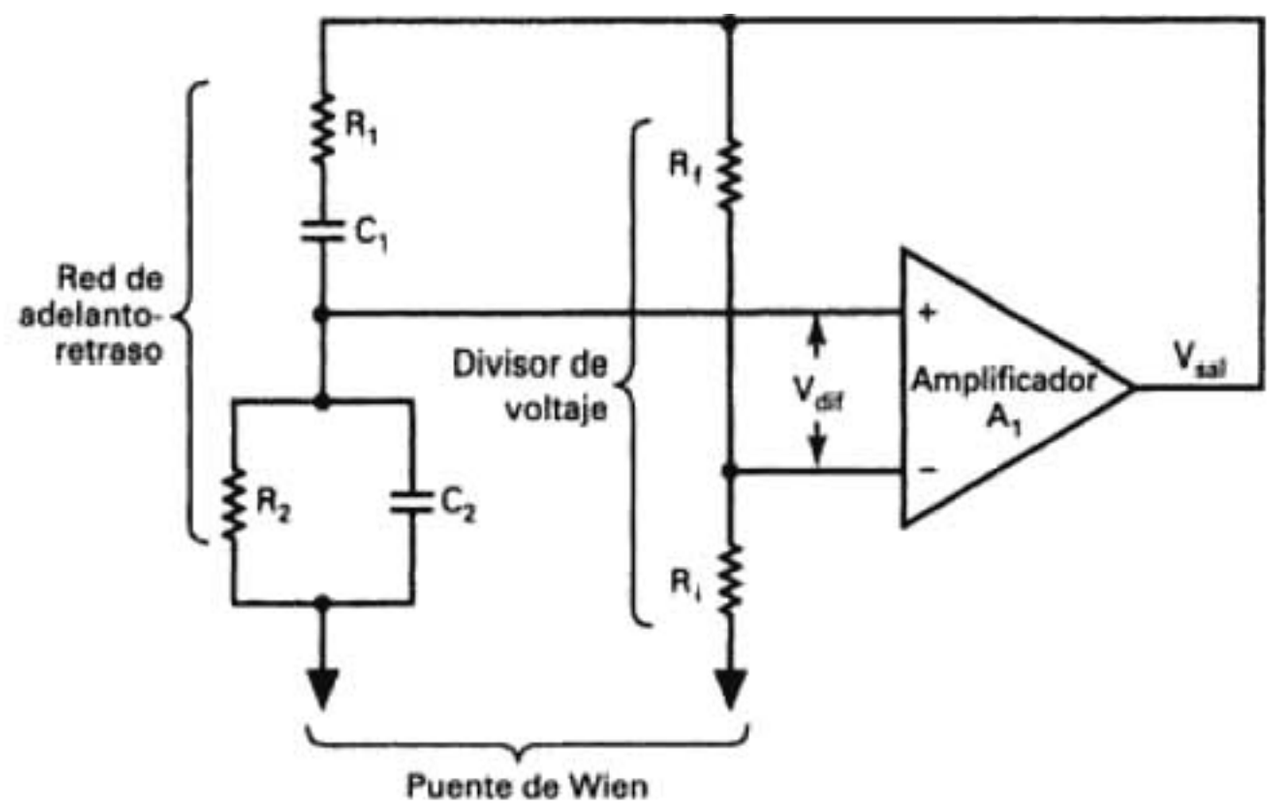


FIGURA 2-3 Oscilador de puente de Wien

modo que cuando V_{sal} aumenta de amplitud, la resistencia del FET aumenta, y cuando disminuye la amplitud de V_{sal}, la resistencia del FET disminuye. Por lo anterior, la ganancia de voltaje del amplificador compensa en forma automática los cambios de amplitud de la señal de salida.

El funcionamiento del circuito de la fig. 2-3 es el siguiente: en el encendido inicial aparece ruido (en todas las frecuencias) en V_{sal}, que se retroalimenta por la red de adelanto-retraso. Sólo pasa ruido de f_o por esa red con un desplazamiento de fase de 0°, y con relación de transferencia de 1/3. En consecuencia, sólo se retroalimenta una sola frecuencia (f_o) en fase, sufre una ganancia de voltaje de 1 en el lazo, y produce oscilaciones autosostenidas.

Osciladores sintonizados

Los osciladores *LC* son circuitos osciladores que usan *circuitos tanque LC* para establecer la frecuencia. El funcionamiento de un circuito tanque implica intercambio de energía entre *cinética* y *potencial*. La fig. 2-4 ilustra el funcionamiento de un circuito tanque *LC*. Como se ve en la fig. 2-4a, una vez que se inyecta corriente en el circuito (momento t_1), se intercambia energía entre el inductor y el capacitor, y se produce un voltaje correspondiente de salida (tiempos t_2 a t_4). La forma de onda de voltaje de salida se muestra en la fig. 2-4b. La frecuencia de funcionamiento de un circuito tanque *LC* no es más que la frecuencia de resonancia de la red *LC* en paralelo, y el ancho de banda es una función de la Q del circuito. La frecuencia de resonancia de un circuito tanque *LC* con $Q \geq 10$ se aproxima bastante con la ecuación

$$f_o = \frac{1}{2\pi\sqrt{(LC)}} \tag{2-2}$$

Entre los osciladores *LC* están el de Hartley y el de Colpitts.

Oscilador de Hartley. La fig. 2-5a muestra el esquema de un *oscilador de Hartley*. El amplificador transistorizado (Q_1) proporciona la amplificación necesaria para tener una ganancia de unidad en el voltaje de lazo a la frecuencia de resonancia. El capacitor de acoplamiento (C_C) proporciona la trayectoria de la retroalimentación regenerativa. Los componentes L_{1a}, L_{1b} y C_1 son los que determinan la frecuencia, y V_{CC} es el voltaje de suministro.

La fig. 2-5b muestra el circuito equivalente de cd para el oscilador de Hartley. C_C es un capacitor de bloqueo que aísla el voltaje de polarización de cd y evita que se ponga en corto a tierra a través de L_{1b}. C_2 también es un capacitor de bloqueo que evita que el voltaje de suministro del colector pase a tierra a través de L_{1a}. La *bobina de radiofrecuencia (o mejor conocida como* RFC, de *radio-frequency choke)* es un corto circuito para cd.

En la fig. 2-5c se ve un circuito equivalente de ca para el oscilador de Hartley. C_C es un capacitor de acoplamiento para ca y proporciona una trayectoria para la retroalimentación regenerativa del circuito tanque a la base de Q_1. C_2 acopla las señales de ca del colector de Q_1 con el circuito tanque. La bobina de radiofrecuencia funciona como un circuito abierto en ca y en consecuencia aísla la fuente de poder de cd y las oscilaciones de ca.

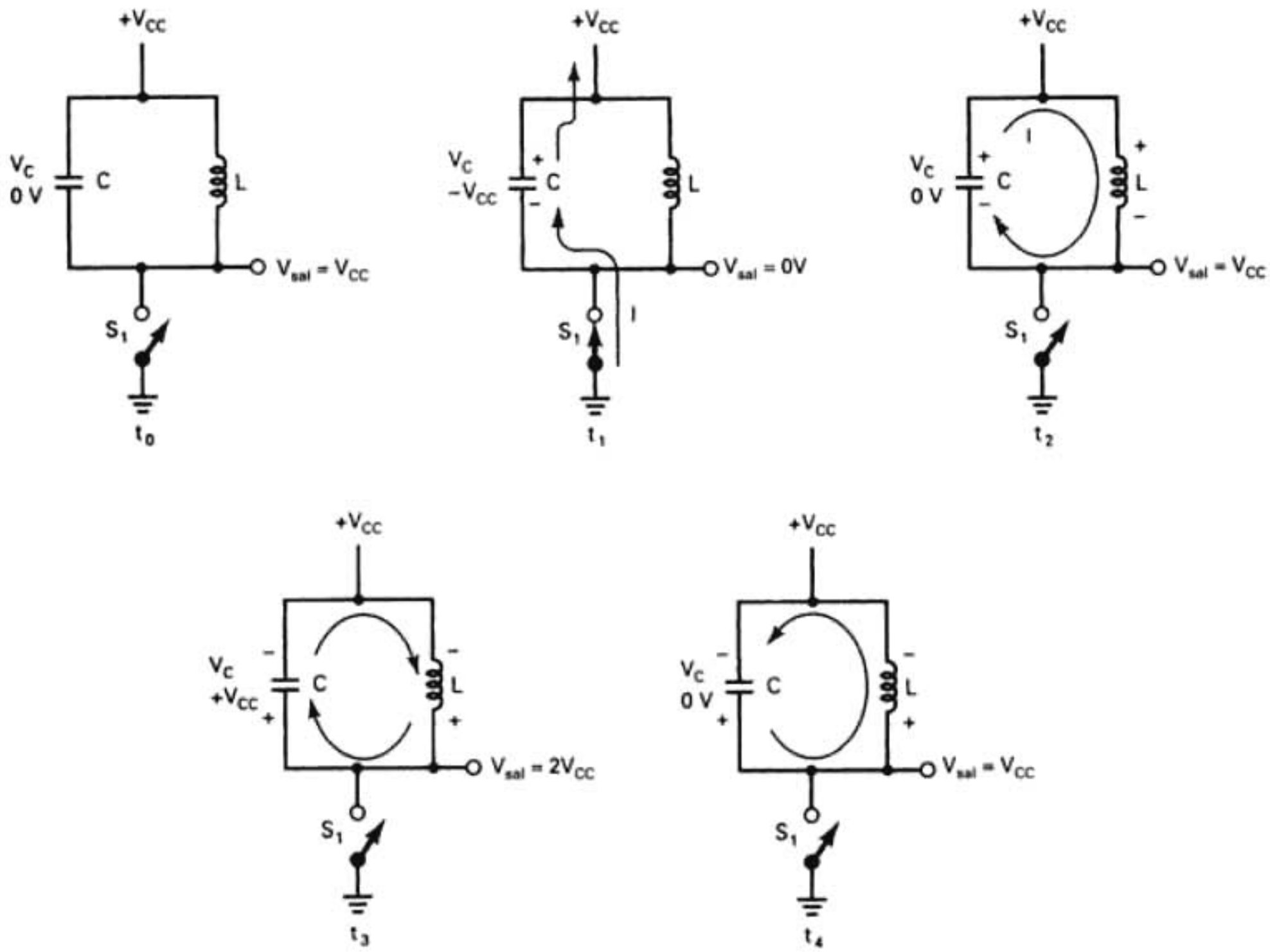


FIGURA 2-4 Circuito tanque *LC*: (a) acción de oscilador y efecto de volante; (b) forma de onda de salida

El oscilador de Hartley funciona como sigue: en el encendido inicial aparece una multitud de frecuencias en el colector de Q_1 y se acoplan al circuito tanque a través de C_2. El ruido inicial proporciona la energía necesaria para cargar a C_1. Una vez que C_1 se carga parcialmente, comienza la acción del oscilador. El circuito tanque sólo oscila con eficiencia a su frecuencia de resonancia. Una parte del voltaje oscilatorio del circuito tanque va a través de L_{1b} y se retroalimenta a la base de Q_1, donde se amplifica. La señal amplificada aparece desfasada 180° en el colector, respecto a la señal de base. Se hace otro desplazamiento de fase de 180° más a través de L_1; en consecuencia, la señal que regresa a la base de Q_1 está amplificada y con su fase desplazada 360°. Por lo anterior, el circuito es regenerativo y sostiene oscilaciones sin señal externa de entrada.

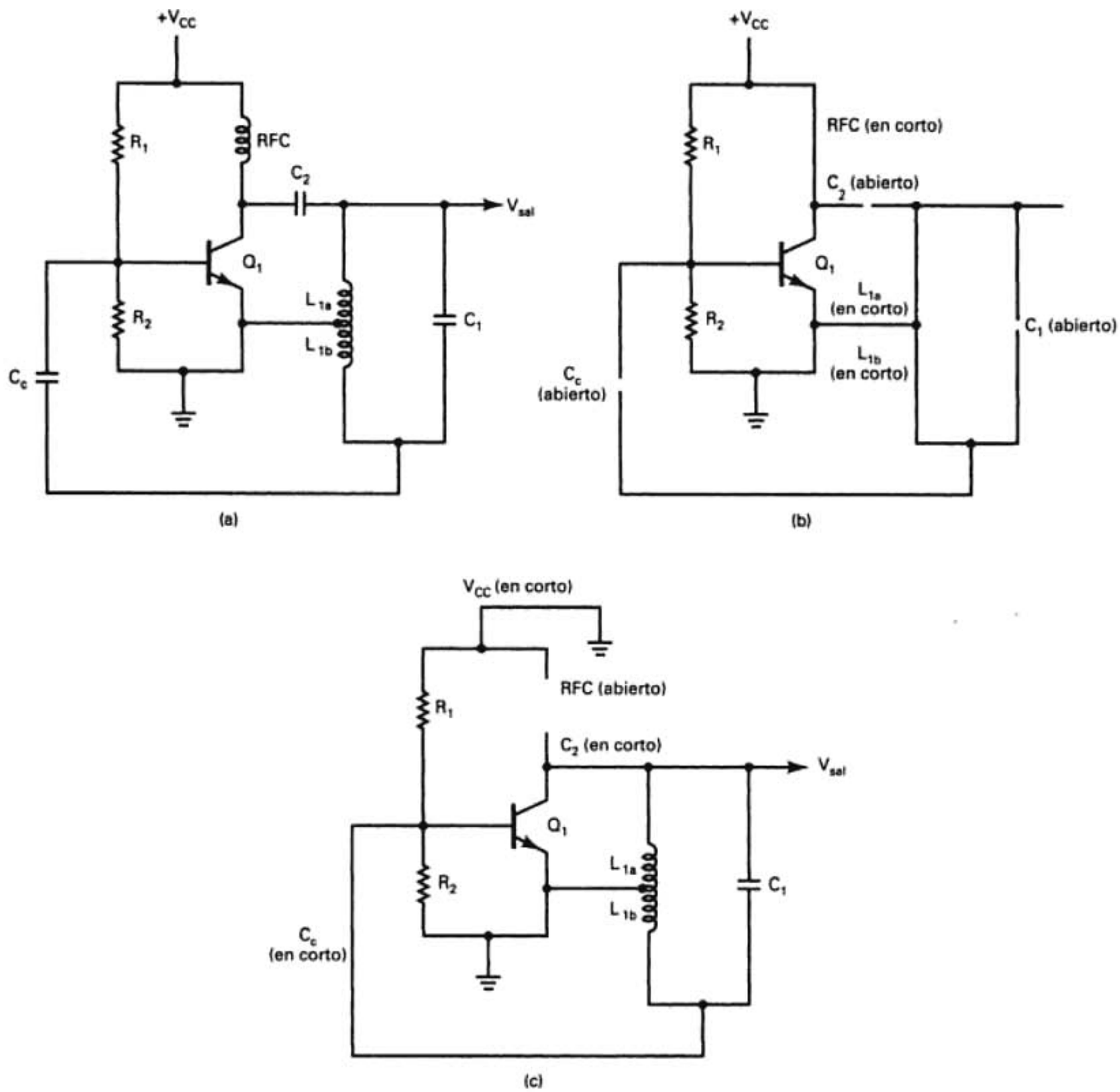


FIGURA 2-5 Oscilador de Hartley: (a) esquema; (b) circuito equivalente de cd; (c) circuito equivalente de ca

La parte de la energía oscilatoria que se retroalimenta a la base de Q_1 se determina con la relación de L_{1b} a la inductancia total, $L_{1a} + L_{1b}$. Si la energía que se retroalimenta es insuficiente, se amortiguan las oscilaciones. Si se retroalimenta demasiada energía, el transistor se satura. Por consiguiente, la posición del contacto móvil en L_1 se ajusta hasta que la cantidad de energía retroalimentada sea exactamente la que se requiere para que la ganancia de voltaje del lazo sea la unidad, y que continúen las oscilaciones.

La frecuencia obtenida en el oscilador de Hartley se aproxima con la siguiente fórmula

$$f_o = \frac{1}{2\pi\sqrt{(LC)}} \tag{2-3}$$

donde $L = L_{1a} + L_{1b}$
$C = C_1$

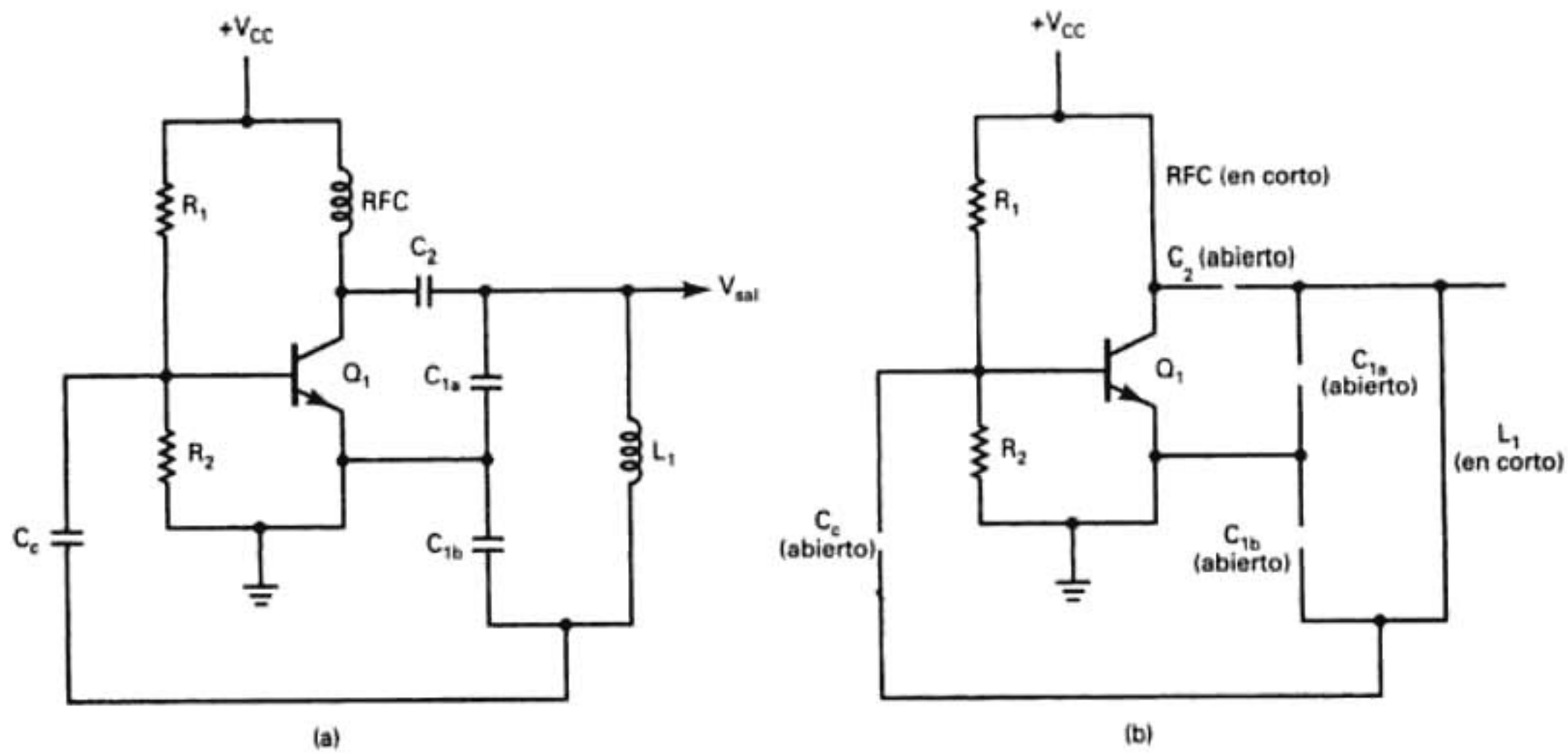


FIGURA 2-6 Oscilador de Colpitts: (a) esquema; (b) circuito equivalente de cd; (c) circuito equivalente de ca

Oscilador de Colpitts. La fig. 2-6a muestra el diagrama de un *oscilador de Colpitts*. El funcionamiento de este oscilador es muy parecido al oscilador de Hartley, con la excepción de que se usa un divisor capacitivo en lugar de una bobina con derivación. El transistor Q_1 proporciona la amplificación, C_C proporciona la trayectoria de retroalimentación regenerativa, L_1, C_{1a} y C_{1b} son los componentes para determinar la frecuencia, y V_{CC} es el voltaje de suministro de cd.

La fig. 2-6b muestra el circuito equivalente de cd del oscilador de Colpitts. C_2 es un capacitor de bloqueo, que evita que aparezca en la salida el voltaje de suministro de colector. La RFC de nuevo es un corto circuito en cd.

La fig. 2-6c muestra el circuito equivalente de ca del oscilador de Colpitts. C_C es un capacitor de acoplamiento para ca, y proporciona la trayectoria de regreso de la retroalimentación regenerativa desde el circuito tanque a la base de Q_1. La RFC es un circuito abierto en ca, y desacopla las oscilaciones de la fuente de poder de cd.

El funcionamiento del oscilador de Colpitts es casi idéntico al del oscilador de Hartley. En el encendido inicial aparece ruido en el colector de Q_1, y suministra la energía al circuito tanque haciéndolo comenzar a oscilar. Los capacitores C_{1a} y C_{1b} forman un divisor de voltaje. La caída de voltaje a través de C_{1b} se retroalimenta a la base de Q_1 a través de C_C. Hay un desplazamiento de fase de 180°, de la base al colector de Q_1, y otro desplazamiento de fase de 180° más a través de C_1. En consecuencia, el desplazamiento total de fase es 360°, y la señal de retroalimentación es regenerativa. La relación de C_{1a} a $C_{1a} + C_{1b}$ determina la amplitud de la señal retroalimentada.

La frecuencia del oscilador de Colpitts se aproxima bastante bien con la ecuación siguiente

$$f_o = \frac{1}{2\pi\sqrt{(LC)}} \tag{2-4}$$

en la que $L = L_1$

$$C = \frac{(C_{1a}C_{1b})}{(C_{1a} + C_{1b})}$$

Oscilador de Clapp. Un circuito oscilador de Clapp es idéntico al oscilador de Colpitts de la fig. 2-6a, con la excepción de la adición de un pequeño capacitor C_S en serie con L_1. La capacitancia de C_S se hace menor que la de C_{1a} o de C_{1b}, y así se obtiene una reactancia grande. En consecuencia, C_S tiene el máximo efecto en la determinación de la frecuencia del circuito tanque. La ventaja de un oscilador de Clapp es que se pueden seleccionar a C_{1a} y C_{1b} para obtener una relación óptima de retroalimentación, mientras que C_S puede ser variable y usarse para establecer la frecuencia de oscilación. En algunas aplicaciones, C_S tiene un coeficiente negativo de temperatura, con lo que se mejora la estabilidad de frecuencia de este oscilador.

Estabilidad de la frecuencia

La *estabilidad de frecuencia* es la capacidad de un oscilador para permanecer en una frecuencia fija, y es de primordial importancia en los sistemas de comunicaciones. Con frecuencia, se especifica la estabilidad de frecuencia como de corto o de largo plazo. La *estabilidad a corto plazo* se afecta principalmente debido a fluctuaciones en los voltajes de operación de cd, mientras que la *estabilidad a largo plazo* es una función del envejecimiento de los componentes y de cambios en la temperatura y la humedad del ambiente. En los osciladores de circuito tanque *LC* y de desplazamiento de fase *RC* que se han mencionado, la estabilidad de frecuencia es inadecuada para la mayoría de las aplicaciones en radiocomunicaciones, porque los osciladores de desplazamiento de fase *RC* son susceptibles a variaciones a corto y a largo plazo. Además, los factores *Q* de los circuitos tanque *LC* son relativamente bajos y permiten que el circuito tanque resonante oscile dentro de un amplio rango de frecuencias.

En general, la estabilidad de frecuencia se caracteriza como un porcentaje de cambio en la frecuencia (tolerancia) respecto al valor deseado. Por ejemplo, un oscilador que funcione a 100 kHz con estabilidad de ±5% trabajará a una frecuencia de 100 ± 5 kHz, o sea, entre 95 y 105 kHz. Las estaciones comerciales de FM deben mantener las frecuencias de sus portadoras dentro de ±2 kHz respecto a su frecuencia asignada, que equivale aproximadamente a una tolerancia de 0.002%. En la difusión comercial en AM, el desplazamiento máximo admisible de la frecuencia portadora sólo es de ±20 Hz.

Son varios los factores que afectan la estabilidad de un oscilador. Los más obvios son los que afectan en forma directa el valor de los componentes que determinan la frecuencia. Se incluyen los cambios de inductancia, capacitancia y resistencia causados por variaciones ambientales de temperatura y humedad, y cambios en el punto de reposo de los transistores normales y de efecto de campo. También las fluctuaciones de las fuentes de poder de cd afectan la estabilidad. La estabilidad de la frecuencia de los osciladores de *RC* o de *LC* se puede mejorar mucho regulando el suministro de cd y minimizando las variaciones ambientales. También se pueden usar componentes especiales, independientes de la temperatura.

La FCC ha establecido reglamentos estrictos acerca de las tolerancias de las portadoras de radiofrecuencia. Siempre que se usa la atmósfera (propagación de radio en espacio libre)

como medio de transmisión, es posible que las transmisiones de una fuente puedan interferir con las de otras fuentes, si se traslapan sus bandas de frecuencia de emisión o de transmisión. Por consiguiente, es importante que todas las fuentes mantengan su frecuencia de operación dentro de una tolerancia especificada.

Osciladores de cristal

Los *osciladores de cristal* son circuitos osciladores de retroalimentación en los que se sustituye el circuito tanque *LC* con un cristal, como componente para determinar la frecuencia. El cristal funciona en forma parecida al tanque *LC*, pero tiene varias ventajas inherentes. A veces, a los cristales se les llama resonadores de cristal, y pueden producir frecuencias precisas y estables para contadores de frecuencia, sistemas de navegación electrónica, radiotransmisores y radiorreceptores, televisiones, grabadoras de video (VCR), relojes para sistemas de cómputo y muchas otras aplicaciones, demasiado numerosas para poder mencionarlas.

La cristalografía es el estudio de la forma, estructura, propiedades y clasificaciones de los cristales. Esta ciencia estudia las redes y los enlaces cristalinos, y el comportamiento de rebanadas de material cristalino, cortadas con diversos ángulos con respecto a los ejes del cristal. Las propiedades mecánicas de las redes cristalinas les permiten mostrar el efecto piezoeléctrico. Las rebanadas de cristales que se han cortado y pulido vibran cuando se aplican voltajes alternos entre sus caras. Las dimensiones físicas de un cristal, en especial su espesor y el lugar y la forma de cortarlo, determinan sus propiedades eléctricas y mecánicas.

Efecto piezoeléctrico. En forma sencilla, el *efecto piezoeléctrico* se presenta cuando se aplican esfuerzos mecánicos oscilatorios a través de una *estructura de red cristalina*, y generan oscilaciones eléctricas, y viceversa. El esfuerzo puede tener la forma de compresión, tensión, torsión o cortante. Si el esfuerzo se aplica en forma periódica, el voltaje de salida es alterno. Al revés, cuando se aplica un voltaje alterno a través de un cristal con la frecuencia natural de resonancia, o cerca de ella, el cristal comenzará a oscilar mecánicamente. A este proceso se le llama *excitación* de *vibraciones mecánicas* en un cristal. Las vibraciones mecánicas se llaman *ondas acústicas en la masa* del cristal (BAW, de *bulk acoustic waves*), y son directamente proporcionales a la amplitud del voltaje aplicado.

Hay varias sustancias cristalinas naturales que tienen propiedades piezoeléctricas: *cuarzo*, *sal de la Rochela* (tartrato de potasio y sodio) y la *turmalina*, así como varias sustancias artificiales, como ADP, EDT y DKT. El efecto piezoeléctrico es más pronunciado en la sal de la Rochela, que por eso es la sustancia que se suele usar en los micrófonos de cristal. Sin embargo, el cuarzo sintético se usa más para control de frecuencia en osciladores, debido a su *permanencia*, su *bajo coeficiente de temperatura* y su alta *Q mecánica*.

Cortes de cristal. En la naturaleza, los cristales completos de cuarzo tienen una sección transversal hexagonal, con extremos en punta, como se ve en la fig. 2-7a. Con un cristal se asocian tres conjuntos de ejes: *óptico, eléctrico* y *mecánico*. El eje longitudinal que une a las puntas en los extremos del cristal se llama *eje óptico* o *eje Z*. Los voltajes eléctricos aplicados en el eje óptico no producen el efecto piezoeléctrico. El *eje eléctrico*, o *eje X*, pasa en dirección diagonal, a través de las esquinas opuestas del hexágono. El eje que es perpendicular a las caras es el *eje mecánico* o *eje Y*. La fig. 2-7b muestra los ejes y el comportamiento básico de un cristal de cuarzo.

Si en un cristal se corta una sección delgada de tal modo que los lados planos sean perpendiculares a un eje eléctrico, los esfuerzos mecánicos a lo largo del eje *Y* producirán cargas eléctricas sobre las caras planas. Cuando el esfuerzo cambie de compresión a tensión y viceversa, se invierte la polaridad de la carga. Al revés, si se deposita una carga eléctrica alterna sobre las caras planas, se produce una vibración mecánica a lo largo del eje *Y*. Éste es el efecto piezoeléctrico, que también se produce cuando se aplican fuerzas mecánicas a través de las caras de un cristal cortado con sus caras planas perpendiculares al eje *Y*. Cuando se corta una oblea de cristal en dirección paralela al eje *Z* con sus caras perpendiculares al eje *X*, se obtiene lo que se llama un cristal de corte *X*. Cuando las caras son perpendiculares al eje *Y*, se obtiene el cristal de corte *Y*. Se pueden obtener diversos cortes girando el plano de corte respecto a uno o más

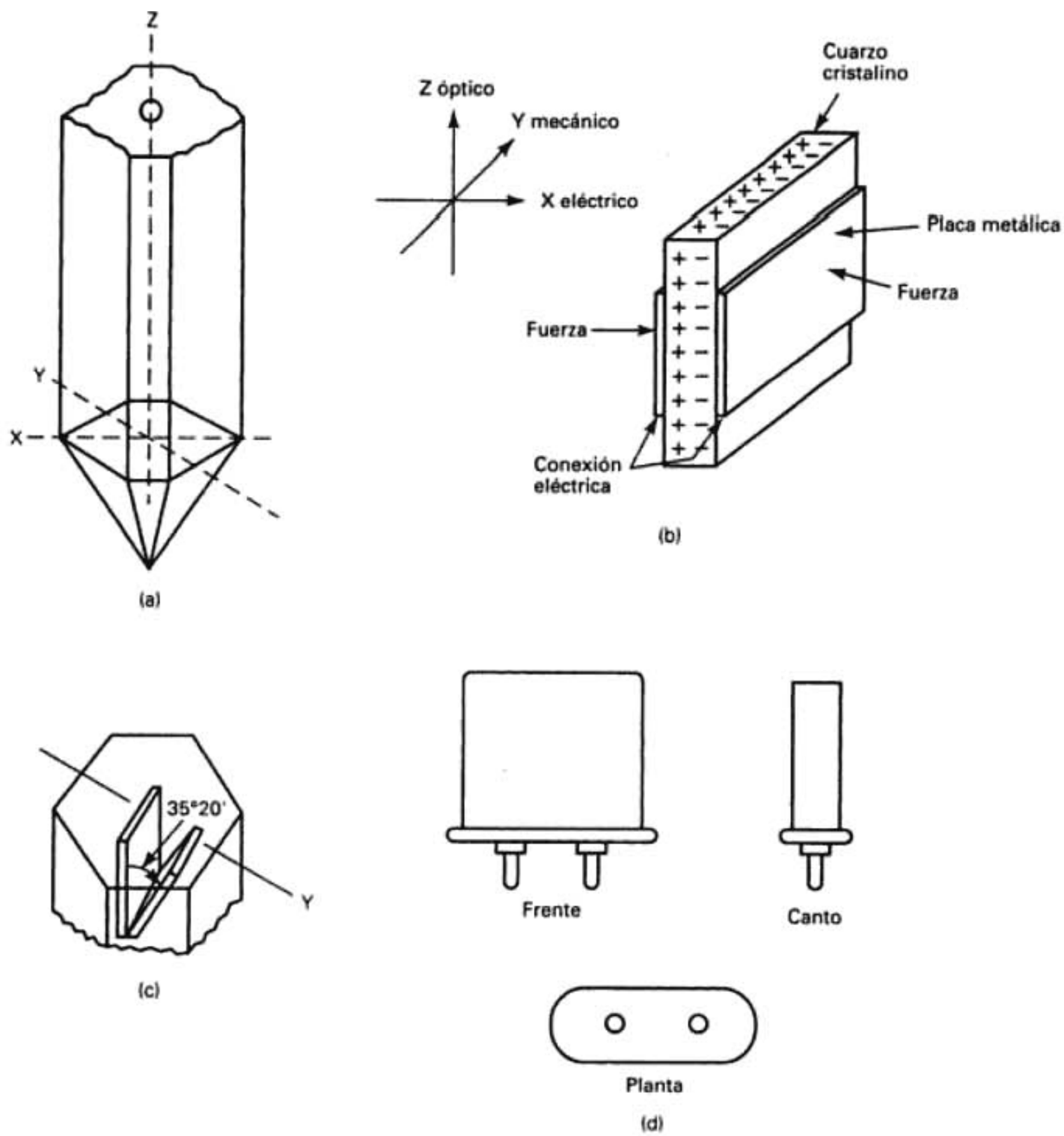


FIGURA 2-7 Cristal de cuarzo: (a) estructura cristalina básica; (b) ejes cristalográficos; (c) cortes del cristal; (d) montura del cristal

ejes. Si el corte *Y* se hace en ángulo de 35° 20′ respecto al eje vertical (fig. 2-7c), se obtiene un corte *AT*. Hay otros tipos de cortes de cristal, que incluyen los cortes *BT, CT, DT, ET, AC, GT, MT, NT* y *JT*. El corte *AT* es el más común en los resonadores de alta y muy alta frecuencia, de cristal. El tipo, longitud y espesor de un corte, y el modo de vibración, determinan la frecuencia natural de resonancia del cristal. Las frecuencias de resonancia de los cristales cortados *AT* van desde unos 800 kHz hasta unos 30 MHz. Los cortes *CT* y *DT* tienen esfuerzo cortante de baja frecuencia, y se usan más en el intervalo de 100 a 500 kHz. El corte *MT* vibra longitudinalmente, y se usa en el intervalo de 50 a 100 kHz, y el corte *NT* tiene un intervalo útil menor que 50 kHz.

Las *obleas* de cristal se suelen montar en *portacristales*, que son los conjuntos de montura y caja. Una *unidad de cristal* es el portacristal con el cristal mismo. La fig. 2-7d muestra una montura normal para un cristal. Como la estabilidad de un cristal depende algo de la temperatura, una unidad de cristal se puede montar en una estufa u horno, para mantener constante la temperatura de funcionamiento.

La relación entre la frecuencia de operación y el espesor de un cristal es la siguiente

$$h = \frac{65.5}{f_n}$$

en donde h = espesor del cristal (pulgadas)
f_n = frecuencia natural de resonancia (hertz)

Esta fórmula indica que, para oscilaciones de alta frecuencia, la oblea de cuarzo debe ser muy delgada. Esto hace difícil la manufactura de osciladores de cristal con frecuencias fundamentales mayores que unos 30 MHz, porque se vuelven tan delgados que son excepcionalmente frágiles, y el corte y el pulido convencionales sólo se pueden hacer con altísimos costos. Se puede aliviar este problema recurriendo al ataque químico, para obtener rebanadas más delgadas. En este proceso es posible obtener cristales con frecuencias fundamentales hasta de 350 MHz.

Oscilador de cristal en sobretono (armónicas). Como se dijo arriba, para aumentar la frecuencia de vibración de un cristal de cuarzo, la oblea de cuarzo se hace más delgada. Esto impone un límite físico obvio: mientras más delgada es la oblea, es más susceptible a dañarse y se hace menos útil. Aunque el límite práctico para los osciladores de cristal en modo fundamental es de unos 30 MHz, es posible hacer trabajar el cristal en modo de sobretonos o armónicas. En este modo se pueden tener vibraciones relacionadas armónicamente en forma simultánea con la vibración fundamental. En el modo de sobretono, el oscilador se sintoniza para trabajar en la tercera, quinta, séptima o hasta en la novena armónica de la frecuencia fundamental del cristal. Las armónicas se llaman sobretonos, porque no son verdaderas armónicas. Los fabricantes pueden procesar cristales de tal modo que se refuerce un sobretono más que los demás. Al usar el modo de sobretono aumenta el límite útil de los osciladores normales de cristal, hasta unos 200 MHz.

Coeficiente de temperatura. La frecuencia natural de resonancia de un cristal está influida un poco por su temperatura de funcionamiento. La relación de la magnitud del cambio de frecuencia (Δf) entre el cambio de temperatura (ΔC) se expresa en cambio de hertz por megahertz de frecuencia de operación del cristal, y por grado Celsius (Hz/MHz/°C). El cambio fraccionario de frecuencia se expresa a menudo en partes por millón por °C. Por ejemplo, un coeficiente de temperatura de +20 Hz/MHz/°C es lo mismo que +20 ppm/°C. Si la dirección del cambio de frecuencia es igual a la del cambio de temperatura, es decir, que un aumento de temperatura cause un aumento de frecuencia, y que una disminución de temperatura cause una disminución de frecuencia, se llama *coeficiente positivo de temperatura*, o *coeficiente térmico positivo*. Si el cambio de frecuencia tiene dirección opuesta a la del cambio de temperatura (un aumento de temperatura causa una disminución de frecuencia y una disminución de temperatura causa un aumento de frecuencia), se llama *coeficiente negativo de temperatura*. La relación matemática del cambio de frecuencia de un cristal con el cambio de temperatura es

$$\Delta f = k(f_n \times \Delta C) \tag{2-5}$$

en donde Δf = cambio de frecuencia (hertz)
k = coeficiente de temperatura (Hz/MHz/°C)
f_n = frecuencia natural del cristal (megahertz)
ΔC = cambio de temperatura (grados Celsius)

y

$$f_o = f_n + \Delta f \tag{2-6}$$

donde f_o es la frecuencia de operación.

El coeficiente de temperatura k de un cristal varía, dependiendo del tipo de corte y de la temperatura de operación. Entre las temperaturas aproximadas de +20° C a +50° C, los cristales cortados X y Y tienen un coeficiente de temperatura que es casi constante. Los cristales cortados X son unas 10 veces más estables que los cortados Y. En forma normal, los cristales cortados X tienen un coeficiente de temperatura que va de −10 a −25 Hz/MHz/°C. El coeficiente de temperatura de los cristales cortados Y va de unos −25 hasta unos +100 Hz/MHz/°C.

Hoy se consiguen cristales con coeficiente cero (cortados GT) cuyos coeficientes de temperatura son de −1 hasta +1 Hz/MHz/°C. El cristal cortado GT es casi un cristal perfecto, de coeficiente cero, desde temperaturas de congelación hasta ebullición, pero sólo es útil en frecuencias menores que algunos cientos de kilohertz.

Ejemplo 2-1

Para un cristal de 10 MHz con coeficiente de temperatura $k = +10$ Hz/MHz/°C, calcular la frecuencia de operación si la temperatura:

(a) Aumenta 10° C.

(b) Disminuye 5° C.

Solución (a) Se sustituyen los valores en las ecuaciones 2-5 y 2-6, para obtener

$$\Delta f = k(f_n \times \Delta C)$$
$$= 10(10 \times 10) = 1 \text{ kHz}$$
$$f_o = f_n + \Delta f$$
$$= 10 \text{ MHz} + 1 \text{ kHz} = 10.001 \text{ MHz}$$

(b) De nuevo se sustituyen los valores en las ecuaciones 2-5 y 2-6

$$\Delta f = 10[10 \times (-5)] = -500 \text{ Hz}$$
$$f_o = 10 \text{ MHz} + (-500 \text{ Hz})$$
$$= 9.9995 \text{ MHz}$$

Circuito equivalente del cristal. La fig. 2-8a muestra el circuito eléctrico equivalente de un cristal. Cada componente eléctrico equivale a una propiedad mecánica del cristal. C_2 es la capacitancia real formada entre los electrodos del cristal, y el cristal mismo es el dieléctrico. C_1 equivale a la docilidad mecánica del cristal (llamada también resiliencia o elasticidad). L equivale a la masa del cristal en vibración y R es la pérdida por fricción mecánica. En un cristal, la *relación de masa a fricción* (L/R), es bastante alta. Los valores característicos van desde 0.1 H hasta bastante más que 100 H; en consecuencia, los factores Q son bastante altos para los cristales. Los valores entre 10,000 a 100,000 o más no son raros (en comparación con los factores Q de 100 a 1000 para los inductores discretos que se usan en los circuitos tanque LC). Esto proporciona la gran estabilidad de los osciladores de cristal, en comparación con los de circuito tanque LC. Los valores de C_1 suelen ser menores que 1 pF, y los de C_2 van de 4 a 40 pF.

Como un cristal tiene un circuito equivalente en serie y uno en paralelo, también hay dos impedancias equivalentes y dos frecuencias de resonancia: en serie y en paralelo. La impedancia en serie es la combinación de R, L y C_1, es decir, $Z_s = R \pm jX$, donde $X = |X_L - X_C|$. La impedancia en paralelo es aproximadamente igual a la impedancia de L y C_2, es decir, $Z_p = (X_L \times X_{C2})/(X_L + X_{C2})$. A muy bajas frecuencias, la impedancia en serie de L, C_1 y R es muy alta y capacitiva (−). Esto se ve en la fig. 2-8c. Al aumentar la frecuencia se alcanza un punto en el que $X_L = X_{C1}$. En esta frecuencia, f_1, la impedancia en serie es mínima, es resistiva y es igual a R. Al aumentar más la frecuencia (f_2), la impedancia en serie se vuelve alta e inductiva (+). La combinación de L y C_2 en paralelo hace que el cristal funcione como un circuito tanque en paralelo, con la impedancia máxima en la resonancia. La diferencia entre f_1 y f_2 suele ser bastante pequeña, normalmente más o menos el 1% de la frecuencia natural del cristal. Un cristal puede funcionar a su frecuencia resonante en serie o en paralelo, dependiendo de la configuración del circuito en que se use. La pendiente relativa de la curva de impedancia que se ve en la fig. 2-8b también se

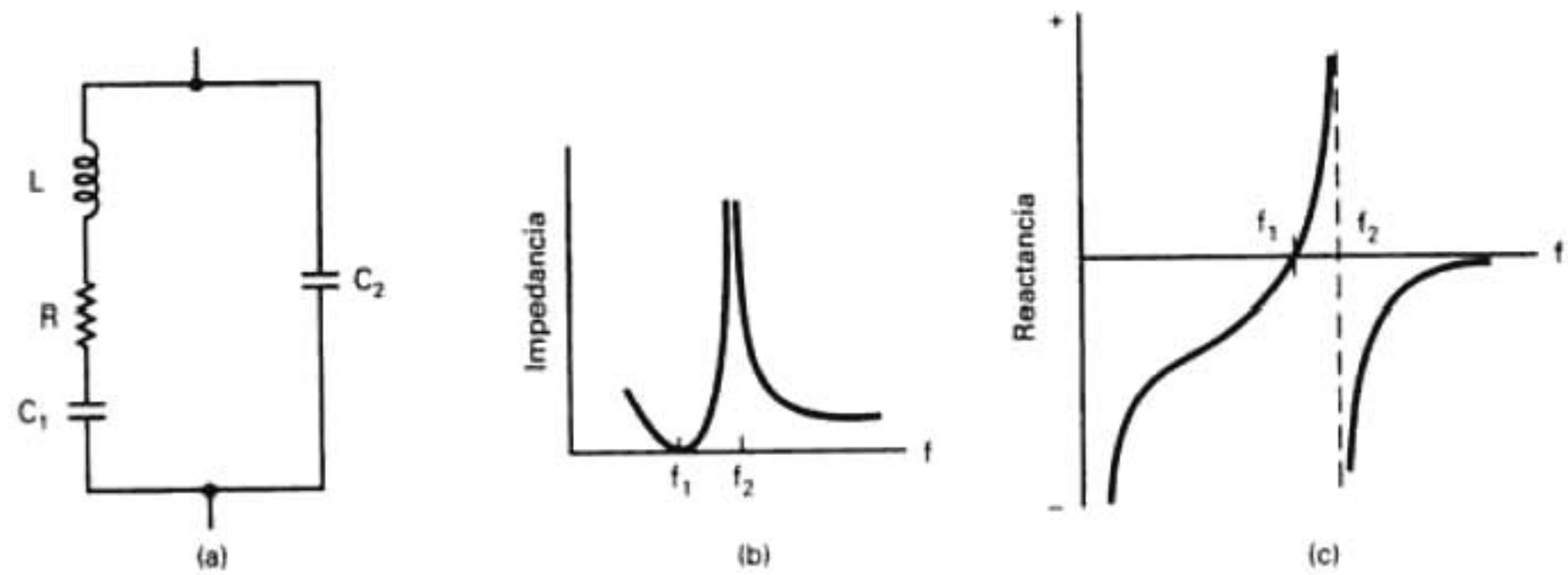


FIGURA 2-8 Circuito equivalente del cristal: (a) circuito equivalente; (b) curva de impedancia; (c) curva de reactancia

atribuye a la estabilidad y la exactitud de un cristal. La frecuencia de resonancia en serie de un cristal de cuarzo es

$$f_1 = \frac{1}{2\pi\sqrt{(LC_1)}}$$

y la frecuencia de resonancia en paralelo es

$$f_2 = \frac{1}{2\pi\sqrt{(LC)}}$$

siendo C la combinación de C_1 y C_2 en serie.

Circuitos con oscilador de cristal. Aunque hay muchas configuraciones de oscilador a base de cristal, las más comunes son las de Pierce, discreta y de circuito integrado, y la de semipuente *RLC*. Si se necesita una muy buena estabilidad de frecuencia y a la vez un circuito razonablemente sencillo, una buena opción es la discreta de Pierce. Si lo que se busca es bajo costo y posibilidades de interconexión digital sencilla, bastará con un oscilador de Pierce de circuito integrado. Sin embargo, para tener la óptima estabilidad de frecuencia, lo más aconsejable es el semipuente *RLC*.

Oscilador discreto de Pierce. El oscilador discreto de Pierce, de cristal, tiene muchas ventajas. Su frecuencia de funcionamiento abarca todo el intervalo de fundamentales del cristal (de 1 kHz a más o menos 30 MHz). Usa circuitos relativamente sencillos que requieren pocos componentes (la mayoría de las versiones para frecuencia intermedia sólo requieren un transistor). El diseño del oscilador de Pierce produce una gran potencia de señal de salida, y al mismo tiempo disipa muy poca potencia en el cristal mismo. Por último, la estabilidad de frecuencia a corto plazo de este oscilador es excelente, porque la Q cargada en el circuito es casi tan alta como la Q interna del cristal. El único inconveniente del oscilador de Pierce es que requiere un amplificador de alta ganancia, de 70 aproximadamente. En consecuencia, se debe usar un solo transistor de alta ganancia, o quizá hasta uno de varias etapas.

La fig. 2-9 muestra un circuito oscilador discreto de Pierce, de 1 MHz. Q_1 proporciona toda la ganancia necesaria para que se produzcan las oscilaciones autosostenidas. R_1 y C_1 proporcionan un retraso de fase de 65° a la señal de retroalimentación. La impedancia del cristal es básicamente resistiva, con un componente inductivo pequeño. Esta impedancia se combina con la reactancia de C_2 para producir un retraso adicional de fase de 115°. El transistor invierte la señal (180° de desplazamiento de fase), dando al circuito los 360° necesarios de desplazamiento total de fase. Como la carga del cristal es principalmente no resistiva (en su mayor parte la combinación de C_1 y C_2 en serie), este tipo de oscilador proporciona una estabilidad de frecuencia a corto plazo muy buena. Desafortunadamente, C_1 y C_2 introducen pérdidas apreciables y, en consecuencia, el transistor debe tener una ganancia relativamente alta de voltaje; esto es su inconveniente obvio.

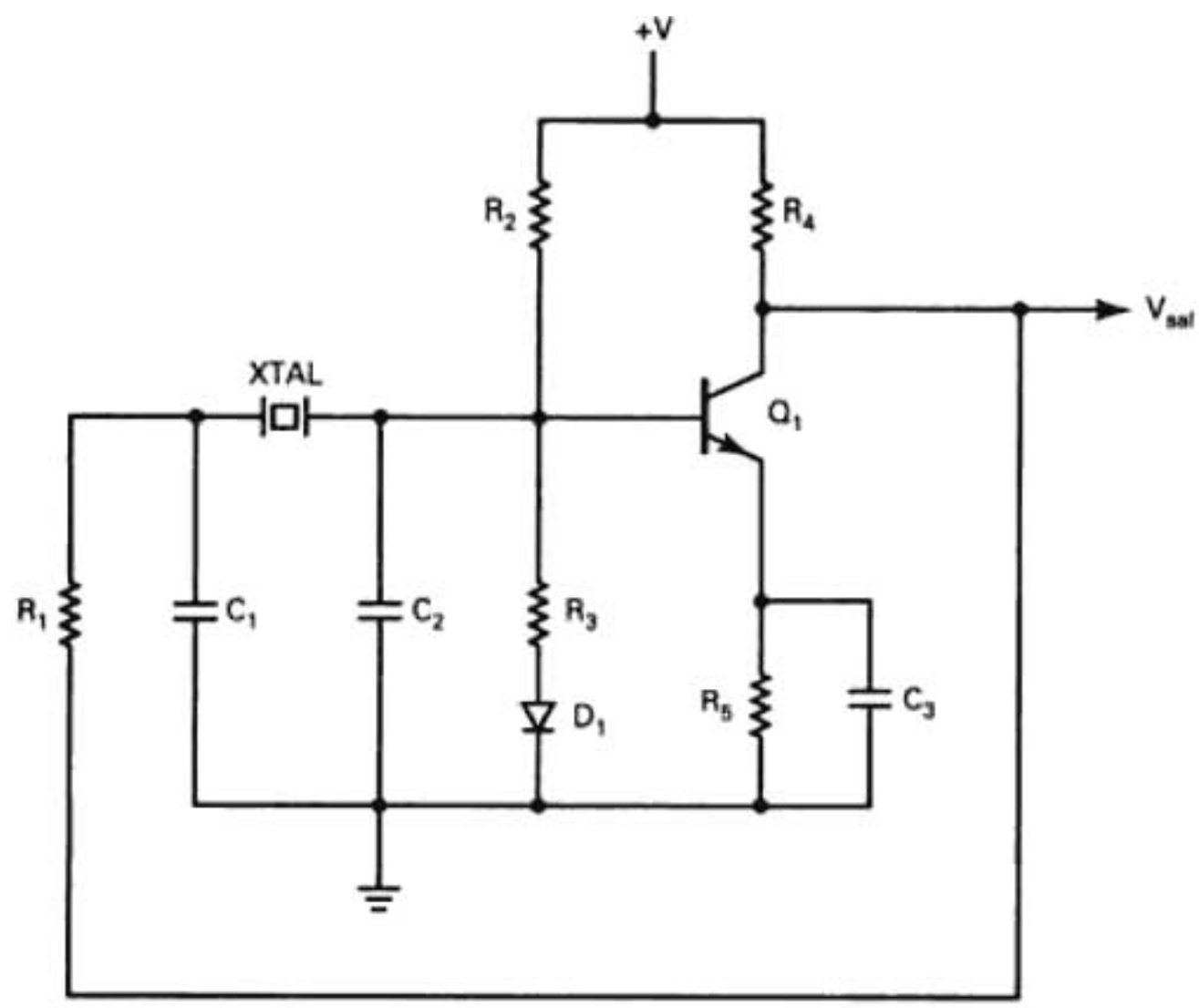


FIGURA 2-9 Oscilador discreto de cristal de Pierce

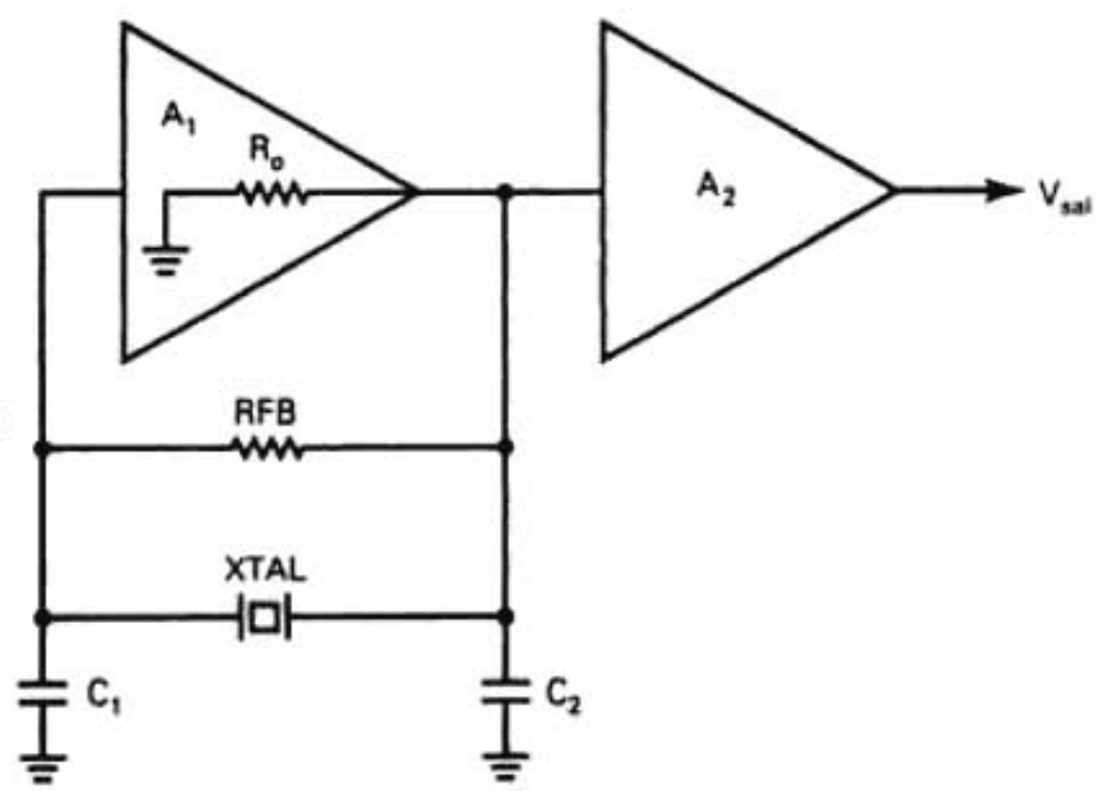


FIGURA 2-10 Oscilador de cristal de Pierce, en circuito integrado

Oscilador de Pierce en circuito integrado. La fig. 2-10 muestra un oscilador de Pierce de cristal en circuito integrado. Aunque proporciona menor estabilidad de frecuencia, se puede implementar con un diseño sencillo de circuito integrado digital, y reduce mucho los costos respecto a los diseños convencionales discretos.

Para asegurar que se inicien las oscilaciones, la cd de retroalimentación regenerativa polariza a la entrada y la salida de A_1 para la operación de clase A. El amplificador A_2 convierte a la salida de A_1 en una oscilación completa de máxima excursión (corte a saturación), reduciendo los tiempos de subida y bajada, y amortiguando la salida de A_1. La resistencia de salida de A_1 se combina con C_2 para formar el retraso de fase RC necesario. Las versiones de semiconductor metal-óxido complementario (CMOS) funcionan hasta unos 2 MHz, y las versiones con lógica acoplada por el emisor (ECL) operan hasta 20 MHz.

***Oscilador de cristal de semipuente* RLC.** La fig. 2-11 muestra la versión Meacham del oscilador de cristal de semipuente *RLC*. El oscilador original Meacham fue desarrollado en la década de 1940, y usaba un puente completo de cuatro brazos y una lámpara de tungsteno con coeficiente negativo. La configuración del circuito de la figura 2-11 sólo usa un puente de dos brazos

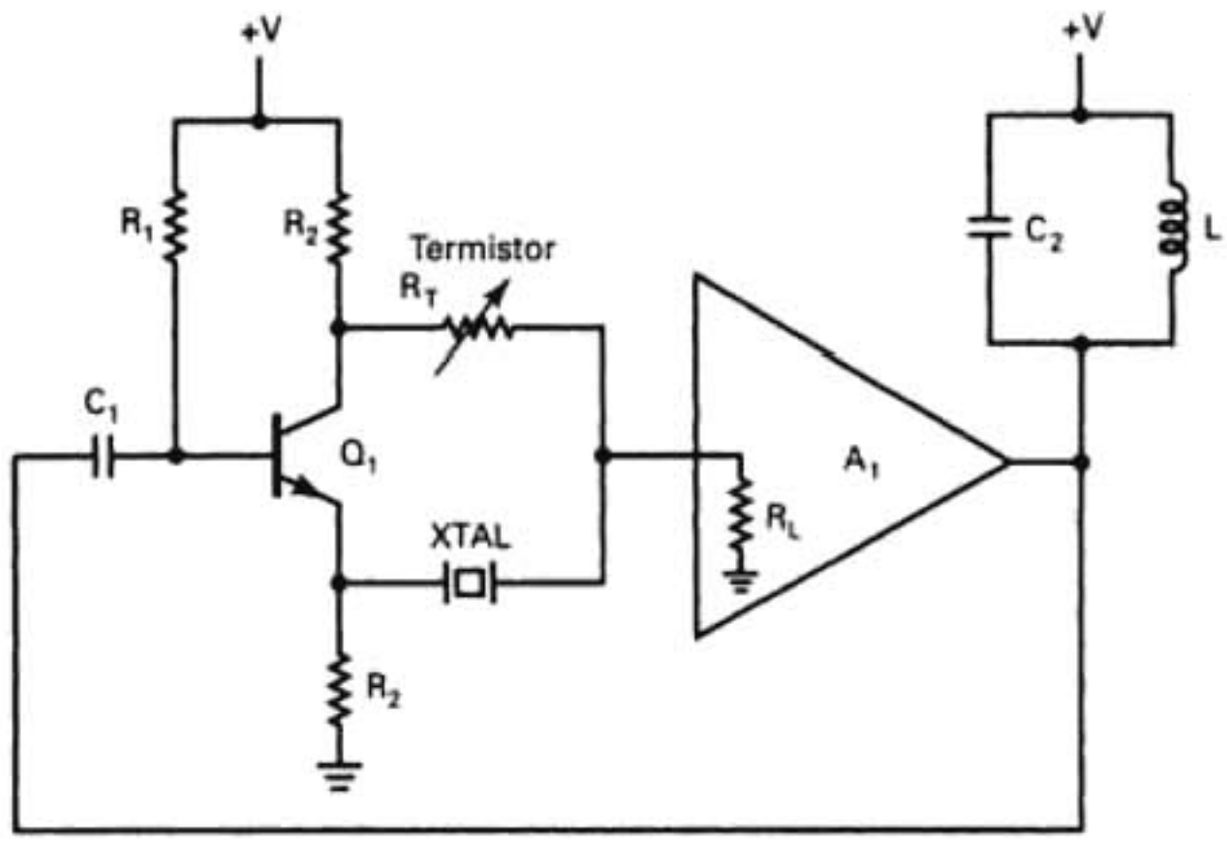


FIGURA 2-11 Oscilador de cristal de semipuente *RLC*

y emplea un termistor con coeficiente negativo de temperatura. Q_1 funciona como divisor de fase y proporciona dos señales desfasadas 180°. El cristal debe funcionar en su frecuencia de resonancia en serie, por lo que su impedancia interna es resistiva y bastante pequeña. Cuando comienzan las oscilaciones, la amplitud de la señal aumenta en forma gradual, disminuyendo la resistencia del termistor hasta que el puente casi se pone en cero. La amplitud de las oscilaciones se estabiliza y determina la resistencia final del termistor. El circuito tanque ***LC*** en la salida se sintoniza a la frecuencia de resonancia en serie del cristal.

Módulo de oscilador de cristal. Un *módulo de oscilador de cristal* consiste en un oscilador controlado por cristal y un componente de voltaje variable, como por ejemplo un *diodo varactor*. Todo el circuito está contenido en un *encapsulado de metal*. En la figura 2-12a se ve un diagrama simplificado de un módulo de oscilador de Colpitts de cristal. X_1 es el cristal mismo y Q_1 es el componente activo para el amplificador. C_1 es un capacitor variable que permite variar la frecuencia del oscilador de cristal dentro de un margen angosto de frecuencias de funcionamiento. VC_1 es un capacitor variable por voltaje (*varicap* o *diodo varactor*). Un diodo varactor es un diodo de construcción especial, cuya capacitancia interna aumenta con polarización inversa; se puede ajustar la capacidad del diodo variando el voltaje de polarización inversa. Un diodo varactor tiene una capa especial de agotamiento entre los materiales tipo *p* y tipo *n* del que está hecho, con diversos grados y clases de material que genera impurezas. A veces se usa el término *unión graduada* para describir la fabricación del diodo varactor. La fig. 2-12b muestra la curva de capacitancia en función del voltaje de polarización inversa para un diodo varactor normal. La capacitancia de uno de estos diodos se calcula aproximadamente con la fórmula

$$C_d = \frac{C}{\sqrt{(1 + 2|V_r|)}} \tag{2-7}$$

donde C = capacitancia del diodo (farads) con 0 V de polarización inversa
$|V_r|$ = magnitud del voltaje de polarización inversa del diodo (volts)
C_d = capacitancia (en farads) del diodo con polarización inversa

La frecuencia con la que oscila el cristal se puede ajustar un poco cambiando la capacitancia de VC_1, es decir, cambiando el valor del voltaje de polarización inversa. El diodo varactor, junto con un módulo de compensación de temperatura, proporciona compensación instantánea de frecuencia por variaciones causadas por cambios de temperatura. El diagrama esquemático de un módulo compensador de temperatura se ve en la fig. 2-13. El módulo de compensación incluye un amplificador separador (Q_1) y una red compensadora de temperatura (T_1). Esta red T_1 es un termistor con coeficiente negativo de temperatura. Cuando la temperatura baja del valor umbral del termistor, el voltaje de compensación aumenta. Este voltaje de compensación se

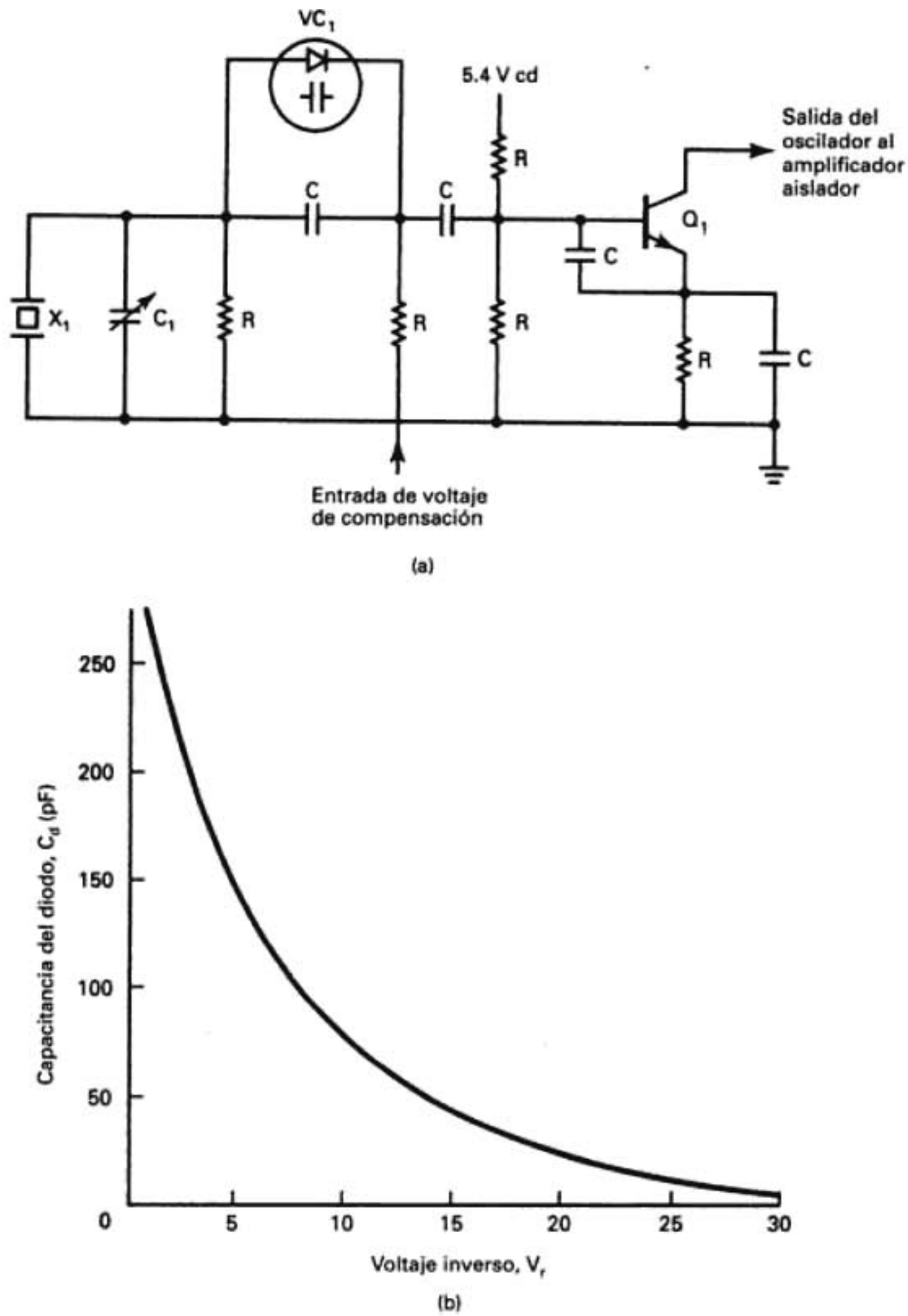


FIGURA 2-12 Módulo de oscilador de cristal: (a) esquema; (b) características del diodo varactor

aplica al módulo oscilador, donde controla la capacitancia del diodo varactor. Se consiguen módulos de compensación cuya acción permite tener una estabilidad de frecuencia de 0.0005%, desde −30° C hasta + 80° C.

OSCILADORES EN GRAN ESCALA DE INTEGRACIÓN

En los últimos años ha aumentado con una rapidez asombrosa el uso de los circuitos con integración a *gran escala* (LSI, de *large scale integration*) para generar frecuencias y formas de onda, porque los osciladores de circuito integrado tienen una excelente estabilidad de frecuencia y un amplio margen de sintonía, y por ser fáciles de usar. Los *generadores de forma de onda* y los *generadores de funciones* se usan en forma extensa en equipos de comunicaciones y de telemetría,

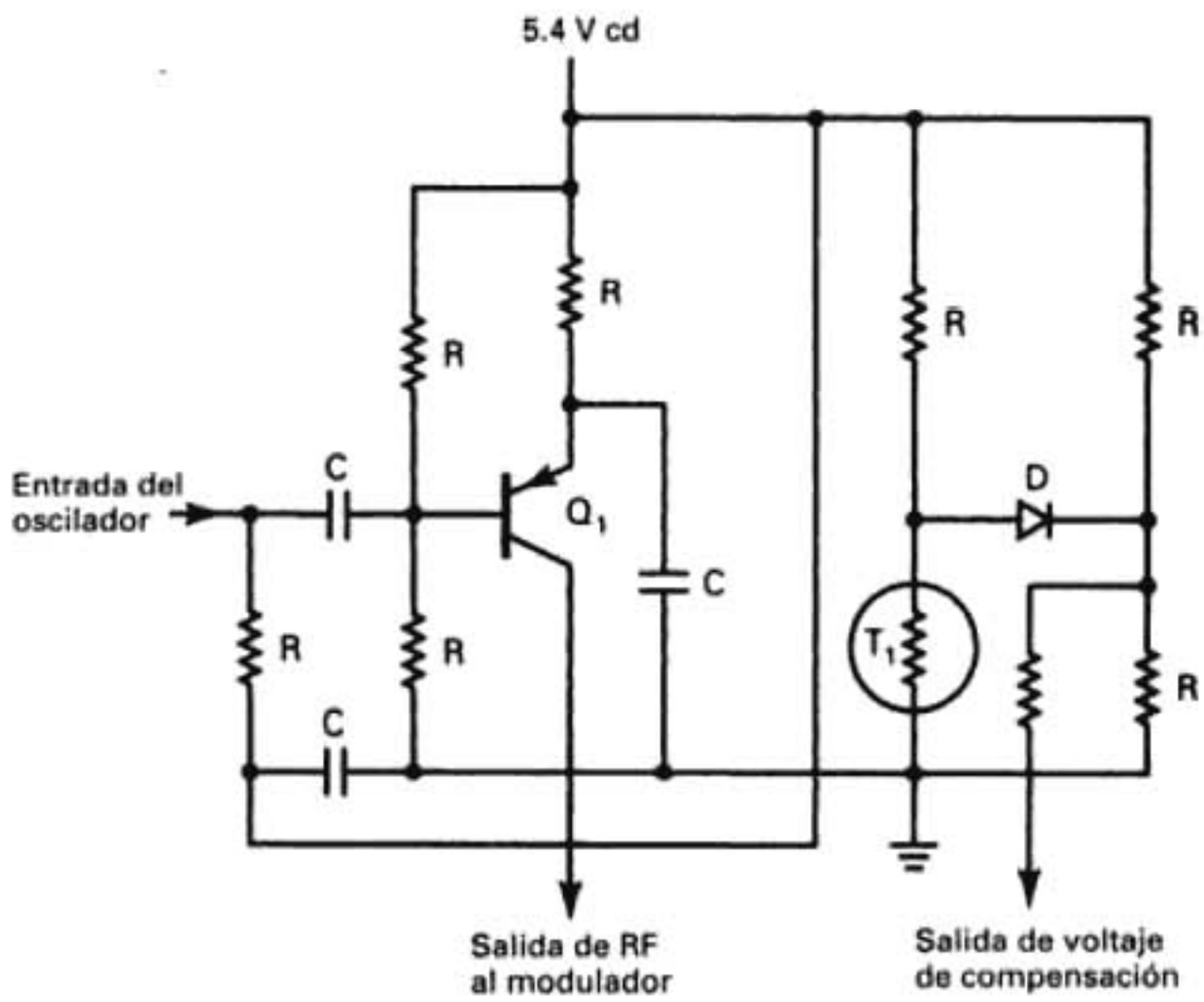


FIGURA 2-13 Circuito de compensación

así como en laboratorios, para equipos de prueba y calibración. En muchas de esas aplicaciones se consiguen osciladores y generadores de funciones comerciales, en circuitos integrados monolíticos, que permiten contar al diseñador con una alternativa de bajo costo para sus contrapartes de circuito no integrado.

Las operaciones básicas necesarias para generar y conformar formas de onda se adaptan bien a la tecnología del circuito integrado monolítico. De hecho, los *circuitos integrados lineales monolíticos* (LIC, de *linear integrated circuits*) tienen varias ventajas inherentes sobre los circuitos discretos, como la disponibilidad de una gran cantidad de dispositivos activos en un solo CI, y la estrecha compensación y seguimiento térmico de los valores de los componentes. Hoy, es posible fabricar generadores de forma de onda con circuito integrado que tienen un desempeño comparable a los generadores discretos complejos, con sólo una fracción de su costo.

Los generadores LSI (con integración en gran escala) que se consiguen hoy incluyen generadores de funciones, temporizadores, relojes programables, osciladores controlados por voltaje, osciladores de precisión y generadores de forma de onda.

Generación de forma de onda con circuito integrado

En su forma más simple, un generador de forma de onda es un circuito oscilador que genera formas de onda bien definidas y estables, que se pueden modular o barrer externamente sobre un intervalo determinado de frecuencia. Un generador de onda típico está formado por cuatro secciones básicas: 1) un oscilador, que genera la forma de onda periódica básica, 2) un conformador de onda, 3) un modulador opcional de AM y 4) un amplificador separador de salida para aislar el oscilador de la carga y proporcionar la corriente necesaria.

La fig. 2-14 muestra un diagrama simplificado de bloques de un circuito generador de forma de onda con circuito integrado que muestra la relación entre las cuatro secciones. Cada sección se ha fabricado por separado en forma monolítica durante varios años; por consiguiente, la fabricación de las cuatro secciones en un solo CI monolítico fue una evolución natural de una tecnología preexistente. La sección del oscilador genera la frecuencia básica de oscilación, y el circuito conformador de onda convierte la salida del oscilador en una forma de onda senoidal, cuadrada, triangular o en rampa. El modulador, cuando se usa, permite que el circuito produzca señales de amplitud modulada, y el amplificador separador aísla al oscilador de su carga y permite contar con un lugar cómodo para sumar valores de cd a la forma de onda en la salida. La salida de sincronización se puede usar como fuente de onda cuadrada o como pulso de sincronización para los circuitos externos de temporización.

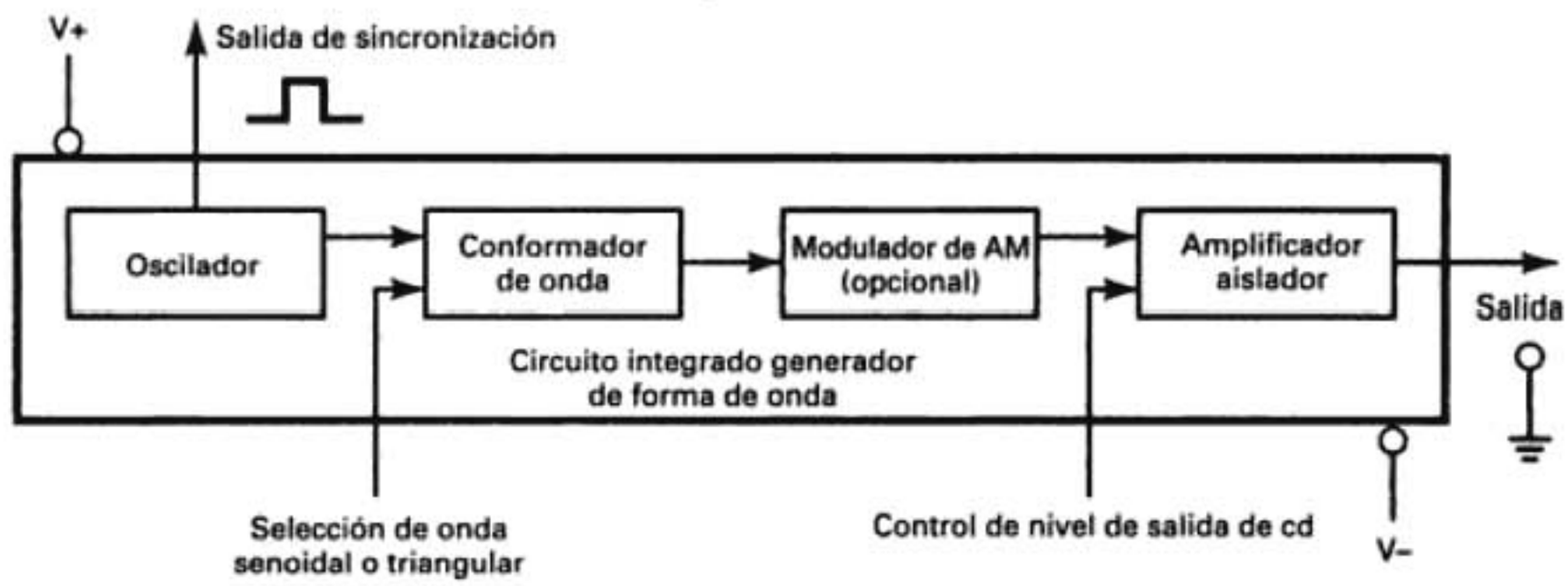


FIGURA 2-14 Generador de forma de onda en circuito integrado

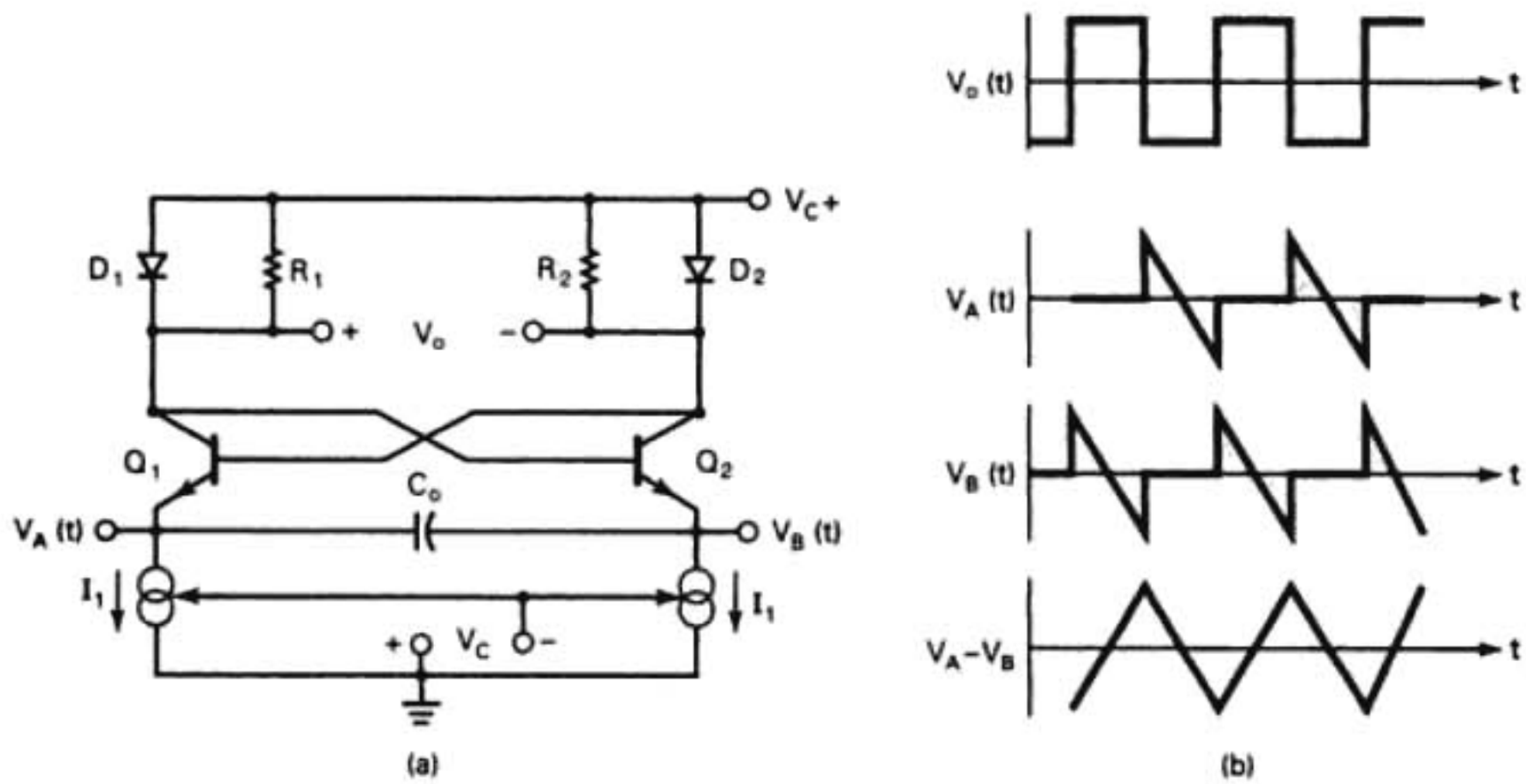


FIGURA 2-15 Generador de forma de onda en circuito integrado, simplificado: (a) esquema; (b) formas de onda

Un circuito integrado oscilador usa la carga y descarga de corriente constante de capacitores externos de temporización. La fig. 2-15a muestra el esquema simplificado de uno de estos generadores de onda, que usa un multivibrador acoplado por el emisor, que es capaz de generar ondas cuadradas, así como las formas triangular y de rampa lineal. El circuito funciona como sigue. Cuando el transistor Q_1 y el diodo D_1 conducen, el transistor Q_2 y el diodo D_2 están apagados y viceversa. Esta acción carga y descarga alternativamente al capacitor C_o con la fuente de corriente constante I_1. El voltaje a través de D_1 y D_2 es una onda cuadrada simétrica con amplitud de pico a pico igual a $2V_{BE}$. V_A es constante cuando Q_1 está activo, pero se transforma en una rampa lineal con pendiente igual a $-I_1/C_o$ cuando Q_1 se apaga. La salida $V_B(t)$ es idéntica a $V_A(t)$, excepto que está demorada medio ciclo. La salida diferencial, $V_A(t) - V_B(t)$ es una onda triangular. La fig. 2-15b muestra las formas de onda de voltaje de salida que se producen en forma normal.

Generadores de funciones con CIs monolíticos. El XR-2206 es un generador de funciones en circuito integrado monolítico, fabricado por EXAR Corporation, que puede producir formas de onda de alta calidad tipo senoidal, cuadradas, triangular, en rampa y de pulso, con un alto grado de estabilidad y exactitud. Las formas de onda de salida del XR-2206 se pueden modular tanto en amplitud como en frecuencia, con una señal moduladora externa; la frecuencia de operación se puede seleccionar externamente dentro de un intervalo de 0.01 Hz hasta más de 1 MHz. El XR-2206 se adapta en forma ideal a las comunicaciones, instrumentación y a las aplicaciones

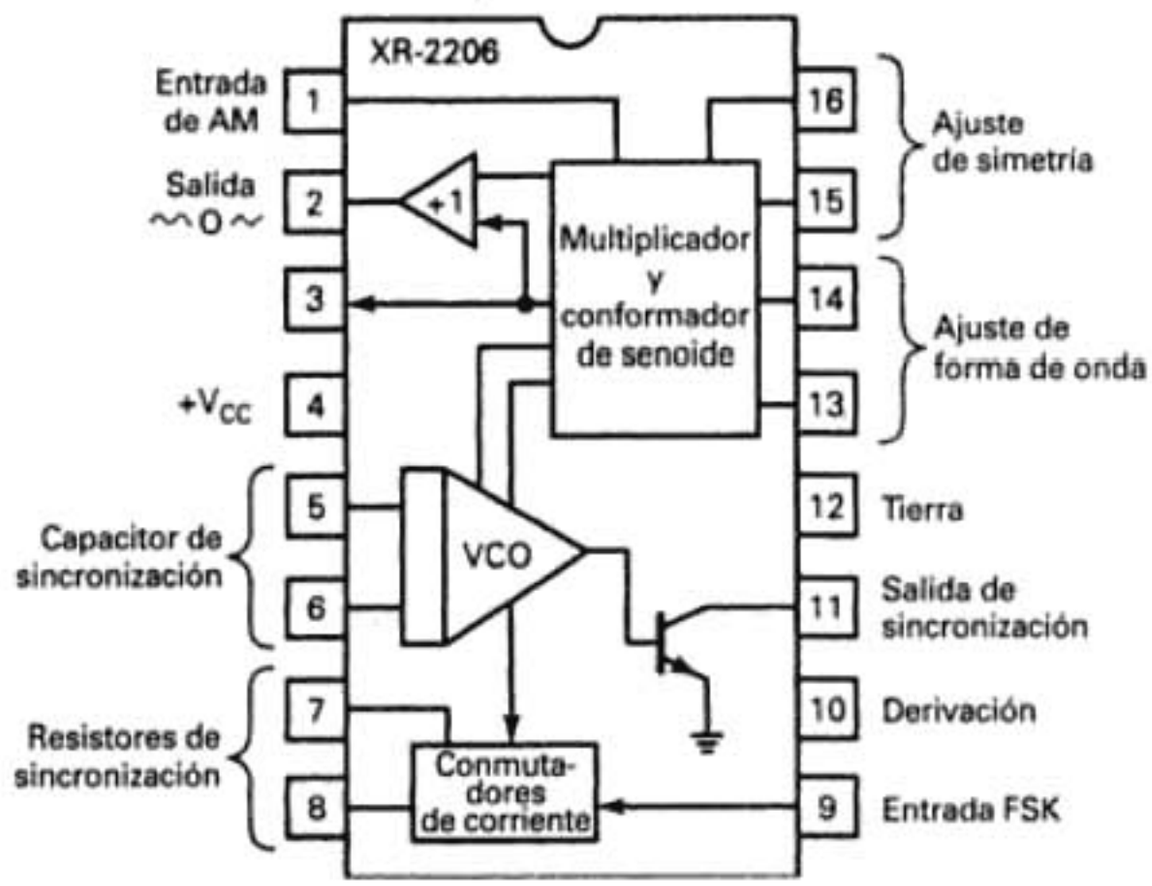


FIGURA 2-16 Diagrama de bloques del generador monolítico de funciones XR-2206

de generador de funciones que requieran tono senoidal, AM o FM. Tiene una estabilidad normal de frecuencia de 20 ppm/°C y se pueden barrer linealmente con un voltaje externo de control, en un intervalo de frecuencias de 2000:1.

El diagrama de bloques del XR-2206 se ve en la fig. 2-16. El generador de funciones está formado por cuatro bloques funcionales: un oscilador controlado por voltaje (VCO), un multiplicador y conformador senoidal analógico, un amplificador separador de ganancia unitaria y un conjunto de interruptores de corriente de entrada. Un *oscilador controlado por voltaje* es un oscilador de funcionamiento autónomo con una frecuencia estable de operación que depende de una capacitancia externa de temporización, resistencia de temporización y un voltaje de control. La salida de un VCO es una frecuencia, y su entrada es una señal de polarización o de control que puede ser un voltaje de cd o de ca. El VCO produce en realidad una frecuencia de salida que es proporcional a una corriente de entrada producida por un resistor desde las terminales de temporización (la 7 o la 8) a tierra. Los interruptores de corriente conducen a ésta de una de las terminales de temporización al VCO. La corriente seleccionada depende del valor del voltaje en la terminal (la 9) de entrada de modulación por conmutación de frecuencia. Por consiguiente, se pueden producir en forma independiente dos frecuencias discretas de salida. Si la terminal 9 está en circuito abierto o está conectada con un voltaje de polarización ≥ 2 V, se selecciona la corriente que pasa por el resistor conectado a la terminal 7. De igual modo, si el valor del voltaje en la terminal 9 es ≤ 1 V, se selecciona la corriente que pasa por el resistor conectado con la terminal 8. De esta manera, la frecuencia de salida se puede cambiar entre f_1 y f_2 tan sólo cambiando el voltaje en la terminal 9. Las fórmulas para determinar las dos frecuencias de operación son

$$f_1 = \frac{1}{R_1 C} \qquad f_2 = \frac{1}{R_2 C}$$

donde R_1 = resistor conectado con la terminal 7
R_2 = resistor conectado con la terminal 8

La frecuencia de oscilación es proporcional a la corriente total de sincronización en la terminal 7 o en la 8. La frecuencia varía en forma lineal con la corriente, dentro de un intervalo de valores de corriente de 1 a 3 μA. Esta frecuencia se puede controlar aplicando un voltaje de control V_C a la terminal seleccionada de sincronización, como se ve en la fig. 2-17. La frecuencia de oscilación se relaciona con V_C mediante la ecuación

$$f = \frac{1}{RC}\left[1 + \frac{R}{R_C}\frac{(1 - V_C)}{3}\right] \text{Hz} \tag{2-8}$$

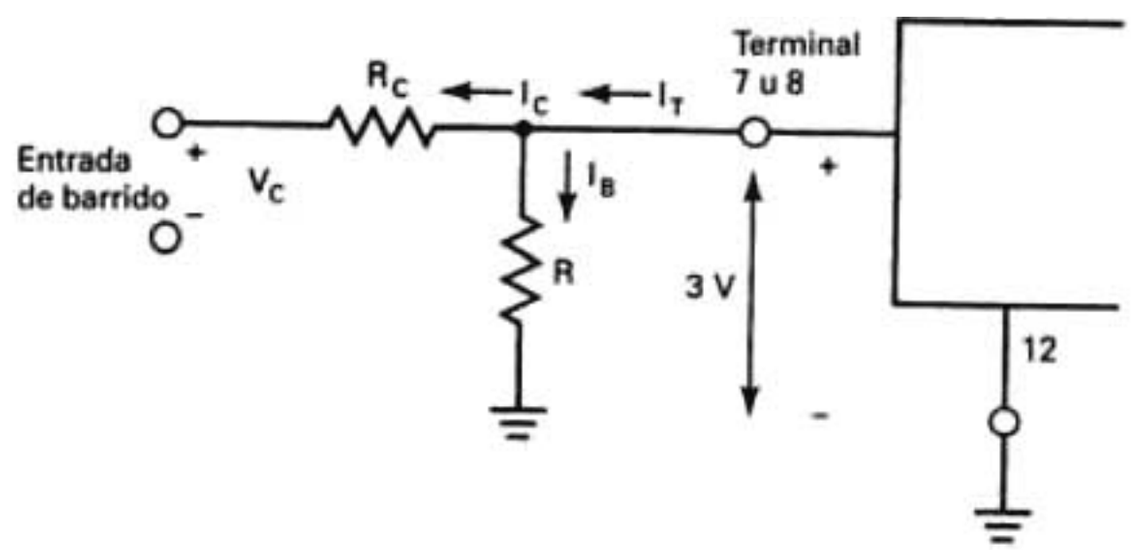


FIGURA 2-17 Conexión del circuito para barrido de frecuencia con voltaje de control para el XR-2206

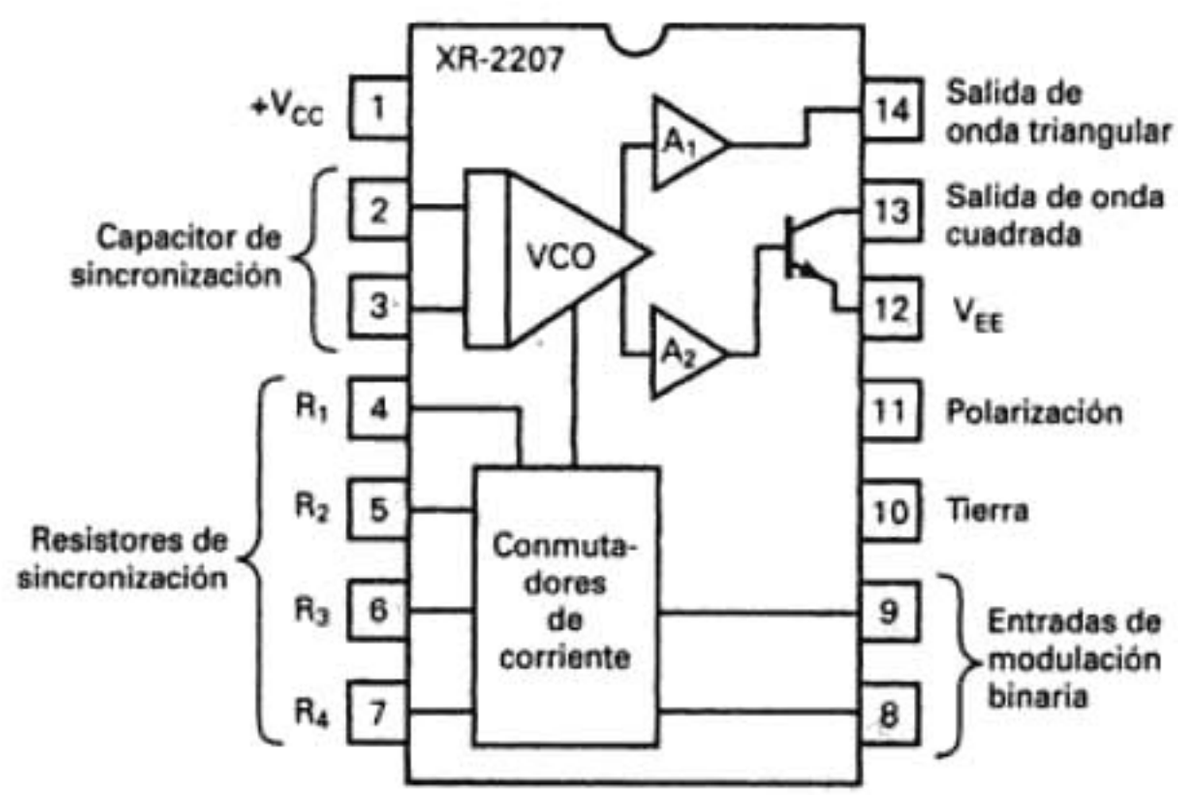


FIGURA 2-18 Diagrama de bloques del oscilador monolítico controlado por voltaje (VCO) XR-2207

La ganancia K, de conversión de voltaje a frecuencia, se determina con

$$K = \frac{\Delta f}{\Delta V_C} = \frac{-0.32}{R_C C} \text{ Hz/V} \tag{2-9}$$

Osciladores con CIs monolíticos controlados por voltaje. El XR-2207 es un oscilador controlado por voltaje (VCO) en circuito integrado monolítico, que produce una excelente estabilidad de frecuencia y un amplio margen de sintonía. El circuito permite tener salidas simultáneas de ondas triangulares y cuadradas, dentro de un intervalo de frecuencias desde 0.01 Hz hasta 1 MHz. Este XR-2207 es ideal para FM, FSK (modulación por conmutación de frecuencia, *frequency-shift keying*) y barrido o generación de tono, así como aplicaciones de lazo de fase cerrada. El circuito tiene una estabilidad normal de frecuencia de 20 ppm/°C y se puede barrer linealmente con un voltaje externo de control, dentro de un intervalo de frecuencias de 1000:1. El ciclo de trabajo de las salidas triangular y cuadrada se puede variar de 0.1% a 99.9%, generando formas de onda estables de pulso y de diente de sierra.

El diagrama de bloques del XR-2207 se ve en la fig. 2-18. El circuito es un multivibrador acoplado por emisor, modificado, que usa cuatro bloques funcionales principales para generación de frecuencia: un oscilador controlado por voltaje (VCO), cuatro conmutadores de corriente que se activan con entradas de manipulación binaria, y dos amplificadores de aislamiento. Dos terminales de entrada primaria (las 8 y 9) determinan cuál de las cuatro corrientes de temporización se canalizan al VCO. Esas corrientes se regulan con resistores a tierra, desde cada una de

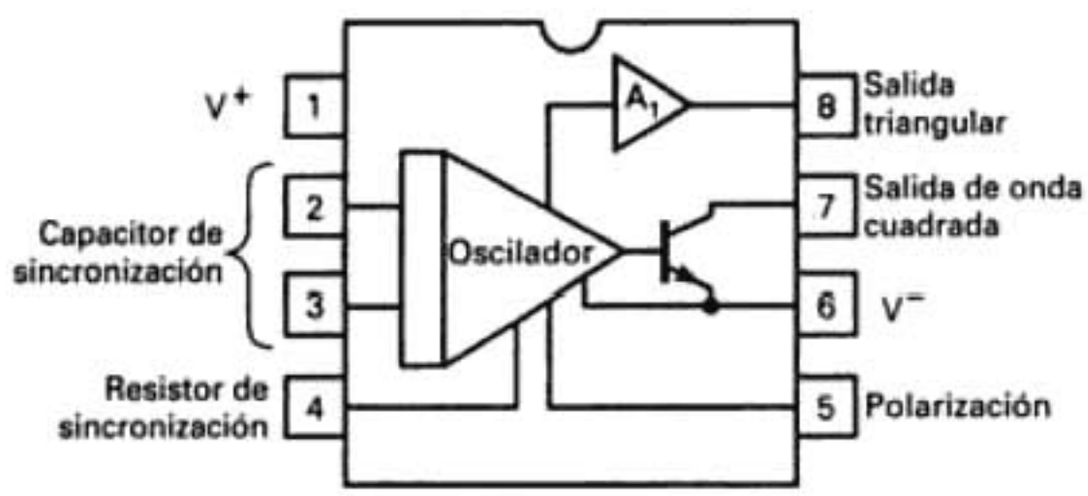


FIGURA 2-19 Diagrama de bloques de un oscilador monolítico de precisión XR-2209

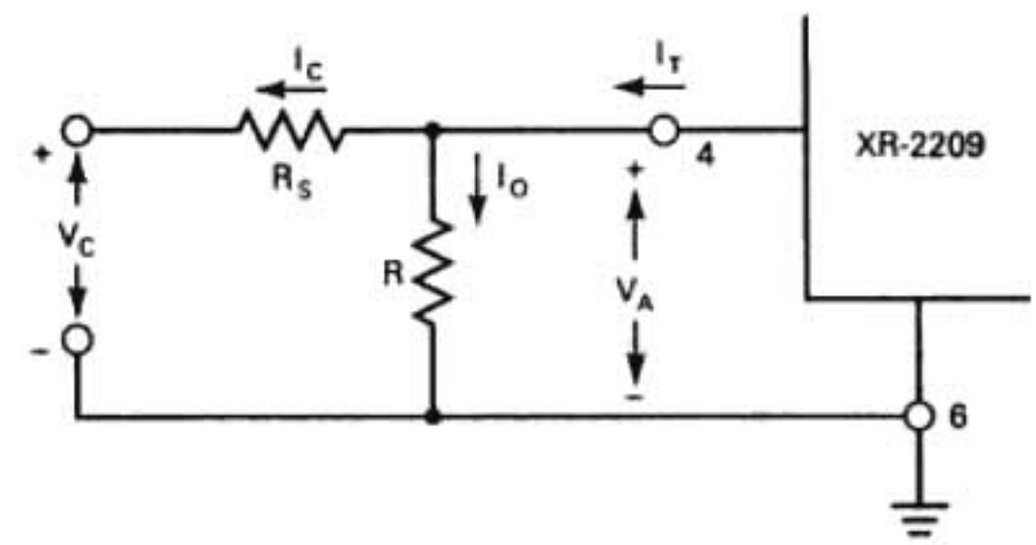


FIGURA 2-20 Conexión de circuito para barrido de frecuencia con voltaje de control en el XR-2209

las cuatro terminales de entrada de temporización (de la 4 a la 7). El separador (búfer) de salida triangular proporciona una salida de baja impedancia (normalmente de 10 W), mientras que la salida de onda cuadrada es de colector abierto.

Osciladores de precisión monolíticos. El XR-2209 es un circuito oscilador de frecuencia variable, monolítico, que tiene una excelente estabilidad de temperatura y un amplio margen de barrido lineal. Proporciona salidas simultáneas de onda triangular y cuadrada, y la frecuencia se ajusta con un producto externo *RC*. El XR-2209 es ideal para modulación de frecuencia, conversión de voltaje a frecuencia y barrido o generación de tono, así como para aplicaciones de lazo de fase cerrada cuando se usa junto con un comparador adecuado de fases.

El diagrama de bloques del oscilador de precisión XR-2209 se ve en la fig. 2-19. El oscilador se compone de tres bloques funcionales: un oscilador de frecuencia variable que genera las formas periódicas básicas de onda, y dos amplificadores de aislamiento para las salidas de onda triangular y cuadrada. La frecuencia del oscilador se ajusta con un capacitor externo y resistor de temporización. El circuito es capaz de operar dentro de ocho décadas de frecuencia, de 0.01 Hz a 1 MHz. Cuando no hay señal externa de barrido ni voltaje de polarización, la frecuencia de oscilación es simplemente $1/RC$.

La frecuencia de operación para el XR-2209 es proporcional a la corriente de temporización, que se toma de la terminal correspondiente. Esta corriente se puede modular aplicando un voltaje de control V_C a la terminal de temporización, a través del resistor R_S en serie, como se ve en la fig. 2-20. Si V_C es negativo con respecto al voltaje en la terminal 4, se toma una corriente I_O adicional de la terminal de sincronización, haciendo aumentar la corriente total de entrada y aumentando de este modo la frecuencia de oscilación. Al revés, si V_C es mayor que el voltaje en la terminal 4, disminuye la frecuencia de oscilación.

LAZOS DE FASE CERRADA

El *lazo de fase cerrada* (PLL, de *phase-locked loop*) o lazo *amarrado por fase* se usa en forma extensa en las comunicaciones electrónicas para modulación, demodulación, generación de frecuencia y síntesis de frecuencia. Los PLL se usan tanto en los transmisores como en los receptores, con modulación analógica o digital, y con la transmisión de pulsos digitales. Los lazos de fase cerrada se usaron por primera vez en 1932 para detección síncrona de señales de radio, circuitos de instrumentación y sistemas de telemetría espacial. Sin embargo, se evitó durante muchos años el uso de los PLL por su gran tamaño, complejidad necesaria, banda angosta y costo. Con la entrada de la integración en gran escala los PLL ocupan poco espacio, son fáciles de usar y son más fiables. Por consiguiente, los PLL cambiaron desde una técnica especializada de diseño y ahora son un elemento constructivo universal, con numerosas aplicaciones. Hoy se consigue más de una docena de productos con PLL distintos, de circuito integrado, con varios fabricantes. Algunos de ellos se consideran como circuitos de propósito general adecuados para una multitud de usos, mientras que otros estan planeados u optimizados para aplicaciones especiales como detección de tono, decodificación estereofónica y síntesis de frecuencias.

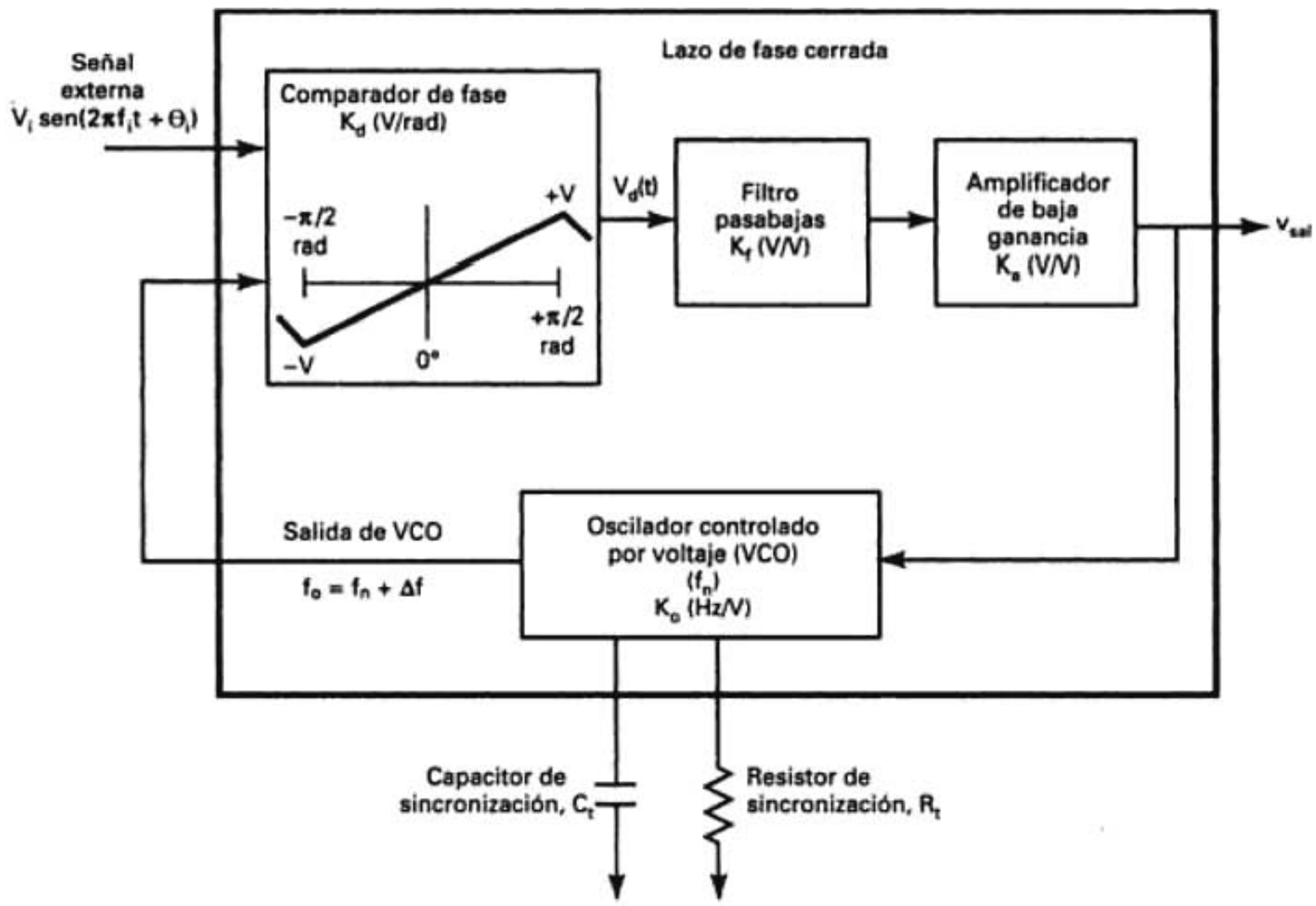


FIGURA 2-21 Diagrama de bloques para el lazo de fase cerrada

En esencia, un PLL es un sistema de control retroalimentado de lazo cerrado en el que la frecuencia de la señal de voltaje retroalimentada es el parámetro de interés, y no sólo el voltaje. El PLL proporciona una sintonía selectiva y filtrado de frecuencia, sin necesidad de bobinas o de inductores. El circuito básico de fase cerrada se ve en la fig. 2-21 y consiste en cuatro bloques primarios: un comparador (mezclador) de fase, un filtro pasabajas, un amplificador de baja ganancia (amplificador operacional) y un oscilador controlado por voltaje (VCO). Cuando no hay señal externa de entrada, el voltaje de salida v_{sal} es igual a cero. El VCO funciona a una frecuencia establecida, llamada su *frecuencia natural* o *de funcionamiento libre* (f_n), que se ajusta con un resistor R_t y un capacitor C_t externos. Si se aplica al sistema una señal de entrada, el comparador de fase compara la fase y la frecuencia de la señal de entrada con la frecuencia natural del VCO y genera un voltaje de error, $V_d(t)$, que se relaciona con la diferencia de fase y frecuencia entre las dos señales. Este voltaje de error se filtra, amplifica y aplica a la terminal de entrada del VCO. Si la frecuencia f_i de entrada es suficientemente cercana a la frecuencia natural del VCO, f_n, la naturaleza de retroalimentación del PLL hace que se sincronice o amarre el VCO con la señal de entrada. Una vez amarrada, la frecuencia del VCO es idéntica a la de la señal de entrada, a excepción de una diferencia finita de fase, que es igual a la fase de la señal de entrada menos la fase de la señal de salida del VCO.

Intervalos de enganche y de captura

Los dos parámetros de los PLL que indican su intervalo de frecuencia útil son el intervalo de enganche y el intervalo de captura.

Intervalo de enganche. Se define al *intervalo de enganche* como el margen de frecuencias cercanas a la frecuencia natural del VCO, f_n, dentro del cual el PLL puede mantener la sincronización con una señal de entrada. Esto presupone que al principio el PLL estaba sincronizado con la entrada. El intervalo de enganche también se llama *intervalo de rastreo*. Es el margen de frecuencias dentro del cual el PLL rastrea o sigue con exactitud a la frecuencia de entrada. El intervalo de enganche aumenta cuando aumenta la ganancia general del lazo del PLL (la ganancia se describe en una sección posterior de este capítulo). El *intervalo de retención* es igual a la mitad del intervalo de

enganche, es decir, intervalo de enganche = 2 × intervalo de retención. La relación entre los intervalos de enganche y de retención se ven en el diagrama de la fig. 2-22. La frecuencia mínima a la que rastrea el PLL se llama *límite inferior de enganche* (f_{ll}), y la frecuencia máxima de rastreo se llama *límite superior de enganche* (f_{lu}). El intervalo de enganche depende de las funciones de transferencia (ganancias) del comparador de fase, del amplificador de baja ganancia y del VCO.

Intervalo de captura. El *intervalo de captura* se define como la banda de frecuencias cercanas a f_n donde el PLL puede establecer o adquirir enganche con una señal de entrada. El intervalo de captura está, en general, entre 1.1 y 1.7 por la frecuencia natural del VCO. El intervalo de captura también se llama *intervalo de adquisición*, y se relaciona con el ancho de banda del filtro de paso bajo, o pasabajas. El intervalo de captura de un PLL disminuye cuando se reduce el ancho de banda del filtro. El semi*intervalo de captura* es el intervalo máximo de captura (es decir, intervalo de captura = 2 × semiintervalo de captura). En el diagrama de frecuencias de la fig. 2-23 se muestran los intervalos y semiintervalos de captura. La frecuencia mínima en la que el PLL se puede sincronizar se llama *límite inferior de captura* (f_{cl}) y la frecuencia máxima a la que se puede enganchar el PLL se llama *límite superior de captura* (f_{cu}).

El intervalo de captura nunca es mayor que, y casi siempre es menor que el intervalo de enganche. La relación entre los intervalos de captura, enganche, retención y semiintervalo de captura se ilustra en el diagrama de frecuencias de la fig. 2-24. Nótese que el intervalo de enganche ≥ intervalo de captura, y que el intervalo de retención ≥ semiintervalo de captura.

Oscilador controlado por voltaje

Un *oscilador controlado por voltaje* (VCO, de *voltage-controlled oscillator*) es un oscilador (en forma más específica, un multivibrador de funcionamiento autónomo) con una frecuencia estable de oscilación, que depende de un voltaje de polarización externo. La salida de un VCO es una frecuencia, y su entrada es una señal de polarización o de control, que puede ser un voltaje de cd o de ca. Cuando se aplica un voltaje de cd o de ca de variación lenta en la entrada del VCO, la

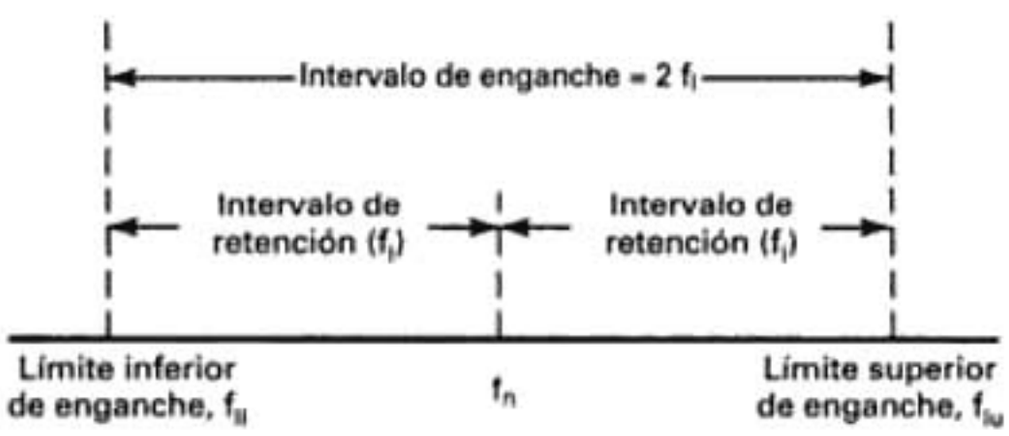


FIGURA 2-22 Intervalo de enganche del PLL

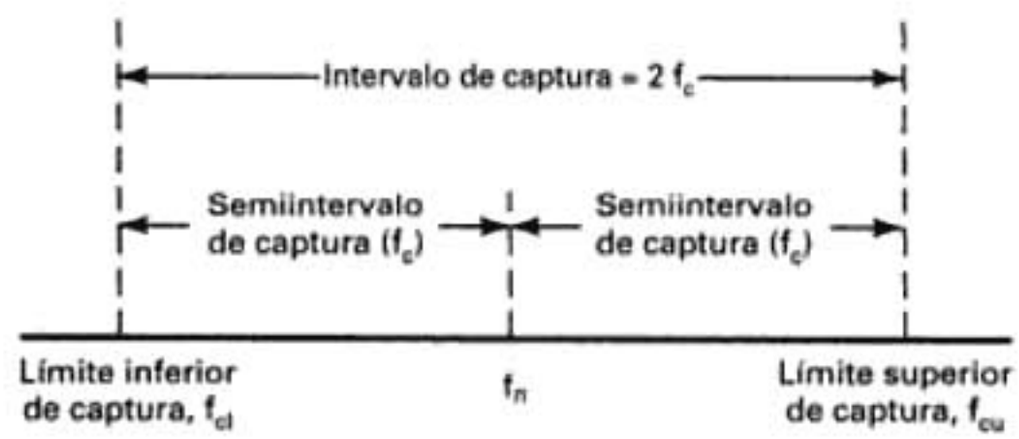


FIGURA 2-23 Intervalo de captura del PLL

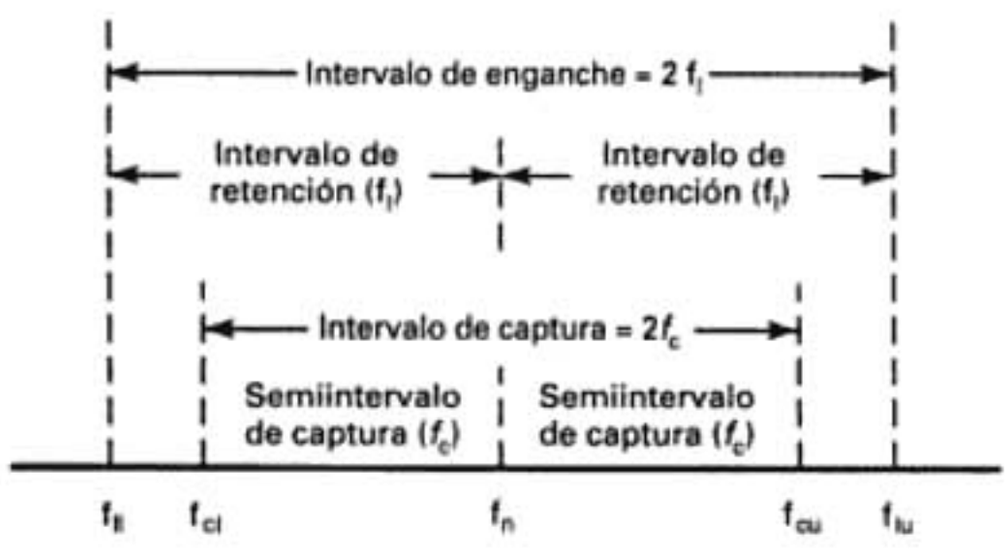


FIGURA 2-24 Intervalos de captura y de enganche del PLL

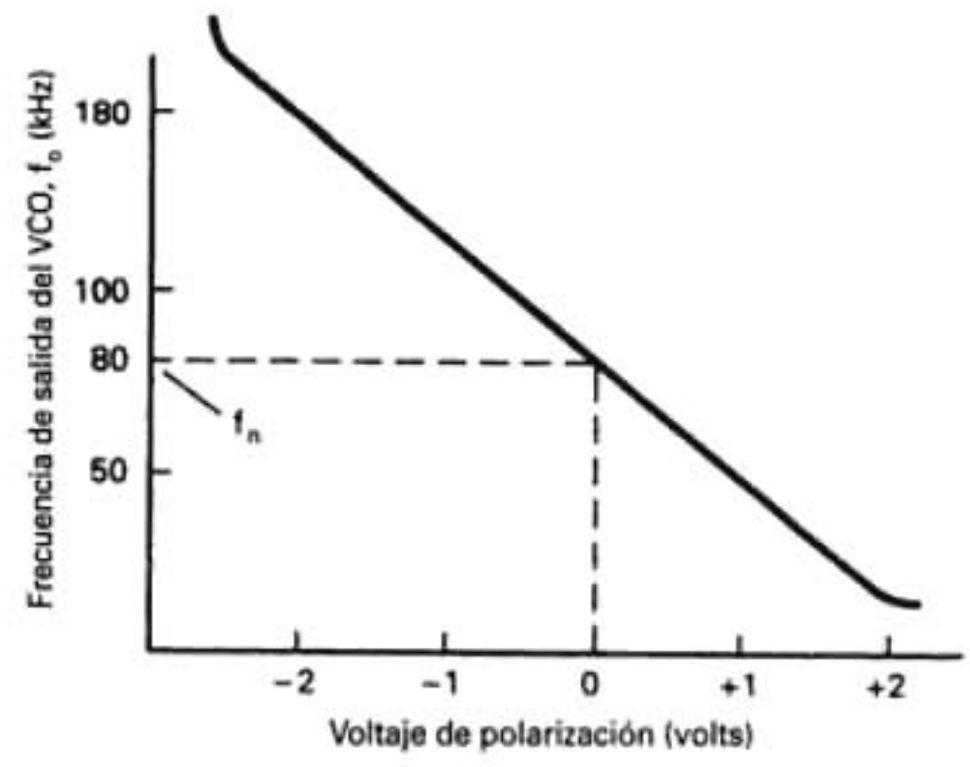


FIGURA 2-25 Característica de la salida de frecuencia de un oscilador controlado por voltaje, en función del voltaje de polarización

frecuencia de salida cambia, o se desvía, en forma proporcional. La fig. 2-25 muestra una curva de transferencia (frecuencia de salida en función de las características del voltaje de polarización en la entrada) de un VCO característico. La frecuencia de salida (f_o) con voltaje de polarización de 0 V en la entrada es la frecuencia natural del VCO, f_n, que está determinada por una red externa *RC*, y el cambio en la frecuencia de salida causado por un cambio de voltaje de entrada se llama desviación de frecuencia, Δf. En consecuencia, $f_o = f_n \times \Delta f$, siendo f_o = la frecuencia de salida del VCO. Para que haya una Δf simétrica, la frecuencia natural del VCO debe estar centrada en la parte lineal de la curva de entrada-salida. La función de transferencia de un VCO es

$$K_o = \frac{\Delta f}{\Delta V} \tag{2-10}$$

en la que K_o = función de transferencia de entrada-salida (hertz por volt)
ΔV = cambio de voltaje de control en la entrada (volts)
Δf = cambio en la frecuencia de salida (hertz)

Comparador de fases

Un comparador de fases, que a veces se le dice *detector de fase*, es un dispositivo no lineal con dos señales de entrada: una frecuencia generada externamente (f_i) y la frecuencia de salida del VCO (f_o). La salida de un comparador de fase es el producto de las dos señales con frecuencias f_i y f_o y, por consiguiente, contiene sus frecuencias de suma y de diferencia ($f_i \pm f_o$). El tema de mezclado se analizará con más detalle más adelante en este capítulo. La fig. 2-26a muestra el diagrama de un comparador de fases sencillo. El voltaje v_o se aplica al mismo tiempo a las dos mitades del transformador T_1 de entrada. Los componentes D_1, R_1 y C_1 forman un rectificador de media onda, al igual que D_2, R_2 y C_2. Nótese que $C_1 = C_2$, y que $R_1 = R_2$. Durante la parte positiva de v_o, D_1 y D_2 se polarizan directamente y están activos o *encendidos*, cargando a C_1 y a C_2 con valores iguales, pero con polaridades opuestas. Por lo anterior, el voltaje promedio de salida es $V_{sal} = V_{C1} + (-V_{C2}) = 0$ V. Esto se ve en la fig. 2-26b. Durante el medio ciclo negativo de v_o, D_1 y D_2 tienen polarización inversa y están *apagados*. Por consiguiente, C_1 y C_2 se descargan por igual a través de R_1 y R_2, respectivamente, manteniendo el voltaje de salida igual a 0 V. Esto se ve en la fig. 2-26c. Los dos rectificadores de media onda producen voltajes de salida de igual magnitud y polaridad inversa. Así, el voltaje de salida debido a v_o es constante e igual a 0 V. Las formas de onda correspondientes en la entrada y la salida, para una señal de onda cuadrada del VCO, se ven en la fig. 2-26d.

Funcionamiento del circuito. Cuando se aplica una señal externa de entrada [$v_{ent} = V_i$ sen($2\pi f_i t$)] al comparador de fases, su voltaje se suma a v_o, haciendo que C_1 y C_2 se carguen y descarguen para producir un cambio proporcional en el voltaje de salida. La fig. 2-27a muestra la forma de onda de salida sin filtrar, sombreada, cuando $f_o = f_i$ y v_o se adelanta 90° a v_i. Para que funcione bien el comparador de fase, v_o debe ser mucho mayor que v_i. Así, D_1 y D_2 se *encienden* sólo durante la parte positiva de v_o y están *apagados* durante la parte negativa. Durante la primera mitad del tiempo *encendido*, o *tiempo de trabajo*, el voltaje aplicado a D_1 es igual a $v_o - v_i$, y el aplicado a D_2 es igual a $v_o + v_i$. Por consiguiente, C_1 descarga mientras C_2 carga. Durante la segunda mitad del tiempo *encendido*, el voltaje aplicado a D_1 es igual a $v_o + v_i$, y el aplicado a D_2 es igual a $v_o - v_i$, y C_1 carga mientras que C_2 descarga. Durante el tiempo *apagado*, C_1 y C_2 no cargan ni descargan. Para cada ciclo completo de v_o, C_1 y c_2 cargan y descargan por igual y el voltaje promedio de salida permanece en 0 V. Por lo anterior, el valor promedio de V_{sal} es 0 V cuando las señales de entrada y de salida del VCO tienen igual frecuencia y están desfasadas 90°.

La fig. 2-27b muestra la forma de onda de voltaje de salida, sin filtrar y sombreada, cuando v_o precede 45° a v_i. Este voltaje v_i es positivo durante el 75% del tiempo *encendido*, y negativo durante el 25% restante. En consecuencia, la salida promedio de voltaje durante un ciclo de v_o es positiva, y aproximadamente igual a 0.3 *V*, donde *V* es el voltaje máximo en la entrada. La fig. 2-27c muestra la forma de onda sin filtrar cuando v_o y v_i están enfasados. Durante todo el tiempo *encendido*, v_i es positivo. En consecuencia, el voltaje de salida es positivo y aproximadamente igual a 0.636 V. Las figs. 2-27d y 2-27e muestran la forma de onda de salida, sin filtrar, cuando v_o precede 135° y 180° a v_i, respectivamente. Se puede ver que el voltaje de salida pasa

a negativo cuando v_o se adelante más de 90°, y llega a su valor máximo cuando el adelanto es de 180°. En esencia, un comparador de fases rectifica a v_i y lo integra para producir un voltaje de salida que es proporcional a la diferencia de fases entre v_o y v_i.

La fig. 2-28 muestra las características de voltaje de salida en función de diferencia de fases, para el comparador de fases de la fig. 2-26a. La fig. 2-28a representa la curva del comparador

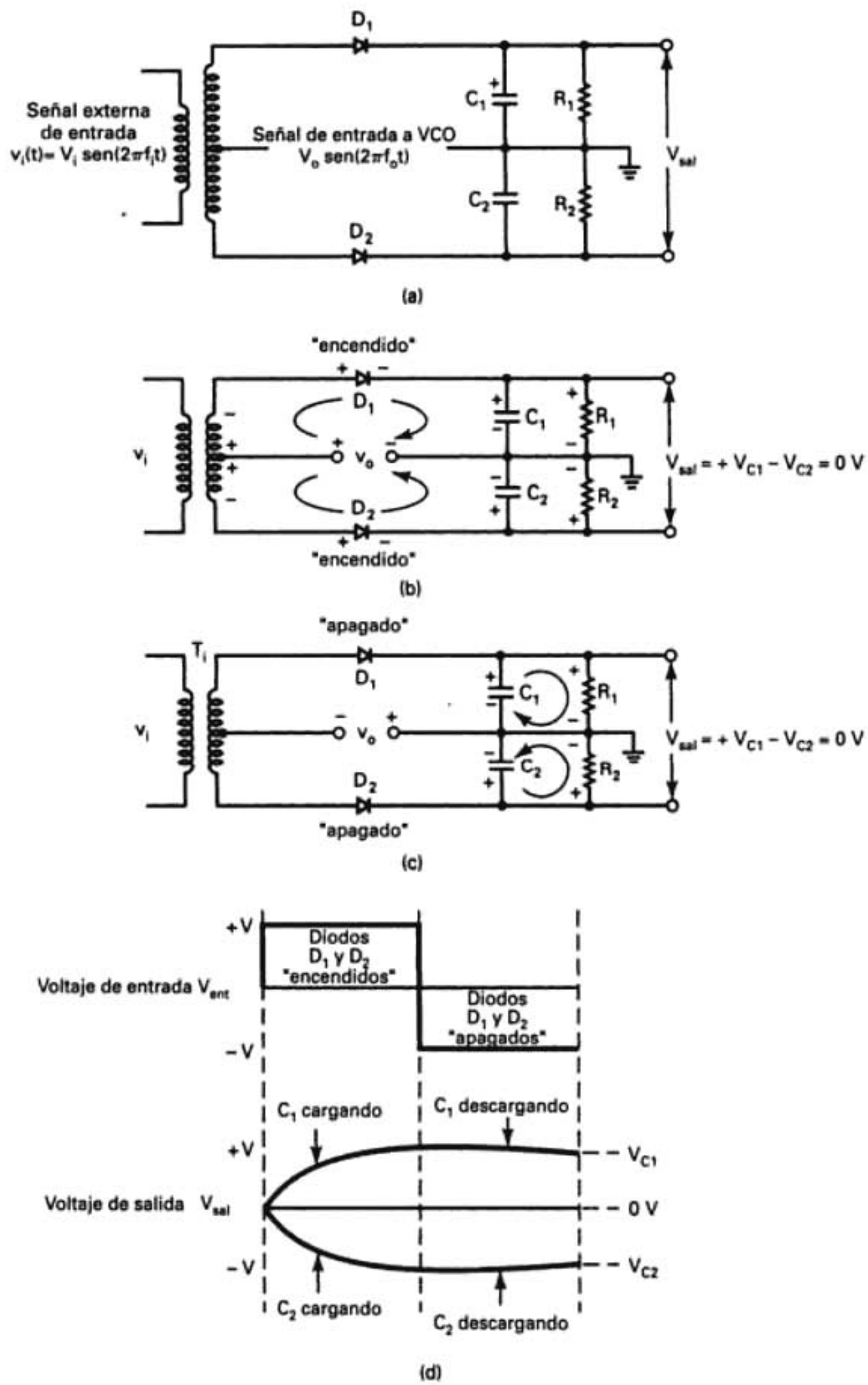
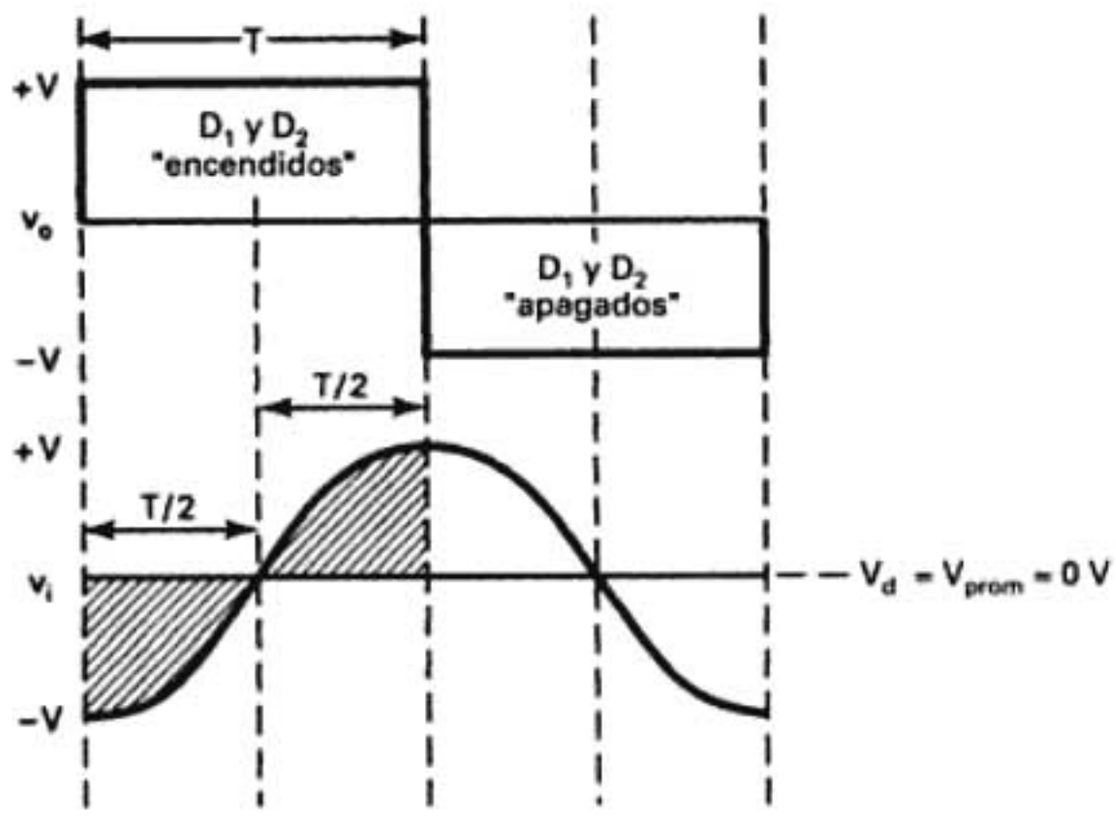


FIGURA 2-26 Comparador de fases: (a) esquema; (b) voltaje de salida debido al medio ciclo positivo de v_o; (c) voltaje de salida debido al medio ciclo negativo de v_o; (d) formas de onda de voltaje de entrada y salida

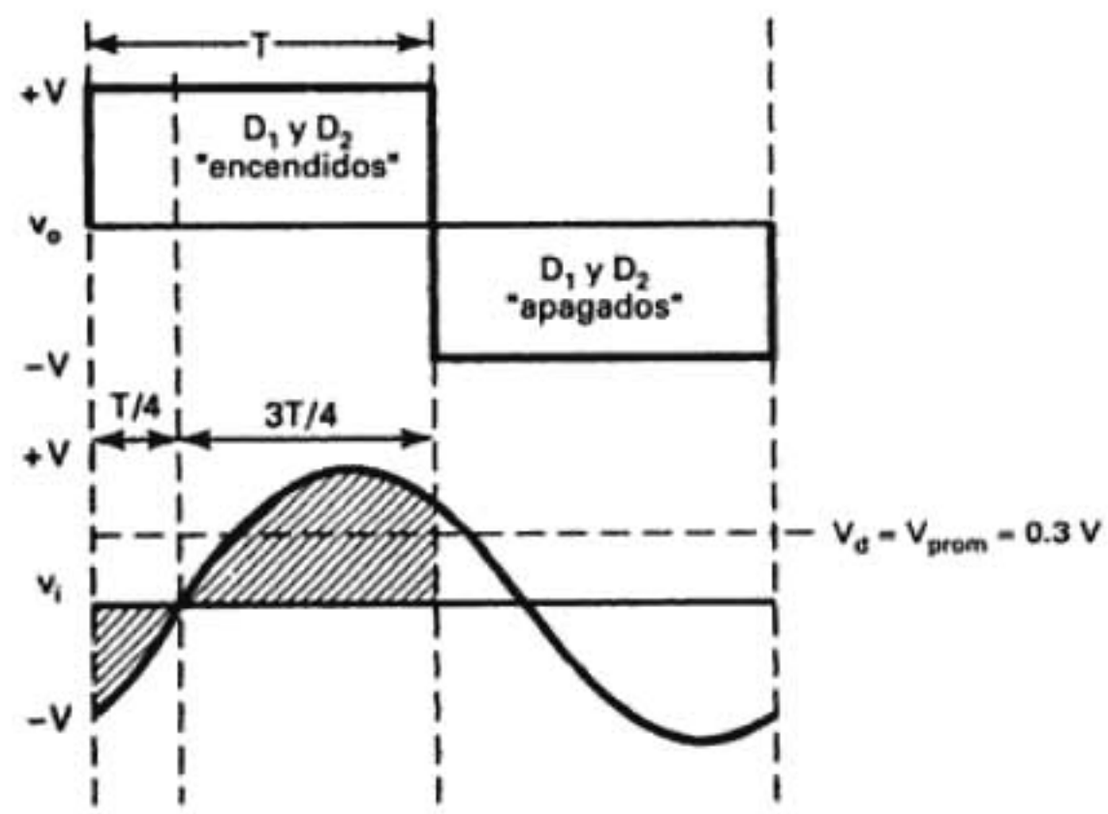
de fases de onda cuadrada. Esta curva tiene forma triangular, con pendiente negativa de 0° a 180°. El voltaje V_{sal} es máximo positivo cuando v_o y v_i están en fase, es 0 V cuando v_o antecede 90° a v_i y es máximo negativo cuando v_o antecede 180° a v_i. Si v_o se adelanta más de 180°, el voltaje de salida se hace menos negativo, y si v_o se atrasa a v_i, el voltaje de salida se hace menos positivo. Por consiguiente, la diferencia máxima de fases que puede detectar el comparador es 90° ± 90°, o sea, de 0° a 180°. Este comparador de fases produce un voltaje de salida que es proporcional a la diferencia de fases entre v_o y v_i. Esta diferencia de fases se llama *error de fase*. La representación matemática del error de fase es

$$\theta_e = \theta_i - \theta_o \tag{2-11}$$

en donde θ_e = error de fase (radianes)
θ_o = fase del voltaje de señal de salida del VCO (radianes)
θ_i = fase del voltaje de señal externa de entrada (radianes)

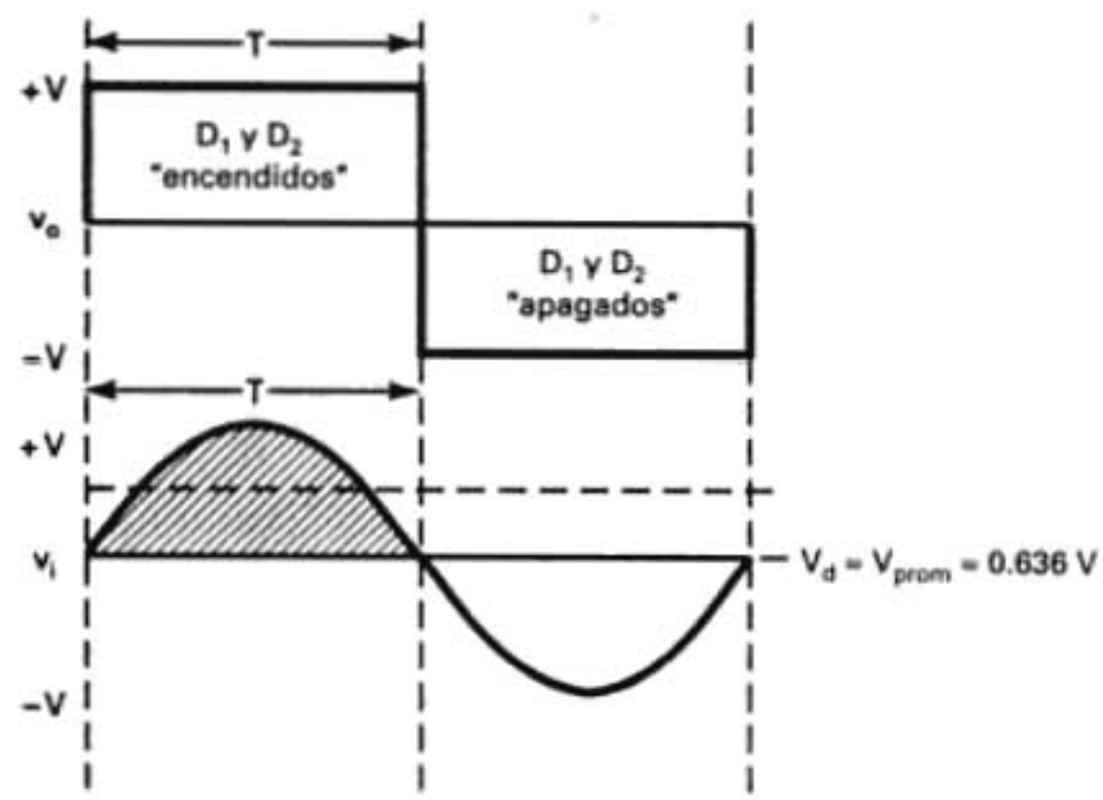


(a)

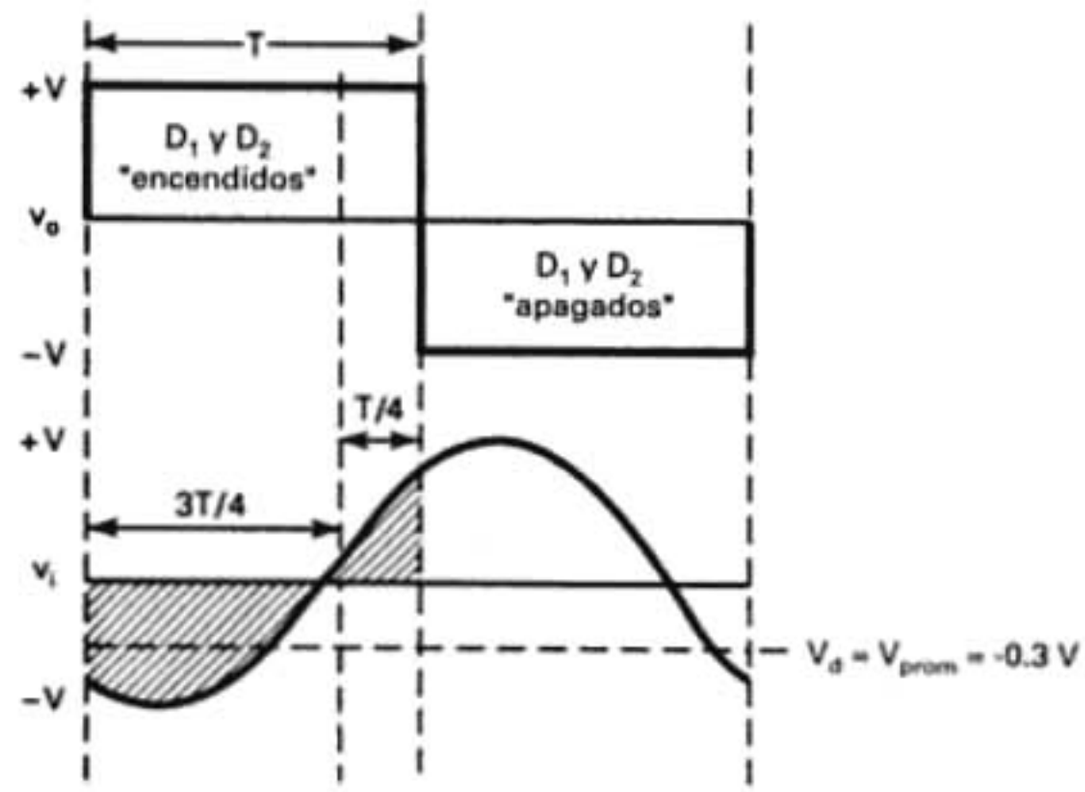


(b)

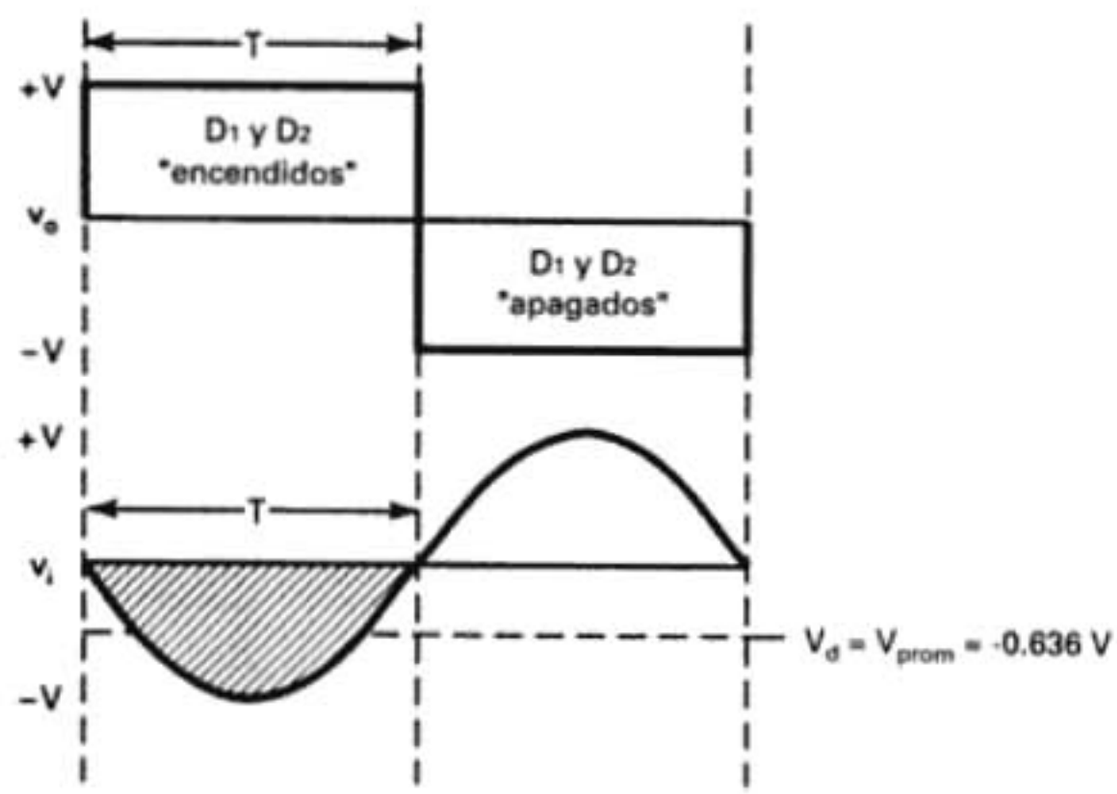
FIGURA 2-27 Formas de onda de voltaje de salida en el comparador de fases: (a) v_o adelantado 90° a v_i; (b) v_o adelantado 45° a v_i; (*continúa*)



(c)



(d)



(e)

FIGURA 2-27 (*Continuación*) (c) v_o y v_i en fase; (d) v_o adelantado 135° a v_i; (e) v_o adelantado 180° a v_i

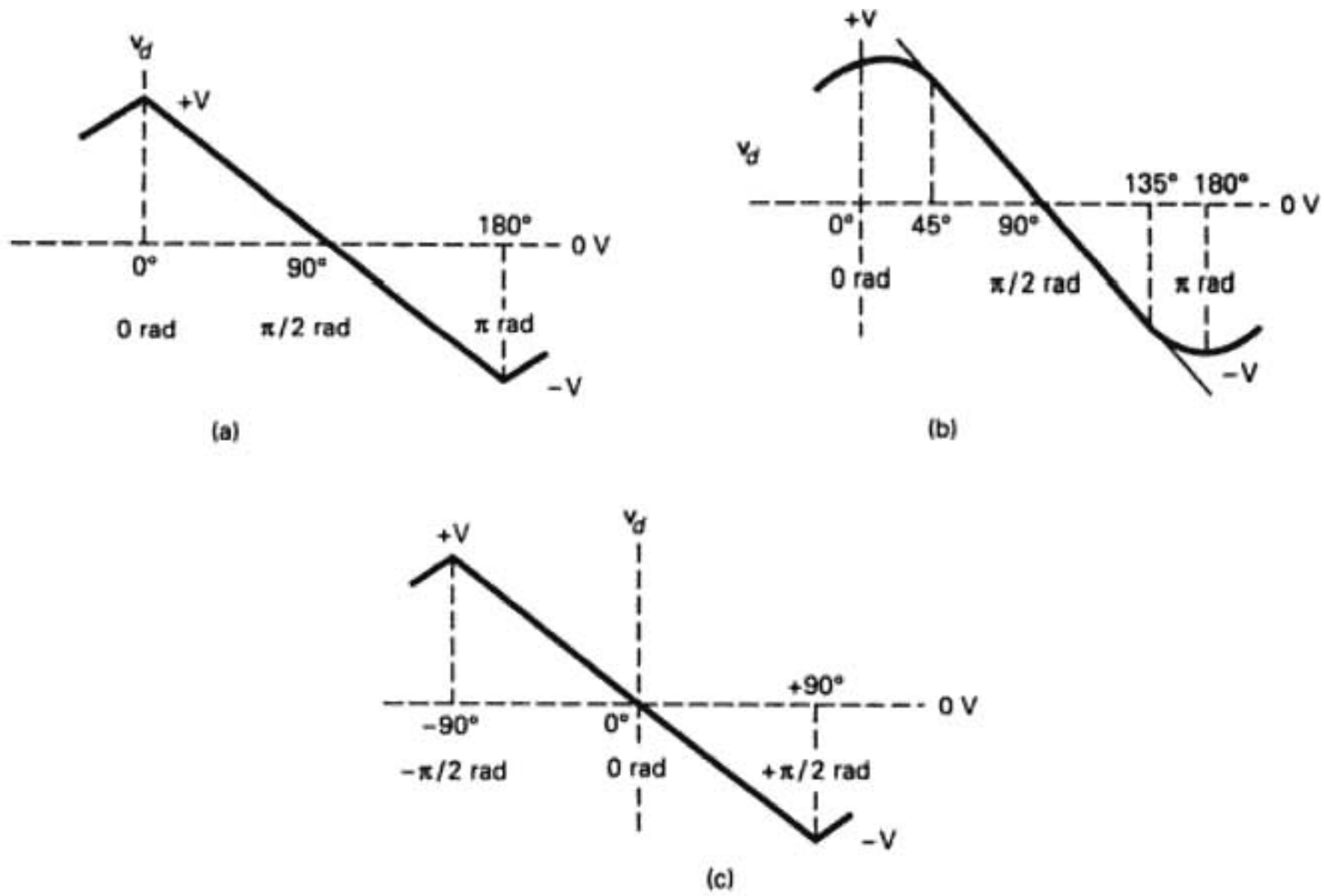


FIGURA 2-28 Voltaje de salida (V_d) del comparador de fase en función de la diferencia de fase (θ_e): (a) entradas de onda cuadrada; (b) entradas senoidales; (c) entradas de onda cuadrada, referencia de polarización de fase

El voltaje de salida del comparador de fases es lineal para errores de fase de 0 a 180° (de 0 a π radianes). Entonces, la función de transferencia de un comparador de fases de onda cuadrada, para errores entre 0 y 180° es

$$K_d = \frac{V_d}{\theta_e} = \frac{2v_i}{\pi} \tag{2-12}$$

siendo K_d = función de transferencia o ganancia (volts por radián)
V_d = voltaje de salida del comparador de fases (volts)
θ_e = error de fase ($\theta_i - \theta_o$) (radianes)
π = 3.14 radianes
v_i = voltaje máximo de la señal de entrada (volts)

La fig. 2-28b muestra la curva de voltaje de salida en función de diferencia de fases en la entrada, de un comparador analógico de fases con características senoidales. El error de fase en función de la salida es casi lineal sólo de 45° a 135°. En consecuencia, la función de transferencia es

$$K_d = \frac{V_d}{\theta_e} \text{ volts por radián} \tag{2-13a}$$

donde K_d = función de transferencia o ganancia (volts por radián)
θ_e = error de fase ($\theta_i - \theta_o$) (radianes)
V_d = voltaje de salida del comparador de fases (volts)

Se puede reordenar la ecuación 2-13a, para despejar a V_d

$$V_d = K_d\theta_e \tag{2-13b}$$

Se puede ver en las figuras 2-28a y b, que el voltaje de salida del comparador de fases es V_{sal} = 0 V cuando $f_o = f_i$, y v_o y v_i están desfasados 90°. Por consiguiente, si la frecuencia de entrada f_i es inicialmente igual a la frecuencia natural del VCO, f_n, se requiere una diferencia de

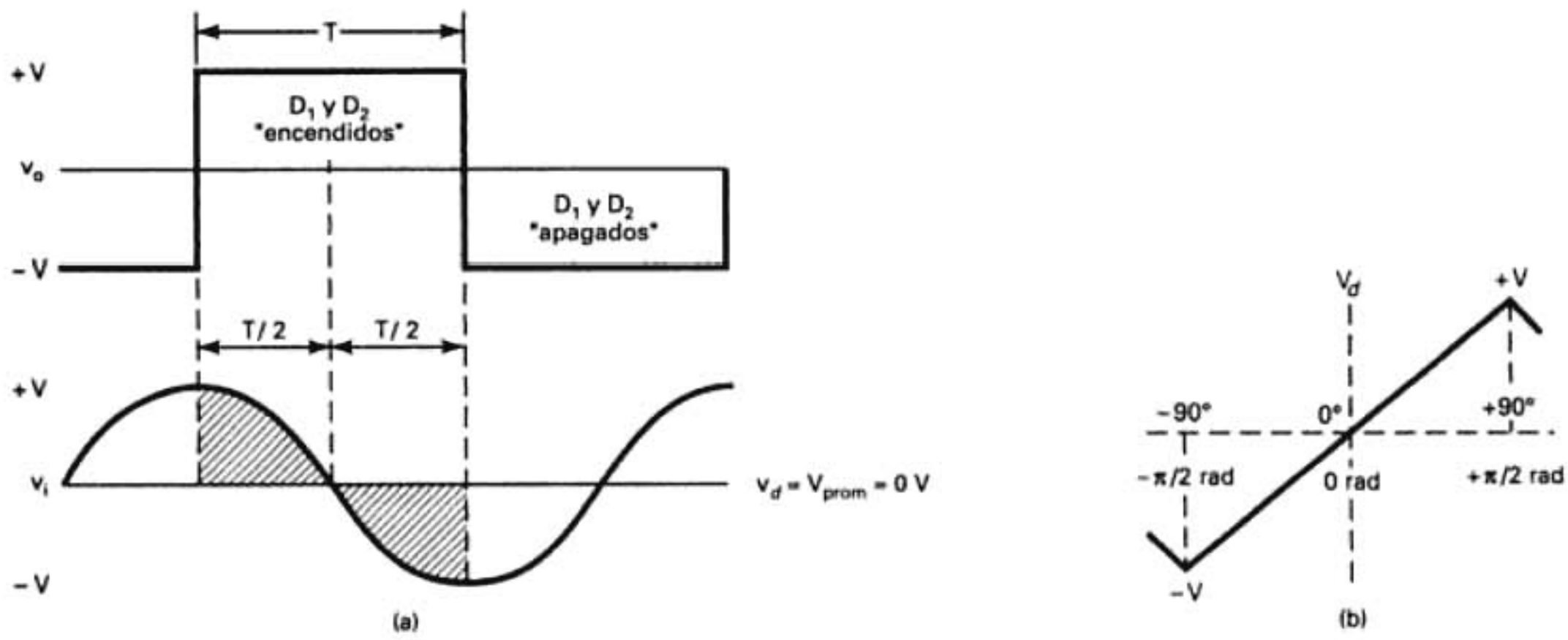


FIGURA 2-29 Voltaje de salida de un comparador de fase: (a) forma de onda de voltaje de salida sin filtrar, cuando v_i se adelanta 90° a v_o; (b) características de voltaje de salida en función de diferencia de fases

fases para mantener en 0 V el voltaje de salida del comparador de fases, y la frecuencia de salida del VCO igual a su frecuencia natural ($f_o = f_n$). Esta diferencia de fases de 90° equivale a una diferencia o desviación de fase. En general, se considera como fase de referencia a la desviación de fases, que se puede variar en $\pm\pi/2$ radianes ($\pm 90°$). Por consiguiente, V_{sal} pasa desde su valor positivo máximo en $-\pi/2$ radianes ($-90°$) hasta su valor negativo máximo en $+p/2$ radianes ($+90°$). La fig. 2-28c muestra las características del voltaje de salida del comparador de fases en función de las características de error de fase, para entradas de onda cuadrada con la diferencia de fases de 90° como referencia.

La fig. 2-29 representa el voltaje de salida no filtrado cuando v_i precede 90° a v_o. Nótese que el valor promedio es 0 V (igual que cuando v_o precedía 90° a v_i). Cuando sucede el enganche de frecuencias, no está definido si el VCO se engancha a la frecuencia de entrada con una diferencia de fases de 90° positiva o negativa. Por consiguiente, hay una ambigüedad de 180° en la fase de la frecuencia de salida del VCO. La fig. 2-29b muestra la curva de voltaje de salida en función de diferencia de fases para entradas de onda cuadrada, cuando la frecuencia de salida del VCO es igual a su frecuencia natural, y se ha enganchado a la señal de entrada con una diferencia de fases de $-90°$. Nótese que cuando el error de fase tiene dirección opuesta, se producen voltajes opuestos, y la pendiente es positiva, más que negativa, desde $-\pi/2$ radianes hasta $+\pi/2$ radianes. Cuando hay enganche de frecuencia, el PLL produce una frecuencia coherente ($f_o = f_i$), pero la fase de la señal de entrada es incierta, ya que f_o precede a f_i en $90° \pm \theta_e$ o viceversa.

Funcionamiento del lazo

Véase la fig. 2-30 para comprender las siguientes explicaciones.

Captura del lazo. Una señal externa de entrada [$(V_i \operatorname{sen}(2\pi f_i t + \theta_i)$] entra al comparador de fases y se mezcla con la señal de salida del VCO, que es una onda cuadrada con frecuencia fundamental f_o. Al principio, las dos frecuencias no son iguales ($f_o \neq f_i$), y el lazo está *desincronizado*. Como el comparador de fases es un dispositivo no lineal, las señales del VCO y de entrada se mezclan, y generan frecuencias de producto cruzado (es decir, suma de frecuencias y diferencia de frecuencias). Así, las frecuencias primarias de salida del comparador de fases son la frecuencia de la entrada externa, f_i, la frecuencia de salida del VCO, f_o, y su suma ($f_i + f_o$) y su diferencia ($f_i - f_o$).

El filtro pasabajas (LPF) bloquea las dos frecuencias originales de entrada, y la frecuencia de suma; así, la entrada al amplificador es sólo la diferencia de frecuencias, $f_i - f_o$. La diferencia de frecuencias se amplifica, para aplicarla a la entrada del oscilador controlado por voltaje,

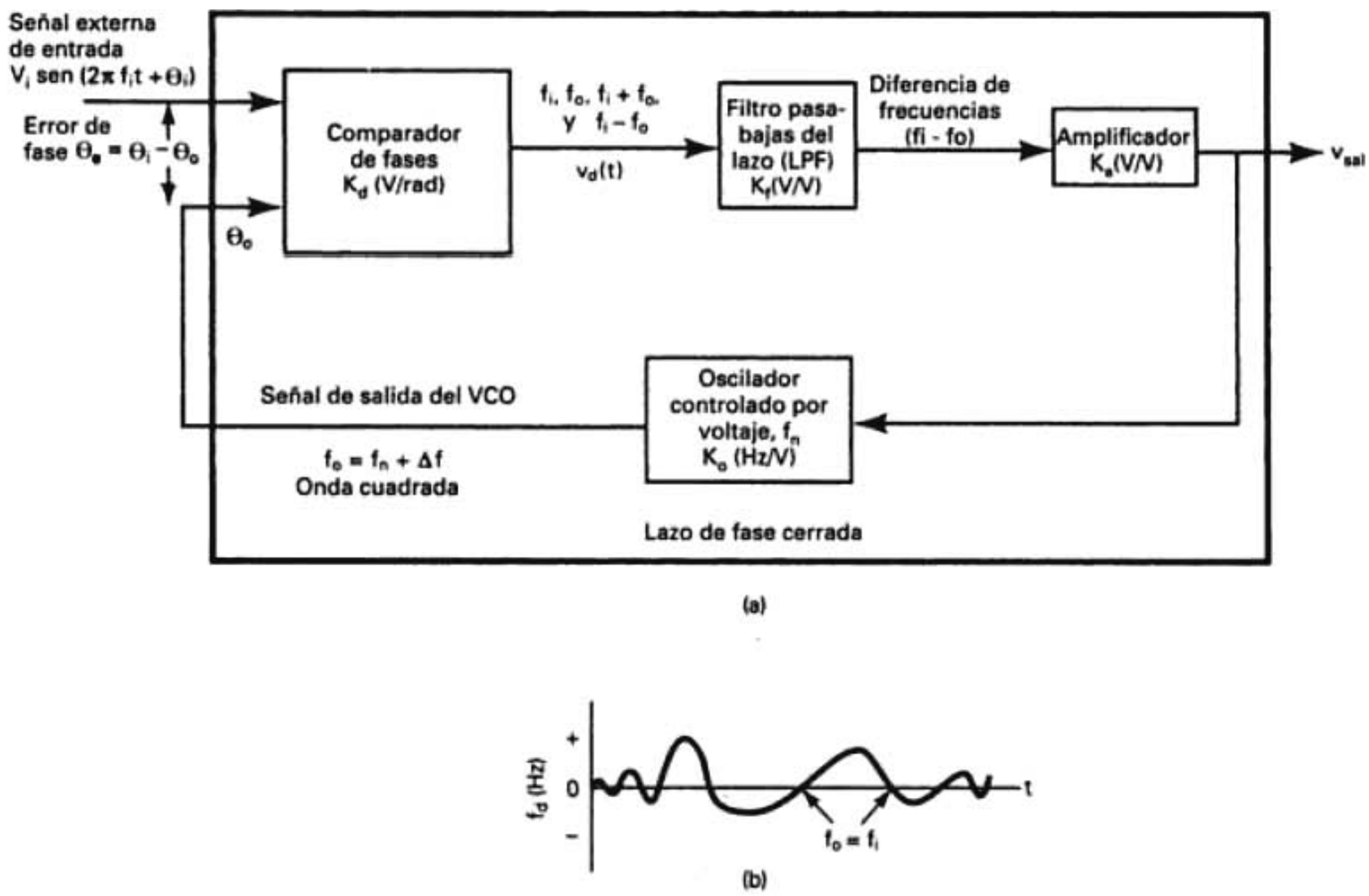


FIGURA 2-30 Funcionamiento del PLL: (a) diagrama de bloques; (b) diferencia de frecuencias

el cual desvía una cantidad proporcional a su polaridad y amplitud. Al cambiar la frecuencia de salida del VCO, la amplitud y la frecuencia de la diferencia de frecuencias cambian en forma proporcional. La fig. 2-30b muestra la diferencia de frecuencias que se produce cuando el VCO es barrido por la diferencia de frecuencias, f_d. Después de algunos ciclos en torno al lazo, la frecuencia de salida del VCO iguala a la frecuencia externa de entrada, y se dice que el lazo está enganchado. Una vez enganchado, la diferencia de frecuencias en la salida del filtro pasabajas es 0 Hz (es un voltaje de cd), que es necesario para polarizar al VCO y mantenerlo enganchado a la frecuencia de entrada externa. En esencia, el comparador de fases es un comparador de frecuencias hasta que se alcanza la frecuencia de captura (cero diferencia), y en adelante se vuelve un comparador de fases. Una vez enganchado el lazo, la diferencia de fases entre la entrada externa y la salida del VCO se convierte en un voltaje de polarización (V_d) en el comparador de fases, se amplifica y después se retroalimenta al VCO para sujetar la sincronía. Por lo anterior, es necesario mantener un error de fase entre la señal de entrada externa y la señal de salida del VCO. El cambio de frecuencia en el VCO es necesario para lograr el enganche, y el tiempo necesario para llegar al mismo (*tiempo de adquisición*) de un PLL sin filtro de lazo (los filtros de lazo se explicarán después en este capítulo) es más o menos igual a $5/K_v$ segundos, donde K_v es la ganancia del PLL en lazo abierto. Una vez enganchado el lazo, todo cambio en la frecuencia de entrada se considera como error de fase, y el comparador produce un cambio en V_d, su voltaje de salida. El cambio de voltaje se amplifica y se retroalimenta al VCO para reestablecer el enganche. Así, el lazo se autoajusta en forma dinámica y sigue los cambios en la frecuencia de entrada.

La ecuación matemática que describe la salida del comparador de fases es (si sólo se considera la frecuencia fundamental en V_o y si se excluye la diferencia de fase de 90°)

$$V_d = [V \operatorname{sen}(2\pi f_o t + \theta_o) \times V \operatorname{sen}(2\pi f_i t + \theta_i)]$$

$$= \frac{V}{2}\cos(2\pi f_o t + \theta_o - 2\pi f_i t - \theta_i) - \frac{V}{2}\cos(2\pi f_o t + \theta_o + 2\pi f_i t - \theta_i)$$

donde V_d = voltaje de salida del detector de fase (volts)
$V = V_o V_i$ (volts máximos)

Cuando $f_o = f_i$,

$$V_d = \frac{V}{2}\cos(\theta_i + \theta_o) \tag{2-14}$$
$$= \frac{V}{2}\cos\theta_e$$

en donde $\theta_i + \theta_o = \theta_e$ (error de fase). El ángulo θ_e es el error de fase que se requiere para cambiar la frecuencia de salida del VCO de f_n a f_i, un cambio igual a Δf, que a veces se llama *error estático de fase*.

Ganancia del lazo. La *ganancia de lazo* para un PLL es sólo el producto de las ganancias o funciones de transferencia individuales en torno al lazo. En la fig. 2-30, la ganancia de lazo abierto es el producto de las ganancias del comparador de fases, del filtro pasabajas, del amplificador y del VCO. Esto se expresa matemáticamente como sigue

$$K_L = K_d K_f K_a K_o \tag{2-15a}$$

donde K_L = ganancia de lazo abierto del PLL (hertz por radián)
K_d = ganancia del comparador de fases (volts por radián)
K_f = ganancia del filtro pasabajas (volts por volt)
K_a = ganancia del amplificador (volts por volt)
K_o = ganancia del VCO (hertz por volt)

y $$K_L = \frac{(\text{volt})(\text{volt})(\text{volt})(\text{hertz})}{(\text{rad})(\text{volt})(\text{volt})(\text{volts})} = \frac{\text{hertz}}{\text{rad}}$$

o bien, la ganancia del PLL con lazo abierto (K_v), en radianes/segundo (s^{-1}) es

$$K_v = \frac{\text{ciclos}/s}{\text{rad}} = \frac{\text{ciclos}}{\text{rad-s}} \times \frac{2\pi\ \text{rad}}{\text{ciclo}} = 2\pi K_L \tag{2-15b}$$

Expresada en decibeles es

$$K_{v(\text{dB})} = 20 \log K_v \tag{2-16}$$

Partiendo de las ecuaciones 2-10, 2-13b y 2-16, se deducen las siguientes relaciones

$$V_d = (\theta_e)(K_d)\ \text{volts} \tag{2-17}$$
$$V_{\text{sal}} = (V_d)(K_f)(K_a)\ \text{volts} \tag{2-18}$$
$$\Delta f = (V_{\text{sal}})(K_o)\ \text{hertz} \tag{2-19}$$

Como se dijo antes, el intervalo de retención de un PLL es el de las frecuencias de entrada en las que el PLL permanece enganchado. Esto supone que antes estaba enganchado el PLL. El intervalo de retención está limitado por la oscilación, de pico a pico, del voltaje de salida del comparador de fases, ΔV_d, y depende de las funciones de transferencia del comparador de fases, el amplificador y el VCO. Se puede ver en la fig. 2-28c que el voltaje de salida V_d del comparador de fases es correctivo para $\pm\pi/2$ radianes ($\pm 90°$). Fuera de esos límites, la polaridad de V_d se invierte y en realidad persigue a la frecuencia del VCO al alejarse de la frecuencia de entrada externa. Por consiguiente, el error máximo de fase (θ_e) que se permite es $\pm\pi/2$ radianes, y el voltaje máximo de salida del comparador de fases es

$$\pm V_{d(\text{máx})} = [\theta_{e(\text{máx})}](K_d) \tag{2-20a}$$
$$= \pm\left(\frac{\pi}{2}\ \text{rad}\right)(K_d) \tag{2-20b}$$

donde $\pm V_{d(\text{máx})}$ = cambio máximo entre picos de voltaje de salida del comparador de fases
K_d = función de transferencia del comparador de fases

En consecuencia, el cambio máximo de la frecuencia de salida del VCO es

$$\pm\Delta f_{\text{máx}} = \pm\left(\frac{\pi}{2}\text{ rad}\right)(K_d)(K_f)(K_a)(K_o) \tag{2-21}$$

en la que $\pm\Delta f_{\text{máx}}$ es el intervalo de retención (cambio máximo entre picos en la frecuencia de salida del VCO).

Se sustituye K_L por $K_dK_fK_aK_o$ para obtener

$$\pm\Delta f_{\text{máx}} = \pm\left(\frac{\pi}{2}\text{ rad}\right)K_L \tag{2-22}$$

Ejemplo 2-2

Para el PLL de la fig. 2-30, la frecuencia natural de un VCO es f_n = 200 kHz, la frecuencia de entrada externa es f_i = 210 kHz y las funciones de transferencia son K_d = 0.2 V/rad, K_f = 1, K_a = 5 y K_o = 20 kHz/V. Calcular lo siguiente:

(a) Ganancia del PLL con lazo abierto, en Hz/rad y en rad/s.
(b) Cambio necesario en la frecuencia del VCO para llegar al amarre (Δf).
(c) Voltaje de salida del PLL (V_{sal}).
(d) Voltaje de salida del detector de fase (V_d).
(e) Error estático de fase (θ_e).
(f) Intervalo de retención ($\Delta f_{\text{máx}}$).

Solución (a) De las ecuaciones 2-15a y 2-15b,

$$K_L = \frac{0.2\text{ V}}{\text{rad}}\frac{1\text{ V}}{\text{V}}\frac{5\text{ V}}{\text{V}}\frac{20\text{ kHz}}{\text{V}} = \frac{20\text{ kHz}}{\text{rad}}$$

$$K_v = 2\pi K_L = \frac{20\text{ kHz}}{\text{rad}} = \frac{20\text{ kilociclos}}{\text{rad - s}} \times \frac{2\pi\text{ rads}}{\text{ciclo}} = 125{,}600\text{ rad /sec}$$

$$K_{v(\text{dB})} = 20\log 125{,}600\text{ k} = 102\text{ dB}$$

(b)
$$\Delta f = f_i - f_n = 210\text{ kHz} - 200\text{ kHz} = 10\text{ kHz}$$

(c) Se arregla la ecuación 2-10 como sigue

$$V_{\text{sal}} = \frac{\Delta f}{K_o} = \frac{10\text{ kHz}}{20\text{ kHz/V}} = 0.5\text{V}$$

(d)
$$V_d = \frac{V_{\text{sal}}}{(K_f)(K_a)} = \frac{0.5}{(1)(5)} = 0.1\text{ V}$$

(e) Rearreglando la ecuación 2-13B se obtiene

$$\theta_e = \frac{V_d}{K_d} = \frac{0.1\text{ V}}{0.2\text{ V/rad}} = 0.5\text{ rad} \quad \text{o} \quad 28.65°$$

(f) Se sustituyen valores en la ecuación 2-22

$$\Delta f_{\text{máx}} = \frac{(\pm\pi/2\text{ rad})(20\text{ kHz})}{\text{rad}} = \pm 31.4\text{ kHz}$$

El intervalo de enganche es el intervalo de frecuencias en el cual el lazo permanece enganchado, o cerrado, con la señal externa de entrada una vez establecido el enganche. Este intervalo de enganche se expresa en rad/s y se relaciona con la ganancia de lazo abierto, K_v, como sigue

$$\text{intervalo de enganche} = 2\Delta f_{\text{máx}} = \pi K_L$$

en donde $K_L = (K_d)(K_f)(K_o)$ para un lazo sencillo con un LPF (filtro lineal de fase), un comparador de fase y un VCO, o bien $K_L = (K_d)(K_f)(K_a)(K_o)$ para un lazo con amplificador.

El intervalo de enganche, en radianes por segundo, es igual a π por la ganancia de voltaje del lazo, y es independiente de la respuesta del LPF. El intervalo de captura depende del intervalo de enganche y de la respuesta del LPF, por lo que cambia con el tipo de filtro usado y con la frecuencia de corte del filtro. Para un LPF *RC* sencillo de un polo, se determina con

$$\text{intervalo de captura} = \frac{2\sqrt{\Delta f_{\text{máx}}}}{RC}$$

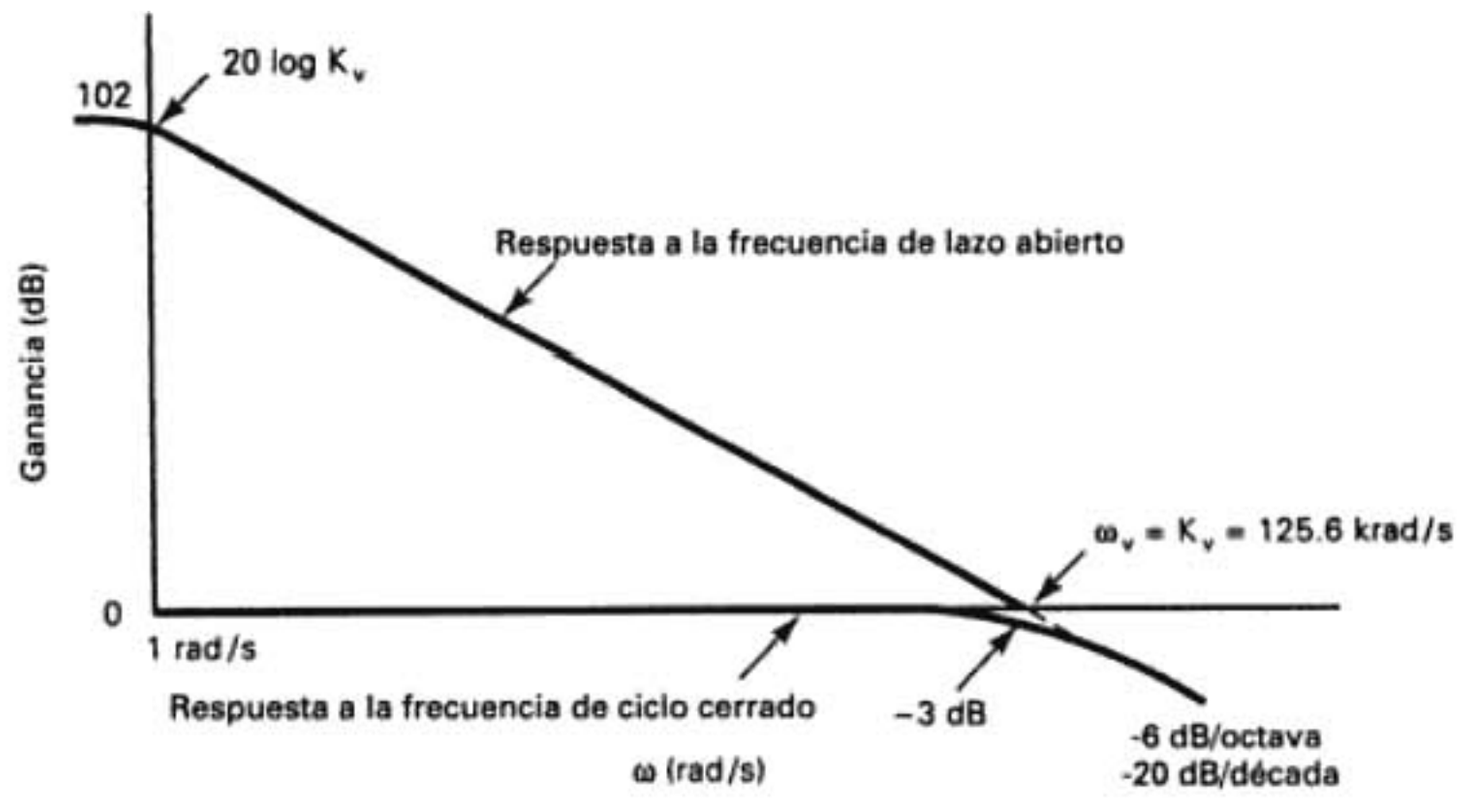


FIGURA 2-31 Respuesta en frecuencia de un lazo de fase cerrada

Respuesta a la frecuencia en ciclo cerrado. La *respuesta a la frecuencia en lazo cerrado* para un PLL *no compensado* (no filtrado) se ve en la fig. 2-31. La ganancia en lazo abierto de un PLL para una frecuencia de 1 rad/s es K_v. La respuesta a la frecuencia de la fig. 2-31 corresponde al circuito y a los parámetros del PLL del ejemplo 2-2. Se puede ver que la ganancia en lazo abierto, K_v, a 1 rad/s es 102 dB, y la ganancia en lazo abierto es igual a 0 dB en la frecuencia de corte del lazo (ω_v). También, la ganancia de lazo cerrado es la unidad hasta ω_v, donde baja a −3 dB y continúa alejándose a 6 dB/octava (20 dB/década). También, $\omega_v = K_v = 125.6$ krad/s, que es el ancho de banda unilateral del lazo cerrado no compensado.

Se ve en la fig. 2-31 que la repuesta en frecuencia de un PLL no compensado es idéntica a la de un filtro pasabajas de un polo (primer orden) con una frecuencia de corte de $\omega_c = 1$ rad/s. En esencia, un PLL es un filtro de rastreo de paso bajo que sigue a los cambios de la frecuencia de entrada que están dentro de un ancho de banda igual a $\pm K_v$.

Si se requiere más limitación de banda, se puede agregar un filtro pasabajas entre el comparador de fases y el amplificador, como se ve en la fig. 2-30. Este filtro puede ser de uno o varios polos. La fig. 2-32 muestra la respuesta en frecuencia del lazo para un filtro *RC* sencillo de un polo, con frecuencia de corte $\omega_c = 100$ rad/s. La respuesta en frecuencia sigue a la de la fig. 2-31 hasta la frecuencia de corte del filtro del lazo; a continuación la respuesta se aparta a 12 dB/octava (40 dB/década). En consecuencia, la frecuencia compensada con ganancia unitaria (ω_c') se reduce a unos ±3.5 krad/s.

Ejemplo 2-3

Graficar la respuesta en frecuencia de un PLL con ganancia de lazo $K_L = 15$ kHz/rad ($\omega_v = 94.3$ krad/s). En el mismo papel logarítmico, graficar la respuesta con la adición de un filtro de lazo de un polo con frecuencia de corte $\omega_c = 1.59$ Hz/rad (10 rad/s) y un filtro de lazo de dos polos, con la misma frecuencia de corte.

Solución En la fig. 2-33 se muestran las curvas de respuesta en frecuencia pedidas. Se puede ver que con el filtro de un polo, la respuesta compensada del lazo es = $\omega_v' = 1$ krad/s, y con el filtro de dos polos es $\omega_v'' = 200$ rad/s.

El ancho de banda del filtro de lazo (si es que se necesita) depende de la aplicación específica.

Circuito integrado para un lazo de fase cerrada de precisión

El XR-215 es un sistema de lazo de fase cerrada, monolítico, diseñado por EXAR Corporation, para una gran variedad de aplicaciones en sistemas de comunicaciones, tanto analógicos como digitales. Se adapta en especial para demodulación FM o FSK, síntesis de frecuencia y aplica-

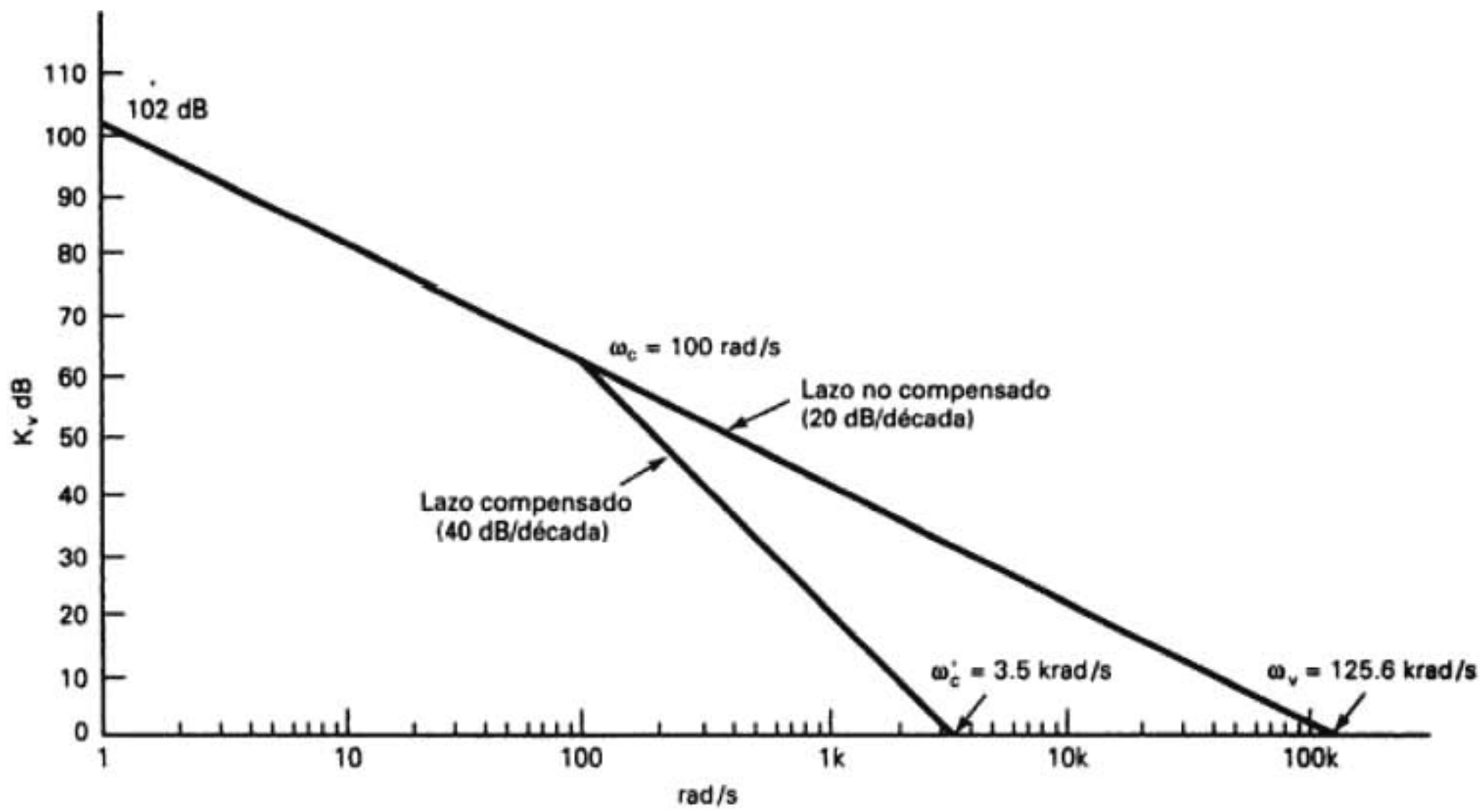


FIGURA 2-32 Respuesta a la frecuencia de un filtro RC de un polo

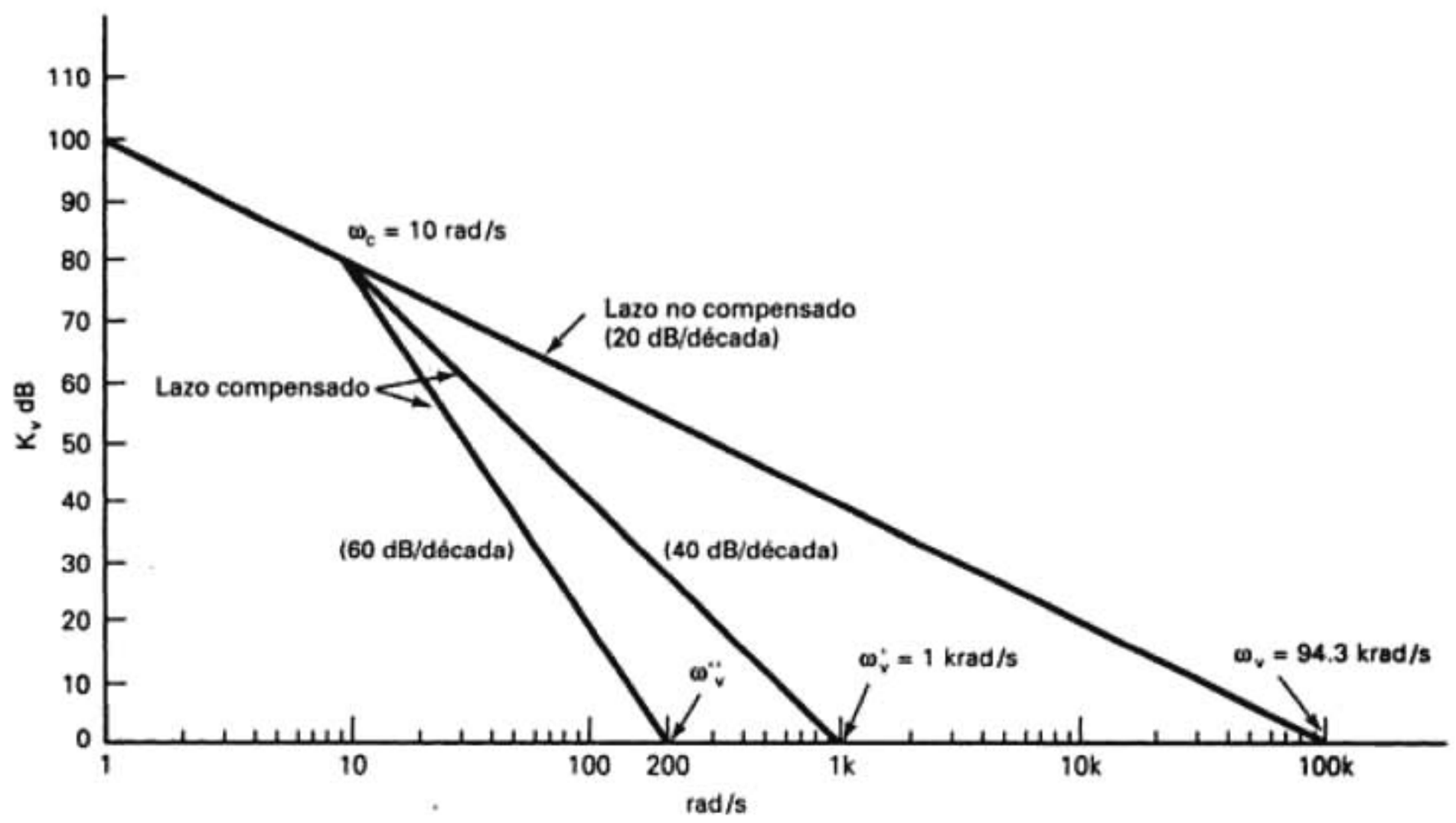


FIGURA 2-33 Respuesta a la frecuencia para el ejemplo 2-3

ciones de filtro de rastreo. Puede operar dentro de un margen relativamente amplio de frecuencias, de 0.5 Hz a 35 MHz, y puede manejar voltajes analógicos de entrada de 300 μV a 3 V. El XR-215 se puede conectar con las familias lógicas DTL, TTL y ECL.

El diagrama de bloques del XR-215 se ve en la fig. 2-34, y consiste en tres secciones principales: un comparador balanceado de fases, un oscilador controlado por voltaje (VCO) muy estable, y un amplificador operacional de alta velocidad. Las salidas del comparador de fases se conectan internamente con las entradas al VCO, y al amplificador no inversor del amplificador

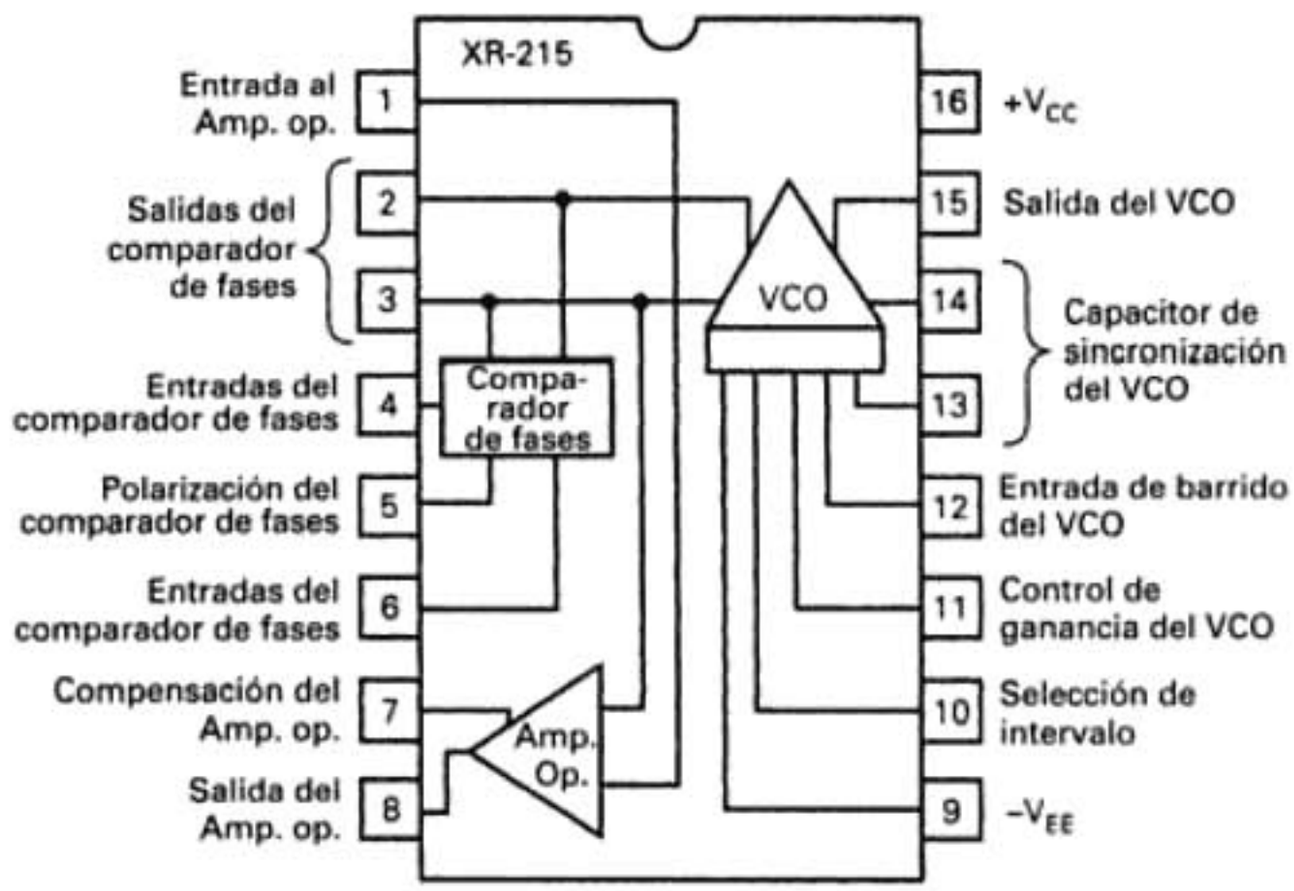


FIGURA 2-34 Diagrama de bloques del lazo de fase cerrada XR-215 monolítico

operacional. Se forma un sistema incorporado de PLL tan sólo acoplando en ca la salida del VCO a cualquiera de las entradas del comparador de fases y agregando un filtro pasabajas a las terminales de salida del comparador.

La sección del VCO tiene barrido de frecuencia, modulación por encendido y apagado, sincronización y posibilidades de programación digital. Su frecuencia es muy estable, y está determinada por un solo capacitor externo. El amplificador operacional se puede usar para preamplificación de audio en aplicaciones de detector de FM, o como amplificador sensor (o comparador) de alta velocidad en demodulación FSK.

Comparador de fase. Una entrada al comparador de fase (terminal 4) se conecta a la señal de entrada externa, y la segunda entrada (terminal 6) se acopla en ca a la terminal de salida del VCO. El voltaje de ca de baja frecuencia (o de cd) a través de las terminales de salida (terminales 2 y 3) del comparador de fases es proporcional a la diferencia entre las fases de las dos señales en las entradas del comparador. Las salidas del comparador de fase se conectan internamente a las terminales de control del VCO. Una salida (terminal 3) se conecta internamente con el amplificador operacional. El filtro pasabajas se consigue conectando una red *RC* a las salidas del comparador de fases, como se ve en la fig. 2-35. Una función de transferencia (ganancia de conversión) típica para el detector de fase es 2 V/rad para voltajes de entrada ≥50 mV.

Oscilador controlado por voltaje (VCO). La frecuencia libre o natural del VCO, f_n, es inversamente proporcional a la capacitancia de un capacitor temporizador, C_o, conectado entre las terminales 13 y 14. El VCO produce una señal de salida con una amplitud aproximada de voltaje de 2.5 V_{p-p} en la terminal 15, con un valor aproximado de cd de salida de 2 V. El VCO se puede barrer dentro de un amplio intervalo de frecuencias de salida, aplicando un voltaje analógico de barrido (V_s) a la terminal 12, como se ve en la fig. 2-36. También se muestran las características normales de barrido. El intervalo de frecuencias del XR-215 se puede ampliar conectándole un resistor externo entre las terminales 9 y 10. La frecuencia de salida del VCO es proporcional a la suma de las corrientes I_1 e I_2, que pasan por los dos transistores internos. La corriente I_1 se establece internamente, mientras que I_2 se ajusta con un resistor externo, R_x. Así, para cualquier valor de C_o, la frecuencia natural del VCO se puede expresar como sigue

$$f_n = f\left(1 + \frac{0.6}{R_x}\right) \tag{2-23}$$

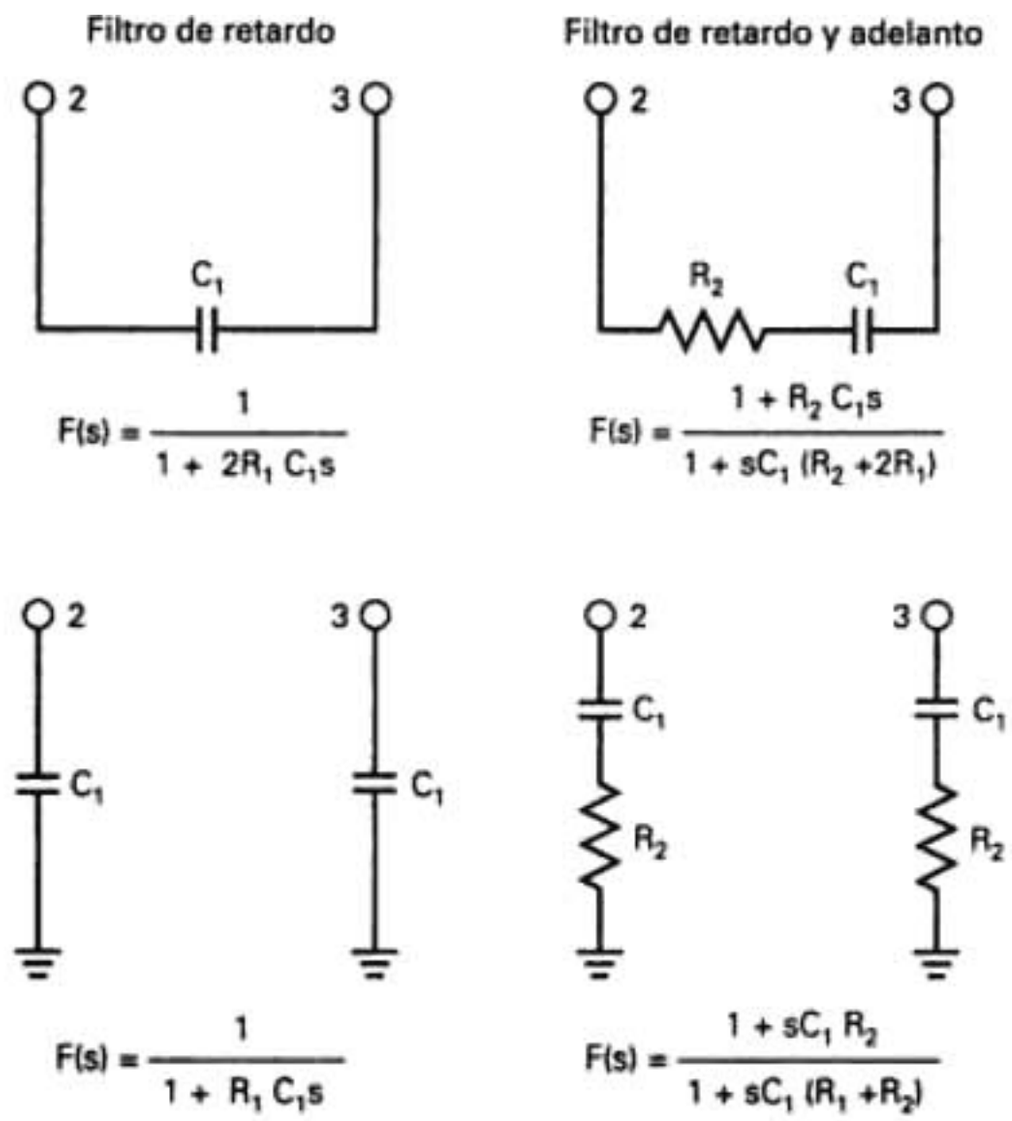


FIGURA 2-35 Conexiones del XR-215 con filtro pasabajas

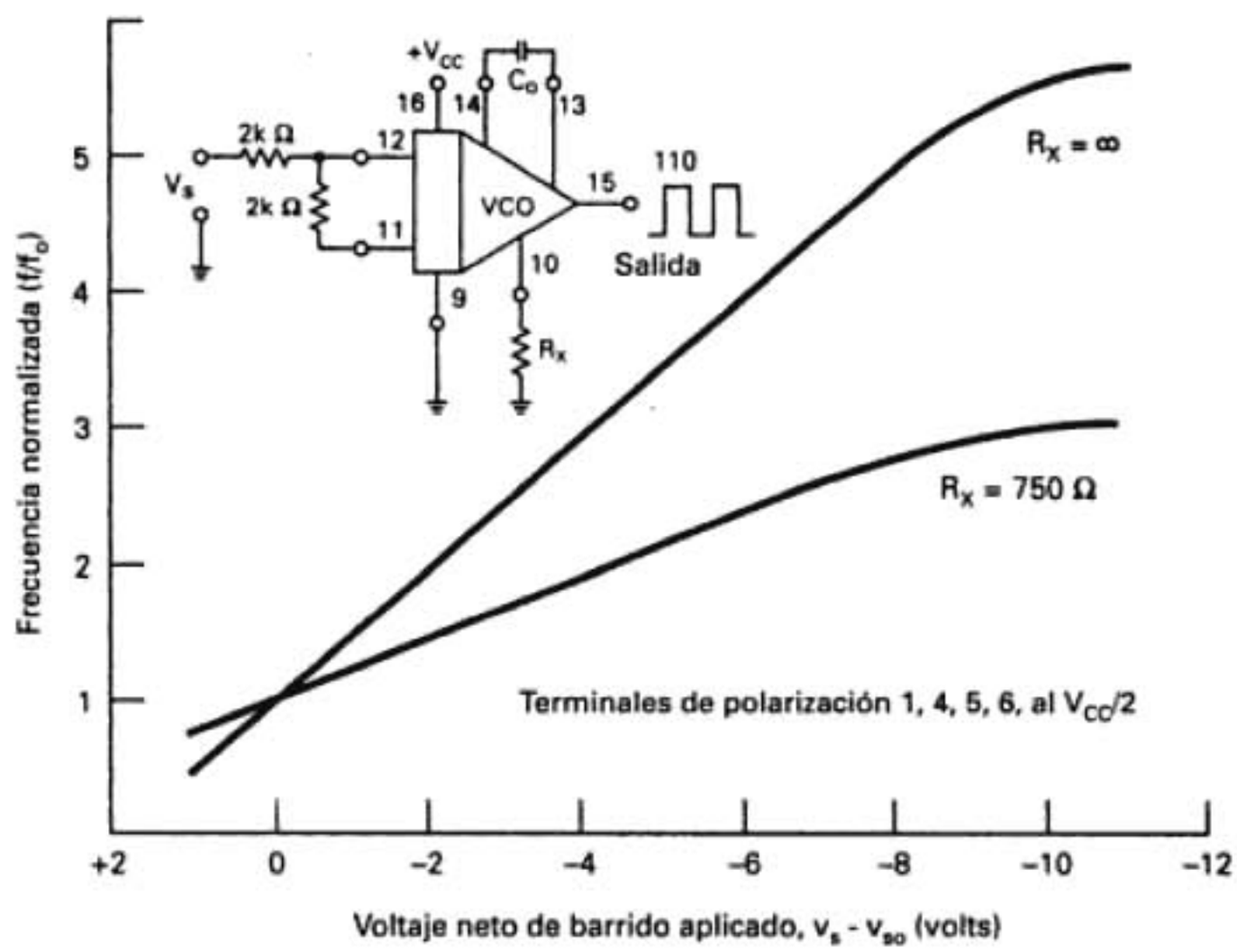


FIGURA 2-36 Características típicas de barrido de frecuencias en función del voltaje de barrido aplicado

en donde f_n = frecuencia natural del VCO (hertz)
f = frecuencia de salida del VCO (hertz) con la terminal 10 en circuito abierto
R_x = resistencia externa (kilohms)

o bien como

$$f_n = \frac{200}{C_o}\left(1 + \frac{0.6}{R_x}\right) \tag{2-24}$$

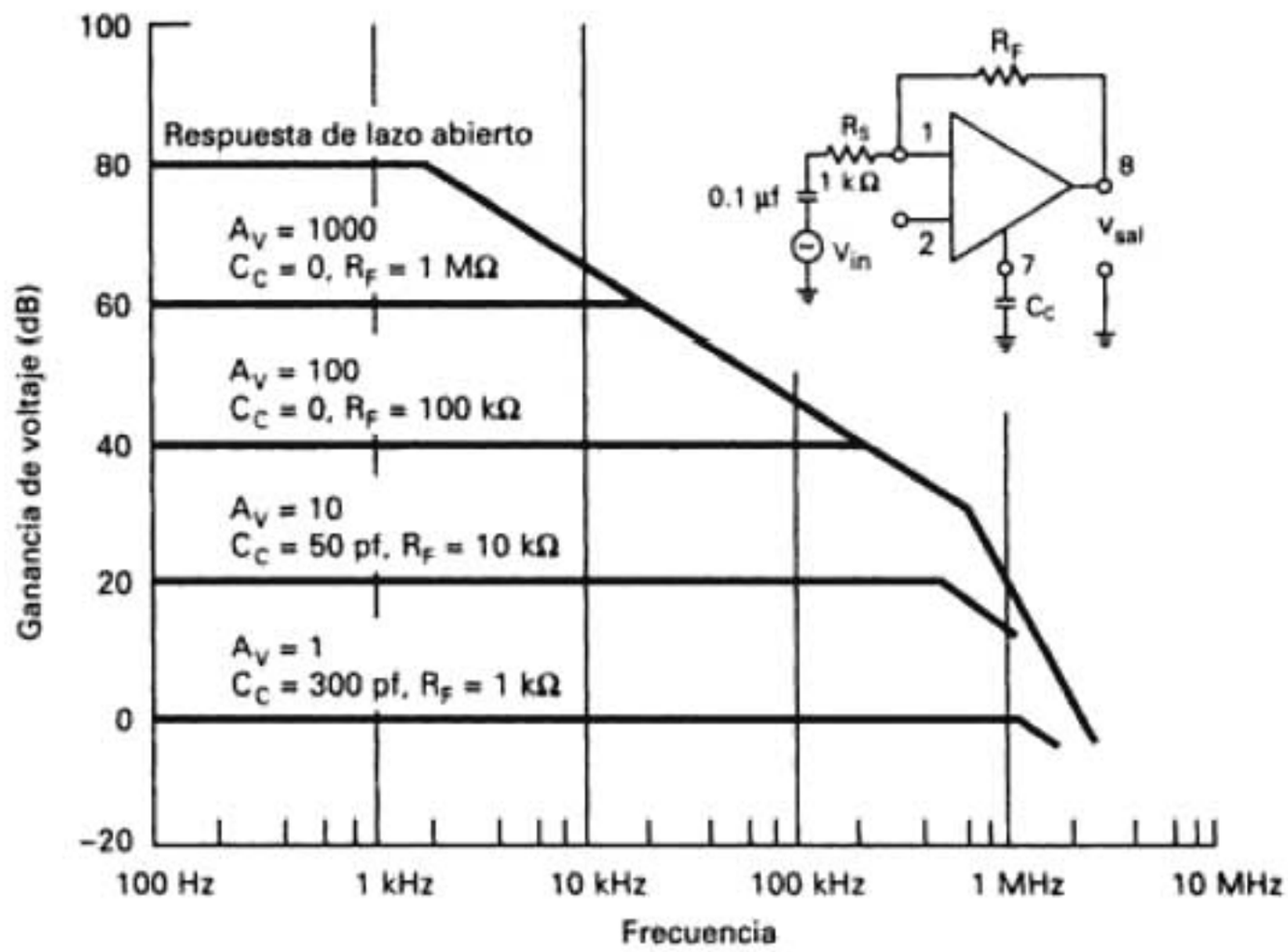


FIGURA 2-37 Respuesta en frecuencia del amplificador operacional XR-215

en donde C_o = capacitancia externa de temporización (microfaradios)
R_x = resistencia externa (kilohms)

La ganancia de conversión (función de transferencia) de voltaje a frecuencia del VCO se determina escogiendo al capacitor de temporización C_o y el resistor de control de ganancia R_o conectados externamente entre las terminales 11 y 12. Matemáticamente, la función de transferencia se expresa como sigue

$$K_o = \frac{700}{C_o R_o} \text{ (rad/s)/V} \tag{2-25}$$

donde K_o = ganancia de conversión del VCO (radianes por segundo por volt)
C_o = capacitancia (microfaradios)
R_o = resistencia (kilohms)

Amplificador operacional. La terminal 1 es la conexión externa a la entrada inversora de la sección del amplificador operacional, y se conecta normalmente con la terminal 2 a través de un resistor de 10 kΩ. La entrada no inversora se conecta internamente con una de las salidas del detector de fase. La terminal 8 se usa como terminal de salida para demodulación FM o FSK. La ganancia de voltaje del amplificador se determina con la resistencia R_f del resistor de retroalimentación, conectado entre las terminales 1 y 8. Las características típicas de respuesta en frecuencia, para un amplificador de éstos, se ven en la fig. 2-37.

La ganancia de voltaje de la sección de amplificador operacional se determina con los resistores de retroalimentación R_f y R_p, entre las terminales 8 y 1 y 2 y 1, respectivamente, y se calcula con la ecuación

$$A_v = \frac{-R_f}{R_s + R_p} \tag{2-26}$$

siendo A_v = ganancia de voltaje (volts por volt)
R_f = resistencia de retroalimentación (ohms)
R_s = resistor externo conectado con la terminal 1 (ohms)
R_p = impedancia interna de 6 kΩ en la terminal 1 (ohms)

Intervalo de enganche. El intervalo de enganche es el intervalo de frecuencias cercanas a la frecuencia natural del VCO dentro del cual el PLL puede mantener el enganche con la señal de entrada externa. Para el XR-215, si no se presenta saturación o limitación, el intervalo de enganche es igual a la ganancia de lazo abierto, o sea

$$\Delta\omega_L = K_v = (K_d)(K_o) \tag{2-27}$$

donde $\Delta\omega_L$ = intervalo de enganche (radianes por segundo)
K_v = ganancia de lazo abierto (radianes por segundo)
K_d = ganancia de conversión del detector de fase (volts por radián)
K_o = ganancia de conversión del VCO (radianes por segundo por volt)

Intervalo de captura. El intervalo de captura es la gama de frecuencias cercanas a la frecuencia natural del VCO dentro de la cual el PLL puede establecer o adquirir enganche con una señal de entrada. Para el XR-215 se puede aproximar mediante una ecuación paramétrica de la forma

$$\Delta\omega_C = \Delta\omega_L |F(j\Delta\omega_C)| \tag{2-28}$$

donde $\Delta\omega_C$ = intervalo de captura (radianes por segundo)
$\Delta\omega_L$ = intervalo de enganche (radianes por segundo)
$|F(j\Delta\omega_C)|$ = magnitud de la respuesta del filtro pasabajas a $\omega = \Delta\omega_C$

SINTETIZADORES DE FRECUENCIAS

Sintetizar quiere decir formar una entidad combinando partes o elementos. Un *sintetizador de frecuencias* se usa para generar muchas frecuencias de salida agregando, restando, multiplicando y dividiendo una cantidad menor de frecuencias fijas. Dicho en modo sencillo, un sintetizador de frecuencias es un generador de frecuencia variable controlado por cristal. El objetivo de un sintetizador de frecuencias es doble. Debe producir tantas frecuencias como sea posible con una cantidad mínima de fuentes, y cada frecuencia producida debe ser tan exacta y estable como las demás. El sintetizador ideal de frecuencias puede generar cientos y hasta miles de frecuencias distintas a partir de un oscilador de un cristal. Un sintetizador de frecuencias puede ser capaz de generar en forma simultánea más de una frecuencia de salida, y cada frecuencia estando sincronizada con un solo oscilador de frecuencia de referencia o maestra. Los sintetizadores de frecuencias se usan mucho en equipo de pruebas y mediciones (generadores de señal de audio y de RF), equipo generador de tono (Touch-Tone), unidades de control remoto (sintonizadores electrónicos), sistemas de comunicaciones de varios canales (telefonía) y sintetizadores de música.

Hay en esencia dos métodos para sintetizar frecuencias: el directo y el indirecto. En la *síntesis directa de frecuencias* se generan varias frecuencias de salida mezclando las salidas de dos o más fuentes de frecuencia controladas por cristal, o dividiendo o multiplicando la frecuencia de salida de un solo oscilador de cristal. En la *síntesis indirecta de frecuencias* se usa un divisor/multiplicador controlado por retroalimentación (como por ejemplo, un PLL), para generar varias frecuencias de salida. La síntesis indirecta de frecuencias es más lenta y más susceptible al ruido; sin embargo, es menos costosa y requiere menos filtros a su vez menos complicados, que la síntesis directa de frecuencias.

Sintetizadores directos de frecuencias

Sintetizador de frecuencias de cristal múltiple. La fig. 2-38 muestra un diagrama de bloques de un *sintetizador de frecuencias de cristal múltiple* que usa mezclado no lineal (heterodinaje) y filtrado, para producir 128 frecuencias distintas con 20 cristales y dos módulos de oscilador. Para los valores indicados de los cristales se sintetiza un intervalo de frecuencias de 510 kHz a 1790 kHz, en incrementos de 10 kHz. Un sintetizador como éste se puede usar para generar las frecuencias de portadora para las 106 estaciones de emisión de AM (de 540 kHz a 1600 kHz). Para las posiciones del conmutador que se indican, están seleccionados los osciladores

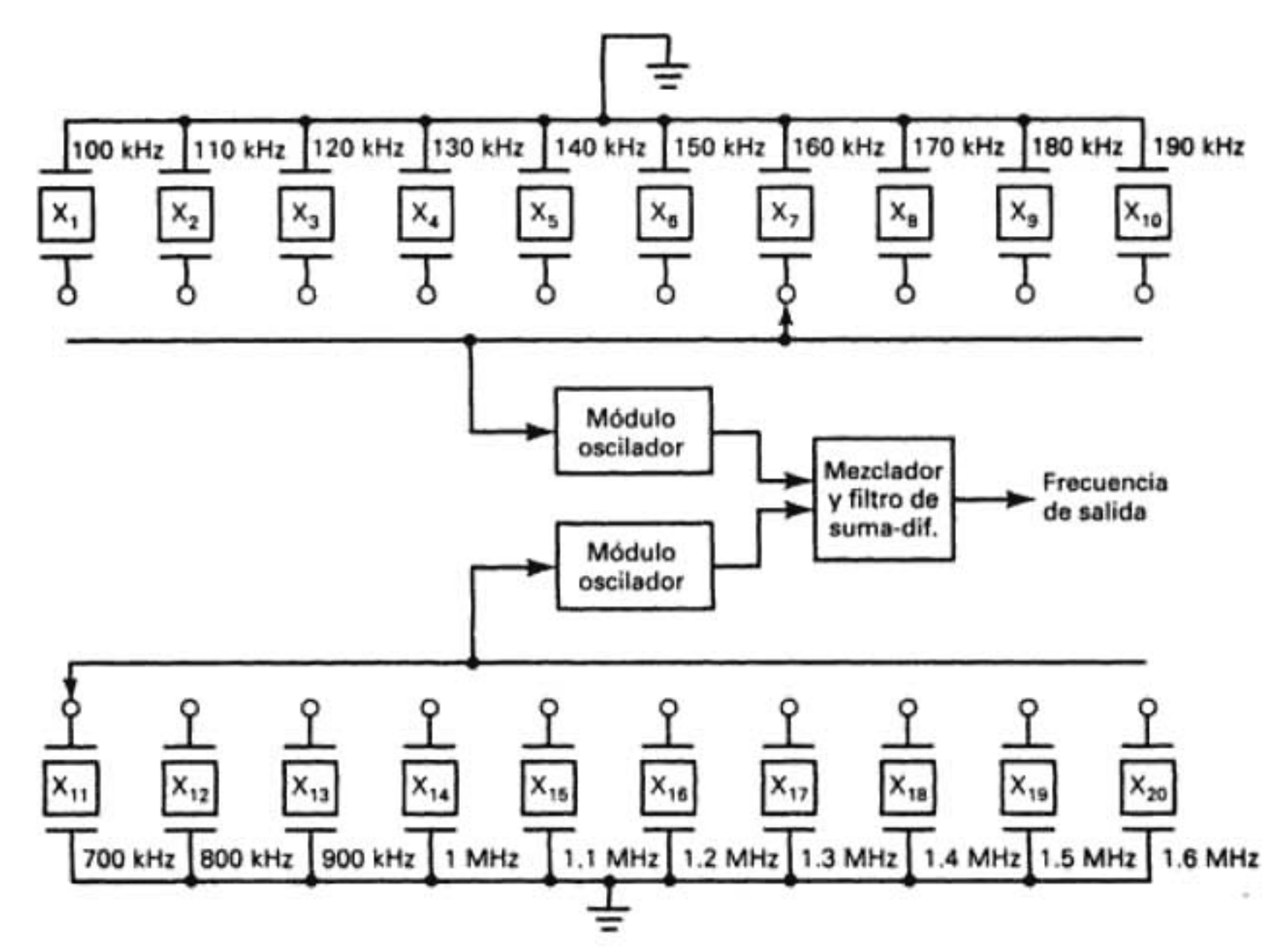


FIGURA 2-38 Sintetizador de frecuencia de cristales múltiples

de 160 kHz y 700 kHz, y las salidas del mezclador balanceado son sus frecuencias de suma y de diferencia (700 kHz ± 160 kHz = 540 kHz y 860 kHz). El filtro de salida se sintoniza a 540 kHz, que es la frecuencia de la portadora para el canal 1. Para generar la frecuencia de portadora para el canal 106 se selecciona el cristal de 100 kHz con el cristal de 1700 kHz (diferencia) o el de 1500 kHz (suma). La separación mínima entre frecuencias de salida en un sintetizador se llama *resolución*. La resolución del sintetizador de la figura 2-38 es 10 kHz.

Sintetizador de frecuencia de un solo cristal. La fig. 2-39 muestra un diagrama de bloques de un *sintetizador de frecuencia de un cristal* que de nuevo usa la suma, resta, multiplicación y división de frecuencias para generar frecuencias de 1 Hz a 999,999 Hz, en incrementos de 1 Hz. Un cristal de 100 kHz es la fuente del oscilador maestro, del que se derivan todas las frecuencias.

La frecuencia del oscilador maestro es una frecuencia base que se divide repetidamente entre 10, para generar cinco frecuencias subbase (10 kHz, 1 kHz, 100 Hz, 10 Hz y 1 Hz). Cada frecuencia subbase se alimenta a un generador separado de armónicas (multiplicador de frecuencias), que consiste en un amplificador no lineal con un filtro sintonizable. El filtro se puede sintonizar con cada una de las primeras nueve armónicas de su frecuencia base. En consecuencia, las frecuencias de salida posibles del generador de armónicas 1 son 0 a 900 kHz en incrementos de 100 kHz; para el generador de armónicas 2, de 10 a 90 kHz en incrementos de 10 kHz, etcétera. La resolución del sintetizador se determina por las veces que se divide la frecuencia del oscilador maestro de cristal. Para el sintetizador de la fig. 2-39, la resolución es de 1 Hz. Los mezcladores que se usan son moduladores balanceados, con filtros de salida que se sintonizan a la suma de las dos frecuencias de entrada. Por ejemplo, las armónicas seleccionadas en la tabla 2-1 producen una frecuencia de salida de 246,313 Hz. Esta tabla muestra una lista de las posiciones del selector para cada generador de armónicas, y las frecuencias de entrada y salida de cada mezclador. Se puede ver que los cinco mezcladores tan sólo suman las frecuencias de salida de los seis generadores de armónicas, con tres niveles de mezclado.

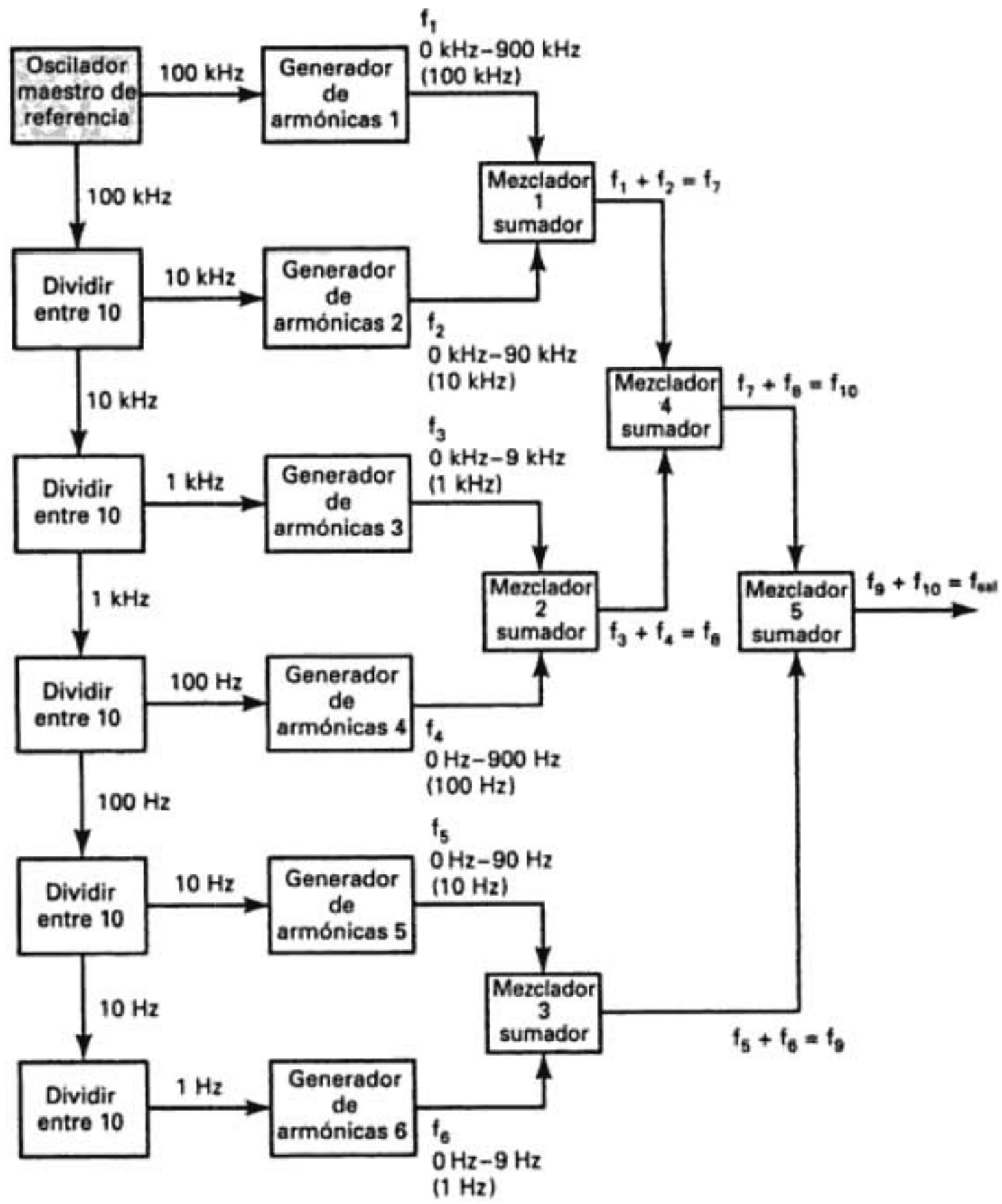


FIGURA 2-39 Sintetizador de frecuencias de un solo cristal

TABLA 2-1 Posiciones de interruptores y armónicas

Generador de armónicas	Salida de frecuencia seleccionada	Frecuencia de salida del mezclador
1	200 kHz	salida del mezclador 1 = 200 kHz + 40 kHz = 240 kHz
2	40 kHz	
3	6 kHz	salida del mezclador 2 = 6 kHz + 0.3 kHz = 6.3 kHz
4	0.3 kHz	
5	10 Hz	salida del mezclador 3 = 10 Hz + 3 Hz = 13 Hz
6	3 Hz	
		salida del mezclador 4 = 240 kHz + 6.3 kHz = 246.3 kHz
		salida del mezclador 5 = 246.3 kHz + 13 Hz = 246.313 kHz

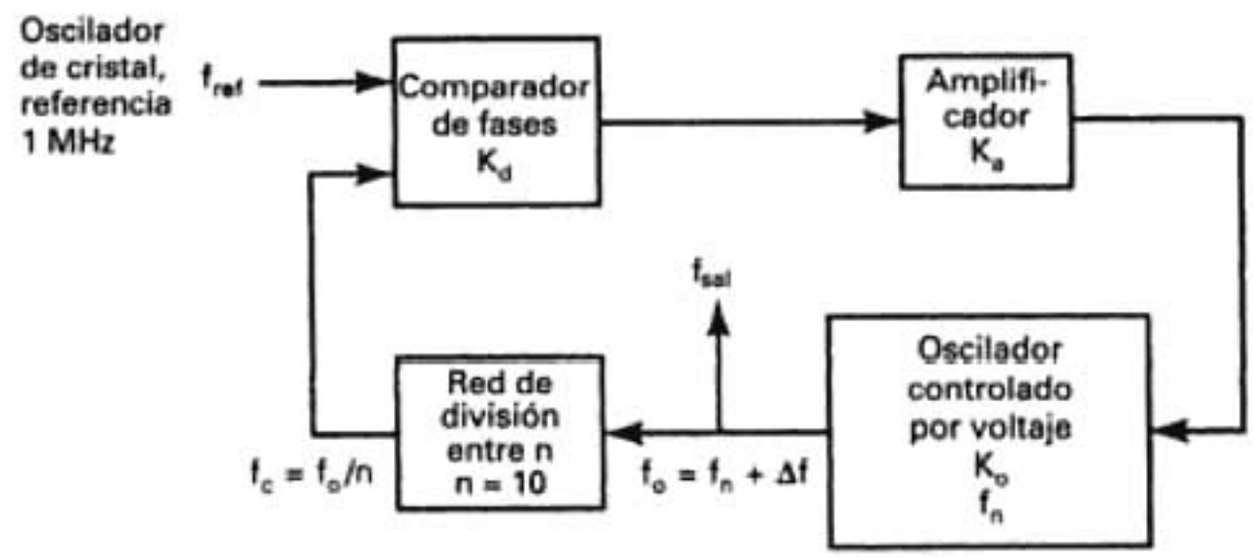


FIGURA 2-40 Sintetizador de frecuencias de PLL con un solo lazo

Sintetizadores indirectos de frecuencias

Sintetizadores de frecuencias con lazo de fase cerrada. En años recientes, los sintetizadores de frecuencias con PLL se han vuelto rápidamente el método más popularizado para síntesis de frecuencias. La fig. 2-40 muestra un diagrama de bloques para un sintetizador PLL de frecuencias de *lazo sencillo*. La referencia estable de frecuencia es un oscilador controlado por cristal. El intervalo de frecuencias generadas, y la resolución, dependen de la red divisora y de la ganancia de lazo abierto. El divisor de frecuencias es un circuito de dividir entre n, donde n es cualquier número entero. La forma más simple de circuito divisor es un *contador digital de subida y bajada* con frecuencia de salida $f_c = f_o/n$, donde f_o es la frecuencia de salida del VCO. Con este arreglo, una vez alcanzado el enganche, $f_c = f_{ref}$, y la frecuencia de salida del VCO y el sintetizador es $f_o = nf_{ref}$. Así, el sintetizador es esencialmente un multiplicador de frecuencia por n. El divisor de frecuencia reduce la ganancia de lazo abierto en un factor n. En consecuencia, los demás circuitos en torno al lazo deben tener ganancias relativamente altas. La ganancia de lazo abierto del sintonizador de frecuencias de la fig. 2-40 es

$$K_v = \frac{(K_d)(K_a)(K_o)}{n} \tag{2-29a}$$

Se puede ver en esta ecuación que cuando cambia n, la ganancia de lazo abierto cambia en proporción inversa. Una forma de remediar este problema es programar la ganancia del amplificador, así como la relación del divisor. Entonces, la ganancia de lazo abierto es

$$K_v = \frac{n(K_d)(K_a)(K_o)}{n} = (K_d)(K_a)(K_o) \tag{2-29b}$$

El intervalo de frecuencias con la frecuencia de referencia y el circuito divisor que se ven en la fig. 2-40, es

$$\begin{aligned} f_o &= nf_{ref} \\ &= f_{ref} \text{ a } 10f_{ref} \\ &= 1 \text{ MHz a } 10 \text{ MHz} \end{aligned}$$

Sintonizador de frecuencias preescalado. La fig. 2-41 muestra el diagrama de bloques de un sintetizador de frecuencias que usa un lazo de fase cerrada y un *preescalador*, para obtener división fraccionaria. También es necesario el preescalado para generar frecuencias mayores que 100 MHz, porque no se dispone de contadores programables que funcionen con eficiencia a esas altas frecuencias. El sintetizador de la fig. 2-41 usa un preescalador *módulo dos*. Tiene dos modos de operación. Un modo proporciona una frecuencia de salida por cada pulso P de entrada, y el otro modo proporciona una salida por cada pulso $P + 1$ de entrada. Cuando el registro m contiene un número distinto de cero, el preescalador cuenta en el modo $P + 1$. En consecuencia,

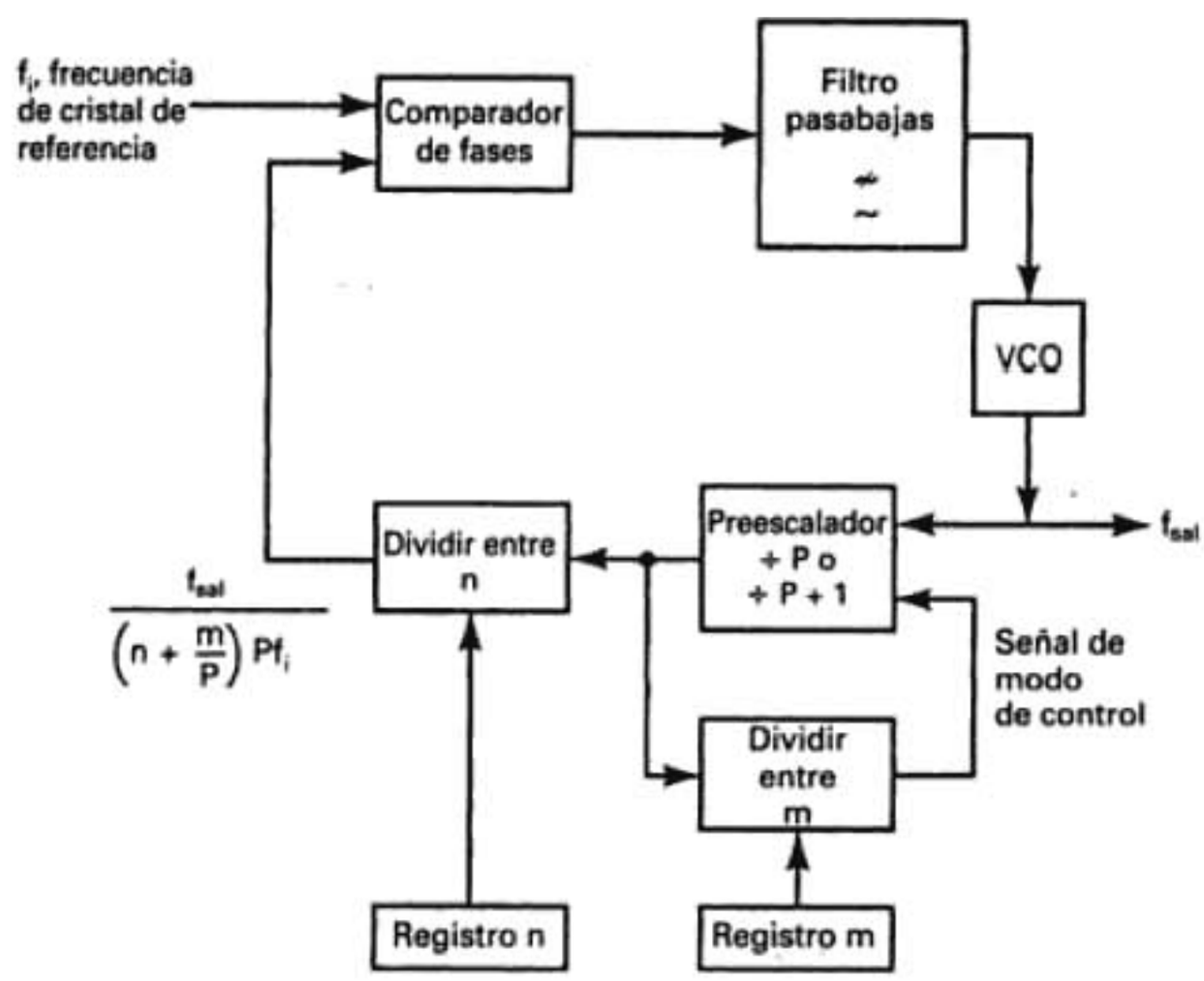


FIGURA 2-41 Sintetizador de frecuencias con preescalador

una vez que los registros m y n se cargan inicialmente, el preescalador contará $(P + 1)m$ veces hacia abajo, hasta que el contador m llegue a cero, el preescalador opera en el modo P y el contador n cuenta $(n - m)$ veces hacia abajo. En este momento los contadores m y n se restablecen en sus valores originales, que se guardaron en los registros m y n, respectivamente, y se repite el proceso. La ecuación de la frecuencia de salida del sintetizador, f_o, es

$$f_o = \left(n + \frac{m}{P}\right)Pf_i \tag{2-30}$$

Preescaladores de circuito integrado. Hoy se consiguen preescaladores avanzados ECL (lógica acoplada por emisor) de módulos duales (dividir entre 128/129 o 64/65) y triples (dividir entre 64/65/72), de circuito integrado, que funcionan a frecuencias de 1 Hz hasta 1.3 GHz. Estos preescaladores de circuito integrado son de tamaño pequeño, funcionan a bajo voltaje, consumen poca corriente y son simples. Son ideales para teléfonos celulares y sin cable, para redes de área local de RF, equipo de pruebas y medición, sistemas militares de radio, radios móviles de VHF/UHF y radios de mano VHF/UHF.

La fig. 2-42 muestra el diagrama de bloques del preescalador NE/SA701, fabricado por Signetics Company. El NE701 es un módulo dual avanzado (dividir entre 128/129 o 64/65), de ECL y de baja potencia. Funciona con un voltaje mínimo de suministro de 2.5 V y tiene un consumo de corriente de 2.8 mA, por lo que permite su uso en equipo de baterías y de baja potencia. La frecuencia máxima de señal de entrada es 1.2 GHz, en aplicaciones celulares y otras móviles terrestres. El circuito se construye en tecnología ECL en el proceso HS4+, y se consigue en paquete SO de 8 terminales.

El NE701 comprende un divisor de frecuencia, construido con un preescalador síncrono divisor entre 4 o 5, seguido de un contador síncrono fijo de cinco etapas. El modo normal de operación es que la entrada SW (conmutador de módulo) sea baja y la entrada MC (control de módulo) sea alta, en cuyo caso funciona como un contador módulo 128. Para que funcione en módulo 129, la entrada MC se coloca en bajo, haciendo que el preescalador pase a la operación de dividir entre 5 para el último ciclo del contador síncrono. De igual modo, para ir a dividir entre 64/65, el NE701 genera los módulos respectivos con la señal SW forzada a ser alta, con lo cual se omite la cuarta etapa del divisor síncrono. Con SW en circuito abierto se selecciona el modo de dividir entre 128/129 y con SW conectada con V_{CC}, se selecciona dividir entre 64/65.

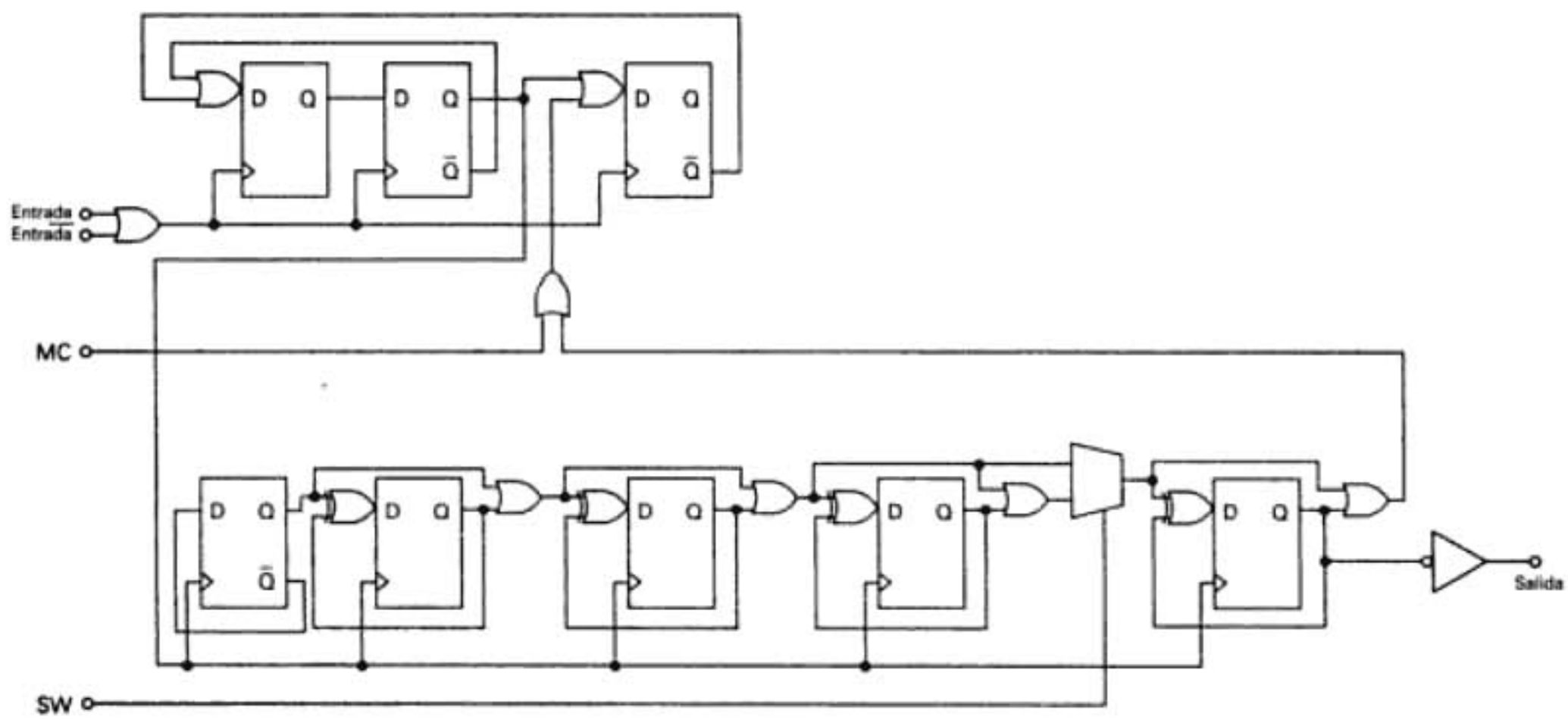


FIGURA 2-42 Diagrama de bloques del preescalador de lógica acoplada por emisor NE/SA701

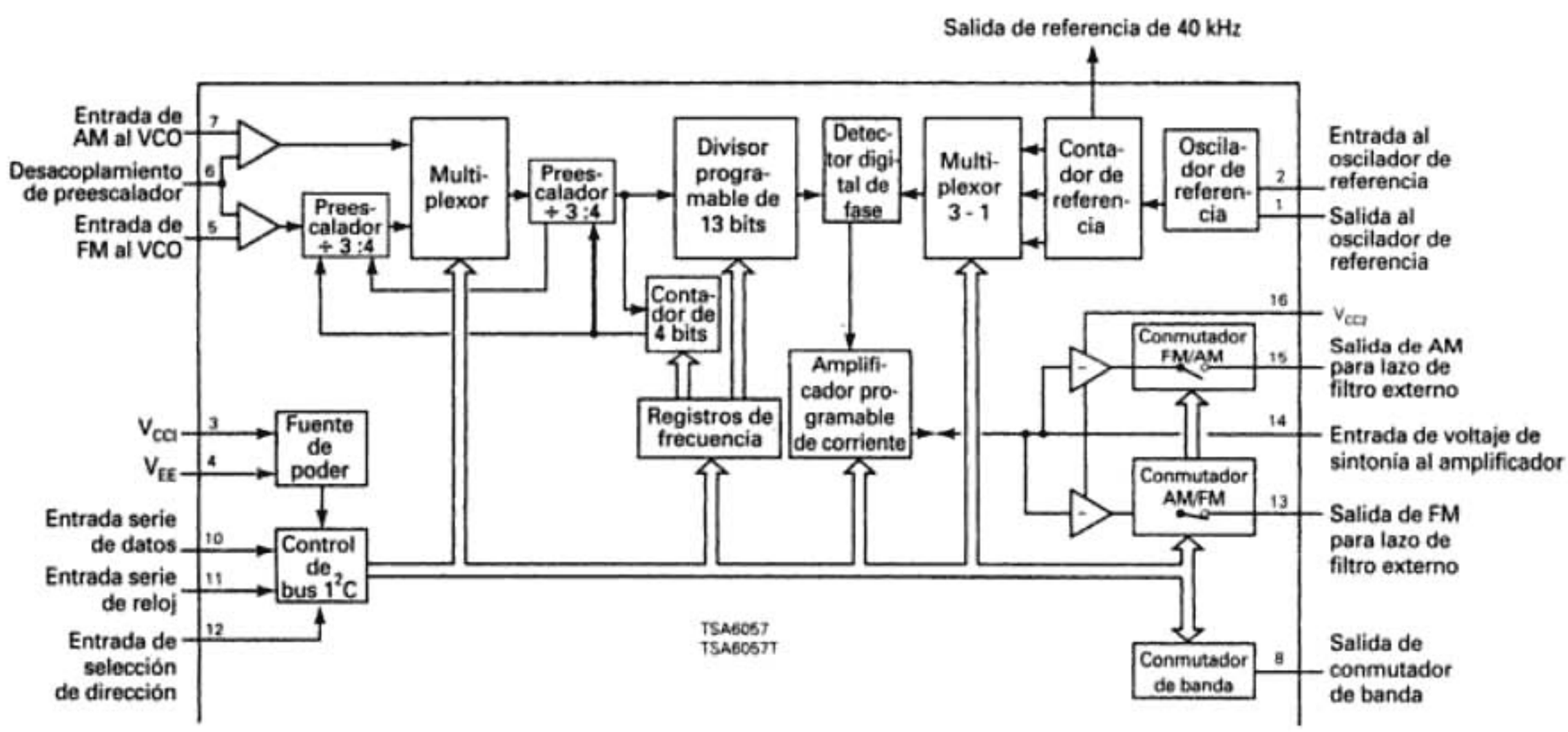


FIGURA 2-43 Diagrama de bloques del sintetizador de frecuencias con PLL para sintonía de radio TSA6057

Circuito integrado sintetizador de frecuencia PLL de sintonía de radio. La fig. 2-43 muestra el diagrama de bloques del sintonizador de frecuencias PLL para sintonía de radio TSA6057/T de Signetics. Es un sintetizador de frecuencias bipolar en CI, fabricado con la tecnología SUBILO-N (componentes separados lateralmente por óxido). Desempeña todas las funciones de sintonía de un sistema sintonizador de radio con PLL. El circuito integrado está diseñado para aplicaciones en todo tipo de radiorreceptores y tiene las siguientes propiedades:

1. Amplificadores separados de entrada para señales de AM, FM y VCO.
2. Preescaladores integrados en el CI de alta sensibilidad de entrada para AM (3:4) y FM (15:16).
3. Sintonía de alta velocidad, por tener un poderoso detector de fase con memoria digital.

4. Amplificador de voltaje de sintonía de alta eficiencia de una entrada (dos salidas) integrado en el CI. Una salida se conecta con el filtro externo de lazo de AM, y la otra salida con el de FM.
5. Amplificador de corriente de dos valores integrado en el CI, que consiste en una fuente de corriente de 5 μA y de 450 μA. Esto permite ajustar la ganancia de lazo y proporciona una sintonía con alta corriente y alta velocidad, y una sintonía estable de baja corriente.
6. Un oscilador de referencia (4 MHz) para AM y FM, seguido de un contador de referencia. La frecuencia de referencia puede ser 1 kHz, 10 kHz o 25 kHz, y se aplica al detector de fase con memoria digital. El contador de referencia también tiene una salida de frecuencia de referencia de 40 kHz en la terminal 9, para cooperación con el sistema FM/IF.
7. La frecuencia del oscilador va de 512 kHz a 30 MHz, y de 30 a 150 MHz.

PREGUNTAS

2-1. Defina *oscilar* y *oscilador.*
2-2. Describa los siguientes términos: *autosostenido, repetitivo, funcionamiento libre* y *con disparo.*
2-3. Describa el proceso regenerativo necesario para que se produzcan las oscilaciones autosostenidas.
2-4. Escriba y describa los cuatro requisitos para que trabaje un *oscilador con retroalimentación.*
2-5. ¿Qué quieren decir los términos *retroalimentación positiva* y *negativa*?
2-6. Defina la *ganancia de lazo cerrado* y *de lazo abierto.*
2-7. Escriba las cuatro configuraciones más comunes de oscilador.
2-8. Describa el funcionamiento de un oscilador de puente de Wien.
2-9. Describa la acción de oscilador en un circuito tanque *LC.*
2-10. ¿Qué quiere decir *oscilación amortiguada*? ¿Qué la causa?
2-11. Describa el funcionamiento de un oscilador de Hartley y el de un oscilador de Colpitts.
2-12. Defina la *estabilidad de frecuencia.*
2-13. Haga una lista de factores que afectan la estabilidad de frecuencia de un oscilador.
2-14. Describa el *efecto piezoeléctrico.*
2-15. ¿Qué quiere decir el término *corte de cristal*? Haga una lista de varios cortes de cristal, descríbalos y compare sus estabilidades.
2-16. Describa cómo funciona un oscilador de cristal de sobretonos.
2-17. ¿Cuál es la ventaja de un oscilador de cristal de sobretonos, respecto a un oscilador convencional de cristal?
2-18. ¿Qué quiere decir *coeficiente positivo de temperatura*? ¿Qué quiere decir *coeficiente negativo de temperatura*?
2-19. ¿Qué es un cristal con *coeficiente cero*?
2-20. Haga un esquema del circuito eléctrico equivalente de un cristal, y describa los diversos componentes y sus contrapartes mecánicas.
2-21. ¿Cuál configuración de oscilador de cristal tiene la mejor estabilidad?
2-22. ¿Cuál configuración de oscilador de cristal es la menos costosa y más adaptable a interconexiones digitales?
2-23. Describa un módulo de oscilador de cristal.
2-24. ¿Cuál es la principal ventaja de los osciladores de cristal en comparación con los de circuito tanque *LC*?
2-25. Describa el funcionamiento de un diodo varactor.
2-26. Describa un lazo de fase cerrada.
2-27. ¿Qué tipos de generadores de forma de onda con integración a gran escala (LSI) se pueden conseguir?
2-28. Describa el funcionamiento básico de un generador de forma de onda en circuito integrado.
2-29. Haga una lista de las ventajas de un generador monolítico de funciones.
2-30. Haga una lista de las ventajas de un oscilador monolítico controlado por voltaje.
2-31. Describa en forma breve el funcionamiento de un oscilador monolítico de precisión.
2-32. Haga una lista de las ventajas de un PLL de circuito integrado respecto a un PLL discreto.
2-33. Describa el funcionamiento de un oscilador controlado por voltaje.
2-34. Describa el funcionamiento de un detector de fase.

2-35. Describa cómo se obtiene la adquisición de lazo con un PLL, desde un estado inicial no sincronizado, hasta que se logra el enganche de frecuencia.

2-36. Defina los siguientes términos: *diferencia de frecuencia, cero diferencia, tiempo de adquisición* y *ganancia de lazo abierto.*

2-37. Compare los siguientes términos e indique cómo se relacionan entre sí: *intervalo de captura, semiintervalo de captura, ganancia de lazo cerrado, intervalo de retención, intervalo de rastreo* e *intervalo de enganche.*

2-38. Defina los siguientes términos: *PLL no compensado, frecuencia de corte del lazo* y *filtro de rastreo.*

2-39. Defina *sintetizar*. ¿Qué es un sintetizador de frecuencias?

2-40. Describa la síntesis directa e indirecta de frecuencias.

2-41. ¿Qué quiere decir resolución de un sintetizador de frecuencias?

2-42. ¿Cuáles son las ventajas de los preescaladores y sintetizadores de frecuencias en circuito integrado sobre sus equivalentes sin circuito integrado?

PROBLEMAS

2-1. Para un cristal de 20 MHz con coeficiente negativo de temperatura $k = -8$ Hz/MHz/°C, calcule la frecuencia de operación con los siguientes cambios de temperatura:
(a) Aumento de 10° C.
(b) Aumento de 20° C.
(c) Disminución de 20° C.

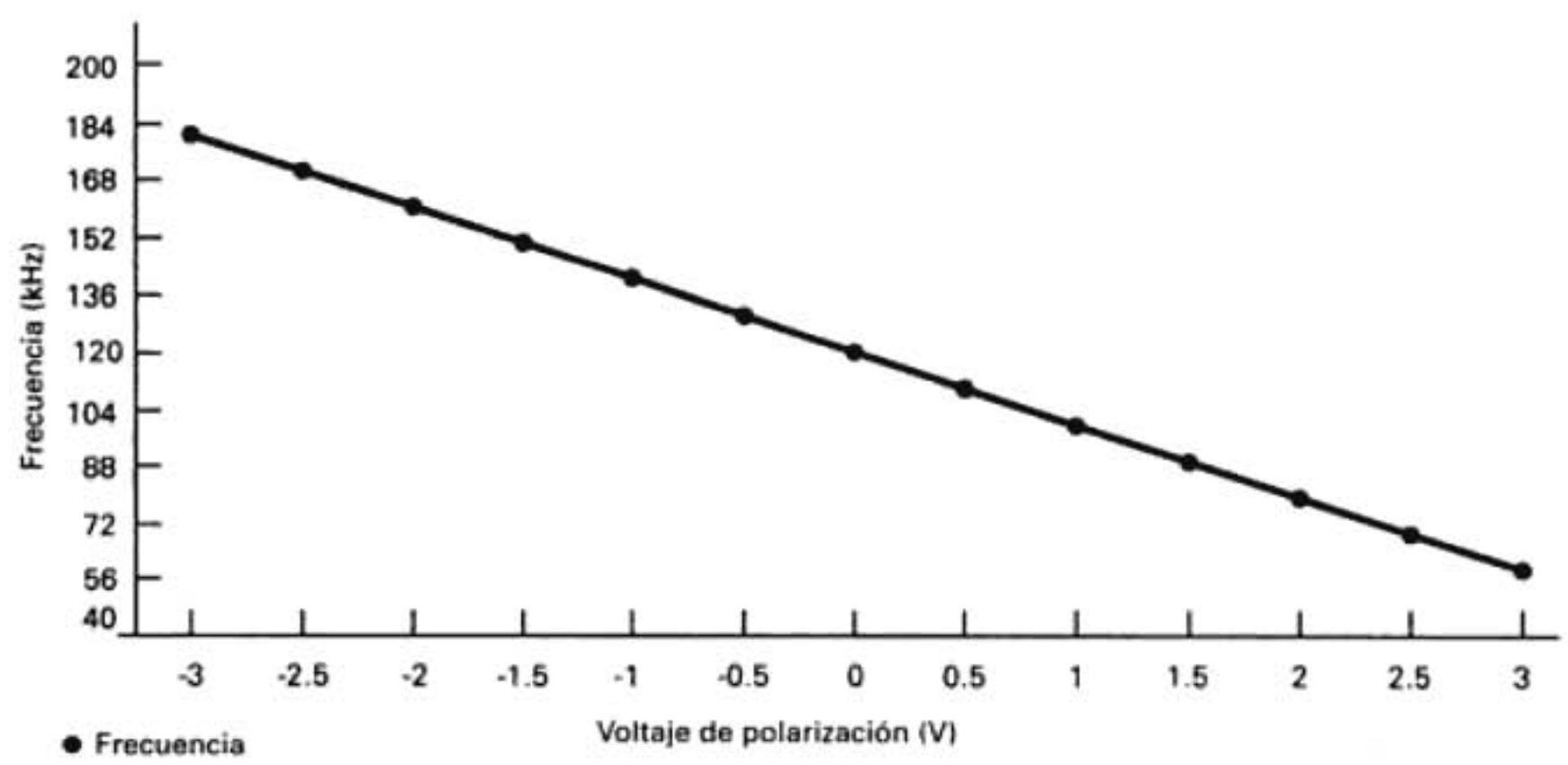
2-2. Para el oscilador de Wien de la fig. 2-3, y con los siguientes valores de componente, calcule la frecuencia de oscilación. $R_1 = R_2 = 1$ kΩ; $C_1 = C_2 = 100$ pF.

2-3. Para el oscilador de Hartley de la fig. 2-5a, y con los siguientes valores de componente, determine la frecuencia de oscilación. $L_{1a} = L_{1b} = 50$ μH; $C_1 = 0.01$ μF.

2-4. Para el oscilador de Colpitts de la fig. 2-6a, y con los siguientes valores de componente, calcule la frecuencia de oscilación. $C_{1a} = C_{1b} = 0.01$ μF; $L_1 = 100$ μH.

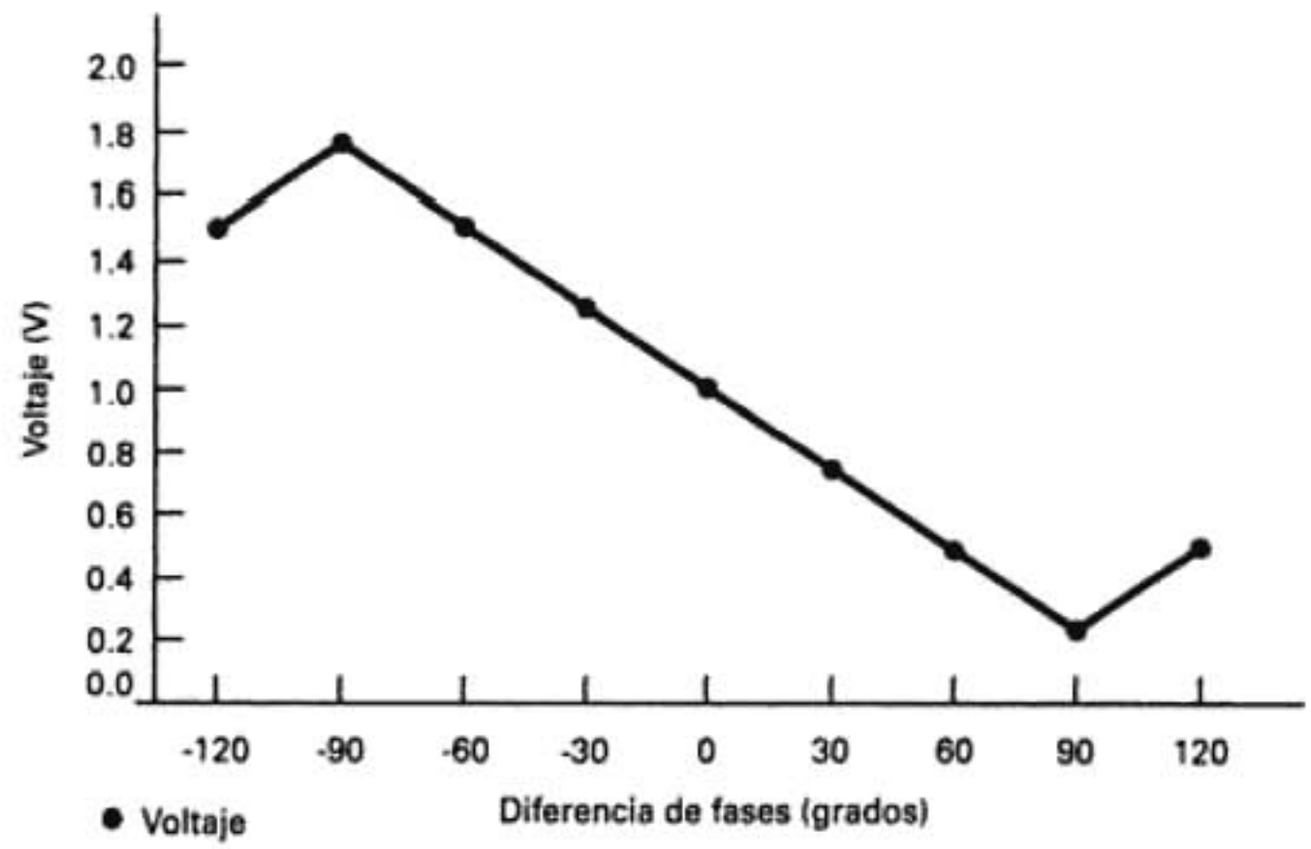
2-5. Calcule la capacitancia de un diodo varactor con los siguientes valores: $C = 0.005$ μF; $V_r = -2$ V.

2-6. Para la curva característica de entrada en función de salida del VCO que se adjunta, calcule:
(a) Frecuencia de operación para una señal de entrada de −2 V.
(b) Desviación de frecuencia para una señal de entrada de $\pm 2\ V_p$.
(c) La función de transferencia, K_o, de la parte lineal de la curva (−3 V a +3 V).



2-7. Con la curva característica voltaje de salida en función de diferencia de fases (θ_e) que se adjunta, determine:
(a) El voltaje de salida para una diferencia de fases de −45°.
(b) El voltaje de salida para una diferencia de fases de +60°.

(c) El voltaje máximo de salida.
(d) La función de transferencia, K_d.



2-8. Para el PLL de la fig. 2-30, con una frecuencia natural de VCO f_n = 150 kHz, una frecuencia de entrada f_i = 160 kHz, y ganancias de circuito K_d = 0.2 V/rad, K_f = 1, K_a = 4 y K_o = 15 kHz/V, calcule:
(a) La ganancia de lazo abierto, K_v.
(b) Δf.
(c) V_{sal}.
(d) V_d.
(e) θ_e.
(f) Intervalo de retención, $\Delta f_{máx}$.

2-9. Grafique la respuesta a la frecuencia de un PLL con ganancia de lazo abierto K_L = 20 kHz/rad. En el mismo papel logarítmico, grafique la respuesta con un filtro de lazo de un polo, con frecuencia de corte ω_c = 100 rad/s, y filtro de dos polos con la misma frecuencia de corte.

2-10. Calcule el cambio de frecuencia, Δf, para un VCO con función de transferencia K_o = 2.5 kHz/V y cambio de voltaje cd en la entrada ΔV = 0.8 V.

2-11. Calcule el voltaje en la entrada de un comparador de fases cuya función de transferencia es K_d = 0.5 V/rad y error de fase θ_e = 0.75 rad.

2-12. Determine el intervalo de retención ($\Delta f_{máx}$) de un PLL con ganancia de circuito abierto K_v = 20 kHz/rad.

2-13. Calcule el error de fase necesario para producir un desplazamiento de frecuencia Δf = 10 kHz en un VCO, con una ganancia de lazo abierto de K_L = 40 kHz/rad.

2-14. Calcule la frecuencia de salida del sintetizador de frecuencias de varios cristales de la fig. 2-38, si se seleccionan los cristales X_8 y X_{18}.

2-15. Determine la frecuencia de salida del sintetizador de frecuencias de cristal único que se ve en la fig. 2-39, para las siguientes armónicas.

Generador de armónicas	Armónica	Generador de armónicas	Armónica
1	6	4	1
2	4	5	2
3	7	6	6

2-16. Calcule f_c para el PLL de la fig. 2-40, con una frecuencia natural f_n = 200 kHz, Δf = 0 Hz y n = 20.

2-17. Determine la frecuencia de operación, para un cristal de 10 MHz con un coeficiente de temperatura k = 12 Hz/MHz/°C, con los cambios siguientes de temperatura:
(a) Aumento de 20° C.
(b) Disminución de 20° C.
(c) Aumento de 10° C.

2-18. Para el oscilador de puente de Wien de la fig. 2-3, con los siguientes valores de componente, calcule la frecuencia de oscilación. $R_1 = R_2$ = 2 k W; $C_1 = C_2$ = 1000 pF.

2-19. Para el oscilador de puente de Wien de la fig. 2-3, y con los valores de componente del problema 2-2, calcule el desplazamiento de fase a través de la red de adelanto y retraso, para frecuencias de una octava arriba y abajo de la frecuencia de oscilación.

2-20. Para el oscilador de Hartley de la fig. 2-5a, y con los siguientes valores de componente, calcule la frecuencia de oscilación. $L_{1a} = L_{1b} = 100\ \mu H$; $C_1 = 0.001\ \mu F$.

2-21. Para el oscilador de Colpitts de la fig. 2-6a, y con los siguientes valores de componente, calcule la frecuencia de oscilación: $C_{1a} = 0.0022\ \mu F$, $C_{1b} = 0.022\ \mu F$ y $L_1 = 3$ mH.

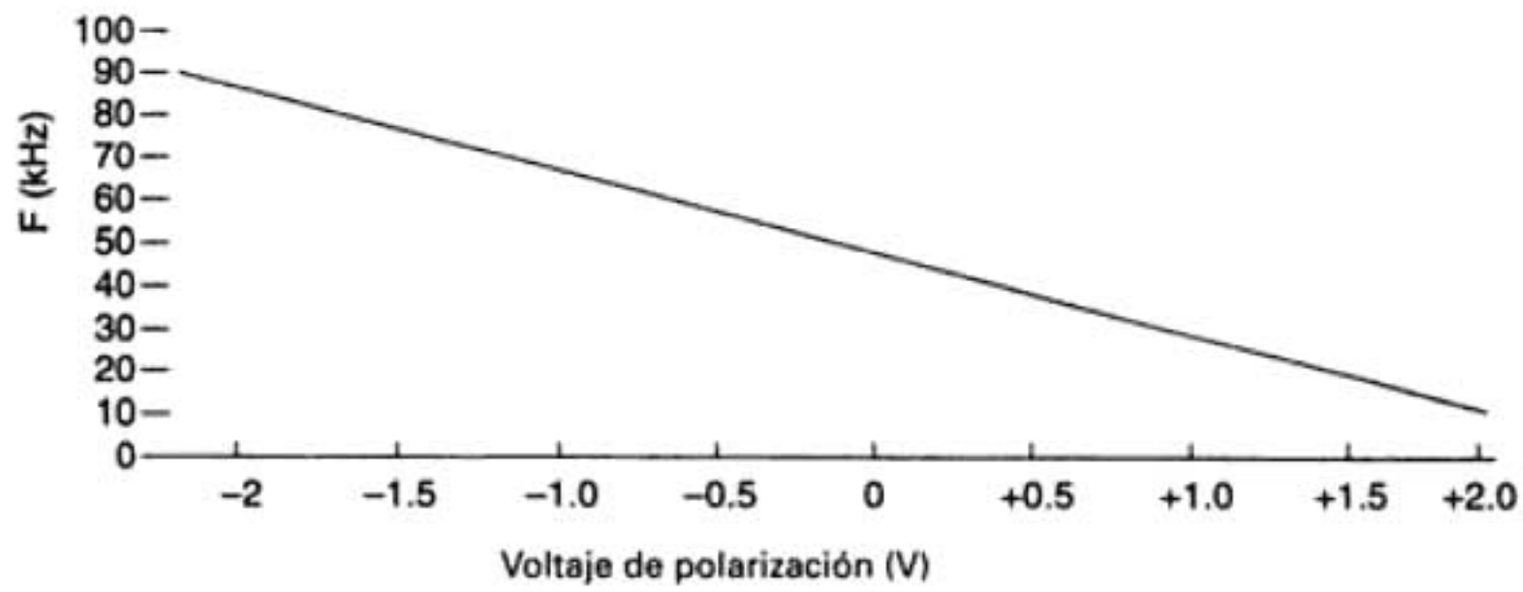
2-22. Calcule la capacitancia de un diodo varactor con los siguientes valores: $C = 0.001\ \mu F$ y $v_r = -1.5$ V.

2-23. Determine, con la curva característica de entrada en función de salida de VCO adjunta:

(a) La frecuencia de operación para una señal de entrada de -1.5 V.

(b) La desviación de frecuencia para una señal de entrada de 2 $V_{p\text{-}p}$.

(c) La función de transferencia K_o para la parte lineal de la curva (-2 V a $+2$V).



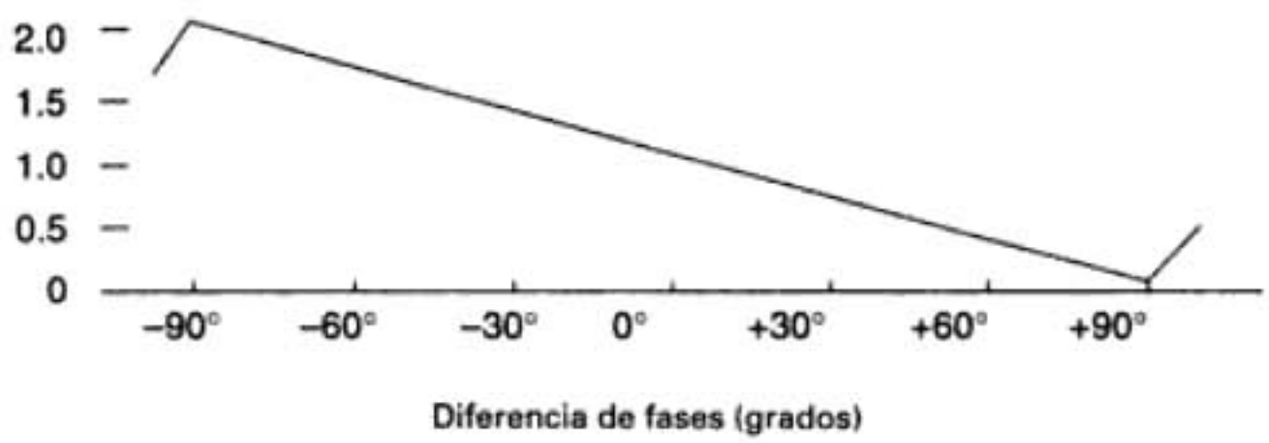
2-24. Para la curva característica de voltaje de salida del detector en función de la diferencia de fases (θ_e) que se adjunta, calcule:

(a) El voltaje de salida para una diferencia de fases de $-45°$.

(b) El voltaje de salida para una diferencia de fases de $+60°$.

(c) El voltaje de salida máximo.

(d) La función de transferencia, K_d.



2-25. Para el PLL de la fig. 2-30a, la frecuencia natural del VCO $f_n = 120$ kHz, la frecuencia de entrada $f_i = 125$ kHz y las siguientes ganancias de circuito: $K_d = 0.2$ V/rad, $K_f = 1$, $K_a = 5$ y $K_o = 12$ kHz/V, calcule:

(a) la ganancia de lazo abierto, K_v.

(b) Δf.

(c) V_{sal}.

(d) V_d.

(e) θ_e.

(f) El intervalo de retención, $\Delta f_{máx}$.

2-26. Grafique la respuesta a la frecuencia de un PLL con ganancia de lazo abierto $K_v = 30$ kHz/rad. En el mismo papel logarítmico, grafique la respuesta con un filtro de un polo con frecuencia de corte $\omega_c = 200$ rad/s, y un filtro de dos polos con la misma frecuencia de corte.

2-27. Calcule el cambio de frecuencia de un VCO con función de transferencia $K_o = 4$ kHz/V y un cambio de voltaje de entrada $\Delta V = 1.2\ V_p$.

2-28. Calcule el voltaje en la salida de un comparador de fases con función de transferencia $K_d = 0.4$ V/rad, y un error de fase $\theta_e = 0.55$ rad.

2-29. Calcule el intervalo de retención de un PLL con ganancia de lazo abierto $K_v = 25$ kHz/rad.

2-30. Calcule el error de fase necesario para producir un desplazamiento de frecuencia de 20 kHz en un PLL con ganancia de lazo abierto $K_v = 50$ kHz/rad.

C A P Í T

Imagen protegida por derechos de autor

Transmisión por modulación de amplitud

INTRODUCCIÓN

Las señales de transmisión se transportan entre un transmisor y un receptor a través de alguna forma de medio de transmisión. Sin embargo, casi nunca tienen las señales de información una forma adecuada para su transmisión. En consecuencia, se deben transformar a una forma más adecuada. El proceso de imprimir señales de información de baja frecuencia en una *señal portadora* de alta frecuencia se llama *modulación*. La *demodulación* es el proceso inverso, donde las señales recibidas se regresan a su forma original. El objetivo de este capítulo es explicar al lector los conceptos fundamentales de la *modulación de amplitud* (AM).

PRINCIPIOS DE MODULACIÓN DE AMPLITUD

La *modulación de amplitud* (AM, por *amplitude modulation*; en español se usa "amplitud modulada") es el proceso de cambiar la amplitud de una señal portadora de frecuencia relativamente alta, en proporción con el valor instantáneo de la señal modulante o moduladora (información). La modulación de amplitud es una forma de modulación relativamente poco costosa y de baja calidad, que se usa para emisiones comerciales de señales de audio y de video. También se usa para radiocomunicaciones móviles en dos sentidos, como por ejemplo los radios de banda civil (CB).

Los moduladores de AM son dispositivos no lineales, con dos entradas y una salida. Una entrada es una sola señal portadora de alta frecuencia y amplitud constante, y la segunda está formada por señales de información, de frecuencia relativamente baja, que puede tener una sola frecuencia, o ser una forma compleja de onda, formada a su vez por muchas frecuencias. Las frecuencias que son lo suficientemente altas como para irradiarse en forma eficiente de una antena, y propagarse por el espacio libre se suelen llamar *radiofrecuencias*, o simplemente RF. En el modulador, la información actúa sobre, o modula, la portadora de RF y produce una forma modulada de onda. La señal de información puede tener una sola frecuencia, o con más probabilidad, puede consistir en un intervalo de frecuencias. Por ejemplo, en un sistema normal de comunicaciones de voz se usa un intervalo de frecuencias de información de 300 a 3000 Hz. A la

forma de onda modulada de salida de un modulador de AM se le llama con frecuencia envolvente de AM.

La envolvente de AM

Aunque hay varias clases de modulación de amplitud, la que probablemente se usa con más frecuencia es la AM de *portadora de máxima potencia y doble banda lateral* (DSBFC, por *double-sideband full carrier*). A este sistema se le llama también AM *convencional* o simplemente AM. La fig. 3-1 ilustra la relación entre la portadora [V_c sen($2\pi f_c t$)], la señal moduladora [V_m sen($2\pi f_m t$)] y la onda modulada [$V_{am}(t)$] en la AM convencional. Allí se ve cómo se produce una forma de onda de AM cuando una señal moduladora de una sola frecuencia actúa sobre una señal portadora de alta frecuencia. La forma de onda de salida contiene todas las frecuencias que forman la señal de AM, y se usa para transportar la información por el sistema. Por consiguiente, la forma de la onda modulada se llama *envolvente de AM*. Nótese que cuando no hay señal modulante, la forma de onda de salida no es más que la señal de la portadora. Sin embargo, cuando se aplica una señal moduladora, varía la amplitud de la onda de salida, de acuerdo con la señal moduladora. Nótese que la frecuencia de repetición de la envolvente es igual a la frecuencia de la señal moduladora, y que la forma de la envolvente es idéntica a la forma de la señal moduladora.

Espectro de frecuencias y ancho de banda de AM

Un modulador de AM es un dispositivo no lineal. En consecuencia, hay mezclado no lineal, y la envolvente de salida es una onda compleja formada por un voltaje de cd, la frecuencia de la portadora y la suma ($f_c + f_m$) y diferencia ($f_c - f_m$) de las frecuencias, es decir, los productos cruzados. Las frecuencias de suma y diferencia están desplazadas respecto a la frecuencia de la portadora una cantidad igual a la frecuencia de la señal moduladora. Por consiguiente, un espectro de señal de AM contiene los componentes de frecuencia apartados f_m Hz a ambos lados de la portadora. Sin embargo, se debe hacer notar que la onda modulada no contiene un componente de frecuencia que sea igual a la de la señal moduladora. El efecto de la modulación es trasladar

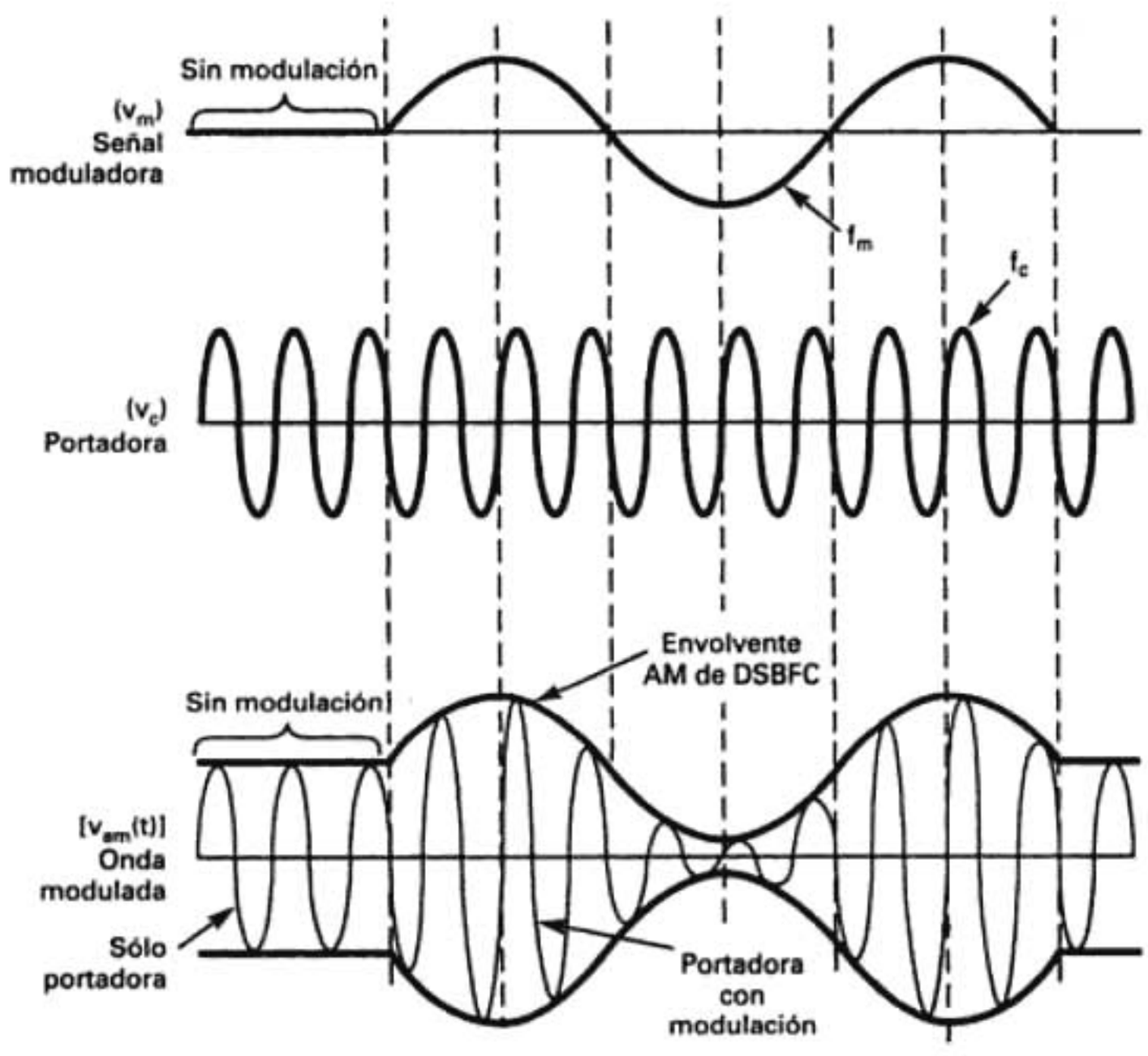


FIGURA 3-1 Generación de AM

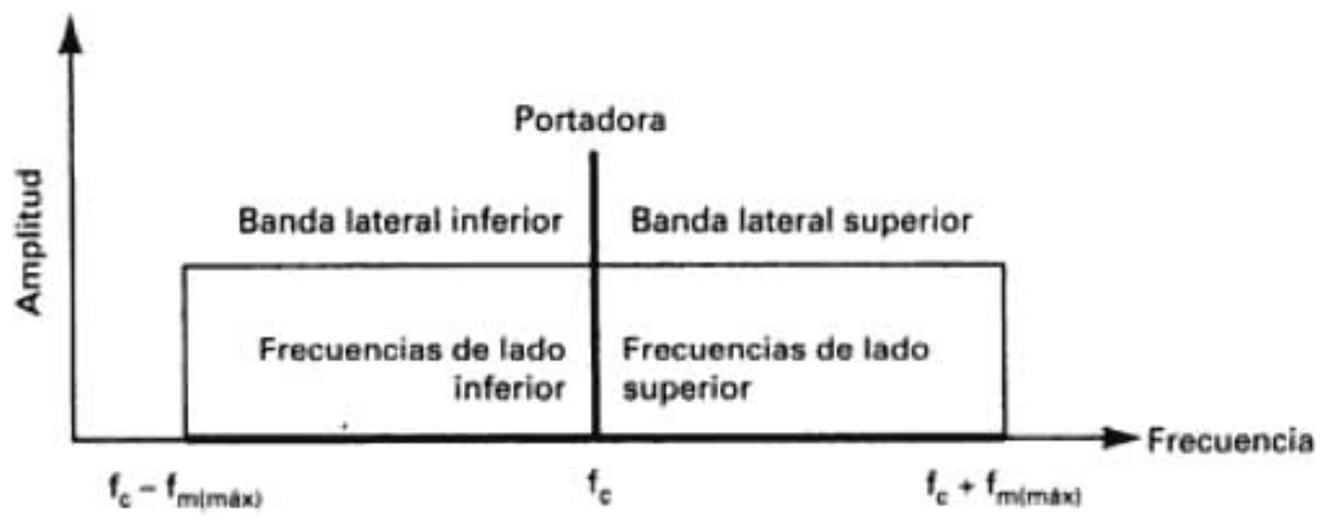


FIGURA 3-2 Espectro de frecuencias de una onda AM de DSBFC

la señal moduladora en el dominio de la frecuencia, de modo que se refleje simétricamente respecto a la frecuencia de la portadora.

La fig. 3-2 muestra el espectro de frecuencias para una onda de AM. Este espectro se extiende desde $f_c - f_{m(máx)}$ hasta $f_c + f_{m(máx)}$, siendo f_c la frecuencia de la portadora y $f_{m(máx)}$ la frecuencia máxima de la señal moduladora. La banda de frecuencias entre $f_c - f_{m(máx)}$ y f_c se llama *banda lateral inferior* (LSB, de *lower sideband*) y toda frecuencia dentro de esta banda es una *frecuencia de lado inferior* (LSF, de *lower side frequency*). La banda de frecuencias entre f_c y $f_c + f_{m(máx)}$ se llama *banda lateral superior* (USB, de *upper sideband*) y las frecuencias dentro de esta banda se llaman *frecuencias de lado superior* (USF, de *upper side frequency*). Por consiguiente, el ancho de banda (B) de una onda DSBFC de AM es igual a la diferencia entre la frecuencia máxima de lado superior y la mínima del lado inferior, o también, igual a dos veces la frecuencia máxima de la señal modulante, es decir, $B = 2f_{m(máx)}$. Para propagación de ondas de radio, la portadora y todas las frecuencias dentro de las bandas laterales superior e inferior deben ser lo bastante elevadas como para poder propagarse lo suficiente a través de la atmósfera terrestre.

Ejemplo 3-1

Para un modulador DSBFC de AM con frecuencia de portadora $f_c = 100$ kHz y una señal moduladora de frecuencia máxima $f_{m(máx)} = 5$ kHz determinar:

(a) Límites de frecuencia de las bandas laterales superior e inferior.
(b) Ancho de banda.
(c) Frecuencias de lado superior e inferior, que se producen cuando la señal moduladora es un tono de frecuencia única de 3 kHz.
(d) Trazar el espectro de frecuencias de salida.

Solución (a) La banda lateral inferior va desde la frecuencia mínima posible de lado inferior hasta la frecuencia de portadora, es decir

$$\begin{aligned} \text{LSB} &= [f_c - f_{m(máx)}] \text{ a } f_c \\ &= (100 - 5) \text{ kHz a } 100 \text{ kHz} = 95 \text{ a } 100 \text{ kHz} \end{aligned}$$

La banda lateral superior va desde la frecuencia de la portadora hasta la frecuencia máxima posible de lado superior, es decir

$$\begin{aligned} \text{USB} &= f_c \text{ a } [f_c + f_{m(máx)}] \\ &= 100 \text{ kHz a } (100 + 5) \text{ kHz} = 100 \text{ a } 105 \text{ kHz} \end{aligned}$$

(b) El ancho de banda es igual a la diferencia entre la frecuencia máxima de lado superior y la frecuencia mínima de lado inferior, es decir

$$\begin{aligned} B &= 2f_{m(máx)} \\ &= 2(5 \text{ kHz}) = 10 \text{ kHz} \end{aligned}$$

(c) La frecuencia de lado superior es la suma de las frecuencias de la portadora y la moduladora

$$f_{fls} = f_c + f_m = 100 \text{ kHz} + 3 \text{ kHz} = 103 \text{ kHz}$$

La frecuencia de lado inferior es la diferencia de frecuencias de la portadora y la moduladora

$$f_{fli} = f_c - f_m = 100 \text{ kHz} - 3 \text{ kHz} = 97 \text{ kHz}$$

(d) El espectro de frecuencias de salida se ve en la fig. 3-3.

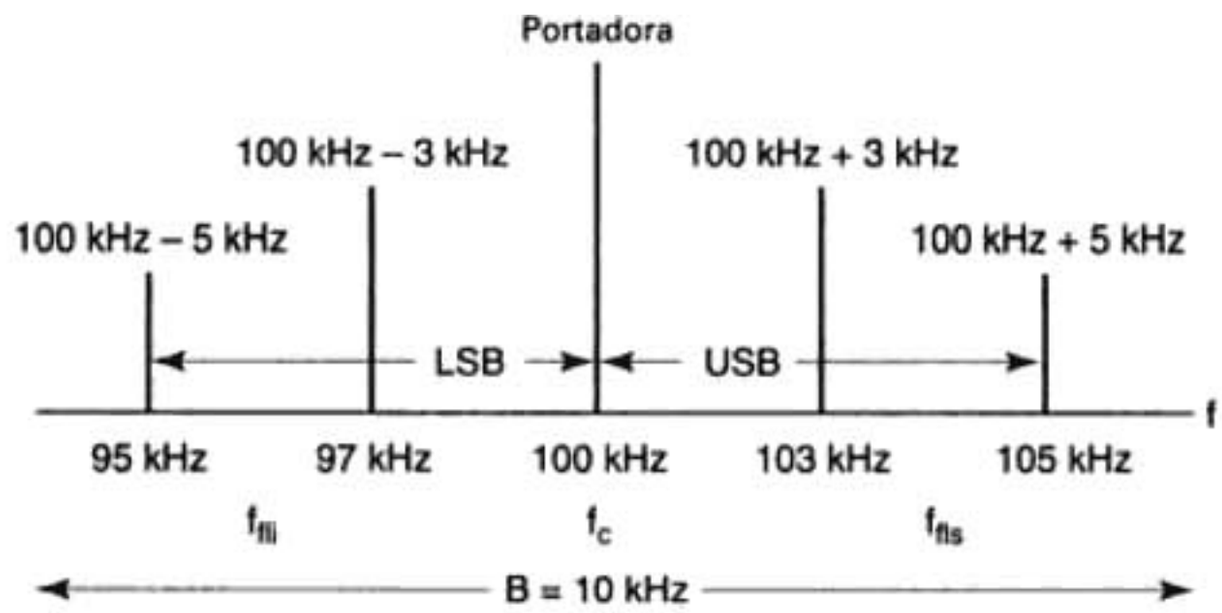


FIGURA 3-3 Espectro de salida para el ejemplo 3-1

Representación fasorial de una onda de amplitud modulada

Con una señal moduladora de frecuencia única, la envolvente de AM se obtiene de la suma vectorial de la portadora y de las frecuencias del lado superior e inferior. Las dos frecuencias laterales se combinan y producen una resultante que se combina con el vector de la portadora. La fig. 3-4a muestra esta suma fasorial. Los fasores de las frecuencias de la portadora y de los lados superior e inferior giran en dirección contraria a la de las manecillas del reloj. Sin embargo, la frecuencia de lado superior gira con más rapidez que la portadora ($\omega_{fls} > \omega_c$) y la frecuencia de lado inferior gira más lento ($\omega_{fli} < \omega_c$). En consecuencia, si se mantiene estacionario el fasor de la portadora, el fasor de la frecuencia de lado superior continúa girando en dirección contraria a las manecillas del reloj en relación con la portadora, y el de la frecuencia del lado inferior gira en dirección de las manecillas del reloj. Los fasores de las frecuencias de portadora y del lado superior e inferior se combinan, a veces en fase (se suman) y a veces desfasados (se restan). Para la forma de onda de la fig. 3-4b, la amplitud positiva máxima de la envolvente se presenta cuando las frecuencias de la portadora y de los lados superior e inferior están en sus valores positivos máximos al mismo tiempo ($+V_{máx} = V_c + V_{fls} + V_{fli}$). La amplitud positiva mínima de la envolvente se produce cuando la portadora tiene su valor máximo positivo, y al mismo tiempo las frecuencias laterales están en sus valores negativos máximos ($+V_{mín} = V_c - V_{fls} - V_{fli}$). La amplitud negativa máxima se produce cuando las frecuencias de la portadora y de las bandas laterales superior e inferior tienen sus valores negativos máximos al mismo tiempo ($-V_{máx} = -V_c - V_{fls} - V_{fli}$). La amplitud negativa mínima se produce cuando la portadora tiene su valor negativo máximo y al mismo tiempo las frecuencias de las bandas superior e inferior tienen sus valores positivos máximos ($-V_{mín} = -V_c + V_{fls} + V_{fli}$).

Coeficiente de modulación y porcentaje de modulación

Un término que describe la cantidad de cambio de amplitud (modulación) que hay en una forma de onda de AM es el *coeficiente de modulación*. El *porcentaje de modulación*, o *modulación porcentual*, es simplemente el coeficiente de modulación expresado como porcentaje. En forma más específica, el porcentaje de modulación indica el cambio porcentual de amplitud de la onda de salida cuando sobre la portadora actúa una señal moduladora. La definición matemática del coeficiente de modulación es

$$m = \frac{E_m}{E_c} \tag{3-1}$$

en la que m = coeficiente de modulación (adimensional)
E_m = cambio máximo de amplitud de la forma de onda de voltaje de salida (volts)
E_c = amplitud máxima del voltaje de la portadora no modulada (volts)

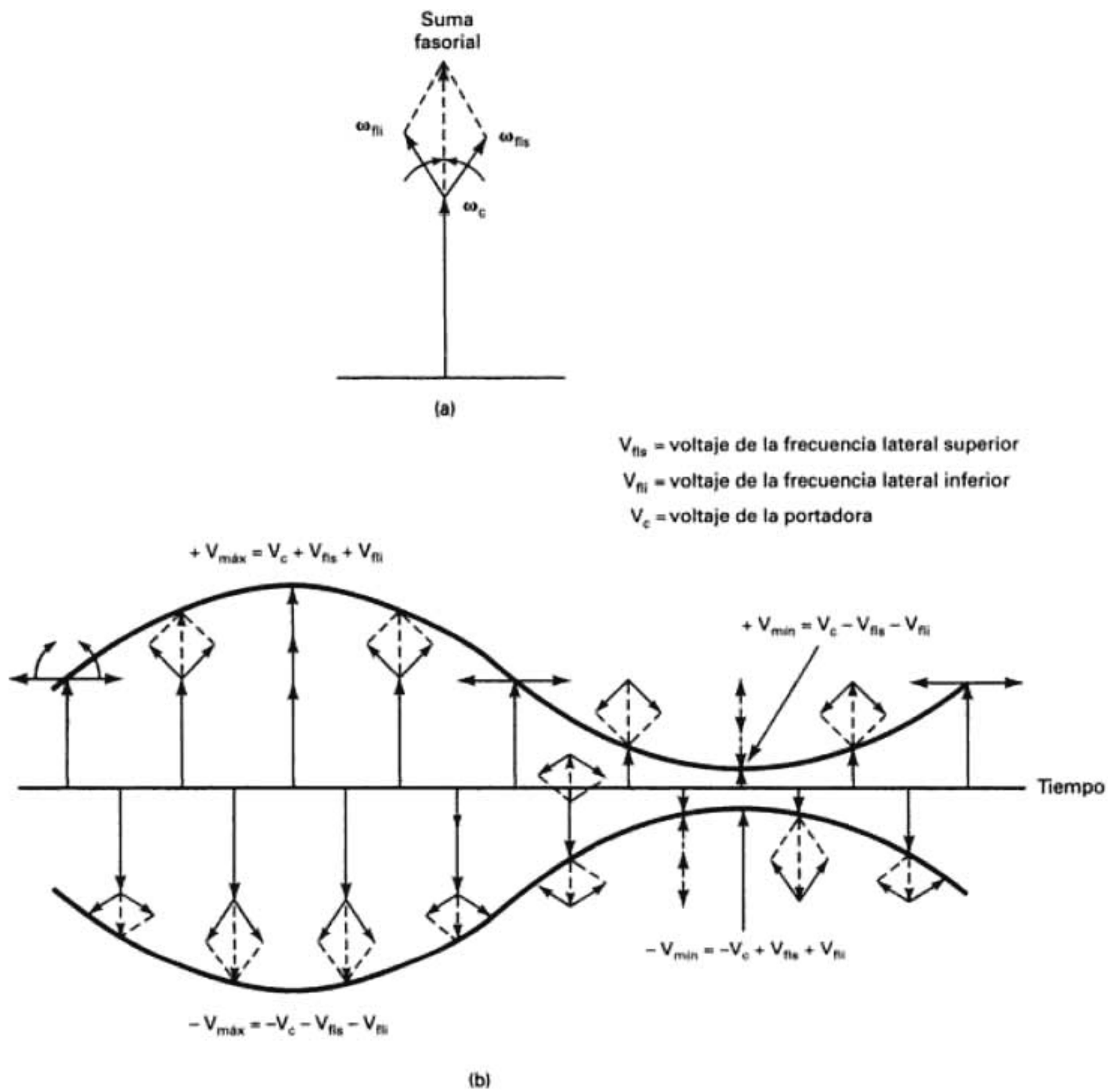


FIGURA 3-4 Suma fasorial en una envolvente de AM de DSBFC: (a) suma fasorial de la portadora y las frecuencias laterales superior e inferior; (b) suma fasorial que produce la envolvente de AM

La ecuación 3-1se puede reordenar, despejando a E_m y a E_c, como sigue

$$E_m = mE_c \tag{3-2}$$

y

$$E_c = \frac{E_m}{m} \tag{3-3}$$

y el porcentaje M de modulación es

$$M = \frac{E_m}{E_c} \times 100 \text{ o simplemente } m \times 100 \tag{3-4}$$

En la fig. 3-5 se ve la relación entre m, E_m y E_c.

Si la señal moduladora es una onda senoidal pura de una sola frecuencia, y el proceso de modulación es simétrico (es decir, las diferencias positiva y negativa de amplitud de la envol-

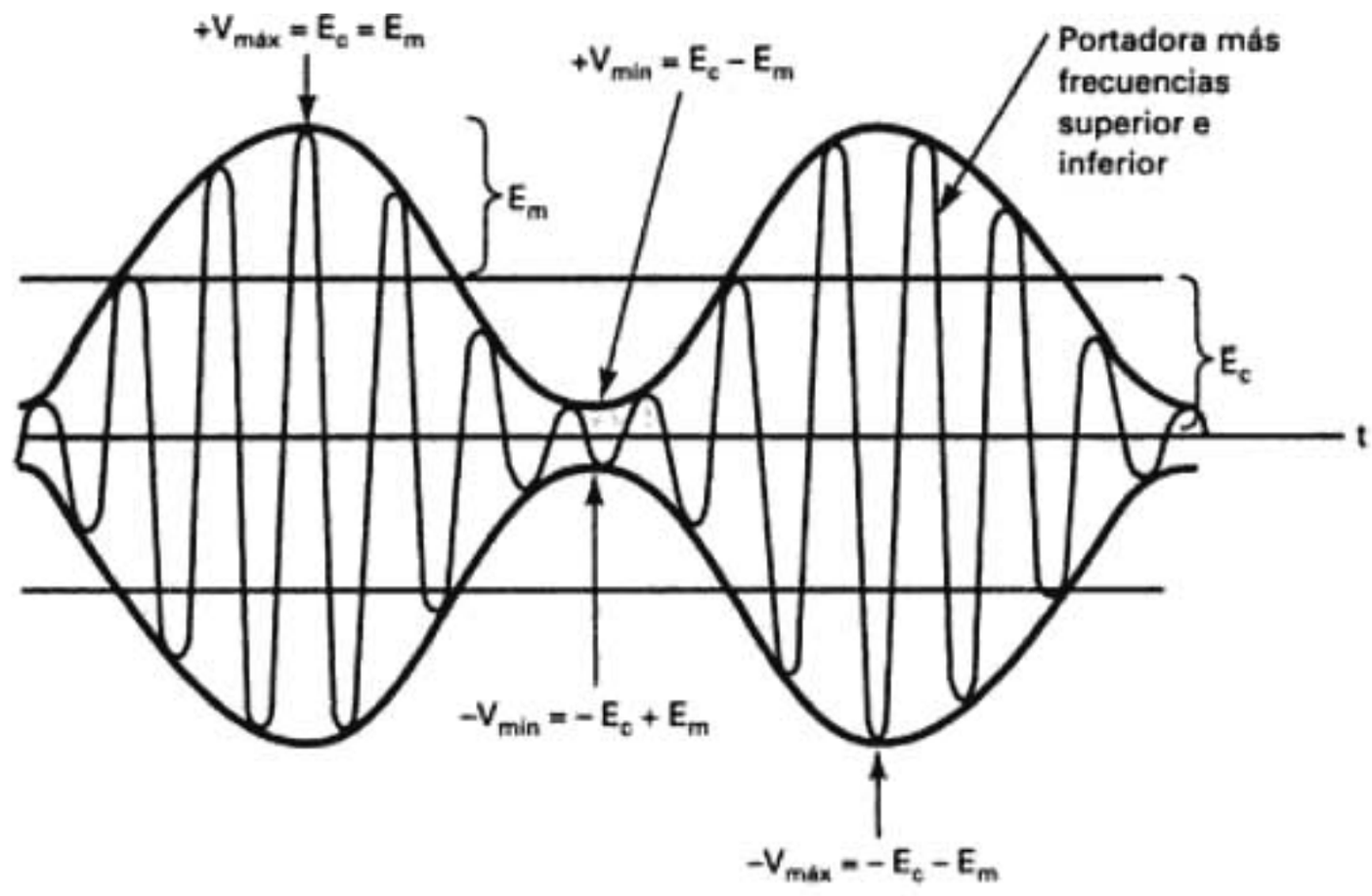


FIGURA 3-5 Coeficiente de modulación, E_m y E_c

vente son iguales), entonces se puede deducir como sigue el porcentaje de modulación (véase la fig. 3-5 durante esta deducción)

$$E_m = \frac{1}{2}(V_{\text{máx}} - V_{\text{mín}}) \tag{3-5}$$

y

$$E_c = \frac{1}{2}(V_{\text{máx}} + V_{\text{mín}}) \tag{3-6}$$

Por consiguiente,

$$M = \frac{1/2(V_{\text{máx}} - V_{\text{mín}})}{1/2(V_{\text{máx}} + V_{\text{mín}})} \times 100$$

$$= \frac{(V_{\text{máx}} - V_{\text{mín}})}{(V_{\text{máx}} + V_{\text{mín}})} \times 100 \tag{3-7}$$

en donde $V_{\text{máx}} = E_c + E_m$
$V_{\text{mín}} = E_c - E_m$

El cambio máximo de amplitud de la onda de salida E_m, es la suma de los voltajes de las frecuencias laterales superior e inferior. Por consiguiente, ya que $E_m = E_{\text{fls}} + E_{\text{fli}}$ y $E_{\text{fls}} = E_{\text{fli}}$, por lo tanto

$$E_{\text{fls}} = E_{\text{fli}} = \frac{E_m}{2} = \frac{1/2(V_{\text{máx}} - V_{\text{mín}})}{2} = \frac{1}{4}(V_{\text{máx}} - V_{\text{mín}}) \tag{3-8}$$

en la que E_{fls} = amplitud máxima de la frecuencia de lado superior (volts)
E_{fli} = amplitud máxima de la frecuencia de lado inferior (volts)

En la ecuación 3-1 se ve que el porcentaje de modulación llega a 100% cuando $E_m = E_c$. Esta condición se muestra en la fig. 3-6d. También se ve que al 100% de modulación, la amplitud mínima de la envolvente es $V_{\text{mín}} = 0$ V. La fig. 3-6c muestra una envolvente modulada 50%. El cambio máximo de amplitud de la envolvente es igual a la mitad de la amplitud de la onda no modulada. La modulación porcentual máxima que se puede aplicar sin causar demasiada distorsión es 100%. A veces, el porcentaje de modulación se expresa como el cambio máximo de voltaje de la onda modulada con respecto a la amplitud máxima de la portadora no modulada (es decir, cambio porcentual = $(\Delta E_c/E_c) \times 100$).

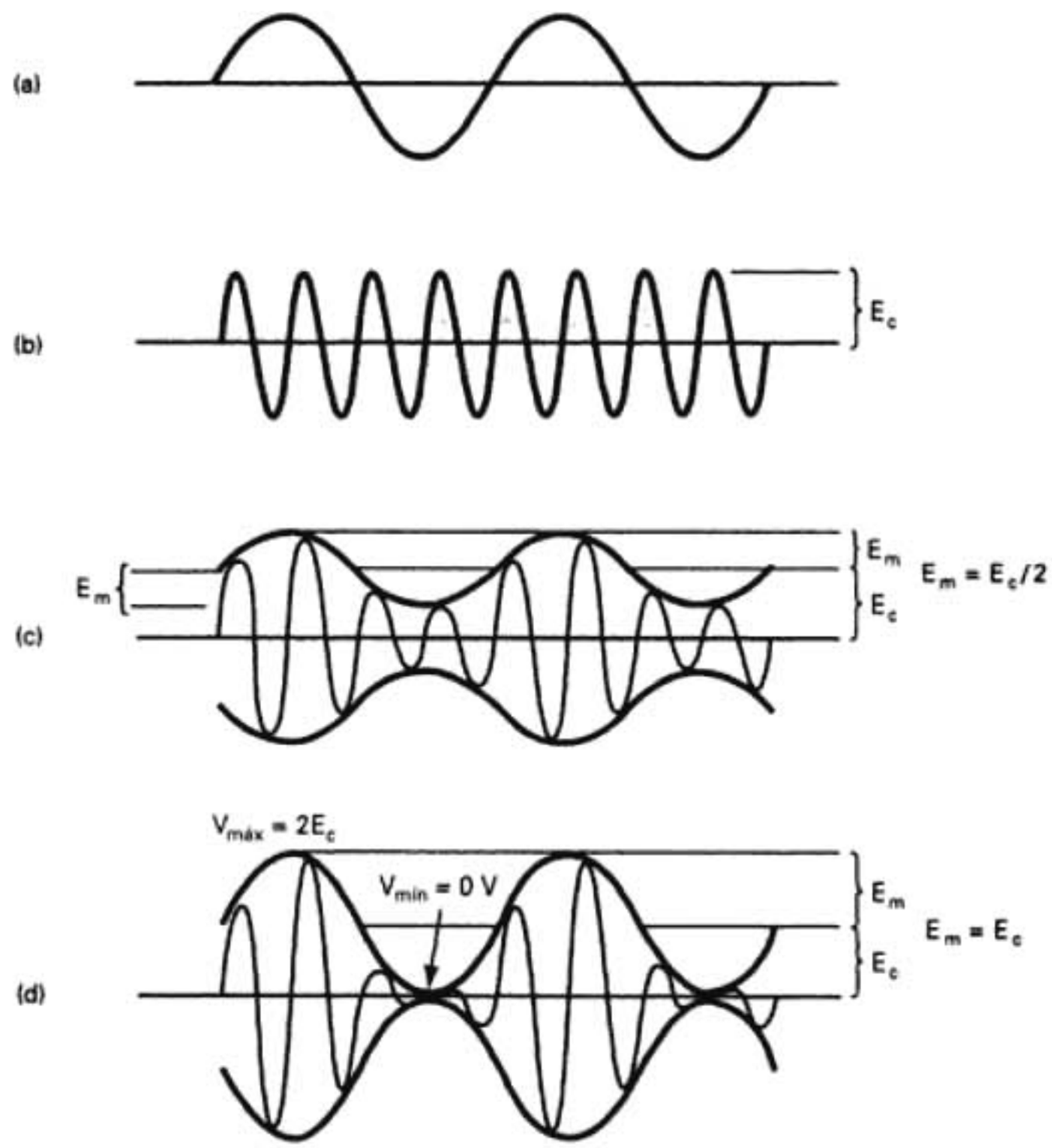


FIGURA 3-6 Porcentaje de modulación de una envolvente de AM de DSBFC: (a) señal moduladora; (b) portadora no modulada; (c) onda modulada 50%; (d) onda modulada 100%

Ejemplo 3-2

Determinar, para la forma de onda de AM en la fig. 3-7:

(a) Amplitud máxima de las frecuencias de lado superior e inferior.
(b) Amplitud máxima de la portadora no modulada.
(c) Cambio máximo de amplitud de la envolvente.
(d) Coeficiente de modulación.
(e) Porcentaje de modulación.

Solución (a) De la ecuación 3-8,

$$E_{fls} = E_{fli} = \frac{1}{4}(18 - 2) = 4 \text{ V}$$

(b) De la ecuación 3-6,

$$E_c = \frac{1}{2}(18 + 2) = 10 \text{ V}$$

(c) De la ecuación 3-5,

$$E_m = \frac{1}{2}(18 - 2) = 8 \text{ V}$$

(d) De la ecuación 3-1,

$$m = \frac{8}{10} = 0.8$$

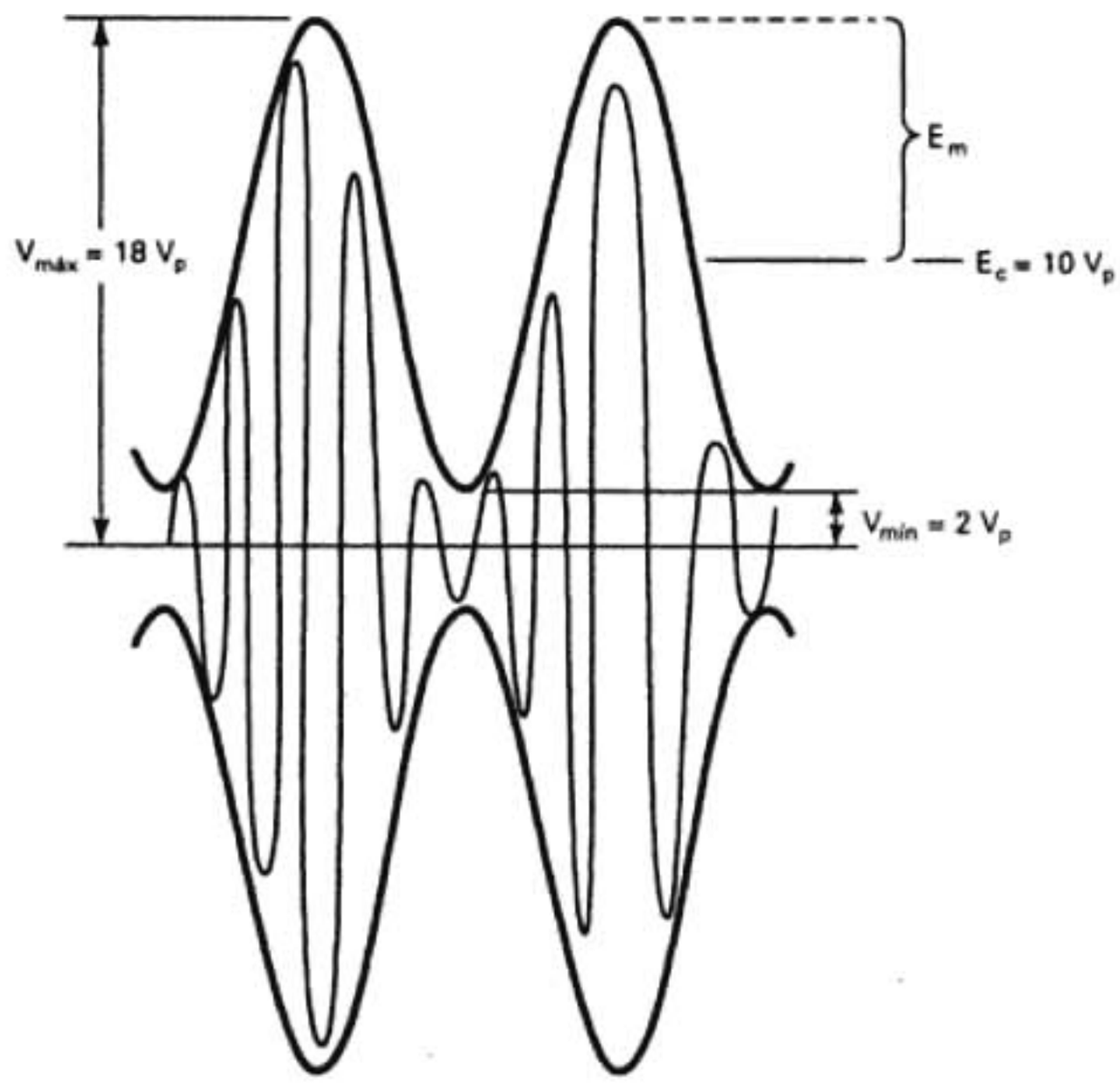


FIGURA 3-7 Envolvente de AM para el ejemplo 3-2

(e) De la ecuación 3-4,

$$M = 0.8 \times 100 = 80\%$$

y de la ecuación 3-7,

$$M = \frac{18 - 2}{18 + 2} \times 100 = 80\%$$

Distribución de voltaje de AM

Una portadora no modulada se puede describir matemáticamente como sigue

$$v_c(t) = E_c \operatorname{sen}(2\pi f_c t)$$

en donde $v_c(t)$ = forma de onda de voltaje de la portadora, variable en el tiempo
E_c = amplitud máxima de la portadora (volts)
f_c = frecuencia de la portadora (hertz)

Se hizo notar, en una sección anterior, que la rapidez de repetición de una envolvente de AM es igual a la frecuencia de la señal moduladora, que la amplitud de la onda de AM varía en proporción con la de la señal moduladora, y que la amplitud máxima de la onda modulada es igual a $E_c + E_m$. Por lo anterior, la amplitud instantánea de la onda modulada se puede expresar como sigue

$$v_{am}(t) = [E_c + E_m \operatorname{sen}(2\pi f_m t)][\operatorname{sen}(2\pi f_c t)] \tag{3-9a}$$

donde $[E_c + E_m \operatorname{sen}(2\pi f_m t)]$ = amplitud de la onda modulada
E_m = cambio máximo de amplitud de la envolvente (volts)
f_m = frecuencia de la señal moduladora (hertz)

Si se sustituye E_m por mE_c,

$$v_{am}(t) = [(E_c + mE_c \operatorname{sen}(2\pi f_m t)][\operatorname{sen}(2\pi f_c t)] \tag{3-9b}$$

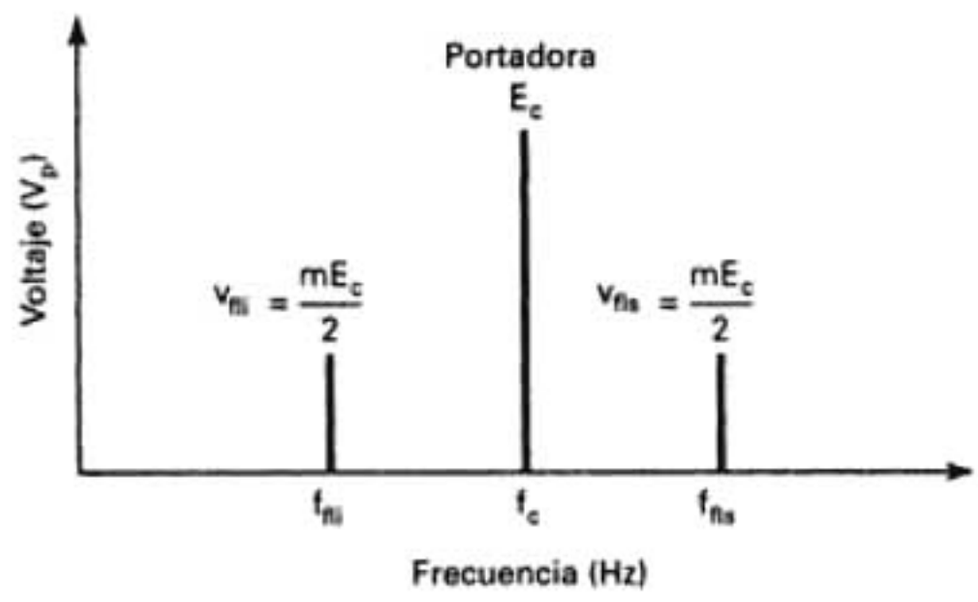


FIGURA 3-8 Espectro de voltajes para una onda DSBFC de AM

siendo $[E_c + mE_c \text{ sen}(2\pi f_m t)]$ la amplitud de la onda modulada.

Se saca a E_c como factor común y se rearregla la ecuación 3-9b:

$$v_{am}(t) = [1 + m \text{ sen}(2\pi f_m t)][E_c \text{ sen}(2\pi f_c t)] \tag{3-9c}$$

donde $[1 + m \text{ sen}(2\pi f_m t)]$ = voltaje constante + señal moduladora
$[E_c \text{ sen}(2\pi f_c t)]$ = voltaje de portadora no modulada

Se ve, en la ecuación 3-9c, que la señal moduladora contiene un componente constante (1) y un componente senoidal con la frecuencia de la señal moduladora $[m \text{ sen}(2\pi f_m t)]$. El siguiente análisis demostrará por qué el componente constante produce el componente de la portadora en la onda modulada, y el componente senoidal produce las frecuencias laterales. Si se hace la multiplicación en la ecuación 3-9b o c, se obtiene

$$v_{am}(t) = E_c \text{ sen}(2\pi f_c t) + [mE_c \text{ sen}(2\pi f_m t)][\text{sen}(2\pi f_c t)]$$

Por consiguiente,

$$v_{am}(t) = E_c \text{ sen}(2\pi f_c t) - \frac{mE_c}{2}\cos[2\pi(f_c + f_m)t] + \frac{mE_c}{2}\cos[2\pi(f_c - f_m)t] \tag{3-10}$$

en donde $E_c \text{ sen}(2\pi f_c t)$ = señal portadora (volts)
$-(mE_c/2)\cos[2\pi(f_c + f_m)t]$ = señal de la frecuencia de lado superior (volts)
$+(mE_c/2)\cos[2\pi(f_c - f_m)t]$ = señal de la frecuencia de lado inferior (volts)

De la ecuación 3-10 se deben hacer notar algunas características interesantes acerca de la modulación de amplitud con portadora completa y doble banda lateral (DSBFC). La primera es que se observa que la amplitud de la portadora después de la modulación es igual a la que era antes de la modulación, E_c. Por consiguiente, la amplitud de la portadora no se altera por el proceso de modulación. En segundo lugar, que la amplitud de las frecuencias de lado superior e inferior depende tanto de la amplitud de la portadora como del coeficiente de modulación. Para 100% de modulación, $m = 1$ y las amplitudes de las bandas superior e inferior de frecuencias son iguales, las dos, a $E_c/2$, la mitad de la amplitud de la portadora. Por consiguiente, con 100% de modulación,

$$V_{(\text{máx})} = E_c - \frac{E_c}{2} - \frac{E_c}{2} = 2E_c$$

y

$$V_{(\text{mín})} = E_c - \frac{E_c}{2} - \frac{E_c}{2} = 0 \text{ V}$$

Según las ecuaciones anteriores y la ecuación 3-10, es evidente que, siempre que no se rebase el 100% de modulación, la amplitud máxima de los máximos de una envolvente de AM es $V_{(\text{máx})} = 2E_c$, y la amplitud mínima del mínimo de una envolvente de AM es $V_{(\text{mín})} = 0$ V. Esta relación se vió en la fig. 3-6d. La fig. 3-8 muestra el espectro de voltajes de una onda DSBFC de AM; nótese que todos los voltajes se indican con valores máximos.

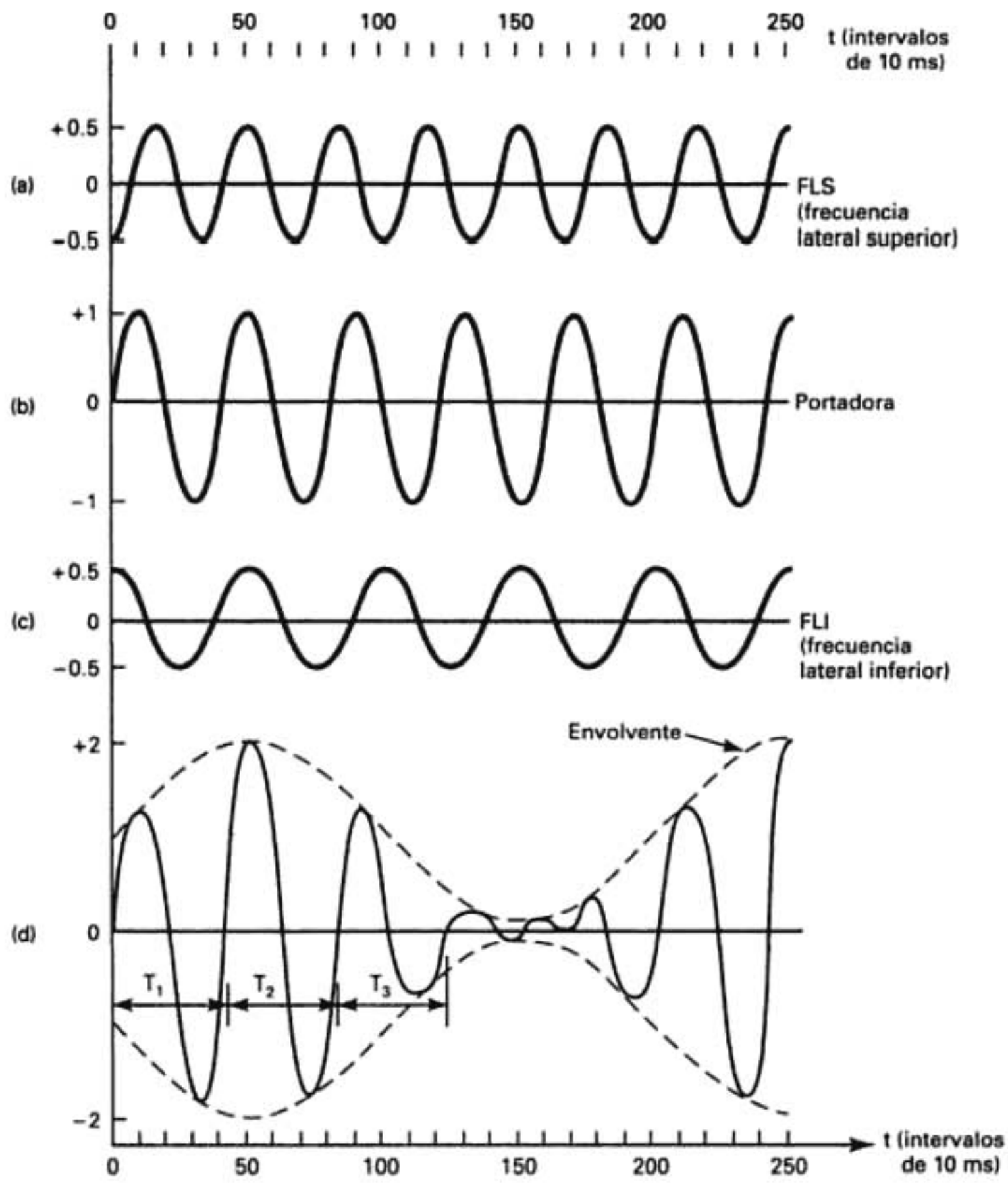


FIGURA 3-9 Generación de una envolvente de AM de DSBFC, representada en el dominio del tiempo: (a) $-\frac{1}{2}(2\pi 30t)$; (b) sen$(2\pi 25t)$; (c) $+\frac{1}{2}\cos(2\pi 20t)$; (d) suma de (a), (b) y (c)

También, de acuerdo con la ecuación 3-10, es evidente la relación de fases relativas de la portadora y las frecuencias laterales superior e inferior. La componente portadora es una función seno +, la frecuencia lateral superior es una función coseno − y la frecuencia de lado inferior es una función coseno +. También, envolvente es una forma de onda repetitiva. Así, al principio de cada ciclo de la envolvente, la portadora está desfasada 90° con respecto a las frecuencias laterales superior e inferior, y éstas están desfasadas 180° entre sí. Esta relación de fases se puede ver en la fig. 3-9 para $f_c = 25$ Hz y $f_m = 5$ Hz.

Ejemplo 3-3

Una entrada a un modulador convencional de AM es una portadora de 500 kHz, con amplitud de 20 V_p de amplitud. La segunda entrada es una señal moduladora de 10 kHz, de amplitud suficiente para causar un cambio de ± 7.5 V_p en la onda de salida. Determinar:

(a) Las frecuencias de lado superior e inferior.
(b) El coeficiente de modulación y el porcentaje de modulación.
(c) La amplitud máxima de la portadora modulada, y los voltajes de frecuencias laterales superior e inferior.
(d) Amplitudes máxima y mínima de la envolvente.
(e) La ecuación de la onda modulada.

A continuación:
(f) Trazar el espectro de salida.
(g) Trazar la envolvente de salida.

Solución (a) Las frecuencias laterales superior e inferior no son más que las frecuencias de suma y de diferencia, respectivamente

$$f_{ls} = 500 + 10 \text{ kHz} = 510 \text{ kHz}$$
$$f_{li} = 500 - 10 \text{ kHz} = 490 \text{ kHz}$$

(b) El coeficiente de modulación se calcula con la ecuación 3-1

$$m = \frac{7.5}{20} = 0.375$$

El porcentaje de modulación se calcula con la ecuación 3-4

$$M = 100 \times 0.375 = 37.5\%$$

(c) La amplitud máxima de la portadora modulada y las frecuencias laterales superior e inferior son

$$E_c \text{ (modulada)} = E_c \text{ (no modulada)} = 20 \text{ V}_p$$
$$E_{ls} = E_{li} = \frac{mE_c}{2} = \frac{(0.375)(20)}{2} = 3.75 \text{ V}_p$$

(d) Las amplitudes máxima y mínima de la envolvente se calculan como sigue

$$V_{(\text{máx})} = E_c + E_m = 20 + 7.5 = 27.5 \text{ V}_p$$
$$V_{(\text{mín})} = E_c - E_m = 20 - 7.5 = 12.5 \text{ V}_p$$

(e) La ecuación de la onda modulada se apega al formato de la ecuación 3-10

$$v_{am}(t) = 20 \text{ sen}(2\pi 600kt) - 3.75 \cos(2\pi 510kt) + 3.75 \cos(2\pi 490kt)$$

(f) El espectro de salida se ve en la fig. 3-10.
(g) La envolvente modulada se ve en la fig. 3-11

Análisis de AM en el dominio del tiempo

La fig. 3-9 muestra cómo se produce una envolvente de DSBFC de AM con la suma algebraica de las formas de onda de la portadora, y de las frecuencias laterales superior e inferior. Para simplificar, se usaron las siguientes ondas como señales de entrada moduladora y portadora

$$\text{portadora} = v_c(t) = E_c \text{ sen}(2\pi 25t) \tag{3-11}$$

$$\text{señal moduladora} = v_m(t) = E_m \text{ sen}(2\pi 5t) \tag{3-12}$$

Si se sustituyen las ecuaciones 3-11 y 3-12 en la ecuación 3-10, la ecuación de la onda modulada es

$$v_{am}(t) = E_c \text{ sen}(2\pi 25t) - \frac{mE_c}{2}\cos(2\pi 30t) + \frac{mE_c}{2}\cos(2\pi 20t) \tag{3-13}$$

en la que
$E_c \text{ sen}(2\pi 25t)$ = portadora (volts)
$-(mE_c/2)\cos(2\pi 30t)$ = frecuencias en la banda lateral superior (volts)
$+(mE_c/2)\cos(2\pi 20t)$ = frecuencias en la banda lateral inferior (volts)

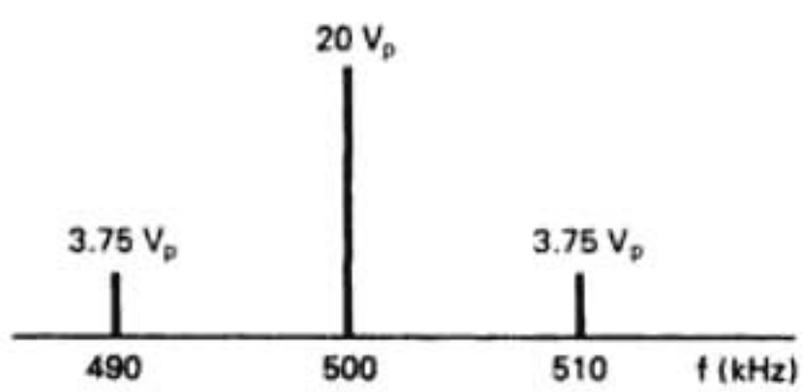


FIGURA 3-10 Espectro de salida para el ejemplo 3-3

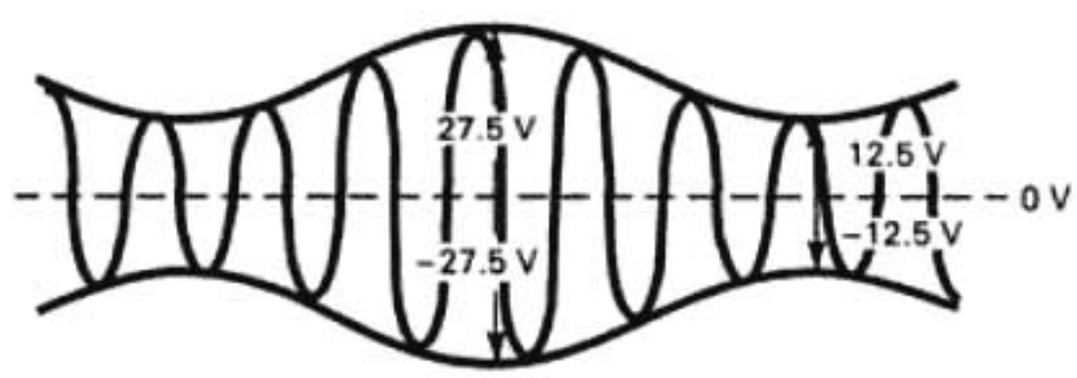


FIGURA 3-11 Envolvente de AM para el ejemplo 3-3

La tabla 3-1 es una lista de valores de los voltajes instantáneos de la portadora, los de la frecuencia lateral superior e inferior y los la onda modulada total, cuando se sustituyen valores de 0 a 250 ms en la ecuación 3-13, en intervalos de 10 ms. El voltaje de portadora no modulada es $E_c = 1\ V_p$, y se alcanza modulación de 100%. En la fig. 3-9 se ven las formas de onda correspondientes. Obsérvese que el voltaje máximo de la envolvente es 2 V ($2E_c$) y que el voltaje mínimo de la envolvente es 0 V.

En la fig. 3-9, nótese que es constante el tiempo entre cruces correspondientes de cero dentro de la envolvente, es decir, que $T_1 = T_2 = T_3$, etcétera. También, nótese que las amplitudes de los máximos sucesivos dentro de la envolvente no son iguales. Esto indica que un ciclo dentro de la envolvente no es una onda senoidal pura y, en consecuencia, que la onda modulada debe estar formada por más de una frecuencia; es la suma de las frecuencias de la portadora y las laterales superior e inferior. La fig. 3-9 también muestra que la amplitud de la portadora no varía, sino más bien la amplitud de la envolvente varía de acuerdo con la señal moduladora. Esto se logra sumando las frecuencias laterales superior e inferior a la forma de onda de la portadora.

Distribución de potencia en AM

En todo circuito eléctrico, la potencia disipada es igual al cuadrado del voltaje dividido entre la resistencia. Así, el promedio de la potencia disipada en una carga, por una portadora no modulada, es igual al cuadrado del voltaje rms (rms = efectivo, o de raíz cuadrática media) de la portadora, dividido entre la resistencia de carga. Esto se expresa con la siguiente ecuación

$$P_c = \frac{(0.707E_c)^2}{R} = \frac{(E_c)^2}{2R} \tag{3-14}$$

TABLA 3-1 Voltajes instantáneos

FLS, $-\frac{1}{2}\cos(2\pi 30t)$	Portadora, $\text{sen}(2\pi 25t)$	FLI, $+\frac{1}{2}\cos(2\pi 20t)$	Envolvente, $v_{am}(t)$	Tiempo, t(ms)
−0.5	0	+0.5	0	0
+0.155	+1	+0.155	+1.31	10
+0.405	0	−0.405	0	20
−0.405	−1	−0.405	−1.81	30
−0.155	0	+0.155	0	40
+0.5	+1	+0.5	2	50
−0.155	0	+0.155	0	60
−0.405	−1	−0.405	−1.81	70
+0.405	0	−0.405	0	80
+0.155	+1	+0.155	+1.31	90
−0.5	0	+0.5	0	100
+0.155	−1	+0.155	−0.69	110
+0.405	0	−0.405	0	120
−0.405	+1	−0.405	+0.19	130
−0.155	0	+0.155	0	140
+0.5	−1	+0.5	0	150
−0.155	0	+0.155	0	160
−0.405	+1	−0.405	+0.19	170
+0.405	0	−0.405	0	180
+0.155	−1	+0.155	−0.69	190
−0.5	0	+0.5	0	200
+0.155	+1	+0.155	+1.31	210
+0.405	0	−0.405	0	220
−0.405	−1	−0.405	−1.81	230
+0.405	0	−0.405	0	240
+0.155	+1	+0.155	+1.31	250

siendo P_c = potencia de la portadora (watts)
E_c = voltaje máximo de la portadora (volts)
R = resistencia de la carga (ohms)

Las potencias en las bandas laterales superior e inferior se determinan con la siguiente ecuación

$$P_{bls} = P_{bli} = \frac{(mE_c/2)^2}{2R}$$

en donde $mE_c/2$ es el voltaje máximo de las frecuencias laterales superior e inferior. Al reordenar esta ecuación se obtiene

$$P_{bls} = P_{bli} = \frac{m^2E_c^2}{8R} \tag{3-15a}$$

en donde P_{bls} = potencia de la banda lateral superior (watts)
P_{bli} = potencia de la banda lateral inferior (watts)

La ecuación 3-15a se reordena como sigue

$$P_{bls} = P_{bli} = \frac{m^2}{4}\left(\frac{E_c^2}{2R}\right) \tag{3-15b}$$

Al sustituir la ecuación 3-14 en la 3-15b, se obtiene

$$P_{bls} = P_{bli} = \frac{m^2P_c}{4} \tag{3-16}$$

Es evidente, en la ecuación 3-16, que para un coeficiente de modulación $m = 0$, la potencia en las bandas laterales superior e inferior es cero, y que la potencia total transmitida es tan sólo la potencia de la portadora.

La potencia total en una onda de amplitud modulada es igual a la suma de las potencias de la portadora y las de la banda lateral superior e inferior. La potencia total en una envolvente DSBFC de AM es

$$P_t = P_c + P_{bls} + P_{bli} \tag{3-17}$$

en donde P_t = potencia total de una envolvente DSBFC de AM (watts)
P_c = potencia de la portadora (watts)
P_{bls} = potencia de la banda lateral superior (watts)
P_{bli} = potencia de la banda lateral inferior (watts)

Al sustituir la ecuación 3-16 en la 3-17 se obtiene

$$P_t = P_c + \frac{m^2P_c}{4} + \frac{m^2P_c}{4} \tag{3-18}$$

es decir,

$$P_t = P_c + \frac{m^2P_c}{2} \tag{3-19}$$

en donde $(m^2P_c)/2$ es la potencia total de las bandas laterales.

Se saca P_c como factor común

$$P_t = P_c\left(1 + \frac{m^2}{2}\right) \tag{3-20}$$

De acuerdo con el análisis anterior, se puede ver que la potencia de portadora en la onda modulada es igual que en la onda no modulada. Por tanto, es evidente que la potencia de la portadora no se afecta en el proceso de modulación. También, como la potencia total en la onda

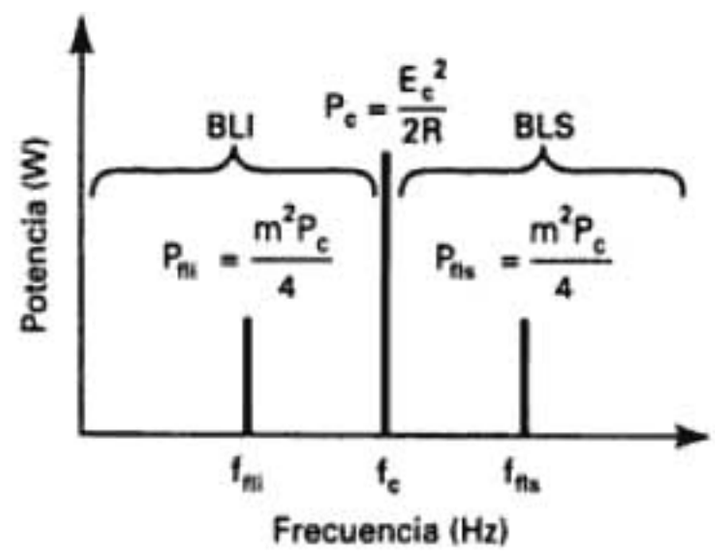


FIGURA 3-12 Espectro de potencias para una onda de AM de DSBFC con una señal moduladora de frecuencia única

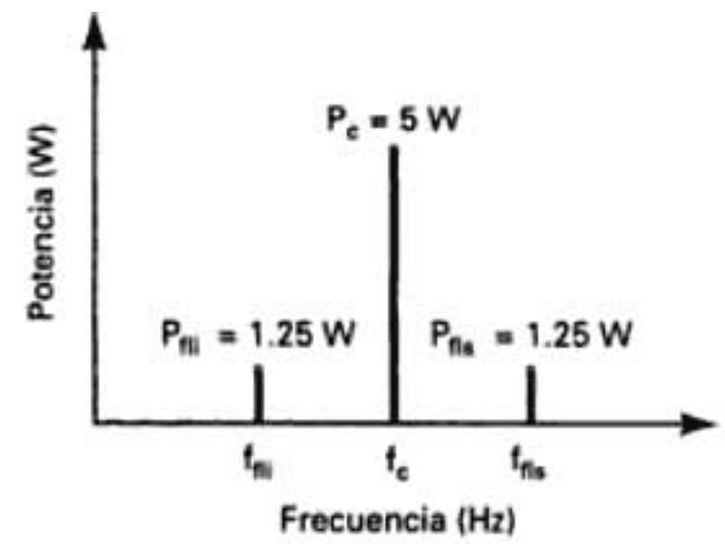


FIGURA 3-13 Espectro de potencias para el ejemplo 3-4d

de AM es igual a la suma de las potencias de la portadora y de las bandas laterales, la potencia total en una envolvente de AM aumenta con la modulación, es decir, cuando aumenta m aumenta P_t.

La fig. 3-12 muestra el espectro de potencias de una onda DSBFC de AM. Nótese que con 100% de modulación, la potencia máxima en la banda lateral superior o inferior, es igual a la cuarta parte de la potencia en la portadora. Así, la potencia máxima total de las bandas laterales es igual a la mitad de la potencia de la portadora. Una de las desventajas más importantes de la transmisión DSBFC de AM es que la información está contenida en las bandas laterales, aunque la mayor parte de la potencia se desperdicia en la portadora. En realidad, no se desperdicia totalmente, porque permite usar circuitos demoduladores relativamente sencillos y poco costosos en el receptor, lo cual es la ventaja principal de la DSBFC de AM.

Ejemplo 3-4

Para una onda DSBFC de AM con voltaje máximo de portadora no modulada $V_c = 10\ V_p$, una resistencia de carga $R_L = 10\ \Omega$ y un coeficiente de modulación $m = 1$, determinar:

(a) Las potencias de la portadora y de las bandas laterales superior e inferior.

(b) La potencia total de las bandas laterales.

(c) La potencia total de la onda modulada.

A continuación

(d) Trazar el espectro de potencias.

(e) Repetir los pasos a) a d) con un índice de modulación $m = 0.5$.

Solución (a) La potencia de la portadora se calcula sustituyendo los valores en la ecuación 3-14

$$P_c = \frac{10^2}{2(10)} = \frac{100}{20} = 5\ \text{W}$$

La potencia de las bandas laterales superior e inferior se calcula con la ecuación 3-16

$$P_{bls} = P_{bli} = \frac{(1^2)(5)}{4} = 1.25\ \text{W}$$

(b) La potencia total de las bandas laterales es

$$P_{tbl} = \frac{m^2P_c}{2} = \frac{(1^2)(5)}{2} = 2.5\ \text{W}$$

(c) La potencia total en la onda modulada se calcula sustituyendo en la ecuación 3-20

$$P_t = 5\left[1 + \frac{(1)^2}{2}\right] = 7.5\ \text{W}$$

(d) El espectro de potencias se ve en la fig. 3-13.

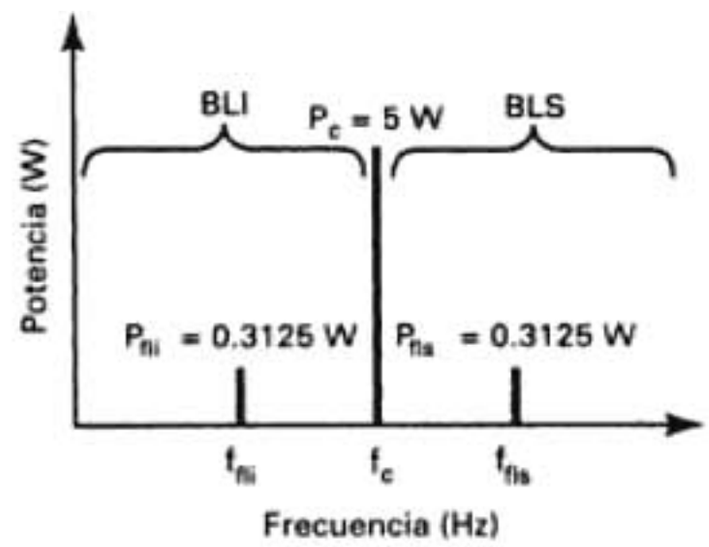


FIGURA 3-14 Espectro de potencias para el ejemplo 3-4

(e) La potencia de la portadora se calcula con la ecuación 3-14

$$P_c = \frac{10^2}{2(10)} = \frac{100}{20} = 5 \text{ W}$$

La potencia de las bandas laterales superior e inferior se calcula con la ecuación 3-16

$$P_{bls} = P_{bli} = \frac{(0.5)^2(5)}{4} = 0.3125 \text{ W}$$

La potencia total de las bandas laterales es

$$P_{tbl} = \frac{m^2 P_c}{2} = \frac{(0.5)^2(5)}{2} = 0.625 \text{ W}$$

La potencia total de la onda modulada se calcula sustituyendo en la ecuación 3-20

$$P_t = 5\left[1 + \frac{(0.5)^2}{2}\right] = 5.625 \text{ W}$$

La fig. 3-14 muestra el espectro de potencias.

En el ejemplo 3-4 se puede ver por qué es importante usar el porcentaje de modulación máximo posible, asegurando a la vez que no haya sobremodulación. Como se ve en el ejemplo, la potencia de portadora queda igual cuando cambia m. Sin embargo, la potencia de las bandas laterales bajó en forma drástica cuando m bajó de 1 a 0.5. Como la potencia de las bandas laterales es proporcional al cuadrado del coeficiente de modulación, una reducción de m a la mitad causa una reducción de la potencia de las bandas laterales a la cuarta parte, es decir, $0.5^2 = 0.25$. La relación entre el coeficiente de modulación a veces puede ser decepcionante, porque la potencia total de transmisión consiste principalmente en la potencia de la portadora y, en consecuencia, no se afecta mucho con los cambios en m. Sin embargo, se debe observar que la potencia en la parte conductora de datos de la señal transmitida (que son las bandas laterales) se afecta en forma drástica con los cambios de m. Por esta razón, en los sistemas DSBFC de AM se trata de mantener un coeficiente de modulación de 0.9 a 0.95 (modulación de 90 a 95%), para tener la máxima amplitud en las señales de información.

Cálculos de corriente en AM

Con la modulación de amplitud se hace necesario con mucha frecuencia, y a veces es deseable, medir la corriente de la onda portadora y la modulada, para después calcular el índice de modulación con estas medidas. Las mediciones se hacen simplemente determinando la corriente en la antena transmisora, con o sin la presencia de la señal moduladora. La relación entre la corriente de la portadora y la de la onda modulada es

$$\frac{P_t}{P_c} = \frac{I_t^2 R}{I_c^2 R} = \frac{I_t^2}{I_c^2} = 1 + \frac{m^2}{2}$$

en donde P_t = potencia total de transmisión (watts)
P_c = potencia de la portadora (watts)
I_t = corriente total de transmisión (amperes)
I_c = corriente de la portadora (amperes)
R = resistencia de la antena (ohms)

y $$\frac{I_t}{I_c} = \sqrt{1 + \frac{m^2}{2}} \tag{3-21a}$$

Por lo que $$I_t = I_c\sqrt{1 + \frac{m^2}{2}} \tag{3-21b}$$

Modulación con una señal compleja de información

En las secciones anteriores de este capítulo se analizaron el espectro de frecuencias, el ancho de banda, el coeficiente de modulación y la distribución de voltaje y potencia para una AM de portadora de máxima potencia y doble banda lateral, con una señal moduladora de una sola frecuencia. Sin embargo, en la práctica, la señal moduladora es una forma de onda compleja, formada por muchas ondas senoidales con distintas amplitudes y frecuencias. En consecuencia, se presentará un breve análisis de los efectos que tiene esa señal moduladora compleja sobre la forma de onda de AM.

Si una señal moduladora contiene dos frecuencias (f_{m1} y f_{m2}), la onda modulada contendrá la portadora y dos conjuntos de frecuencias laterales ubicadas en forma simétrica en torno a la portadora. Se puede representar esa onda por

$$v_{am}(t) = \text{sen}(2\pi f_c t) + \frac{1}{2}\cos[2\pi(f_c - f_{m1})t] - \frac{1}{2}\cos[2\pi(f_c + f_{m1})t]$$

$$+ \frac{1}{2}\cos[2\pi(f_c - f_{m2})t] - \frac{1}{2}\cos[2\pi(f_c + f_{m2})t]$$

Cuando varias frecuencias modulan en forma simultánea la amplitud de una portadora, el coeficiente combinado de modulación es igual a la raíz cuadrada de la suma de cuadrados de los índices individuales de modulación

$$m_t = \sqrt{m_1^2 + m_2^2 + m_3^2 + m_n^2} \tag{3-22}$$

en la que m_t = coeficiente total de modulación
m_1, m_2, m_3, y m_n = coeficientes de modulación para las señales de entrada 1, 2, 3 y n

Con el efecto combinado de la modulación se pueden determinar las potencias totales de las bandas laterales y de transmisión, como sigue

$$P_{\text{tbls}} = P_{\text{tbli}} = \frac{P_c m_t^2}{2} \tag{3-23}$$

y $$P_{\text{tbl}} \frac{P_c m_t^2}{2} \tag{3-24}$$

Por lo que $$P_t = P_c\left(1 + \frac{m_t^2}{2}\right) \tag{3-25}$$

en la que P_{tbls} = potencia total de la banda lateral superior (watts)
P_{tbli} = potencia total de la banda lateral inferior (watts)
P_{tbl} = potencia total de las bandas laterales (watts)
P_t = potencia total transmitida (watts)

En un transmisor de AM se debe tener cuidado en asegurar que los voltajes combinados de todas las señales moduladoras no sobremodulen a la portadora.

Ejemplo 3-5

Un transmisor DSBFC de AM tiene potencia de portadora no modulada $P_c = 100$ W, que se modula en forma simultánea por tres señales cuyos coeficientes de modulación son $m_1 = 0.2$, $m_2 = 0.4$, y $m_3 = 0.5$, determinar:

(a) El coeficiente total de modulación.
(b) La potencia de las bandas laterales superior e inferior.
(c) La potencia total transmitida.

Solución (a) El coeficiente total de modulación se calcula sustituyendo en la ecuación 3-22.

$$m_t = \sqrt{0.2^2 + 0.4^2 + 0.5^2}$$
$$= \sqrt{0.04 + 0.16 + 0.25} = 0.67$$

(b) La potencia total de las bandas laterales se calcula sustituyendo el resultado del paso a) en la ecuación 3-24.

$$P_{\text{tbl}} = \frac{(0.67^2)100}{2} = 22.445 \text{ W}$$

(c) La potencia total transmitida se calcula sustituyendo en la ecuación 3-25.

$$P_t = 100\left(1 + \frac{0.67^2}{2}\right) = 122.445 \text{ W}$$

CIRCUITOS MODULADORES DE AM

En un transmisor, el lugar donde se hace la modulación determina si el circuito es un *transmisor de bajo* o *de alto nivel*. Con modulación de bajo nivel, ésta se hace antes del elemento de salida de la etapa final del transmisor; en otras palabras, antes del colector del transistor de salida en un transmisor transistorizado, antes del drenaje del FET de salida en un transmisor de FET, o antes de la placa del tubo de salida en un transmisor con tubos al vacío.

Una ventaja de la modulación de bajo nivel es que se requiere menos potencia de señal moduladora para lograr modulación de alto porcentaje. En los moduladores de alto nivel, la modulación se hace en el elemento final de la etapa final, donde la señal portadora tiene su amplitud máxima y por ello requiere una señal moduladora de mucho mayor amplitud para lograr un porcentaje razonable de modulación. Con la modulación de alto nivel, el amplificador final de señal moduladora debe suministrar toda la potencia de banda lateral, que podría ser hasta 33% de la potencia total de transmisión. Una desventaja obvia de la modulación de bajo nivel se presenta en las aplicaciones de gran potencia, donde todos los amplificadores que siguen a la etapa moduladora deben ser amplificadores lineales, lo cual es extremadamente ineficiente.

Modulador de AM de bajo nivel

Un amplificador de señal pequeña, de clase A, como el de la fig. 3-15a, se puede usar para la modulación de amplitud; sin embargo, el amplificador debe tener dos entradas: una para la señal portadora y otra para la señal moduladora. Cuando no hay señal moduladora presente, el circuito funciona como un amplificador lineal de clase A, y la salida no es más que la portadora amplificada por la ganancia de voltaje en reposo. Sin embargo, cuando se aplica una señal moduladora, el amplificador funciona en forma no lineal y se produce la multiplicación de señal que se describe con la ecuación 3-9a. En la fig. 3-15a, la portadora se aplica a la base, y la señal moduladora al emisor. En consecuencia, a esta configuración de circuito se le llama *modulación por emisor*. La señal moduladora varía la ganancia del amplificador con tasa senoidal igual a su frecuencia. La cantidad de modulación que se obtiene es proporcional a la amplitud de la señal moduladora. La ganancia de voltaje para un modulador por emisor se describe con la ecuación

$$A_v = A_q[1 + m \operatorname{sen}(2\pi f_m t)] \tag{3-26}$$

en la que A_v = ganancia de voltaje del amplificador con modulación (adimensional)
A_q = ganancia de voltaje del amplificador en reposo (sin modulación, adimensional)

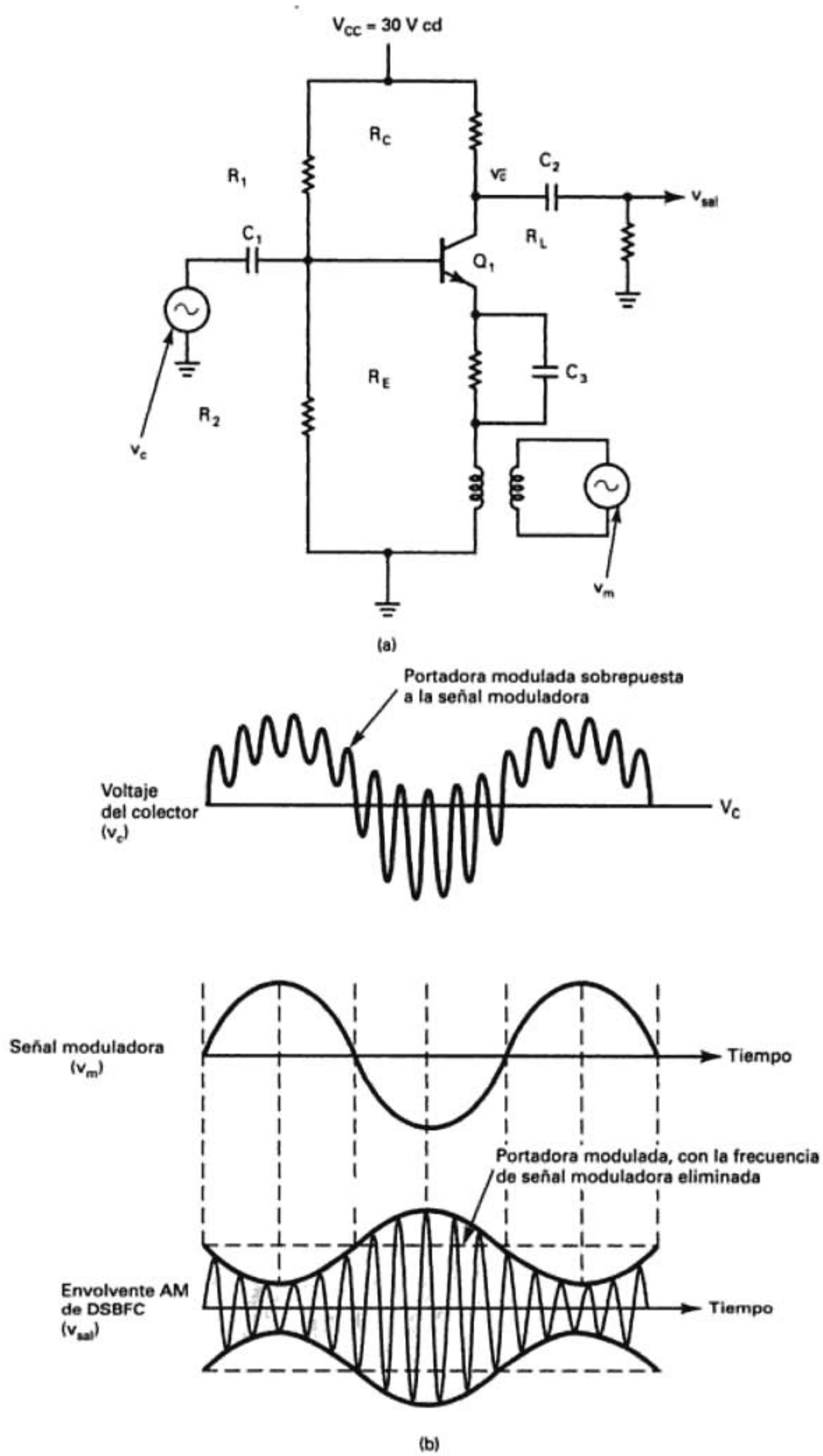


FIGURA 3-15 (a) Transistor único, modulador por emisor; (b) Formas de onda de salida

El sen($2\pi f_m t$) va desde un valor máximo de $+1$ hasta uno mínimo de -1. Así, la ecuación (3-26) se reduce a

$$A_v = A_q(1 \pm m) \tag{3-27}$$

donde m es igual al coeficiente de modulación. En 100% de modulación, $m = 1$ y la ecuación 3-27 se reduce a

$$A_{v(\text{máx})} = 2A_q$$
$$A_{v(\text{mín})} = 0$$

La fig. 3-15b muestra las formas de onda para el circuito de la fig. 3-15a. La señal moduladora se aplica a través de un transformador T_1 de aislamiento, al emisor de Q_1, y la portadora se aplica a la base en forma directa. La señal moduladora lleva al circuito tanto a saturación como a corte y produce así la amplificación no lineal necesaria para que se produzca la modulación. La forma de onda del colector incluye a la portadora y a las frecuencias laterales superior e inferior, así como un componente a la frecuencia de la señal moduladora. El capacitor de acoplamiento C_2 quita la frecuencia de la señal moduladora de la forma de onda de AM y produce así una envolvente simétrica de AM en V_{sal}.

Con modulación por emisor, la amplitud de la señal de salida depende de la amplitud de la portadora de entrada y de la ganancia de voltaje del amplificador. El coeficiente de modulación depende en su totalidad de la amplitud de la señal moduladora. La desventaja principal de la modulación por emisor es que el amplificador funciona como clase A, que es ineficiente en extremo. Los moduladores por emisor tampoco son capaces de producir formas de onda de salida de gran potencia.

Ejemplo 3-6

Para un modulador de AM de bajo nivel, parecido al de la fig. 3-15, con un coeficiente de modulación $m = 0.8$, una ganancia de voltaje en reposo $A_q = 100$, una frecuencia de portadora de entrada $f_c = 500$ kHz con amplitud $V_c = 5$ mV y una señal moduladora de 1000 Hz, determinar lo siguiente:

(a) Ganancias máxima y mínima de voltaje.
(b) Amplitudes máxima y mínima de V_{sal}.
Después:
(c) Trazar la envolvente de la AM de salida.

Solución (a) Se sustituye en la ecuación 3-26

$$A_{\text{máx}} = 100(1 + 0.8) = 180$$
$$A_{\text{mín}} = 100(1 - 0.8) = 20$$

(b)

$$V_{\text{sal(máx)}} = 180(0.005) = 0.9 \text{ V}$$
$$V_{\text{sal(mín)}} = 20(0.005) = 0.1 \text{ V}$$

(c) La envolvente de AM se ve en la fig. 3-16

Modulador de AM de potencia intermedia

Los primeros transmisores de AM de potencia intermedia y alta se limitaban a usar tubos al vacío como dispositivos activos. Sin embargo, a partir de mediados de la década de 1970 se dispone

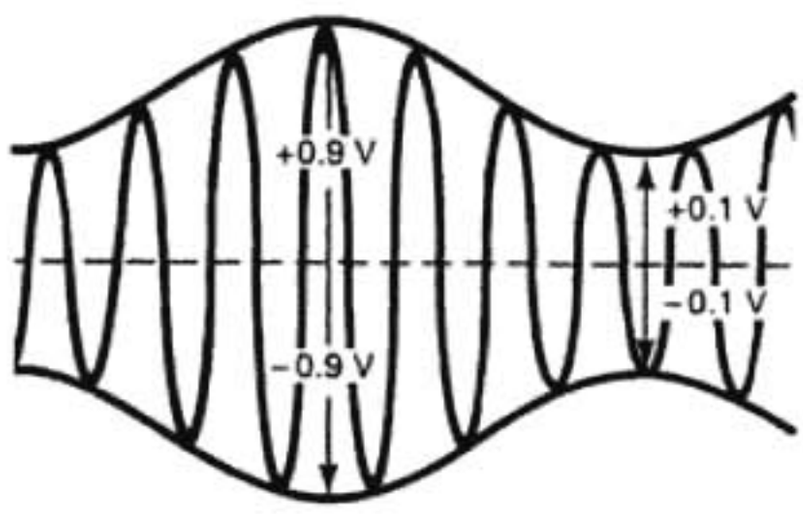


FIGURA 3-16 Envolvente de AM para el ejemplo 3-6

de transmisores de estado sólido, con potencias de salida tan altas como algunos miles de watts. Esto se logra poniendo al final en paralelo, varios amplificadores de potencia, de tal manera que se combinen sus señales de salida en fase y sean, por consiguiente, aditivas.

La fig. 3-17a muestra el diagrama de un modulador de AM de potencia intermedia con un solo transistor. La modulación se hace en el colector, que es el elemento de salida del transistor. Por consiguiente, si es la etapa activa final del transmisor (esto es, no hay amplificadores entre él y la antena), el modulador es de alto nivel.

Para alcanzar alta eficiencia de potencia, los moduladores de potencia intermedia y alta de AM trabajan en general en clase C. Por consiguiente, es posible una eficiencia práctica hasta de 80%. El circuito de la fig. 3-17a es un amplificador de clase C con dos entradas: una portadora

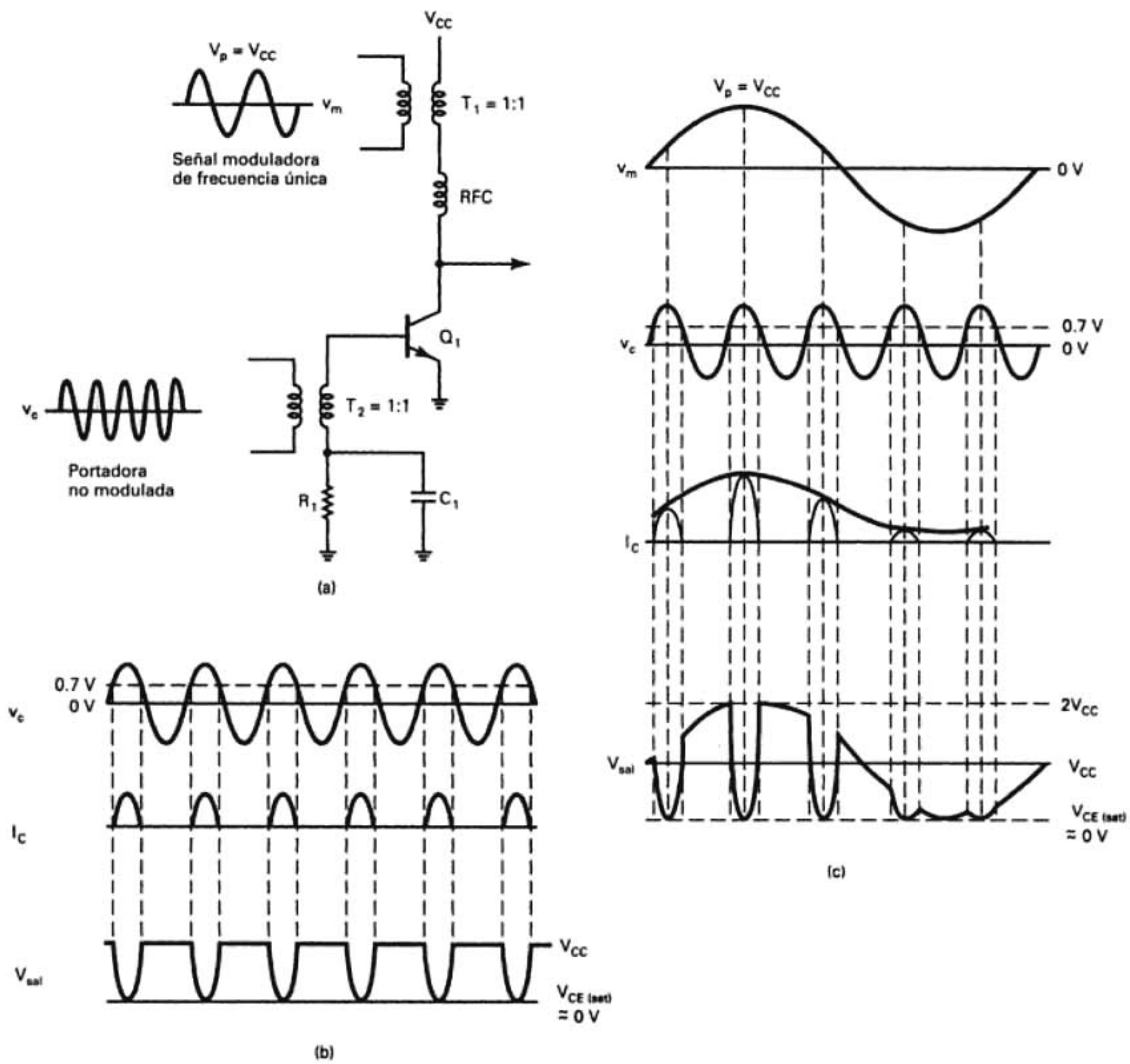


FIGURA 3-17 Modulador AM de DSBFC simplificado, con transistor de potencia intermedia: (a) diagrama; (b) formas de onda del colector sin señal moduladora; (c) formas de onda del colector con una señal moduladora

(v_c) y una señal moduladora de frecuencia única (v_m). Como el transistor está polarizado en clase C, es no lineal y es capaz de mezclado (modulación) no lineal. Este circuito se llama *modulador por colector*, porque la señal moduladora se aplica en forma directa al colector. El RFC, de *radio-frequency choke*, es una bobina de radiofrecuencia que funciona como corto circuito a la cd y circuito abierto en altas frecuencias. Por consiguiente, el RFC aísla la fuente de poder de cd de la portadora de alta frecuencia y de las frecuencias laterales, y al mismo tiempo permite que las señales de datos, de baja frecuencia, modulen al colector de Q_1.

Funcionamiento del circuito. Durante la siguiente explicación, consúltese el circuito de la fig. 3-17a y las formas de onda en la fig. 3-17b. Cuando la amplitud de la portadora es mayor que la barrera de potencial de la unión de base-emisor (unos 0.7 V para un transistor de silicio), Q_1 se activa y pasa la corriente del colector. Cuando la amplitud de la portadora baja de 0.7 V, Q_1 se apaga y cesa la corriente del colector. En consecuencia, Q_1 cambia entre saturación y corte controlado por la señal portadora, la corriente del colector pasa durante menos de 180° de cada ciclo de portadora y se alcanza la operación de clase C. Cada ciclo sucesivo de la portadora enciende a Q_1 durante un instante y deja pasar la corriente un tiempo corto, produciendo una forma de onda que pasa a negativa en el colector. Las formas de onda de corriente y voltaje del colector se ven en la fig. 3-17b. La forma de onda de voltaje del colector se asemeja a una señal rectificada de media onda, con frecuencia fundamental igual a f_c.

Cuando se aplica la señal moduladora al colector en serie con el voltaje de suministro de cd, se suma y se resta de V_{CC}. Se producen las formas de onda de la fig. 3-17c cuando la amplitud máxima del máximo de señal moduladora es igual a V_{CC}. Se puede ver que la forma de onda de voltaje de salida pasa de un valor máximo de 2 V_{CC} a casi 0 V [$V_{CE(sat)}$]. El cambio máximo

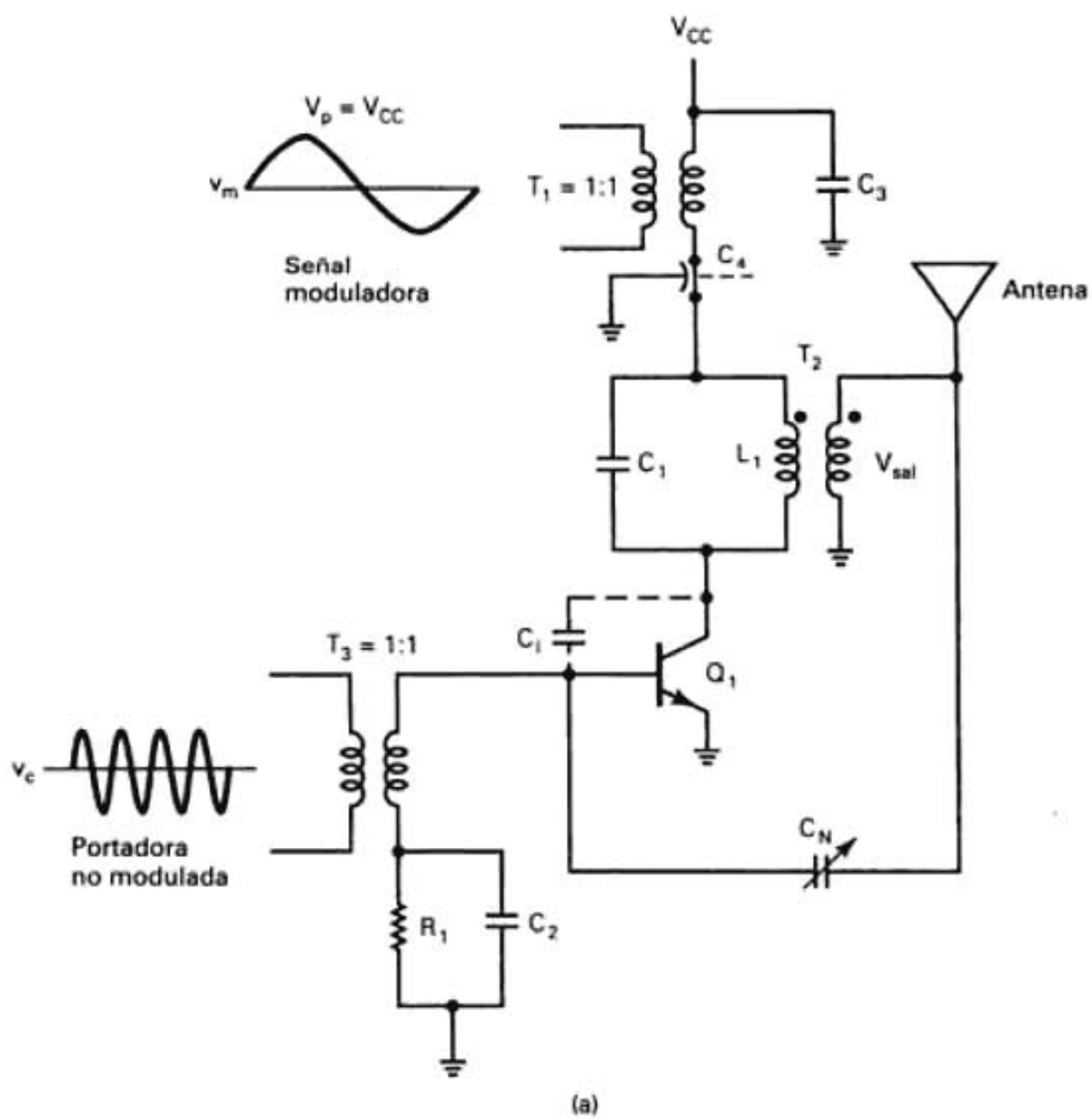


FIGURA 3-18 Modulador AM de DSBFC con transistor de potencia intermedia: (a) diagrama; *(continúa)*

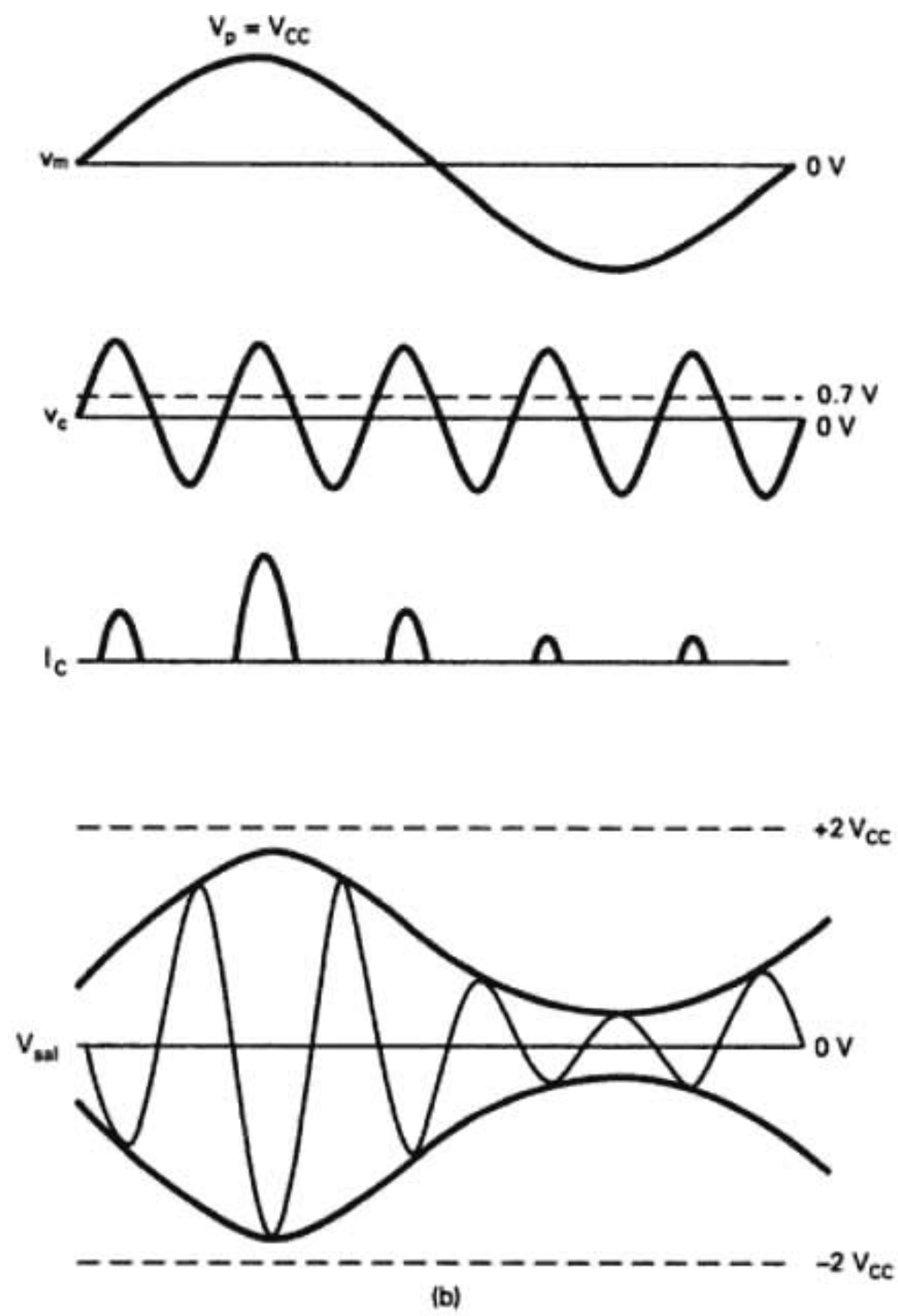


FIGURA 3-18 *(Continuación)* (b) Formas de onda del colector y de salida

de voltaje de colector es igual a V_{CC}. De nuevo, la forma de onda se parece a una portadora rectificada de media onda, sobrepuesta a una señal de datos de ca de baja frecuencia.

Como Q_1 trabaja en forma no lineal, la forma de onda del colector contiene las dos frecuencias originales de entrada, f_c y f_m, y sus frecuencias de suma y diferencia ($f_c \pm f_m$). Ya que la forma de onda de salida también contiene las armónicas de orden superior y los componentes de intermodulación, antes de transmitirla se debe limitar su ancho de banda a $f_c \pm f_m$.

Un circuito más práctico para producir una señal DSBFC de AM y de potencia intermedia se ve en la fig. 3-18a, con las formas de onda correspondientes en la fig. 3-18b. Este circuito también es un modulador por colector con una amplitud máxima "pico" de $V_{m(\text{máx})} = V_{CC}$. El funcionamiento de este circuito es casi idéntico al del circuito de la fig. 3-17a, excepto por la adición de un circuito tanque (C_1 y L_1) en el colector de Q_1. Como el transistor funciona entre saturación y corte, la corriente del colector no depende del voltaje de excitación de la base. El voltaje desarrollado en el circuito tanque se determina con la componente alterna de la corriente del colector, y por la impedancia de resonancia del circuito tanque. Esta impedancia depende a su vez del factor de calidad (Q) de la bobina. Las formas de onda de la señal moduladora, la portadora y la corriente del colector son idénticas a las del ejemplo anterior. El voltaje de salida es una señal simétrica DSBFC de AM con un voltaje promedio de 0 V, amplitud pico máxima positiva igual a $2V_{CC}$ y amplitud pico máxima negativa de $-2V_{CC}$. El medio ciclo positivo de la onda lo produce el circuito tanque mediante el *efecto inercial (flywheel effect)*. Cuando Q_1 conduce, C_1 se carga a $V_{CC} + V_m$ (un valor máximo de $2V_{CC}$), y cuando Q_1

está apagado, C_1 se descarga a través de L_1. Cuando L_1 descarga, C_1 se carga a un valor mínimo de $-2V_{CC}$. Eso produce el medio ciclo positivo de la envolvente de AM. La frecuencia de resonancia del circuito tanque es igual a la frecuencia de portadora, y el ancho de banda va desde $f_c - f_m$ hasta $f_c + f_m$. En consecuencia, la señal moduladora, las armónicas y todos los productos cruzados de orden superior se eliminan de la forma de onda y queda una onda DSBFC simétrica de AM. Se obtiene modulación de 100% cuando la amplitud máxima de la señal moduladora es igual a V_{CC}.

En la fig. 3-18a hay varios componentes que no se han explicado. R_1 es el resistor de polarización de Q_1. Este resistor y C_2 forman un circuito fijador de nivel que produce una "auto" polarización inversa, y junto con la barrera de potencial del transistor determina el voltaje de activación de Q_1. En consecuencia, se puede polarizar a Q_1 para que sólo se encienda durante los picos más positivos del voltaje de portadora. Esto produce una forma angosta de onda de corriente de colector, y amplía la eficiencia de clase C.

El capacitor C_3 es de derivación, que parece cerrado a las frecuencias de la señal moduladora, evitando que las señales de información entren a la fuente de poder de cd. El capacitor C_{bc} *es la capacitancia de unión de base a colector de* Q_1. A las radiofrecuencias, las capacitancias de unión relativamente pequeñas, dentro del transistor, son insignificantes. Si es apreciable la reactancia capacitiva de C_{bc}, se puede retrasar la señal del colector hasta la base, con amplitud suficiente para hacer que Q_1 comience a oscilar. En consecuencia, se debe retroalimentar a la base una señal de igual amplitud y frecuencia, y desfasada 180° para anular o *neutralizar la retroalimentación debido a la capacitancia entre electrodos*. El capacitor C_N es un *capacitor neutralizador*. Su objetivo es proporcionar un camino a la retroalimentación de una señal que tenga amplitud igual, pero frecuencia desfasada 180° respecto a la señal que se retroalimenta a través de C_{bc}. El condensador C_4 es un capacitor de derivación de RF. Su objeto es aislar la fuente de poder de cd de las radiofrecuencias. Su funcionamiento es bastante sencillo; en la frecuencia de la portadora, C_4 parece conducir y evita que la portadora *se fugue* hacia la fuente de poder o a los circuitos de señal moduladora, y que se distribuya por el transmisor.

Modulación simultánea de base y colector

Los moduladores por colector producen una envolvente más simétrica que los moduladores por emisor de bajo nivel, y los moduladores por colector tienen más eficiencia energética. Sin embargo, los moduladores por colector requieren una señal moduladora de mayor amplitud, y no pueden llegar a una oscilación de voltaje de salida desde la saturación total hasta el corte, con lo que evitan la obtención de modulación de 100%. Por lo anterior, para lograr la modulación simétrica, trabajar a eficiencia máxima, obtener gran potencia de salida y necesitar un mínimo de potencia en la señal moduladora, a veces se usan modulaciones simultáneas por emisor y por colector.

Funcionamiento del circuito. La fig. 3-19 muestra un modulador de AM que usa una combinación de modulaciones por base y por colector. La señal moduladora se alimenta en forma simultánea a los colectores de los moduladores en contrafase (push-pull) (Q_2 y Q_3) y al colector del amplificador excitador (Q_1). En Q_1 se hace modulación por colector; en consecuencia, la señal portadora en la base de Q_2 y Q_3 ya se ha modulado parcialmente, y se puede reducir la potencia de la señal moduladora. También, no se necesita que los moduladores trabajen en toda su curva de operación para alcanzar una modulación de 100%.

Moduladores de AM en circuito integrado lineal

Los *generadores de función en circuito integrado lineal* usan un arreglo exclusivo de transistores y FET para hacer la multiplicación de señal, que es una característica que los hace ideales para generar formas de onda de AM. Los circuitos integrados, a diferencia de sus contrapartes discretas, pueden compensar con precisión el flujo de corriente, la ganancia de voltaje del amplificador y las variaciones de temperatura. También, estos moduladores de AM en circuito integrado lineal ofrecen una excelente estabilidad de frecuencia, características simétricas de modulación, miniaturización de circuitos, menos componentes, inmunidad a la temperatura y simplicidad de diseño y de localización de fallas. Entre sus desventajas están la baja potencia de

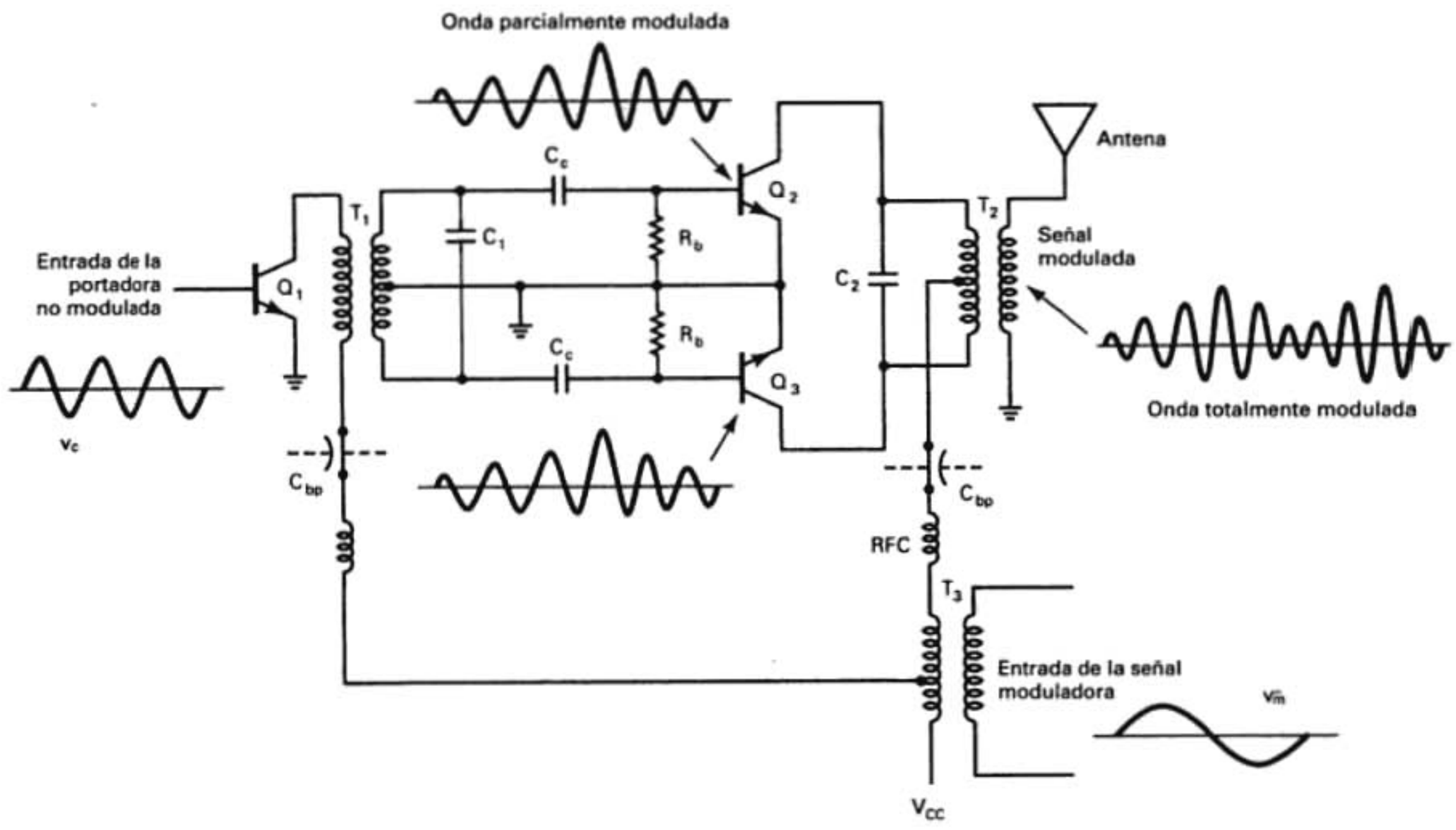


FIGURA 3-19 Modulador transistorizado AM de DSBFC de alta potencia

salida, un intervalo de frecuencia útil relativamente angosto, y la susceptibilidad a fluctuaciones en la cd suministrada.

El *generador monolítico de funciones* XR-2206 es ideal para hacer modulación de amplitud. La fig. 3-20a muestra su diagrama de bloques y la fig. 3.20b muestra su diagrama. El XR-2206 consiste en cuatro bloques funcionales: un oscilador controlado por voltaje (VCO),

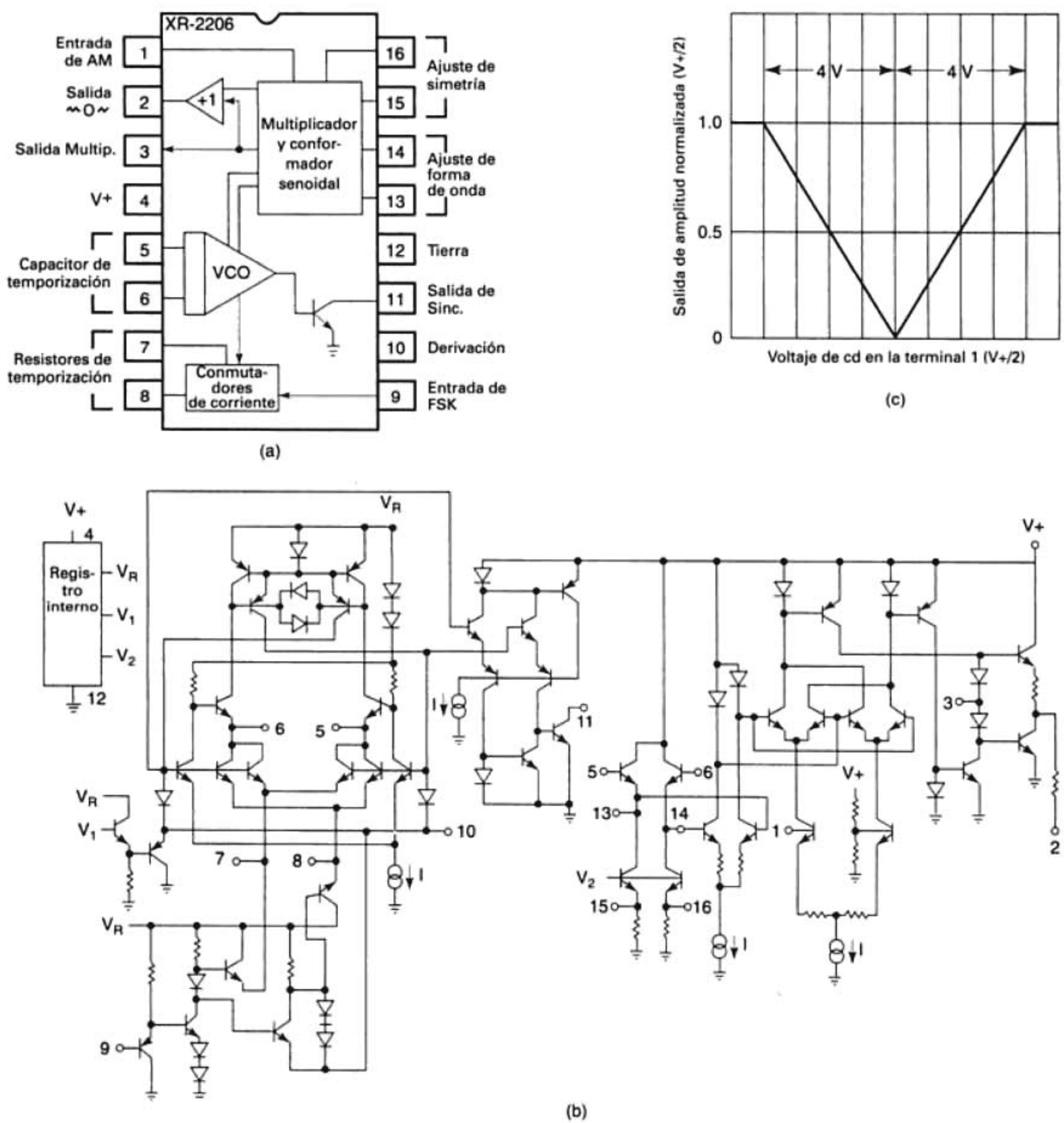


FIGURA 3-20 XR-2206: (a) diagrama de bloques; (b) diagrama; (c) curva de voltaje de salida en función del voltaje de entrada

un multiplicador analógico y un conformador senoidal, un separador de ganancia unitaria y un conjunto de conmutadores de corriente. La frecuencia de oscilación del VCO, f_c, se determina con el capacitor externo C_1 de temporización, entre las terminales 5 y 6, y con el resistor de temporización, R_1, conectado entre las terminales 7 u 8, y tierra. La selección de las terminales 7 u 8 depende del nivel de voltaje de la terminal 9. Si la terminal 9 está abierta o conectada a un voltaje externo ≥32 V, se selecciona la terminal 7. Si el voltaje de la terminal 9 es ≤1 V, se selecciona la terminal 8. La frecuencia de oscilación de determina con

$$f_c = \frac{1}{R_1 C_1} \text{ Hz} \tag{3-28}$$

La amplitud de la salida en la terminal 2 se puede modular aplicando una polarización de cd y una señal moduladora a la terminal 1. La fig. 3-20c muestra la amplitud normalizada de salida en función de la polarización de cd. Una salida normalizada de 1 corresponde al voltaje máximo de salida; un valor normalizado de 0.5 corresponde a un voltaje de salida igual a la mitad del valor máximo, y un valor normalizado de 0 corresponde a cero señal de salida. Como se ve en la figura, la amplitud de la salida varía en forma lineal con la polarización de entrada, para voltajes dentro de ±4 volts respecto a $V^+/2$. Un voltaje de entrada igual a $V^+/2$ hace que la amplitud de salida vaya a 0 V, y un voltaje de entrada 4 V mayor o 4 V menor que $V^+/2$ produce la amplitud máxima de salida.

La fig. 3-21 muestra el diagrama de un modulador de AM con circuito integrado lineal, que usa el XR-2206. La frecuencia de salida del VCO es la señal portadora. La señal moduladora y el voltaje de polarización se aplican al circuito multiplicador (modulador) interno por la terminal 1. La señal moduladora se mezcla con la señal del VCO y produce una onda de AM en V_{sal}. La onda de salida es una envolvente simétrica de AM que contiene a la portadora y a las frecuencias laterales superior e inferior.

Ejemplo 3-7

Se tiene un modulador XR-2206 de circuito lineal integrado como el de la fig. 3-21, con un voltaje de suministro $V^+ = 12$ V cd, amplitud de señal moduladora $V_m = 2$ V_p, frecuencia de señal moduladora $f_m = 4$ kHz, polarización de cd $V_{pol} = +4$ V cd, resistor de sincronización $R_1 = 10$ kΩ y capacitor de sincronización $C_1 = 0.001$ μF. Determinar:

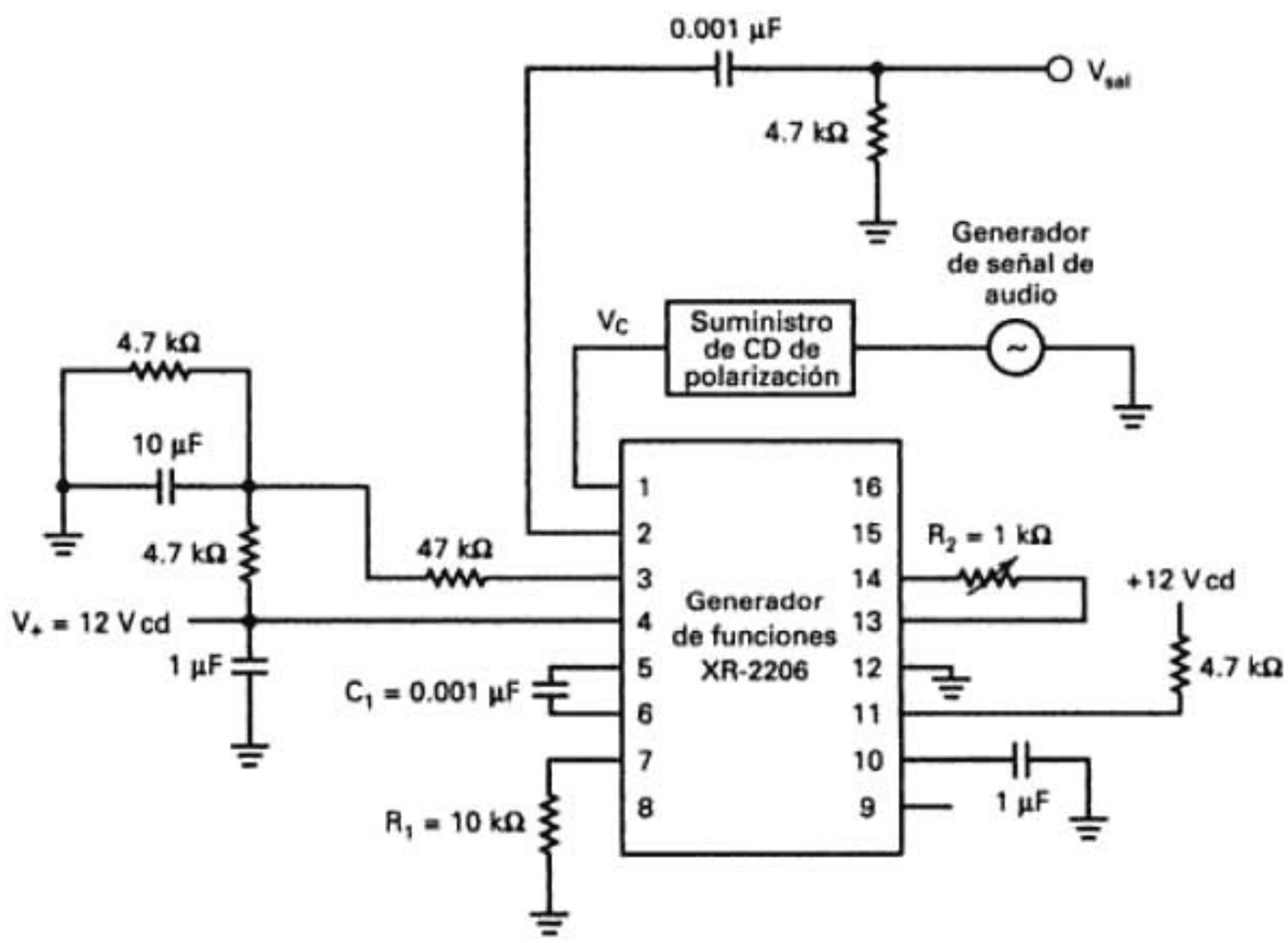


FIGURA 3-21 Modulador de AM con circuito lineal integrado

(a) La frecuencia de la portadora.
(b) Las frecuencias laterales superior e inferior.
Después
(c) Trazar la onda de salida.
(d) De acuerdo con la forma de onda de salida, determinar el coeficiente de modulación y el porcentaje de modulación.

Solución (a) La frecuencia de portadora se calcula con la ecuación 3-28

$$f_c = \frac{1}{(10\ \text{k}\Omega)(0.001\ \mu\text{F})} = 100\ \text{kHz}$$

(b) Las frecuencias laterales superior e inferior son tan sólo las frecuencias de suma y diferencia de las de la portadora y la señal moduladora

$$f_{\text{fls}} = 100\ \text{kHz} + 4\ \text{kHz} = 104\ \text{kHz}$$
$$f_{\text{fli}} = 100\ \text{kHz} - 4\ \text{kHz} = 96\ \text{kHz}$$

(c) En la fig. 3-22 se ve cómo se produce la envolvente de AM para las características de voltaje de salida en función de voltaje de entrada del XR-2206.
(d) El porcentaje de modulación se calcula con la ecuación 3-7, a partir de la envolvente de AM de la fig. 3-22

$$V_{\text{máx}} = 10\ \text{V}_\text{p} \qquad V_{\text{mín}} = 0\ \text{V}$$
$$m = \frac{10 - 0}{10 + 0} = 1$$
$$M = 1 \times 100 = 100\%$$

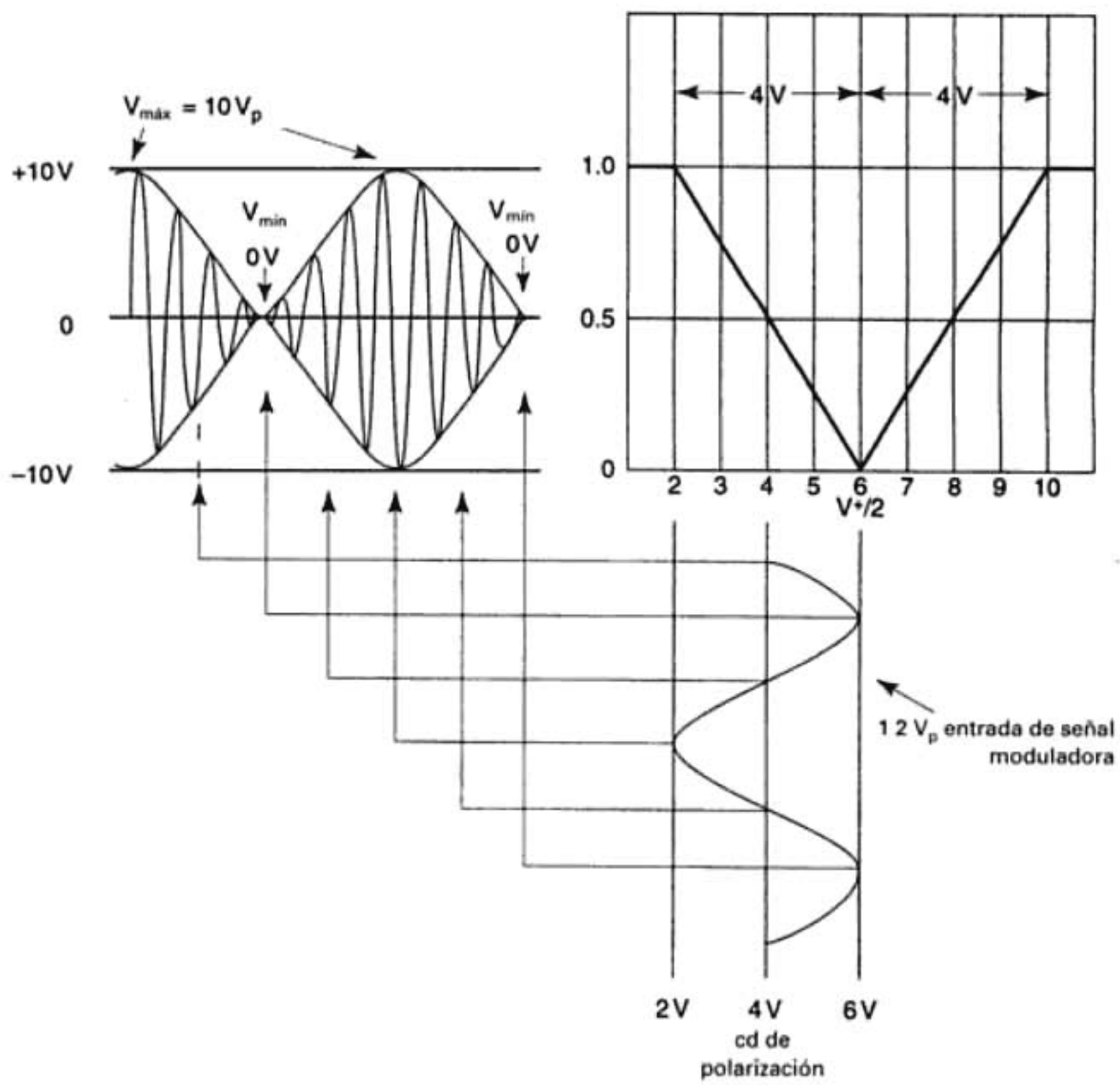


FIGURA 3-22 Envolvente de AM para el ejemplo 3-7

Transmisores de bajo nivel

La fig. 3-23 muestra un diagrama de bloques de un transmisor DSBFC de AM, de bajo nivel. Cuando se transmite voz o música, la fuente de señal moduladora es en general un transductor acústico, como un micrófono, una cinta magnética o un disco CD o de fonógrafo. El *preamplificador* suele ser un amplificador sensible y lineal de voltaje de clase A, con alta impedancia de entrada. La función del preamplificador es elevar la amplitud de la señal de la fuente hasta un valor útil, produciendo a la vez una distorsión no lineal mínima, y también agregar el menor ruido térmico que sea posible. El excitador de la señal moduladora también es un amplificador lineal que sólo amplifica la señal de información hasta un valor adecuado para la excitación suficiente del modulador. Se puede requerir más de un amplificador de excitación.

El *oscilador de portadora* de RF puede ser cualquiera de las configuraciones de oscilador que se describieron en el capítulo 2. La FCC establece estrictos requisitos para la exactitud y estabilidad del transmisor y, en consecuencia, los circuitos de uso más frecuente son los osciladores controlados por cristal. El *amplificador separador* es un amplificador lineal de baja ganancia y alta impedancia de entrada. Su función es aislar al oscilador de los amplificadores de alta potencia. Este separador proporciona una carga relativamente constante al oscilador, que ayuda a reducir la ocurrencia y magnitud de variaciones de frecuencia de corto plazo. Con frecuencia se usan seguidores de emisor o amplificadores operacionales de circuito integrado como separadores. El modulador puede tener modulación por emisor o por colector. Los amplificadores de potencia intermedia y final son moduladores en contrafase (push-pull) lineales de clase A o de clase B. Se requieren con los transmisores de bajo nivel, para mantener la simetría de la envolvente de AM. La red de acoplamiento con la antena compensa, o iguala, la impedancia de salida del amplificador final de potencia, con la de la línea de transmisión y la antena.

Los transmisores de bajo nivel como el de la fig. 3-23 se usan principalmente en sistemas de baja potencia y baja capacidad, como intercomunicaciones inalámbricas, unidades de control remoto, localizadores de personas y radioteléfonos de corto alcance.

Transmisores de alto nivel

La fig. 3-24 muestra el diagrama de bloques de un transmisor DSBFC de AM de alto nivel. La señal moduladora se procesa de la misma forma que en el transmisor de bajo nivel, excepto por la adición de un amplificador de potencia. En los transmisores de alto nivel, la potencia de la señal moduladora debe ser mucho más alta que la que se usa en los de bajo nivel. Esto se debe a que la portadora tiene toda la potencia en el punto del transmisor donde se hace la modulación y, en consecuencia, requiere una señal moduladora de gran amplitud para producir 100% de modulación.

El oscilador de la portadora de RF, su separador asociado y el excitador de portadora también son en esencia los mismos circuitos que los de los transmisores de bajo nivel. Sin embargo, en los de alto nivel la portadora de RF sufre una amplificación adicional de potencia, antes de la etapa moduladora, y el amplificador final de potencia también es el modulador. En consecuencia, el modulador suele ser un amplificador de clase C, modulado en el drenaje, la placa o el colector.

Con los transmisores de alto nivel, el circuito modulador tiene tres funciones primarias. Proporciona los circuitos necesarios para efectuar la modulación (es decir, la no linealidad), es el amplificador final de potencia (clase C, para tener eficiencia), y un convertidor elevador de frecuencia. Un convertidor elevador traslada la señal de datos de baja frecuencia a las señales de radiofrecuencia que se puedan irradiar con eficiencia de una antena, y después propagar por el espacio libre también con eficiencia.

Patrones trapezoidales

Se usan los *patrones trapezoidales* para observar las características de modulación de los transmisores de AM (coeficiente de modulación y simetría de modulación). Aunque se pueden examinar las características de modulación con un osciloscopio normal, se interpretan con más facilidad y exactitud en un patrón trapezoidal. La fig. 3-25 muestra un dispositivo básico de prueba

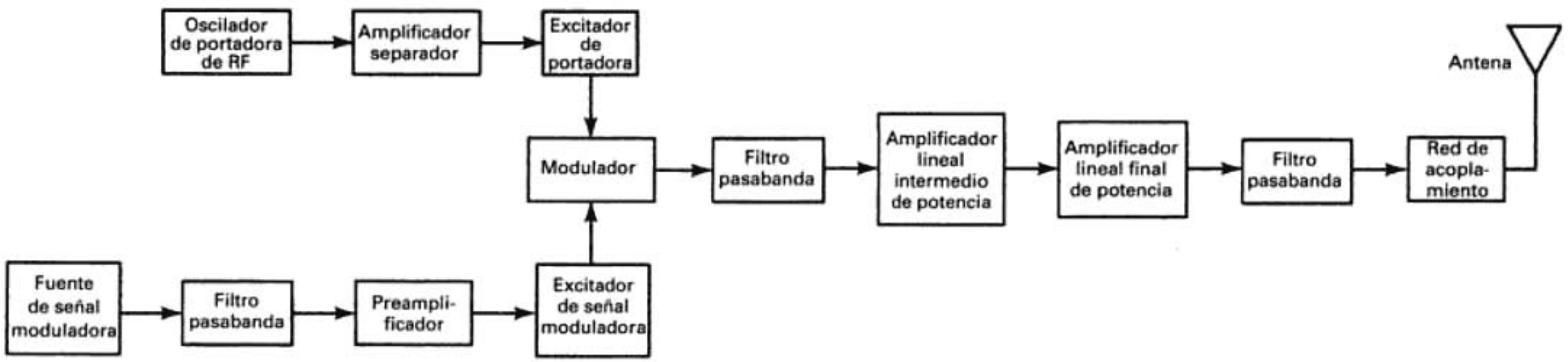


FIGURA 3-23 Diagrama de bloques de un transmisor AM de DSBFC

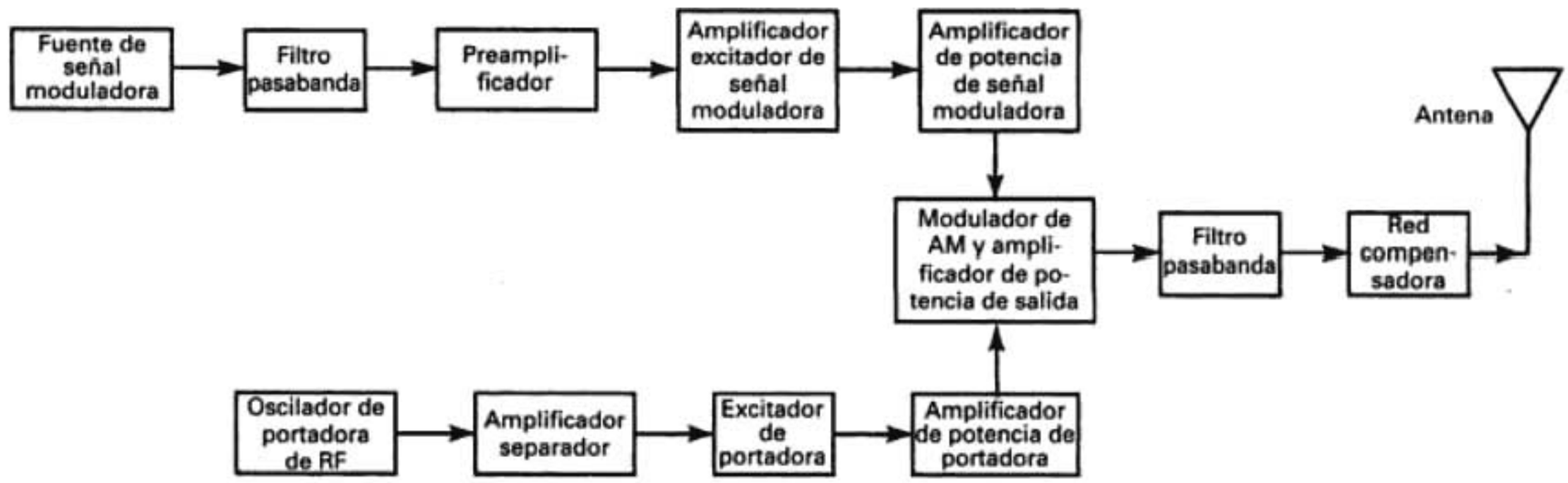


FIGURA 3-24 Diagrama de bloques de un transmisor DSBFC de AM de alto nivel

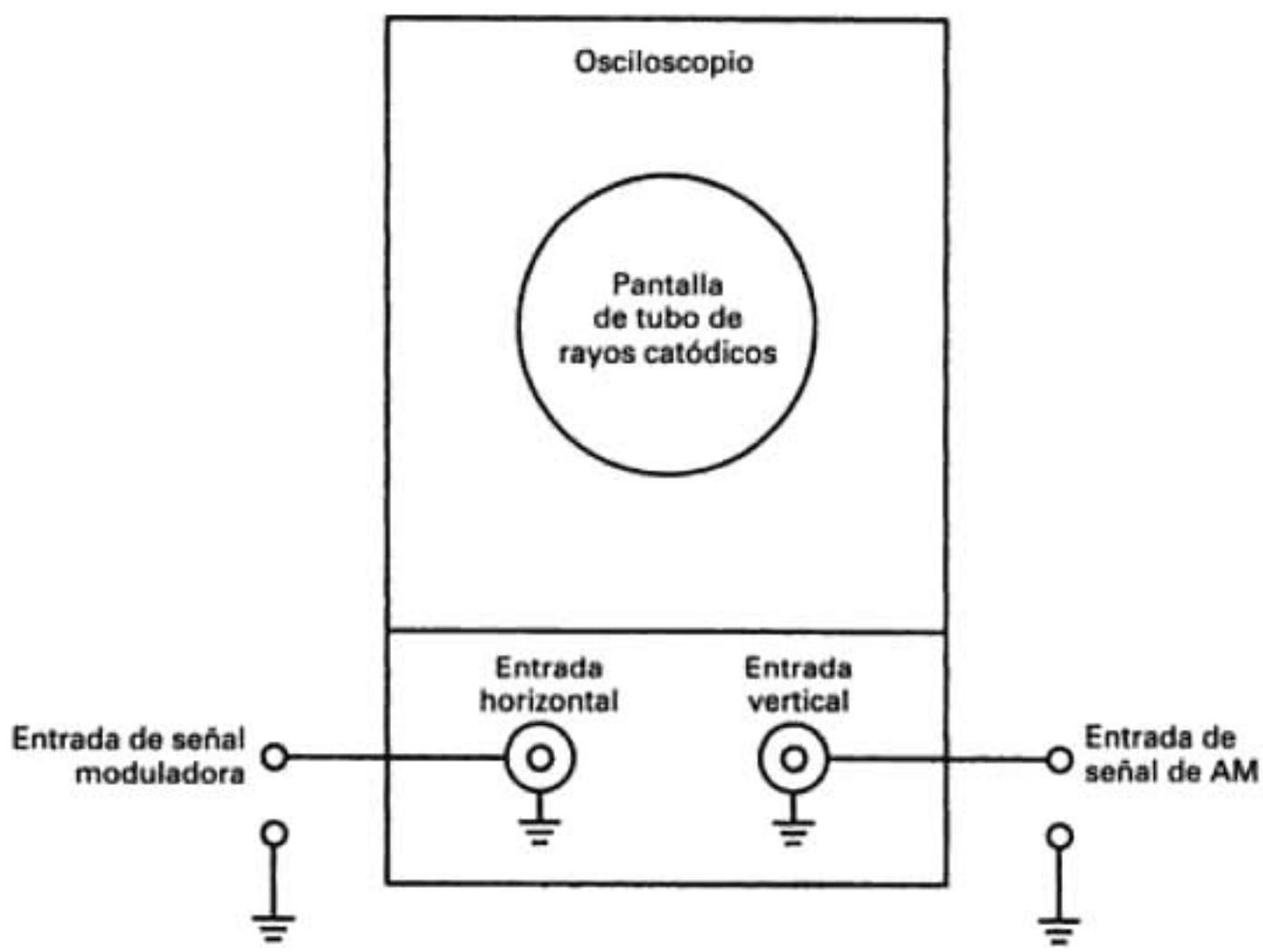


FIGURA 3-25 Aparato de pruebas para formar patrones trapezoidales en un osciloscopio

para producir un patrón trapezoidal en el tubo de rayos catódicos de un osciloscopio normal. La onda de AM se aplica a la entrada vertical del osciloscopio, y la señal moduladora se aplica a la entrada externa horizontal, con el barrido horizontal interno desactivado. En consecuencia, la frecuencia de barrido horizontal queda determinada por la frecuencia de la señal moduladora, y la magnitud de la desviación horizontal es proporcional a la amplitud de la señal moduladora. La desviación vertical depende totalmente de la amplitud y la rapidez de cambio de la señal modulada. En esencia, el haz de electrones emitido del cátodo del tubo queda sometido a la acción simultánea en planos horizontal y vertical.

La fig. 3-26 muestra cómo producen, la señal modulada y la señal moduladora, un patrón trapezoidal. Con un osciloscopio, cuando se aplican 0 V a la entrada horizontal externa, el haz de electrones se centra horizontalmente en la pantalla. Cuando se aplica un voltaje distinto de 0 V a las entradas vertical u horizontal, el haz se desvía en dirección vertical y horizontal, respectivamente. Si se comienza con la onda modulada y la señal moduladora en 0 V ambas (t_0), el haz de electrones se ubica en el centro de la pantalla. A medida que se hace positiva la señal moduladora, el haz se desvía hacia la derecha. Al mismo tiempo, la señal modulada se hace positiva, y

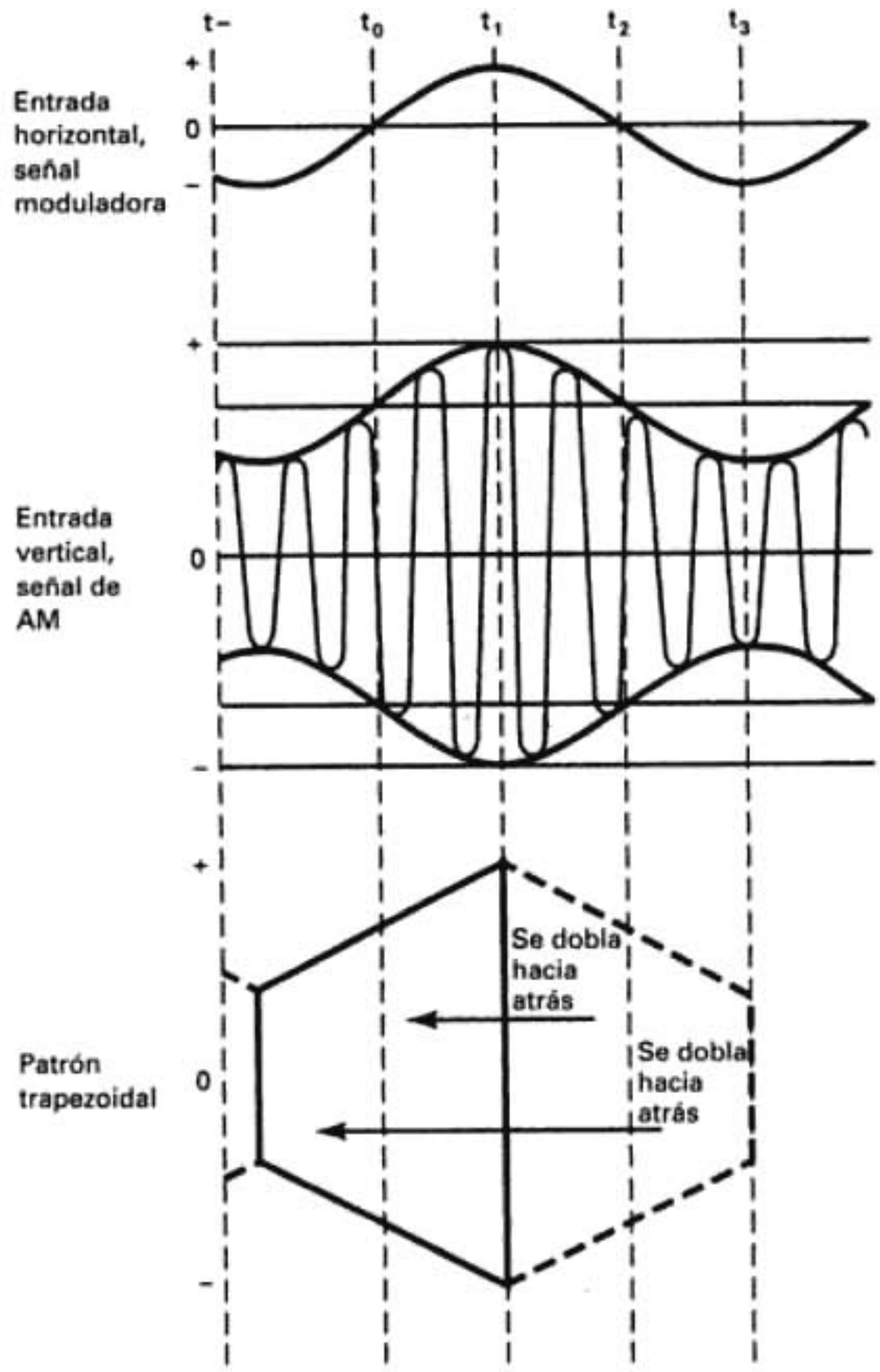


FIGURA 3-26 Producción de un patrón trapezoidal

el rayo se desvía hacia arriba. El haz se continúa desviando hacia la derecha hasta que la señal moduladora llega a su valor máximo positivo, t_1. Mientras el rayo se mueve hacia la derecha, también se desvía hacia arriba y hacia abajo a medida que la señal modulada pasa de positiva a negativa. Obsérvese que en cada cambio sucesivo, la señal modulada llega a una magnitud mayor que en la alternancia anterior. En consecuencia, a medida que el rayo del osciloscopio se desvía hacia la derecha, aumenta su desviación vertical, entre un máximo y otro, en cada ciclo sucesivo de la señal moduladora. Cuando la señal moduladora se hace menos positiva, el rayo se desvía hacia la izquierda (hacia el centro de la pantalla). Al mismo tiempo, la señal modulada pasa alternativamente entre positiva y negativa, desviando el haz hacia arriba y hacia abajo, pero ahora cada ciclo sucesivo tiene menor amplitud que el anterior. En consecuencia, cuando el rayo se mueve en dirección horizontal hacia el centro de la pantalla, disminuye la desviación vertical. La señal moduladora y la señal modulada pasan por 0 V al mismo tiempo, y el rayo está en el centro de la pantalla (t_2). Cuando la señal moduladora se hace negativa, el rayo se desvía hacia el lado izquierdo de la pantalla. Al mismo tiempo, la señal modulada disminuye de magnitud en cada alternancia sucesiva. La señal moduladora llega a su valor máximo negativo en el momento en que la señal modulada llega a su amplitud mínima (t_3). El patrón trapezoidal que se ve entre los tiempos t_1 y t_3 se dobla o refleja a la figura que se produce entre los momentos t– y t_1. Así, se muestra en la pantalla un patrón trapezoidal completo, al terminar los barridos horizontales de izquierda a derecha y derecha a izquierda.

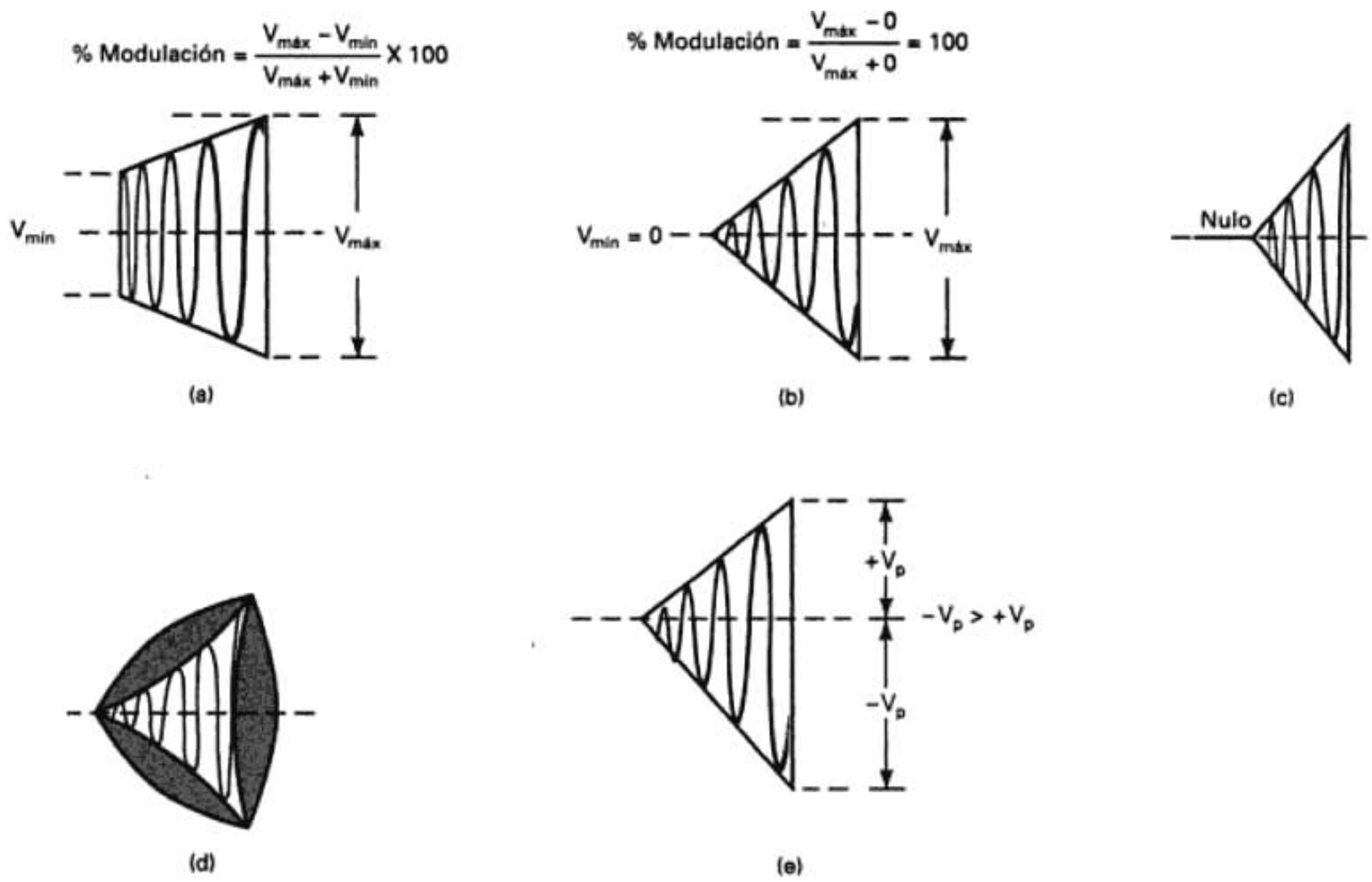


FIGURA 3-27 Patrones trapezoidales: (a) modulación lineal de AM de 50%; (b) modulación de AM de 100%; (c) más de 100% de modulación AM; (d) relación inadecuada de fases; (e) envolvente de AM asimétrica

Si la modulación es simétrica, la mitad superior de la señal modulada es una imagen especular de la mitad inferior, y se produce una figura trapezoidal como la de la fig. 3-27a. Cuando la modulación es de 100%, la amplitud mínima de la señal modulada es cero, y el patrón trapezoidal se junta en un punto en un extremo, como se ve en la fig. 3-27b. Si la modulación es mayor que el 100%, se produce el patrón de la fig. 3-27c. El patrón de la fig. 3-27a es una onda modulada 50%. Si la señal moduladora y la señal modulada están desfasadas, se produce un patrón parecido al de la fig. 3-27d. Si la magnitud de las amplitudes positiva y negativa de la señal modulada no son iguales, resulta un patrón como el de la fig. 3-27e. Si la fase de la señal moduladora se desplaza 180° (se invierte), los patrones trapezoidales tan sólo apuntan en dirección contraria. Como se puede ver, la modulación porcentual y la simetría de modulación se observan con más facilidad en un patrón trapezoidal que en una presentación en pantalla normal de osciloscopio.

Desplazamiento de la portadora

Con frecuencia, se comprende mal o se interpreta mal el término desplazamiento de la portadora. A veces se le llama *modulación ascendente* o *descendente*, y no tiene absolutamente nada que ver con la frecuencia de la portadora. El *desplazamiento de portadora* es una forma de distorsión de amplitud, que se presenta cuando las partes positiva y negativa de la señal modulada de AM no son iguales, es decir, la modulación es asimétrica. El desplazamiento de la portadora puede ser positivo o negativo. Si la parte positiva de la señal modulada tiene mayor amplitud que la negativa, se produce un desplazamiento positivo de la portadora. Si la parte negativa es mayor que la positiva, se produce el desplazamiento negativo de la portadora.

El desplazamiento de la portadora es una indicación del voltaje promedio de una señal modulada de AM. Si las mitades positiva y negativa de la señal modulada son iguales, el voltaje promedio es 0 V. Si la mitad positiva es mayor, el voltaje promedio es positivo, y si la mitad negativa es mayor, el voltaje promedio es negativo. La fig. 3-28a muestra una envolvente simétrica

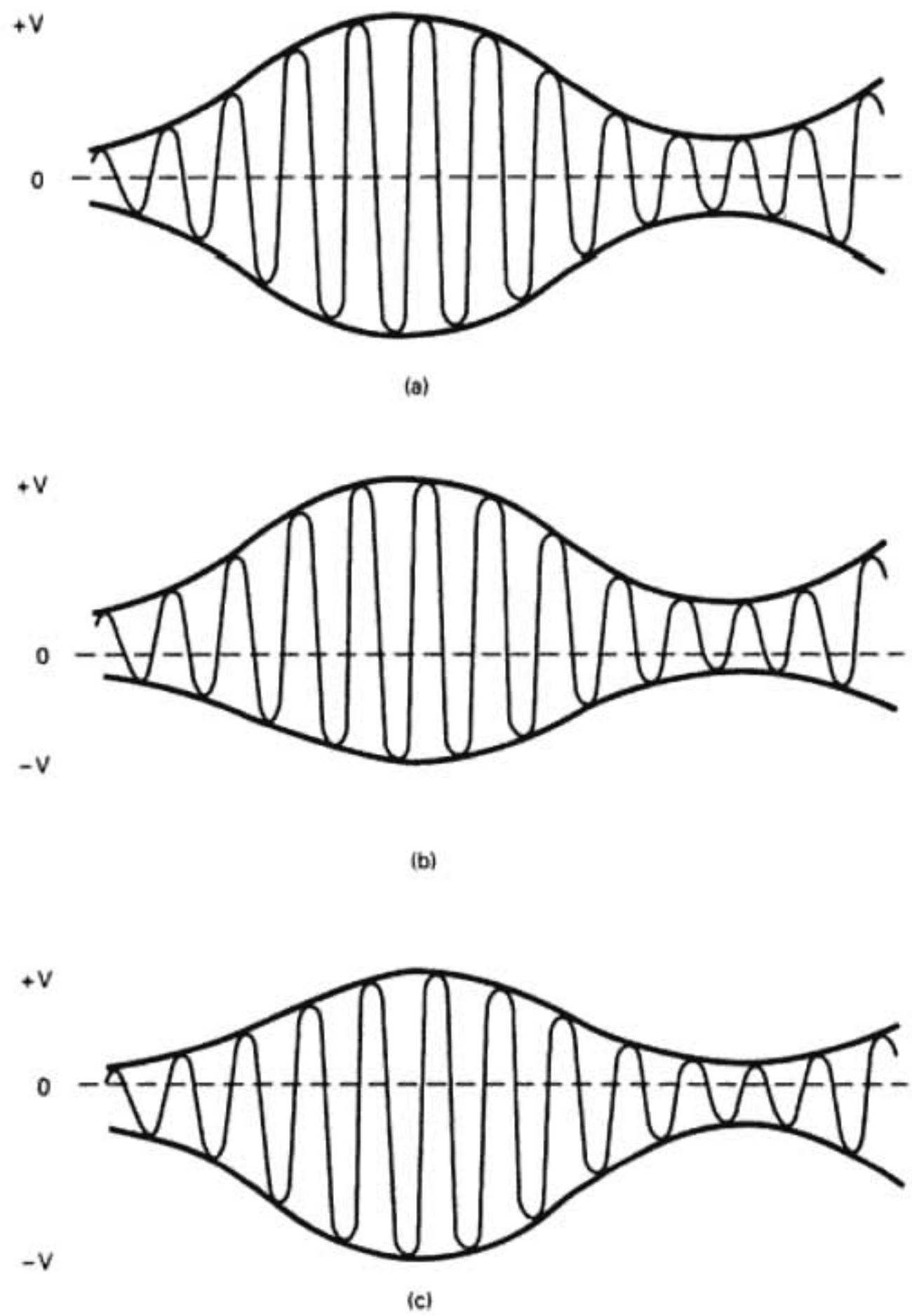


FIGURA 3-28 Desplazamiento de portadora: (a) modulación lineal; (b) desplazamiento positivo de portadora; (c) desplazamiento negativo de portadora

de AM, sin desplazamiento de la portadora; el voltaje promedio es 0 V. Las figuras 3-28b y c representan desplazamientos positivos y negativos de la portadora, respectivamente.

Envolventes de AM producidas por señales complejas no senoidales

Las señales no senoidales son formas de onda compleja, formadas por dos o más frecuencias. Las formas de onda repetitivas complejas son ondas formadas por dos o más ondas senoidales relacionadas armónicamente, y pueden ser ondas cuadradas, rectangulares y triangulares. También, las señales moduladoras complejas pueden contener dos o más frecuencias no relacionadas, como por ejemplo las señales de voz que se originan en fuentes distintas. Cuando se modula una portadora con señales que no sean ondas senoidales o cosenoidales puras, la envolvente modulada contiene frecuencias de banda lateral superior e inferior comparables con las que contiene la señal moduladora y, en consecuencia, la forma de la envolvente se asemeja a la de la onda moduladora original.

La fig. 3-29 muestra señales moduladoras complejas, y sus envolventes respectivas de AM.

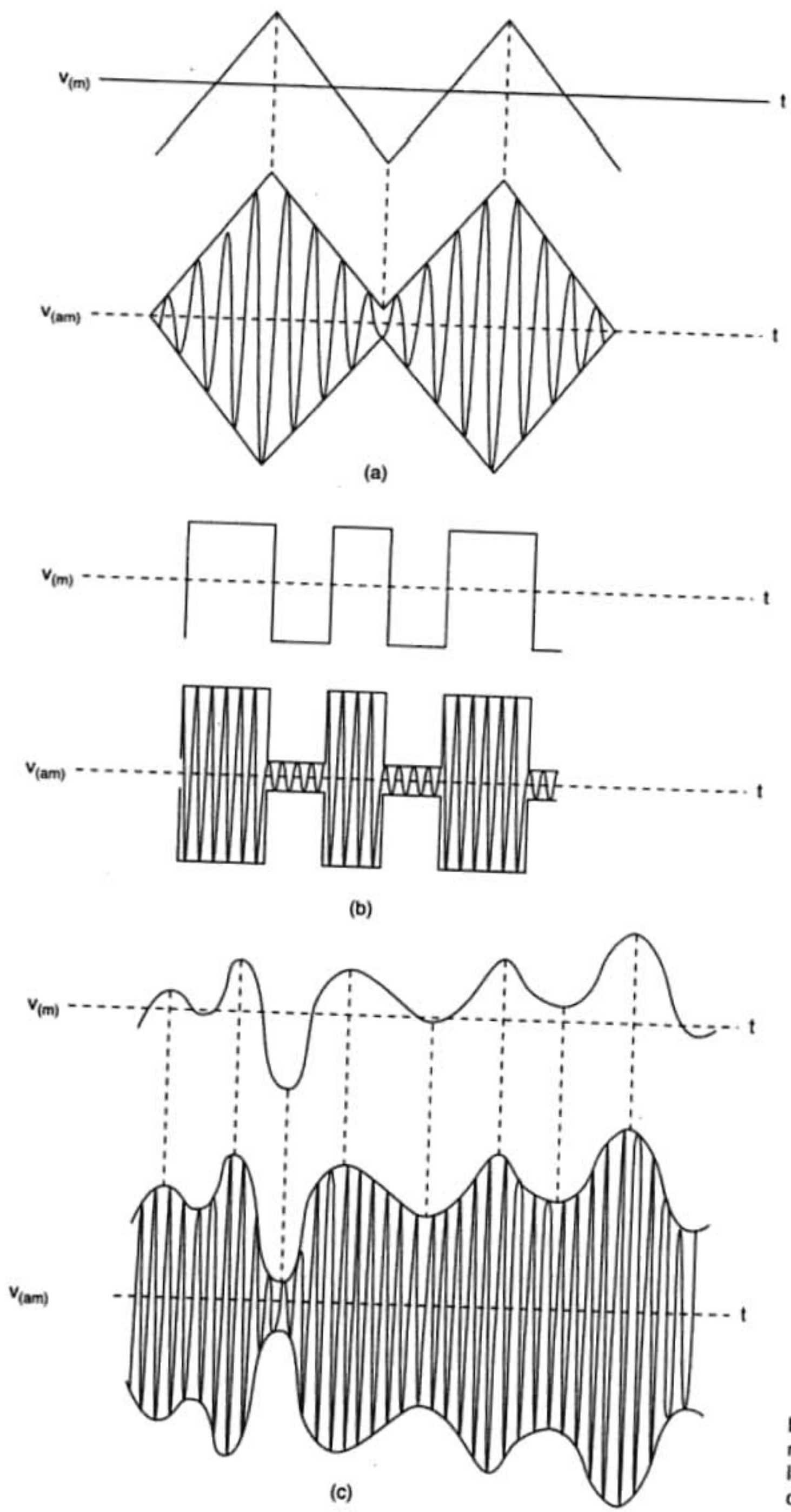
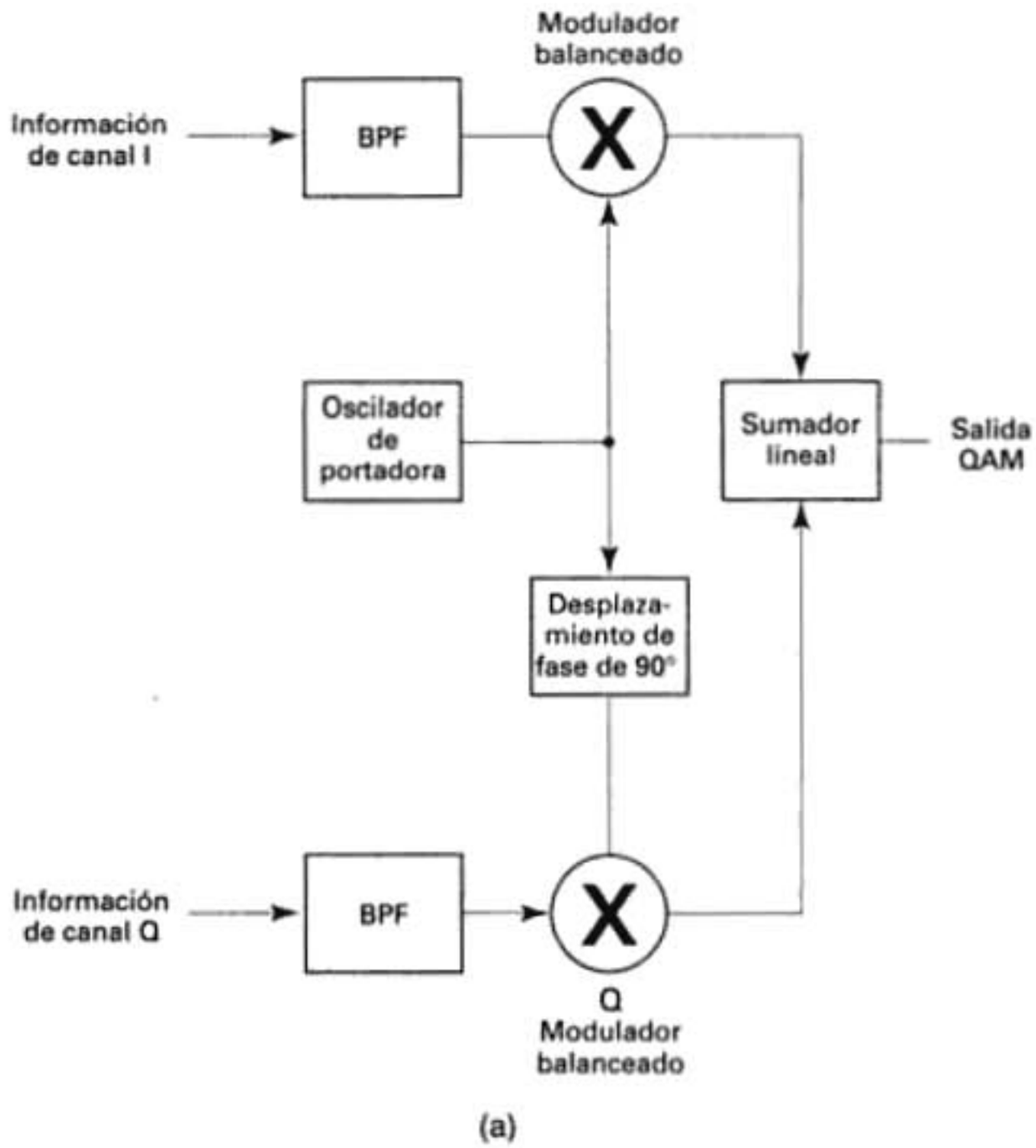
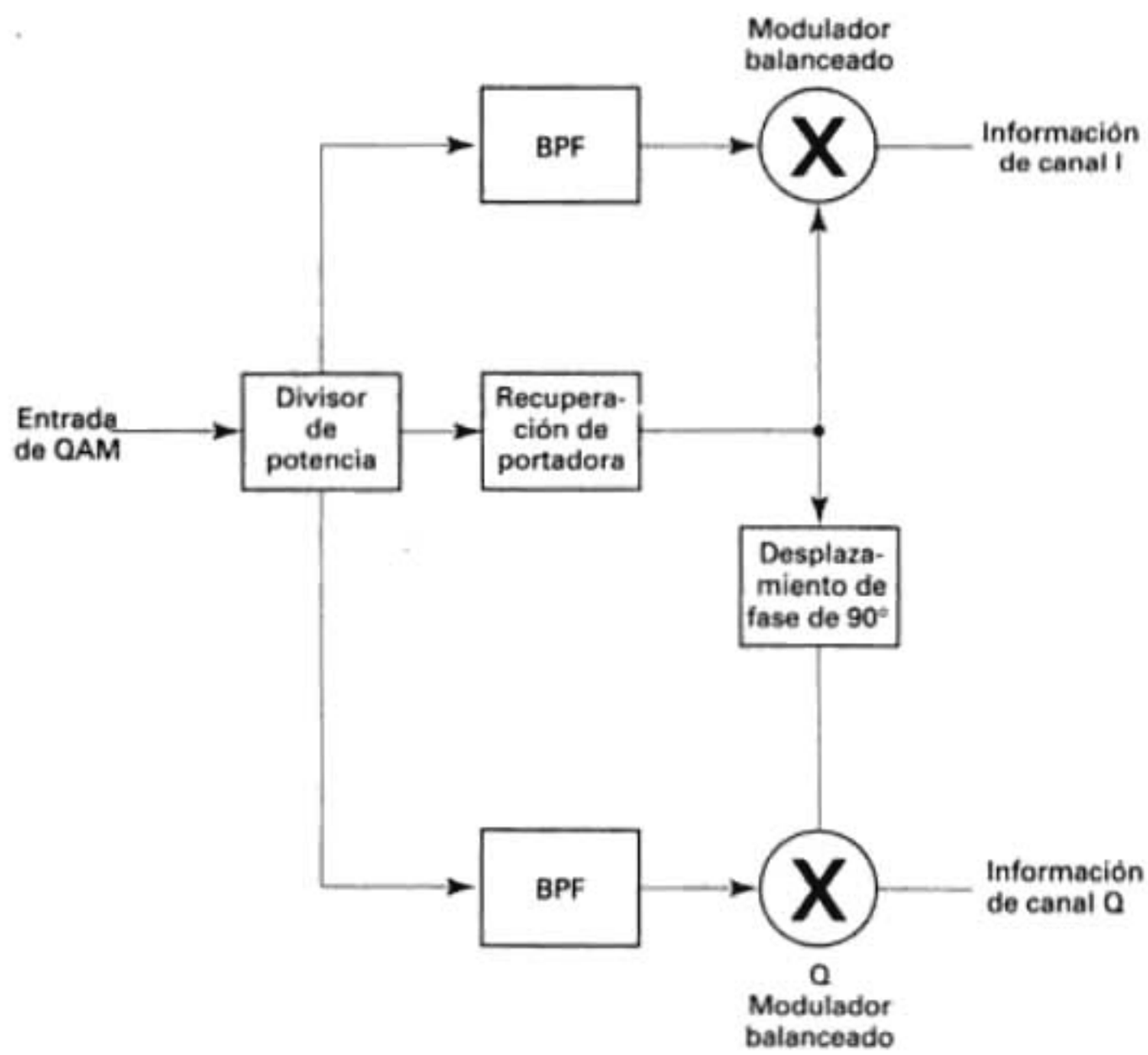


FIGURA 3-29 AM con señal moduladora compleja: (a) modulación con onda triangular; (b) modulación con onda rectangular; (c) modulación de voz



(a)



(b)

FIGURA 3-30 AM en cuadratura: (a) modulador; (b) demodulador

MODULACIÓN DE AMPLITUD EN CUADRATURA

La *modulación de amplitud en cuadratura* es una forma de AM en la que dos señales de fuentes separadas de información (es decir, dos canales) modulan a la misma frecuencia de portadora, al mismo tiempo, sin interferir entre sí. Las fuentes de información modulan la misma portadora, después de haberse separado en dos señales portadoras desfasadas 90° entre sí. Este esquema se llama, a veces, *AM en cuadratura* (QUAM o QAM).

En la fig. 3-30a se ve un diagrama de bloques simplificado de un modulador de AM en cuadratura. Como allí se ve, hay un solo oscilador de portadora, que produce una portadora enfasada al modulador *I*, y a continuación desplaza 90° a la portadora y proporciona una segunda portadora en cuadratura al modulador *Q*. Las salidas de los dos moduladores se suman linealmente, antes de pasar por más etapas de elevación de frecuencia y de amplificación de potencia.

La fig. 3-30b muestra un diagrama simplificado de bloques de un demodulador de AM en cuadratura. Como se ve, para demodular las señales de AM en cuadratura se necesita un circuito recuperador de portadora, que reproduzca la frecuencia y la fase originales de la portadora, y dos moduladores balanceados, para demodular realmente las señales. A este sistema se le llama *detección sincrónica* o *síncrona*, y hace que la demodulación de señales de AM en cuadratura sea bastante costosa con respecto a los circuitos convencionales demoduladores de AM. Se puede ver que la AM en cuadratura es mucho más compleja que la AM convencional, cuesta más realizarla y produce señales demoduladas más o menos con la misma calidad. Sin embargo, la ventaja principal de la AM en cuadratura es la conservación del ancho de banda. La AM en cuadratura sólo necesita la mitad del ancho de banda que la AM convencional, y con dos canales separados se puede modular la misma portadora. A veces, a la QAM se le llama *multiplexado por división de fase*, y fue una de las técnicas de modulación que se consideraron para emitir señales estereofónicas de AM. Por ahora, la AM en cuadratura es el esquema de modulación que se usa para codificar señales de color en los sistemas analógicos de emisoras de televisión.

Hoy, la AM en cuadratura se usa casi exclusivamente para modulación digital de portadoras analógicas, en módems de datos, para mandar datos a través de la red telefónica pública. También se usa en sistemas digitales de comunicaciones satelitales. La modulación digital en cuadratura y la detección sincrónica son temas que se describen con más detalle en capítulos posteriores de este libro.

PREGUNTAS

3-1. Defina *modulación de amplitud.*

3-2. Describa el funcionamiento básico de un *modulador de AM.*

3-3. ¿Qué quiere decir el término *RF*?

3-4. ¿Cuántas entradas hay en un *modulador de amplitud*? ¿Cuáles son?

3-5. En un sistema de comunicaciones de AM, ¿qué significan los términos *señal moduladora, portadora, onda modulada* y *envolvente de AM*?

3-6. ¿Qué quiere decir *frecuencia de repetición* de la envolvente de AM?

3-7. Describa las *bandas laterales superior* e *inferior*, y las *frecuencias laterales superior* e *inferior.*

3-8. ¿Cuál es la relación entre la *frecuencia de la señal moduladora* y el *ancho de banda* en un *sistema convencional de AM*?

3-9. Defina el *coeficiente de modulación* y el *porcentaje de modulación.*

3-10. ¿Cuáles son el coeficiente de modulación y el porcentaje de modulación máximos posibles con un sistema convencional de AM, sin causar demasiada distorsión?

3-11. Para una modulación de 100%, ¿cuál es la relación entre las amplitudes de voltaje de las frecuencias laterales y de la portadora?

3-12. Describa el significado de la siguiente ecuación:

$$v_{am}(t) = E_c \operatorname{sen}(2\pi f_c t) - \frac{mE_c}{2}\cos[2\pi(f_c + f_m)t] + \frac{mE_c}{2}\cos[2\pi(f_c - f_m)t]$$

3-13. Describa el significado de cada término en la siguiente ecuación:

$$v_{am}(t) = 10 \text{ sen}(2\pi 500kt) - 5 \cos(2\pi 515kt) + 5 \cos(2\pi 485kt)$$

3-14. ¿Cuál es el efecto de la modulación sobre la amplitud de la componente portadora del espectro original de señal modulada?

3-15. Describa el significado de la siguiente fórmula:

$$P_t = P_c\left(1 + \frac{m^2}{2}\right)$$

3-16. ¿Qué quiere decir AM de DSBFC?

3-17. Describa la relación entre las potencias de la *portadora* y de la *banda lateral* en una onda DSBFC de AM.

3-18. ¿Cuál es la principal desventaja de la AM de DSBFC?

3-19. ¿Cuál es la principal ventaja de la AM de DSBFC?

3-20. ¿Cuál es la principal desventaja de la AM de bajo nivel?

3-21. ¿Por qué cualquier amplificador que siga al circuito modulador en un transmisor de AM de DSBFC tiene que ser lineal?

3-22. Describa las diferencias entre los *moduladores de bajo* y *alto nivel*.

3-23. Haga una lista de las ventajas de la modulación de bajo nivel, y de la modulación de alto nivel.

3-24. ¿Cuáles son las ventajas de usar *moduladores de circuito integrado lineal* para AM?

3-25. ¿Cuál es la ventaja de usar un *patrón trapezoidal* para evaluar una envolvente de AM?

PROBLEMAS

3-1. Un modulador de AM de DSBFC tiene frecuencia de portadora $f_c = 100$ kHz y una frecuencia máxima de señal moduladora $f_{m(\text{máx})} = 5$ kHz. Calcule:

(a) Límites de frecuencia para las bandas laterales superior e inferior.

(b) Ancho de banda.

(c) Frecuencias laterales superior e inferior producidas cuando la señal moduladora es un tono de frecuencia única de 3 kHz.

A continuación:

(d) Trace el espectro de frecuencias de salida.

3-2. ¿Cuál es la frecuencia máxima de señal moduladora que se puede usar con un sistema de AM de DSBFC, con ancho de banda de 20 kHz?

3-3. Si una onda modulada con voltaje promedio de 20 V_p cambia de amplitud en ± 5 V, determine las amplitudes máxima y mínima de envolvente, el coeficiente de modulación y el porcentaje de modulación.

3-4. Trace la envolvente del problema 3-3. Identifique todos los voltajes pertinentes.

3-5. Para una amplitud de portadora de 30 V_p, calcule las amplitudes máximas de frecuencia lateral superior e inferior de la envolvente de AM de DSBFC.

3-6. Con un voltaje máximo positivo de envolvente de +12 V y una amplitud mínima positiva de envolvente de +4 V, determine el coeficiente de modulación y el porcentaje de modulación.

3-7. Trace la envolvente del problema 3-6. Identifique todos los voltajes pertinentes.

3-8. Una envolvente de AM de DSBFC tiene $+V_{\text{máx}} = 40$ V y $+V_{\text{mín}} = 10$ V. Calcule:

(a) La amplitud de la portadora no modulada.

(b) El cambio máximo de amplitud de la onda modulada.

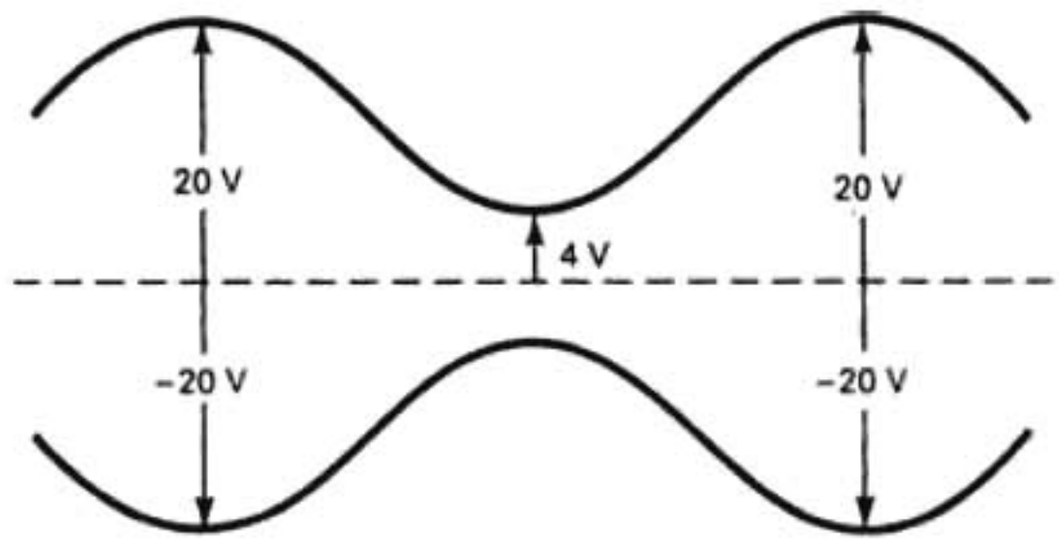
(c) El coeficiente de modulación y el porcentaje de modulación.

3-9. Para una amplitud de portadora no modulada de 16 V_p y un coeficiente de modulación $m = 0.4$, determine las amplitudes de la portadora modulada y las frecuencias laterales.

3-10. Trace la envolvente del problema 3-9. Identifique todos los voltajes pertinentes.

3-11. Para la envolvente de AM que se adjunta, determine:

(a) La amplitud máxima de las frecuencias laterales superior e inferior.
(b) La amplitud máxima de la portadora.
(c) El cambio máximo de amplitud de la envolvente.
(d) El coeficiente de modulación.
(e) El porcentaje de modulación.



3-12. Una entrada a un modulador de AM de DSBFC es una portadora de 800 kHz, con amplitud de 40 V_p. La segunda entrada es una señal moduladora de 25 kHz, cuya amplitud es suficiente para producir un cambio de ±10 V en la amplitud de la envolvente. Calcule:

(a) Las frecuencias laterales superior e inferior.
(b) El coeficiente de modulación y el porcentaje de modulación.
(c) Las amplitudes máximas de los máximos y mínimos de la envolvente.

A continuación:

(d) Trace el espectro de frecuencias de salida.
(e) Trace la envolvente. Identifique todos los voltajes pertinentes.

3-13. Determine, para un coeficiente de modulación $m = 0.2$ y una potencia de portadora no modulada $P_c = 1000$ W:

(a) La potencia total de banda lateral.
(b) La potencia de banda lateral superior e inferior.
(c) La potencia de la portadora modulada.
(d) La potencia total transmitida.

3-14. Determine la potencia máxima de la banda lateral superior, inferior y total para una potencia de portadora no modulada $P_c = 2000$ W.

3-15. Calcule la potencia total máxima transmitida (P_t) con el sistema de AM descrito en el problema 3-14.

3-16. Calcule, para una onda de AM de DSBFC con voltaje de portadora no modulada de 25 V_p y una resistencia de carga de 50 Ω, lo siguiente:

(a) Potencia en la portadora no modulada.
(b) Potencia de la portadora modulada, de las bandas laterales superior e inferior, y potencia total transmitida con un coeficiente de modulación $m = 0.6$.

3-17. Un modulador transistorizado de baja potencia tiene coeficiente de modulación $m = 0.4$, ganancia de voltaje en reposo $A_q = 80$ y amplitud de voltaje de portadora de entrada de 0.002 V. Calcule:

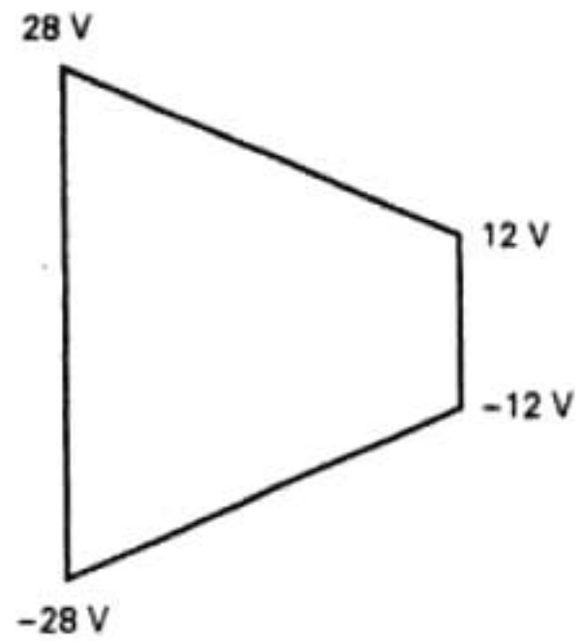
(a) Las ganancias de voltaje máxima y mínima.
(b) Los voltajes máximo y mínimo para v_{sal}.

A continuación:

(c) Trace la envolvente modulada.

3-18. Para el patrón trapezoidal adjunto, determine:

(a) El coeficiente de modulación.
(b) El porcentaje de modulación.
(c) La amplitud de la portadora.
(d) Las amplitudes de las frecuencias laterales superior e inferior.



3-19. Trace los patrones trapezoidales aproximados para las siguientes modulaciones porcentuales y condiciones de modulación:
(a) 100%.
(b) 50%.
(c) >100%.
(d) Relación incorrecta de fase.
(e) Modulación AM no simétrica.

3-20. Para un modulador de AM con frecuencia de portadora f_c = 200 kHz y una frecuencia máxima de señal moduladora $f_{m(\text{máx})}$ = 10 kHz, determine:
(a) Los límites de frecuencia para las bandas laterales superior e inferior.
(b) Las frecuencias de banda superior e inferior producidas cuando la señal moduladora es un tono de frecuencia único de 7 kHz.
(c) El ancho de banda necesario para pasar la frecuencia máxima de la señal moduladora.
A continuación:
(d) Trace el espectro de salida.

3-21. Para un voltaje de portadora no modulada de 10 V_p y un cambio de amplitud de la envolvente de ±4 V, calcule:
(a) El coeficiente de modulación.
(b) El porcentaje de modulación.

3-22. Una envolvente tiene voltaje máximo positivo $V_{\text{máx}}$ = +20 V, y voltaje mínimo positivo de +6 V. Determine:
(a) El coeficiente de modulación.
(b) El porcentaje de modulación.
(c) La amplitud de la portadora.

3-23. Una envolvente tiene $+V_{\text{máx}}$ = +30 V_p y $+V_{\text{mín}}$ = +10 V_p. Calcule:
(a) La amplitud de la portadora no modulada.
(b) La amplitud de la portadora modulada.
(c) El cambio máximo de amplitud de la envolvente.
(d) El coeficiente de modulación.
(e) El porcentaje de modulación.

3-24. Escriba una ecuación de una onda de voltaje de AM con los siguientes valores:
Portadora no modulada = 20 V_p.
Coeficiente de modulación = 0.4.
Frecuencia de la señal moduladora = 5 kHz.
Frecuencia de la portadora = 200 kHz.

3-25. Para una amplitud de portadora no modulada de 12 V_p y un coeficiente de modulación de 0.5, determine lo siguiente:
(a) Porcentaje de modulación.
(b) Voltajes máximos de las frecuencias de portadora y laterales.
(c) Voltaje máximo positivo de la envolvente.
(d) Voltaje mínimo positivo de la envolvente.

3-26. Trace la envolvente del problema 3-25.

3-27. Para una envolvente con voltaje máximo de pico de 52 V y un voltaje mínimo de pico a pico de 24 V, determine lo siguiente:

(a) Modulación porcentual.
(b) Voltajes máximos de las frecuencias de portadora y laterales.
(c) Voltaje máximo positivo de la envolvente.
(d) Voltaje mínimo positivo de la envolvente.

3-28. Una entrada de un modulador AM de DSBFC es una portadora de 500 kHz, con amplitud máxima de 32 V. La segunda entrada es de señal moduladora de 12 kHz, suficiente para producir un cambio de $\pm 14\ V_p$ en la amplitud de la envolvente. Determine lo siguiente:

(a) Frecuencias laterales superior e inferior.
(b) Coeficiente de modulación y porcentaje de modulación.
(c) Amplitudes máxima y mínima de la envolvente.

A continuación:

(d) Trace la envolvente de salida.
(e) Trace el espectro de frecuencias de salida.

3-29. Para un coeficiente de modulación de 0.4 y una potencia de portadora de 400 W, determine:

(a) La potencia total de las bandas laterales.
(b) La potencia total transmitida.

3-30. Una onda de AM de DSBFC tiene voltaje no modulado de portadora de 18 V_p, y resistencia de carga de 72 Ω. Determine:

(a) La potencia de portadora no modulada.
(b) La potencia de la portadora modulada.
(c) La potencia total de las bandas laterales.
(d) Las potencias de las bandas laterales superior e inferior.
(e) La potencia total transmitida.

3-31. Para un modulador de AM de baja potencia, con coeficiente de modulación 0.8, ganancia en reposo 90 y amplitud de portadora de entrada de 10 mV_p, determine:

(a) Las ganancias de voltaje máxima y mínima.
(b) Los voltajes máximo y mínimo de envolvente.

A continuación:

(c) Trace la envolvente de AM.

C A P Í T

Imagen protegida por derechos de autor

Recepción de amplitud modulada

INTRODUCCIÓN

La demodulación de AM es el proceso inverso de modulación de AM. Un receptor convencional de AM de doble banda lateral tan sólo reconvierte una onda de amplitud modulada a la información de la fuente original. Para hacerlo, el receptor debe ser capaz de recibir, amplificar y demodular una onda de AM. También debe ser capaz de limitar la banda del espectro total de radiofrecuencias a determinada banda deseada de frecuencias. El proceso de selección se llama *sintonía del receptor*.

Para comprender bien el proceso de demodulación primero es necesario tener una comprensión básica de la terminología más usada para describir los radiorreceptores y sus características. La fig. 4-1 muestra un diagrama simplificado de bloques de un receptor normal de AM. La *sección de RF* es la primera etapa del receptor, y en consecuencia se le llama también el *frente del receptor*. Las funciones principales de la sección de RF (radiofrecuencia) son detectar, limitar la banda y amplificar las señales de RF recibidas. La *sección de mezclador/convertidor* es la siguiente etapa. Esta sección hace conversión descendente de las frecuencias de RF recibidas en *frecuencias intermedias* (FI), que tan sólo son frecuencias que están en algún lugar entre la RF y las frecuencias de información; de ahí el nombre de *intermedias*. Las funciones principales de la *sección FI* (o *sección IF*, por "*intermediate frequencies*") son de amplificación y selectividad. El *detector de AM* demodula la onda de AM y la convierte en la señal original de información, y la *sección de audio* sólo amplifica la información recuperada.

PARÁMETROS DEL RECEPTOR

Hay varios parámetros de uso común para evaluar las posibilidades de un receptor para demodular bien una señal de radio. Los más importantes son la selectividad y la sensibilidad, que se usan con frecuencia para comparar la calidad de dos radiorreceptores.

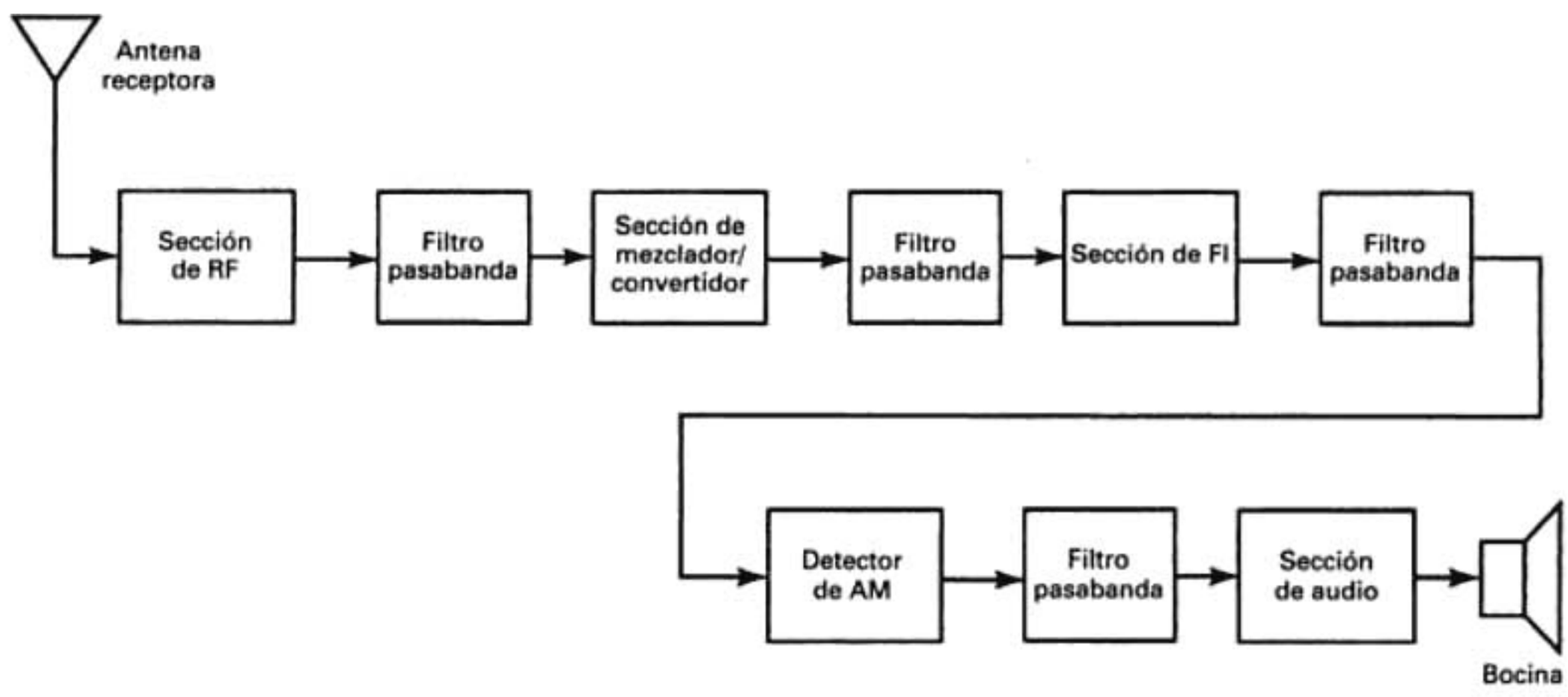


FIGURA 4-1 Diagrama de bloques simplificado de un receptor de AM

Selectividad

La selectividad es un parámetro del receptor con el que se mide la capacidad de éste para aceptar una determinada banda de frecuencias y rechazar las demás. Por ejemplo, en la banda comercial de emisión de AM, a cada estación transmisora se le asigna un ancho de banda de 10 kHz. En consecuencia, para que un receptor seleccione sólo aquellas frecuencias asignadas a un solo canal, debe limitar su ancho de banda a 10 kHz. Si la banda de paso es mayor que 10 kHz, se puede recibir y demodular en forma simultánea más de un canal. Si la banda de paso de un receptor es menor que 10 kHz, se rechaza o bloquea una parte de la información de la señal moduladora para ese canal y, en consecuencia, se pierde.

Hay varias formas aceptables de describir la selectividad de un receptor de radio. Una forma frecuente es simplemente especificar el ancho de banda del receptor en los puntos de −3 dB. Sin embargo, este ancho de banda no es necesariamente una buena forma de determinar lo bien que el receptor rechaza las frecuencias no deseadas. En consecuencia, se acostumbra especificar el ancho de banda en dos niveles de atenuación, por ejemplo, −3 dB y −60 dB. La relación de esos dos anchos de banda se llama *factor de forma*, y se define con la siguiente ecuación

$$SF = \frac{B_{(-60\text{ dB})}}{B_{(-3\text{ dB})}} \tag{4-1}$$

en donde SF = factor de forma (adimensional)
$B_{(-60\text{ dB})}$ = ancho de banda 60 dB abajo del nivel máximo de la señal
$B_{(-3\text{ dB})}$ = ancho de banda 3 dB abajo del nivel máximo de la señal

En el caso ideal, el ancho de banda en los puntos de −3 dB y −60 dB sería igual, y el factor de forma sería 1. Es imposible, naturalmente, lograr este valor con un circuito práctico. Un radiorreceptor normal de AM podría tener un ancho de banda en −3 dB de 10 kHz, y uno de 20 kHz a −60 dB, con lo que el factor de forma sería 2. Los radiorreceptores más costosos y complicados de satélite, microondas y radioteléfonos tienen factores de forma más cercanos al valor ideal de 1.

En el congestionado espectro actual de radiofrecuencias, la FCC hace las asignaciones de canal adyacente tan cercanas entre sí como sea posible, y sólo 10 kHz separan los canales de emisión comercial de AM. La distancia entre canales comerciales adyacentes de emisión en FM

es 200 kHz, y la separación de los canales comerciales de TV es 6 MHz. Un radiorreceptor debe ser capaz de separar las señales del canal deseado sin permitir que entre la interferencia de un canal adyacente en la banda de paso del canal deseado.

Mejoramiento del ancho de banda

Como se dijo en el capítulo 1 y de acuerdo con la ecuación 1-24, el ruido térmico es la forma más prevaleciente de ruido, y es directamente proporcional al ancho de banda. En consecuencia, si se puede reducir el ancho de banda, el ruido también se reducirá en la misma proporción y aumentará la relación de potencias de señal a ruido, y mejorará la eficiencia del sistema. Naturalmente hay un límite de la eficiencia del sistema respecto a lo que se puede reducir el ancho de banda. El límite inferior es que el ancho de banda del circuito debe ser mayor que el ancho de banda de la señal de información, porque de otro modo se reducirán la potencia de información y/o el contenido de frecuencias de la señal de información, con lo que desmerecerá el desempeño del sistema. Cuando una señal se propaga de la antena a través de la sección de RF, de la de mezclador/convertidor y de la de FI, se reduce el ancho de banda y en consecuencia se reduce el ruido. El problema teórico es cuánto se debe reducir el ancho de banda, y el problema práctico está en la dificultad de fabricar filtros estables de banda angosta.

La relación de señal a ruido en la entrada se calcula en el frente de un receptor, usando el ancho de banda de RF para medir la potencia del ruido. Sin embargo, el ancho de banda de RF en general es mayor que el del resto del receptor; es decir, el ancho de banda de FI es más angosto que el de RF, por razones que explicaremos en las secciones siguientes de este capítulo. Reducir el ancho de banda equivale a reducir (mejorar) la cantidad de ruido del receptor. La relación de reducción de ruido alcanzada reduciendo el ancho de banda se llama *mejoramiento del ancho de banda* (BI, de *bandwidth improvement*) y se define matemáticamente como sigue

$$\mathrm{BI} = \frac{B_{\mathrm{RF}}}{B_{\mathrm{IF}}} \tag{4-2}$$

en donde BI = mejoramiento del ancho de banda (adimensional)
B_{RF} = ancho de banda de RF (hertz)
B_{IF} = ancho de banda de FI (hertz)

La reducción correspondiente de ruido debida a la reducción de ancho de banda se llama *mejoramiento de la cifra de ruido* y se expresa en dB como sigue

$$\mathrm{NF}_{\mathrm{mejoramiento}} = 10 \log \mathrm{BI} \tag{4-3}$$

Ejemplo 4-1

Determinar el mejoramiento de la cifra de ruido para un receptor con un ancho de banda de RF igual a 200 kHz y un ancho de banda de FI igual a 10 kHz.

Solución El mejoramiento del ancho de banda se calcula sustituyendo en la ecuación 4-2

$$\mathrm{BI} = \frac{200\ \mathrm{kHz}}{10\ \mathrm{kHz}} = 20$$

y el mejoramiento de la cifra de ruido se calcula con la ecuación 4-3

$$\mathrm{NF}_{\mathrm{mejoramiento}} = 10 \log 20 = 13\ \mathrm{dB}$$

Sensibilidad

La *sensibilidad* de un receptor es el nivel mínimo de la señal de RF que se puede detectar a la entrada del receptor y producir una señal útil de información demodulada. Es algo arbitrario la definición de qué es una señal útil de información. En general, se usan la relación de señal a ruido y la potencia de la señal a la salida de la sección de audio, para determinar la calidad de una señal recibida, y determinar si es útil. Para los receptores en la banda comercial de emisión en AM, se considera como útil una relación de señal a ruido de 10 dB o más, con 1/2 W (27 dBm) de potencia a la salida de la sección de audio. Sin embargo, para los receptores de microondas de banda ancha, el valor mínimo aceptable tiene una relación de señal a ruido de 40 dB o más, con

unos 5 mW (7 dBm) de potencia de señal. La sensibilidad de un receptor se suele especificar en microvolts de señal recibida. Por ejemplo, una sensibilidad normal de un receptor de banda comercial de emisión en AM es de 50 μV, y en general un radioteléfono móvil tiene una sensibilidad de 0.1 a 10 μV. También, a la sensibilidad del receptor se le llama *umbral* del receptor. La sensibilidad de un receptor de AM depende de la potencia de ruido presente en la entrada al receptor, la cifra de ruido (indicación del ruido generado en el frente del receptor), la sensibilidad del detector de AM y el factor de mejoramiento de ancho de banda del receptor. La mejor manera de mejorar la sensibilidad de un receptor es reducir el nivel de ruido. Esto se logra reduciendo la temperatura o el ancho de banda del receptor, o mejorando la cifra de ruido del receptor.

Margen dinámico

El *margen dinámico* de un receptor se define como la diferencia en decibeles entre el nivel de entrada mínimo necesario para discernir una señal, y el valor de entrada que sobreexcita, o satura, al receptor, y produce distorsión. En términos sencillos, el margen dinámico es el intervalo de potencias de entrada dentro del cual el receptor es útil. El nivel mínimo de recepción es una función del ruido frontal, de la cifra de ruido y de la calidad deseada de señal. El nivel de señal de entrada que produce distorsión por sobreexcitación es una función de la ganancia neta del receptor, es decir, de la ganancia total de todas las etapas del receptor. El límite de alta potencia de un receptor depende de si trabajará con una señal de entrada de una o de varias frecuencias. Si se usa operación con frecuencia única, se suele usar como límite superior de utilidad el *punto de compresión de 1 dB*. Este punto se define como la potencia de salida cuando la respuesta del amplificador de RF es 1 dB menor que la respuesta ideal con ganancia lineal. La fig. 4-2 muestra la ganancia lineal y el punto de compresión de 1 dB para un amplificador normal en el que la ganancia lineal cae justo antes de la saturación. El punto de compresión de 1 dB se mide con frecuencia en forma directa, como el punto en el que un aumento de 10 dB en la potencia de entrada da como resultado un aumento de 9 dB en la potencia de salida.

Se considera como aproximadamente el máximo posible, un margen dinámico de 100 dB. Un bajo margen dinámico puede provocar una desensibilización de los amplificadores de RF, y causar gran distorsión por intermodulación en las señales más débiles de entrada. Más adelante en este capítulo se describirán las medidas de sensibilidad.

Fidelidad

La *fidelidad* es una medida de la capacidad de un sistema de comunicaciones para producir, a la salida del receptor, una réplica exacta de la información de la fuente original. Toda variación de

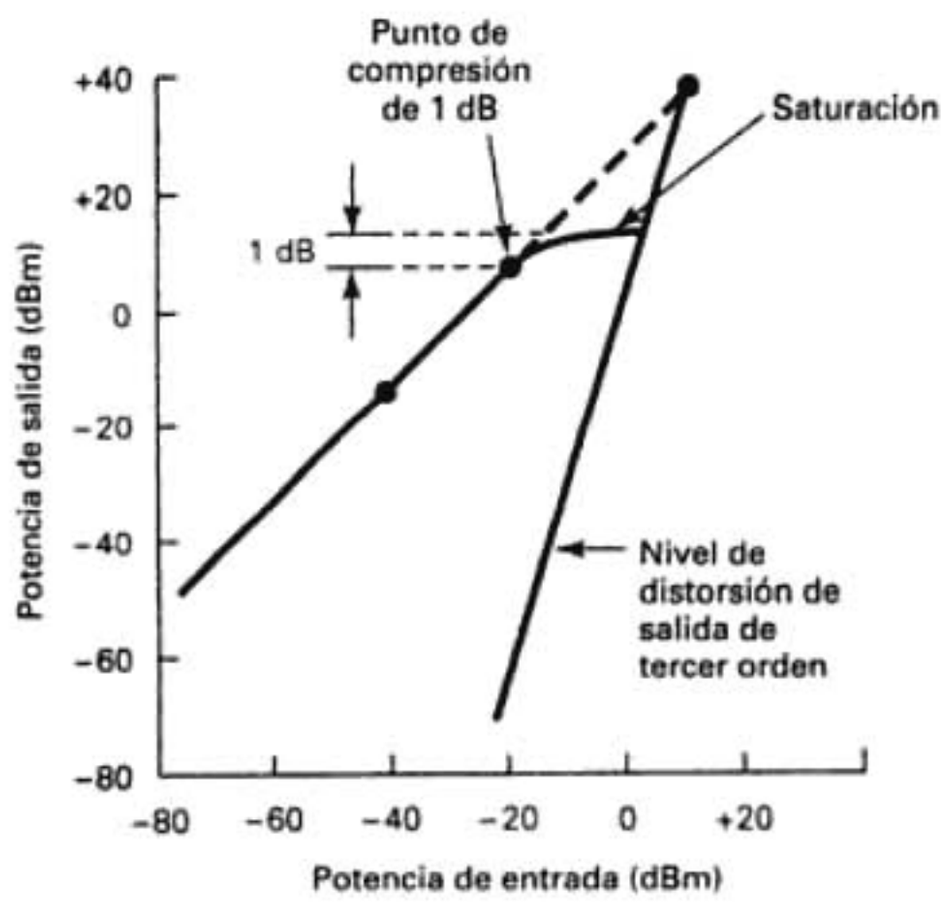


FIGURA 4-2 Ganancia lineal, punto de compresión de 1 dB y distorsión de intersección de tercer orden en un amplificador normal

fase, frecuencia o amplitud que haya en la forma de onda demodulada y que no había en la señal original de información se considera como distorsión.

En esencia, hay tres formas de distorsión que pueden deteriorar la fidelidad de un sistema de comunicaciones: de *amplitud, frecuencia* y *fase*. La distorsión de fase no tiene mucha importancia en la transmisión de voz, porque el oído humano es relativamente insensible hacia las variaciones de fase. Sin embargo, la distorsión de fase puede ser devastadora en la transmisión de datos. La causa principal de la distorsión de fase es el filtrado, tanto intencional como indeseado. Las frecuencias iguales o cercanas a la frecuencia de corte de un filtro sufren diversos valores de desplazamiento de fase. En consecuencia, la frecuencia de corte de un filtro se ajusta, casi siempre, más allá del valor mínimo necesario para pasar las señales de información de frecuencias máximas (normalmente, la frecuencia superior de corte de un filtro pasabajas es más o menos 1.3 por el valor mínimo). El *desplazamiento absoluto de fase* es el desplazamiento total de fase sufrido por una señal, y en general se puede tolerar siempre que todas las frecuencias sufran la misma cantidad de retardo de fase. El *desplazamiento diferencial de fase* se presenta cuando distintas frecuencias sufren distintos desplazamientos de fase, y puede tener un efecto negativo sobre una forma de onda complicada, en especial si la información está codificada en la fase de la portadora, como en la modulación por conmutación de fase. Si el desplazamiento de fase es lineal respecto a la frecuencia, el retardo es constante respecto a la frecuencia. Si no todas las frecuencias se retardan la misma cantidad de tiempo, la relación entre frecuencia y fase de la forma de onda recibida no es consistente con la información de la fuente original, y la información recuperada se distorsiona.

La distorsión de amplitud se presenta cuando las características de amplitud en función de frecuencia de una señal a la salida del receptor difieren de las de la señal original de información. La distorsión de amplitud es el resultado de *ganancia no uniforme* en amplificadores y filtros.

La distorsión de frecuencia se presenta cuando hay frecuencias en una señal recibida que no había en la fuente original de información. La distorsión de frecuencia se debe a distorsión por armónicas y por intermodulación, causada por amplificación no lineal. Los *productos de segundo orden* ($2f_1$, $2f_2$, $f_1 \pm f_2$, etcétera) en general sólo son problema en los sistemas de banda ancha, porque suelen caer fuera del ancho de banda cuando el sistema es de banda angosta. Sin embargo, los *productos de tercer orden* sí caen con frecuencia dentro del ancho de banda del sistema, y producen una distorsión llamada *distorsión de intersección de tercer orden.* Esta distorsión de tercer orden es un caso especial de la distorsión por intermodulación, y es la forma principal de distorsión de frecuencia. Los componentes de intermodulación de tercer orden son las frecuencias de producto cruzado que se producen cuando la segunda armónica de una señal se suma a la frecuencia fundamental de otra señal (es decir, $2f_1 \pm f_2$, $f_1 \pm 2f_2$, etcétera). La distorsión de frecuencia se puede reducir usando un *dispositivo de ley cuadrada*, como un FET, en el frente de un receptor. Los dispositivos de ley cuadrada tienen una ventaja única sobre los BJT, porque sólo producen componentes de segunda armónica y de intermodulación. La fig. 4-2 muestra una curva característica de distorsión de tercer orden, típica, en función de la potencia en la entrada del amplificador y de su ganancia.

Pérdida de inserción

La pérdida de inserción (IL, de *insertion loss*) es un parámetro asociado con las frecuencias que caen en la banda de paso de un filtro, y en general se define como la relación de la potencia transferida a una carga, con un filtro en el circuito, entre la potencia transferida a una carga sin el filtro. Como los filtros se suelen fabricar con componentes propensos a pérdidas, como resistores y capacitores imperfectos, hasta las señales que caen dentro de la banda de paso de un filtro son atenuadas (su magnitud se reduce). Las pérdidas normales de inserción están entre unas décimas de decibel hasta varios decibeles. En esencia, la pérdida de inserción no es más que la relación de la potencia de salida de un filtro a la potencia de entrada, para frecuencias dentro de la banda de paso del filtro, y se expresa matemáticamente, en decibeles, como

$$\mathrm{IL}_{(\mathrm{dB})} = 10 \log \frac{P_{\mathrm{sal}}}{P_{\mathrm{ent}}} \tag{4-4}$$

TABLA 4-1 (T = 17° C)

NF (dB)	F (adimensional)	T_e (°K)
0.8	1.2	58
1.17	1.31	90
1.5	1.41	119
2.0	1.58	168

Temperatura de ruido y temperatura equivalente de ruido

Como el ruido térmico es directamente proporcional a la temperatura, es razonable expresar al ruido en grados, igual que en watts o en volts. Si se reordena la ecuación 1-24 se obtiene

$$T = \frac{N}{KB} \tag{4-5}$$

en donde T = temperatura ambiente (grados kelvin)
N = potencia del ruido (watts)
K = constante de Boltzmann (1.38×10^{-23} J/K)
B = ancho de banda (*hertz*)

La *temperatura equivalente de ruido*, T_e, es un valor hipotético que no se puede medir en forma directa. T_e es un parámetro que se usa con frecuencia en radiorreceptores complicados con bajo ruido, y no es una cifra de ruido. La T_e es una indicación de la reducción de señal a ruido a medida que una señal se propaga a través de un receptor. Mientras menor sea la temperatura equivalente de ruido, la calidad del receptor es mejor. Los valores normales de T_e van de 20° para los receptores *fríos*, hasta 1000° para los receptores *ruidosos*. La ecuación matemática de T_e en la entrada a un receptor se expresa como sigue

$$T_e = T(\mathrm{F} - 1) \tag{4-6}$$

siendo T_e = temperatura equivalente de ruido (grados kelvin)
T = temperatura ambiente (grados kelvin)
F = factor de ruido (adimensional)

La tabla 4-1 muestra algunos valores de cifra de ruido, factor de ruido y temperatura equivalente de ruido para una temperatura ambiente de 17° C (290° K).

RECEPTORES DE AM

Hay dos tipos básicos de radiorreceptores: *coherentes* y *no coherentes*. Con un receptor coherente o *síncrono*, las frecuencias generadas en el receptor, que se usan para demodulación, se sincronizan con las frecuencias de un oscilador, generadas en el transmisor (el receptor debe tener algún medio de recuperar la portadora recibida y sincronizarse a ella). Con los receptores no coherentes o *asíncronos*, no se generan frecuencias en el receptor, o bien las frecuencias que se usan para demodular son totalmente independientes de la frecuencia de la portadora del transmisor. A la *detección no coherente* se le llama con frecuencia *detección de envolvente*, porque la información se recupera a partir de la onda recibida, detectando la forma de la envolvente modulada. Los receptores que se describen en este capítulo son no coherentes. Los receptores coherentes se describen en el capítulo 5.

Receptor de radiofrecuencia sintonizada

El *receptor de radiofrecuencia sintonizada* (TRF, de *tuned radio-frequency*) fue una de las primeras clases de receptores de AM. Es probable que los receptores TRF sean los de diseño más sencillo que se consiguen en la actualidad; sin embargo tienen algunos inconvenientes que limitan su empleo a algunas aplicaciones. La fig. 4-3 muestra un diagrama de bloques de un radiorreceptor TRF de tres etapas que incluye una etapa de RF, una etapa detectora y una etapa

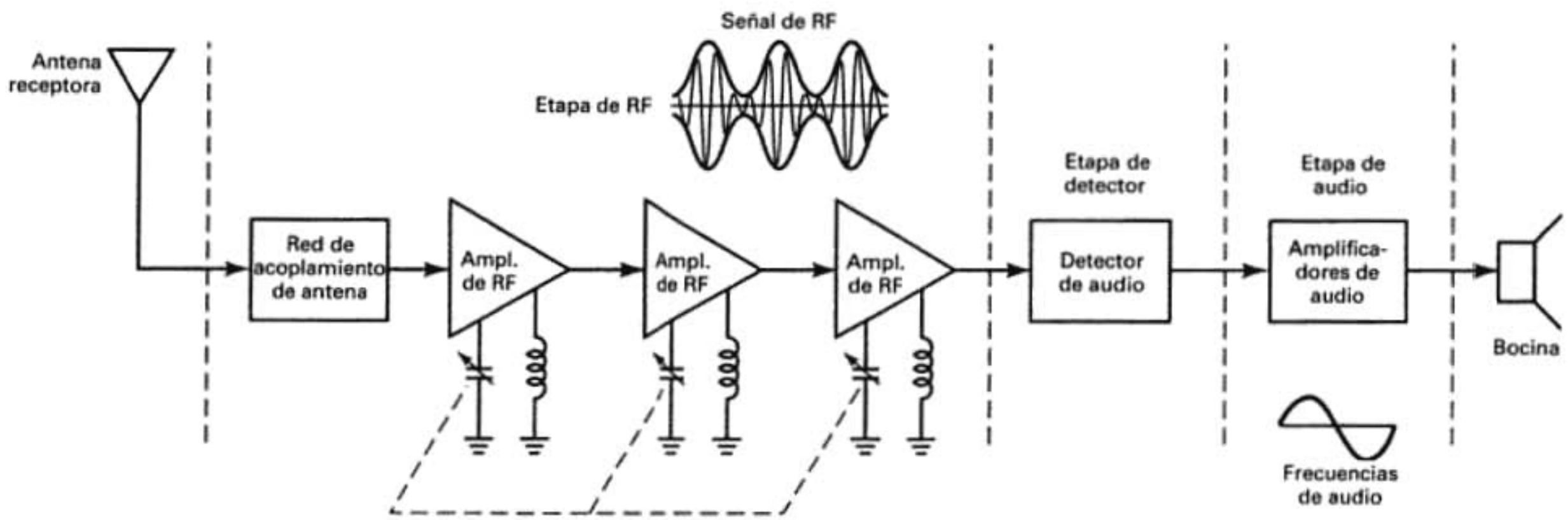


FIGURA 4-3 Diagrama de bloques de un receptor no coherente sintonizado de radiofrecuencias

de audio. En general, se requieren dos o tres amplificadores de RF para filtrar y amplificar la señal recibida hasta un nivel suficiente para excitar la etapa del detector. El detector convierte las señales de RF en forma directa a información, y la etapa de audio amplifica las señales de información hasta un nivel útil.

Aunque los receptores TRF son sencillos y tienen sensibilidad relativamente alta, tienen tres desventajas definidas que limitan su utilidad a aplicaciones de un solo canal y baja frecuencia. La principal desventaja es que su ancho de banda es inconsistente, y varía con la frecuencia central cuando se sintoniza dentro de un amplio margen de frecuencias de entrada. Esto se debe a un fenómeno llamado *efecto película (skin effect)*. A las radiofrecuencias, el flujo de corriente se limita a la zona externa de un conductor, y entonces, mientras mayor sea la frecuencia, el área efectiva es menor y la resistencia es mayor. En consecuencia, el *factor de calidad* ($Q = R/X_L$) de los circuitos tanque permanece relativamente constante dentro de un amplio margen de frecuencias, haciendo que aumente el ancho de banda (f/Q) con la frecuencia. Como resultado, la selectividad del filtro de entrada cambia dentro de límites apreciables de frecuencias de entrada. Si el ancho de banda se ajusta al valor deseado para señales de RF de baja frecuencia, será excesivo para señales de alta frecuencia.

La segunda desventaja de los receptores TRF es la inestabilidad debida a la gran cantidad de amplificadores de RF, todos sintonizados a la misma frecuencia central. Los amplificadores de varias etapas para alta frecuencia son susceptibles de oscilar. Este problema se puede reducir algo sintonizando cada amplificador a una frecuencia un poco distinta, ligeramente arriba o abajo de la frecuencia central deseada. A esta técnica se le llama *sintonía escalonada*. La tercera desventaja de los receptores TRF es que sus ganancias no son uniformes dentro de un margen muy amplio de frecuencias, debido a las relaciones L/C no uniformes de los circuitos tanque acoplados con transformador en los amplificadores de RF.

Con el desarrollo del *receptor superheterodino*, los receptores TRF casi no se usan, excepto para receptores de propósito especial y de una sola estación, y en consecuencia no se describirán más.

Ejemplo 4-2

Para un receptor de AM en la banda comercial de emisiones (535 kHz a 1605 kHz) con un factor Q del filtro de entrada igual a 54, calcular el ancho de banda, y los extremos bajo y alto del espectro de RF.

Solución El ancho de banda en el extremo bajo del espectro de AM se centra en torno a una frecuencia de portadora de 540 kHz, y es

$$B = \frac{f}{Q} = \frac{540 \text{ kHz}}{54} = 10 \text{ kHz}$$

El ancho de banda en el extremo de alta frecuencia del espectro de AM está centrado respecto a una frecuencia de portadora de 1600 kHz, y es

$$B = \frac{1600 \text{ kHz}}{54} = 29.630 \text{ Hz}$$

El ancho de banda de −3-dB en el extremo de baja frecuencia del espectro de AM es exactamente 10 kHz, que es el valor deseado. Sin embargo, en el extremo de alta frecuencia es casi 30 kHz, tres veces el intervalo deseado. En consecuencia, al sintonizar estaciones en el extremo alto del espectro se recibirían en forma simultánea tres estaciones.

Para lograr un ancho de banda de 10 kHz en el extremo de alta frecuencia del espectro, se requiere una Q de 160 (1600 kHz/10 kHz). Con una Q de 160, el ancho de banda en el extremo de baja frecuencia es

$$B = \frac{540 \text{ kHz}}{160} = 3.375 \text{ Hz}$$

que naturalmente es demasiado selectivo, porque suprimiría unas dos terceras partes del ancho de banda de información.

Receptor superheterodino

La selectividad no uniforme del TRF condujo al desarrollo del *receptor superheterodino*, cerca del final de la Primera Guerra Mundial. Aunque la calidad del receptor superheterodino ha mejorado

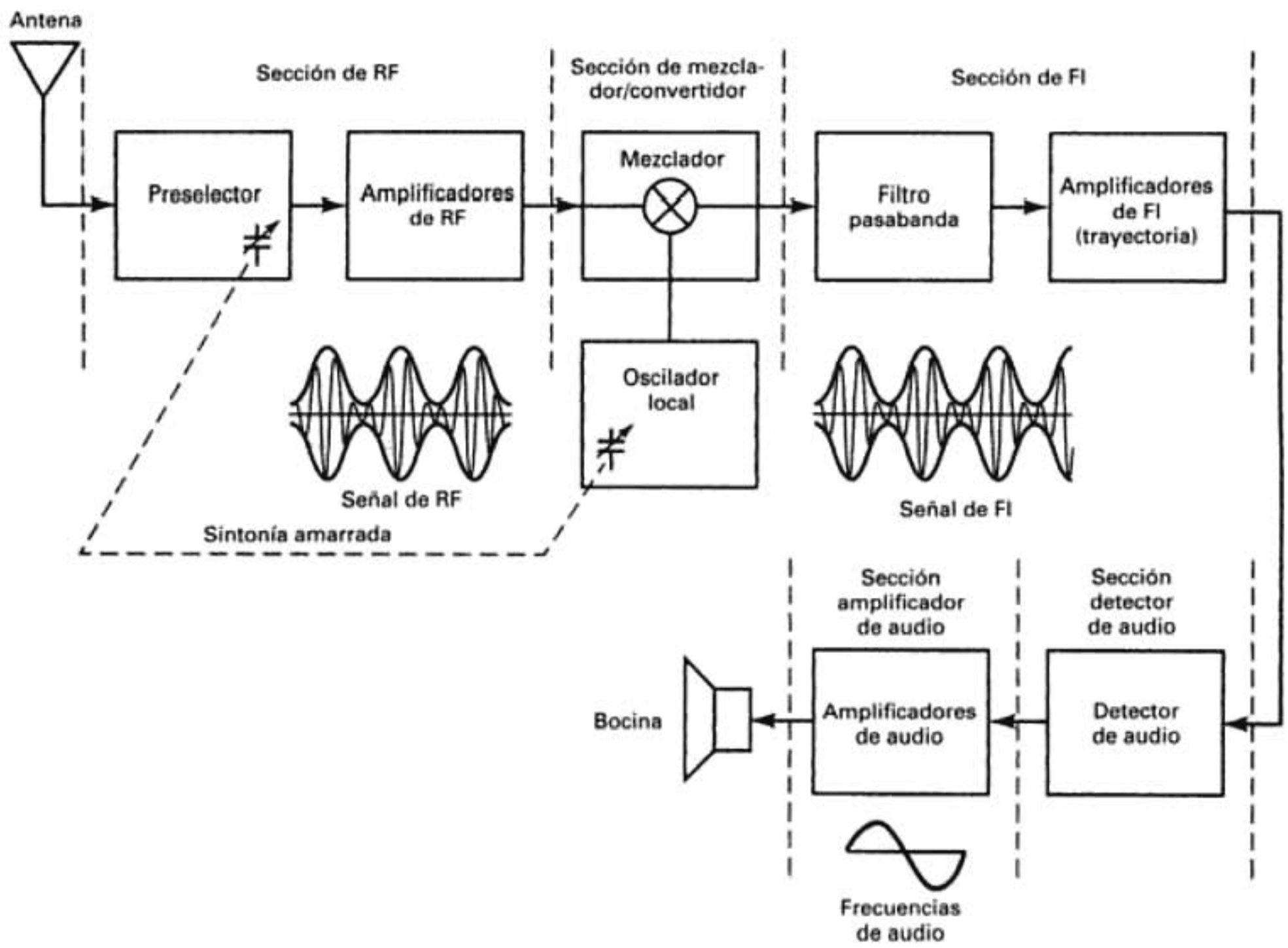


FIGURA 4-4 Diagrama de bloques de un receptor superheterodino

mucho respecto a su diseño original, no ha cambiado mucho su configuración básica, y se sigue usando hoy en una gran variedad de servicios de radiocomunicación. Se ha seguido usando porque su ganancia, selectividad y sensibilidad son mejores que las de otras configuraciones de receptor.

Heterodinar quiere decir mezclar dos frecuencias en un dispositivo no lineal, o trasladar una frecuencia a otra usando mezclado no lineal. En la fig. 4-4 se ve un diagrama de bloques de un receptor superheterodino no coherente. En esencia, un receptor superheterodino tiene cinco secciones: la sección de RF, la sección de mezclador/convertidor, la sección de FI, la sección del detector de audio y la sección del amplificador de audio.

Sección de RF. La sección de RF consiste en general en una etapa preselectora y en una amplificadora. Pueden ser circuitos separados, o un solo circuito combinado. El preselector es un filtro pasabanda sintonizado a banda ancha, con frecuencia central ajustable, que se sintoniza con la frecuencia portadora deseada. El objetivo principal del preselector es proporcionar suficiente límite inicial de banda para evitar que entre una radiofrecuencia específica no deseada, llamada *frecuencia imagen* (la cual se explicará más adelante en esta sección). El preselector también reduce el ancho de banda de ruido del receptor, y proporciona el paso inicial para reducir el ancho general de banda al mínimo requerido para pasar las señales de información. El amplificador de RF determina la sensibilidad del receptor (es decir, establece el umbral de señal). También, como el amplificador de RF es el primer dispositivo activo que encuentra una señal recibida, es el principal generador de ruido y por consiguiente es un factor predominante para determinar la cifra de ruido del receptor. Un receptor puede tener uno o varios amplificadores de RF, o puede no tener ninguno, dependiendo de la sensibilidad deseada. Algunas de las ventajas de incluir amplificadores de RF en un receptor son las siguientes:

1. Mayor ganancia y por consiguiente mayor sensibilidad.
2. Mejor rechazo de frecuencia imagen.
3. Mejor relación de señal a ruido.
4. Mejor selectividad.

Sección de mezclador/convertidor. En la sección de mezclador/convertidor está una etapa de oscilador de radiofrecuencia (que se suele llamar *oscilador local*) y una etapa de mezclador/convertidor (que se suele llamar *primer detector*). El oscilador local puede ser cualquiera de los circuitos osciladores que se describieron en el capítulo 2, dependiendo de la estabilidad y la exactitud deseadas. La etapa de mezclador es un dispositivo no lineal, y su objetivo es convertir las radiofrecuencias en frecuencias intermedias (traslación de RF a FI). El heterodinado se lleva a cabo en la etapa de mezclador, y las radiofrecuencias se bajan a frecuencias intermedias. Aunque las frecuencias de portadora y de las bandas laterales se van de RF a FI, la forma de la envolvente permanece igual y, en consecuencia, la información original que contiene la envolvente permanece sin cambios. Es importante observar que, aunque la portadora y las frecuencias laterales superior e inferior cambian de frecuencia, el ancho de banda no cambia en el proceso de heterodinado. La frecuencia intermedia más usada en los receptores de la banda de emisión en AM es 455 kHz.

Sección de FI. La sección de FI consiste en una serie de amplificadores y filtros pasabanda de FI que se llama con frecuencia la *trayectoria de FI*. La mayor parte de la ganancia y la selectividad del receptor se hacen en la sección de FI. La frecuencia central y el ancho de banda de FI son constantes para todas las estaciones, y se escogen de tal manera que su frecuencia sea menor que cualquiera de las señales de RF que se van a recibir. La FI siempre tiene menor frecuencia que la RF, porque es más fácil y menos costoso fabricar amplificaciones estables de alta ganancia para señales de baja frecuencia. Por lo anterior no es raro ver un receptor con cinco o seis amplificadores de FI y un solo amplificador de RF, o quizá sin amplificador de RF.

Sección de detector. El objetivo de la sección de detector es regresar las señales de FI a la información de la fuente original. El detector se suele llamar *detector de audio*, o *segundo detector* en receptores de banda de emisión, porque las señales de información tienen frecuencias de audio. El detector puede ser tan sencillo como un solo diodo, o tan complejo como un lazo de fase cerrada o un demodulador balanceado.

Sección de amplificador de audio. La sección de audio abarca varios amplificadores de audio en cascada, y una o más bocinas o altoparlantes. La cantidad de amplificadores que se usen depende de la potencia deseada en la señal de audio.

Funcionamiento del receptor. En un receptor superheterodino, durante el proceso de demodulación, las señales recibidas pasan por dos o más traslaciones de frecuencia: primero, la RF se convierte a FI, después la FI se convierte a la información original. Los términos RF y FI dependen del sistema, y con frecuencia son engañosos porque no necesariamente indican un intervalo específico de frecuencias. Por ejemplo, la RF en la banda de emisión comercial en AM tiene frecuencias entre 535 y 1605 kHz, y las señales de FI tienen frecuencias de 450 a 460 kHz. En los receptores de banda comercial de FM se usan frecuencias intermedias hasta de 10.7 MHz, bastante mayores que las señales de RF de banda de emisión. Frecuencias intermedias sólo se refieren a las que se usan dentro de un transmisor o receptor, que están entre las radiofrecuencias y las frecuencias de la información original.

Conversión de frecuencias. La *conversión de frecuencia* en la etapa de mezclador/convertidor es idéntica a la que se hace en la etapa moduladora de un transmisor, pero en el receptor las frecuencias tienen conversión descendente, y no ascendente como en el transmisor. En el mezclador/convertidor, las señales de RF se combinan con la frecuencia del oscilador local,

en un dispositivo no lineal. La salida del mezclador contiene una cantidad infinita de frecuencias armónicas y de producto cruzado, que incluyen las frecuencias de suma y diferencia entre la portadora deseada de RF y la del oscilador local. Los filtros de FI se sintonizan a las frecuencias de diferencia. El oscilador local se diseña de tal manera que su frecuencia de oscilación siempre sea mayor o menor que la de la portadora deseada de RF, una cantidad igual a la frecuencia central de FI. Por lo anterior, la diferencia entre la frecuencia de RF y la del oscilador local siempre es igual a la FI. El ajuste de la frecuencia central del preselector, y el ajuste de la frecuencia del oscilador local se *sintonizan amarradas*. Sintonía amarrada quiere decir que los dos ajustes se enlazan mecánicamente de tal manera que con un solo ajuste se cambie la frecuencia central del preselector y, al mismo tiempo, se cambie la frecuencia del oscilador local. Cuando la frecuencia del oscilador local se sintoniza arriba de la RF, se llama *inyección lateral alta*, o *inyección de oscilación superior*. Cuando el oscilador local se sintoniza abajo de la RF, se llama *inyección lateral inferior*, o *inyección de oscilación inferior*. En los receptores de banda de AM, siempre se usa la inyección lateral superior, por una razón que explicaremos más adelante en esta sección. La ecuación de la frecuencia del oscilador local es

Para inyección lateral superior: $$f_{lo} = f_{RF} + f_{FI} \tag{4-7a}$$

Para inyección lateral inferior: $$f_{lo} = f_{RF} - f_{FI} \tag{4-7b}$$

donde f_{lo} = frecuencia del oscilador local (hertz)
f_{RF} = radiofrecuencia (hertz)
f_{FI} = frecuencia intermedia (hertz)

Ejemplo 4-3

Para un receptor superheterodino de AM que usa inyección lateral superior y tiene una frecuencia de oscilador local de 1355 kHz, calcular la portadora de FI, la frecuencia de lado superior y la de lado inferior, para una onda de RF formada por una portadora y frecuencias laterales superior e inferior (fls y fli) de 900 kHz, 905 kHz y 895 kHz, respectivamente.

Solución Véase la fig. 4-5. Como se usa inyección lateral superior, las frecuencias intermedias son la diferencia entre las radiofrecuencias y la frecuencia del oscilador local. Al reordenar la ecuación 4-7a se obtiene

$$f_{FI} = f_{lo} - f_{RF} = 1355 \text{ kHz} - 900 \text{ kHz} = 455 \text{ kHz}$$

Las frecuencias intermedias superior e inferior son

$$f_{FI(fls)} = f_{lo} - f_{RF(fli)} = 1355 \text{ kHz} - 895 \text{ kHz} = 460 \text{ kHz}$$
$$f_{FI(fli)} = f_{lo} - f_{RF(fls)} = 1355 \text{ kHz} - 905 \text{ kHz} = 450 \text{ kHz}$$

Nótese que las frecuencias laterales sufren una inversión de banda lateral durante el proceso de heterodinado, es decir, que la frecuencia RF de banda superior se traslada a una frecuencia FI de lado inferior, y la frecuencia de RF de lado inferior se traslada a una frecuencia FI de lado superior. A esto se le suele llama *inversión de banda lateral*. La inversión de banda lateral no es perjudicial en la AM convencional de doble banda lateral, porque ambas bandas laterales contienen exactamente la misma información.

Rastreo de oscilador local. El *rastreo* es la capacidad del oscilador local de un receptor, de oscilar arriba o abajo de la portadora de radiofrecuencia seleccionada, una cantidad igual a la frecuencia intermedia en toda la banda de radiofrecuencias. Con inyección lateral superior, el oscilador debe rastrear arriba de la portadora de RF de entrada, una frecuencia fija igual a $f_{RF} + f_{FI}$, y con inyección lateral inferior, el oscilador local debe rastrear abajo de la portadora de RF, una frecuencia fija igual a $f_{RF} - f_{FI}$.

La fig. 4-6a muestra el esquema del circuito preselector y oscilador local sintonizado, en un receptor de banda de emisión en AM. Las líneas interrumpidas que unen a los dos capacitores de sintonía indican que están *amarradas* entre sí, es decir, conectadas con un solo control de sintonía. El circuito sintonizado en el preselector es sintonizable desde una frecuencia central

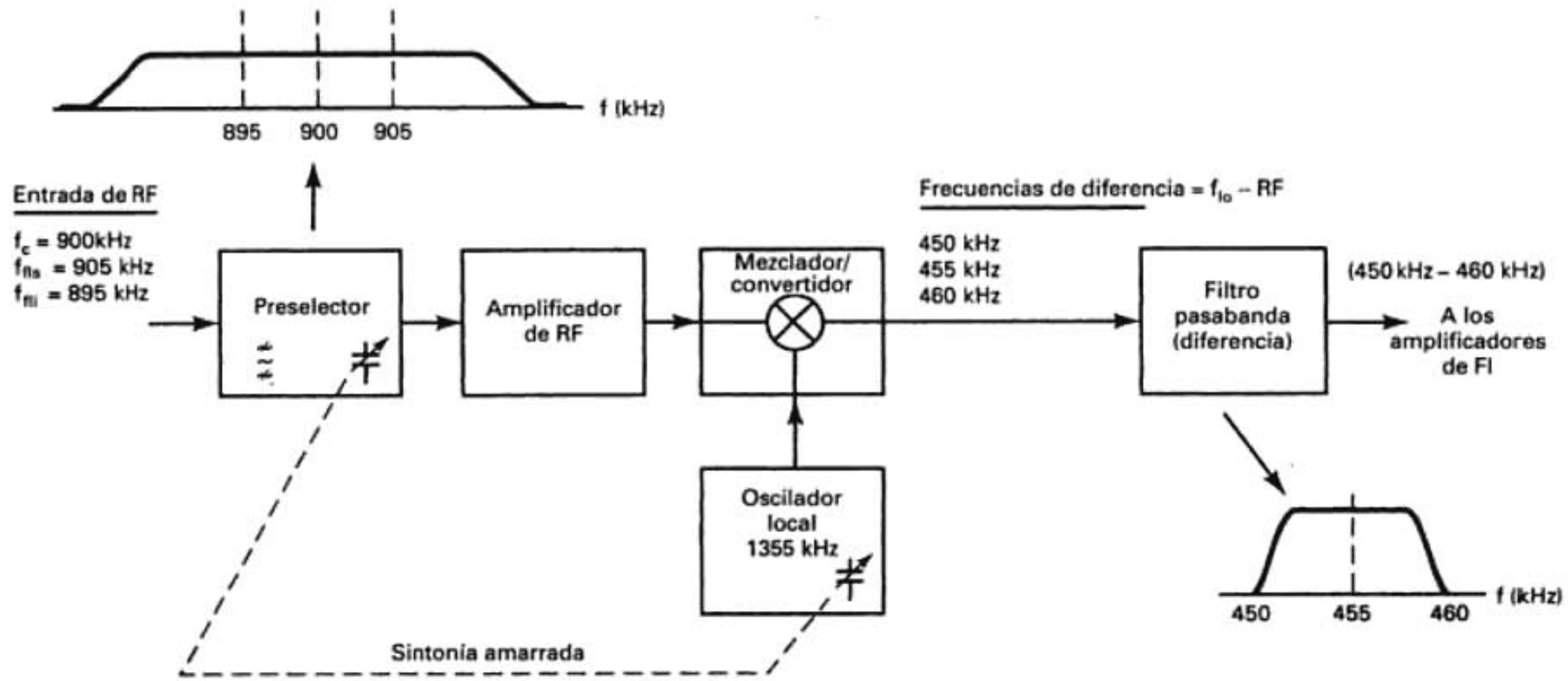


FIGURA 4-5 Figura para el ejemplo 4-3

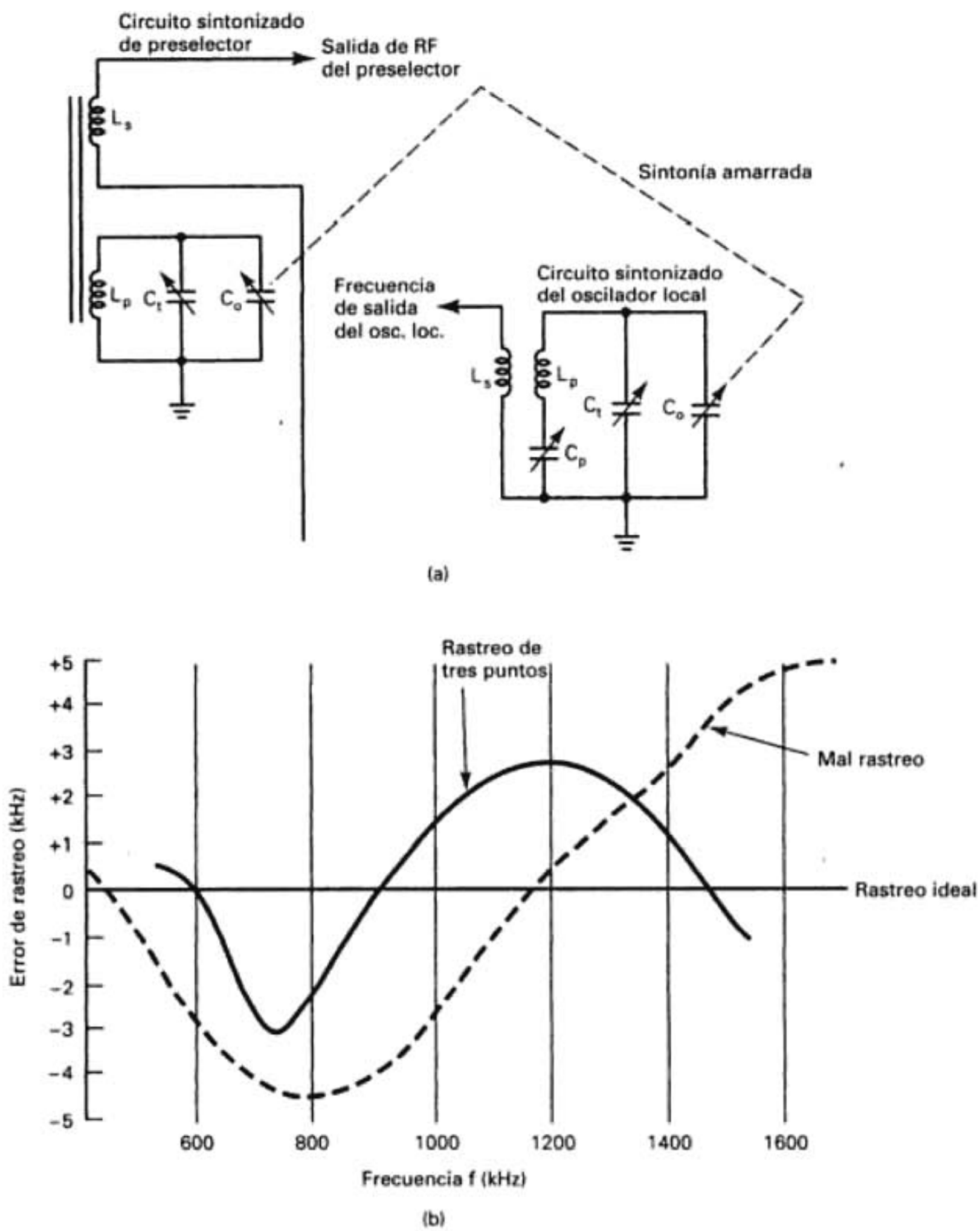


FIGURA 4-6 Rastreo de receptor: (a) esquema de preselector y oscilador local; (b) curvas de rastreo

de 540 kHz hasta 1600 kHz (una relación de 2.96 a 1), y el oscilador local se puede sintonizar desde 995 kHz hasta 2055 kHz (una relación de 2.06 a 1). Como la frecuencia de resonancia de un circuito sintonizado es inversamente proporcional a la raíz cuadrada de la capacitancia, la capacitancia en el circuito preselector debe cambiar en un factor de 8.8, mientras que al mismo tiempo, la del oscilador local debe cambiar en un factor de sólo 4.26. El oscilador local debe oscilar 455 kHz arriba de la frecuencia central del preselector, en toda la banda de frecuencias de AM, y debe haber un solo control de sintonía. Es difícil, pero no imposible, fabricar ese circuito. Por consiguiente, es improbable que se tenga un rastreo perfecto sobre toda la banda de AM. La diferencia entre la frecuencia real del oscilador local y la frecuencia deseada se llama *error de rastreo*. Normalmente ese error no es uniforme dentro de todo el espectro de RF. Más o menos lo mejor que se puede esperar en un receptor doméstico de banda de emisión de AM es

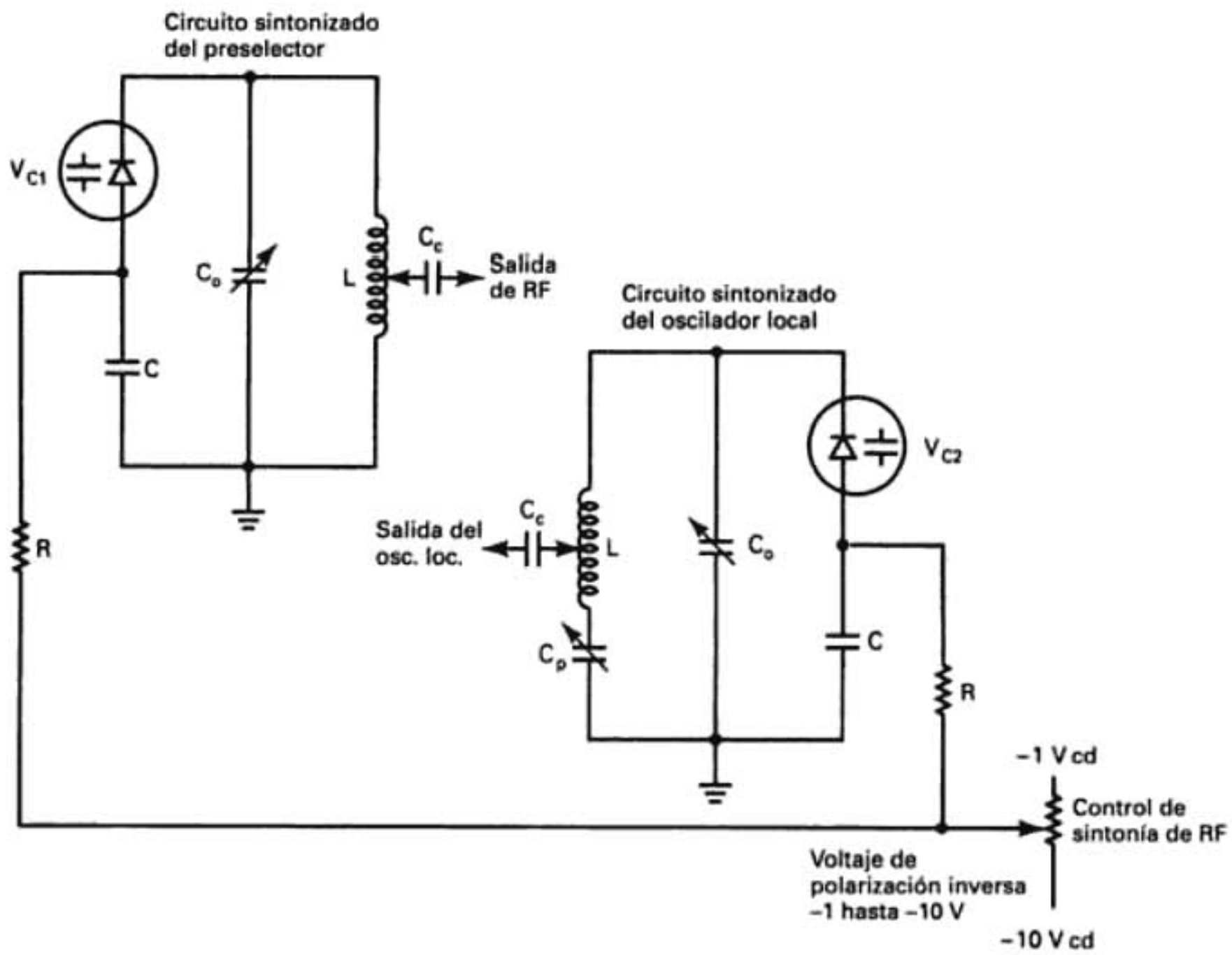


FIGURA 4-7 Sintonía electrónica

un error de rastreo de ±3 kHz, cuando la frecuencia intermedia es de 455 kHz. La fig. 4-6b muestra una curva típica de rastreo. Un error de rastreo de +3 kHz corresponde a una frecuencia FI central de 458 kHz, y uno de −3 kHz, a una de 452 kHz.

El error de rastreo se reduce con una técnica llamada *rastreo de tres puntos*. El preselector y el oscilador local tienen un capacitor de ajuste (C_t), cada uno, en paralelo con el capacitor primario de sintonía (C_o), que compensa los errores pequeños de rastreo en el extremo alto del espectro de AM. El oscilador local tiene un capacitor de compensación (C_p) en serie con la bobina de sintonía, que compensa los errores pequeños de rastreo en el extremo bajo del espectro de AM. Con rastreo de tres puntos, el error de rastreo se ajusta a 0 Hz más o menos en 600 kHz, 950 kHz y 1500 kHz.

Con inyección lateral inferior, el oscilador local debería poder sintonizarse desde 85 kHz hasta 1145 kHz (una relación de 13.5 a 1). En consecuencia, la capacitancia debe cambiar en un factor de 182. Los capacitores variables normales casi nunca tienen un intervalo mayor que de 10 a 1. Ésta es la causa por la que la inyección lateral inferior no es práctica para los receptores de banda comercial de AM. Con la inyección lateral superior, el oscilador local debe ser sintonizable desde 995 kHz hasta 2055 kHz, que corresponde a una relación de capacitancia sólo de 4.63 a 1.

Los capacitores amarrados son relativamente grandes, costosos e inexactos, y son algo difíciles de compensar. En consecuencia, se están reemplazando por circuitos de estado sólido, sintonizados electrónicamente. Estos circuitos son menores, menos costosos, más exactos y relativamente inmunes a cambios del ambiente, se compensan con mayor facilidad y se adaptan más fácilmente a la sintonía digital por control remoto y por botones que sus contrapartes mecánicas. Como en el caso de los módulos de oscilador de cristal que se describieron en el capítulo 2, los circuitos sintonizados electrónicamente usan diodos de estado sólido y de capacitancia variable (diodos varactor). La fig. 4-7 muestra un diagrama de un preselector y oscilador local sintonizado electrónicamente. El voltaje de polarización inversa de −1 V a −10 V proviene de un solo control de sintonía. Al cambiar de posición el selector de un resistor variable de

precisión, cambia la polarización inversa para los dos diodos de sintonía, V_{C1} y V_{C2}. La capacitancia del diodo y, por consiguiente, la frecuencia de resonancia del circuito sintonizado varían de acuerdo con la polarización inversa. Se logra una compensación de tres puntos en la sintonía electrónica, igual que en la sintonía mecánica.

En un receptor superheterodino, la mayor parte de la selectividad del receptor se logra en la etapa de FI. Para tener una reducción máxima de ruido, el ancho de banda de los filtros de FI es igual al ancho mínimo de banda necesario para pasar la señal de información, que con la transmisión de doble banda lateral es igual a dos veces la frecuencia máxima de la señal moduladora. Para una frecuencia máxima de señal moduladora de 5 kHz, el ancho de banda mínimo de FI con rastreo perfecto es 10 kHz. Para una frecuencia central de FI de 455 kHz, es necesaria una banda de paso de 450 kHz a 460 kHz. Sin embargo, en realidad algunas portadoras de RF se rastrean hasta ±3 kHz, arriba o abajo de 455 kHz. Por consiguiente, se debe ampliar el ancho de banda de RF para permitir que pasen las señales de FI de las estaciones fuera del rastreo a través de los filtros de FI.

Ejemplo 4-4

Para la curva de rastreo de la fig. 4-8a, determinar el ancho mínimo de banda de FI con 455 kHz de frecuencia FI central y una frecuencia máxima de señal moduladora de 5 kHz.

Solución Una señal de AM de doble banda lateral con frecuencia máxima de señal moduladora de 5 kHz necesitaría un ancho de banda de 10 kHz. Por consiguiente, un receptor de 455 kHz de frecuencia central FI intermedia y rastreo ideal produciría señales entre 450 kHz y 460 kHz. La curva de rastreo de la fig. 4-8a es para un receptor con rastreo perfecto sólo para frecuencias de portadora de RF de 600 kHz y 1000 kHz. La banda de paso ideal de FI se ve en la fig. 4-8b.

La curva de rastreo de la fig. 4-8a también muestra que el error máximo positivo de rastreo es –3 kHz, para una frecuencia de portadora de RF de 800 kHz, y que el error máximo negativo de

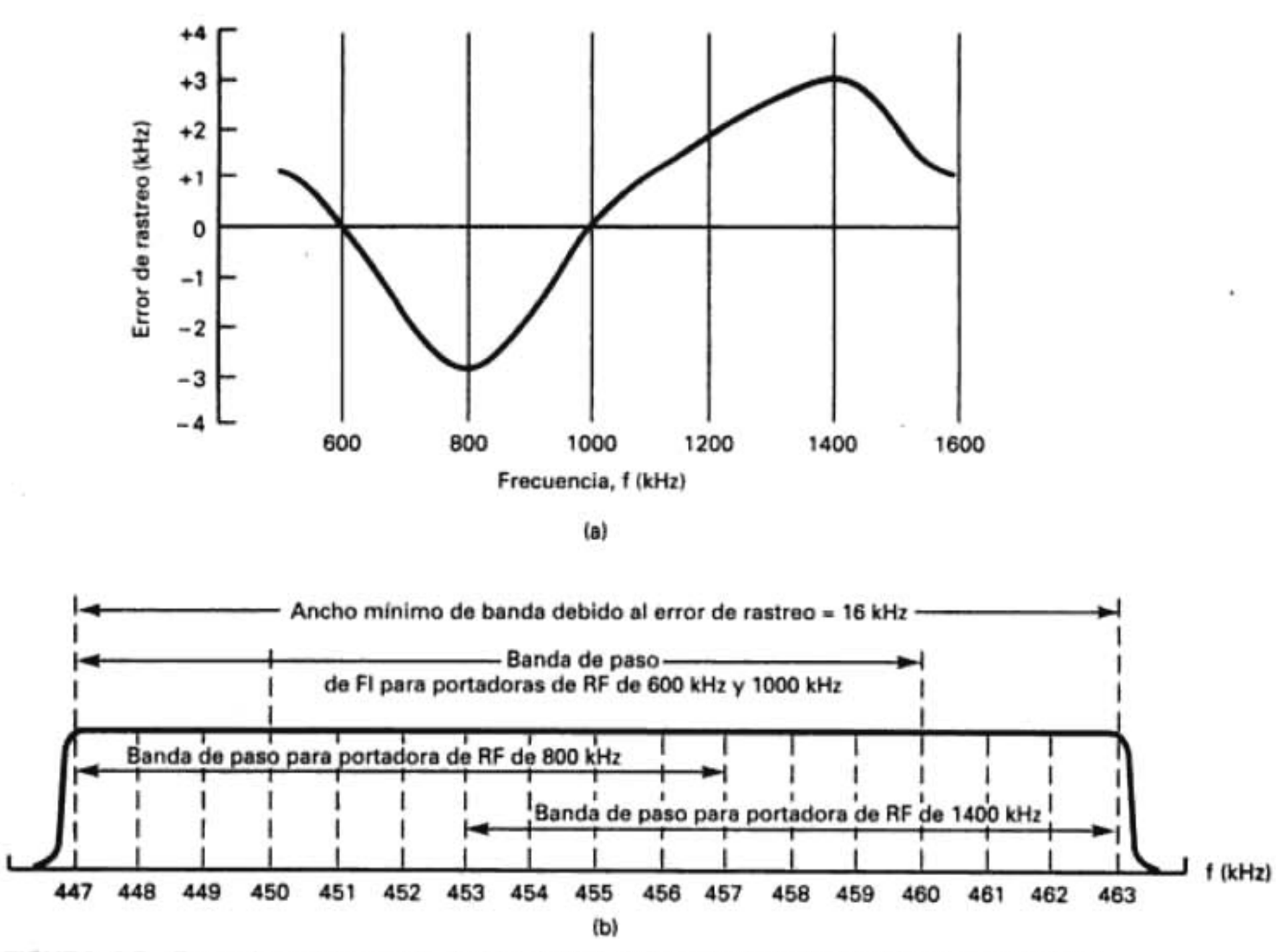


FIGURA 4-8 Error de rastreo para el ejemplo 4-4: (a) curva de rastreo; (b) características de la banda de paso

rastreo es +3 kHz, para una frecuencia de portadora de RF de 1400 kHz. En consecuencia, como se ve en la fig. 4-8b, el espectro de frecuencias producido por una portadora de 800 kHz va de 447 kHz a 457 kHz, y el espectro de radiofrecuencias de FI producido con una portadora de 1400 kHz iría de 453 a 463 kHz.

Por lo anterior, la frecuencia intermedia máxima se tiene cuando la portadora de RF tiene el error de rastreo más positivo (1400 kHz) y con una señal moduladora de 5 kHz.

$$f_{\text{FI(máx)}} = f_{\text{FI}} + \text{error de rastreo} + f_{m\text{(máx)}}$$
$$= 455 \text{ kHz} + 3 \text{ kHz} + 5 \text{ kHz} = 463 \text{ kHz}$$

La frecuencia intermedia mínima se tiene cuando la portadora de RF tiene el error más negativo de rastreo (a 800 kHz) y con una señal moduladora de 5 kHz.

$$f_{\text{FI(mín)}} = f_{\text{FI}} + \text{error de rastreo} - f_{m\text{(máx)}}$$
$$= 455 \text{ kHz} + (-3 \text{ kHz}) - 5 \text{ kHz} = 447 \text{ kHz}$$

El ancho mínimo de banda de FI necesario para pasar las dos bandas laterales es la diferencia entre las frecuencias intermedias máxima y mínima, es decir

$$B_{\text{mín}} = 463 \text{ kHz} - 447 \text{ kHz} = 16 \text{ kHz}$$

La fig. 4-8b muestra las características de la banda pasante de FI, para el ejemplo 4-4.

Frecuencia imagen. Una *frecuencia imagen* es una frecuencia distinta a la frecuencia de la portadora seleccionada que, si se deja entrar a un receptor y mezclarse con el oscilador local, produce una frecuencia de producto cruzado que es igual a la frecuencia intermedia. Una frecuencia imagen equivale a una segunda radiofrecuencia que producirá una FI que interfiere con la FI debida a la radiofrecuencia deseada. Una vez que se ha mezclado una frecuencia imagen hasta la FI, no se puede filtrar ni suprimir. Si la portadora de RF seleccionada y su frecuencia imagen entran al mismo tiempo a un receptor, se mezclan con la frecuencia del oscilador local, y producen frecuencias de diferencia que son iguales a la FI. En consecuencia, se reciben y demodulan dos estaciones diferentes al mismo tiempo, y producen dos conjuntos de información. Para que una radiofrecuencia produzca un producto cruzado igual a la FI, debe estar desplazada respecto a la frecuencia del oscilador local un valor igual al de la FI. Con inyección lateral superior, la RF seleccionada está abajo de la del oscilador local una frecuencia igual a la FI. Por lo anterior, la frecuencia imagen es la radiofrecuencia que está ubicada en la FI arriba del oscilador local. La ecuación de la frecuencia imagen (f_{im}), para la inyección lateral superior, es

$$f_{\text{im}} = f_{\text{lo}} + f_{\text{FI}} \tag{4-8a}$$

y como la RF deseada es igual a la frecuencia del oscilador local menos la FI,

$$f_{\text{im}} = f_{\text{RF}} + 2f_{\text{FI}} \tag{4-8b}$$

La fig. 4-9 muestra el espectro de frecuencias relativas de RF, FI, oscilador local e imagen, para un receptor superheterodino con inyección lateral superior. Se ve que mientras mayor sea la FI, la frecuencia de imagen está más alejada, en el espectro de frecuencias, de la RF deseada. Por lo anterior, para un mejor *rechazo de frecuencia imagen* se prefiere una frecuencia intermedia alta. Sin embargo, mientras más alta sea la FI, es más difícil fabricar amplificadores

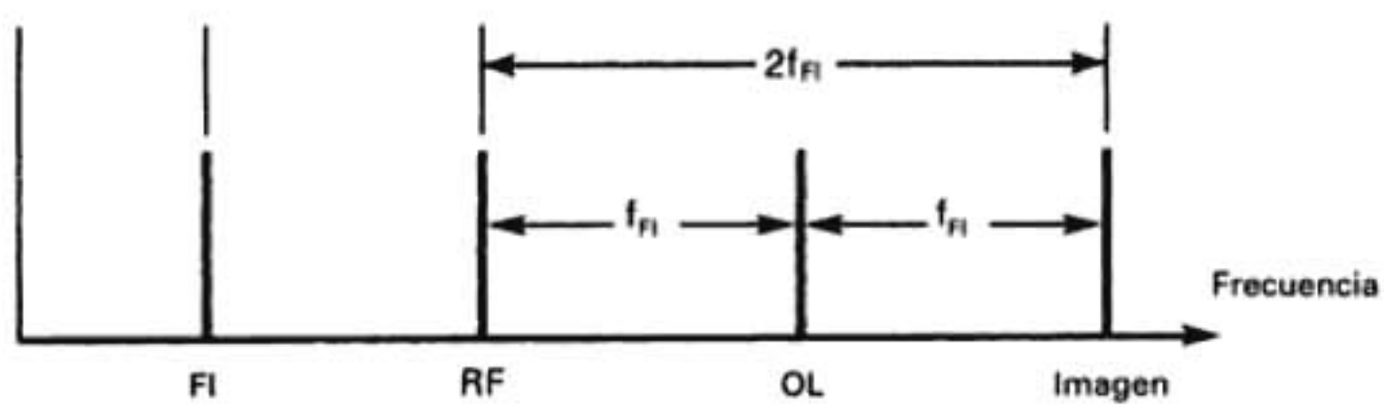


FIGURA 4-9 Frecuencia imagen

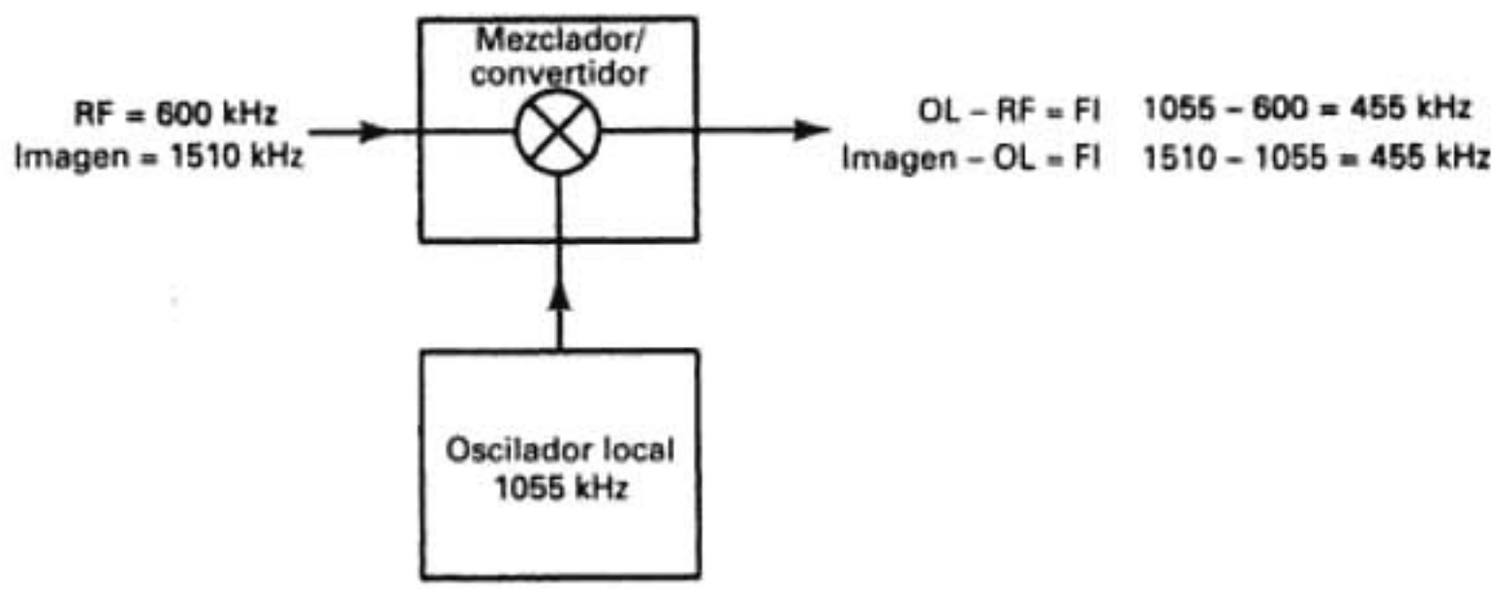


FIGURA 4-10 Conversión de frecuencia para el ejemplo 4-5

estables con alta ganancia. Por consiguiente, hay un compromiso de selección de la FI para un radiorreceptor, entre el rechazo de la frecuencia imagen y la ganancia y estabilidad de la FI.

Relación de rechazo de frecuencia imagen. La *relación de rechazo de frecuencia imagen* (IFRR, por *image-frequency rejection ratio*) es una medida numérica de la capacidad de un preselector para rechazar la frecuencia de imagen. Para un preselector de sintonía única, la relación de su ganancia en la RF deseada, entre la ganancia a la frecuencia imagen, es la IFRR. La definición de la IFRR es

$$\text{IFRR} = \sqrt{(1 + Q^2\rho^2)} \tag{4-9a}$$

en donde $\rho = (f_{im}/f_{RF}) - (f_{RF}/f_{im})$.

$$\text{IFRR}_{(dB)} = 10 \log \text{IFRR} \tag{4-9b}$$

Si hay más de un circuito sintonizado en el frente de un receptor (quizá un filtro preselector y un amplificador RF sintonizado por separado), la IFRR es simplemente el producto de las dos relaciones.

Ejemplo 4-5

Para un receptor superheterodino de banda comercial de AM con frecuencias FI, RF y de oscilador local de 455 kHz, 600 kHz y 1055 kHz respectivamente, determine lo siguiente consultando la fig. 4-10:
(a) La frecuencia imagen.
(b) La IFRR para una Q del preselector de 100.

Solución (a) Según la ecuación 4-8a,

$$f_{im} = 1055 \text{ kHz} + 455 \text{ kHz} = 1510 \text{ kHz}$$

o de acuerdo con la ecuación 4-8b.

$$f_{im} = 600 \text{ kHz} + 2(455 \text{ kHz}) = 1510 \text{ kHz}$$

(b) De las ecuaciones 4-9a y 4-9b,

$$\rho = \frac{1510 \text{ kHz}}{600 \text{ kHz}} - \frac{600 \text{ kHz}}{1510 \text{ kHz}} = 2.51 - 0.397 = 2.113$$

$$\text{IFRR} = \sqrt{1 + (100^2)(2.113^2)} = 212.15 \text{ o } 23.25 \text{ dB}$$

Una vez que una frecuencia imagen tuvo conversión descendente hasta la FI, no se puede eliminar. Por consiguiente, para rechazar la frecuencia imagen se debe bloquear antes de la etapa de mezclador/convertidor. El rechazo de la frecuencia imagen es el objetivo principal del preselector de RF. Si el ancho de banda del preselector es suficientemente bajo, se evita que la frecuencia imagen entre al receptor. La fig. 4-11 ilustra cómo un buen filtrado de FI y RF puede evitar que una frecuencia imagen interfiera con la radiofrecuencia deseada.

La relación de la RF a la FI es una consideración importante para rechazar la frecuencia imagen. Cuanto más cerca esté la RF a la FI, más cercana estará la RF a la frecuencia imagen.

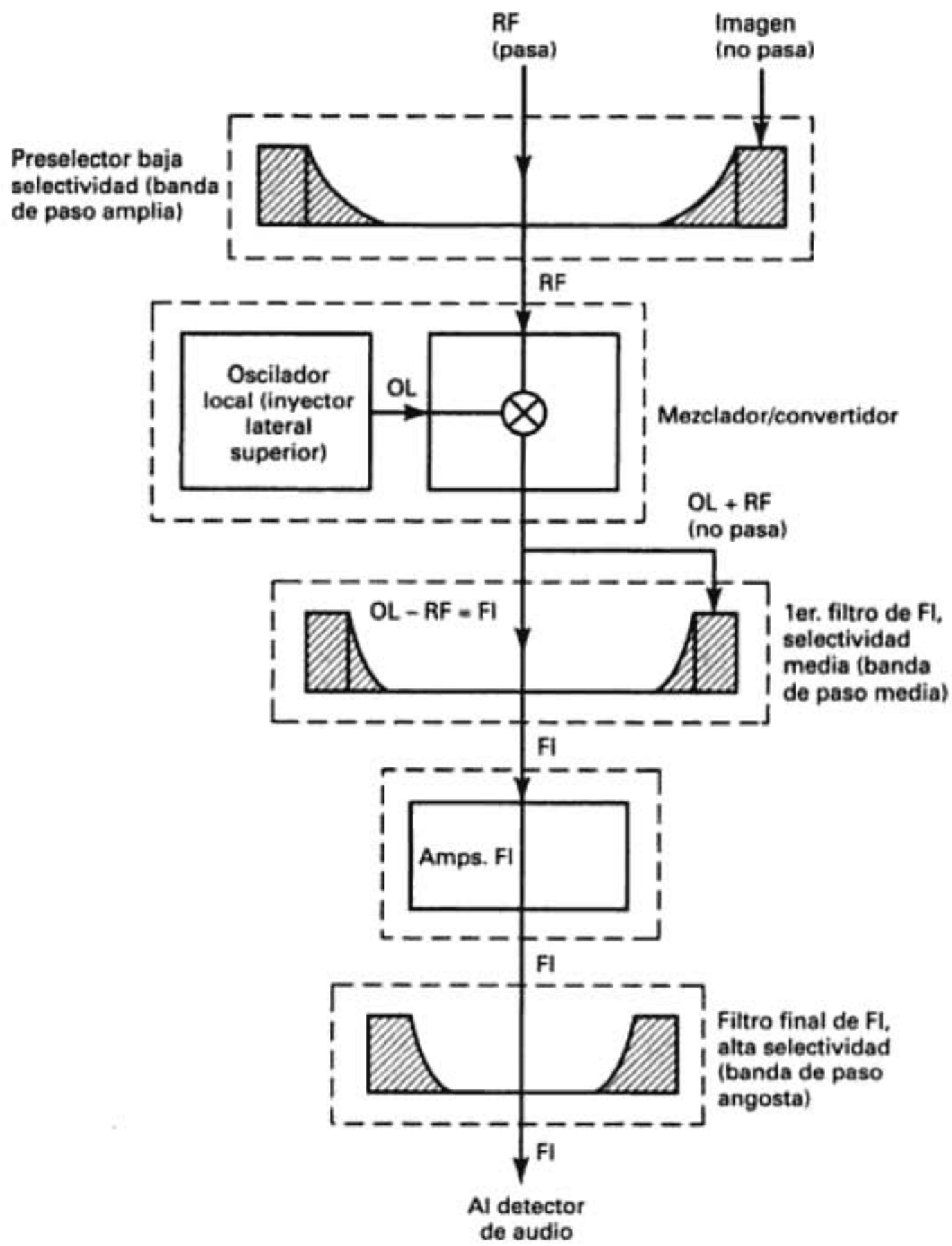


FIGURA 4-11 Rechazo de la frecuencia imagen

Ejemplo 4-6

Para un receptor de banda civil con inyección lateral superior y una portadora de RF de 27 MHz y frecuencia central de FI de 455 kHz, determinar:

(a) La frecuencia del oscilador local.
(b) La frecuencia imagen.
(c) La IFRR para una Q del selector igual a 100.
(d) La Q del preselector necesaria para lograr la misma IFRR que la alcanzada con una portadora de RF de 600 kHz, del ejemplo 4-5.

Solución (a) Según la ecuación 4-7a,

$$f_{lo} = 27 \text{ MHz} + 455 \text{ kHz} = 27.455 \text{ MHz}$$

(b) De la ecuación 4-8a,

$$f_{im} = 27.455 \text{ MHz} + 455 \text{ kHz} = 27.91 \text{ MHz}$$

(c) De acuerdo con las ecuaciones 4-9a y 4-9b,

$$\text{IFRR} = 6.7 \text{ o sea } 16.5 \text{ dB}$$

(d) Reordenando la ecuación 4-9a,

$$Q = \frac{\sqrt{(\text{IFRR}^2 - 1)}}{\rho} = 3536$$

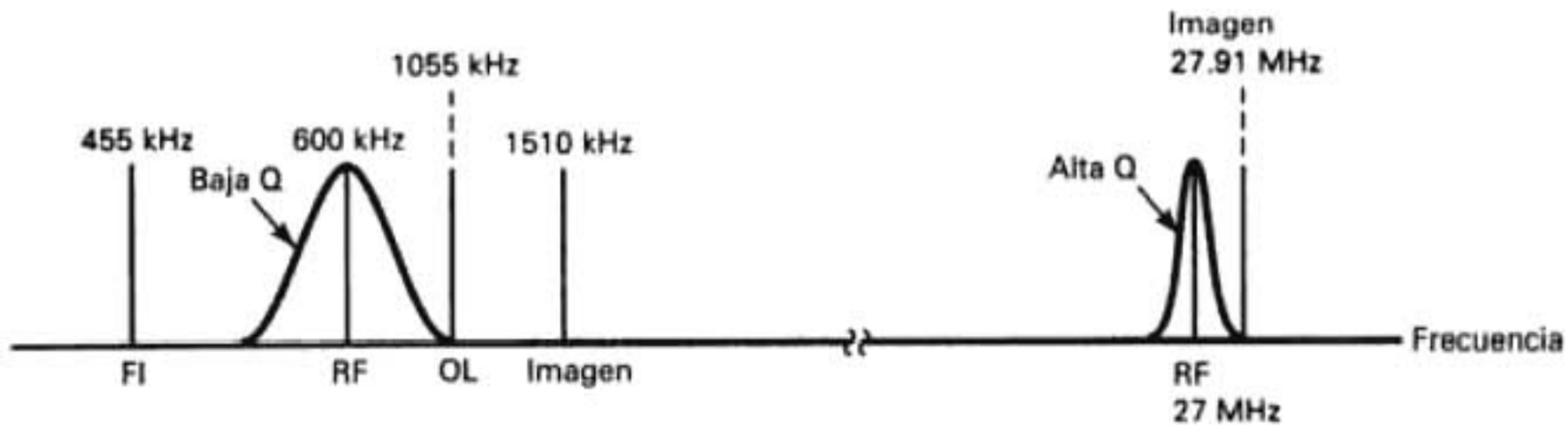


FIGURA 4-12 Espectro de frecuencias para el ejemplo 4-6

En los ejemplos 4-5 y 4-6 se ve que mientras mayor sea la frecuencia de la portadora de RF, es más difícil evitar que la frecuencia imagen entre al receptor. Para la misma IFRR, las portadoras de RF de mayor frecuencia requieren un filtro preselector de mucho mayor calidad. Esto se ve en la fig. 4-12.

Se ve, en los ejemplos 4-5 y 4-6, que para una FI de 455 kHz es más difícil evitar que la frecuencia imagen entre al receptor con altas frecuencias de portadora de RF que con bajas frecuencias de portadora. Para la misma IFRR, las portadoras de mayor RF requieren un preselector de mucho mejor calidad; de hecho, se necesitan en muchos casos valores irreales o inalcanzables de Q. Una solución más sencilla del problema es usar mayores frecuencias FI cuando se reciben portadoras de mayores frecuencias RF. Por ejemplo, si se usara una FI de 5 MHz para el receptor de banda civil del ejemplo 4-6, la frecuencia imagen sería 37 MHz, con suficiente alejamiento de la portadora de 27 MHz como para que un preselector con Q realista evite con facilidad que entre la frecuencia de imagen al receptor.

Doble sintonía. Hay *doble sintonía* cuando un receptor recoge la misma estación en dos puntos cercanos de su escala de sintonía. Un punto está en el lugar deseado, y el otro es el llamado *punto espurio*. La doble sintonía se debe a mala selectividad en el frente o a un rechazo inadecuado de la frecuencia imagen.

Es perjudicial la doble sintonía, porque las estaciones débiles pueden quedar ocultas por la recepción de una estación fuerte cercana en el punto espurio del espectro de frecuencias. Se puede usar la doble sintonía para determinar la frecuencia intermedia de un receptor desconocido, porque el punto espurio en la escala está exactamente al doble de la frecuencia FI central, por abajo de la frecuencia correcta de recepción.

CIRCUITOS RECEPTORES DE AM

Circuitos amplificadores de RF

Un amplificador de RF es de alta ganancia, bajo ruido y sintonizado que, cuando se usa, es la primera etapa activa que encuentra la señal recibida. Los objetivos primarios de una etapa de RF son selectividad, amplificación y sensibilidad. Por lo anterior, los amplificadores de RF deben tener:

1. Bajo ruido térmico.
2. Baja cifra de ruido.
3. Ganancia de moderada a alta.
4. Baja distorsión por intermodulación y armónica (es decir, tener operación lineal).
5. Selectividad moderada.
6. Alta relación de rechazo de frecuencia imagen.

Dos de los parámetros más importantes de un receptor son la amplificación y la cifra de ruido que dependen, ambos, de la etapa de RF. Un demodulador (o detector, como se le dice a veces) de AM detecta variaciones de amplitud en la onda modulada, y las convierte en cambios de amplitud de su señal de salida. Por consiguiente, las variaciones de amplitud que fueron

causadas por el ruido se convierten en fluctuaciones erróneas en la salida del demodulador, y se degrada la calidad de la salida de éste. Mientras más ganancia tenga una señal al pasar por un receptor, las variaciones de amplitud son más pronunciadas en la entrada del demodulador, y las variaciones causadas por el ruido son menos notables. Mientras más angosta sea la banda, el ruido que se propaga por el receptor es menor, y en consecuencia, el detector demodula menos ruido. De acuerdo con la ecuación 1-28 ($V_N = \sqrt{4RKTB}$), el voltaje de ruido es directamente proporcional a la raíz cuadrada de la temperatura, el ancho de banda y la resistencia equivalente de ruido. Entonces, si se minimizan estos tres parámetros, el ruido térmico se reduce. Se puede reducir la temperatura de una etapa de RF enfriando, artificialmente, el frente del receptor con ventiladores de aire o hasta con helio líquido en los receptores más costosos. El ancho de banda se reduce mediante amplificadores y filtros, y la resistencia equivalente de ruido se reduce con componentes de fabricación especial en los dispositivos activos. La cifra de ruido es esencialmente una medida del ruido que agrega un amplificador. Por consiguiente, la cifra de ruido mejora (se reduce) reduciendo el ruido interno del amplificador.

La intermodulación y la distorsión armónica son formas de distorsión no lineal que aumentan la magnitud de la cifra de ruido al sumar ruido correlacionado al espectro total de ruido. Mientras más lineal es el funcionamiento de un amplificador, se produce menos distorsión no lineal y la cifra de ruido del receptor es mejor. La reducción de la frecuencia imagen por el amplificador de RF se combina con la del preselector, para reducir el ancho de banda de entrada del receptor lo suficiente para ayudar a evitar que la frecuencia imagen entre a la etapa de mezclador/convertidor. En consecuencia, todo lo que se requiere en la etapa de RF es una selectividad moderada.

La fig. 4-13 muestra varios circuitos amplificadores de RF de uso frecuente. Téngase en cuenta que RF es un término relativo, y que tan sólo indica que la frecuencia es suficientemente alta como para ser irradiada con eficacia desde una antena y ser propagada por el espacio libre en forma de una onda electromagnética. La RF para la banda de emisión en AM está entre 535 kHz y 1605 kHz, mientras que para la radio de microondas es mayor que 1 GHz (1000 MHz). Una frecuencia intermedia de uso común en receptores de FM comercial es de 10.7 MHz, bastante mayor que las radiofrecuencias asociadas con la banda de emisión de AM. RF no es más que la frecuencia de la señal irradiada o recibida, y FI es la de una señal intermedia, dentro de un transmisor o de un receptor. Por lo anterior, muchas de las consideraciones para los amplificadores de RF también se aplican a los de FI, como por ejemplo, la neutralización, el filtrado y el acoplamiento.

La fig. 4-13a muestra el esquema de un amplificador de RF con transistor bipolar. Los elementos C_a, C_b, C_c y L_1 forman el circuito de acoplamiento de la antena. Q_1 está polarizado en clase A para reducir la distorsión no lineal. El circuito del colector se acopla con el mezclador/convertidor a través del transformador T_1, con doble sintonía para tener más selectividad. C_x y C_y son capacitores de derivación de RF. Sus símbolos indican que están fabricados especialmente como capacitores *de paso*. Estos capacitores tienen menos inductancia, con lo que se evita que una parte de la señal sea irradiada por sus terminales. El capacitor C_n es de *neutralización*. Una parte de la señal del colector se retroalimenta al circuito de la base, para compensar (o neutralizar) la señal que se regresó por la capacitancia de adelanto entre el colector y la base para evitar que haya oscilaciones. Las capacitancias C_x y C_n forman un divisor de voltaje de ca, para la señal de retroalimentación. Esta configuración de neutralización se llama neutralización *respecto a tierra*.

La fig. 4-13b muestra un amplificador de RF con transistores de efecto de campo de doble compuerta. Esta configuración usa varios FET de metal-óxido semiconductor con aumento de deserción (DEMOS FETs, de *depletion-enhancement metal-oxide semiconductor FETs*). Los FET tienen alta impedancia de entrada y bajo ruido. Un FET es un dispositivo de ley cuadrática que sólo genera componentes de distorsión por segunda armónica y por intermodulación, y en consecuencia produce menos distorsión no lineal que un transistor bipolar. Q_1 vuelve a estar polarizado en clase A para tener funcionamiento lineal. El transformador T_1 tiene sintonía única a la frecuencia deseada de portadora de RF, para aumentar la selectividad del receptor y mejorar el IFRR. El inductor L_5 es una bobina de radiofrecuencia, y junto con C_5, desacopla las señales de RF del suministro de corriente directa.

La fig. 4-13c muestra el esquema de configuración de un amplificador especial de RF, llamado amplificador *cascode*. Un amplificador cascode tiene alta ganancia y menos ruido que los amplificadores convencionales en cascada. Los dispositivos activos pueden ser transistores bipolares

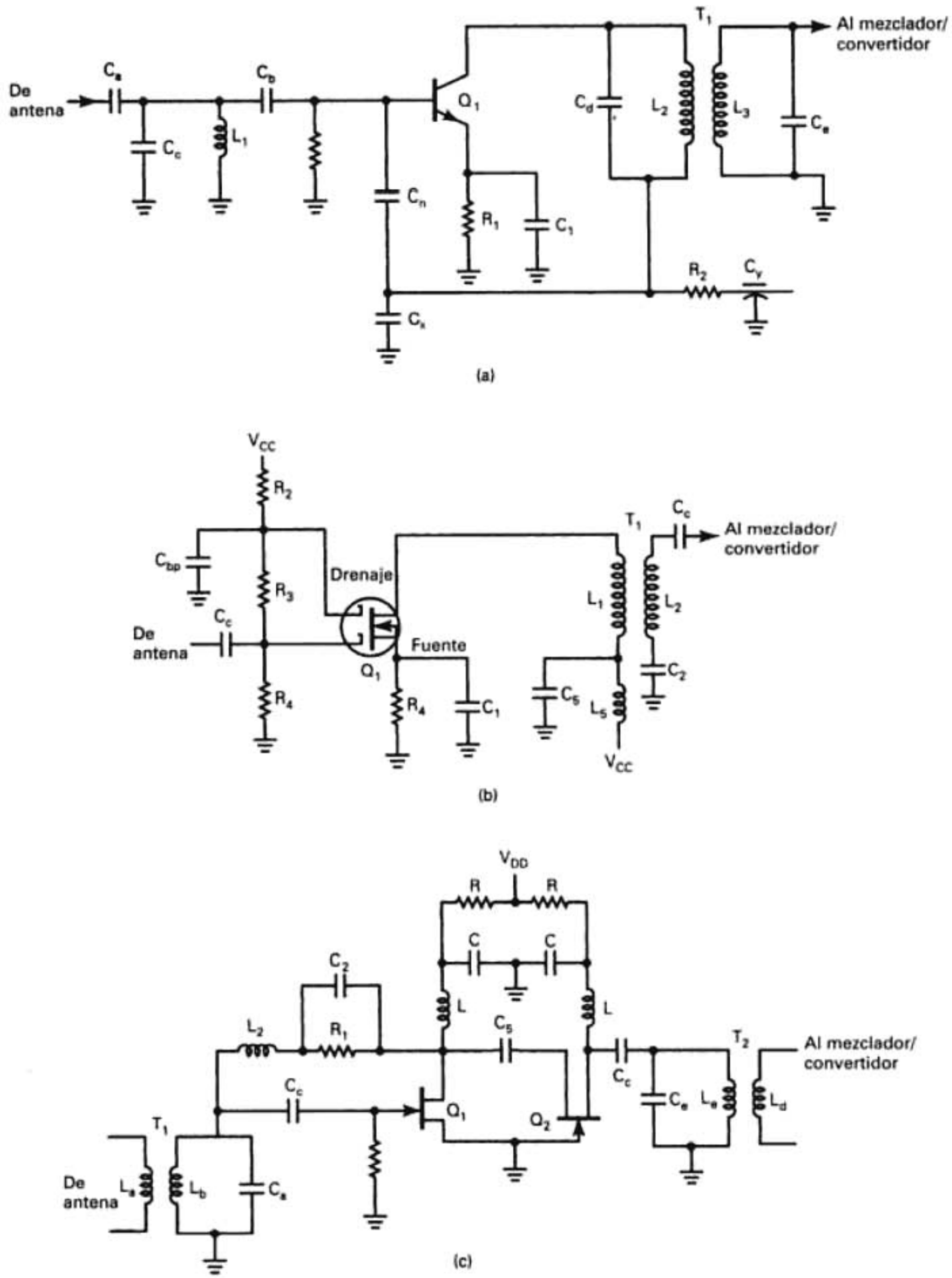


FIGURA 4-13 Configuraciones de amplificador de RF: (a) amplificador de transistor bipolar; (b) amplificador de FET DEMOS; (c) amplificador cascode

o FET. El amplificador Q_1 es de fuente común, cuya salida se acopla con la fuente de Q_2 a través de una impedancia. Debido a la baja impedancia de entrada de Q_2, Q_1 no necesita ser neutralizado; sin embargo, la neutralización reduce todavía más la cifra de ruido. En consecuencia, L_2, R_1 y C_2 constituyen la trayectoria de retroalimentación para la neutralización. El amplificador Q_2 es de compuerta común que, debido a su baja impedancia de entrada, no requiere neutralización.

Amplificadores de bajo ruido

Los receptores de microondas de alto rendimiento requieren un *amplificador de bajo ruido* (LNA, por *low-noise amplifier*) como etapa de entrada a la sección de RF, para optimizar su cifra de ruido. La ecuación 1-38 indicaba que el primer amplificador en un receptor es el más importante para determinar la cifra de ruido del receptor. La primera etapa debe tener bajo ruido y alta ganancia. Desafortunadamente es difícil lograr esto con una sola etapa de amplificación, y en consecuencia, los amplificadores de bajo ruido suelen tener dos etapas de amplificación, junto con redes de compensación de impedancia, para aumentar su eficiencia. La primera etapa tiene ganancia moderada y ruido mínimo, y la segunda etapa tiene alta ganancia y ruido moderado.

Los amplificadores de RF de bajo ruido se polarizan en clase A, y suelen usar transistores bipolares o de efecto de campo hasta unos 2 GHz, y FET de arseniuro de galio arriba de esta frecuencia. Un tipo especial de FET de arseniuro de galio, muy utilizado, es el MESFET (FET de unión semiconductor metal). Un MESFET es un FET con unión metal-semiconductor en la compuerta del dispositivo, llamada barrera de Schottky. Los amplificadores de bajo ruido se describen con detalle en un capítulo posterior de este libro.

Amplificadores de RF en circuito integrado. El NE/SA5200 (fig. 4-14) es un amplificador de RF en circuito integrado lineal de banda ancha, incondicionalmente estable, de baja potencia y de ganancia dual, fabricado por Signetics Corporation. Trabaja desde cd hasta 1200 MHz, y tiene baja cifra de ruido. Tiene varias ventajas inherentes respecto a realizaciones con partes discretas: no necesita componentes externos de polarización, ocupa poco espacio en una tarjeta de circuito impreso, y su alto nivel de integración mejora su confiabilidad en comparación con sus contrapartes discretas. El NE/SA5200 también tiene un modo inactivo que ayuda a reducir el consumo de potencia, cuando se pueden desactivar los amplificadores.

En la fig. 4-14a se muestra el diagrama de bloques del SA5200, y en la fig. 4-14b aparece su diagrama simplificado. Nótese que los dos amplificadores de banda ancha se polarizan con un mismo generador de polarización. Cada etapa amplificadora tiene una cifra de ruido aproximada de 3.6 dB, y una ganancia aproximada de 11 dB. Se pueden conectar en cascada varias etapas de NE/SA5200, y usarse como una trayectoria de FI, y la terminal de habilitación se puede usar para mejorar el margen dinámico del receptor. Para niveles de entrada extremadamente altos, se pueden desactivar los amplificadores del NE/SA5200. Cuando se desactivan, la señal de entrada se atenúa 13 dB y se evita la sobrecarga del receptor.

Circuitos de mezclador/convertidor

El objeto de la etapa de mezclador/convertidor es la conversión descendente de las radiofrecuencias que llegan hasta las frecuencias intermedias. Esto se logra mezclando las señales de RF con la frecuencia del oscilador local, en un dispositivo no lineal. En esencia eso es el heterodinaje. Un mezclador es un amplificador no lineal, parecido a un modulador, pero la salida se sintoniza con la diferencia entre la RF y las frecuencias del oscilador local. La fig. 4-15 muestra un diagrama de bloques de una etapa de mezclador/convertidor. La salida de un mezclador balanceado es el producto de las frecuencias de RF y del oscilador local, y se expresa matemáticamente como sigue

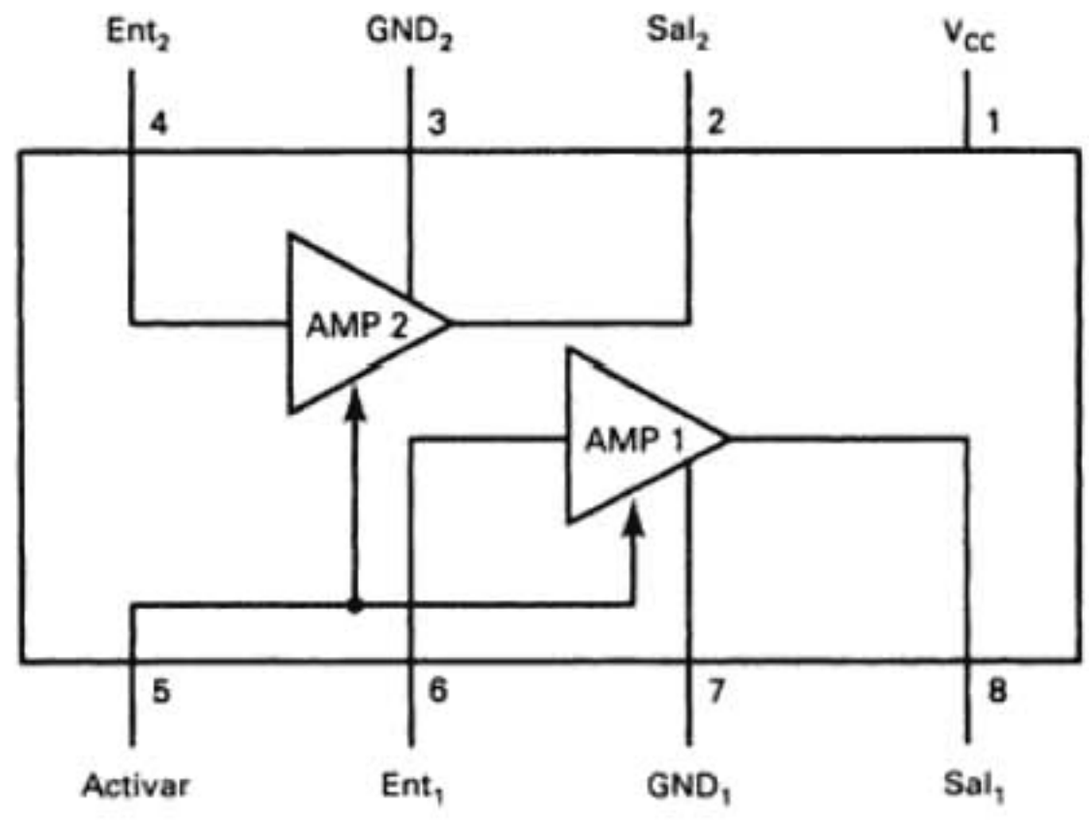
$$V_{sal} = (\text{sen } 2\pi f_{RF}t)(\text{sen } 2\pi f_{lo}t)$$

siendo f_{RF} = radiofrecuencia que llega (hertz)
f_{lo} = frecuencia del oscilador local (hertz)

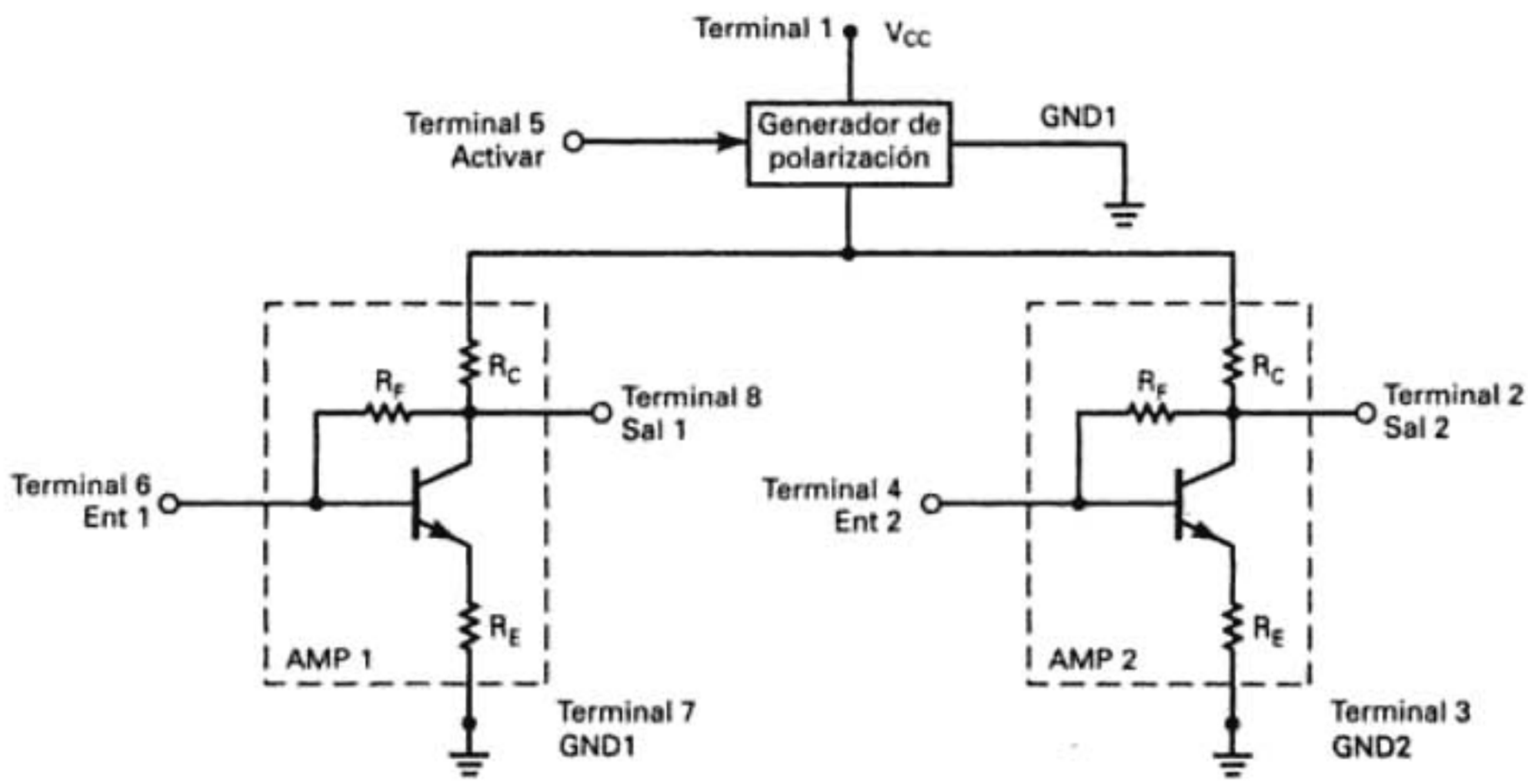
Así, al aplicar la identidad trigonométrica del producto de dos senos, la salida de un mezclador es

$$V_{sal} = \frac{1}{2}\cos[2\pi(f_{RF} - f_{lo})t] - \frac{1}{2}\cos[2\pi(f_{RF} + f_{lo})t]$$

El valor absoluto de la diferencia de frecuencias, $|f_{RF} - f_{lo}|$ es la frecuencia intermedia.



(a)



(b)

FIGURA 4-14 Etapa de ganancia doble de RF en un circuito NE/SA5200: (a) diagrama de bloques; (b) diagrama simplificado

Aunque se puede usar cualquier dispositivo no lineal como mezclador, se prefiere en general un transistor o un FET respecto a un diodo simple porque también es capaz de amplificación. Sin embargo, como la salida real de un mezclador es una frecuencia de producto cruzado, hay una pérdida neta de la señal. Ésta es la llamada *pérdida por conversión* (que a veces se le llama *ganancia de conversión*), porque se ha hecho una conversión de frecuencias y, al mismo tiempo, la señal de salida de FI tiene menor amplitud que la señal de entrada de RF. En general, la pérdida por conversión es 6 dB, que corresponde a una ganancia de conversión de −6 dB. La ganancia de conversión es la diferencia entre el nivel de salida de FI con una señal de entrada de RF, entre el nivel de salida de FI con una señal de entrada de FI.

La fig. 4-16 muestra los diagramas de varios circuitos comunes de mezclador/convertidor. En la fig. 4-16a se ve lo que posiblemente sea el circuito mezclador más sencillo disponible (además de un mezclador de un solo diodo), que se usa exclusivamente para receptores de banda de AM poco costosos. Las señales de radiofrecuencia de la antena se filtran mediante el circuito sintonizado del preselector (L_1 y C_1), y después se acoplan con un transformador a la base de Q_1. El

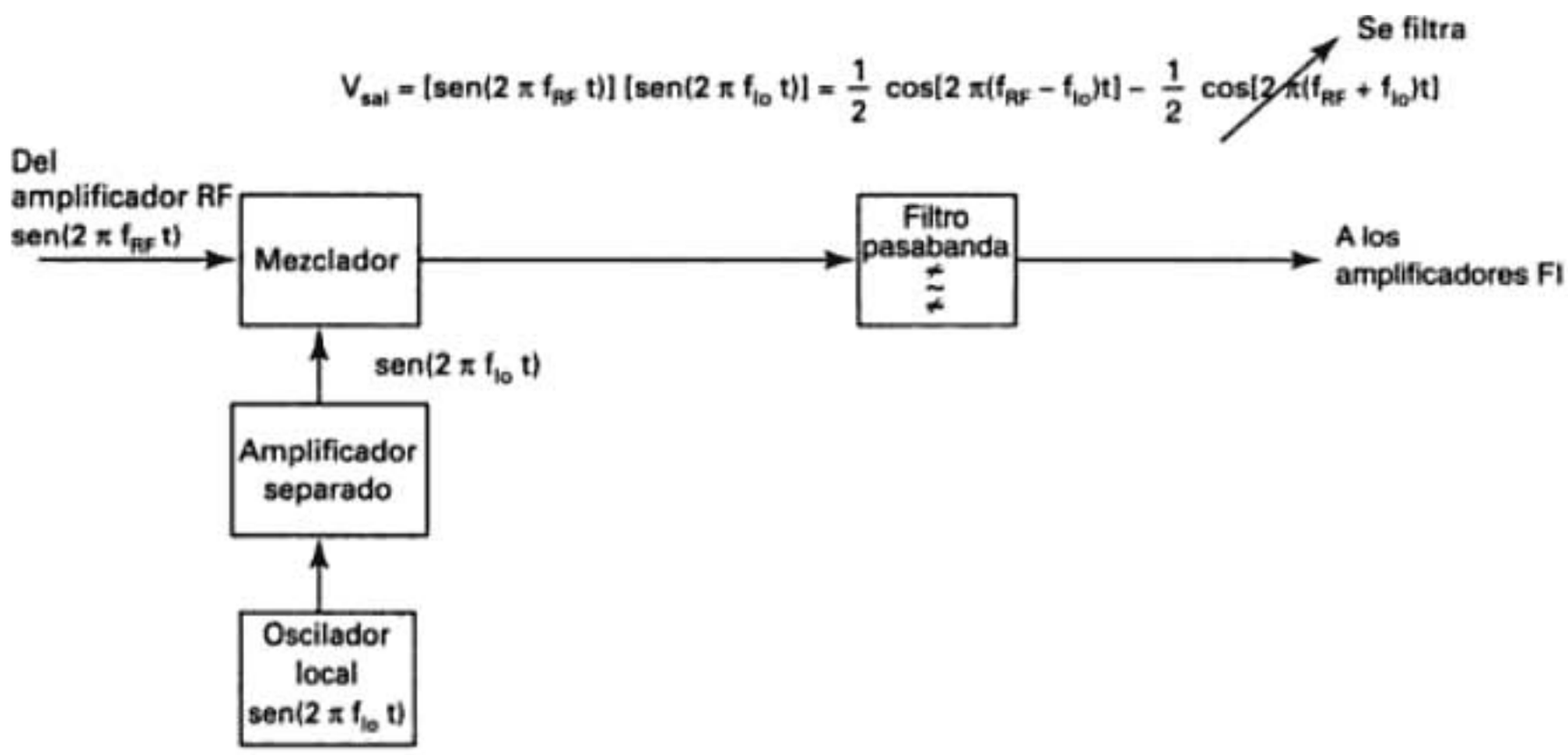


FIGURA 4-15 Diagrama de bloques de un mezclador/convertido

dispositivo activo del mezclador (Q_1) también proporciona amplificación para el oscilador local. A esta configuración se le suele llamar de mezclador *autoexcitado*, porque el mezclador se excita a sí mismo retroalimentando energía al oscilador de circuito tanque (C_2 y L_2), para sostener las oscilaciones. Cuando se aplica inicialmente la corriente, Q_1 amplifica tanto las señales de RF que llegan, como cualquier ruido presente, y abastece al oscilador de circuito tanque con suficiente energía para iniciar la acción de oscilador. La frecuencia del oscilador local es la frecuencia de resonancia del circuito tanque. Una parte de la energía de resonancia del circuito tanque se acopla a través de L_2 y L_5 al emisor de Q_1. Esta señal activa a Q_1 en su región de operación no lineal y, en consecuencia, produce frecuencias de suma y diferencia en su colector. La frecuencia de diferencia es la FI. El circuito tanque de salida (L_3 y C_3) se sintoniza con la banda de FI. Por consiguiente, la señal de FI se acopla con transformador a la entrada del primer amplificador de FI. El proceso es regenerativo mientras haya una señal de entrada de RF. Los capacitores de sintonía en los circuitos tanque de RF y del oscilador local se agrupan en un solo control de sintonía. Los capacitores C_p y C_t son para rastreo de tres puntos. La configuración tiene mala selectividad y mal rechazo de frecuencia imagen, porque no hay amplificador sintonizado con la frecuencia de señal de RF y, en consecuencia, la única selectividad de RF está en el preselector. Además, esencialmente no hay ganancia de RF, y las no linealidades del transistor producen componentes armónicos y de intermodulación que pueden caer dentro de la banda de paso de FI.

El circuito de mezclador/convertidor de la fig. 4-16b es un mezclador *con excitación separada*. Su funcionamiento es esencialmente igual al del mezclador autoexcitado, pero el oscilador local y el mezclador tienen sus propios dispositivos de ganancia. El mezclador mismo es un FET, que tiene características no lineales que se adecuan mejor para conversión de FI que las de un transistor bipolar. La retroalimentación es de L_2 a L_3 del transformador en la fuente de Q_1. Este circuito se usa mucho en receptores de alta frecuencia (HF, de *high-frequency*) y de muy alta frecuencia (VHF).

El circuito de mezclador/convertidor de la fig. 4-16c es un mezclador de *diodo único*. El concepto es muy sencillo: las señales de RF y del oscilador local se acoplan en el diodo, que es un dispositivo no lineal. Por consiguiente, se hace un mezclado no lineal, y se producen las frecuencias de suma y de diferencia. El circuito tanque de salida (C_3 y L_3) esta entonado a la diferencia de frecuencia (FI). Un mezclador de diodo único es ineficiente porque tiene una pérdida neta. Sin embargo, se usa el mezclador de diodo en detección de audio en receptores de AM, y para producir la subportadora de audio en los receptores de televisión.

La fig. 4-16d muestra el esquema de un mezclador de *diodo balanceado*. Los mezcladores balanceados son de los circuitos más importantes en los sistemas actuales de comunicaciones. También se les llama *moduladores balanceados*, *moduladores ideales* y *detectores de producto*.

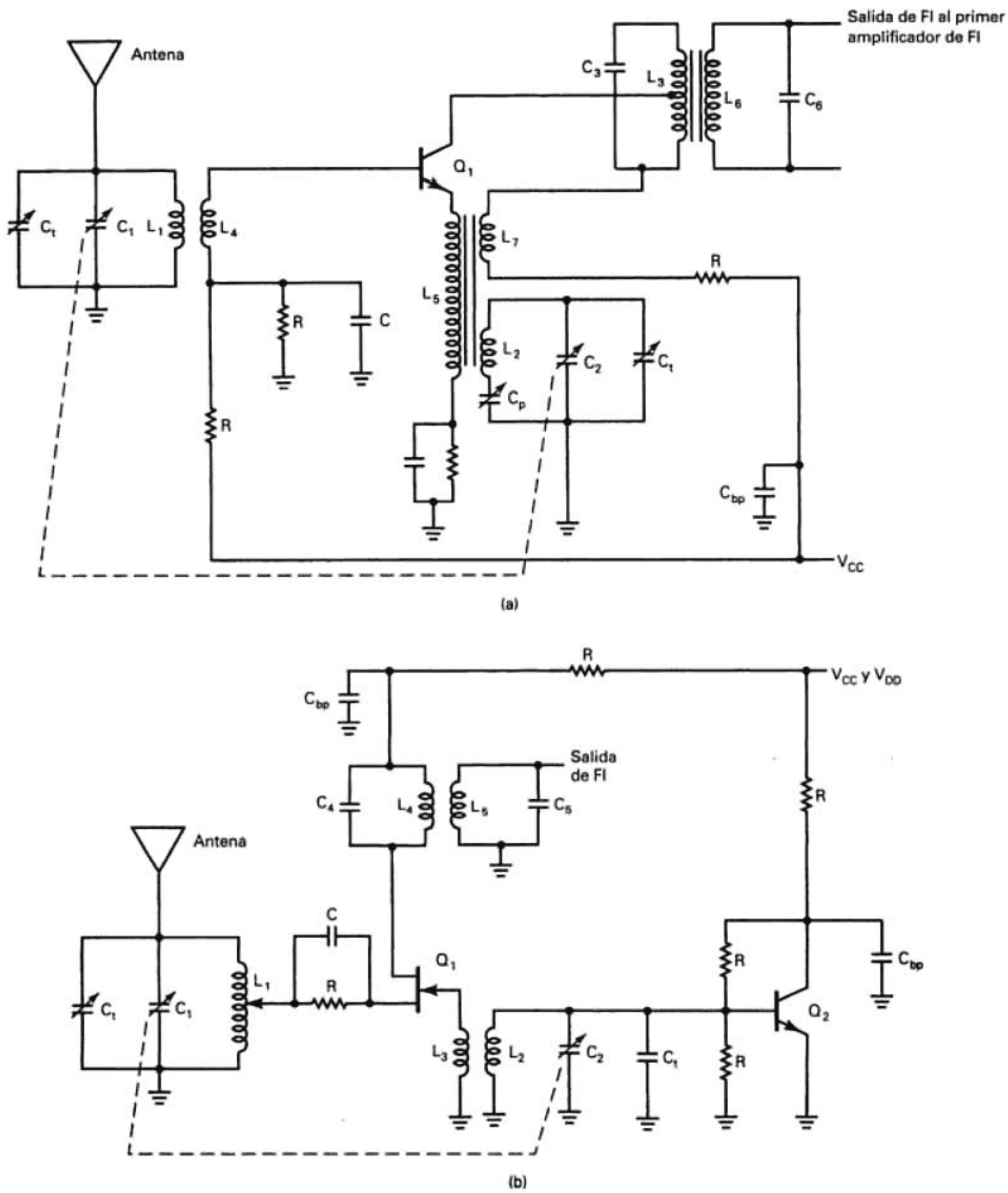


FIGURA 4-16 Circuitos de mezclador/convertidor: (a) mezclador autoexcitado; (b) mezclador excitado en forma independiente; (*continúa*)

Los detectores de fase que se usan en los lazos de fase cerrada, que se explicaron en el capítulo 2, son moduladores balanceados. Los mezcladores balanceados se usan mucho tanto en transmisores como en receptores para AM, FM y muchos de los esquemas de modulación digital, como PSK y QAM. Los mezcladores balanceados tienen dos ventajas inherentes sobre otras clases de mezcladores: reducción de ruido y supresión de portadora.

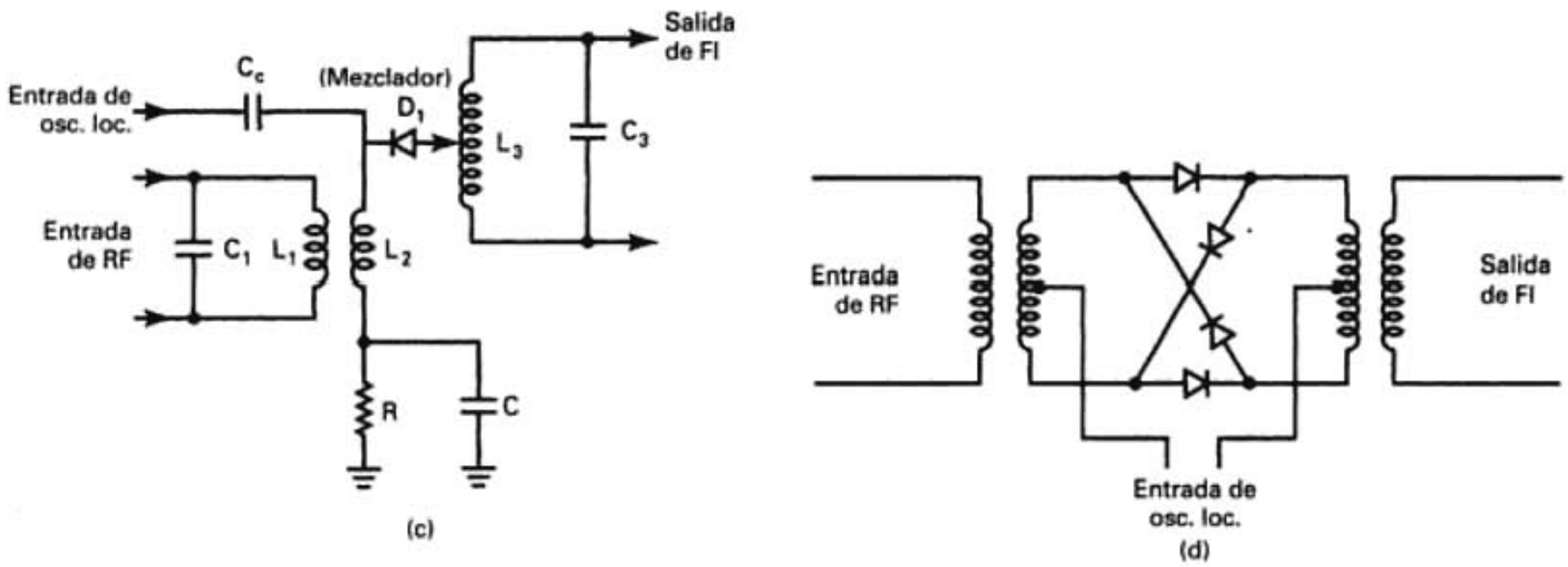


FIGURA 4-16 (*Continuación*) (c) mezclador de diodo; (d) mezclador de diodo balanceado

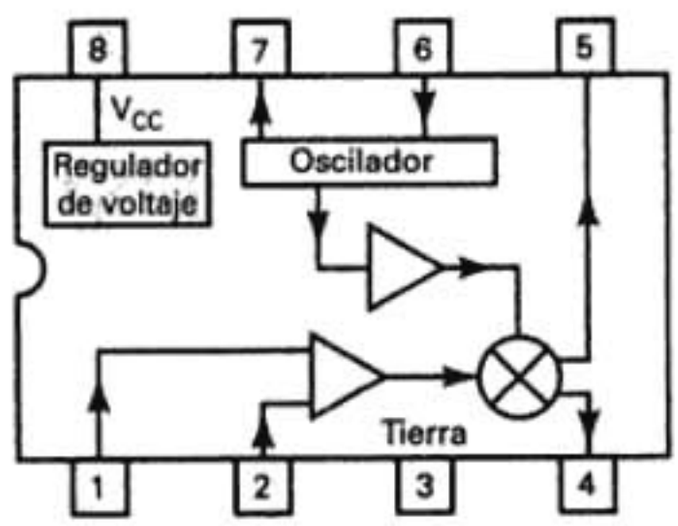


FIGURA 4-17 Diagrama de bloques del circuito NE/SA602A de mezclador doblemente balanceado y oscilador

Mezclador/oscilador en circuito integrado. La fig. 4-17 muestra el diagrama de bloques del *mezclador doblemente balanceado y oscilador* NE/SA602A de Signetics. Es un mezclador monolítico de baja potencia y doblemente balanceado, con amplificador de entrada, oscilador incorporado y regulador de voltaje. Es para usarse en sistemas de comunicaciones de alto rendimiento y baja potencia, y se adapta en especial para aplicaciones de *radio celular*. El mezclador tiene una configuración del tipo celda multiplicadora *de Gilbert*, que suele proporcionar 18 dB de ganancia a 45 MHz. Una celda de Gilbert es un amplificador diferencial que activa a una celda de conmutación balanceada. La etapa diferencial de entrada proporciona la ganancia, y determina el rendimiento de índice de ruido y de manejo de señal del sistema. El oscilador trabaja hasta 200 MHz y se puede configurar como oscilador de cristal, o de circuito tanque LC sintonizado de un amplificador separador, para un oscilador externo. La cifra de ruido del NE/SA602A a 45 MHz suele ser menor de 5 dB. La ganancia, la eficiencia de intersección de tercer orden y las características de baja potencia y bajo ruido hacen del NE/SA602A una buena elección para equipos de alto rendimiento operados con baterías. Las conexiones de entrada, de salida del mezclador de RF y del oscilador respaldan diversas configuraciones de entrada. Las entradas de RF (terminales 1 y 2) están polarizadas internamente y son simétricas. La fig. 4-18 muestra tres configuraciones normales de entrada: entrada asimétrica sintonizada, entrada balanceada y asimétrica no sintonizada.

Circuitos amplificadores de FI

Los amplificadores de frecuencia intermedia (FI) son amplificadores sintonizados, con ganancia relativamente alta, muy parecidos a los amplificadores de RF, pero los de FI trabajan dentro de una banda de frecuencias fija y relativamente angosta. En consecuencia, son fáciles de diseñar y fabricar para que sean estables, no irradien y se neutralicen con facilidad. Como los amplificadores de FI trabajan dentro de una banda fija de frecuencias, se pueden acoplar los amplificadores sucesivos en forma inductiva con circuitos *de doble resonancia* (en los circuitos de doble resonancia, o *doblemente sintonizados*, los lados primario y secundario del transformador son circuitos tanque sintonizados). Por lo anterior, es más fácil alcanzar un factor de forma

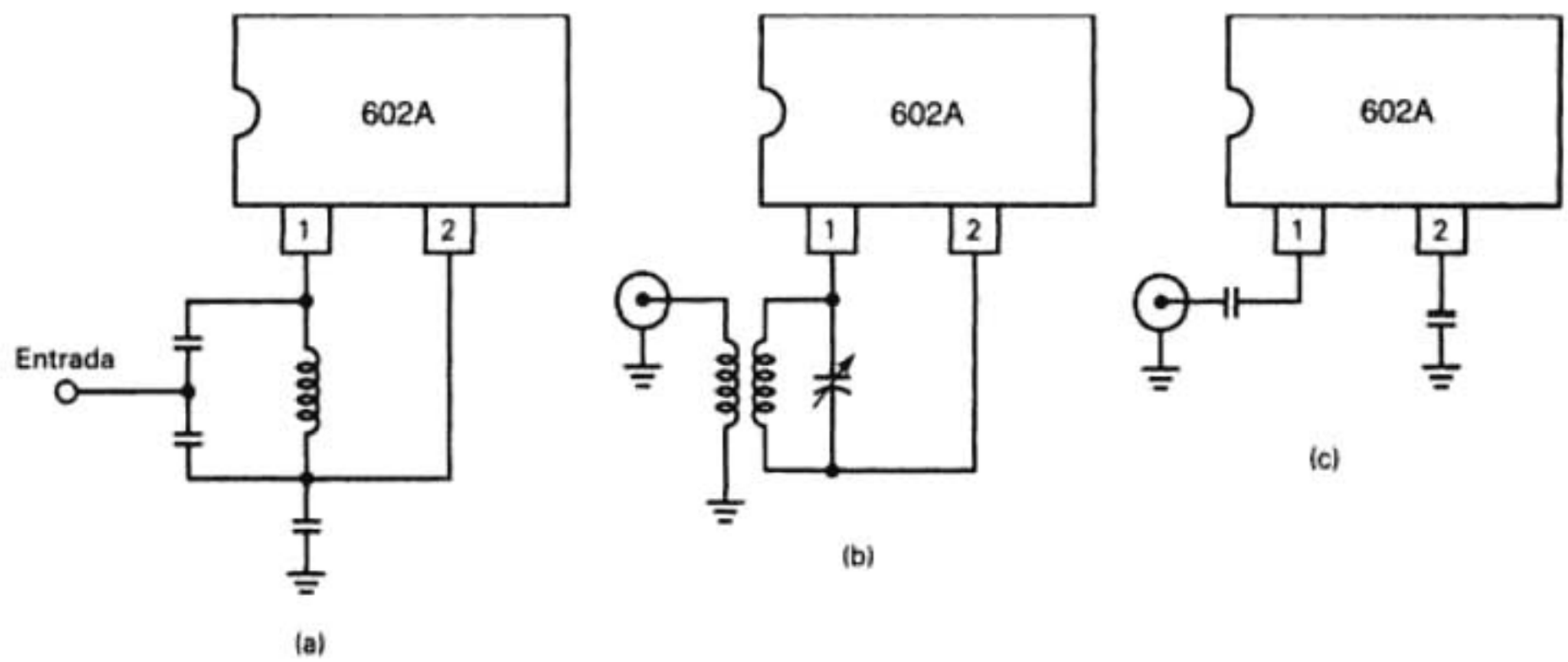


FIGURA 4-18 Circuito NE/SA602A. Configuraciones normales de entrada: (a) entrada asimétrica sintonizada; (b) entrada balanceada; (c) entrada asimétrica no sintonizada

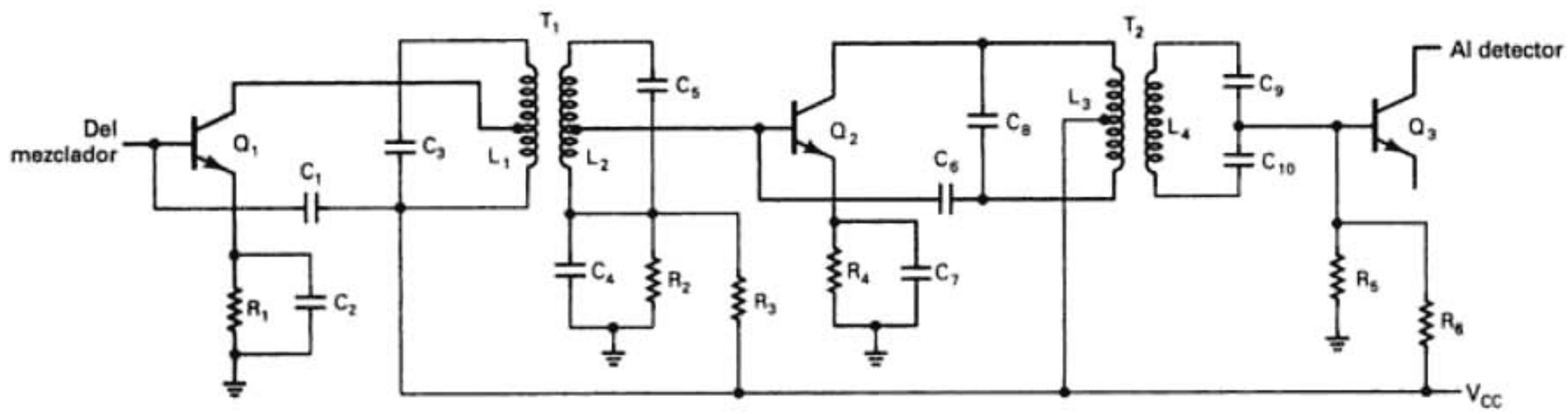


FIGURA 4-19 Sección de FI de tres etapas

óptimo (bajo) y buena selectividad. La mayor parte de la ganancia y la selectividad de un receptor se logran en la sección del amplificador de FI. Una etapa de FI tiene, en general, entre dos y cinco amplificadores de FI. La fig. 4-19 muestra un esquema de una sección de FI de tres etapas. Los transformadores T_1 y T_2 son de doble sintonía y L_1, L_2 y L_3 tienen derivaciones para reducir los efectos de la carga. La base de Q_3 se alimenta del par de capacitores C_9 y C_{10} con derivación, por la misma razón. C_1 y C_6 son capacitores de neutralización.

Acoplamiento inductivo. El *acoplamiento inductivo*, o *por transformador* es la técnica de uso más frecuente para acoplar los amplificadores de FI. En este acoplamiento, el voltaje que se aplica a los devanados del primario de un transformador se transfiere a los devanados secundarios. La proporción de voltaje primario que se acopla al secundario depende de la cantidad de vueltas en los devanados primario y secundario (la relación de vueltas), la cantidad de *flujo magnético* en el devanado primario, el *coeficiente de acoplamiento* y de la rapidez con que cambia el flujo (velocidad angular). La ecuación de la magnitud del voltaje inducido en los devanados secundarios es

$$E_s = \omega M I_p \tag{4-10}$$

en donde E_s = magnitud del voltaje inducido en los devanados secundarios (volts)
ω = velocidad angular de la onda de voltaje primario (radianes por segundo)
M = inductancia mutua (henrys)
I_p = corriente en el primario (amperes)

La propiedad de una bobina de inducir un voltaje dentro de sí misma se llama *autoinductancia*, o simplemente *inductancia*, L. Cuando una bobina induce un voltaje por *inducción magnética* en otra bobina, se dice que las dos bobinas están *acopladas* entre sí. La capacidad de una bobina para inducir un voltaje en otra bobina se llama *inductancia mutua*, M. La inductancia mutua en un transformador se debe a las *líneas de fuerza magnética (el flujo magnético)* que se producen en los devanados primarios y pasan a los secundarios, y es directamente proporcional al coeficiente de acoplamiento. Coeficiente de acoplamiento es la relación del flujo en secundario entre el flujo en primario, y se expresa como sigue

$$k = \frac{\phi_s}{\phi_p} \tag{4-11}$$

en donde k = coeficiente de acoplamiento (adimensional)
ϕ_p = flujo en primario (webers)
ϕ_s = flujo en secundario (webers)

Si todo el flujo producido en el devanado primario pasa al devanado secundario, el coeficiente de acoplamiento es 1. Si nada del flujo primario pasa a los devanados del secundario, el coeficiente de acoplamiento es 0. Es casi imposible alcanzar un coeficiente de acoplamiento de 1, a menos que las dos bobinas se instalen en torno a un núcleo de hierro de alta permeabilidad. En el caso típico, el coeficiente de acoplamiento de los transformadores normales de FI es mucho menor que 1. La transferencia de flujo de los devanados del primario a los del secundario se llama *enlace de flujo*, y es directamente proporcional al coeficiente de acoplamiento. La inductancia mutua de un transformador es directamente proporcional al coeficiente de acoplamiento y a la raíz cuadrada del producto de las inductancias primaria y secundaria, esto es

$$M = k\sqrt{L_s L_p} \tag{4-12}$$

siendo M = inductancia mutua (henrys)
L_s = inductancia del devanado secundario (henrys)
L_p = inductancia del devanado primario (henrys)
k = coeficiente de acoplamiento (adimensional)

Los amplificadores con acoplamiento inductivo se dividen en dos categorías generales: de una y de doble sintonía.

Transformadores sintonizados simples. La fig. 4-20a muestra el esquema de un *amplificador sintonizado simple con acoplamiento inductivo*. Esta configuración se llama de *primario no sintonizado-secundario sintonizado*. El lado primario de T_1 es sólo la inductancia del devanado primario, mientras que hay un capacitor en paralelo con el devanado secundario, y se forma un secundario sintonizado. Los devanados del transformador no tienen tomas, porque el efecto de carga del FET es insignificante. La fig. 4-20b es la curva de respuesta de un transformador con primario no sintonizado y secundario sintonizado. E_s aumenta con la frecuencia hasta que se llega a la frecuencia de resonancia del secundario, f_o; en adelante, E_s comienza a disminuir al aumentar la frecuencia. El máximo de la curva de respuesta en la frecuencia de resonancia se debe a la impedancia reflejada. La impedancia del secundario se regresa al primario, debido a la inductancia mutua entre los dos devanados. Para frecuencias menores que la de resonancia, el aumento de ωM es mayor que la disminución de I_p, y en consecuencia E_s aumenta. Para frecuencias mayores que la de resonancia, el aumento en ωM es menor que la disminución en I_p y por consiguiente disminuye E_s.

La fig. 4-20c muestra el efecto del acoplamiento sobre la curva de respuesta de un transformador de primario no sintonizado y secundario sintonizado. Con *acoplamiento débil* (bajo coeficiente de acoplamiento), el voltaje del secundario es relativamente bajo, y el ancho de banda es reducido. Al aumentar el grado de acoplamiento (aumenta el coeficiente de acoplamiento), el voltaje inducido en el secundario aumenta y el ancho de banda se amplía. Por lo anterior, para tener una selectividad alta, se desea un acoplamiento débil; sin embargo, se sacrifica la amplitud

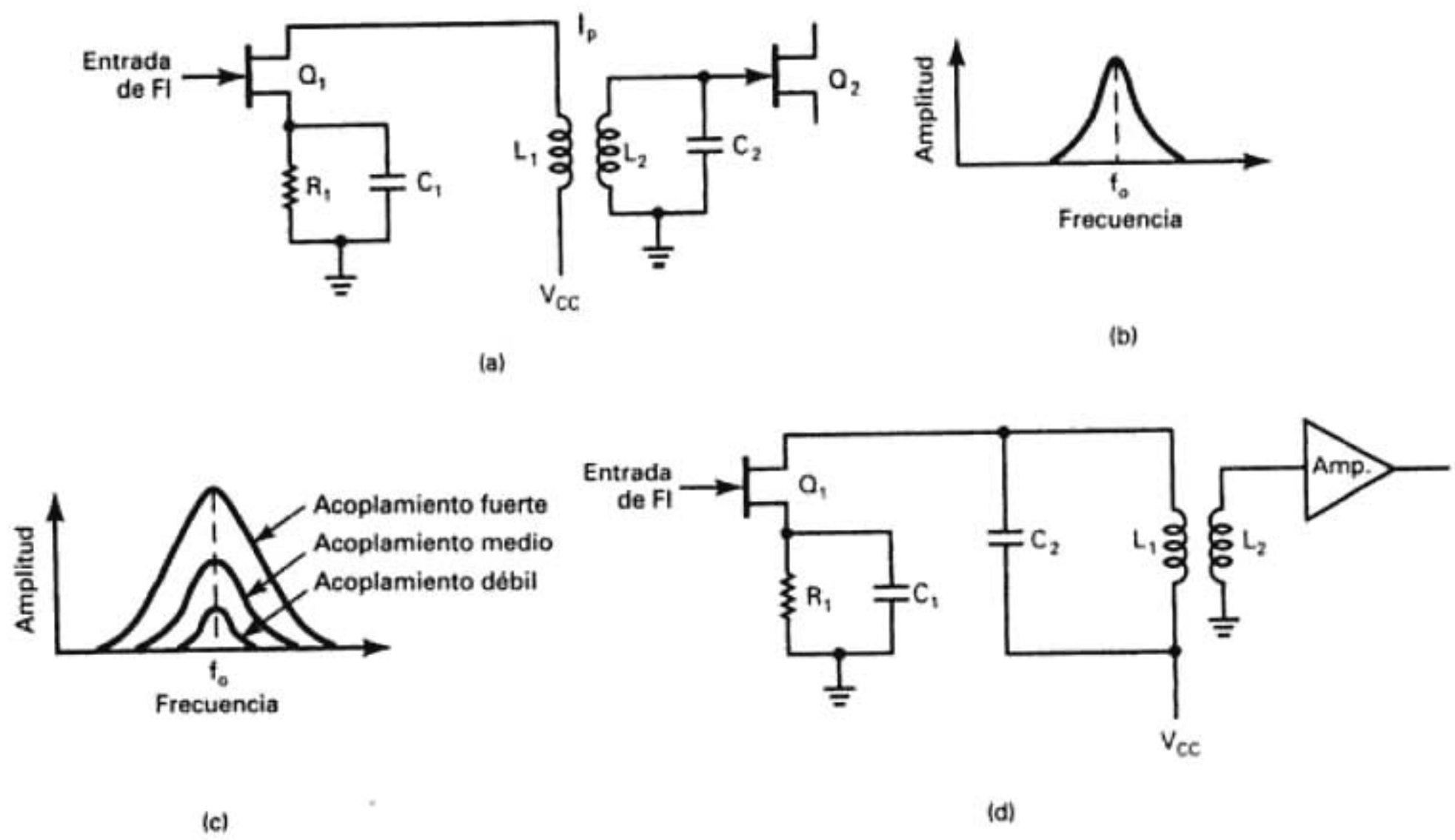


FIGURA 4-20 Transformador sintonizado simple: (a) diagrama; (b) curva de respuesta; (c) efectos del acoplamiento; (d) primario sintonizado-secundario no sintonizado

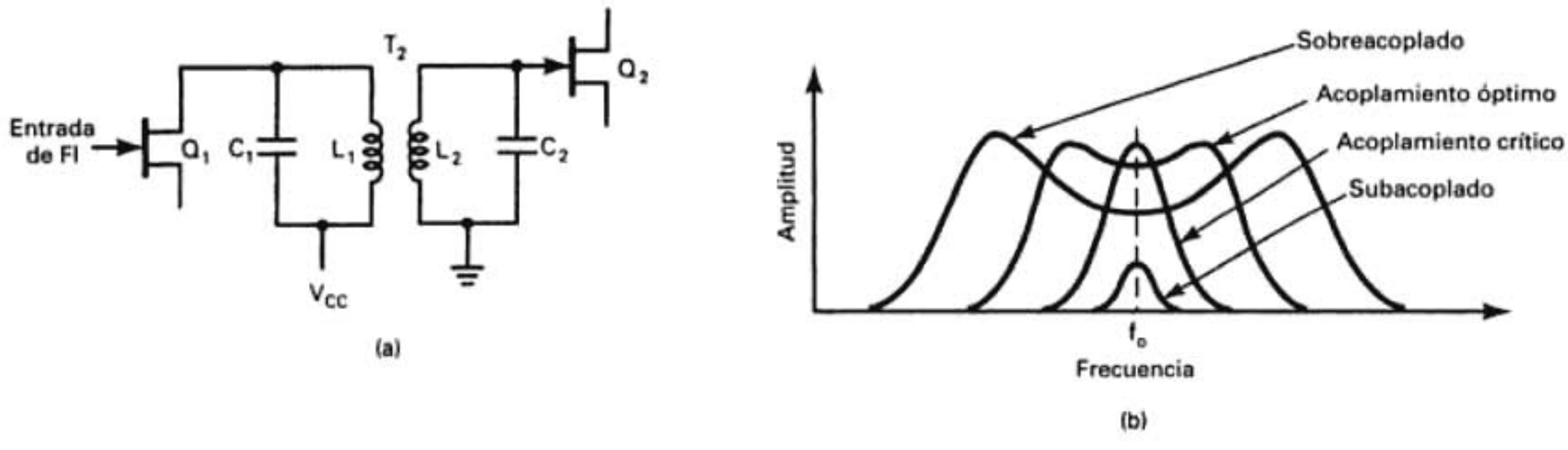


FIGURA 4-21 Transformador sintonizado doble: (a) diagrama; (b) curva de respuesta

de la señal. Para una alta ganancia y un amplio ancho de banda, es necesario el *acoplamiento fuerte*. Otra configuración de amplificador sintonizado simple (de sintonía simple) es la de *primario sintonizado y secundario no sintonizado*, que se ve en la fig. 4-20d.

Transformadores de doble sintonía. La fig. 4-21a muestra el esquema de un *amplificador de acoplamiento inductivo y doble sintonía*, o *doble resonancia*. Esta configuración se llama de *primario sintonizado y secundario sintonizado, porque los devanados del primario y del secundario del transformador* T_1 son circuitos tanque sintonizados. La fig. 4-21b muestra el efecto del acoplamiento sobre la curva de respuesta de un transformador de acoplamiento inductivo y doble sintonía. Esta curva se parece mucho a la de un circuito sintonizado simple para valores de acoplamiento menores al de *acoplamiento crítico*, k_c. El acoplamiento crítico es el punto en el que la resistencia reflejada es igual a la del primario, y la Q del circuito tanque primario disminuye a la mitad y el ancho de banda sube al doble. Si aumenta el coeficiente de acoplamiento más allá del punto crítico, disminuye la respuesta en la frecuencia de resonancia, y se presentan dos nuevos máximos, uno en cada lado de la frecuencia de resonancia. Este

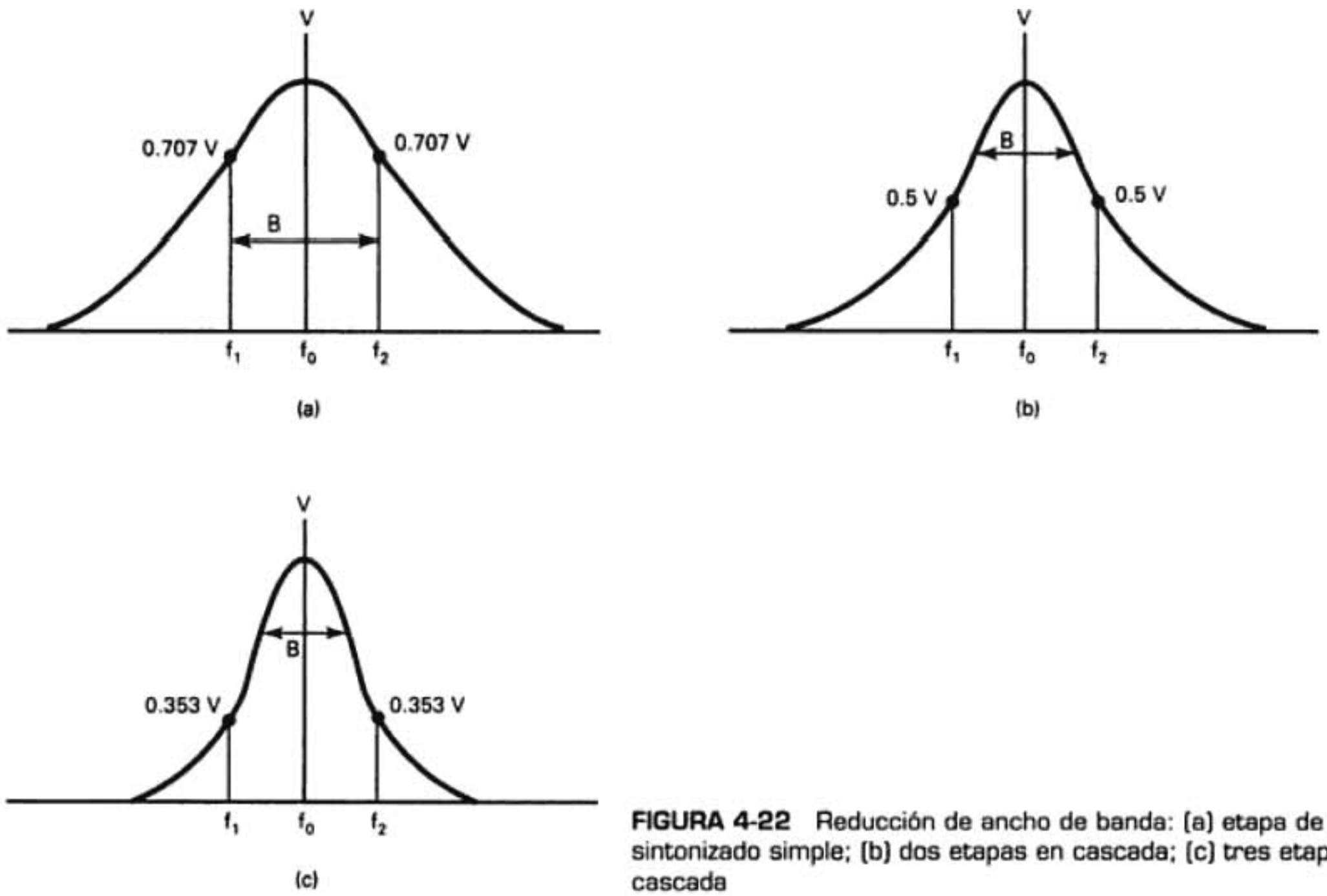


FIGURA 4-22 Reducción de ancho de banda: (a) etapa de sintonizado simple; (b) dos etapas en cascada; (c) tres etapas en cascada

doble máximo se debe al elemento reactivo de la impedancia reflejada, que tiene la importancia suficiente como para cambiar la frecuencia de resonancia del circuito primario sintonizado. Si el coeficiente de acoplamiento aumenta más, la baja en la resonancia se hace más pronunciada, y los dos máximos se apartan todavía más de la frecuencia de resonancia. Al aumentar el acoplamiento fuera del valor crítico se amplía la banda, pero al mismo tiempo se forma una ondulación en la curva de respuesta. Una curva de respuesta ideal tiene una forma rectangular: con meseta plana y faldas verticales. Se puede ver en la fig. 4-21b que con un coeficiente de acoplamiento más o menos 50% mayor que el valor crítico proporciona un buen balance entre la respuesta plana y las faldas verticales. Ese valor de acoplamiento se llama *acoplamiento óptimo*, k_{opt}, y se define como sigue

$$k_{opt} = 1.5k_c \tag{4-13}$$

en donde k_{opt} = acoplamiento óptimo
k_c = acoplamiento crítico = $1/\sqrt{Q_pQ_s}$

siendo Q_p y Q_s los valores sin acoplamiento.

El ancho de banda de un amplificador de doble resonancia es

$$B_{dt} = kf_o \tag{4-14}$$

Reducción de ancho de banda. Cuando se conectan en cascada varios amplificadores sintonizados, la respuesta total es igual al producto de las frecuencias individuales de los amplificadores. La fig. 4-22a muestra la curva de respuesta de un amplificador sintonizado. La ganancia a f_1 y f_2 es 0.707 de la ganancia a f_o. Si se conectan en cascada dos amplificadores sintonizados idénticos, la ganancia a f_1 y f_2 se reduce a 0.5 (=0.707 × 0.707), y si se conectan en cascada tres amplificadores sintonizados idénticos, la ganancia a f_1 y f_2 baja a 0.353. En consecuencia, al agregar más amplificadores, la curva general de respuesta se angosta y se reduce el ancho de

banda. Esta reducción de ancho de banda se ve en las figs. 4-22b y c. El ancho de banda general con n etapas de sintonía simple es

$$B_n = B_1(\sqrt{2^{1/n}-1}) \tag{4-15}$$

en donde B_n = ancho de banda de n etapas de sintonía simple (hertz)
B_1 = ancho de banda de una etapa de sintonía simple (hertz)
n = cantidad de etapas (cualquier entero positivo)

El ancho de banda con n etapas de doble sintonía es

$$B_{ndt} = B_{1dt}[2^{1/n}-1]^{1/4} \tag{4-16}$$

en donde B_{ndt} = ancho de banda general de n amplificadores de doble sintonía (hertz)
B_{1dt} = ancho de banda de un amplificador de doble sintonía (hertz)
n = cantidad de etapas de doble sintonía (cualquier entero positivo)

Ejemplo 4-7

Calcular el ancho de banda general para:

(a) Dos amplificadores de sintonía simple, cada uno con 10 kHz de ancho de banda.
(b) Tres amplificadores de sintonía simple, cada uno con 10 kHz de ancho de banda.
(c) Cuatro amplificadores de sintonía simple, cada uno con 10 kHz de ancho de banda.
(d) Un amplificador de doble sintonía con acoplamiento óptimo, acoplamiento crítico de 0.02 y frecuencia de resonancia de 1 MHz.
(e) Repetir las partes a), b) y c) para el amplificador de doble sintonía de la parte d)

Solución (a) De acuerdo con la ecuación 4-15,

$$B_2 = 10\text{ kHz}(\sqrt{2^{1/2}-1}) = 6436\text{ Hz}$$

(b) De nuevo con la ecuación 4-15,

$$B_3 = 10\text{ kHz}(\sqrt{2^{1/3}-1}) = 5098\text{ Hz}$$

(c) De nuevo con la ecuación 4-15,

$$B_4 = 10\text{ kHz}(\sqrt{2^{1/4}-1}) = 4350\text{ Hz}$$

(d) De acuerdo con la ecuación, $K_{opt} = 1.5(0.02) = 0.03$

De la ecuación 4-14, $B_{dt} = 0.03(1\text{ MHz}) = 30\text{ kHz}$

(e) Según la ecuación 4-16,

n	B (Hz)
2	24,067
3	21,420
4	19,786

Los transformadores de FI se consiguen como circuitos sintonizados de diseño especial, en paquetes metálicos conectables a tierra, llamados *transformadores de FI*. Las figs. 4-23a y b muestran los esquemas físicos y de circuito de un transformador de FI típico. El devanado primario viene con un capacitor de 125 pF conectado en paralelo. La flecha vertical entre los devanados primario y secundario indica que el núcleo de ferrita se puede sintonizar con un destornillador no metálico, o herramienta de sintonía. Al ajustar el núcleo de ferrita se cambia la inductancia mutua que controla la magnitud del voltaje inducido en los devanados secundarios. La toma en el devanado primario se puede usar para aumentar la Q del circuito de colector del transistor de activación. Si no se usa la toma, el circuito equivalente es el que muestra la fig. 4-23b; la Q efectiva es $Q = R_L/X_L$, y el ancho de banda $B = f/Q$, donde f es la frecuencia de resonancia. Si se conecta la toma a tierra, el circuito equivalente es el de la fig. 4-23c. Con la toma a tierra aumenta la Q efectiva y la respuesta general es más selectiva.

Amplificadores de FI en circuito integrado. En años recientes, los amplificadores de FI en circuito integrado han tenido aceptación universal en los sistemas de radio móviles, como por ejemplo en los radioteléfonos. Los circuitos integrados tienen las ventajas obvias de tamaño pequeño y bajo consumo de energía. Uno de los amplificadores en CI para frecuencias interme-

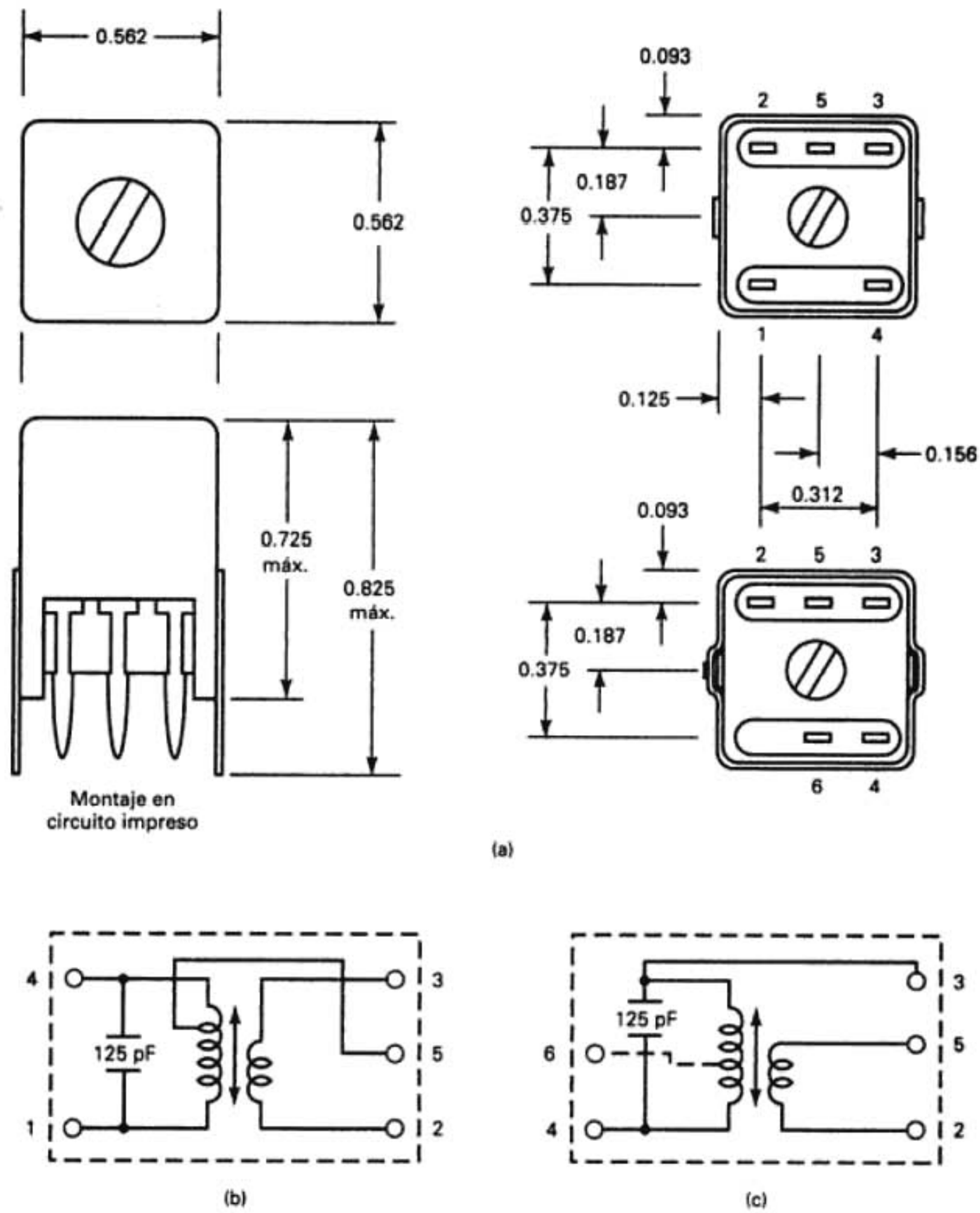


FIGURA 4-23 Transformador de FI: (a) diagrama físico; (b) diagrama eléctrico con bobina sin tomas; (c) diagrama con bobina con tomas

dias más usados es el CA3028A, que es diferencial en cascode, diseñado para las comunicaciones y el equipo industrial como amplificador de FI o de RF, a frecuencias desde cd hasta 120 MHz. El CA3028A tiene voltaje de desplazamiento, de polarización, corriente de desplazamiento y corriente de entrada de controlados. Usa una configuración de amplificador diferencial balanceado con fuente de corriente constante controlada, y se puede usar para operar con terminal sencilla o dual. El CA3028A tiene funciones de control automático de ganancia (CAG) y un amplio intervalo de corriente de operación.

La fig. 4-24a muestra el diagrama del CA3028A, y la fig. 4-24b lo muestra usado como amplificador de FI en cascode. La entrada de FI se aplica a la terminal 2 y la salida se toma de la terminal 6. Cuando el voltaje de CAG en la terminal 1 es igual al de referencia de CAG, en la terminal 5, las corrientes de emisor que circulan en Q_1 y Q_2 son iguales, y el amplificador tiene alta ganancia. Si aumenta el voltaje del CAG en la terminal 1, disminuye la corriente Q_2 y baja la ganancia de etapa.

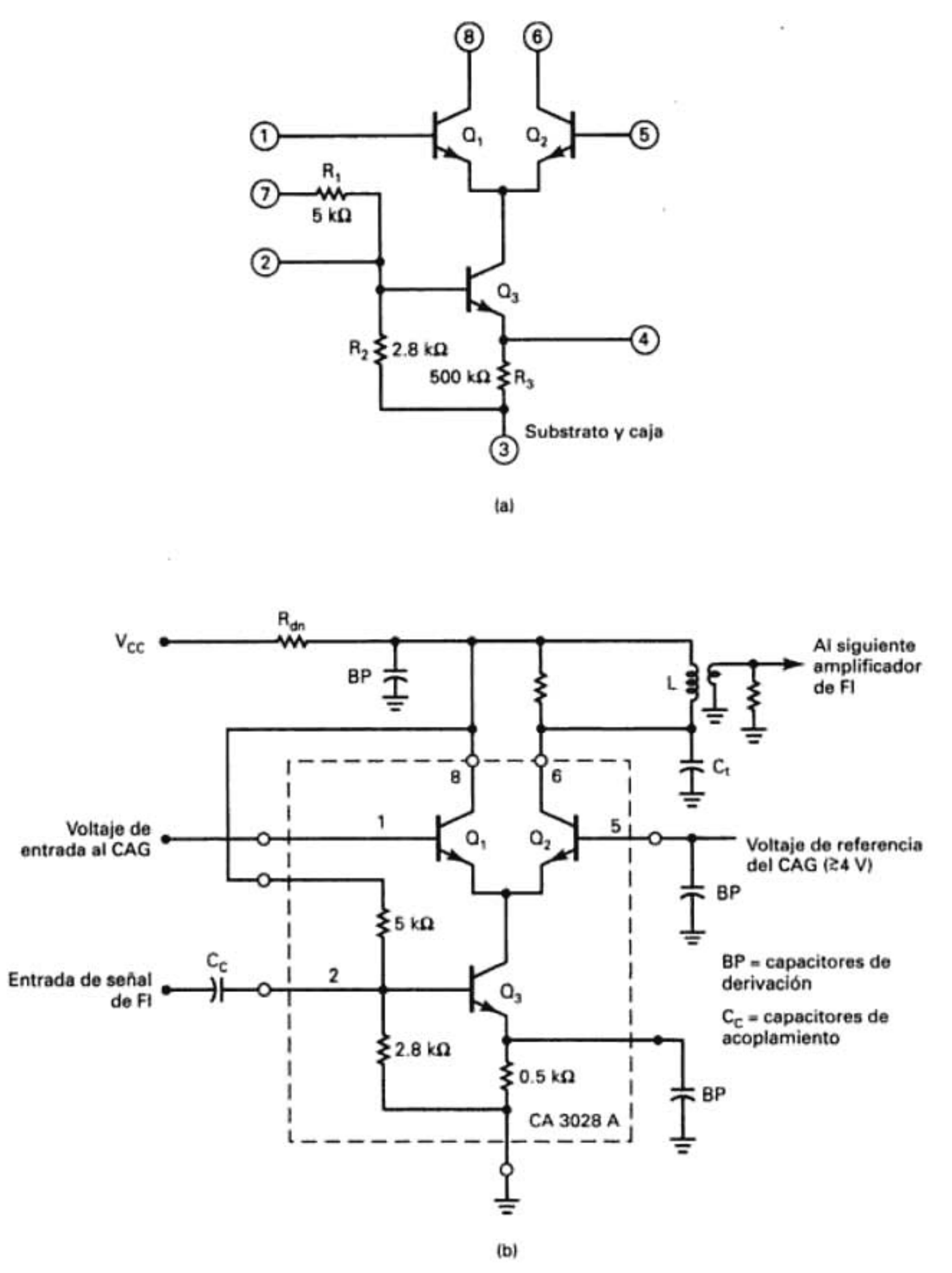


FIGURA 4-24 Amplificador de circuito integrado lineal, diferencial/cascode CA3028A: (a) diagrama esquemático; (b) configuración del amplificador en cascode

Circuitos detectores de AM

La función de un detector de AM es demodular la señal de AM y recuperar o reproducir la información de la fuente original. La señal recuperada debe contener las mismas frecuencias que las de la información original, y tener las mismas características de amplitud relativa. A veces, al detector de AM se le llama *segundo detector*, ya que el mezclador/convertidor es el primer detector, porque antecede al detector de AM.

Detector de picos. La fig. 4-25a muestra el esquema de un demodulador de AM sencillo, no coherente, que se llama a menudo *detector de picos*. Como un diodo es un dispositivo no

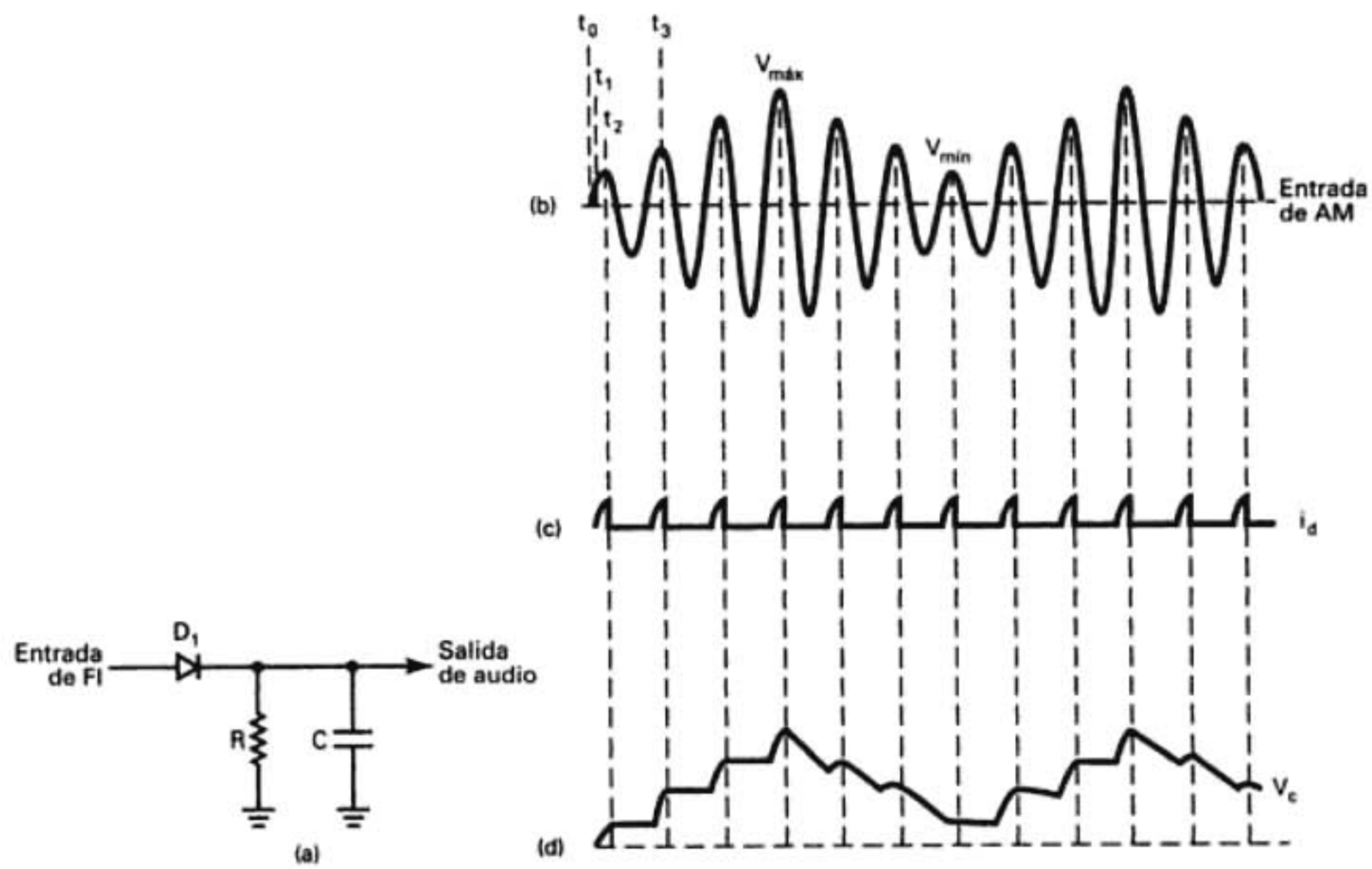


FIGURA 4-25 Detector de picos: (a) diagrama; (b) forma de onda de AM de entrada; (c) forma de onda de corriente de diodo; (d) forma de onda de voltaje de salida

lineal, se hace mezclado no lineal en D_1 cuando se aplican a su entrada dos o más señales. Por lo anterior, la salida contiene las frecuencias de entrada originales, sus armónicas y sus productos cruzados. Si una portadora de 300 kHz es modulada en amplitud por una onda senoidal de 2 kHz, la onda modulada está formada por una frecuencia de lado inferior, la portadora y la frecuencia de lado superior, de 298 kHz, 300 kHz y 302 kHz, respectivamente. Si la señal resultante es la entrada al detector de AM de la fig. 4-25a, la salida abarcará las tres frecuencias de entrada, las armónicas de las tres frecuencias, y los productos cruzados de todas las combinaciones posibles de las tres frecuencias y sus armónicas. Esa salida es la siguiente

$$f_{sal} = \text{frecuencias de entrada} + \text{armónicas} + \text{sumas y diferencias}$$

Como la red *RC* es un filtro pasabajas, sólo pasan a la sección de audio las frecuencias de diferencia, y entonces, la salida es simplemente

$$\begin{aligned} f_{sal} &= 300 - 298 = 2 \text{ kHz} \\ &= 302 - 300 = 2 \text{ kHz} \\ &= 302 - 298 = 4 \text{ kHz} \end{aligned}$$

Debido a las características de amplitud relativa de las frecuencias laterales superior e inferior y de la portadora, la diferencia entre la frecuencia de la portadora y de las laterales es la señal predominante de salida. En consecuencia, para fines prácticos, la señal moduladora original (2 kHz) es el único componente que tiene la salida del detector de picos.

En el análisis anterior se consideró al detector de diodo como un mezclador simple, y lo es. En esencia, la diferencia entre un modulador de AM y un demodulador de AM es que la salida de un modulador se sintoniza a las frecuencias de suma (convertidor ascendente), mientras que la salida de un demodulador se sintoniza a las frecuencias de diferencia (convertidor descendente). El circuito demodulador de la fig. 4-25a se suele llamar *detector de diodo*, porque el dispositivo no lineal es un diodo, o *detector de picos*, porque detecta los máximos de la envolvente de entrada. También se le llama *detector de forma* o *detector de envolvente*, porque detecta la forma

de la envolvente de entrada. En esencia, la señal portadora *captura* al diodo y lo obliga a encenderse y apagarse (a rectificar) en sincronía, tanto en frecuencia como en fase. Así, las frecuencias laterales se mezclan con la portadora y se recuperan las señales originales en la banda base.

Las figs. 4-25b, c y d muestran una forma de onda de voltaje de entrada al detector, la forma de onda de corriente correspondiente en el diodo y la forma de onda de voltaje de salida del detector. En el momento t_0, el diodo tiene polarización inversa y está apagado ($i_d = 0$ A), el capacitor está totalmente descargado ($V_C = 0$ V) y así la salida es 0 V. El diodo permanece apagado hasta que el voltaje de entrada es mayor que el potencial de barrera de D_1, aproximadamente 0.3 V. Cuando V_{ent} llega a 0.3 V (t_1), el diodo se enciende y comienza a pasar corriente, cargando al capacitor. El voltaje del capacitor permanece en 0.3 V abajo del voltaje de entrada, hasta que V_{ent} llega a su valor máximo. Cuando el voltaje de entrada comienza a disminuir, el diodo se apaga e i_d pasa a 0 A (t_2). El capacitor comienza a descargarse por el resistor, pero la constante del tiempo RC se hace suficientemente grande para que el capacitor no se pueda descargar con tanta rapidez como la de bajada de V_{ent}. El diodo permanece apagado hasta el siguiente ciclo de entrada, cuando V_{ent} llega a 0.3 V más positivo que V_C (en t_3). En ese momento el diodo se enciende, pasa la corriente y el capacitor comienza a cargar de nuevo. Es relativamente fácil que el capacitor se cargue al nuevo valor, porque la constante de tiempo RC de carga es R_dC, donde R_d es la resistencia del diodo *activo* o *encendido*, que es bastante baja. Esta secuencia se repite en cada pico positivo sucesivo de V_{ent}, y el voltaje del capacitor sigue a los picos positivos de V_{ent}, de aquí su nombre de detector de picos. La forma de onda de salida se asemeja a la envolvente de entrada, y de aquí el nombre de detector de forma. La forma de onda de salida tiene una ondulación de alta frecuencia, igual a la frecuencia de la portadora. Esto es porque el diodo se enciende durante los picos positivos de la envolvente. La ondulación residual se elimina con facilidad con los amplificadores de audio, porque la frecuencia de la portadora es mucho mayor que la máxima frecuencia de la señal moduladora. El circuito de la fig. 4-25 responde sólo a los picos positivos de V_{ent} y en consecuencia se llama *detector de picos positivos*. Si sencillamente se voltea el diodo, el circuito se transforma en detector de picos negativos. El voltaje de salida llega a su amplitud positiva máxima al mismo tiempo que la envolvente de salida alcanza su valor máximo positivo ($V_{máx}$), y el voltaje de salida llega a su amplitud mínima al mismo tiempo que el voltaje de entrada llega a su valor mínimo ($V_{mín}$). Para una modulación de 100%, V_{sal} va de 0 V a un valor igual a $V_{máx} - 0.3$ V.

La fig. 4-26 muestra las formas de onda de entrada y de salida de un detector de picos con diversos porcentajes de modulación. Sin modulación, un detector de picos no es más que un rectificador de media onda filtrada, y el voltaje de salida es aproximadamente igual al voltaje máximo de entrada menos los 0.3 V. Al cambiar el porcentaje de modulación, aumentan y disminuyen proporcionalmente las variaciones de voltaje de salida; la forma de onda de salida sigue la forma

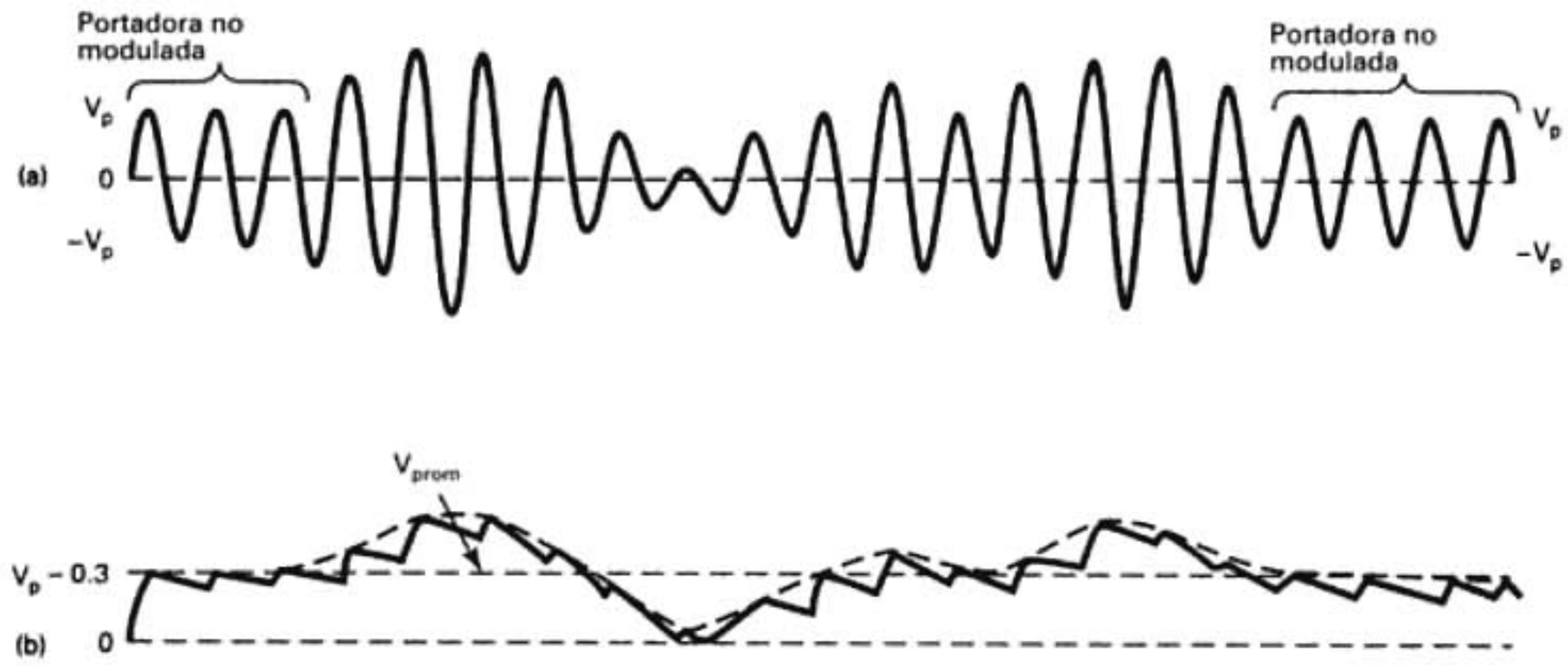


FIGURA 4-26 Detector de picos positivos: (a) forma de onda de entrada; (b) forma de onda de salida

de la envolvente de AM. Sin embargo, independiente de si haya modulación, el valor promedio del voltaje de salida es aproximadamente igual al valor máximo de la portadora no modulada.

Distorsión por detector. Cuando aumentan los picos sucesivos de la forma de onda de entrada al detector, es importante que el capacitor mantenga su carga entre ellos, es decir, es necesaria una constante de tiempo *RC* relativamente grande. Sin embargo, cuando va disminuyendo la amplitud de los picos positivos, es importante que el capacitor se descargue entre los picos sucesivos, a valores menores que el del pico siguiente; en consecuencia se necesita una constante de tiempo *RC* pequeña. Es obvio que se debe hacer un balance entre constante de tiempo grande y pequeña. Si la constante de tiempo *RC* es demasiado pequeña, la forma de onda de salida se parece a una señal rectificada de media onda. A esto se le llama a veces *distorsión por rectificación*, y se ve en la fig. 4-27b. Si la constante de tiempo *RC* es demasiado grande, la pendiente de la forma de onda de salida no puede seguir la pendiente trasera de la envolvente. A este tipo de distorsión se le llama *recorte diagonal*, y se ve en la fig. 4-27c.

La red *RC* que sigue al diodo en un detector de picos es un filtro pasabajas. La pendiente de la envolvente depende tanto de la frecuencia de la señal moduladora como del coeficiente de modulación, *m*. En consecuencia, la pendiente máxima (máxima rapidez de cambio) se tiene cuando la envolvente cruza su eje cero en la dirección negativa. La frecuencia máxima de la señal moduladora que se puede demodular con un detector de picos sin atenuación es

$$f_{m(\text{máx})} = \frac{\sqrt{(1/m^2) - 1}}{2\pi RC} \tag{4-17a}$$

en donde $f_{m(\text{máx})}$ = frecuencia máxima de la señal moduladora (hertz)
m = coeficiente de modulación (adimensional)
RC = constante de tiempo (segundos)

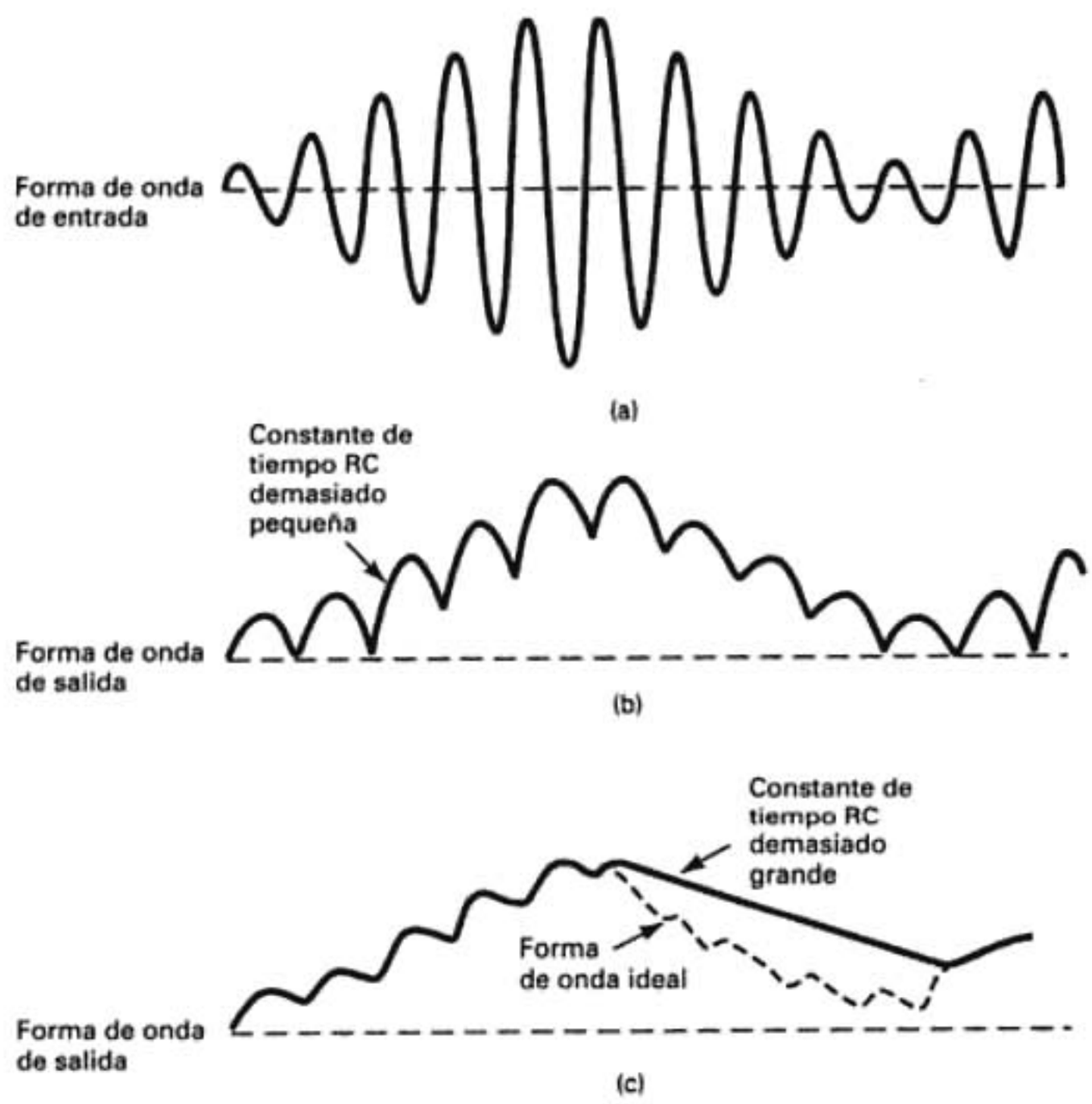


FIGURA 4-27 Distorsión por detector: (a) envolvente de entrada; (b) distorsión en rectificador; (c) recorte diagonal

Para tener modulación de 100%, el numerador de la ecuación 4-17a es cero, y eso significa en esencia que todas las frecuencias de la señal moduladora se atenúan cuando se demodulan. En el caso normal, la amplitud de la señal moduladora en un transmisor se limita, o se comprime, de tal modo que lo máximo que se alcanza es aproximadamente un 90% de modulación. Para una modulación de 70.7%, la ecuación 4.17a se reduce a

$$f_{m(\max)} = \frac{1}{2\pi RC} \tag{4-17b}$$

Se suele usar la ecuación 4-17b para diseñar detectores de picos, y determinar aproximadamente una señal moduladora máxima.

Controles automáticos de ganancia

Un circuito de *control automático de ganancia CAG* (AGC, de *automatic gain control*) compensa variaciones pequeñas en el nivel de la señal de RF recibida. Aumenta en forma automática la ganancia del receptor con valores bajos de entrada de RF, y disminuye en forma automática la ganancia del receptor cuando se recibe una señal fuerte de RF. Las señales débiles pueden quedar enterradas en ruido del receptor y, en consecuencia, pueden ser imposibles de detectar. Una señal demasiado fuerte puede sobreexcitar a los amplificadores de RF y/o de FI, y producir demasiada distorsión no lineal, y hasta saturación. Hay varias clases de CAG, que incluyen el simple, el demorado y el directo.

CAG simple. La fig. 4-28 muestra un diagrama de bloques de un receptor superheterodino de AM con CAG simple. El circuito de control automático de ganancia vigila el nivel de la señal recibida y manda de regreso una señal a los amplificadores de RF y FI, para que ajusten en forma automática su ganancia. La del CAG es una forma de retroalimentación degenerativa, o negativa. El objeto del CAG es permitir que un receptor detecte y demodule, con igual calidad, las señales que se transmiten desde distintas estaciones, cuya potencia de salida y distancia al receptor varía. Por ejemplo, la radio AM de un vehículo no recibe la misma intensidad de señal de todas las estaciones transmisoras de la zona o, lo que es igual, de una sola estación, cuando se mueve el automóvil. El circuito de CAG produce un voltaje que ajusta la ganancia del receptor y mantiene la potencia de portadora de FI, en la entrada del detector de AM, en un valor relativamente constante. El circuito de CAG no es una forma de *control automático de volumen* (AVC, por *automatic volume control*); el CAG es independiente de la modulación, y es inmune a los cambios normales de amplitud de señal moduladora.

La fig. 4-29 muestra un esquema de un circuito de CAG simple. Como se puede ver, el circuito es en esencia un detector de picos. De hecho, con mucha frecuencia el voltaje de corrección CAG se toma de la salida del detector de audio. En la fig. 4-26 se indicó que el voltaje de cd a la salida de un detector de picos es igual a la amplitud máxima de la portadora no modulada, menos el potencial de barrera del diodo, y es totalmente independiente del grado de modulación. Si aumenta la amplitud de la portadora, aumenta el voltaje de CAG, y si disminuye la amplitud de la portadora, el voltaje de CAG disminuye. El circuito de la fig. 4-29 es un detector de picos negativos, y produce un voltaje negativo en su salida. Mientras mayor sea la amplitud de la portadora en la entrada, el voltaje de salida será más negativo. El voltaje negativo del detector del CAG se retroalimenta a la etapa de FI, donde controla el voltaje de polarización en la base de Q_1. Cuando aumenta la amplitud de la portadora, el voltaje en la base de Q_1 se vuelve menos positivo, y hace que baje la corriente del emisor. Como resultado, aumenta r'_e y disminuye la ganancia del amplificador, r_c/r'_e, lo cual a su vez hace que disminuya la amplitud de la portadora. Cuando disminuye la amplitud de la portadora, el voltaje del CAG se hace menos negativo, aumenta la corriente del emisor, disminuye r'_e y aumenta la ganancia del amplificador. El capacitor C_1 es de derivación de audio, que evita que los cambios de voltaje de CAG debidos a la modulación afecten a la ganancia de Q_1.

CAG demorado. El CAG simple se usa en los receptores de radio AM comercial menos costosos. Sin embargo, con ese sistema, la polarización de CAG comienza a aumentar tan pronto como la intensidad de la señal recibida es mayor que el ruido térmico del receptor. En

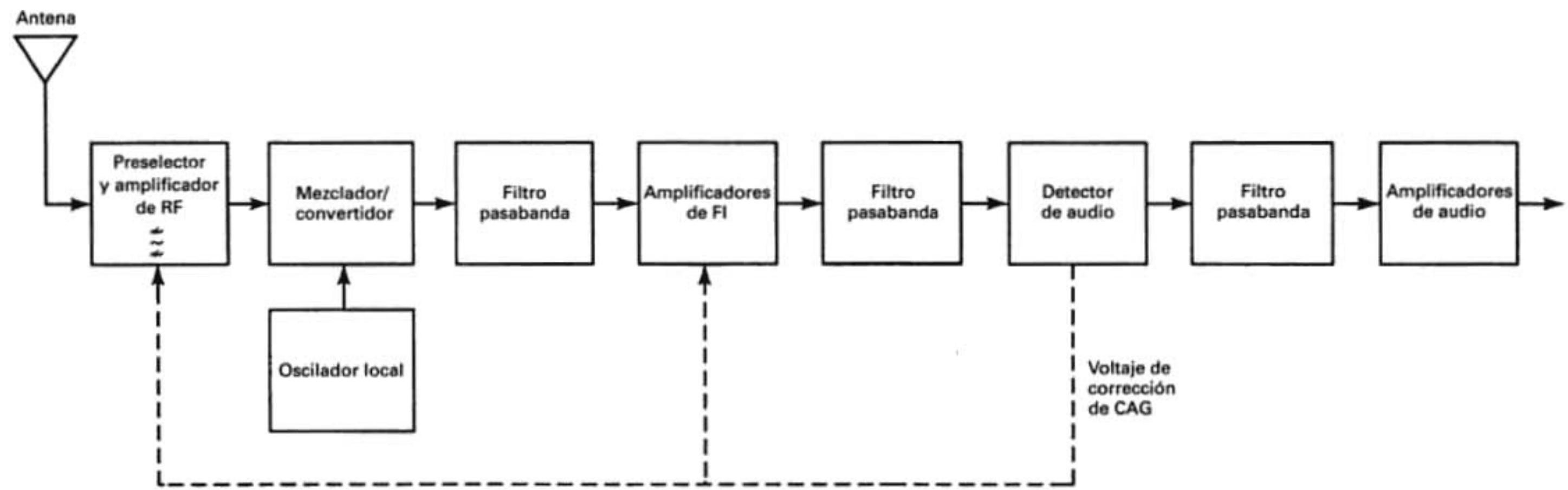


FIGURA 4-28 Receptor de AM con CAG simple

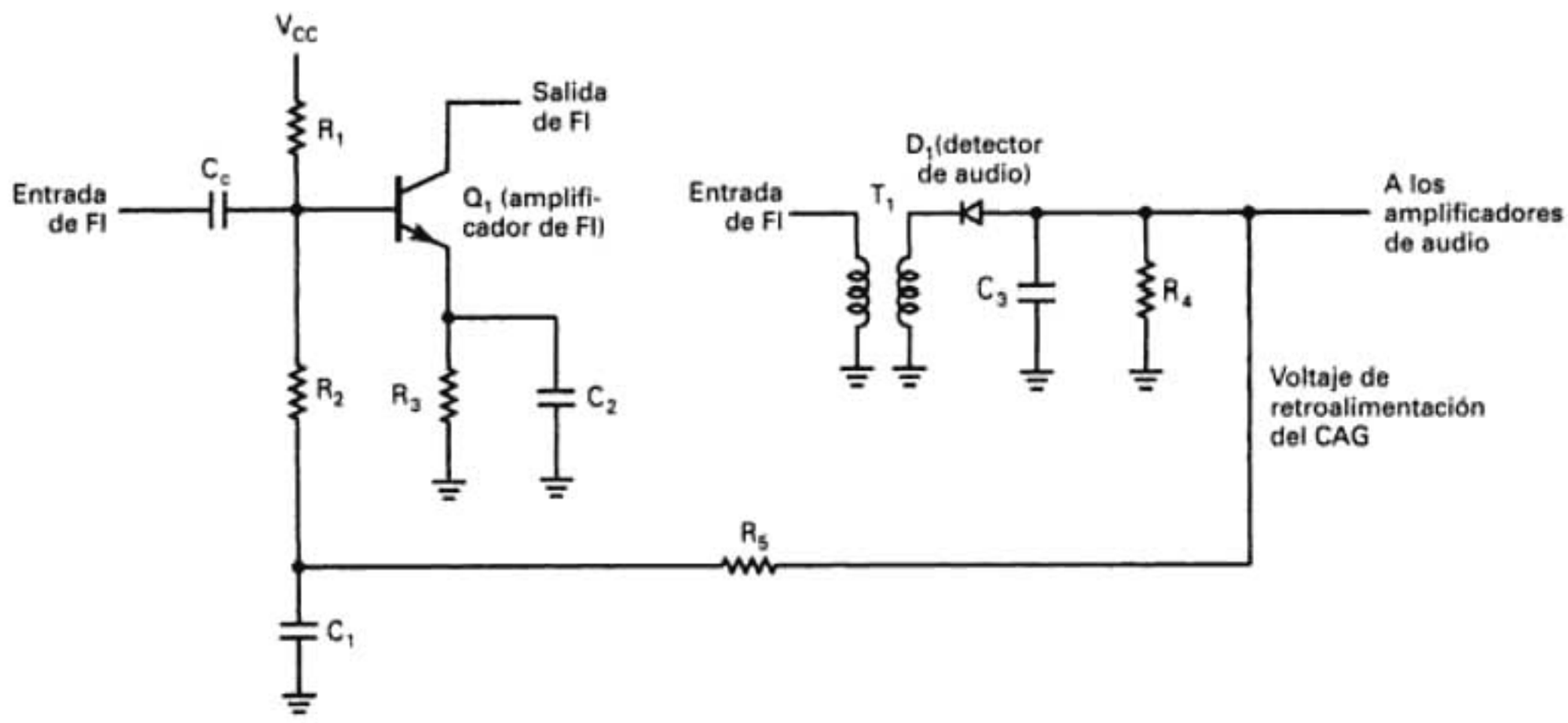


FIGURA 4-29 Circuito de CAG simple

consecuencia, el receptor se hace menos sensible (a esto a veces se le llama *desensibilización automática*). El *CAG demorado* evita que el voltaje de retroalimentación del CAG llegue a los amplificadores de RF o FI, hasta que el nivel de RF pase de una magnitud predeterminada. Una vez que la señal portadora ha rebasado el nivel umbral, el voltaje de CAG demorado es proporcional a la intensidad de la señal portadora. La fig. 4-30a muestra las características de respuesta del CAG simple y del demorado. Se ve que con CAG demorado, la ganancia del receptor no se afecta, sino hasta que se rebasa el nivel umbral, mientras que en el CAG simple, la ganancia del receptor se afecta de inmediato. El CAG demorado se usa en los receptores de comunicaciones más complicados. La fig. 4-30b muestra la ganancia de FI en función del nivel de señal de RF en la entrada, tanto para CAG simple como demorado.

CAG directo. Un problema inherente al CAG, tanto simple como demorado, es que los dos son formas de compensación *post-CAG* (después de lo sucedido). Con post-CAG, el circuito que vigila la intensidad de la portadora y suministra el voltaje de corrección CAG, está después de los amplificadores de FI, y en consecuencia, el solo hecho de que cambió el voltaje de CAG indica que puede ser demasiado tarde: la intensidad de la portadora ya cambió, y se ha propagado por el receptor. Así, ni el CAG simple ni el demorado pueden compensar con exactitud los cambios rápidos de amplitud de la portadora. El *CAG directo* es parecido al CAG convencional, pero la señal de recepción se vigila más cerca del frente del receptor, y el voltaje de corrección se alimenta positivamente a los amplificadores de FI. En consecuencia, cuando se detecta un cambio en el nivel de la señal, se puede compensar en las etapas siguientes. La fig. 4-31 muestra un receptor superheterodino de AM con CAG directo. Para conocer un método más complicado de lograr el CAG, véase la sección "Radiorreceptor de FM en dos sentidos", en el capítulo 7.

Circuitos de reducción de ruido

El objeto de un *circuito reductor de ruido* es *silenciar* a un receptor cuando no hay recepción de señal. Si un receptor de AM se sintoniza en un lugar del espectro de RF donde no hay señal de RF, el circuito de CAG ajusta al receptor para tener ganancia máxima. En consecuencia, el receptor amplifica y demodula su propio ruido interno. Es el conocido ruido de frituras y estática que tiene la bocina cuando no hay recepción de portadora. En los sistemas domésticos de AM, cada estación transmite en forma continua una portadora, independientemente de si hay modulación alguna. Por consiguiente, la única vez que se oye el ruido del receptor inactivo es cuando se sintoniza entre las estaciones. Sin embargo, en los sistemas de radio en dos sentidos, en general la portadora del transmisor se desconecta, o se apaga, excepto cuando hay una señal moduladora. Por lo anterior, durante los tiempos de transmisión inactiva, un receptor sólo

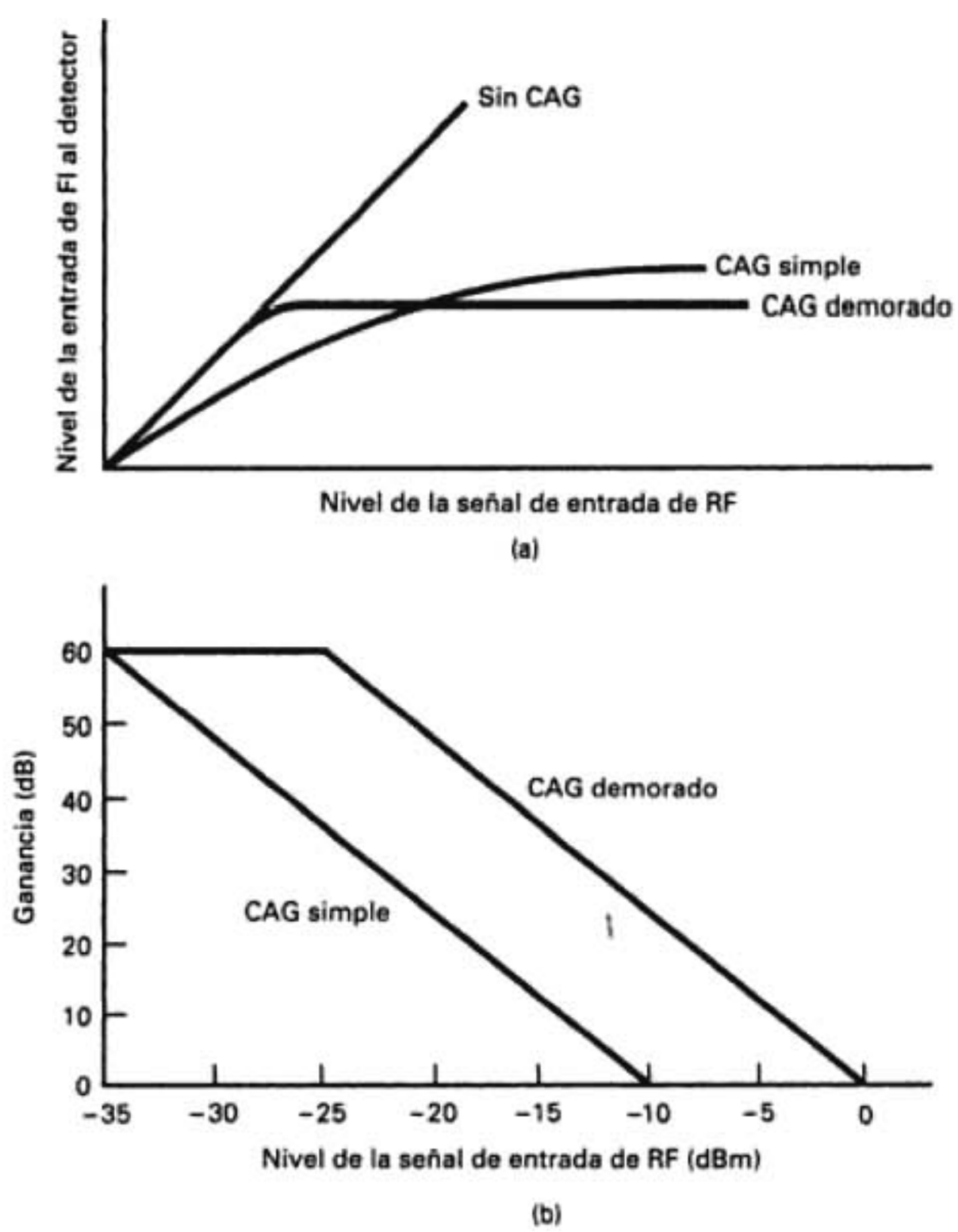


FIGURA 4-30 Control automático de ganancia (CAG): (a) características de respuesta; (b) ganancia de FI en función del nivel de la señal de entrada de RF

amplifica y demodula ruido. Un circuito supresor de ruidos mantiene apagada o *enmudecida* la sección de audio del receptor cuando no hay señal recibida (el receptor está suprimido). Una desventaja de un circuito eliminador de ruido es que las señales débiles de RF no producen salida de audio.

La fig. 4-32 muestra un diagrama de un circuito reductor de ruidos. Ese circuito usa el voltaje de CAG para vigilar la intensidad de la señal de RF recibida. Mientras mayor sea el voltaje de CAG, la señal de RF es más fuerte. Cuando el voltaje de CAG baja de un nivel preestablecido, el circuito de eliminación se activa, y desactiva la sección de audio del receptor. En la fig. 4-32 se puede ver que el detector de reducción de ruido usa un divisor resistivo de voltaje, para vigilar el voltaje del CAG. Cuando la señal de RF baja del umbral del reductor, se enciende Q_2 y apaga los amplificadores de audio. Cuando la señal de RF aumenta sobre el umbral del reductor, el voltaje del CAG se hace más negativo, apaga Q_2 y activa los amplificadores de audio. Se puede ajustar el valor umbral del supresor con R_3. Un método más complicado de enmudecer a un receptor se describirá en el capítulo 7, en la sección "Radiorreceptor de FM en dos sentidos".

Limitadores y eliminadores de ruido

Se pueden eliminar estados transitorios de ruido de gran amplitud y corta duración, como por ejemplo el ruido impulsivo, con *limitadores* o *recortadores* de diodo, en la sección de audio de un receptor. El nivel de umbral de limitación o recorte se establece normalmente justo arriba del nivel máximo de la señal de audio. Por consiguiente, los pulsos de ruido prácticamente no afectan la señal, pero se limitan a ser del mismo nivel aproximado que la señal. Los pulsos de ruido

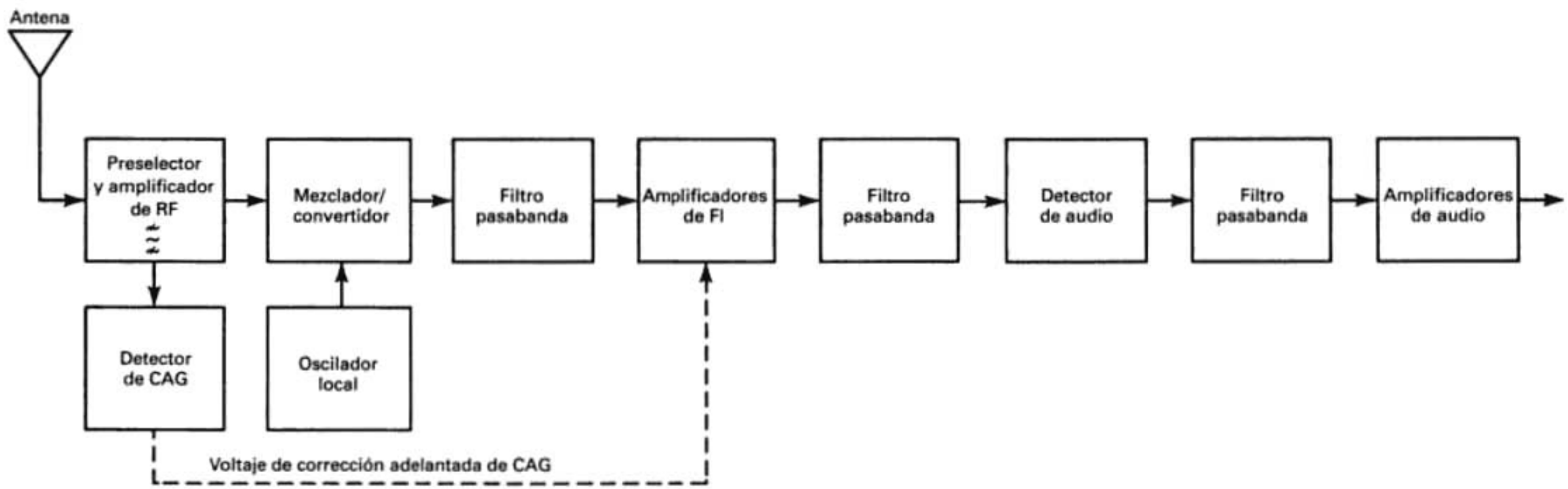


FIGURA 4-31 CAG directo

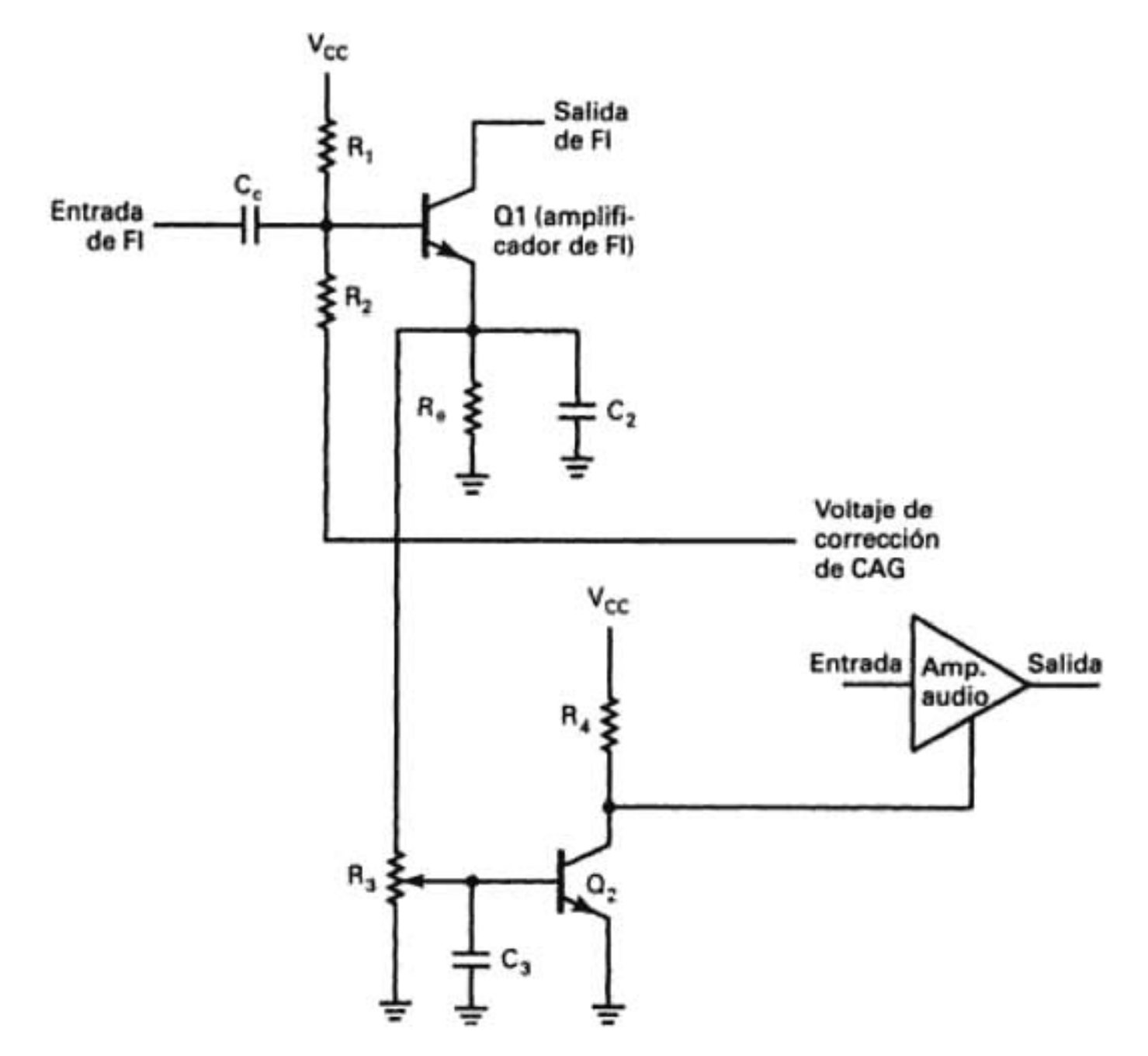


FIGURA 4-32 Circuito reductor de ruidos

son, en general, señales de amplitud grande y corta duración y por consiguiente, al limitarlos se quita mucha de su energía, y se hacen mucho menos molestos.

Un *circuito eliminador* es otra opción que se usa con frecuencia para reducir los efectos del ruido impulsivo de gran amplitud. En esencia, un circuito eliminador detecta la presencia de un pico de gran amplitud y corta duración, y enmudece al receptor apagando una parte del mismo, durante la duración del pulso. Por ejemplo, cuando se detecta un pulso de ruido a la entrada de la sección amplificadora de FI en un receptor, el circuito eliminador apaga los amplificadores de FI durante el pulso, aquietando así al receptor. Ha demostrado ser más efectivo apagar los amplificadores de FI que apagar la sección de audio, porque los filtros pasabanda más anchos de la etapa de FI tienen una tendencia a ensanchar el pulso de ruido.

Sin embargo, los circuitos limitadores y eliminadores tienen poco efecto sobre el ruido blanco, porque la potencia de ruido blanco es en general mucho menor que la potencia de la señal, y los limitadores y los eliminadores sólo trabajan cuando el ruido da un salto hasta un valor superior al de la señal.

Medidas alternas de señal a ruido

La sensibilidad tiene poca importancia, a menos que vaya acompañada de una relación de señal a ruido y, como es difícil separar la señal del ruido y viceversa, con frecuencia la sensibilidad se acompaña de una *indicación de señal más ruido entre ruido*, (S + N)/N. La sensibilidad de un receptor de AM se define en general como el valor mínimo de la señal en la entrada del receptor con modulación de 30%, necesaria para producir una potencia de salida de audio cuando menos de 500 mW con una relación de (S + N)/N de 10 dB.

Para medir (S + N)/N, se aplica una portadora modulada 30% por un tono de 1 kHz en la entrada del receptor. La potencia total de audio se mide en la salida del receptor, que incluye la señal de audio y el ruido presente en el ancho de banda de audio. A continuación se elimina la modulación de la señal de RF y se vuelve a medir la potencia total de audio. Sin embargo, esta vez sólo hay señales de ruido. El objeto de dejar activa la portadora, más que quitar la entrada

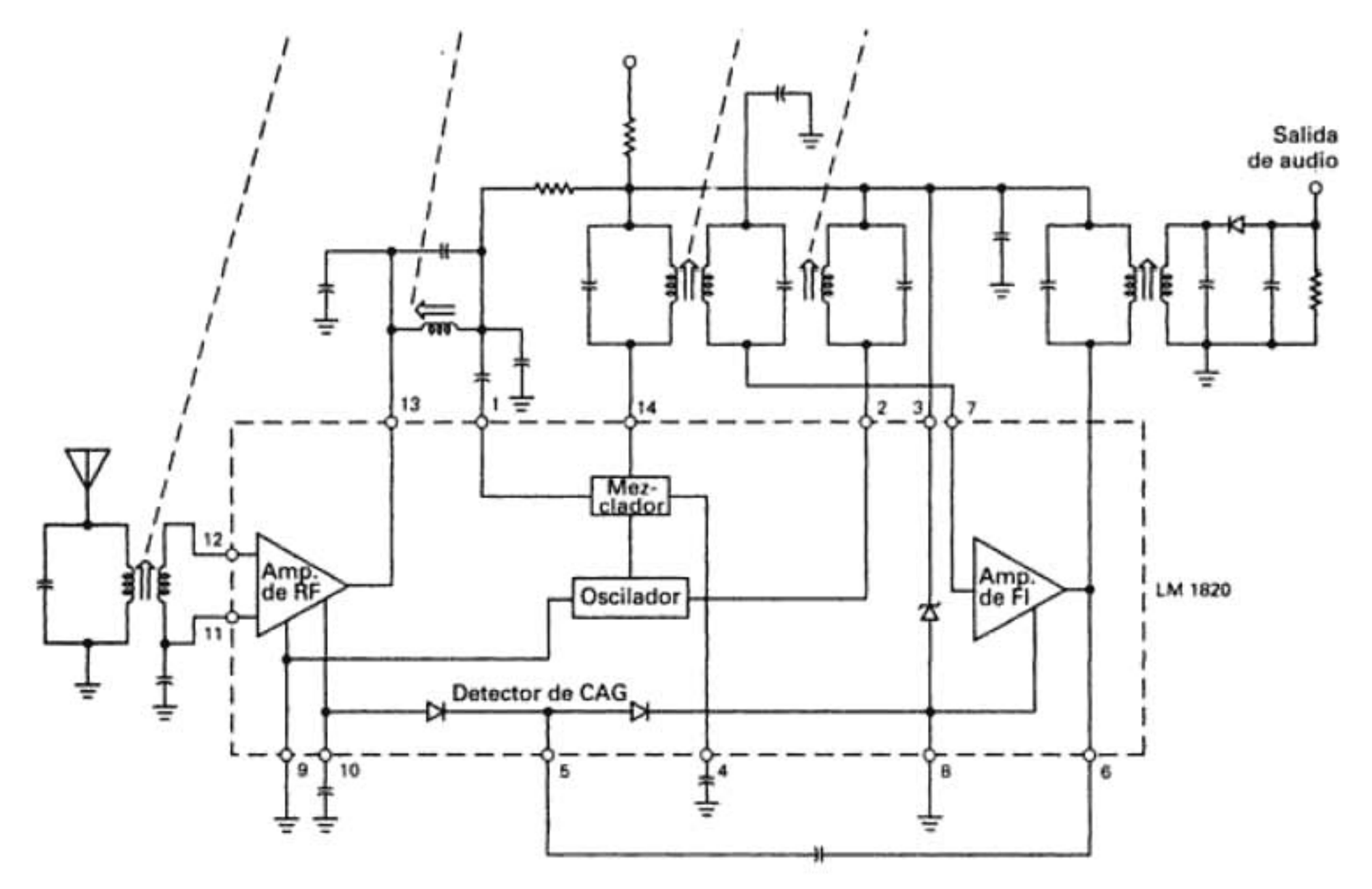


FIGURA 4-33 Sistema de radio AM con circuito integrado lineal LM1820

al receptor, es que la portadora es necesaria para evitar que el circuito de CAG detecte ausencia de portadora y ponga la ganancia del receptor en su máximo. La indicación de ruido cuando el receptor está funcionando con ganancia máxima amplificaría el ruido generado internamente, más allá de su nivel normal, produciendo así una indicación de ruido carente de sentido.

Otro método de medir las intensidades relativas de señal y ruido se llama relación de *señal a ruido de banda angosta*. De nuevo, la portadora de RF, modulada 30% por un tono de 1 kHz, se aplica a la entrada del receptor. A la salida del mismo se mide la potencia total de audio más ruido. A continuación se introduce un filtro rechaza banda de banda angosta de 1 kHz entre la salida del receptor y el medidor de potencia, y se toma otra medida de la potencia. De nuevo, la indicación de potencia sólo incluirá ruido. Sin embargo, esta vez todo el receptor funciona en condiciones casi normales de señal, porque recibía y demodulaba una portadora modulada. Las relaciones de señal a ruido de banda angosta sólo tienen sentido si el filtro rechaza banda tiene un ancho de banda extremadamente bajo, de unos cuantos hertz, e introduce 40 dB o más de atenuación a la señal.

Receptores de AM en circuito integrado lineal

Hoy se consiguen circuitos integrados lineales en un solo CI, con diversos fabricantes, que efectúan todas las funciones de un receptor, excepto filtrado de RF y de FI, y control de volumen. La fig. 4-33 muestra el esquema de un receptor que usa el CI de circuito integrado lineal LM1820 de National Semiconductor Corporation, para radio de AM. Tiene a bordo las etapas de amplificador de RF, mezclador, oscilador local y amplificador de FI. Sin embargo, la selectividad de RF e FI se logra ajustando las bobinas de sintonía en circuitos sintonizados o transformadores, conectados externamente. También, es necesario un amplificador de circuito lineal integrado, como el LM386, y una bocina, para completar un receptor que funcione.

Los receptores de radio con circuito integrado lineal de AM no se usan mucho, porque la reducción del tamaño físico que fue posible al reducir la cantidad de componentes mediante la integración, se eliminó por el tamaño de los componentes externos necesarios para proporcionar limitación de banda y selección de canal. En el futuro próximo se podrán integrar alternativas

de circuitos tanque *LC* y de transformadores de FI, como por ejemplo, filtros de cerámica. También se podrán tener, en el futuro, nuevas configuraciones de receptor, distintas del TRF o de superheterodino, usando tecnología de lazo de fase cerrada. Estos receptores de lazo de fase cerrada sólo necesitarían dos componentes externos: un control de volumen y un control de sintonía de estación.

RECEPTORES DE AM DE DOBLE CONVERSIÓN

Para un buen rechazo de frecuencia imagen, se desea tener una frecuencia intermedia relativamente alta. Sin embargo, para los amplificadores selectivos de alta ganancia que son estables y se neutralicen con facilidad, es necesario una frecuencia intermedia baja. La solución es usar dos frecuencias intermedias. La *primera FI* es una frecuencia relativamente alta para tener un buen rechazo de frecuencia imagen, y la *segunda FI* tiene frecuencia relativamente baja, para facilidad de amplificación. La fig. 4-34 muestra el diagrama de bloques de un receptor de AM de *doble conversión*. La primera FI es de 10.625 MHz, que aparta la frecuencia imagen a 21.25 MHz de la RF deseada. La primera FI tiene conversión descendente inmediata a 455 kHz, y se alimenta a una serie de amplificadores de FI de alta ganancia. La fig. 4-35 muestra los requisitos de filtrado para un receptor de AM de doble conversión.

GANANCIA NETA DEL RECEPTOR

Hasta ahora se ha descrito la ganancia de RF, la de conversión y la de FI. Sin embargo, es probable que la ganancia más importante sea la *ganancia neta del receptor*. Esta ganancia no es más que la relación del nivel de la señal demodulada a la salida del receptor (salida de audio) entre el nivel de señal de radiofrecuencia a la entrada del receptor, o sea, la diferencia entre la intensidad de la señal de audio en dBm y la de la señal de RF, en dBm.

En esencia, la ganancia neta del receptor es la suma, en dB, de todas las ganancias del receptor, menos la suma, en dB, de todas las pérdidas. Entre las pérdidas normales del receptor están la del preselector, del mezclador (es decir, la ganancia de conversión) y las de los detectores. Entre las ganancias están la de RF, de FI y del amplificador de audio. La fig. 4-36 muestra las ganancias y las pérdidas de un radiorreceptor normal.

La ganancia neta del receptor es

$$G_{dB} = \text{ganancias}_{dB} - \text{pérdidas}_{dB}$$

en donde ganancias = ganancia del amplificador de RF + ganancia del amplificador de FI + ganancia del amplificador de audio

pérdidas = pérdida en preselector + pérdida del mezclador + pérdida en detector

Ejemplo 4-8

Para un receptor de AM con nivel de señal de entrada de RF de −80 dBm, con las siguientes ganancias y pérdidas, calcular la ganancia neta del receptor y la intensidad de la señal de audio.

Ganancias: amplificador de RF = 33 dB, amplificador de FI = 47 dB, amplificador de audio = 25 dB

Pérdidas: pérdida en preselector = 3 dB, pérdida en mezclador = 6 dB, pérdida en detector = 8 dB

Solución La suma de las ganancias es

$$33 + 47 + 25 = 105 \text{ dB}$$

La suma de las pérdidas es

$$3 + 6 + 8 = 17 \text{ dB}$$

Entonces, la ganancia neta del receptor es

$$G = 105 - 17 = 88 \text{ dB}$$

y la intensidad o nivel de la señal de audio es

$$-80 \text{ dBm} + 88 \text{ dB} = 8 \text{ dBm}$$

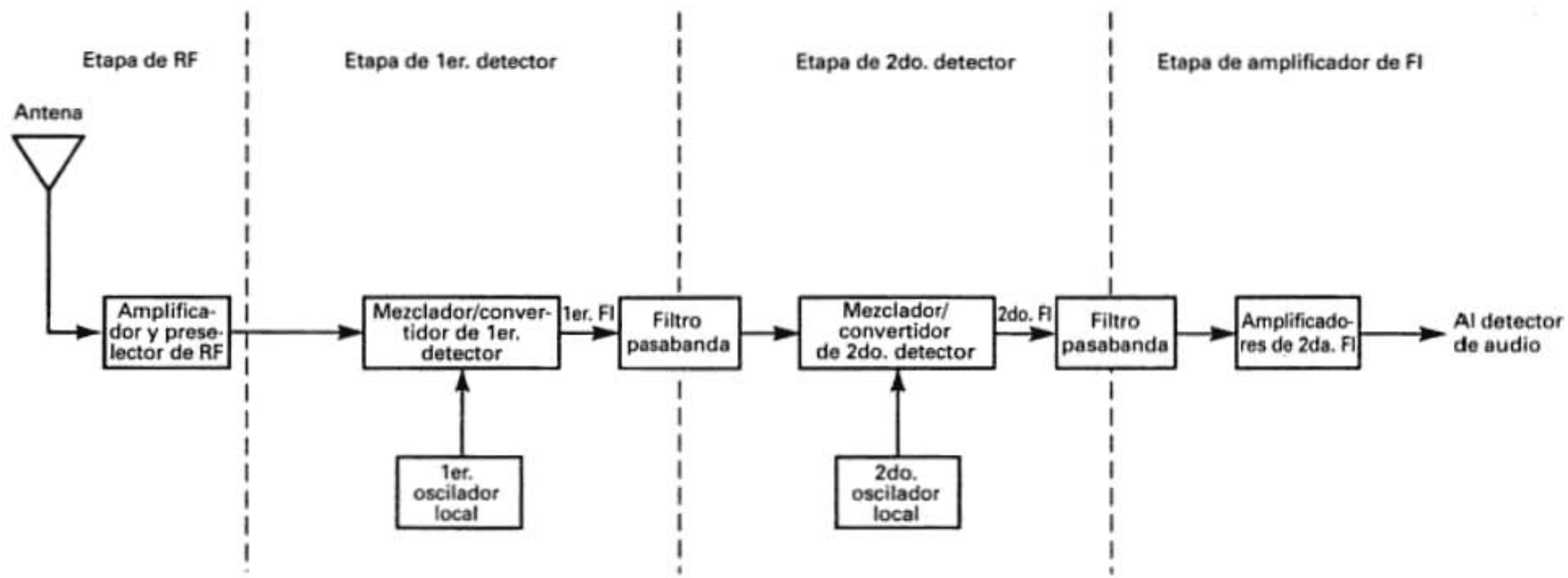


FIGURA 4-34 Receptor de AM de doble conversión

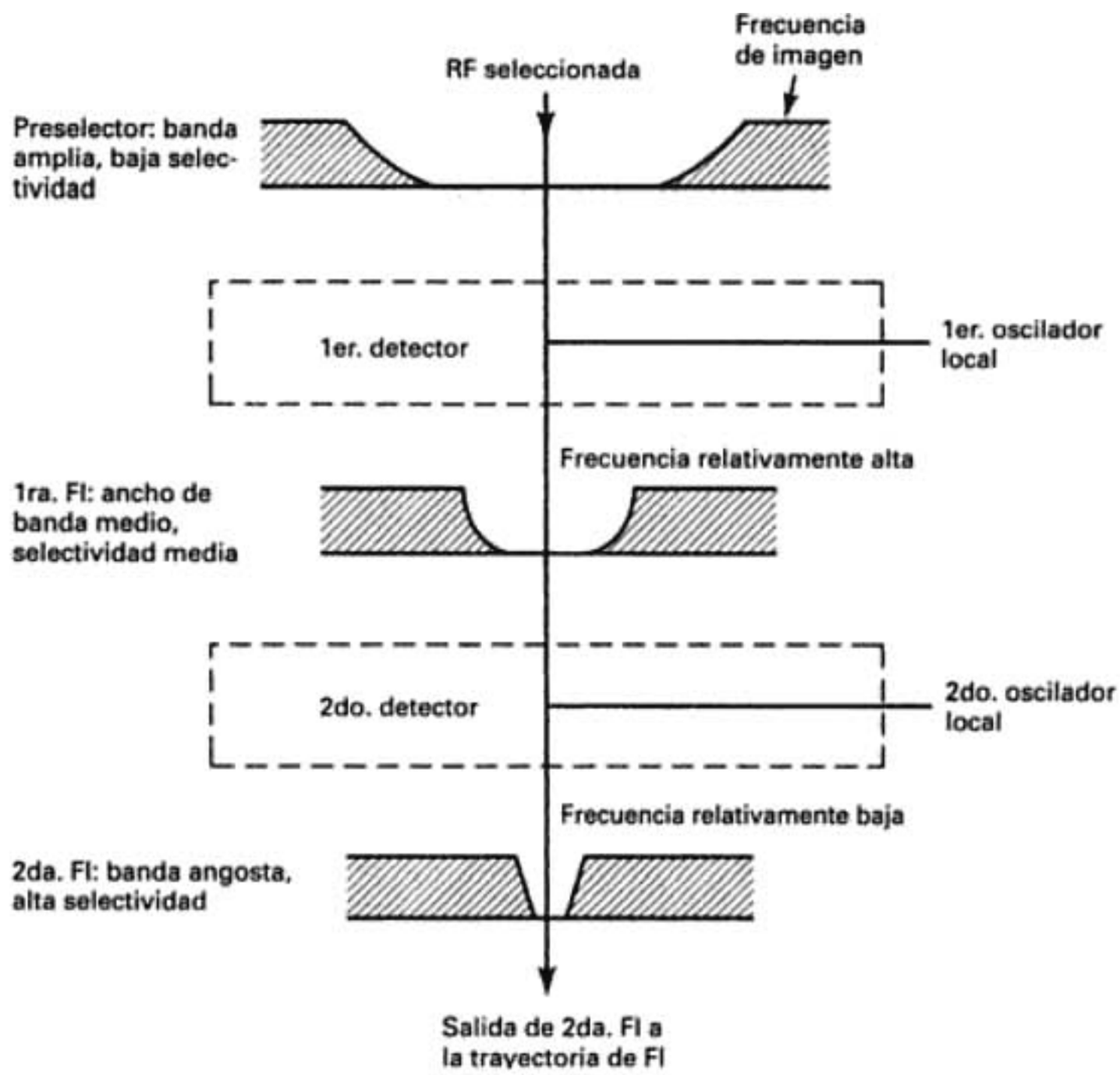


FIGURA 4-35 Requisitos de filtrado en el receptor de AM de doble conversión de la fig. 4-34

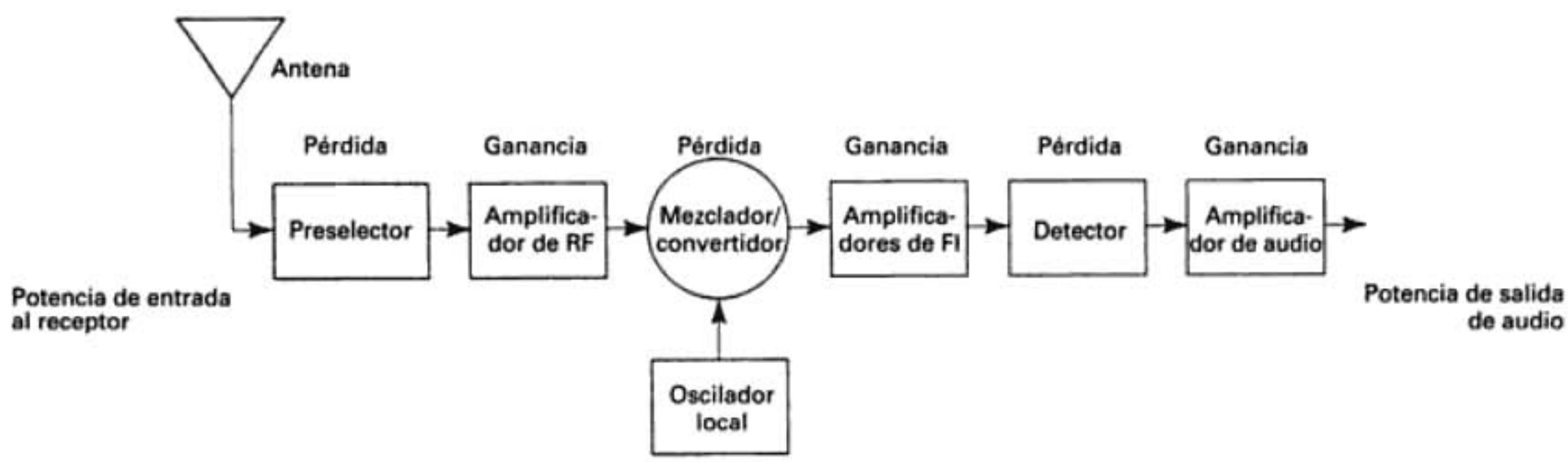


FIGURA 4-36 Ganancias y pérdidas de un receptor

Sin embargo, es importante notar que, debido a los efectos del CAG, la ganancia de FI y/o de RF de un receptor se podrían especificar como valor máximo, mínimo o promedio. Por ejemplo, la ganancia en FI del receptor en el ejemplo 4-8 podría variar, debido al CAG, entre 20 y 60 dB, dependiendo de la intensidad de la señal de entrada. Por consiguiente, la ganancia neta del receptor podría variar entre un valor máximo de 101 dB y uno mínimo de 61 dB, con valor promedio de 81 dB.

No se debe confundir la ganancia neta del receptor con la *ganancia general del sistema*. En la primera sólo se incluyen los componentes dentro del receptor, comenzando en la entrada al preselector. La ganancia del sistema incluye todas las ganancias y las pérdidas que tiene una

señal al propagarse de la etapa de salida del transmisor hasta llegar a la salida del detector en el receptor, e incluye la ganancia de la antena y las pérdidas en líneas de transmisión y por propagación. En el capítulo 17 se describirá con detalle la ganancia del sistema.

PREGUNTAS

4-1. ¿Qué quiere decir *frente* de un receptor?

4-2. ¿Cuáles son las funciones principales del frente de un receptor?

4-3. Defina la *selectividad* y el *factor de forma*. ¿Cuál es la relación entre el ruido del receptor y la selectividad?

4-4. Defina el *mejoramiento del ancho de banda*. ¿Cuál es la relación entre el mejoramiento del ancho de banda y el ruido del receptor?

4-5. Defina la *sensibilidad*.

4-6. ¿Cuál es la relación entre ruido de receptor, ancho de banda y temperatura?

4-7. Defina la *fidelidad*.

4-8. Describa los tres tipos de distorsión que reducen la fidelidad de un receptor.

4-9. Defina la *pérdida de inserción*.

4-10. Defina la *temperatura de ruido* y la *temperatura equivalente de ruido*.

4-11. Describa la diferencia entre un radiorreceptor *coherente* y uno *no coherente*.

4-12. Trace el diagrama de bloques de un radiorreceptor TRF, y describa su funcionamiento en forma breve.

4-13. ¿Cuáles son las desventajas principales de un receptor TRF?

4-14. Trace el diagrama de bloques de un receptor superheterodino de AM y describa su funcionamiento y las funciones principales de cada etapa.

4-15. Defina *heterodinaje*.

4-16. ¿Qué quieren decir los términos *inyección lateral superior* e *inferior*?

4-17. Defina lo que es *rastreo con oscilador local* y *error de rastreo*.

4-18. Describa lo que es *rastreo de tres puntos*.

4-19. ¿Qué quiere decir *sintonía múltiple*?

4-20. Defina qué es *frecuencia imagen*.

4-21. Defina la *relación de rechazo de frecuencia imagen*.

4-22. Haga una lista de las seis características favorables en un amplificador de RF.

4-23. ¿Qué ventaja tienen los amplificadores de RF con FET sobre los de BJT?

4-24. Defina qué es *neutralización*. Describa el proceso de neutralización.

4-25. ¿Qué es un amplificador *cascode*?

4-26. Defina la *ganancia de conversión*.

4-27. ¿Cuál es la ventaja de una frecuencia intermedia relativamente alta? ¿De una frecuencia intermedia relativamente baja?

4-28. Defina los siguientes términos: *acoplamiento inductivo, autoinductancia, inductancia mutua, coeficiente de acoplamiento, acoplamiento crítico* y *acoplamiento óptimo*.

4-29. Describa qué es acoplamiento *débil* y acoplamiento *fuerte*.

4-30. Describa el funcionamiento de un *detector de picos*.

4-31. Describa la *distorsión debida a rectificador* y sus causas.

4-32. Describa el *recorte diagonal* y sus causas.

4-33. Describa los siguientes términos: *CAG simple, CAG demorado, CAG directo*.

4-34. ¿Cuál es la finalidad de un *circuito reductor de ruido*?

4-35. Explique el funcionamiento de un receptor superheterodino de doble conversión.

PROBLEMAS

4-1. Calcule el ancho de banda de FI necesario para lograr un mejoramiento del ancho de banda de 16 dB, en un radiorreceptor con ancho de banda de RF de 320 kHz.

4-2. Calcule el mejoramiento de la cifra de ruido en un receptor con ancho de banda de RF igual a 40 kHz y ancho de banda de FI de 16 kHz.

4-3. Calcule la temperatura equivalente de ruido para un amplificador con cifra de ruido de 6 dB, y una temperatura ambiente $T = 27°$ C.

4-4. Para un receptor de banda comercial de AM, con factor Q del filtro de entrada igual a 85, determine el ancho de banda en los extremos bajo y alto del espectro de RF.

4-5. Para un receptor superheterodino de AM con inyección lateral superior, con una frecuencia de oscilador local de 1200 kHz, calcule las frecuencias de portadora y de lado superior e inferior, con una envolvente de RF formada por una portadora y frecuencias de banda lateral superior e inferior de 600 kHz, 604 kHz y 596 kHz, respectivamente.

4-6. Calcule el ancho de banda mínimo de FI para un receptor con error de rastreo de ± 2.5 kHz, FI de 455 kHz y frecuencia máxima de señal moduladora $f_m = 6$ kHz.

4-7. Para un receptor, las frecuencias de FI, RF y de oscilador local son 455 kHz, 900 kHz y 1355 kHz, respectivamente. Calcule:

(a) La frecuencia imagen.

(b) El IFRR para una Q del preselector igual a 80.

4-8. Un receptor de banda civil tiene inyección lateral superior, con una portadora de RF de 27.04 MHz y primera FI de 10.645 MHz. Calcule:

(a) La frecuencia del oscilador local.

(b) La frecuencia imagen.

4-9. Un amplificador de RF de tres etapas y doble sintonía, tiene portadora de RF de 800 kHz, y su coeficiente de acoplamiento $k_{ópt} = 0.025$. Determine:

(a) El ancho de banda para cada etapa individual.

(b) El ancho de banda general con las tres etapas.

4-10. Calcule la frecuencia máxima de señal moduladora para un detector de picos con los siguientes parámetros: $C = 1000$ pF, $R = 10$ kΩ y $m = 0.5$. Repita el problema con $m = 0.707$.

4-11. Determine el mejoramiento del ancho de banda para un radiorreceptor con ancho de banda de RF de 60 kHz y uno de FI de 15 kHz.

4-12. Determine la temperatura equivalente de ruido para un amplificador con cifra de ruido $F = 8$ dB y una temperatura ambiente $T = 122°$ C.

4-13. Para un receptor de banda comercial de AM con factor Q del filtro de entrada igual a 60, determine el ancho de banda en los extremos bajo y alto del espectro de RF.

4-14. Para un receptor superheterodino de AM que usa inyección lateral superior, con una frecuencia de oscilador local de 1400 kHz, determine las frecuencias de portadora de FI, y las laterales superior e inferior para una envolvente de RF formada por una portadora y frecuencias laterales superior e inferior de 800 kHz, 806 kHz y 794 kHz, respectivamente.

4-15. Determine el ancho mínimo de banda de FI para un receptor con error de rastreo de ± 2800 Hz y una frecuencia máxima de señal moduladora $f_m = 4$ kHz.

4-16. Determine lo siguiente, para un receptor con frecuencias de FI, RF y de oscilador local de 455 kHz, 1100 kHz y 1555 kHz, respectivamente:

(a) Frecuencia imagen.

(b) Relación de rechazo de frecuencia imagen para una $Q = 100$ de preselector.

(c) Relación de rechazo de frecuencia imagen para una $Q = 50$.

4-17. Para un receptor de banda civil con inyección lateral superior, con portadora de RF de 27.04 MHz y FI de 10.645 MHz, determine:

(a) Frecuencia del oscilador local.

(b) Frecuencia imagen.

4-18. Para un amplificador de tres etapas de RF de doble sintonía, con RF de 1000 kHz y un coeficiente de acoplamiento $k_{\text{óp}t} = 0.01$, determine:

(a) Ancho de banda para cada etapa individual.

(b) Ancho de banda para las tres etapas conectadas en cascada.

4-19. Determine la frecuencia máxima de señal moduladora para un detector de picos con los siguientes parámetros: $C = 1000$ pF, $R = 6.8$ kΩ y $m = 0.5$. Repita el problema con $m = 0.707$.

4-20. Determine la ganancia neta de receptor de AM con potencia de señal de entrada de RF de −87 dBm y una potencia de señal de audio de 10 dBm.

4-21. Determine la ganancia neta de un receptor de AM con las siguientes ganancias y pérdidas:

Ganancias: amplificador de RF = 30 dB, amplificador de FI = 44 dB, amplificador de audio = 24 dB.

Pérdidas: pérdidas de preselector = 2 dB, pérdida de mezclador = 6 dB, pérdida de detector = 8 dB.

4-22. Determine la potencia mínima de señal de entrada de RF necesaria para producir una potencia de señal de audio de 10 dBm, para el receptor descrito en el problema 4-21.

4-23. Determine la ganancia neta de un receptor de AM con las siguientes ganancias y pérdidas:

Ganancias: amplificador de RF = 33 dB, amplificador de FI = 44 dB, amplificador de audio = 22 dB.

Pérdidas: pérdidas de preselector = 3.5 dB, pérdida de mezclador = 5 dB, pérdida de detector = 9 dB.

CAPÍ

Imagen protegida por derechos de autor

Sistemas de comunicaciones de banda lateral única

INTRODUCCIÓN

Los sistemas convencionales de comunicaciones de doble banda lateral de AM, como los que se describieron en los capítulos 3 y 4, tienen dos desventajas inherentes. La primera es que con la AM convencional, la potencia de la portadora tiene dos tercios o más de la potencia total transmitida. Es un gran inconveniente, porque la portadora no contiene información; las bandas laterales son las que contienen la información. La segunda es que los sistemas convencionales de AM usan el doble de ancho de banda que el necesario en los sistemas de banda lateral única. Con la transmisión con doble banda lateral, la información contenida en la banda superior es idéntica a la de la banda inferior. Por consiguiente, es una redundancia transmitir ambas bandas laterales, y la AM convencional es ineficiente respecto a la potencia y al ancho de banda, que son las dos consideraciones más importantes al diseñar los sistemas modernos de comunicaciones electrónicas.

El objeto de este capítulo es presentar al lector varios sistemas de AM de banda lateral única, y explicar las ventajas y desventajas de escogerlos, en comparación con la AM convencional, de portadora de máxima potencia y doble banda lateral.

El uso más importante de los sistemas de banda lateral única con supresión de portadora es en comunicaciones de multicanal que emplean multiplexado por división de frecuencia (FDM, de *frequency-division multiplexing*), como son los sistemas telefónicos de larga distancia. Después en el capítulo se describirá el multiplexado por división de frecuencia, y todavía se describirá con más detalle en el capítulo 16.

SISTEMAS DE BANDA LATERAL ÚNICA

Ya desde 1914 se reconoció y comprendió matemáticamente la banda lateral única; sin embargo, no fue sino hasta 1923 que se otorgó la primera patente y se estableció un buen enlace de comunicaciones entre Inglaterra y Estados Unidos. Hay muchas clases distintas de sistemas de comunicaciones de *banda lateral*. Algunos de ellos conservan el ancho de banda, algunos conservan la potencia y otros más conservan ambas cosas. La fig. 5-1 compara los espectros de frecuencia y las distribuciones relativas de potencia en la AM convencional, y en algunos de los sistemas más frecuentes de banda lateral única (SSB, de *single-sideband*).

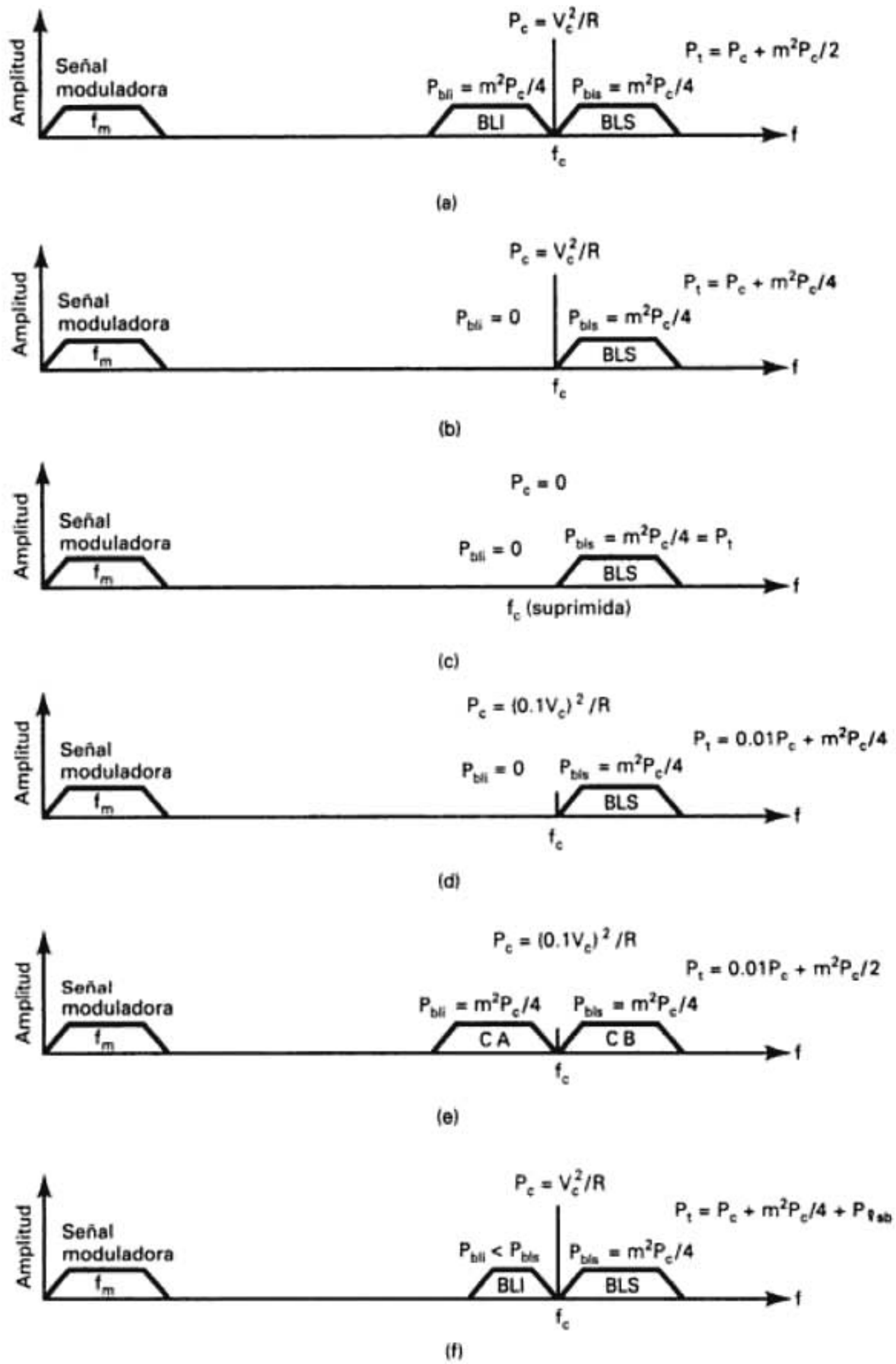


FIGURA 5-1 Sistemas de banda lateral única (SSB): (a) DSBFC de AM convencional; (b) banda lateral única con portadora de máxima potencia; (c) banda lateral única con portadora suprimida; (d) banda lateral única con portadora reducida; (e) banda lateral independiente; (f) banda lateral residual.

AM de banda lateral única y portadora de máxima potencia

La AM de *banda lateral única y portadora de máxima potencia* (SSBFC, de *single-sideband full carrier*) es una forma de modulación de amplitud en la que la portadora se transmite con potencia máxima, pero sólo se transmite una de las bandas laterales. En consecuencia, las transmisiones SSBFC sólo necesitan la mitad del ancho de banda que la AM convencional con doble banda lateral. En la fig. 5-1b se ven el espectro de frecuencias y la distribución relativa de potencia en el sistema

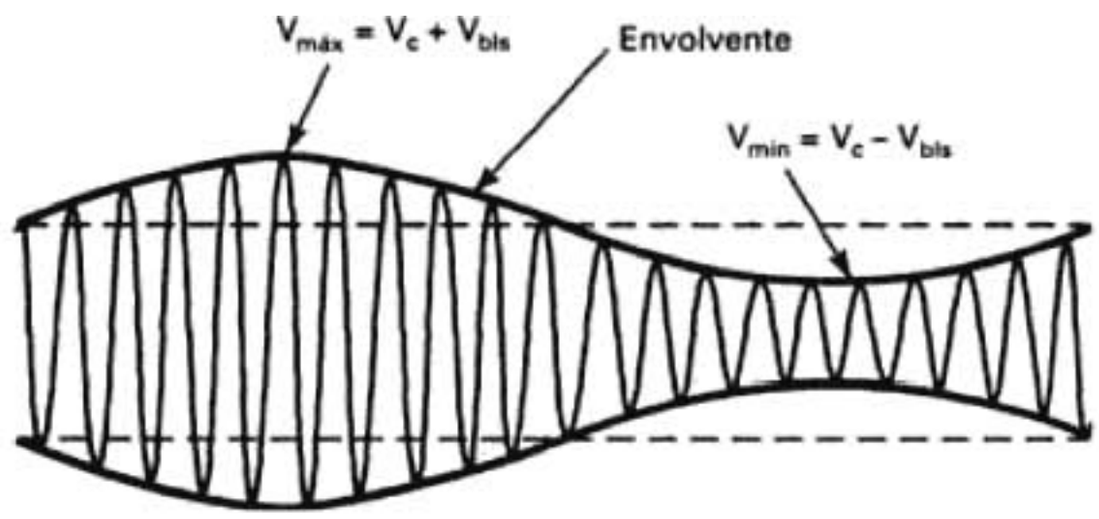


FIGURA 5-2 Forma de onda SSBFC, con 100% de modulación

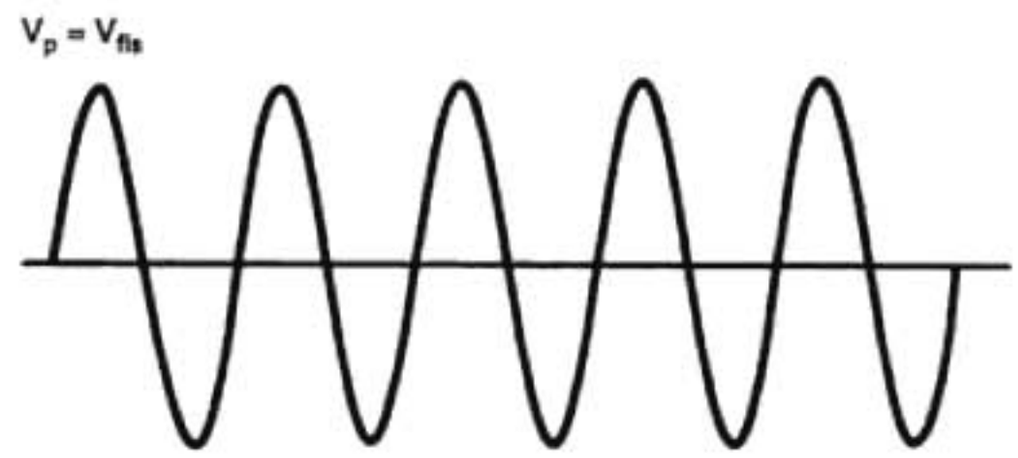


FIGURA 5-3 Forma de onda SSBSC

SSBFC. Nótese que con 100% de modulación, la potencia de la portadora (P_c) constituye las cuatro quintas partes (el 80%) de la potencia total transmitida P_t, y sólo la quinta parte (20%) de la potencia total está en la banda lateral. En la AM convencional con doble banda lateral, dos terceras partes (67%) de la potencia total transmitida está en la portadora y un tercio (33%) está en las bandas laterales. Por lo anterior, aunque en el sistema SSBFC se requiere menos potencia, en realidad se usa menor porcentaje de esa potencia en la parte de la señal que es portadora de información.

La fig. 5-2 muestra la forma de una onda SSBFC modulada 100%, con una sola señal moduladora. La envolvente 100% modulada de portadora de máxima potencia y banda lateral única parece idéntica a una 50% modulada de banda lateral doble y portadora de máxima potencia. Recuérdese que en el capítulo 3 se explicó que los picos máximos positivos y negativos de una onda DSBFC de AM se presentan cuando la portadora y las dos bandas laterales llegan a sus máximos respectivos al mismo tiempo, y el cambio máximo en la envolvente es igual a la suma de las amplitudes de las frecuencias laterales superior e inferior. En la transmisión de banda lateral única, sólo hay una banda lateral, que puede ser la superior o la inferior, que se suma a la portadora. Por consiguiente, el cambio máximo en la envolvente es la mitad del que hay en la transmisión con doble banda lateral. Así, con transmisión de banda lateral única y portadora de máxima potencia, las señales demoduladas sólo tienen la mitad de la amplitud que una onda demodulada con doble banda lateral. Por lo anterior es necesario hacer un compromiso. El SSBFC requiere menos ancho de banda que el DSBFC, pero también produce una señal demodulada con amplitud menor. Sin embargo, cuando baja el ancho de banda a la mitad, también se reduce a la mitad la potencia de ruido (es decir, se reduce 3 dB), y si se elimina una banda lateral, la potencia en la parte de información de la onda también baja a la mitad. En consecuencia, las relaciones de señal a ruido con banda lateral única y doble banda lateral son iguales.

Con el sistema SSBFC, la frecuencia de repetición de la envolvente es igual a la frecuencia de la señal moduladora, y el grado de modulación es proporcional a la amplitud de la señal moduladora. Por lo anterior, como en la transmisión con doble banda lateral, la información está en la envolvente de la señal modulada con portadora de máxima potencia.

AM de banda lateral única y portadora suprimida

La AM *de banda lateral única y portadora suprimida* (SSBSC, de *single-sideband suppressed carrier*) es una forma de modulación de amplitud en la que la portadora se suprime en su totalidad y se quita una de las bandas laterales. Por consiguiente, en el SSBSC se requiere la mitad del ancho de banda que en la AM convencional de doble banda lateral, y bastante menos potencia de transmisión. El espectro de frecuencias y la distribución de potencia relativa en un SSBSC con transmisión de banda lateral superior se ven en la fig. 5-1c. Se puede apreciar que la potencia de la banda lateral es el 100% de la potencia total transmitida. En la fig. 5-3 se ve una forma de onda de SSBSC cuando la señal moduladora es de una sola frecuencia. Se ve que la forma de onda no es una envolvente, simplemente es una senoide de una sola frecuencia, igual a la frecuencia de la portadora más la frecuencia de la señal moduladora, o la frecuencia de la portadora menos la frecuencia de la señal moduladora, dependiendo de cuál banda lateral se transmite.

AM de banda lateral única y portadora reducida

La AM de *banda lateral única y portadora reducida* (SSBRC, de *single-sideband reduced carrier*) es una forma de modulación de amplitud, en la que se quita totalmente una banda lateral y el

voltaje de la portadora se reduce más o menos hasta el 10% de su amplitud no modulada. En consecuencia, hasta el 96% de la potencia total transmitida está en la banda lateral no suprimida. Para producir un componente reducido de portadora, se suprime totalmente la portadora durante la modulación y después se vuelve a insertar con una amplitud reducida. Por lo anterior, a veces se dice que el sistema SSBRC es de banda lateral única y portadora *reinsertada*. Con frecuencia se llama portadora piloto a la portadora reinsertada, y se reinserta para fines de demodulación; eso se explicará más adelante en este capítulo. El espectro de frecuencias y la distribución relativa de potencia para el SSBRC se ven en la fig. 5-1d. Allí se ve que la potencia de la portadora constituye casi el 100% de la potencia transmitida. Como en el caso de la AM de portadora de máxima potencia y doble banda lateral, la rapidez de repetición de la envolvente es igual a la frecuencia de la señal moduladora. Para demodular una forma de onda con portadora reducida mediante un detector convencional de picos, la portadora se debe separar, amplificar y después reinsertar con intensidad mayor en el receptor. En consecuencia, a veces a la transmisión con portadora reducida se le llama portadora *incrementada*, porque se eleva en el receptor antes de la demodulación. Con la detección de portadora incrementada, la amplificación de la portadora en el receptor debe ser suficiente para aumentar su amplitud hasta un valor mayor que el de la señal de banda lateral. En el sistema SSBRC se requiere la mitad del ancho de banda que en la AM convencional y, como la portadora se transmite a un nivel reducido, también economiza mucha potencia.

AM de banda lateral independiente

La *AM de banda lateral independiente* (ISB, de *independent sideband*) es una forma de modulación de amplitud en la que una sola frecuencia de portadora se modula en forma independiente con dos señales moduladoras diferentes. En esencia, el sistema ISB es una forma de transmisión de doble banda lateral en el que el transmisor consiste en dos moduladores independientes de banda lateral única y portadora suprimida. Un modulador sólo produce la banda lateral superior y el otro sólo la banda lateral inferior. Las señales de salida de banda lateral única procedentes de los dos moduladores se combinan para formar una señal de doble banda lateral, en la que las dos bandas laterales son totalmente independientes entre sí, excepto que son simétricas respecto a una frecuencia portadora común. Una banda lateral está arriba de la portadora, en el espectro de frecuencias, y una está abajo. Para fines de demodulación, se reinserta la portadora generalmente con un nivel reducido, como en la transmisión SSBRC. La fig. 5-1e muestra el espectro de frecuencias y la distribución de potencia en ISB, y la fig. 5-4 muestra la forma de onda transmitida para dos señales de información independientes, de una frecuencia (f_{m1} y f_{m2}). Las dos señales de información tienen la misma frecuencia y, por consiguiente, la forma de onda es idéntica a la de una de doble banda lateral y portadora suprimida, pero con una rapidez de repetición igual a dos veces la frecuencia de la señal moduladora. En el ISB se conservan tanto la potencia de transmisión como el ancho de banda, porque se transmiten dos fuentes de información dentro del mismo espectro de frecuencias que se necesitaría para una sola fuente con el uso de transmisión convencional de doble banda lateral. El ISB es una técnica que se usa en Estados

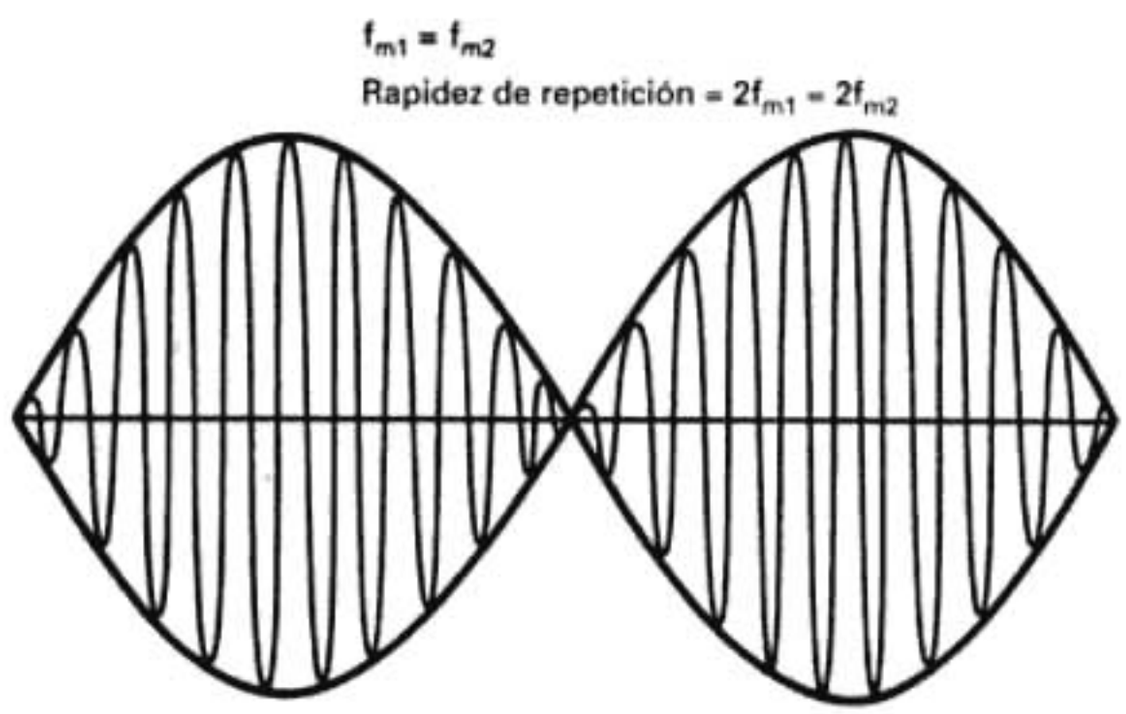


FIGURA 5-4 Forma de onda ISB

Unidos para transmisión estereofónica en AM. Un canal (el izquierdo) se transmite en la banda lateral inferior, y el otro (el derecho) en la banda lateral superior.

AM de banda lateral residual

La AM de *banda lateral residual* (VSB, de *vestigial sideband*) es una forma de modulación de amplitud donde se transmiten la portadora y una banda lateral de máxima potencia, pero sólo una parte de la segunda banda lateral. La portadora se transmite con potencia máxima. En el sistema VSB, las señales moduladoras de menor frecuencia se transmiten en doble banda lateral, y las de mayor frecuencia se transmiten en banda lateral única. En consecuencia, las frecuencias menores pueden tener el 100% de modulación, mientras que las mayores no pueden alcanzar efecto mayor que el 50% de modulación. Así, las frecuencias moduladoras menores se hacen destacar y producen señales de mayor amplitud en el demodulador que las altas frecuencias. El espectro de frecuencias y la distribución relativa de potencias en el sistema VSB se ven en la figura 5-1f. Es probable que el sistema VSB más conocido es el de la parte de imagen en las señales de televisión comercial, la FCC designa a ese sistema como A5C.

Comparación de la transmisión con banda lateral única y la AM convencional

Según la descripción anterior y la fig. 5-1, se puede ver que los ahorros de ancho de banda y de potencia son ventajas obvias de la transmisión con banda lateral única y portadora suprimida o reducida, en comparación con la transmisión convencional con doble banda lateral y portadora de máxima potencia (es decir, la AM convencional). En la transmisión de banda lateral única sólo se requiere la mitad del ancho de banda que el de la banda lateral doble, y en las transmisiones con portadora suprimida y reducida se requiere bastante menos potencia total de transmisión que en la AM de portadora de máxima potencia.

La potencia total transmitida necesaria para producir determinada relación de señal a ruido en la salida de un receptor es un medio cómodo y útil para comparar los requisitos de potencia y la eficiencia relativa de los sistemas de banda lateral única y los convencionales de AM. La relación de señal a ruido determina el grado de inteligibilidad de una señal recibida.

La fig. 5-5 es un resumen de las formas de onda producida para determinada señal moduladora, con tres de los sistemas más comunes de transmisión por AM: doble banda lateral y portadora de máxima potencia (DSBFC), doble banda lateral y portadora suprimida (DSBSC) y banda lateral única, portadora suprimida (SSBSC). Como se ve en la figura, la rapidez de repetición en la envolvente de DSBFC es igual a la frecuencia de la señal moduladora; en la envolvente de DSBSC es igual al doble de la señal moduladora, y en la forma de onda de SSBSC no es portadora alguna, sino más bien una senoide de una sola frecuencia, igual a la frecuencia de la banda lateral no suprimida, es decir, la frecuencia lateral superior o inferior.

Una onda de AM convencional con 100% de modulación contiene 1 unidad de potencia de portadora y 0.25 unidades de potencia en cada banda lateral, lo que hacen un total de 1.5 unidades de potencia total transmitida máxima. Un transmisor de banda lateral única con potencia nominal de 0.5 unidades produce la misma relación de señal a ruido a la salida del receptor que 1.5 unidades de potencia de portadora más banda lateral en una señal de doble banda lateral y portadora de máxima potencia. En otras palabras, se tiene la misma eficiencia con el sistema SSBSC usando sólo un tercio de la potencia transmitida y la mitad del ancho de banda. En la tabla 5-1 se compara la AM convencional con la de banda lateral única y portadora suprimida, para una señal moduladora de frecuencia única. La *potencia máxima de envolvente* (PEP, de *peak envelope power*) es la potencia rms desarrollada en la cresta de la envolvente de modulación, es decir, cuando los componentes de señal moduladora tienen sus amplitudes máximas.

También se ven los vectores de voltaje para las potencias requeridas mencionadas. Se ve que en la AM convencional se requiere 0.5 unidad de voltaje por banda lateral, y 1 unidad para la portadora, haciendo un total de 2 volts máximos de envolvente (PEV, de *peak envelope volts*), y sólo 0.707 PEV para la banda lateral única. También se muestran las envolventes de RF, que corresponden a las relaciones de voltaje y potencia descritas arriba. La señal demodulada en la salida de un receptor convencional de AM es proporcional a la suma de cuadrados de los voltajes de las señales de banda lateral superior e inferior, que es igual a 1 unidad de PEV. Para

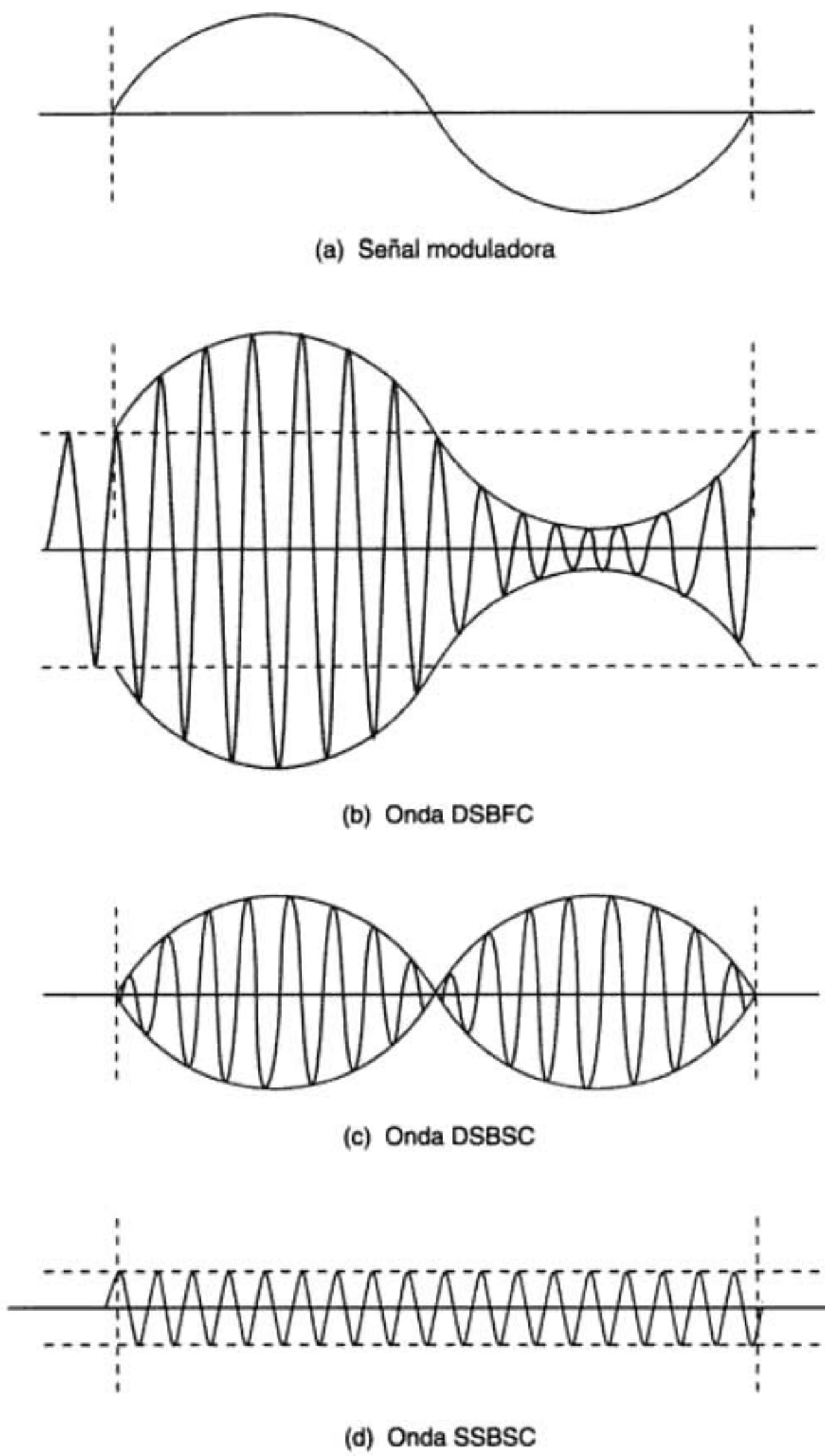


FIGURA 5-5 Comparación de tres sistemas comunes de transmisión de AM: (a) señal moduladora; (b) onda DSBFC; (c) onda DSBSC; (d) onda SSBSC

recepción con banda lateral única, la señal demodulada tiene $0.707 \times 1 = 0.707$ PEV. Si se escoge arbitrariamente que el voltaje de ruido para la AM convencional sea 0.1 V/kHz, el voltaje de ruido para la señal de banda lateral única con la mitad del ancho de banda, es 0.0707 V/kHz. En consecuencia, la eficiencia del sistema SSBSC en cuanto a señal/ruido, es igual a la de la AM convencional.

Ventajas de la transmisión con banda lateral única. A continuación se describen las cuatro ventajas principales de la transmisión con banda lateral única y portadora suprimida o reducida, respecto a la transmisión con portadora de máxima potencia y doble banda lateral.

TABLA 5-1 Comparación de AM convencional y banda lateral única

Potencia nominal (en unidades)	$P_t = 1.0 + 0.5 = 1.5$ 1; 0.25; 0.25 BLI C BLS	$P_t = 0.5 = PEP$ 0.5 C BLS
Vector de voltaje (Mod. 100%)	BLI BLS; 0.5 0.5; C 2; 1; C	PEV = 0.707 0.707 BLS
Señal de RF	PEV = 2 PEP = 4	PEV = 0.707 PEP = 0.5
Señal demodulada de información	BLS + BLI = 1	0.707
Voltaje arbitrario de ruido por kHz de ancho de banda	0.1 V/kHz	0.0707 V/kHz
Relación S/N (señal a ruido) = 20 log S/N	$20 \log \frac{1}{0.1} = 20$ dB	$20 \log \frac{0.707}{0.0707} = 20$ dB

Ahorro de energía. En el caso normal, con transmisión de banda lateral única, sólo se transmite una banda lateral, y la portadora se suprime o se reduce bastante. En consecuencia, se necesita mucho menor potencia transmitida para producir esencialmente la misma calidad de señal en el receptor que la que se alcanza con la transmisión con portadora de máxima potencia y doble banda lateral. En la portadora están al menos las dos terceras partes de la potencia de una señal normal de AM con doble banda lateral y portadora de máxima potencia, y la potencia máxima contenida en cualquiera de las bandas laterales sólo es la sexta parte de la potencia total.

Ahorro de ancho de banda. En la transmisión de banda lateral única se requiere la mitad de ancho de banda que en la trasmisión convencional de AM con doble banda lateral. Esto es una ventaja de bastante importancia en la actualidad, cuando el espectro de radiofrecuencias ya está sobresaturado. Al eliminar una banda lateral se reduce en realidad el ancho de banda necesario en un factor mayor que dos, porque la mayoría de las señales moduladoras, incluyendo las de audio, casi nunca llegan a 0 Hz (cd). Un límite inferior más práctico para las señales de audio es el de 300 Hz; así, un canal de audio de 3 kHz en realidad tiene un ancho de banda aproximado de 2700 Hz (300 a 3000 Hz). En consecuencia, un canal de audio de 2700 Hz

transmitido por un sistema de AM de doble banda lateral requeriría un ancho de banda de 6 kHz, mientras que la misma información de audio sólo requeriría 2700 Hz de ancho de banda, con un sistema de banda lateral única. Por consiguiente, el sistema con banda lateral única que se describió tiene un mejoramiento de ancho de banda igual a 10 log (6000/2700), o sea una reducción de 3.5 dB en la potencia de ruido. Una aproximación general del lado de la seguridad es que el ancho de banda se reduce 50% con la banda lateral única, respecto a la banda lateral doble, y eso significa un mejoramiento de 3 dB en la relación de señal a ruido.

Si se combina el mejoramiento de ancho de banda logrado al transmitir sólo una banda lateral, y la ventaja de eliminar la portadora, en términos de potencia, el mejoramiento general en la relación de señal a ruido al usar portadora suprimida y banda lateral única, es más o menos 7.8 dB (3 + 4.8), mejor que con portadora de máxima potencia y doble banda lateral.

Desvanecimiento selectivo. En la transmisión con doble banda lateral, las dos bandas laterales y la portadora se pueden propagar con distintas trayectorias a través de los medios de transmisión y, en consecuencia, pueden sufrir distintos deterioros por transmisión. A este caso se le llama *desvanecimiento selectivo.* Un tipo de desvanecimiento selectivo se llama *desvanecimiento de banda lateral.* En ese caso se atenúa mucho una banda lateral. Esta pérdida da como resultado menor amplitud de señal a la salida del demodulador del receptor y, en consecuencia, una relación de señal a ruido menor en 3 dB. Esa pérdida causa cierta distorsión, pero no es muy perjudicial a la señal, porque las dos bandas laterales contienen la misma información.

La forma más común y más grave de desvanecimiento selectivo es el *desvanecimiento de amplitud de la portadora.* La reducción de la intensidad de la portadora en una onda 100% modulada causa que el voltaje de portadora sea menor que la suma vectorial de las dos bandas laterales. En consecuencia, la envolvente parece la de una sobremodulación y causa distorsión grave en la señal demodulada.

Una tercera causa de desvanecimiento selectivo es el desplazamiento de fase de la portadora o de la banda lateral. Cuando cambian las posiciones relativas de los vectores de la portadora y de la banda lateral en la señal recibida, se produce un cambio notable en la forma de la portadora, que causa una señal demodulada muy distorsionada.

Cuando se transmite sólo una banda lateral con portadora reducida o suprimida, no se pueden presentar el desplazamiento de fase ni el desvanecimiento de la portadora, y el desvanecimiento de la banda lateral sólo cambia la respuesta de amplitud y frecuencia de la señal demodulada. En general, estos cambios no producen distorsión suficiente para causar pérdida de inteligibilidad en la señal recibida. En la transmisión con banda lateral única no es necesario mantener determinada relación de amplitud o fase entre las señales de portadora y de banda lateral.

Reducción de ruido. En vista de que un sistema de banda lateral única usa la mitad del ancho de banda que la AM convencional, la potencia de ruido térmico se reduce a la mitad de la de un sistema con doble banda lateral. Si se toma en cuenta tanto la reducción de ancho de banda como la inmunidad frente al desvanecimiento selectivo, los sistemas SSB tienen una ventaja en relación de señal a ruido de unos 12 dB respecto a la AM convencional; es decir, un sistema convencional de AM debe transmitir una señal 12 dB más potente, para alcanzar el mismo desempeño que un sistema comparable de banda lateral única.

Desventajas de la transmisión con banda lateral única. A continuación se describen dos desventajas principales de la transmisión con banda lateral única y portadora reducida o suprimida, en comparación con la transmisión con doble banda lateral y portadora de máxima potencia.

Receptores complicados. Los sistemas de banda lateral única requieren receptores más complejos y costosos que los convencionales de AM, porque la mayoría de las transmisiones con banda lateral única tienen portadora reducida o suprimida. Por lo anterior, no se puede usar detección de envolvente a menos que se regenere la portadora a un nivel suficiente. Los receptores de banda lateral única requieren un circuito de recuperación y sincronización de la portadora, como por ejemplo un sintetizador de frecuencias PLL, que aumenta su costo, complejidad y tamaño.

Dificultades de sintonización. Los receptores de banda lateral única requieren una sintonización más compleja y precisa que los receptores convencionales de AM. Esto es una desventaja para el usuario promedio. Esta desventaja se puede superar usando circuitos de sintonía más exactos, complejos y costosos.

ANÁLISIS MATEMÁTICO DE AM CON PORTADORA SUPRIMIDA

Un modulador de AM es un *modulador de producto*; la señal de salida es el producto de la señal moduladora y de la portadora. En esencia, la portadora se multiplica por la señal moduladora. Se presentó la ecuación 3-9c como sigue

$$v_{am}(t) = [1 + m \operatorname{sen}(2\pi f_m t)]\,[E_c \operatorname{sen}(2\pi f_c t)]$$

en donde $1 + m \operatorname{sen}(2\pi f_m t)$ = voltaje constante + señal moduladora
$E_c \operatorname{sen}(2\pi f_c t)$ = voltaje de portadora no modulada

Si se elimina el componente constante de la señal moduladora, entonces

$$v_{am}(t) = [m \operatorname{sen}(2\pi f_m t)]\,[E_c \operatorname{sen}(2\pi f_c t)]$$

Al hacer la multiplicación se obtiene

$$v_{am}(t) = -\frac{mE_c}{2}\cos[2\pi(f_c + f_m)t] + \frac{mE_c}{2}\cos[2\pi(f_c - f_m)t]$$

en donde $-(mE_c/2)\cos[2\pi(f_c + f_m)t]$ = componente de frecuencia de lado superior
$+(mE_c/2)\cos[2\pi(f_c - f_m)t]$ = componente de frecuencia de lado inferior

Se puede ver en la operación matemática anterior que, si se quita el componente constante antes de hacer la multiplicación, el componente de portadora se quita de la onda modulada, y la señal de salida está formada tan sólo por dos ondas cosenoidales, una en la frecuencia de suma ($f_c + f_m = f_{bls}$) y la otra en la frecuencia de diferencia ($f_c - f_m = f_{bli}$). En el modulador se ha suprimido la portadora. Para convertir a banda lateral única sólo se quita la frecuencia de suma o la de diferencia.

GENERACIÓN DE BANDA LATERAL ÚNICA

En las secciones anteriores se demostró que en la mayoría de los sistemas de banda lateral única, la portadora se suprime totalmente o se reduce a tan sólo una fracción de su valor original, y se quita una banda lateral. Es en extremo difícil, si no imposible, quitar la portadora de la onda modulada, o reducir su amplitud, usando filtros convencionales de banda angosta, porque esos filtros simplemente no tienen factores Q suficientes como para quitar la portadora sin quitar también una parte de la banda lateral. Sin embargo, también se indicó que al quitar la componente constante se suprimía la portadora en el modulador mismo. En consecuencia, se han desarrollado circuitos moduladores que en forma inherente quitan la portadora durante el proceso de modulación. Esos circuitos se llaman *moduladores de portadora suprimida y doble banda lateral* (DSBSC, de *double-sideband suppressed-carrier*). Se mostrará más adelante en este capítulo cómo se puede quitar una de las bandas laterales una vez que se ha suprimido la portadora.

Un circuito que produce una señal DSBSC es un *modulador balanceado*. El modulador balanceado ha llegado a ser rápidamente uno de los circuitos más útiles y muy usados en las comunicaciones electrónicas. Además de usarse en los sistemas AM con portadora suprimida, se usan mucho en sistemas de modulación de fase y frecuencia, así como en los sistemas de modulación digital, como por ejemplo en la modulación por conmutación fase y en la modulación de amplitud en cuadratura.

Modulador de anillo balanceado

Las figs. 5-6 y 5-7 muestran los diagramas y las formas de onda de un *modulador de anillo balanceado*. El esquema de la fig. 5-6a se forma con diodos y transformadores. Los diodos de semiconductor son ideales para usarse en los circuitos de modulador balanceado, porque son estables, no requieren fuente externa de energía, tienen larga vida y virtualmente no requieren mantenimiento alguno. El modulador de anillo balanceado se llama a veces *modulador de red balanceada* o simplemente *modulador balanceado*. Un modulador balanceado tiene dos entradas: una portadora de frecuencia única y la señal moduladora, que puede ser de una frecuencia o una

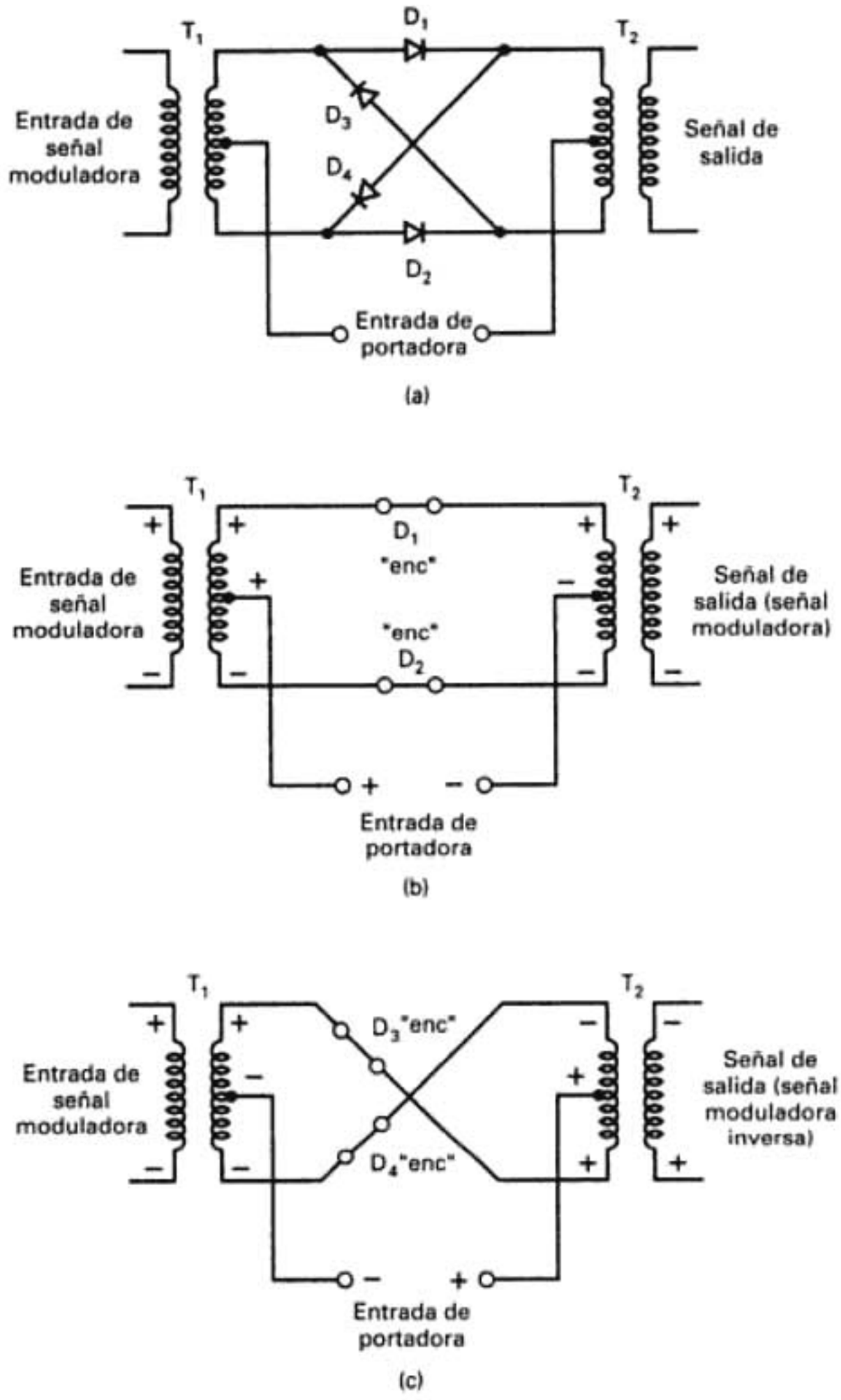


FIGURA 5-6 Modulador de anillo balanceado: (a) diagrama; (b) D_1 y D_2 polarizados directamente; (c) D_3 y D_4 polarizados directamente

forma compleja de onda. Para que trabaje bien un modulador balanceado, la amplitud de la portadora debe ser mayor que la amplitud de la señal moduladora, de seis a siete veces mayor. Así se asegura que la portadora sea la que controle el estado activado o desactivado (encendido o apagado) de los cuatro conmutadores de diodo (D_1 a D_4), y no la señal moduladora.

Funcionamiento del circuito. En esencia, los diodos D_1 a D_4 son conmutadores electrónicos que controlan si pasa la señal moduladora del transformador de entrada T_1 al transformador de salida T_2, y está encendida con 180° de desplazamiento de fase. Con la polaridad de la portadora que se indica en la fig. 5-6b, los interruptores de diodo D_1 y D_2 tienen polarización directa y están encendidos, mientras que los interruptores de diodo D_3 y D_4 tienen polarización inversa y están apagados. En consecuencia, la señal moduladora pasa a través de los interruptores cerrados a T_2 sin invertir su fase. Cuando se invierte la polaridad de la portadora, como se ve en

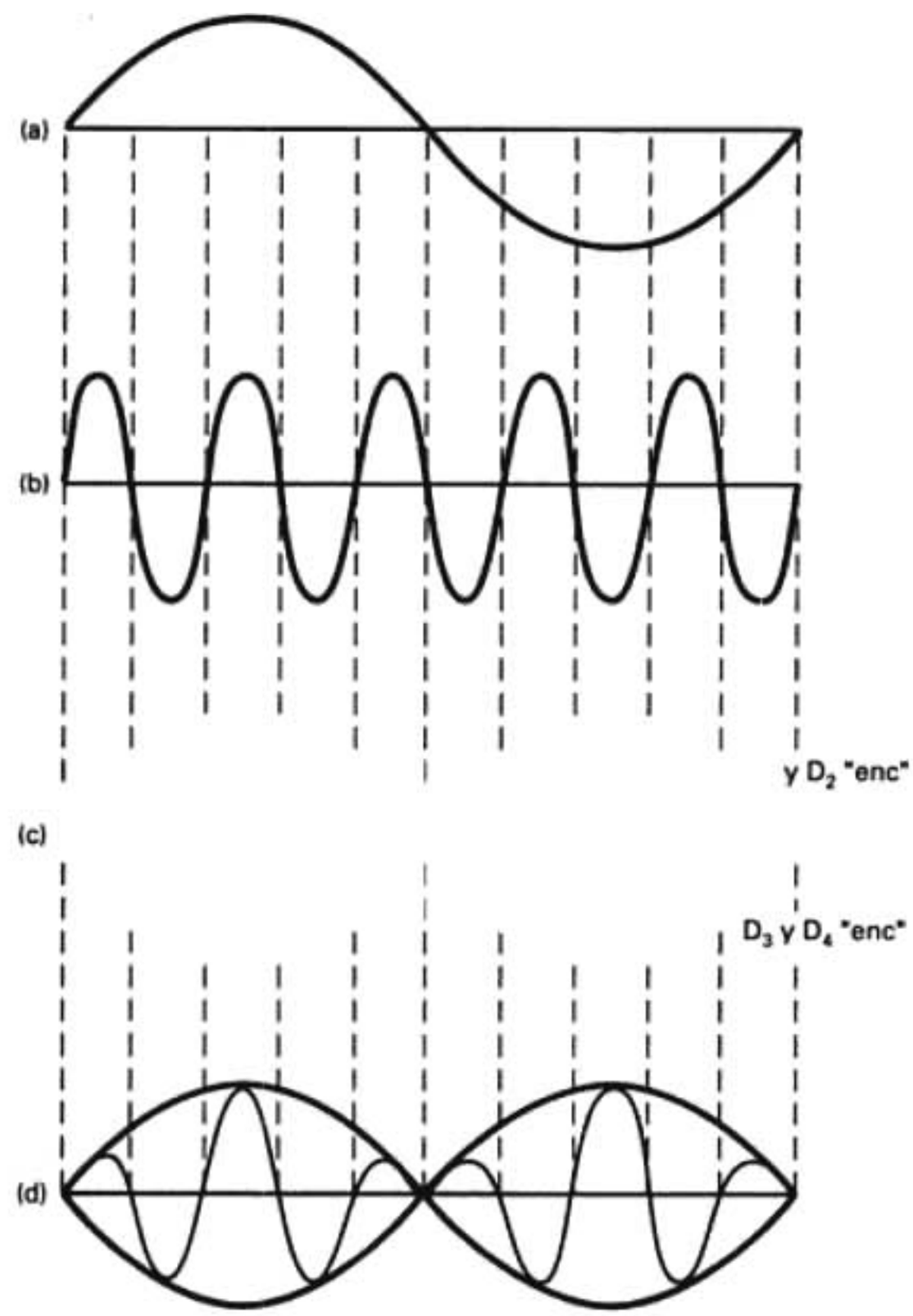


FIGURA 5-7 Formas de onda en un modulador balanceado: (a) señal moduladora; (b) señal portadora; (c) onda de salida antes de filtrarla; (d) onda de salida después de filtrada

la fig. 5-6c, los interruptores de diodo D_1 y D_2 tienen polarización inversa y están apagados, mientras que los interruptores de diodo D_3 y D_4 tienen polarización directa y están encendidos. En consecuencia, la señal moduladora sufre una inversión de fase de 180° antes de llegar a T_2. La corriente de la portadora pasa desde su fuente hasta las tomas centrales de T_1 y T_2, donde se divide y va en direcciones opuestas a través de las mitades superior e inferior de los transformadores. Así, sus campos magnéticos se anulan en los devanados secundarios del transformador, y se suprime la portadora. Si los diodos no están perfectamente equilibrados, o si las tomas de los transformadores no están exactamente en el centro, el circuito está fuera de balance y la portadora no se suprime totalmente. Es virtualmente imposible alcanzar un balance perfecto y, por tanto, siempre hay un pequeño componente de portadora en la señal de salida. A esto se le suele llamar *fuga de portadora*. La cantidad normal de supresión de portadora va de 40 a 60 dB.

La fig. 5-7 muestra las formas de onda de entrada y salida asociadas con un modulador balanceado, para una señal moduladora de una sola frecuencia. Se ve que D_1 y D_2 sólo conducen durante los medios ciclos positivos de la señal portadora de entrada, y que D_3 y D_4 sólo conducen durante los medios ciclos negativos. La salida de un modulador balanceado consiste en una serie de impulsos de RF, cuya frecuencia de repetición queda determinada por la frecuencia de conmutación de la portadora de RF, y la amplitud se controla con el nivel de la señal moduladora. En consecuencia, la forma de onda de salida toma la forma de la señal moduladora, excepto con polaridades alternas positivas y negativas que corresponden a la polaridad de la señal portadora.

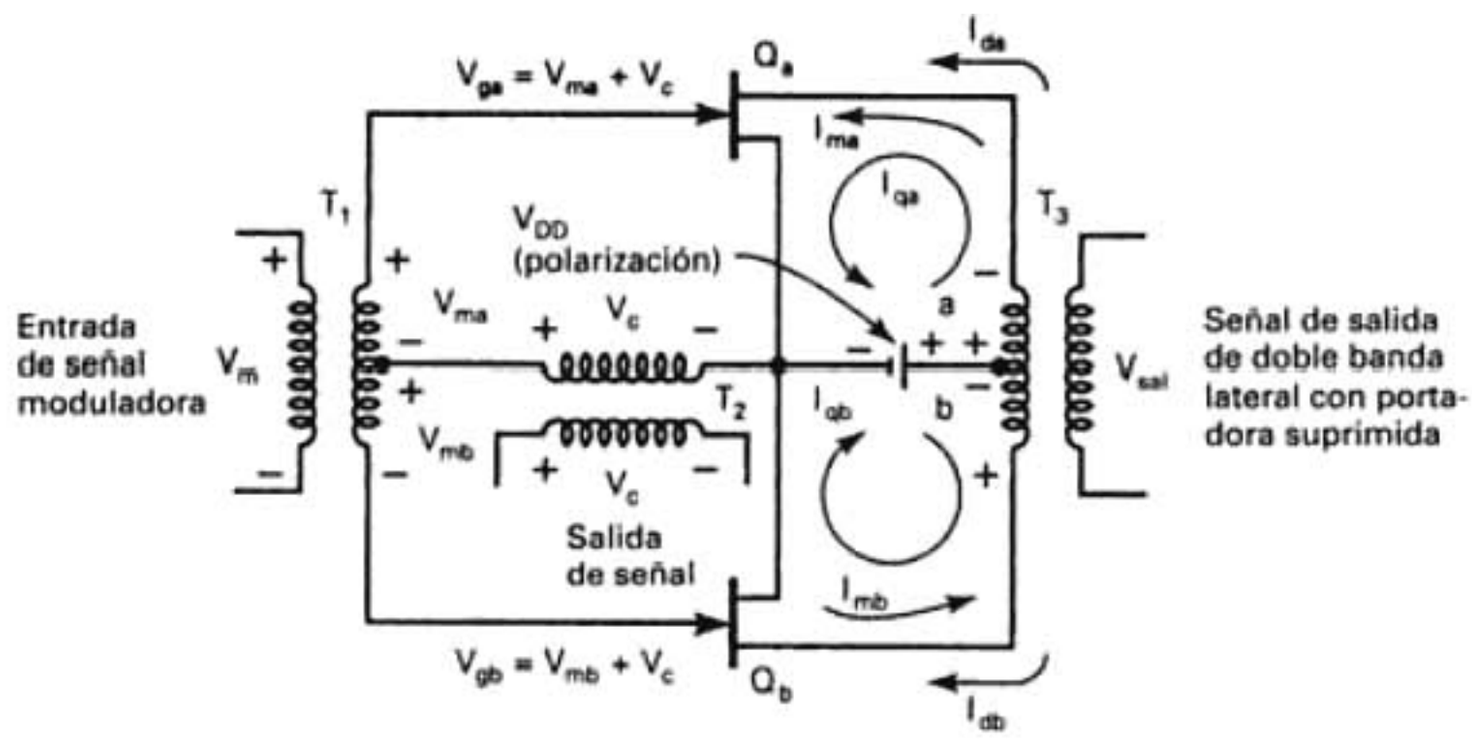


FIGURA 5-8 Modulador balanceado con FET. Para las polaridades indicadas

$$I_{ta} = I_{qa} + I_{da} + I_{ma}$$
$$I_{tb} = -I_{Qb} - I_{db} + I_{mb}$$
$$I_t = I_{ma} + I_{mb} = -2I_m$$

V_{SAL} es proporcional a la corriente modulante (I_{ma} e I_{mb})

Modulador balanceado en contrafase (push-pull) con FET

La fig. 5-8 muestra un diagrama de un modulador balanceado que usa FET y no diodos, como elementos no lineales. Un FET es un dispositivo no lineal que tiene propiedades de ley cuadrática y sólo produce frecuencias de producto cruzado de segundo orden. Como un modulador de diodo balanceado, uno de FET es modulador de producto, y sólo produce las bandas laterales en su salida, y suprime la portadora. El modulador balanceado de FET se parece a un amplificador normal en contrafase, excepto que el circuito modulador tiene dos entradas (la portadora y la señal moduladora).

Funcionamiento del circuito. La portadora entra al circuito de tal manera que se aplica en forma simultánea y en fase a las compuertas de ambos amplificadores FET (Q_1 y Q_2). La portadora produce corrientes en las mitades superior e inferior del transformador de salida T_3, iguales en magnitud, pero desfasadas 180°. Por consiguiente, se anulan y no aparece componente de portadora en la forma de onda de salida. La señal moduladora se aplica al circuito de tal manera que entra en forma simultánea a las compuertas de los dos FETs, 180° fuera de fase. La señal moduladora causa un aumento en la corriente de drenaje en un FET, y una disminución de la corriente de drenaje en el otro FET.

La fig. 5-9 muestra el diagrama fasorial de las corrientes producidas en el transformador de salida de un modulador balanceado con FET. En la fig. 5-9a se ve que las corrientes cd de drenaje en reposo de Q_a y Q_b (I_{qa} e I_{qb}) pasan por sus respectivas mitades del devanado primario de T_3 desfasadas 180° entre sí. La fig. 5-9a también muestra que un aumento en la corriente de drenaje, debido a la señal de portadora (I_{da} e I_{db}) se suma a la corriente en reposo en ambas mitades de los devanados del transformador, y produce corrientes iguales, I_{qa} e I_{qb}, que son sólo la suma de las corrientes de reposo y de portadora. Las corrientes I_{qa} e I_{qb} son iguales, pero tienen direcciones opuestas; en consecuencia se anulan entre sí. La fig. 5-9b muestra la suma fasorial de las corrientes de reposo y de portadora cuando las corrientes de portadora van en dirección contraria a las corrientes de reposo. Las corrientes totales en ambas mitades de los devanados siguen siendo de igual magnitud, pero ahora son iguales a la diferencia entre las corrientes de reposo y de portadora. La fig. 5-9c muestra el diagrama fasorial cuando se suma un componente de corriente debido a una señal moduladora. Las corrientes de señal moduladora, I_{ma} e I_{mb}, producen corrientes en sus respectivas mitades del transformador de salida, que están en fase entre sí. Sin embargo, se puede ver que en una mitad de los devanados, la corriente total es igual a la diferencia entre las corrientes de cd y de portadora, y la corriente de señal moduladora,

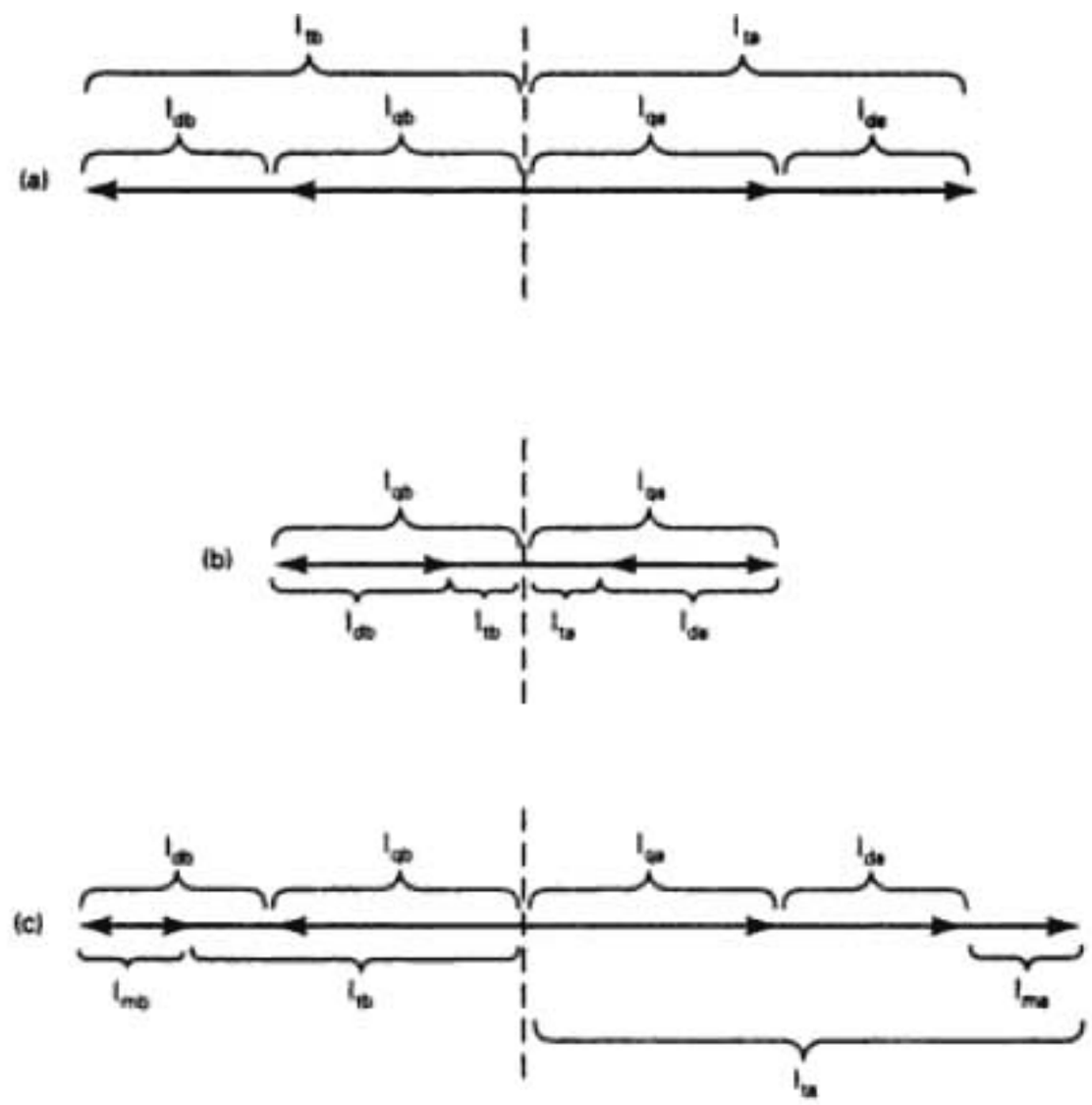


FIGURA 5-9 Diagramas fasoriales de un modulador balanceado con FET: (a) suma enfasada de las corrientes cd y de portadora; (b) suma desfasada de cd y corrientes de portadora; (c) suma de las corrientes de cd, portadora y señal moduladora

y en la otra mitad del devanado, la corriente total es igual a la suma de las corrientes de cd, de portadora y de señal moduladora. Así, las corrientes cd y de portadora se anulan en los devanados secundarios, mientras que los componentes de diferencia se suman. Las corrientes de portadora y de señal moduladora, en cambio continuo, producen las frecuencias de producto cruzado.

Las polaridades de las señales de portadora y moduladora que muestra la fig. 5-8 producen una corriente de salida que es proporcional a los voltajes de esas señales. La señal de portadora (V_c) produce una corriente en ambos FET (I_{da} e I_{db}) que tiene la misma dirección que las corrientes en reposo (I_{qa} e I_{qb}). La señal moduladora (V_{ma} y V_{mb}) produce una corriente I_{ma} en Q_a que tiene la misma dirección que I_{da} e I_{qa}, y una corriente I_{mb} en Q_b que tiene dirección contraria a la de I_{db} e I_{qb}. Por lo anterior, la corriente total a través del lado a de T_3 es $I_{ta} = I_{da} + I_{qa} + I_{ma}$, y la corriente total por el lado b de T_3 es $I_{tb} = -I_{db} - I_{qb} + I_{mb}$. Así, la corriente neta que pasa por el devanado primario de T_3 es $I_{ta} + I_{tb} = I_{ma} + I_{mb}$. Para una señal moduladora con polaridad opuesta, la corriente de drenaje en Q_b aumenta, y disminuye en Q_a. Sin tener en cuenta la corriente cd en reposo, I_{qa} e I_{qb}, la corriente de drenaje en un FET es la suma de las corrientes de portadora y de señal moduladora, $I_d + I_m$, y la corriente de drenaje en el otro FET es la diferencia, $I_d - I_m$.

T_1 es un transformador de audio, mientras que T_2 y T_3 son transformadores de radiofrecuencia. En consecuencia, cualquier componente de audio que aparezca en los circuitos de drenaje de Q_1 y Q_2 no pasa a la salida. Para lograr una supresión total de portadora, Q_a y Q_b deben ser iguales, y T_1 y T_3 deben tener una derivación exactamente al centro. Como en el caso de los moduladores de diodo balanceados, el modulador balanceado de FET suele agregar de 40 a 60 dB de atenuación a la portadora.

Modulador de puente balanceado

La fig. 5-10a muestra el diagrama de un *modulador de puente balanceado*. El funcionamiento del modulador de puente como el del modulador de anillo balanceado, depende por completo de la acción de conmutación de los diodos D_1 a D_4, bajo la influencia de los voltajes de la portadora y de la señal moduladora. De nuevo, el voltaje de la portadora controla el estado encendido

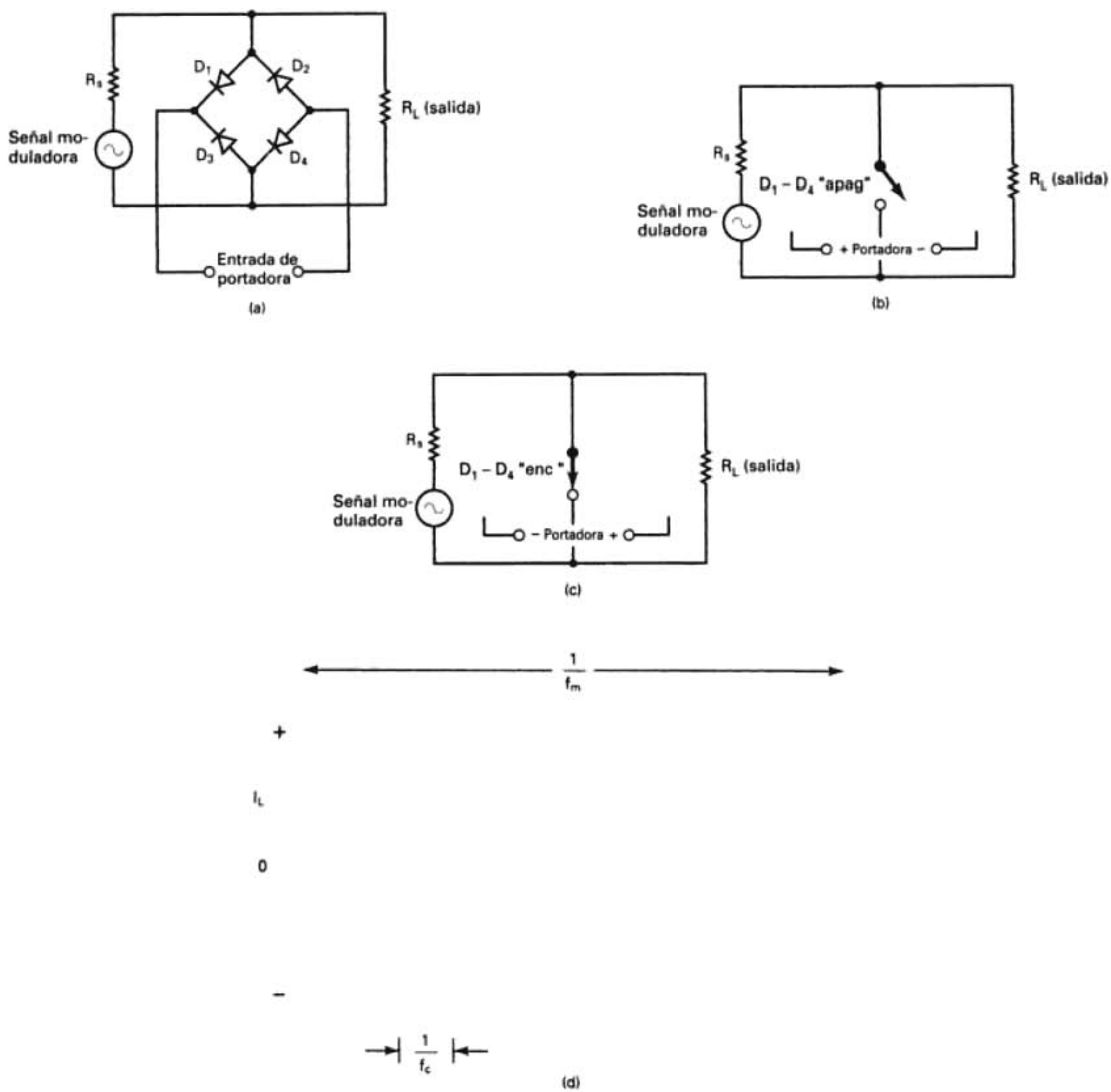


FIGURA 5-10 Modulador de puente balanceado: (a) esquema; (b) diodos polarizados en forma inversa; (c) diodos polarizados en forma directa; (d) forma de onda de salida

o apagado de los diodos y, en consecuencia, debe ser bastante mayor que el voltaje de la señal moduladora.

Funcionamiento del circuito. Para las polaridades de portadora de la fig. 5-10b, los cuatro diodos tienen polarización inversa y están apagados. En consecuencia, el voltaje de señal de audio pasa en forma directa al resistor de carga R_L. La fig. 5-10c muestra el circuito equivalente para una portadora con polaridad opuesta. Los cuatro diodos tienen polarización directa y están activos, y el resistor de carga se *pone en corto*. Al cambiar el voltaje de portadora de positivo a negativo y viceversa, la forma de onda de salida contiene una serie de pulsos forma-

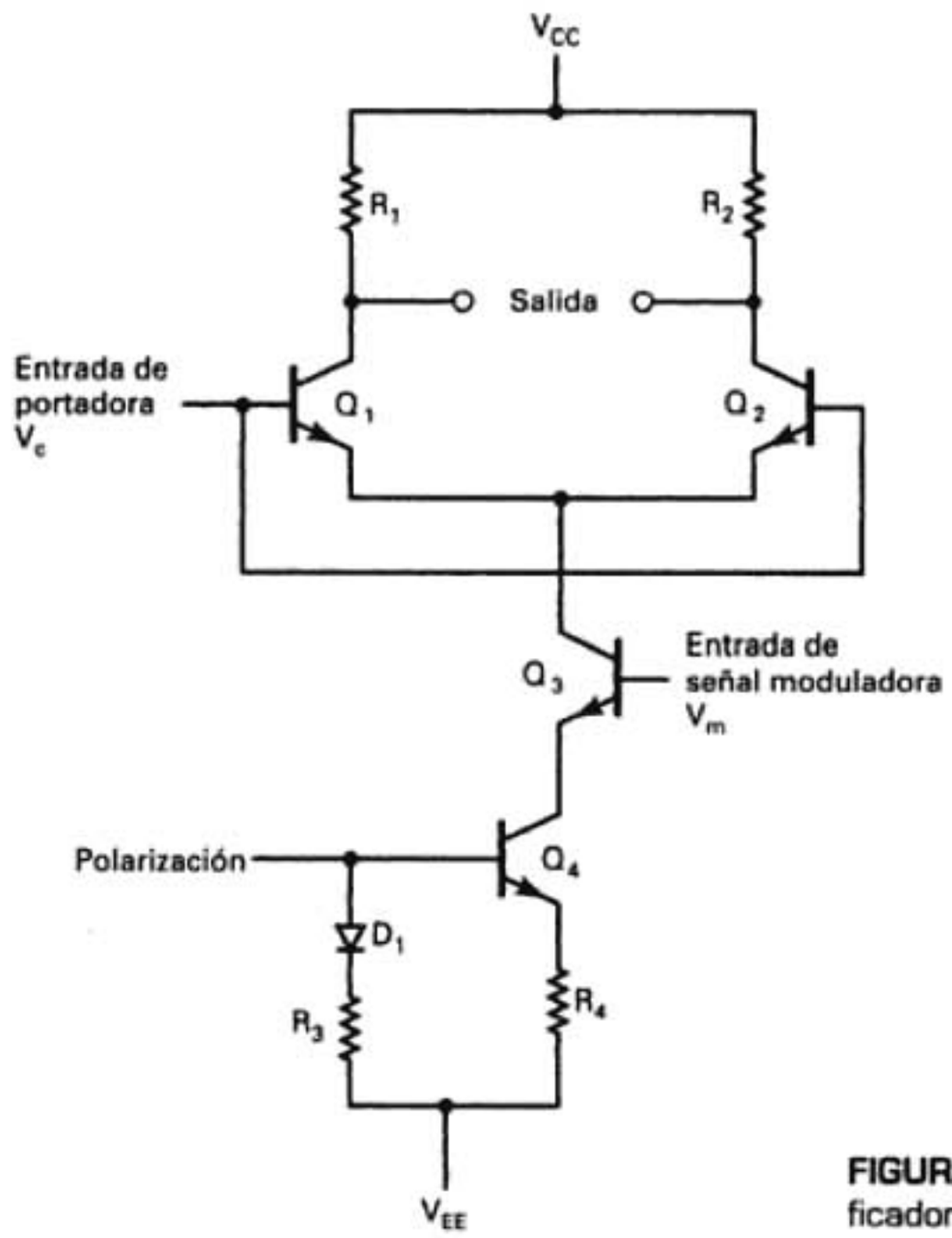


FIGURA 5-11 Esquema de un amplificador diferencial

da principalmente por frecuencias de banda lateral superior e inferior. La forma de onda de salida se ve en la fig. 5-10d. La serie de pulsos se indica allí como área sombreada.

Moduladores balanceados en circuito integrado lineal

Los *moduladores balanceados en circuito integrado lineal* (CI lineal) se consiguen hasta para 100 MHz, como por ejemplo el LM1496/1596, que puede proporcionar supresión de portadora de 50 dB a 10 MHz, y hasta 65 dB a 500 kHz. Este modulador balanceado en circuito integrado lineal es un *modulador/demodulador doblemente balanceado*, que produce una señal de salida que es proporcional al producto de sus señales de entrada. Los circuitos integrados son ideales para aplicaciones donde se requiere una operación balanceada.

Funcionamiento del circuito. La fig. 5-11 muestra un diagrama simplificado de un amplificador diferencial, que es el circuito fundamental de un modulador balanceado en CI lineal dado, por su excelente *relación de rechazo en modo común*, normalmente de 85 dB o más. Cuando se aplica una señal de portadora a la base de Q_1, las corrientes de emisor en ambos transistores varían la misma cantidad. Como la corriente de emisor para Q_1 y Q_2 proviene de una fuente común de corriente constante, Q_4, todo aumento en la corriente del emisor Q_1 da como resultado una disminución correspondiente de la corriente de emisor en Q_2, y viceversa. De igual modo, cuando se aplica una señal de portadora a la base de Q_2, las corrientes de emisor de Q_1 y Q_2 varían la misma magnitud, pero con direcciones opuestas. En consecuencia, si se alimenta la misma señal de portadora a las bases de Q_1 y Q_2, los aumentos y disminuciones respectivos son iguales y por tanto se anulan. Entonces, las corrientes de colector y el voltaje de salida permanecen invariables. Si se aplica una señal moduladora a la base de Q_3, causa un aumento o disminución correspondiente (dependiendo de su polaridad) en las corrientes de colector de Q_1 y Q_2. Sin embargo, las frecuencias de portadora y de señal moduladora se mezclan en los transistores y producen frecuencias de producto cruzado en la salida. Entonces, las frecuencias de portadora y de señal moduladora se anulan en los transistores balanceados, mientras que en la salida aparecen las frecuencias de suma y de diferencia.

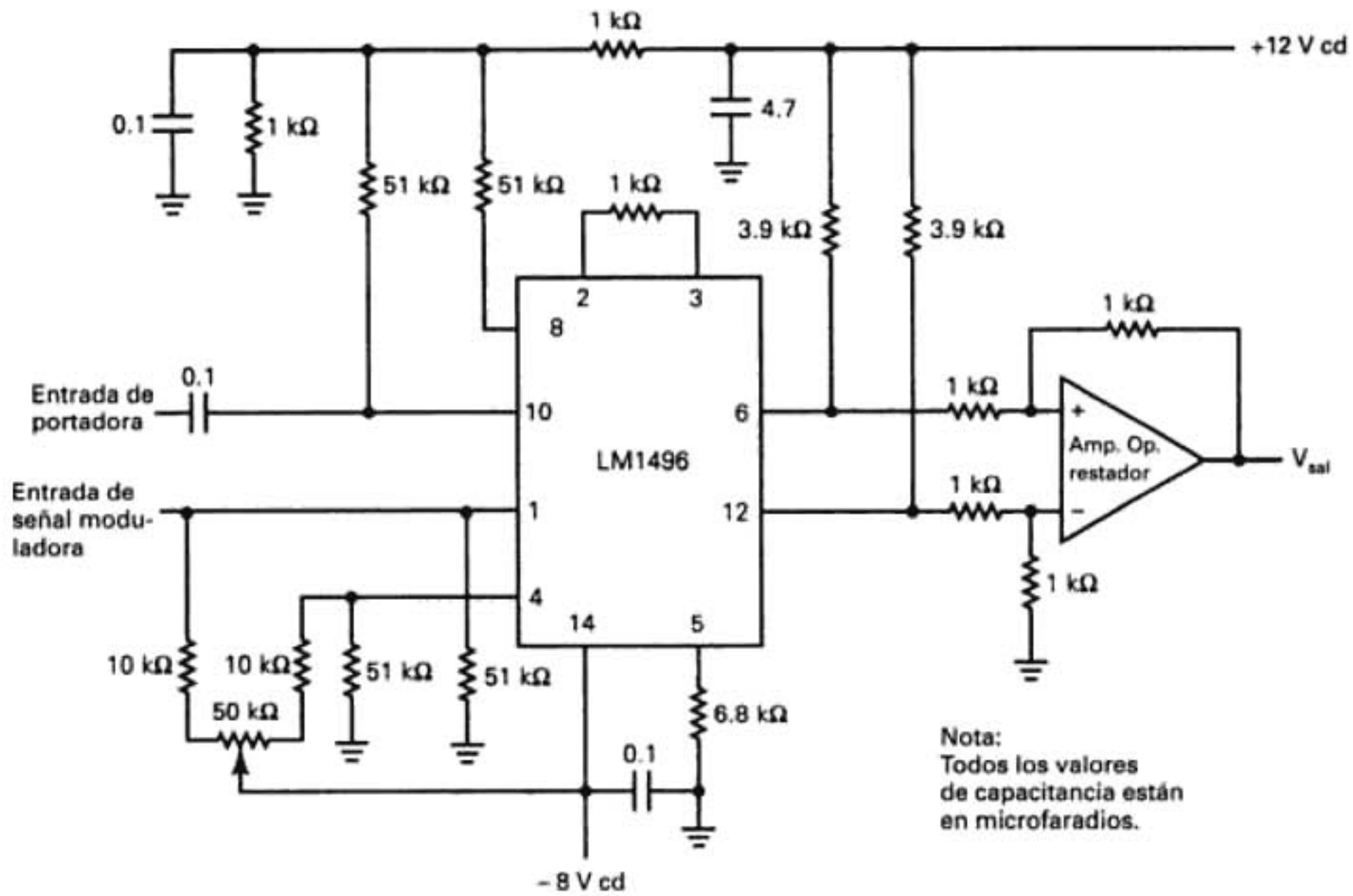


FIGURA 5-12 Modulador DSBSC de AM con circuito integrado lineal LM1496/1596

La fig. 5-12 muestra el diagrama de un modulador típico DSBSC de AM, que usa el circuito integrado LM1496/1596. Este circuito es un modulador/demodulador balanceado, en el que la salida es el producto de sus dos entradas. Tiene una supresión excelente de portadora (65 dB a 0.5 MHz), ganancia ajustable, entradas y salidas balanceadas y una alta relación de rechazo de modo común (85 dB). Cuando se usa como detector de producto tiene una sensibilidad de 3.0 μV y un intervalo dinámico de 90 dB trabajando a una frecuencia intermedia de 9 MHz.

La señal de portadora se aplica a la terminal 10 que, en conjunto con la terminal 8, proporciona una entrada a un amplificador de acoplamiento cruzado y salida diferencial en cuadratura. Esta configuración es para asegurar la obtención de multiplicación de onda de máxima potencia para la portadora y la señal moduladora. La señal moduladora se aplica a la terminal 1, que junto con la terminal 4 proporciona una entrada diferencial a los transistores activadores de corriente para el amplificador de diferencia de salida. El potenciómetro de 50 kΩ, junto con V_{EE}, de −8 V cd, se usa para balancear las corrientes de polarización para los amplificadores de diferencia y para anular la portadora. Las terminales 6 y 12 son de salida asimétrica que contienen componentes de portadora y de banda lateral. Cuando una de las salidas se invierte y se suma a la otra, se suprime la portadora y se produce una onda de doble banda lateral y portadora suprimida. Ese proceso se efectúa en el restador de amplificador operacional. Ese restador invierte la señal en la entrada inversora (−) y la suma a la señal en la entrada no inversora (+). Así, aparece en la salida del amplificador operacional una onda de doble banda lateral y portadora suprimida. El resistor de 6.8 kΩ conectado a la terminal 5 es de polarización, para el suministro interno de corriente constante.

El circuito lineal integrado XR-2206 de modulador DSBFC de AM en circuito integrado que se describió en el capítulo 3, y que se ve en la fig. 3-20a, también se puede usar para producir una onda de doble banda lateral y portadora suprimida, tan sólo con ajustar la polarización de cd a $V^+/2$ y limitar la amplitud de señal moduladora a $\pm 4\ V_p$. Al pasar la señal moduladora por sus ceros, la fase de la portadora sufre una inversión de fase de 180°. Esta propiedad también hace que el XR-2206 sea ideal como modulador por desplazamiento de fase. El intervalo dinámico de modulación de amplitud de este circuito es de 55 dB, aproximadamente.

TRANSMISORES DE BANDA LATERAL ÚNICA

Los sistemas que se usan para transmisión de banda lateral única (SSB, por *single-sideband*) con portadora suprimida o reducida son idénticos, excepto que los de portadora reinsertada tienen un circuito adicional que suma una portadora de baja amplitud a la forma de onda de banda lateral única después de haber hecho la modulación de portadora suprimida, y haber eliminado una de las bandas laterales. La portadora reinsertada se llama *portadora piloto*. El circuito en el que se reinserta la portadora se llama *sumador lineal*, si es una red resistiva, y *bobina híbrida* si la forma de onda SSB y la portadora piloto se combinan inductivamente en un circuito de puente con transformador. Se suelen usar tres configuraciones de transmisor para generar la banda lateral única: el método con filtro, el de desplazamiento de fase y el llamado *tercer método*.

Transmisor de banda lateral única: método con filtro

La fig. 5-13 muestra un diagrama de bloques de un transmisor SSB que usa moduladores balanceados para suprimir la portadora indeseada, y filtros para suprimir la banda lateral indeseada. Ahí se ve un transmisor que usa tres etapas de conversión elevadora de frecuencia. La señal moduladora es una señal de audio cuyo espectro va de 0 a 5 kHz. Se mezcla con una portadora de baja frecuencia (LF, de *low-frequency*) de 100 kHz en el modulador balanceado 1 para producir un espectro de frecuencias de banda lateral doble, centrada alrededor de la portadora suprimida de FI, de 100 kHz. El filtro pasabanda 1 (BPF 1) se sintoniza a un ancho de banda de 5 kHz centrado en 102.5 kHz, que es el centro del espectro de frecuencias de la banda lateral superior. La señal piloto o portadora de amplitud reducida se suma a la forma de onda de la banda lateral única en la etapa de reinserción de portadora, que no es más que un sumador lineal. Es un circuito sencillo que combina la portadora piloto de 100 kHz con el espectro de frecuencias de banda lateral superior, de 100 a 105 kHz. Así, la salida del sumador es una forma de onda SSBRC (*banda lateral única y portadora reducida*). Si se desea una transmisión con portadora suprimida se pueden omitir la portadora piloto y el circuito sumador.

La FI de baja frecuencia se convierte a la banda de frecuencias de operación a través de una serie de translaciones de frecuencia. Primero, se mezcla la forma de onda SSBRC en el modulador balanceado 2 con una portadora de frecuencia media (MF, por *medium frequency*) de 2 MHz. La salida es una señal de doble banda lateral y portadora suprimida, en la que las bandas laterales superior e inferior contienen cada una el espectro de frecuencias SSBRC original. Las bandas laterales superior e inferior están separadas por una banda de 200 kHz que no contiene información. La frecuencia central del filtro pasabanda (BPF, de *bandpass filter*) 2 es 2.105 MHz con ancho de banda de 5 kHz. En consecuencia, la salida de BPF 2 es de nuevo una forma de onda de banda lateral única y portadora suprimida (SSBRC). Su espectro de frecuencias comprende una segunda portadora reducida de FI de 2.1 MHz, y una banda lateral superior de 5 kHz de ancho. La salida del BPF 2 se mezcla con una portadora de alta frecuencia (HF, por *high-frequency*) de 20 MHz en el modulador balanceado 3. La salida es una señal de doble banda lateral y portadora suprimida en la que cada una de las bandas laterales superior e inferior contienen de nuevo el espectro de frecuencias SSBRC original. Las bandas laterales están separadas por una banda de frecuencias de 4.2 MHz, sin información. El BPF 3 está centrado en 22.105 MHz, con un ancho de banda de 5 kHz. Por consiguiente, la salida de ese BPF 3 es de nuevo una forma de onda de banda lateral única con portadora de RF reducida a 22.1 MHz, y una banda lateral superior de 5 kHz de ancho. La forma de onda de salida se amplifica en el amplificador lineal de potencia, para después ser transmitida.

En el transmisor que se acaba de describir, el espectro original de frecuencias de señal moduladora tuvo conversión elevadora en tres etapas de modulación hasta llegar a la frecuencia final de portadora de 22.1 MHz, y una banda lateral superior única que abarcó desde la portadora hasta 22.105 MHz. Después de cada conversión elevadora (traslación de frecuencia), la banda lateral deseada se separa del espectro de doble banda lateral con un filtro pasabanda. Se puede producir el mismo espectro final de salida con un proceso único de heterodinado: un modulador balanceado, un filtro pasabanda y una sola fuente de portadora de alta frecuencia. La fig. 5-14a muestra el diagrama de bloques y el espectro de frecuencias de salida para un transmisor con una sola conversión. La salida del modulador balanceado es un espectro de frecuencias de doble banda lateral, centrado en una frecuencia de portadora suprimida de 22.1 MHz. Para separar la banda lateral su-

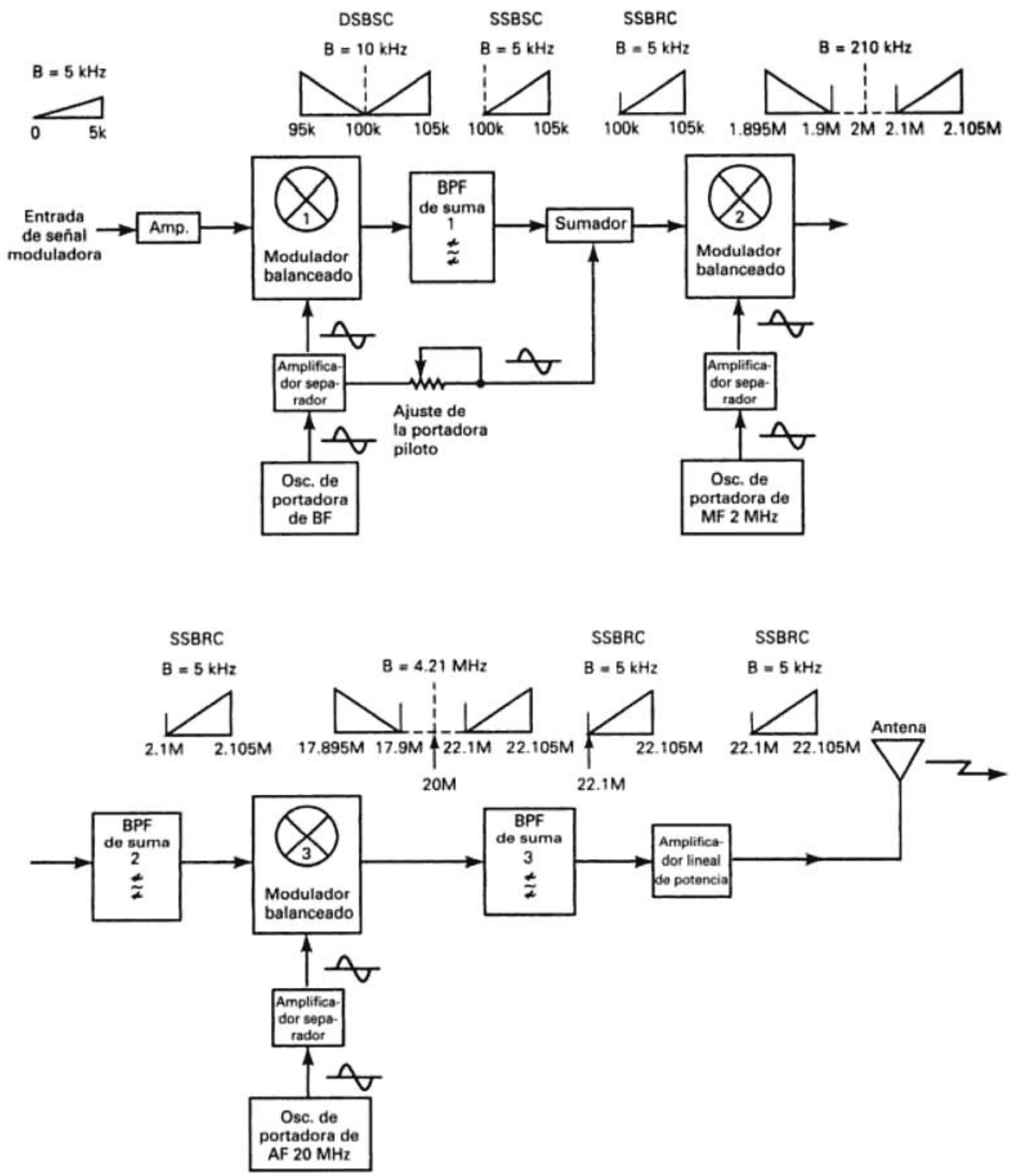


FIGURA 5-13 Transmisor de banda lateral única: método del filtro

perior de 5 kHz de ancho y el espectro de frecuencias compuestas, se requiere un filtro pasabandas de varios polos con una Q extremadamente alta. Es difícil fabricar un filtro que llene estos criterios, pero supongamos que fuera un transmisor multicanal, y que fuera sintonizable la frecuencia de la portadora. En ese caso, el BPF también debe ser sintonizable. La fabricación de un filtro pasabanda sintonizable en la región de frecuencias de megahertz con una banda de paso de sólo 5 kHz cae fuera de la factibilidad técnico económica. El único filtro pasabanda que se ve en la fig. 5-13, que tiene dos bandas laterales separadas inmediatamente adyacentes entre sí es el BPF 1. Es

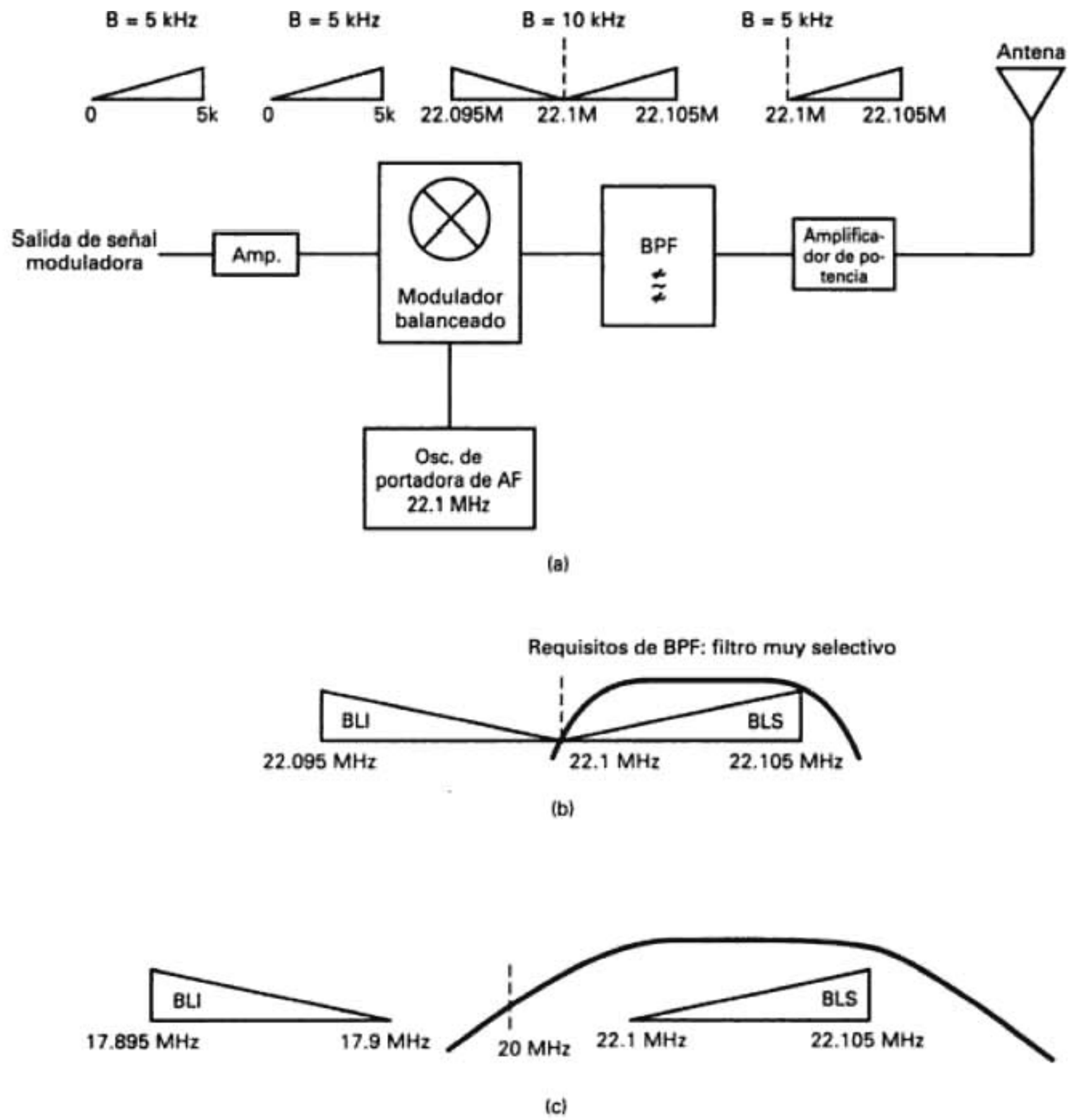


FIGURA 5-14 Transmisor SSBSC por conversión única, método del filtro: (a) diagrama de bloques; (b) espectro de frecuencias y requisitos de filtrado para un transmisor de conversión única; (c) espectro de salida y requisitos de filtrado para un transmisor de triple conversión

relativamente fácil fabricar un filtro pasabandas de 5 kHz de ancho con faldas verticales, en 100 kHz, porque sólo se requiere una Q moderada. Las bandas laterales separadas por el BPF2 están separadas 200 kHz y, por lo tanto, se puede usar un filtro de baja Q con características graduales de atenuación progresiva, sin peligro de pasar parte alguna de la banda lateral no deseada. El BPF 3 separa las bandas laterales que están a 4.2 MHz de distancia. Si se usan canales múltiples y si la portadora de alta frecuencia es sintonizable, se puede usar un solo filtro de banda ancha como BPF 3, sin peligro de que pase por él parte alguna de la banda lateral indeseada. Para operación con un solo canal, el diseño más sencillo es el transmisor de conversión única, pero para operación en multicanal es más práctico el sistema de tres conversiones. Las Figs. 5-14b y c muestran el espectro de frecuencias y los requisitos de filtrado para ambos métodos.

Filtros de banda lateral única. Es evidente que los filtros son parte esencial de cualquier sistema de comunicaciones electrónicas, y en especial de los sistemas de banda lateral única. En los transmisores y los receptores hay requisitos de redes muy selectivas, para limitar los espectros de frecuencia tanto de la señal como del ruido. El factor de calidad Q de un filtro de banda lateral única depende de la frecuencia de la portadora, la frecuencia que separa a ambas

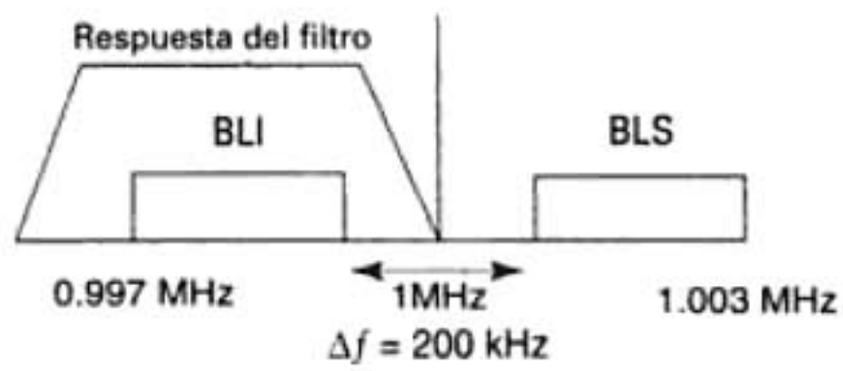
bandas laterales y del nivel de atenuación de la banda lateral no deseada que se quiere alcanzar. La ecuación para calcular Q se puede expresar como sigue

$$Q = \frac{f_c(\log^{-1} S/20)^{1/2}}{4\Delta f} \tag{5-1}$$

en la que Q = factor de calidad
f_c = frecuencia central o de la portadora
S = grado de supresión en dB, de la banda lateral no deseada
Δf = separación de frecuencias máxima de banda lateral inferior y mínima de banda lateral superior

Ejemplo 5-1

Determinar el factor de calidad Q necesario para un filtro de banda lateral única con frecuencia de portadora de 1 MHz, 80 dB de supresión de banda lateral no deseada, y el siguiente espectro de frecuencias



Solución Al sustituir en la ecuación 5-1 se obtiene

$$Q = \frac{1\text{ MHz}(\log^{-1} 80/20)^{1/2}}{4(200)} = 125{,}000$$

Las Q de los filtros convencionales LC son relativamente bajas y en consecuencia esos filtros no tienen la selectividad suficiente para la mayoría de las aplicaciones de banda lateral única. Así, los filtros que se usan para generar banda lateral única se suelen fabricar con materiales de cristal o cerámicos, con filtros mecánicos o con filtros de onda acústica superficial (SAW, de *surface acoustic wave*).

Filtros de cristal. El *filtro de red cristalina* se usa mucho en sistemas de banda lateral única. En la fig. 5-15a se ve el diagrama de un filtro pasabanda de red cristalina. La red comprende dos juegos de pares iguales de cristal, X_1 y X_2; X_3 y X_4, conectados entre los transformadores sintonizados de entrada y salida T_1 y T_2. Los cristales X_1 y X_2 están conectados en serie, mientras que X_3 y X_4 están conectados en paralelo. Cada par de cristales se ajusta en frecuencia, con un margen de 10 a 20 Hz. Los cristales X_1 y X_2 se cortan para que funcionen en la frecuencia inferior de corte del filtro, y X_3 y X_4 para que operen en la frecuencia superior de corte. Los transformadores de entrada y salida se sintonizan al centro de la banda de paso deseada, y eso tiende a dispersar la diferencia entre las frecuencias de resonancia en serie y en paralelo. Los capacitores C_1 y C_2 se usan para corregir cualquier sobredispersión de diferencia de frecuencias, bajo condiciones de cristales iguales.

El funcionamiento del filtro de cristal se parece al de un circuito puente. Cuando las reactancias de las ramas del puente son iguales y tienen el mismo signo (inductivas o capacitivas), las señales que se propagan por las dos trayectorias posibles del puente se anulan entre sí. En la frecuencia en la que las reactancias tienen magnitud igual y signos contrarios (una inductiva y otra capacitiva), la señal se propaga por la red, con amplitud máxima.

La fig. 5-15b muestra una curva característica normal para un filtro pasabanda de red cristalina. Los filtros de cristal se consiguen con Q tan alta como 100,000. El filtro de la fig. 5-15a es de un elemento. Sin embargo, para que un filtro de cristal pase en forma adecuada una banda específica de frecuencias y rechace todas las demás, son necesarios al menos dos elementos. Las pérdidas normales de inserción para los filtros de cristal son de 1.5 dB a 3 dB.

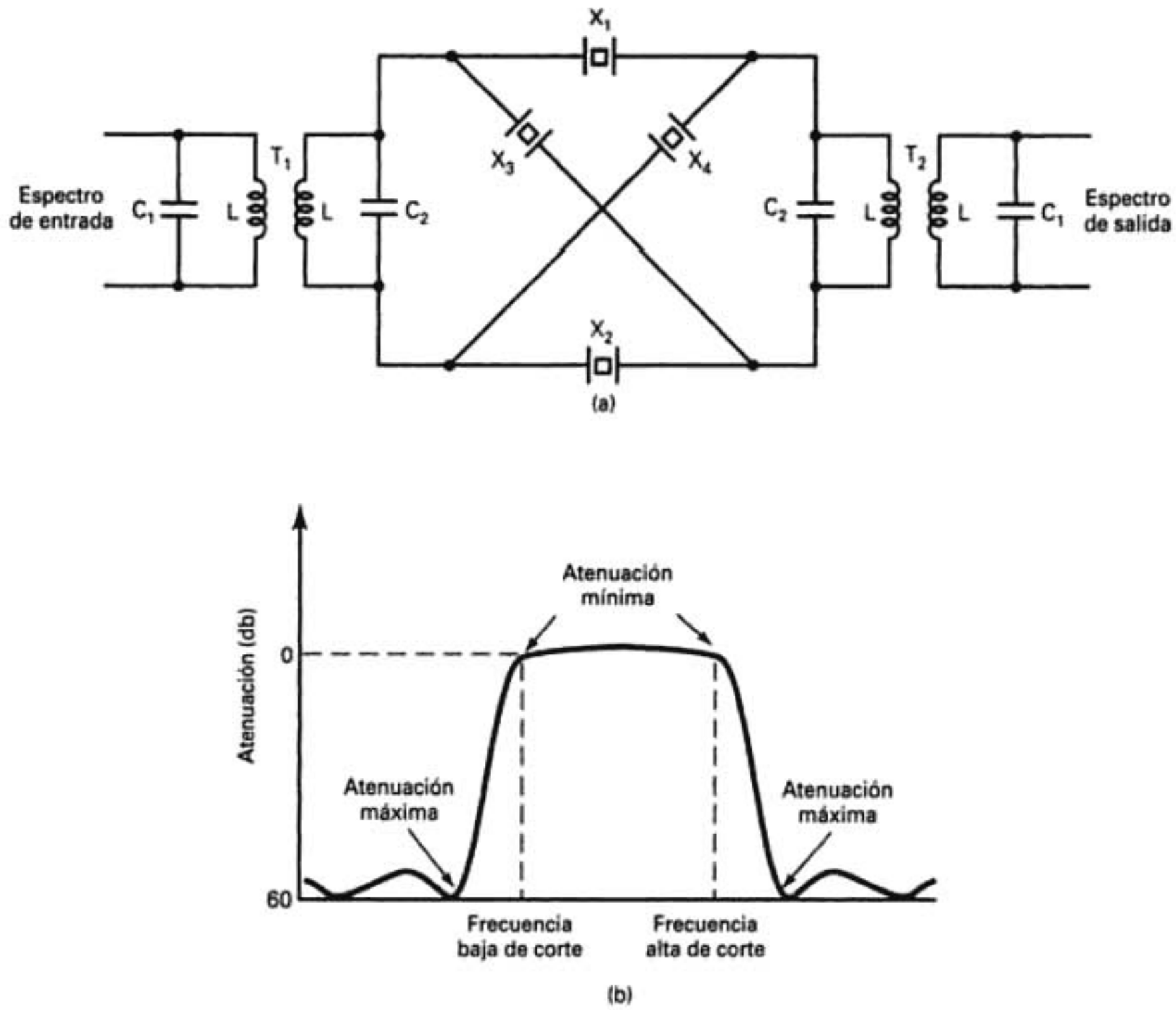


FIGURA 5-15 Filtro de red cristalina: a) diagrama eléctrico; (b) curva característica

Filtros de cerámica. Los *filtros de cerámica* se fabrican con zirconato-titanato de plomo, que posee el efecto piezoeléctrico. En consecuencia funcionan en forma muy parecida a los filtros de cristal, pero los de cerámica no tienen un factor Q tan alto. Los valores Q normales en los filtros de cerámica suben hasta 2000, aproximadamente. Estos filtros son menos costosos, más pequeños y más robustos que sus contrapartes de red cristalina. Sin embargo, los filtros de cerámica tienen más pérdida. Su pérdida normal de inserción está entre 2 y 4 dB.

Los filtros de cerámica se suelen conseguir en paquetes de un elemento y tres terminales, dos elementos y ocho terminales y cuatro elementos y catorce terminales. Tienen tamaño pequeño, perfil bajo, características simétricas de selectividad, baja respuesta espuria y excelente inmunidad contra variaciones de condiciones ambientales, con variación mínima de características de operación. Sin embargo se deben tener con ellos algunas precauciones, que incluyen las siguientes:

1. ***Condiciones de adaptación de impedancia y carga.*** Los filtros de cerámica se distinguen de las bobinas en que su impedancia no se puede cambiar con facilidad. Al usar filtros de cerámica es muy importante tener bien acopladas las impedancias.
2. ***Señales espurias.*** Prácticamente en todos los casos en que se usan los filtros de cerámica, se generan señales espurias. Para suprimir esas respuestas, el método más sencillo y efectivo es el balance de impedancia con transformadores de FI.
3. ***Bobinas de balance.*** Cuando surgen dificultades en la supresión de respuesta espuria, o para mejorar la selectividad o balance de impedancia en las etapas de FI, se aconseja usar una bobina de balance de impedancia.

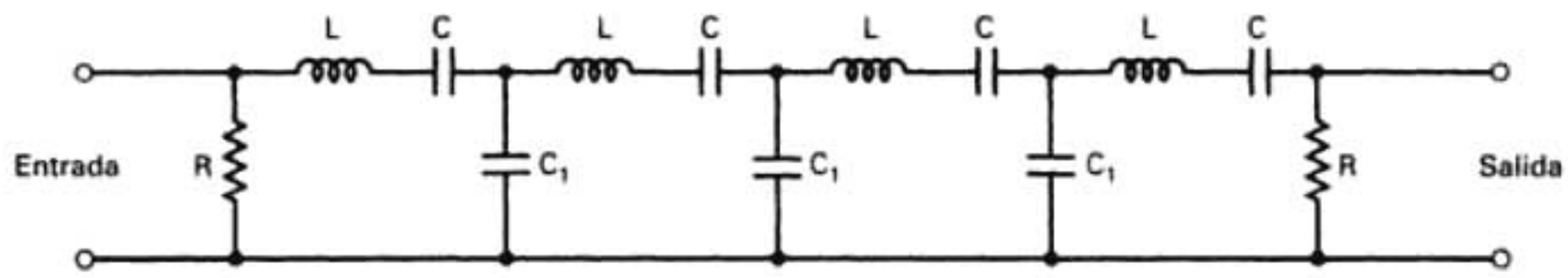


FIGURA 5-16 Circuito equivalente a un filtro mecánico

4. ***Error en conexiones de conductores de entrada y salida.*** Se debe tener cuidado al conectar las terminales de entrada y salida de un filtro de cerámica. Cualquier error causará distorsiones de forma de onda, y quizá desviación de frecuencia de la señal.
5. ***Uso de dos filtros de cerámica en cascada.*** Para la mayor eficiencia se debe usar una bobina entre dos unidades de filtro de cerámica. Cuando el costo es importante y es necesaria una conexión directa, se puede usar un capacitor o un resistor adecuado.

Filtros mecánicos. Un *filtro mecánico* es un *transductor resonante mecánicamente*. Recibe energía eléctrica, la convierte en vibraciones mecánicas, y a continuación reconvierte las vibraciones en energía eléctrica en su salida. En esencia, un filtro mecánico comprende cuatro elementos: un transductor de entrada, que convierte la energía eléctrica que entra en vibraciones mecánicas, una serie de discos metálicos, con resonancia mecánica, que vibran a la frecuencia de resonancia deseada, una varilla de acoplamiento que agrupa a los discos metálicos, y un transductor de salida que reconvierte las vibraciones mecánicas en energía eléctrica. La fig. 5-16 muestra el circuito eléctrico equivalente de un filtro mecánico. Los circuitos resonantes en serie (combinaciones *LC*) representan los discos metálicos; el capacitor de acoplamiento C_1 representa la varilla de acoplamiento y *R* representa las cargas mecánicas. La frecuencia de resonancia del filtro se determina con los discos *LC* en serie, y C_1 determina el ancho de banda. Los filtros mecánicos son más robustos que los de cerámica o de cristal, y tienen características semejantes de respuesta a la frecuencia. Sin embargo, son más grandes y pesados y, en consecuencia, no son prácticos en los equipos móviles de comunicaciones.

Filtros de onda acústica superficial. Los *filtros de onda acústica superficial* (SAW, por *surface acoustic wave*) se desarrollaron en la década de 1960, pero no estuvieron disponibles sino hasta la década siguiente. Usan energía acústica, más que energía electromecánica, para obtener una eficiencia excelente de filtrado pasabanda. En esencia, los filtros SAW atrapan, o guían, ondas acústicas en una superficie. Pueden trabajar con frecuencias centrales hasta de varios gigahertz, y anchos de banda hasta de 50 MHz, con más exactitud y confiabilidad que el filtro mecánico; además lo hacen con menor costo. Los filtros SAW tienen características extremadamente verticales de atenuación progresiva, y suelen atenuar frecuencias fuera de su pasabanda entre 30 y 50 dB más que las señales dentro de su banda de paso. Los filtros SAW se usan en receptores superheterodinos, tanto de conversión única como de varias conversiones, como filtros de RF y de FI, y en sistemas de banda lateral única para una multitud de aplicaciones de filtrado.

Un filtro de onda acústica superficial está formado por transductores recortados en una película delgada de aluminio depositada en la superficie de un material cristalino semiconductor que posea el efecto piezoeléctrico. Esto da como resultado una deformación física (ondulaciones) sobre la superficie del substrato. Esas ondulaciones varían en la frecuencia de la señal aplicada, pero se mueven por la superficie del material a la velocidad del sonido. Con los filtros SAW se aplica una señal eléctrica oscilatoria a través de una pieza pequeña de cristal de semiconductor, que es parte de una superficie plana mayor, como se ve en la fig. 5-17a. El efecto piezoeléctrico hace que vibre el material del cristal. Esas vibraciones están en forma de energía acústica, que viaja por la superficie del substrato hasta llegar a un segundo cristal, en el extremo opuesto, donde la energía acústica se reconvierte en energía eléctrica.

Para lograr la acción de filtro se deposita una fila de láminas interdigitales metálicas, a distancias precisas, sobre la superficie plana del substrato, como se ve en la fig. 5-17b. Los cen-

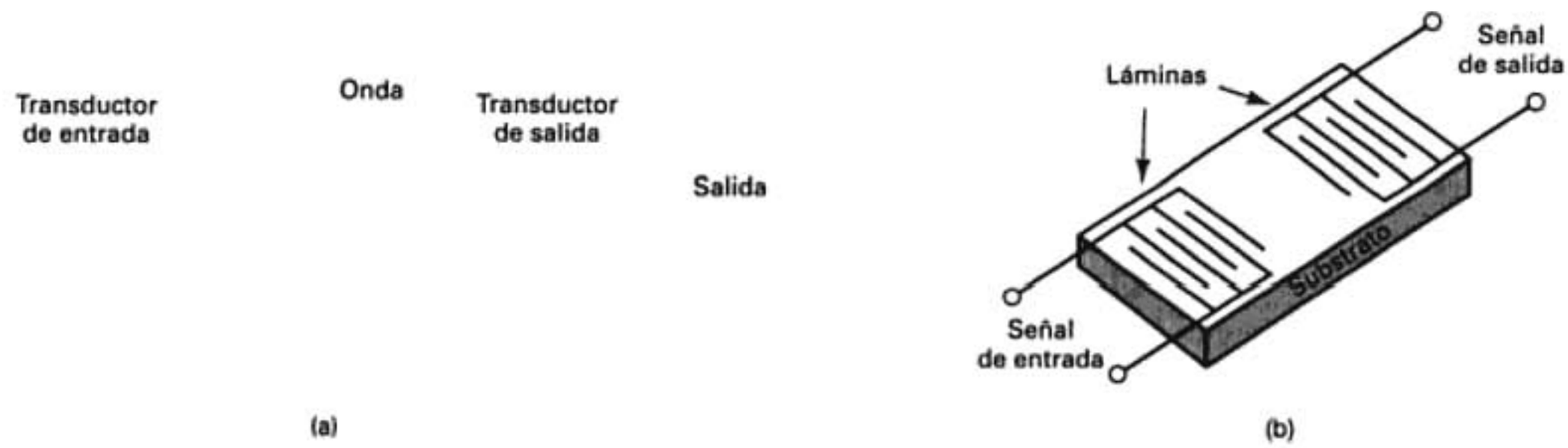


FIGURA 5-17 Filtro de onda acústica superficial (SAW): (a) onda superficial; (b) láminas metálicas

tros de las láminas están espaciados a la mitad o a la cuarta parte de la longitud de onda de la frecuencia central deseada. Cuando las ondas acústicas viajan a través de la superficie del substrato, se reflejan y van y vienen al chocar con las láminas. Según la longitud de onda acústica y de la distancia entre las láminas, algo de la energía reflejada se anula, y atenúa la energía de la onda incidente; a esto se le llama *interferencia destructiva*, mientras que algo de la energía se suma (*interferencia constructiva*). Las frecuencias exactas de energía acústica que se anulan dependen de la distancia entre las láminas. El ancho de banda del filtro está determinado por el espesor y la cantidad de láminas.

El filtro SAW básico es *bidireccional*. Esto es, la mitad de la potencia es irradiada hacia el transductor de salida, y la otra mitad es irradiada hacia el extremo del substrato cristalino y se pierde. Por reciprocidad, la mitad de la potencia se pierde en el transductor de salida. En consecuencia, los filtros SAW tienen una pérdida de inserción relativamente alta. Esta desventaja se puede superar hasta cierto grado, usando una estructura más complicada llamada *transductor unidireccional*, que lanza la onda acústica sólo en una dirección.

Los filtros SAW son muy robustos y confiables. Como sus frecuencias de operación y respuestas pasabanda se establecen con el proceso fotolitográfico, no necesitan operaciones complicadas de sintonización ni se desintonizan al paso del tiempo. Las técnicas de procesamiento de obleas de semiconductor que se usan en la fabricación de estos filtros permiten la producción en gran escala, con dispositivos económicos y reproducibles. Por último, sus excelentes cualidades de funcionamiento se logran con un tamaño y peso bastante reducido, en comparación con otras tecnologías.

La mayor desventaja de los filtros SAW es su pérdida de inserción extremadamente alta, que suele ser de 25 a 35 dB. Por esta razón no se pueden usar para filtrar señales de bajo nivel. También los filtros SAW tienen un tiempo de demora mucho mayor que sus contrapartes electrónicas, unas 20,000 veces mayor. En consecuencia, a veces se usan los filtros SAW en *líneas de retardo*.

Transmisor de banda lateral única: método de desplazamiento de fase

Con el método de desplazamiento de fase para generar banda lateral única (SSB), la banda lateral no deseada se anula en la salida del modulador y, en consecuencia, no es necesario un filtrado exacto. La fig. 5-18 muestra un diagrama de bloques de un transmisor de SSB donde se usa el método de desplazamiento de fase para quitar la banda lateral superior. En esencia, hay dos moduladores separados de doble banda lateral (moduladores balanceados 1 y 2). Las señales moduladora y portadora se aplican en forma directa a uno de los moduladores, y a continuación las dos se desplazan 90° y se aplican al segundo modulador. Las salidas de los dos moduladores balanceados son señales de doble banda lateral y portadora suprimida, con la fase adecuada de tal manera que cuando se combinan en un sumador lineal, la banda lateral superior se anula.

Representación fasorial. Los fasores de la fig. 5-18 ilustran cómo se anula la banda lateral superior, girando 90° tanto la portadora como la señal moduladora, antes de la modulación.

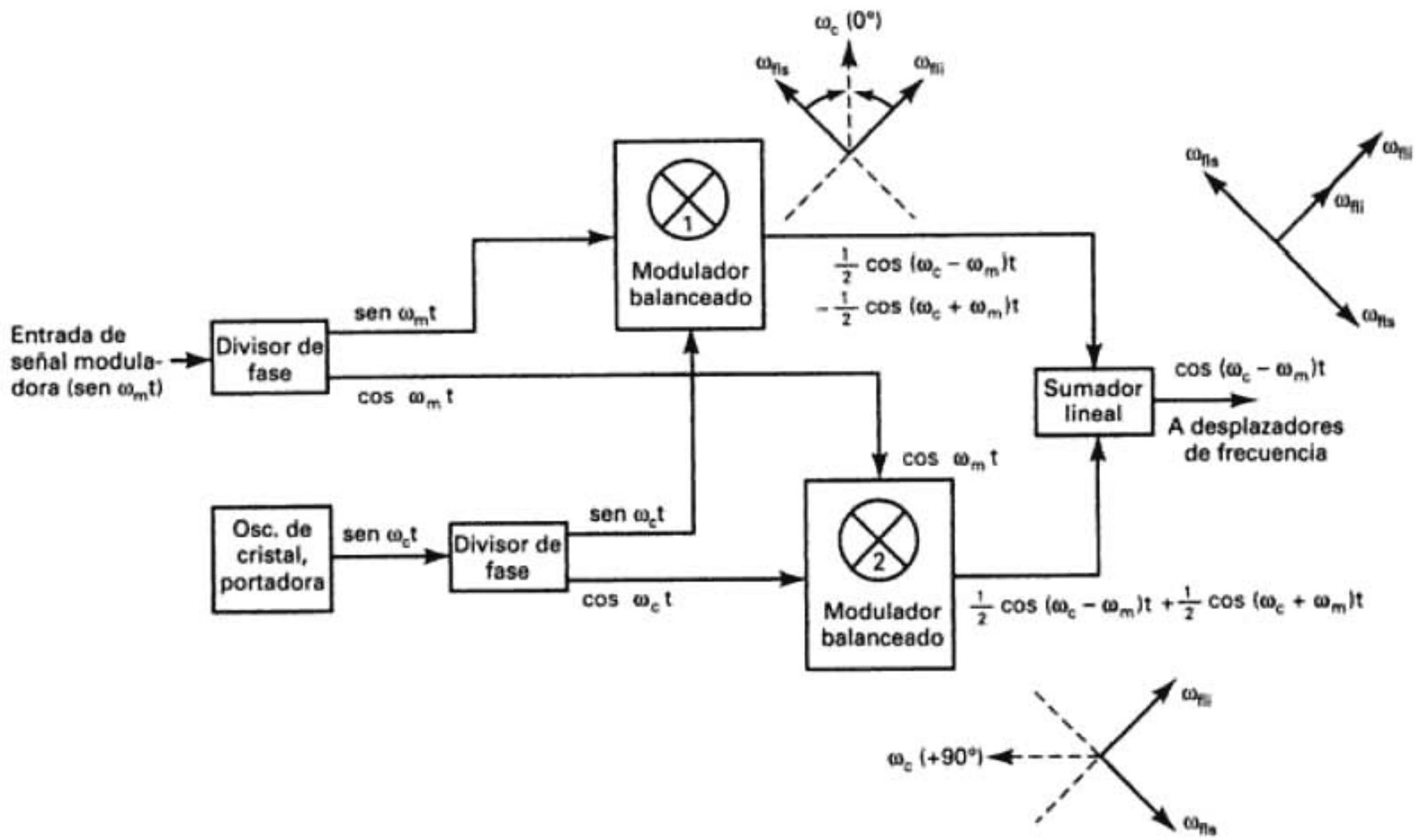


FIGURA 5-18 Transmisor SSB: método del desplazamiento de fase

La fase de salida del modulador balanceado 1 muestra la posición relativa y la dirección de giro de las frecuencias laterales superior (ω_{fls}) e inferior (ω_{fli}) respecto a la portadora suprimida (ω_c). Los fasores en la salida del modulador balanceado 2 son esencialmente iguales, excepto que la fase de la portadora y la de la señal moduladora se giran 90°. La salida del sumador es la suma de los fasores procedentes de los dos moduladores balanceados. Los dos fasores de la banda lateral inferior están en fase y son aditivos, mientras que los de la banda lateral superior están desfasados 180° y por consiguiente se anulan. Por lo anterior, sólo aparece la banda lateral inferior en la salida del sumador.

Análisis matemático. En la fig. 5-18 se alimenta la señal moduladora de entrada, sen $\omega_m t$, en forma directa al modulador balanceado 1, se desplaza 90° (cos $\omega_m t$) y alimenta al modulador balanceado 2. La portadora de baja frecuencia (sen $\omega_c t$) también se alimenta en forma directa al modulador balanceado 1 y se desplaza 90° (cos $\omega_c t$) y pasa al modulador balanceado 2. Los moduladores balanceados son moduladores de producto, y sus salidas se describen matemáticamente como sigue

$$\text{salida del modulador balanceado 1} = (\text{sen } \omega_m t)(\text{sen } \omega_c t)$$

$$= \frac{1}{2}\cos(\omega_c - \omega_m)t - \frac{1}{2}\cos(\omega_c + \omega_m)t$$

$$\text{salida del modulador balanceado 2} = (\cos \omega_m t)(\cos \omega_c t)$$

$$= \frac{1}{2}\cos(\omega_c - \omega_m)t + \frac{1}{2}\cos(\omega_c + \omega_m)t$$

y la salida del sumador lineal es

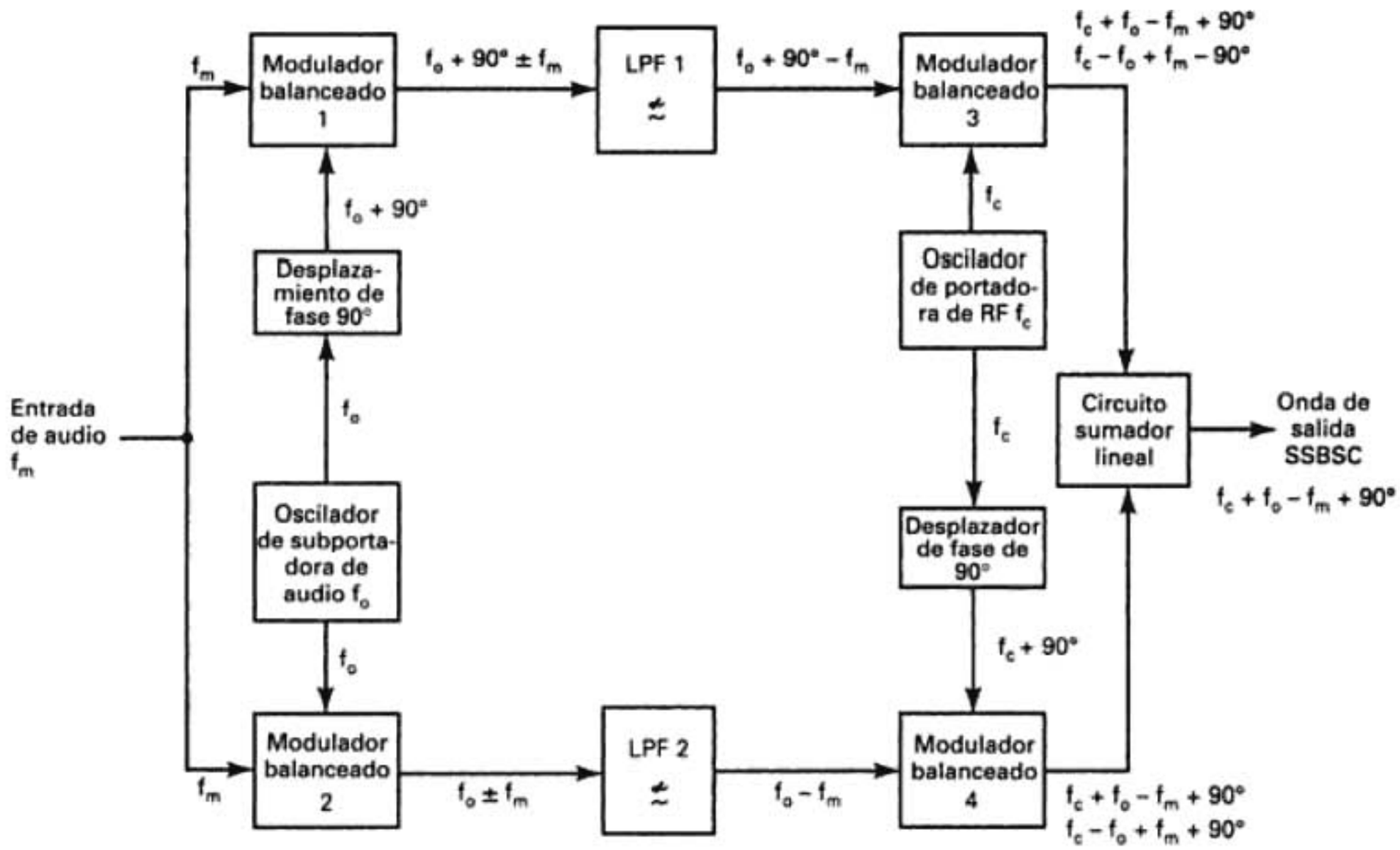


FIGURA 5-19 Modulador de banda lateral única y portadora suprimida: el "tercer método"

$$\frac{1}{2}\cos(\omega_c - \omega_m)t - \frac{1}{2}\cos(\omega_c + \omega_m)t$$
$$+\frac{1}{2}\cos(\omega_c - \omega_m)t + \frac{1}{2}\cos(\omega_c + \omega_m)t$$
$$\cos(\omega_c - \omega_m)t \qquad \text{se anulan}$$

banda lateral inferior
(señal de diferencia)

Transmisor de banda lateral única: tercer método

El *tercer método* de generación de banda lateral única, desarrollado por D. K. Weaver en la década de 1950, se parece al de desplazamiento de fase porque usa desplazamiento y suma de fase para anular la banda lateral no deseada. Sin embargo, tiene la ventaja de que la señal de información se modula al principio en una *subportadora* de audio, y se elimina así la necesidad de un desfasador de *banda amplia* (un desfasador, o variador de fase, que debe desplazar la misma cantidad una banda de frecuencias, lo cual es difícil de fabricar en la práctica). El diagrama de bloques de un modulador de SSB de tercer método se ve en la fig. 5-19. Nótese que las entradas y salidas de los dos desfasadores son frecuencias únicas (f_o, $f_o + 90°$, f_c y $f_c + 90°$). Las señales de audio que entran se mezclan con la subportadora de audio en los moduladores balanceados 1 y 2, que son alimentados con señales subportadoras en cuadratura (desfasadas 90°), f_o y $f_o + 90°$. La salida del modulador balanceado 2 contiene las bandas laterales superior e inferior ($f_o \pm f_m$), mientras que la salida del modulador balanceado 1 contiene las bandas laterales superior e inferior desplazadas, cada una, 90° ($f_o \pm f_m + 90°$). Las bandas laterales superiores se eliminan con sus filtros pasabajas (LPF, de *low-pass filter*) respectivos, cuya frecuencia de corte superior es igual a la de la subportadora de audio suprimida. La salida del LPF 1, $f_o - f_m + 90°$, se mezcla con la portadora de RF, f_c, en el modulador balanceado 3, y la salida del LPF 2, $f_o - f_m$, se mezcla con una portadora de fase desplazada 90° ($f_c + 90°$) en el modulador balanceado 4. Las portadoras de RF se suprimen en los moduladores balanceados 3 y 4. Entonces, las señales de salida de suma y diferencia, del modulador 3, $(f_c + f_o - f_m + 90°) + (f_c - f_o + f_m - 90°)$ se

combinan en el sumador lineal, con las señales de salida del modulador balanceado 4, $(f_c + f_o - f_m + 90°) + (f_c - f_o + f_m + 90°)$. La salida del sumador es

$$\begin{array}{l} (f_c + f_o - f_m + 90°) + (f_c - f_o + f_m - 90°) \\ + (f_c + f_o - f_m + 90°) + (f_c - f_o + f_m + 90°) \\ \hline (f_c + f_o - f_m + 90°) \quad \text{se anulan} \end{array}$$

La frecuencia final de salida de RF es $f_c + f_o - f_m$, que es esencialmente la banda lateral inferior de la portadora de RF, $f_c + f_o$. La fase desplazada 90° es un desplazamiento absoluto de fase que sufren todas las frecuencias y que, por consiguiente, no es importante. Si se desea la banda lateral superior de RF, tan sólo se intercambian las entradas de portadora a los moduladores balanceados 3 y 4, en cuyo caso la portadora final de RF es $f_c - f_o$.

Transmisor de banda lateral independiente

La fig. 5-20 muestra un diagrama de bloques de un transmisor de *banda lateral independiente* (ISB, de *independent sideband*) con tres etapas de modulación. Usa el método del filtro para producir dos canales de banda lateral única (canal A y canal B). Los dos canales se combinan y a continuación se reinserta una portadora piloto. La forma de onda de ISB con portadora reducida se eleva a RF con dos etapas más de traslación de frecuencia. Hay dos señales de información de 5 kHz que se originan en dos fuentes independientes. Las señales de información del canal A modulan a una portadora de baja frecuencia, de 100 kHz, en el modulador balanceado A. La salida del modulador balanceado A pasa por el filtro pasabanda (BPF) A, que se sintoniza a la banda lateral inferior (de 95 a 100 kHz). Las señales del canal de información B modulan la misma portadora de baja frecuencia de 100 kHz en el modulador balanceado B. La salida de este modulador B pasa por el BPF B, que se sintoniza a la banda lateral superior (de 100 a 105 kHz). Los dos espectros de frecuencia de banda lateral única se combinan en una red híbrida para formar un espectro compuesto ISB de portadora suprimida (de 95 a 105 kHz. La portadora de baja frecuencia de 100 kHz se reinserta en el sumador lineal para formar una forma de onda ISB de portadora reducida. El espectro ISB se mezcla con una portadora de frecuencia media de 2.7 MHz en el modulador balanceado 3. La salida de este modulador 3 pasa por el BPF 3 para producir un espectro ISB de portadora reducida que va desde 2.795 hasta 2.805 MHz con una portadora reducida piloto de 2.8 MHz. El modulador balanceado 4, el BPF 4 y la portadora de alta frecuencia trasladan al espectro de frecuencia media a una banda de radiofrecuencia que va de 27.795 a 27.8 MHz (canal A) y de 27.8 a 27.805 MHz (canal B), con una portadora de amplitud reducida de 27.8 MHz.

RECEPTORES DE BANDA LATERAL ÚNICA

Receptor BFO de banda lateral única

La fig. 5-21 muestra el diagrama de bloques de un sencillo *receptor BFO* no coherente de banda lateral única. El espectro seleccionado de radiofrecuencias se amplifica y a continuación se mezcla y reduce a frecuencias intermedias para su posterior amplificación y reducción de banda. La salida de la etapa amplificadora de frecuencia intermedia se heterodina (se pulsa) con la salida de un *oscilador de frecuencia de pulso* (BFO, de *beat frequency oscillator*). La frecuencia BFO es igual a la frecuencia intermedia de la portadora; de este modo, la diferencia entre las frecuencias intermedia y BFO es la señal de información. La demodulación se logra con varias etapas de mezclado y filtrado. El receptor es no coherente, porque el oscilador local de RF y las señales BFO no están sincronizados entre sí o con los osciladores en el transmisor. En consecuencia, toda diferencia entre las frecuencias de transmisión y las frecuencias del oscilador local produce un error de desplazamiento de frecuencia en la señal demodulada de información. Por ejemplo, si el oscilador local de recepción tiene 100 Hz más que su frecuencia designada y el BFO está a 50 Hz arriba de su frecuencia designada, la información restaurada está desplazada 150 Hz respecto a su espectro original de frecuencias. Un escucha normal percibe un corrimiento de 50 Hz o más como una variación de tono.

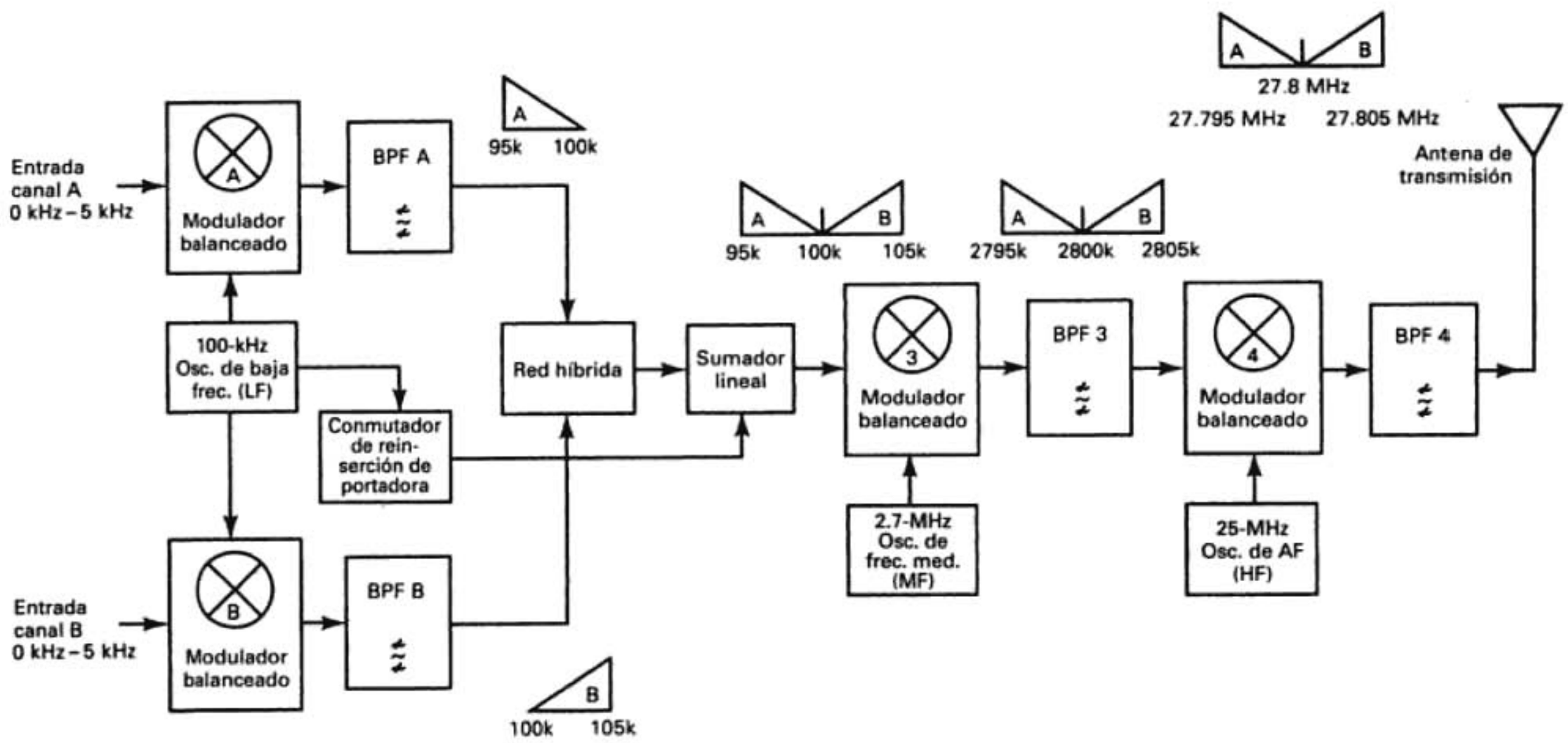


FIGURA 5-20 Transmisor ISB: diagrama de bloques

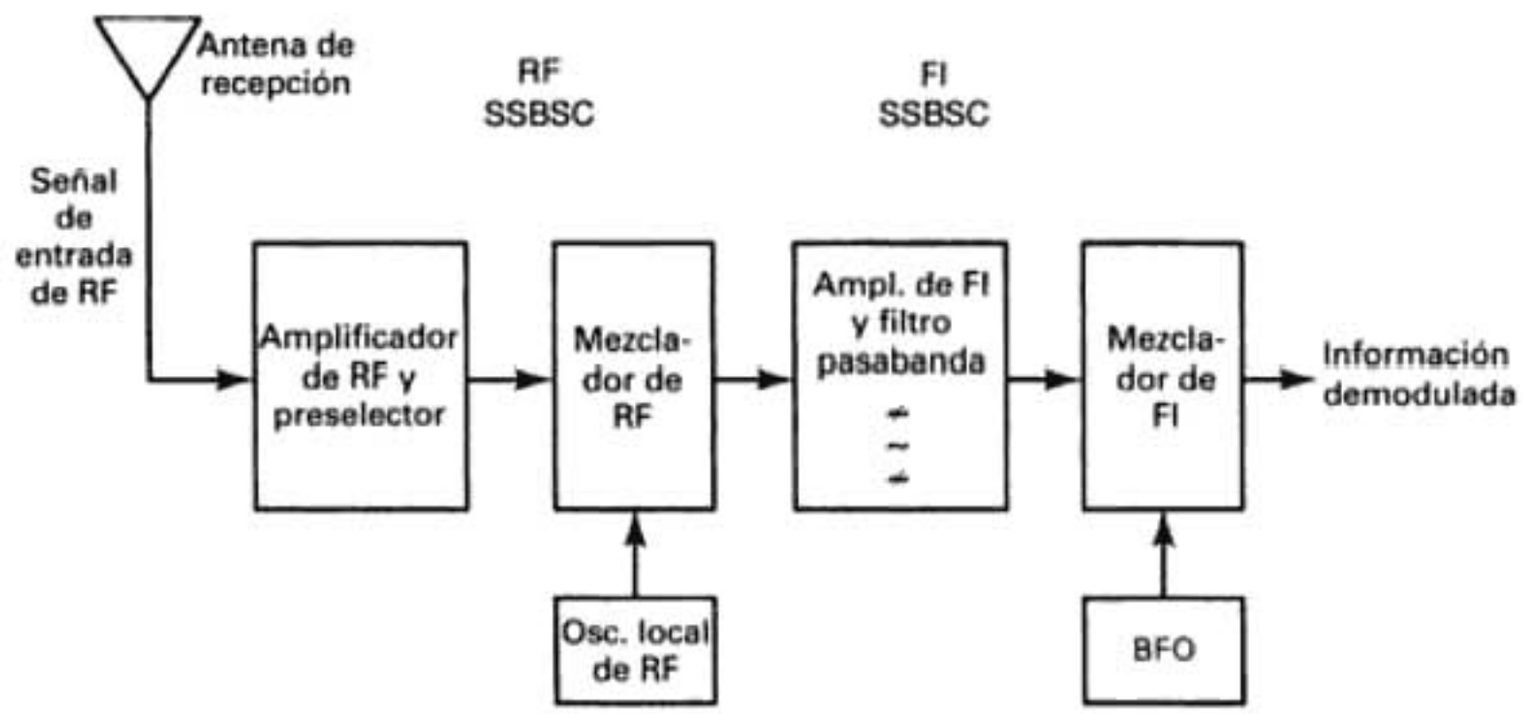


FIGURA 5-21 Receptor SSB BFO no coherente

El mezclador de RF y el segundo detector de la fig. 5-21 son detectores de producto. Como en los moduladores balanceados en el transmisor, sus salidas son el producto de sus entradas. Un modulador de producto y un detector de producto son el mismo circuito, en esencia. La única diferencia es que la entrada al modulador de producto se sintoniza a una señal moduladora de baja frecuencia, y la salida a una portadora de alta frecuencia, mientras que en un detector de producto la entrada se sintoniza a una portadora modulada de alta frecuencia y la salida a una señal de información de baja frecuencia. Tanto en el modulador como el detector, la portadora de frecuencia única es la señal de conmutación. En un receptor, la señal de entrada, que es una portadora de RF suprimida y reducida y una banda lateral, se mezcla con la RF del oscilador local para producir una frecuencia intermedia. La salida del segundo detector de producto tiene las frecuencias de suma y diferencia entre la FI y la frecuencia de pulso. La banda de frecuencias de diferencia es la información original de entrada.

Ejemplo 5-2

El receptor BFO de la fig. 5-21 tiene una banda de RF de recepción de 30 a 30.005 MHz, oscilador local de RF de 20 MHz, una banda de FI de 10 a 10.005 MHz y una frecuencia BFO de 10 MHz. Determinar:

(a) La primera banda de FI demodulada y la banda de frecuencias de información demodulada.

(b) Banda de frecuencias de información demoduladas si la RF del oscilador local se corre hacia abajo 0.001%.

Solución (a) La salida de FI del mezclador de RF es la diferencia entre la frecuencia de la señal recibida y la frecuencia del oscilador local de RF, es decir

$$f_{FI} = (30 \text{ MHz a } 30.005 \text{ MHz}) - 20 \text{ MHz} = 10 \text{ MHz a } 10.005 \text{ MHz}$$

El espectro de la señal de información demodulada es la diferencia entre la banda de frecuencia intermedia y la frecuencia BFO, o sea

$$f_m = (10 \text{ MHz a } 10.005 \text{ MHz}) - 10 \text{ MHz} = 0 \text{ kHz a } 5 \text{ kHz}$$

(b) Un desplazamiento de 0.001% causaría una disminución en la RF del oscilador local de

$$\Delta f = (0.00001)(20 \text{ MHz}) = 200 \text{ Hz}$$

Así, la RF del oscilador local se desplazaría hacia abajo hasta 19.9998 MHz, y la salida del mezclador de RF es

$$\begin{aligned} f_{FI} &= (30 \text{ MHz a } 30.005 \text{ MHz}) - 19.9998 \text{ MHz} \\ &= 10.0002 \text{ MHz a } 10.0052 \text{ MHz} \end{aligned}$$

El espectro de la señal demodulada de información es la diferencia entre la banda de frecuencia intermedia y el BFO, es decir

$$\begin{aligned} f_m &= (10.0002 \text{ MHz a } 10.0052 \text{ MHz}) - 10 \text{ MHz} \\ &= 200 \text{ Hz to } 5200 \text{ Hz} \end{aligned}$$

El desplazamiento de 0.001% de la RF del oscilador local causó un error de 200 Hz en el espectro de la señal demodulada de información.

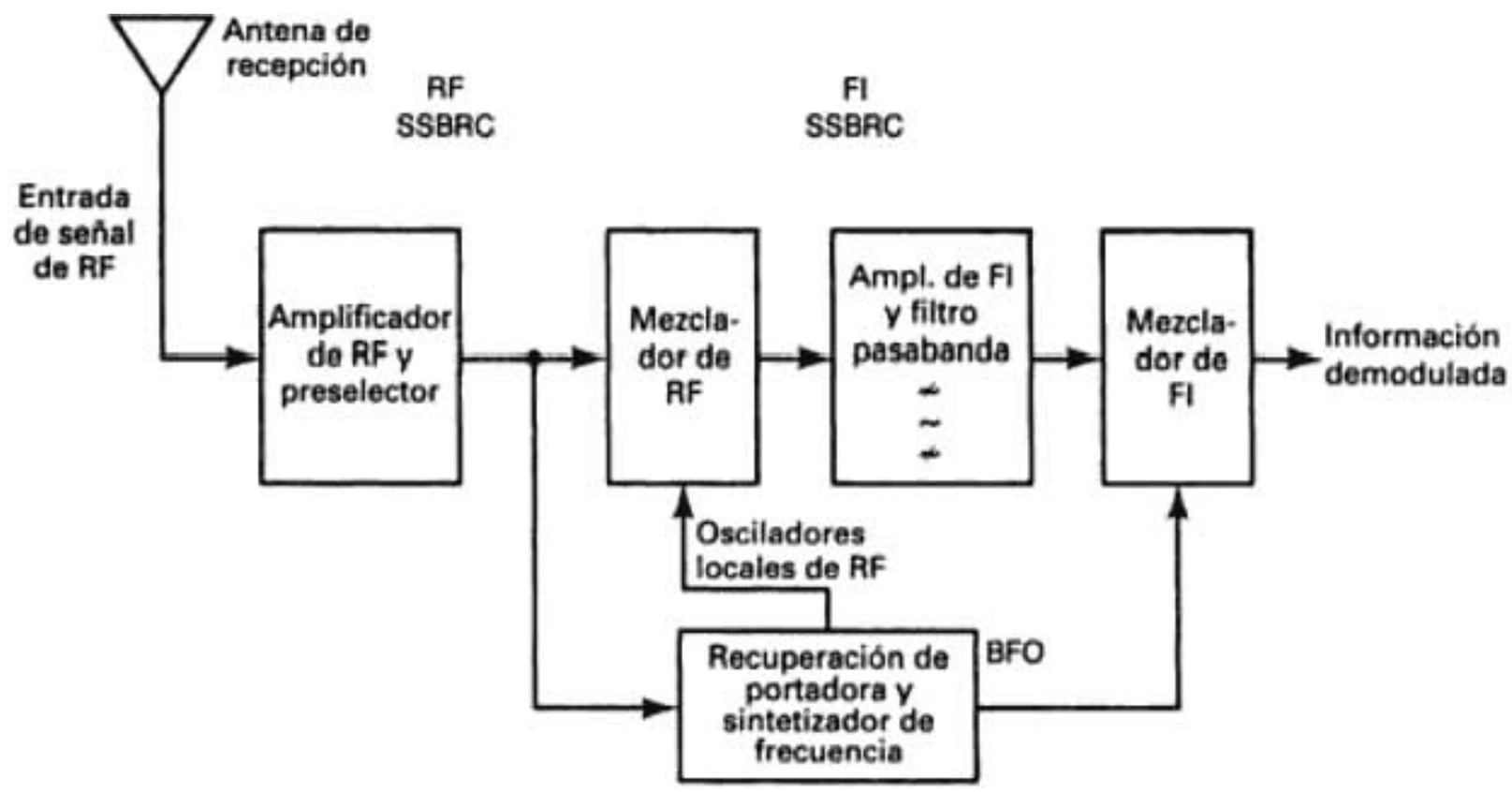


FIGURA 5-22 Receptor SSB BFO coherente

Receptor BFO coherente de banda lateral única

La fig. 5-22 muestra el diagrama de bloques de un receptor BFO coherente de banda lateral única. Este receptor es idéntico al de la fig. 5-21, excepto que las frecuencias de oscilador local y BFO se sincronizan a los osciladores de portadora en el transmisor. El *circuito de recuperación* de portadora es un PLL (lazo de fase cerrada) que rastrea la portadora piloto en la señal compuesta del receptor SSBRC y usa la portadora recuperada para regenerar frecuencias coherentes de oscilador local en el sintetizador. El circuito sintetizador produce una RF coherente de oscilador local y frecuencia BFO. El circuito de recuperación de portadora rastrea la portadora piloto recibida. En consecuencia, se compensan cambios pequeños de frecuencia de la portadora en el transmisor, y se elimina el error de desplazamiento de frecuencia. Si el receptor coherente de la fig. 5-22 se hubiera usado en el ejemplo 5-2, no se habría permitido desplazarse en forma independiente al oscilador local de RF.

Ejemplo 5-3

Para el receptor BFO coherente de banda lateral única de la fig. 5-22, una RF de portadora reducida de 30 MHz con una banda lateral superior que va desde justo arriba de 30 hasta 30.005 MHz, una RF del oscilador local de 20 MHz, una FI central de 10 MHz y una frecuencia de salida BFO de 10 MHz, determinar:

(a) La primera banda de FI demodulada y la banda de frecuencias de información demoduladas.

(b) La banda de frecuencias de información demodulada, si la RF de entrada de portadora subiera 60 Hz, produciendo una RF de portadora de 30.000000 MHz y una banda lateral superior que va hasta 30.000060 MHz.

Solución (a) La solución es idéntica a la que se presentó en el ejemplo 5-2. La única diferencia es el método con el que se producen las frecuencias de RF del oscilador y de BFO. En el receptor coherente, la RF del oscilador local y la BFO se producen en el circuito de recuperación de portadora y, en consecuencia, están sincronizadas a la RF de la portadora recibida.

$$f_{FI} = (30 \text{ MHz a } 30.005 \text{ MHz}) - 20 \text{ MHz} = 10 \text{ MHz a } 10.005 \text{ MHz}$$

El espectro de señal demodulada de información no es más que la diferencia entre la banda de frecuencias intermedias y la frecuencia BFO

$$f_m = (10 \text{ MHz a } 10.005 \text{ MHz}) - 10 \text{ MHz} = 0 \text{ Hz a } 5 \text{ kHz}$$

(a) Como las frecuencias RF del oscilador local y BFO se sincronizan a la señal recibida de portadora de RF, el oscilador local de RF se desplazará en forma proporcional al cambio de la RF de la señal de entrada. En consecuencia, la RF del oscilador local se ajustará en forma automática a 20.0004 MHz, y producirá un espectro de FI de

$$f_{FI} = (30.00006 \text{ MHz a } 30.00506 \text{ MHz}) - 20.0004 \text{ MHz} = 10.0002 \text{ MHz a } 10.00016 \text{ MHz}$$

La frecuencia de salida del BFO también se ajustará en forma automática y proporcional a 10.0002 MHz, produciendo una señal de información demodulada de

$$f_m = (10.0002 \text{ MHz a } 10.0052 \text{ MHz}) - 10.0002 \text{ MHz} = 0 \text{ Hz a } 5 \text{ kHz}$$

Se puede ver en el ejemplo 5-3 que el receptor coherente de banda lateral única se ajusta en forma automática a desplazamientos de frecuencia de la portadora transmitida. Por consiguiente, este receptor coherente es inmune a desplazamientos de portadora, siempre que la magnitud del desplazamiento esté dentro de los límites del circuito de recuperación de portadora.

Receptor de detección de envolvente con banda lateral única

La fig. 5-23 muestra el diagrama de bloques de un receptor de banda lateral única que usa portadoras síncronas y detección de envolvente para demodular las señales recibidas. Se detecta la portadora reducida piloto, se separa del espectro demodulado y se regenera en el circuito de recuperación de portadora. La piloto regenerada se divide y usa como fuente de frecuencia estable para un sintetizador de frecuencias, que suministra al receptor con osciladores locales de frecuencia coherente. La RF de recepción se baja a FI en el primer detector. Una portadora regenerada de FI se agrega al espectro de FI en el último sumador lineal, que produce una envolvente de portadora de máxima potencia y banda lateral única. La envolvente se demodula en un detector convencional de picos de diodo, para producir el espectro de la señal original de información. A este tipo de receptor se le llama a menudo receptor de portadora *exaltada* o *reacondicionada*.

Receptor de banda lateral única con portadora piloto en multicanales

La fig. 5-24 es un diagrama de bloques de un receptor SSB de portadora piloto multicanal, que usa un circuito de lazo de fase cerrada (PLL) para recuperación de portadora, y un sintetizador de frecuencia para producir frecuencias de oscilador local coherente y de pulso. El intervalo de entrada de RF va de 4 a 30 MHz, y la frecuencia natural del VCO (*oscilador controlado por voltaje*) se ajusta aproximadamente con un selector externo de canal al intervalo de frecuencias de 6 a 32 MHz. La frecuencia del VCO rastrea 2 MHz arriba de la RF de entrada, que es la primera FI. Una frecuencia de pulso de 1.8 MHz establece la segunda FI en 200 kHz.

La frecuencia VCO se establece en forma aproximada con el selector de canales, y a continuación se mezcla con la señal de RF de entrada en el primer detector, para producir una

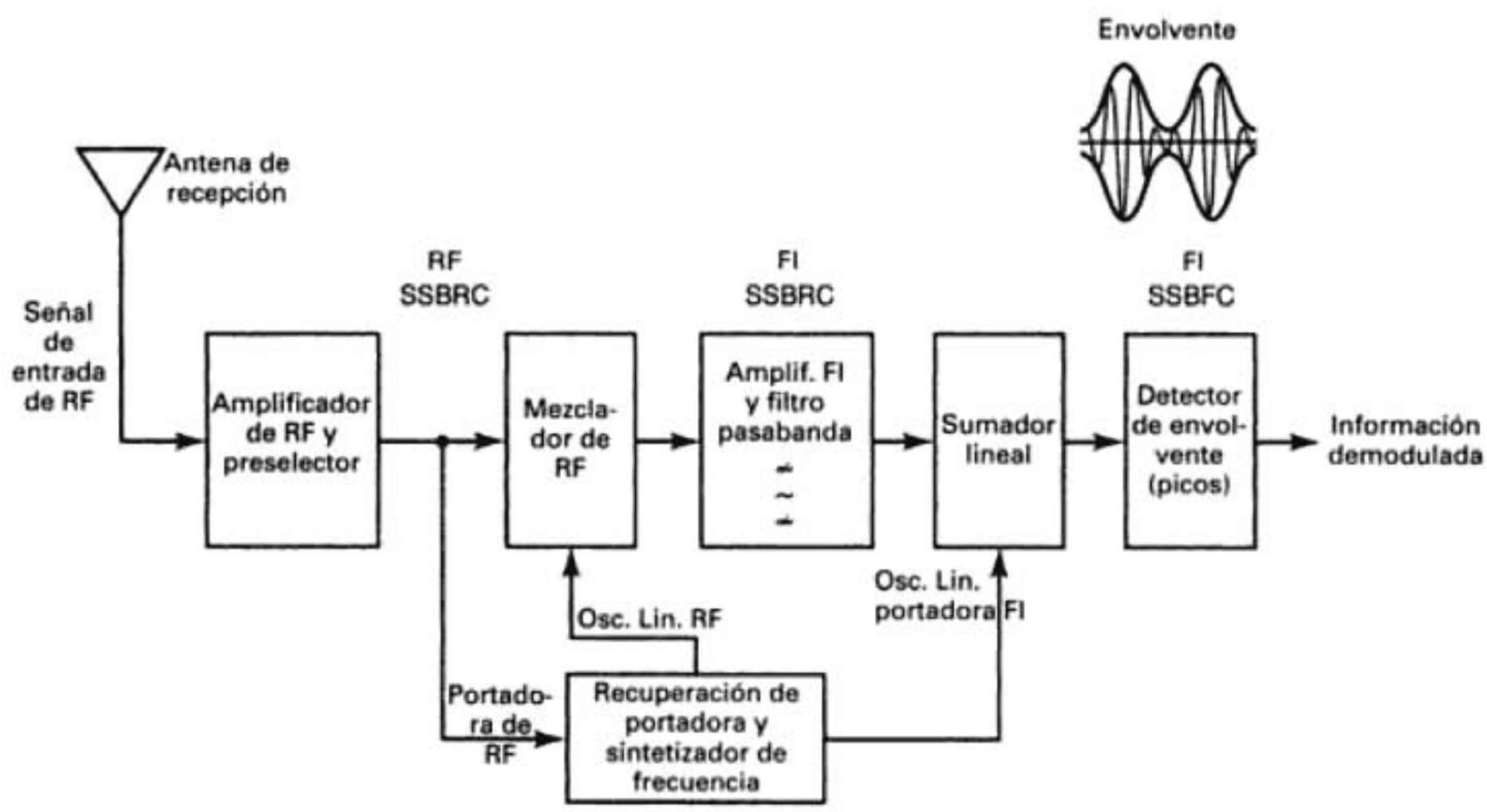


FIGURA 5-23 Receptor de banda lateral única con detección de envolvente

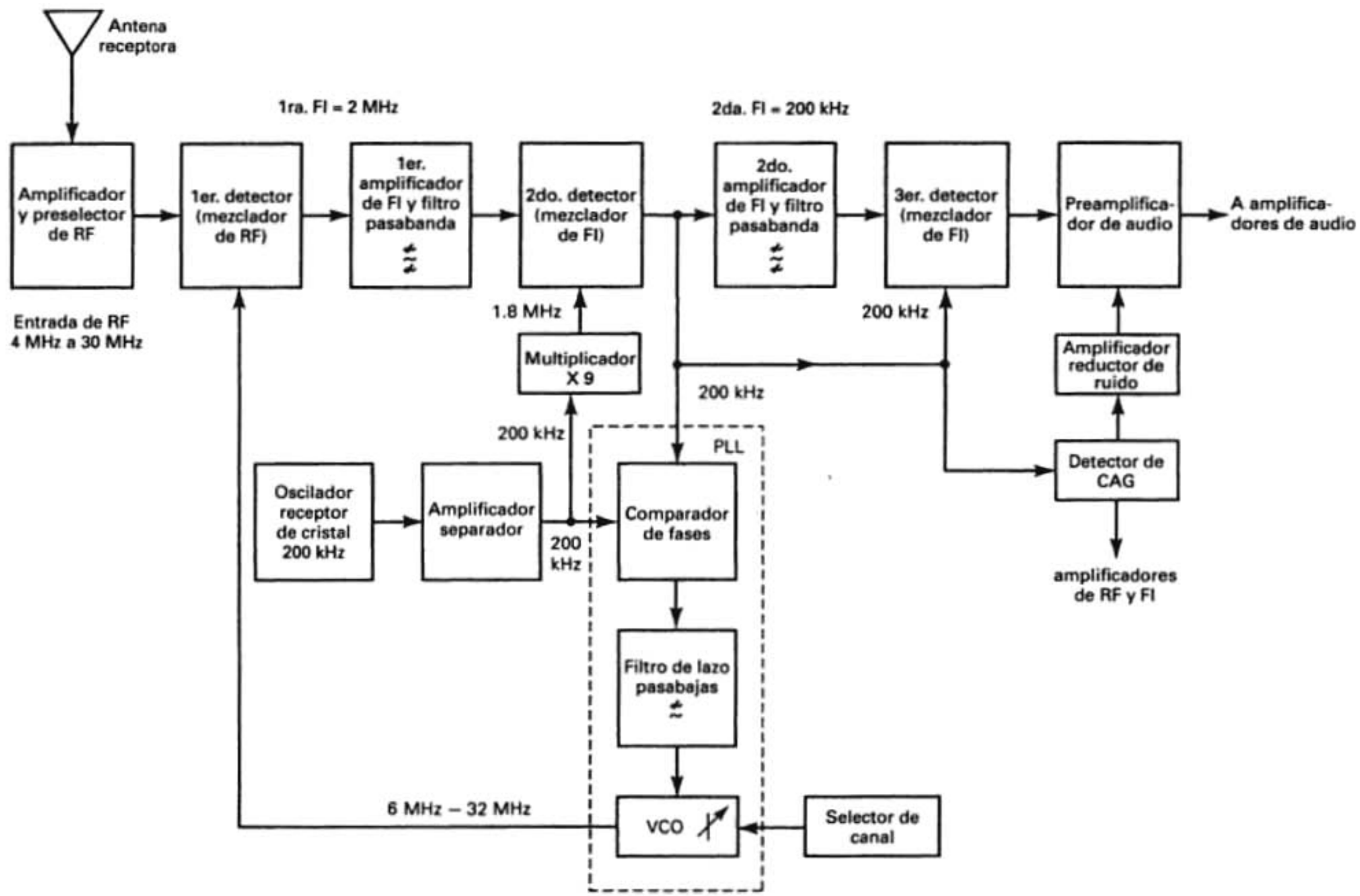


FIGURA 5-24 Receptor SSB de portadora piloto y varios canales

frecuencia FI de diferencia de 2 MHz. La primera FI se mezcla con la frecuencia de pulso local de 1.8 MHz para producir una segunda FI de 200 kHz. El PLL se ajusta a la piloto de 200 kHz y produce un voltaje de corrección que hace el ajuste fino del VCO. La segunda FI se reduce a audio en el tercer detector, y la señal pasa al preamplificador de audio para su procesamiento posterior. El detector del control automático de ganancia produce un voltaje CAG que es proporcional a la amplitud de la piloto de 200 kHz. El voltaje CAG se retroalimenta a los amplificadores de RF y/o de FI, para ajustar sus ganancias en proporción con el valor piloto recibido, y al circuito reductor de ruido para apagar el preamplificador de audio en ausencia de recepción de piloto. El PLL compara la piloto de 200 kHz con una referencia estable controlada por cristal. En consecuencia, aunque no se sincroniza en forma directa la portadora del receptor a los osciladores de transmisión, la primera y segunda FI compensan de esta manera cualquier desplazamiento de frecuencia en el espectro de audio demodulado.

COMPRESIÓN Y EXPANSIÓN DE AMPLITUD DE BANDA LATERAL ÚNICA

Los sistemas de *compresión y expansión de amplitud de banda lateral única* (ACSSB, de *amplitude compandoring single-sideband*) proporcionan comunicaciones de voz en banda angosta, para servicios móviles terrestres casi con la calidad que se logra en los sistemas de FM, y lo hacen usando menos de la tercera parte del ancho de banda. Con ACSSB, las señales de audio se comprimen antes de la modulación, amplificando las de mayor magnitud menos que las de menor magnitud. Después de la demodulación en el receptor, las señales de audio se expanden amplificando las de mayor magnitud más que las de menor magnitud. Un dispositivo que hace compresión y expansión se llama *compansor*.

Al comprimir y expandir una señal de información se aumenta el intervalo dinámico de un sistema, al reducir el de las señales de información antes de su transmisión, para después expandirlas luego de la demodulación. Por ejemplo, cuando se usa *compansión* (compresión y expansión), se pueden propagar señales dentro de un intervalo dinámico de 80 dB a través de un sistema de comunicaciones que sólo tenga 50 dB de intervalo dinámico. La compresión y expansión reduce un poco las relaciones de señal a ruido de las señales de gran amplitud, mientras que aumenta mucho las de las señales de baja amplitud.

Los sistemas ACSSB requieren transmitir una señal portadora piloto de amplitud reducida, con las señales de información. El piloto se usa para sincronizar los osciladores en el receptor y proporciona una señal para el CAG, que vigila y ajusta la ganancia del receptor, y lo calla cuando no se recibe piloto.

Con la ACSSB se reduce bastante el intervalo dinámico, permitiendo que las señales de bajo nivel se transmitan con mayor potencia, a la vez que se permanece dentro de los límites de potencia del transmisor cuando hay señales de alto nivel. En consecuencia, se mejora mucho la relación de señal a ruido, para señales de bajo nivel, y se introduce también un aumento insignificante en los niveles de ruido para las señales de más intensidad.

La fig. 5-25 muestra la ubicación y amplitudes relativas de la banda de paso de audio y el tono piloto de un sistema ACSSB. Se ve que la piloto se transmite a 10 dB menos que el nivel máximo de potencia, para señales dentro de la banda de paso de audio.

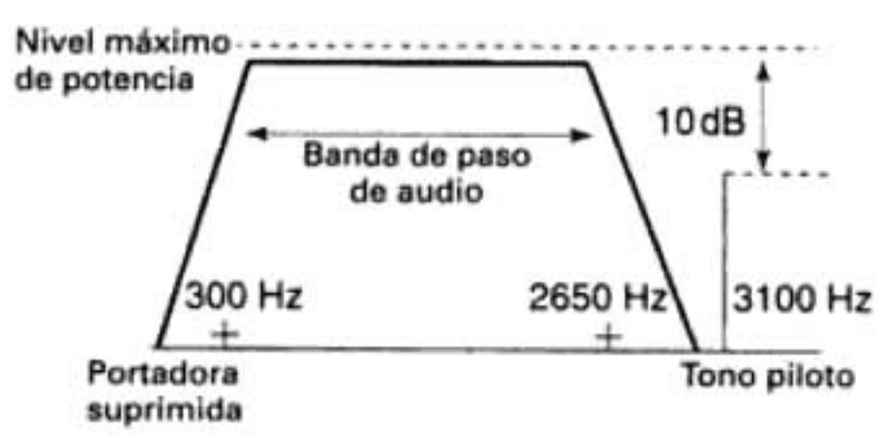


FIGURA 5-25 Señal ACSSB

BANDA LATERAL UNICA CON PORTADORA SUPRIMIDA Y MULTIPLEXADO POR DIVISIÓN DE FRECUENCIAS

A causa del ancho de banda y las eficiencias de potencia inherentes a la banda lateral única con portadora suprimida, su aplicación más común es el *multiplexado por división de frecuencias* (FDM, de *frequency-division multiplexing*). En general el *multiplexado*, es el proceso de combinar transmisiones procedentes de más de una fuente y transmitirlas en un solo sistema, como por ejemplo un cable metálico o uno de fibra óptica, o por un canal de radiofrecuencia. El multiplexado por división de frecuencia es un método analógico de combinar dos o más fuentes analógicas que ocupaban originalmente la misma banda de frecuencias, de tal manera que los canales no interfieran entre sí. El FDM se usa mucho para combinar muchas fuentes de banda relativamente angosta y formar un solo canal de banda ancha, como por ejemplo en los sistemas telefónicos públicos.

Con el FDM, cada canal de banda angosta pasa a un lugar distinto en el espectro total de frecuencias. Los canales se apilan, esencialmente, uno sobre otro en el dominio de la frecuencia. La fig. 5-26 muestra un sistema FDM sencillo, en el que se multiplexan cuatro canales de 5 kHz, por división de frecuencia, y forman un solo canal combinado de 20 kHz. Como se ve en la fig. 5-26a, las señales del canal 1 modulan a una portadora de 100 kHz en un modulador balanceado, que en forma inherente suprimió la portadora de 100 kHz. La salida del modulador balanceado es una forma de onda con portadora suprimida y doble banda lateral, con ancho de banda de 10 kHz. La onda DSBSC pasa por un filtro pasabanda (BPF) donde se convierte en señal SSBSC. Para este ejemplo se bloquea la banda lateral inferior, así que la salida del BPF ocupa la banda de frecuencias de 100 a 105 kHz (un ancho de banda de 5 kHz).

Las señales del canal 2 modulan una portadora de 105 kHz en un modulador balanceado, y vuelven a producir una forma de onda DSBSC, que se convierte a SSBSC pasándola por un filtro pasabanda sintonizado para pasar sólo las frecuencias de banda lateral superior. Así, la salida del BPF ocupa una banda de frecuencias de 105 a 110 kHz. Con el mismo proceso se convierten las señales de los canales 3 y 4 a las bandas de frecuencia de 110 a 115 kHz, y de 115 a 120 kHz, respectivamente. El espectro combinado de frecuencias, con las salidas de los cuatro filtros pasabanda, se ve en la fig. 5-26b. Ahí se ve que el ancho total combinado de banda es igual a 20 kHz, y cada canal ocupa una parte distinta de 5 kHz del ancho total de 20 kHz. Además, las cuatro portadoras se han suprimido y permiten concentrar la potencia disponible en las señales de banda lateral.

Se puede usar la transmisión de banda lateral única y portadora suprimida para combinar cientos, o hasta miles de canales de banda angosta, como circuitos de voz o de datos de baja velocidad, en un solo canal compuesto de banda ancha, sin que los canales interfieran entre sí. Véase el capítulo 16, donde se presenta una descripción más detallada del multiplexado por división de frecuencias.

DOBLE BANDA LATERAL CON PORTADORA SUPRIMIDA Y MULTIPLEXADO EN CUADRATURA

El *multiplexado en cuadratura* (QM, de *quadrature multiplexing*) es un método de multiplexado que usa transmisión de doble banda lateral y portadora suprimida para combinar dos fuentes de información en una sola forma de onda compuesta, que a continuación se transmite por un solo sistema, sin que los dos canales interfieran entre sí.

La fig. 5-27 muestra cómo se combinan dos fuentes de información en un solo canal de comunicaciones, mediante multiplexado en cuadratura. Como allí se ve, la información de cada canal modula la misma señal portadora en moduladores balanceados. La principal diferencia entre los dos moduladores es que la portadora del canal Q se ha desplazado 90° en fase respecto a la portadora del canal 1. Se dice que las dos portadoras están *en cuadratura* entre sí. Así, la salida del modulador del canal 1 es el producto de las señales de información de la fuente 1 [$v_{m1}(t)$] y la señal de portadora en fase [sen($\omega_c t$)]. La salida del modulador del canal 2 es el producto de las señales de información de la fuente 2 [$v_{m2}(t)$] y una portadora de señal cuya fase se ha desplazado

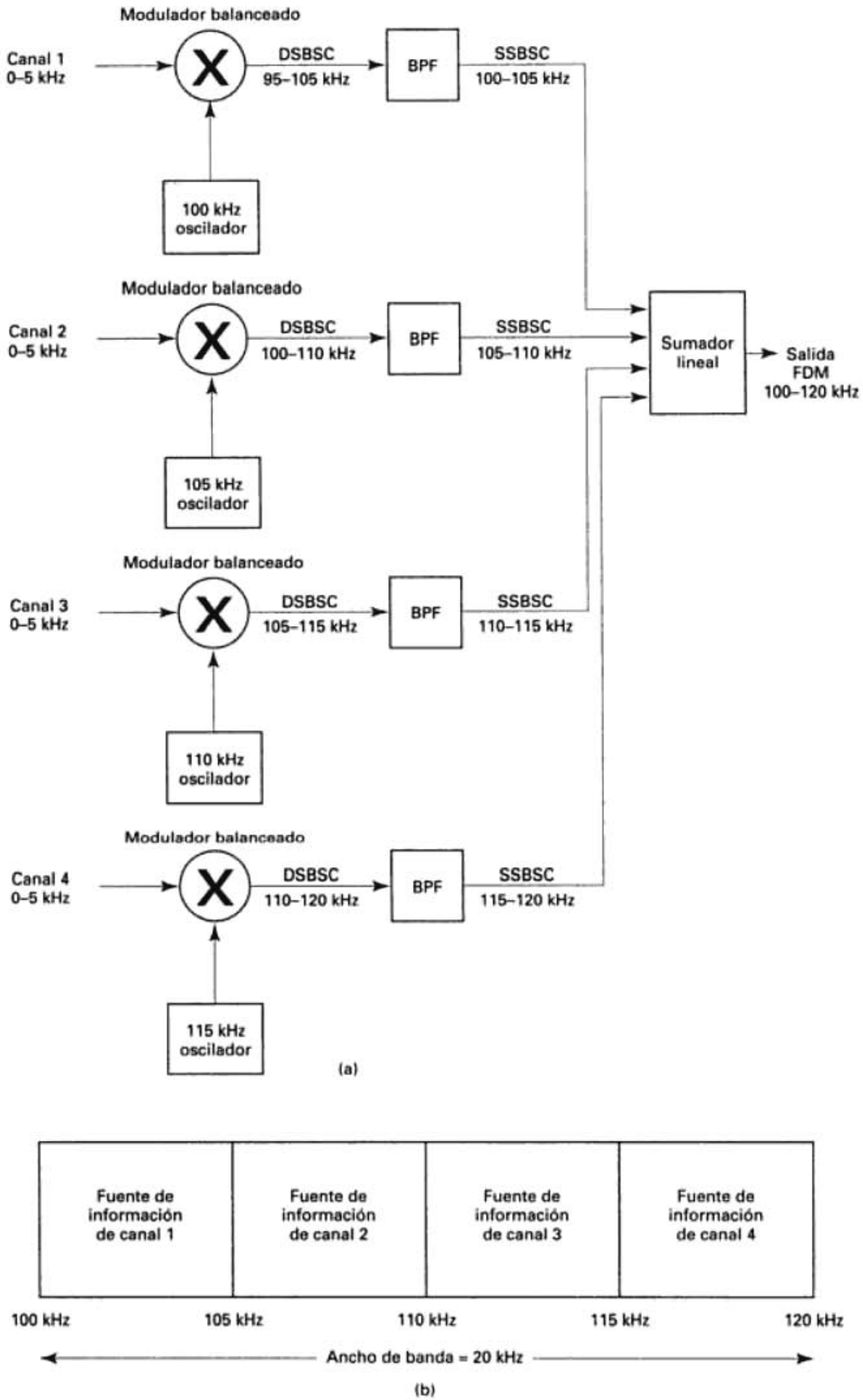


FIGURA 5-26 Sistema SSBSC FDM: (a) diagrama de bloques; (b) espectro de frecuencias de salida

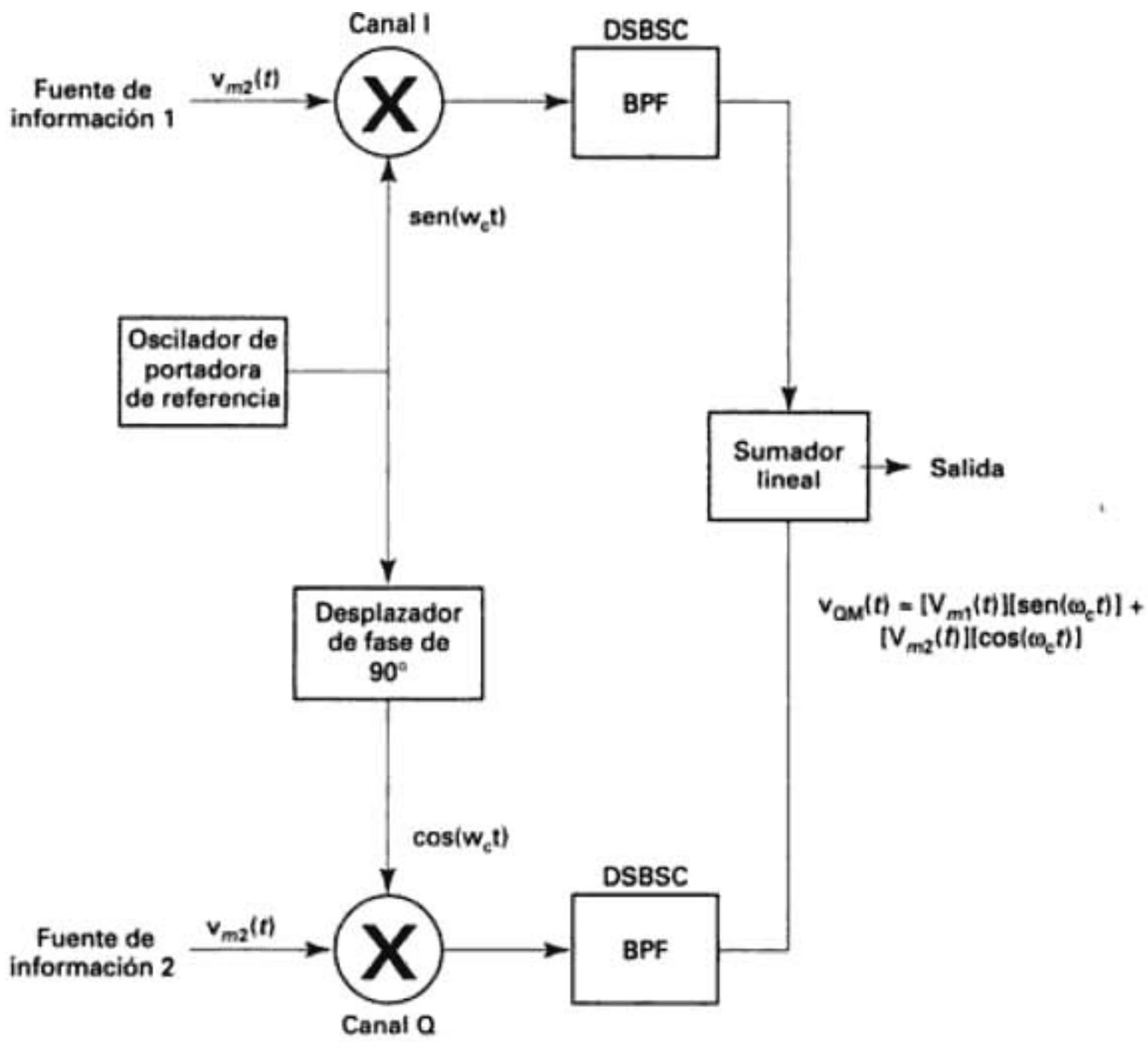


FIGURA 5-27 Diagrama de bloques del sistema DSBSC QM

90° respecto al oscilador de referencia [cos($\omega_c(t)$]. Las salidas de los dos filtros pasabanda se combinan en un sumador lineal y producen una forma de onda compuesta, formada por las dos señales ortogonales (a 90°) de doble banda lateral, simétricas respecto a la misma portadora suprimida, $v_{QM}(t) = [v_{m1}(t)][\text{sen}(\omega_c t)] + [v_{m2}(t)][\cos(\omega_c t)]$. Así, los dos canales están separados en el dominio de fase. El multiplexado en cuadratura se usa en forma normal para combinar canales de información en módems de datos (capítulo 13) y para multiplexar señales de color en televisión de banda ancha (capítulo 11).

MEDICIONES DE BANDA LATERAL ÚNICA

Los transmisores de banda lateral única se especifican por la *potencia máxima de envolvente* (PEP, de *peak envelope power*) y por el *voltaje máximo de envolvente* (PEV, de *peak envelope voltage*), y no tan sólo por la potencia y el voltaje efectivos (rms). Para una señal moduladora de frecuencia única, la señal modulada de salida con transmisión de banda lateral única y portadora suprimida no es una envolvente, sino más bien una señal continua de una sola frecuencia. Una sola frecuencia no representa una señal normal de información. Así, para fines de prueba, se usa una señal de prueba de *dos frecuencias*, como señal moduladora en la cual los dos tonos tienen amplitudes iguales. La fig. 5-28a muestra la forma de onda producida en un modulador SSBSC con una señal moduladora de dos tonos. Esa forma de onda es la suma vectorial de las dos frecuencias laterales de igual amplitud, y se parece a la forma de onda de AM convencional, pero la rapidez de repetición es igual a la diferencia de las frecuencias de las dos señales moduladoras. La fig. 5-28b muestra la envolvente de una señal de prueba de dos tonos, cuando se agrega una portadora piloto de poca amplitud. La envolvente tiene básicamente la misma forma, excepto por la adición de una ondulación senoidal de baja amplitud a la frecuencia de portadora.

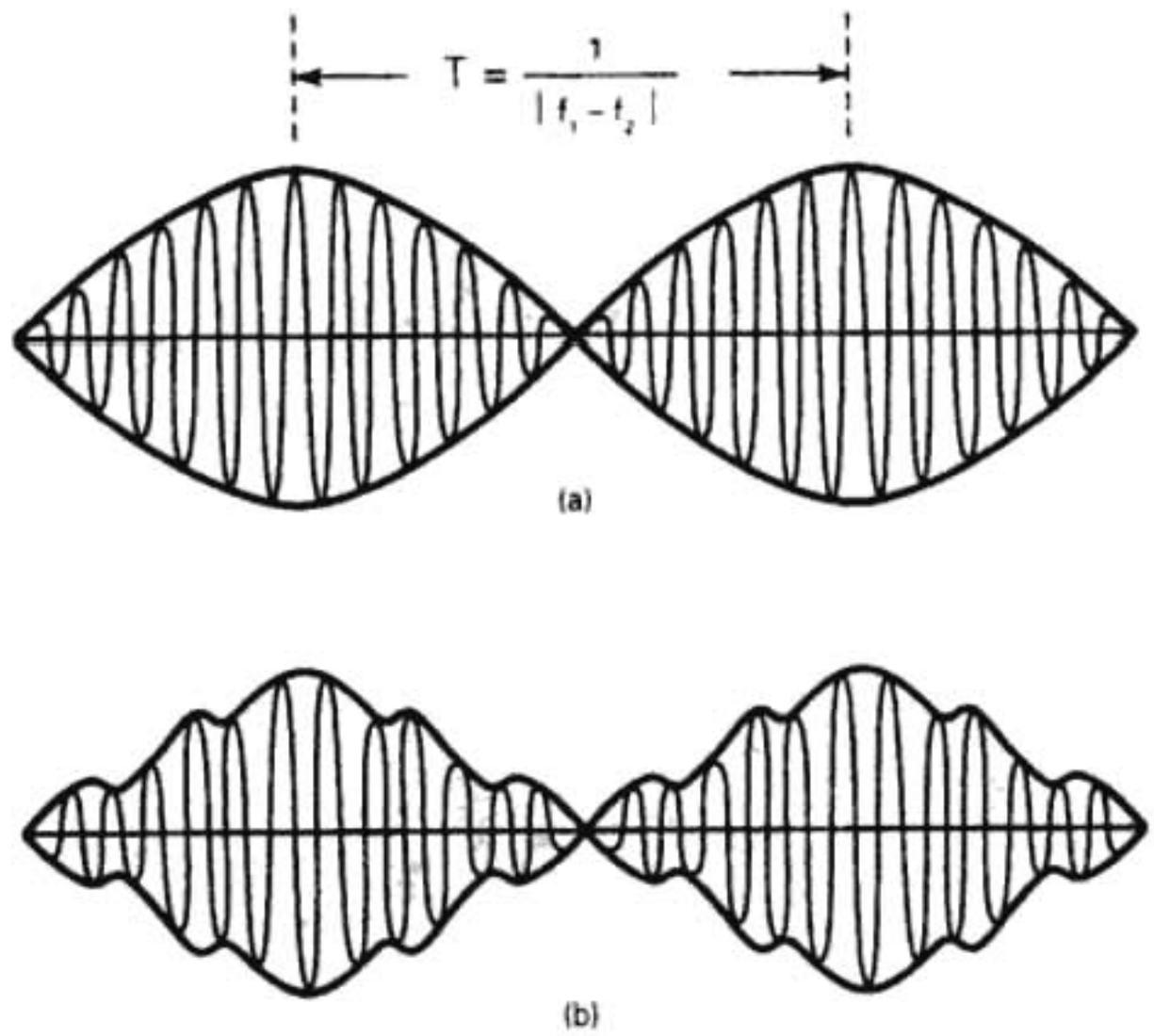


FIGURA 5-28 Señal de prueba SSB de dos tonos: (a) sin portadora reinsertada; (b) con portadora reinsertada

La envolvente de salida de SSB con dos tonos es una consideración importante, porque de esta envolvente se determina la potencia de salida de un transmisor SSB. La PEP de un transmisor SSBSC es análoga a la potencia total de salida de un transmisor convencional de doble banda lateral y portadora de máxima potencia. La PEP nominal es la potencia de salida, medida en el máximo de la envolvente, cuando la entrada es una señal de prueba de dos tonos, y los dos tonos tienen igual magnitud. Con esa señal de salida, la potencia real disipada en la carga es igual a la mitad de la PEP. Por lo anterior, el voltaje desarrollado a través de la carga es

$$e_{\text{total}} = \sqrt{E_1^2 + E_2^2}$$

siendo E_1 y E_2 los voltajes rms de los dos tonos de prueba. Entonces,

$$\text{PEP} = \frac{(\sqrt{E_1^2 + E_2^2})^2}{R}$$

y como $E_1 = E_2$,

$$\text{PEP} = \frac{2E^2}{R} \tag{5-2}$$

Sin embargo, la potencia promedio disipada en la carga es igual a la suma de las potencias de los dos tonos:

$$P_{\text{prom}} = \frac{E_1^2}{2R} + \frac{E_2^2}{2R} = \frac{2E^2}{2R} = \frac{E^2}{R} \tag{5-3}$$

que se simplifica a

$$P_{\text{prom}} = \frac{\text{PEP}}{2} \tag{5-4}$$

En esta prueba se usan dos tonos de igual amplitud, por las siguientes razones:

1. Un tono produce una salida continua de una sola frecuencia que no produce intermodulación.
2. Una señal de salida de una sola frecuencia no es análoga a una señal normal de información.

3. El análisis es impráctico con más de dos tonos.
4. Dos tonos de igual amplitud imponen al transmisor requisitos más ambiciosos de los que probablemente se presenten durante la operación normal.

Ejemplo 5-4

Una señal de prueba es de dos tonos, de 1.5 y 3 kHz, y la frecuencia de la portadora es 100 kHz. Determinar lo siguiente, para una transmisión de banda lateral única con portadora suprimida.

(a) Espectro de frecuencias de salida si sólo se transmite la banda lateral superior.

(b) Para $E_1 = E_2 = 5$ V, y resistencia de carga de 50 Ω, la PEP y la potencia promedio de salida.

Solución (a) El espectro de frecuencias de salida contiene las dos frecuencias laterales superiores

$$f_{fls1} = 100 \text{ kHz} + 1.5 \text{ kHz} = 101.5 \text{ kHz}$$
$$f_{fls2} = 100 \text{ kHz} + 3 \text{ kHz} = 103 \text{ kHz}$$

(b) Al sustituir en la ecuación 5-2 se obtiene

$$\text{PEP} = \frac{2(0.707 \times 5)^2}{50} = 0.5 \text{ W}$$

De acuerdo con la ecuación 5-4 obtiene

$$P_{\text{prom}} = \frac{\text{PEP}}{2} = \frac{0.5}{2} = 0.25 \text{ W}$$

PREGUNTAS

5-1. Describa la SSBFC de AM. Compare la SSBFC con la AM convencional.
5-2. Describa la SSBSC de AM. Compare la SSBSC con la AM convencional.
5-3. Describa la SSBRC de AM. Compare la SSBRC con la AM convencional.
5-4. ¿Qué es una *portadora piloto*?
5-5. ¿Qué es una *portadora exaltada*?
5-6. Describa la AM de ISB. Compare ISB con la AM convencional.
5-7. Describa la AM de VSB. Compare VSB con la AM convencional.
5-8. Defina *potencia máxima de envolvente*.
5-9. Describa el funcionamiento de un modulador de anillo balanceado.
5-10. ¿Qué es un modulador de producto?
5-11. Describa el funcionamiento de un modulador balanceado en contrafase con FET.
5-12. Describa la operación de un modulador de puente balanceado.
5-13. ¿Cuáles son las ventajas de un modulador balanceado en circuito integrado lineal en comparación con uno de circuito discreto?
5-14. Describa el funcionamiento de un transmisor SSB tipo filtro.
5-15. Describa las diferencias entre los filtros de cristal, de cerámica y mecánicos.
5-16. Describa el funcionamiento de un transmisor SSB del tipo por desplazamiento de fase.
5-17. Describa la operación de un transmisor SSB del "tercer tipo".
5-18. Describa el funcionamiento de un transmisor de banda lateral independiente.
5-19. ¿Cuál es la diferencia entre un modulador de producto y un detector de producto?
5-20. ¿Cuál es la diferencia entre un receptor coherente y uno no coherente?
5-21. Describa el funcionamiento de un receptor SSBRC con portadora piloto y multicanal.
5-22. ¿Por qué se usa una señal de prueba de dos tonos para hacer mediciones de PEP?

PROBLEMAS

5-1. Para el modulador de anillo balanceado de la fig. 5-6a, con una frecuencia de entrada de portadora $f_c = 400$ kHz y un intervalo de frecuencias de señal moduladora $f_m = 0$ a 4 kHz, determine:

(a) El espectro de frecuencias de salida.
(b) La frecuencia de salida para una entrada de frecuencia única $f_m = 2.8$ kHz.

5-2. Para el modulador LIC balanceado de la fig. 5-12, con frecuencia de entrada de portadora de 200 kHz e intervalo de frecuencias de señal moduladora $f_m = 0$ a 3 kHz, determine:
(a) El espectro de frecuencias de salida.
(b) La frecuencia de salida para una entrada de frecuencia única $f_m = 1.2$ kHz.

5-3. Un transmisor de SSB como el de la fig. 5-13 tiene portadora de frecuencia baja de 100 kHz, portadora de frecuencia media de 4 MHz, portadora de alta frecuencia de 30 MHz e intervalo de frecuencias de señal moduladora de 0 a 4 kHz.
(a) Trace los espectros de frecuencia en los siguientes puntos: salida del modulador balanceado 1, salida del BPF 1, salida del sumador, salida del modulador balanceado 2, salida del BPF 2, salida del modulador balanceado 3 y salida del BPF 3.
(b) Para la entrada de frecuencia única $f_m = 1.5$ kHz, determine la frecuencia trasladada en los siguientes puntos: salida de BPF 1, salida de BPF 2 y salida de BPF 3.

5-4. Repita el problema 5-3, pero cambie a 500 kHz la portadora de baja frecuencia. ¿Cuál transmisor tiene los requisitos de filtrado más rigurosos?

5-5. En el transmisor SSB de la fig. 5-14a se usa intervalo de frecuencias moduladoras de entrada de 0 a 3 kHz y una portadora de alta frecuencia de 28 MHz.
(a) Trace el espectro de frecuencias de salida.
(b) Determine la frecuencia de salida para una señal de entrada de frecuencia moduladora única de 2.2 kHz.

5-6. Repita el problema 5-5, pero cambie el intervalo de frecuencias de entrada de audio de 300 a 5000 Hz.

5-7. Para el transmisor SSB de la fig. 5-18, con frecuencia de portadora de entrada de 500 kHz y un intervalo de frecuencias de señal moduladora de 0 a 4 kHz.
(a) Trace el espectro de frecuencias a la salida del sumador lineal.
(b) Determine la frecuencia de salida con una señal moduladora de frecuencia única de 3 kHz en la entrada.

5-8. Repita el problema 5-7, pero cambie la frecuencia de portadora de entrada a 400 kHz y el intervalo de frecuencia moduladora para que abarque de 300 a 5000 Hz.

5-9. Para el transmisor ISB de la fig. 5-20, el intervalo de frecuencias de entrada al canal A es de 0 a 4 kHz, el de entrada al canal B es de 0 a 4 kHz, la portadora de baja frecuencia tiene 200 kHz, la de frecuencia media tiene 4 MHz y la portadora de alta frecuencia es de 32 MHz.
(a) Trace los espectros de frecuencia en los siguientes puntos: salida de modulador balanceado A, salida del BPF A, salida del modulador balanceado B, salida del BPF B, salida de la red híbrida, salida del sumador lineal, salida del modulador balanceado 3, salida del BPF 3, salida del modulador balanceado 4 y salida del BPF 4.
(b) Para una frecuencia de entrada al canal A de 2.5 kHz y al canal B de 3 kHz, determine los componentes de frecuencia en los siguientes puntos: salida de BPF A, salida de BPF B, salida de BPF 3 y salida de BPF 4.

5-10. Repita el problema 5-9, pero cambiando el intervalo de frecuencias de entrada del canal A de 0 a 10 kHz, y el intervalo de entrada al canal B de 0 a 6 kHz.

5-11. Para el receptor SSB de la fig. 5-21, con RF de entrada de 35.602 MHz, RF de oscilador local de 25 MHz y frecuencia de señal moduladora de 2 kHz, determine las frecuencias FI y BFO.

5-12. Para el receptor SSB de portadora piloto y varios canales de la fig. 5-24, con una frecuencia de oscilador de cristal de 300 kHz, una primera FI de 3.3 MHz, una RF de entrada de 23.303 MHz, frecuencia de señal moduladora de 3 kHz, determine lo siguiente: frecuencia de salida de VCO, factor de multiplicación y segunda FI.

5-13. Una señal de prueba tiene 2 y 3 kHz; la frecuencia de la portadora es 200 kHz.
(a) Determine el espectro de frecuencias de salida.
(b) Para $E_1 = E_2 = 12\ V_p$, y resistencia de carga $R_L = 50\ \Omega$, calcule la PEP y la potencia promedio.

5-14. Para el modulador de anillo balanceado de la fig. 5-6a, con una frecuencia de portadora de entrada $f_c = 500$ kHz, y frecuencia de señal moduladora de entrada $f_m = 0$ a 5 kHz, determine:
(a) El intervalo de frecuencias de salida.
(b) La frecuencia de salida para una sola frecuencia de entrada $f_m = 3.4$ kHz.

5-15. Para el circuito integrado lineal de modulador balanceado de la fig. 5-12, con una frecuencia de portadora de entrada f_c = 300 kHz y una frecuencia de señal moduladora de entrada f_m = 0 a 6 kHz, determine:

(a) El intervalo de frecuencias de salida.

(b) La frecuencia de salida para una entrada de frecuencia única f_m = 4.5 kHz.

5-16. Para el transmisor SSB de la fig. 5-13, con frecuencia baja de portadora f_{LF} = 120 kHz, media de portadora f_{MF} = 3 MHz, alta de portadora f_{HF} = 28 MHz y espectro de frecuencias de audio en la entrada f_m = 0 a 5 kHz.

(a) Trace los espectros de frecuencia en los siguientes puntos: salida del BPF 1, salida del BPF 2 y salida del BPF 3.

(b) Para una frecuencia única de entrada f_m = 2.5 kHz, determine la frecuencia trasladada en los siguientes puntos: salida del BPF 1, salida del BPF 2 y salida del BPF 3.

5-17. Repita el problema 5-16, pero cambie la frecuencia baja de portadora a 500 kHz. ¿Cuál transmisor tiene los requisitos más estrictos de filtrado?

5-18. Para el transmisor SSB de la fig. 5-14a, con una frecuencia de audio en la entrada f_m = 0 a 4 kHz y una frecuencia alta de portadora f_{HF} = 27 MHz.

(a) Trace el espectro de frecuencias de salida.

(b) Determine la frecuencia de salida cuando la frecuencia de entrada es única f_m = 1.8 kHz.

5-19. Repita el problema 5-18, pero cambie el espectro de frecuencias de audio en la entrada f_m de 300 a 4000 Hz.

5-20. Para el transmisor SSB de la fig. 5-18, con una frecuencia de portadora f_c = 400 kHz y un espectro de frecuencias de entrada f_m = 0 a 5 kHz.

(a) Trace el espectro de frecuencias en la salida del sumador lineal.

(b) Para una sola frecuencia de audio de entrada f_m = 2.5 kHz, determine la frecuencia de salida.

5-21. Repita el problema 5-20, pero cambie la frecuencia de la portadora de entrada a 600 kHz y el espectro de frecuencias de entrada de 300 a 6000 Hz.

5-22. Para el transmisor ISB de la fig. 5-20, la frecuencia de entrada en el canal A es f_a = 0 a 5 kHz, la frecuencia de entrada en el canal B es f_b = 0 a 5 kHz, la frecuencia de portadora baja es f_{LF} = 180 kHz, media de portadora f_{MF} = 3 MHz y alta de portadora f_{HF} = 30 MHz.

(a) Trace el espectro de frecuencias en los siguientes puntos: salida de BPF A, salida de BPF B, salida de BPF 3 y salida de BPF 4.

(b) Para una frecuencia de entrada del canal A f_a = 2.5 kHz y de entrada del canal B f_b = 2 kHz, determine los componentes de frecuencia en los siguientes puntos: salida de BPF A, salida de BPF B, salida de BPF 3 y salida de BPF 4.

5-23. Repita el problema 5-22, pero cambie el espectro de frecuencias de entrada en el canal A de 0 a 8 kHz, y de entrada en canal B de 0 a 6 kHz.

5-24. Para el receptor SSB de la fig. 5-24, la RF de entrada f_{RF} = 36.803 MHz, RF del oscilador local f_{lo} = 26 MHz y frecuencia de señal moduladora de 3 kHz, determine lo siguiente: frecuencia de salida de BFO y frecuencia de información detectada.

5-25. Para el receptor de SSB con portadora piloto y varios canales de la fig. 5-24, con frecuencia de oscilador de cristal f_{co} = 400 kHz, primera FI f_{FI} = 4.4 MHz, RF de entrada f_{RF} = 23.403 MHz y frecuencia de señal moduladora f_m = 3 kHz, determine lo siguiente: frecuencia de salida de VCO, factor de multiplicación y segunda FI.

5-26. Para una señal de prueba de dos tonos, de 3 y 4 kHz, y una frecuencia de portadora de 400 kHz, determine:

(a) El espectro de frecuencias de salida.

(b) Para E_1 y E_2 = 20 V_p, y resistor de carga R_L = 100 Ω, determine la PEP y la potencia promedio.

CAPÍT

Imagen protegida por derechos de autor

Transmisión por modulación angular

INTRODUCCIÓN

Como se dijo antes, hay tres propiedades de una señal analógica que se pueden variar, o modular, mediante la señal de información. Estas propiedades son la amplitud, la frecuencia y la fase. En los capítulos 3, 4 y 5 se describió la modulación de amplitud. Este capítulo y el capítulo 7 describen la *modulación de frecuencia* (FM) y la *modulación de fase* (PM), dos formas de *modulación angular* o *modulación de ángulo*. Desafortunadamente, con frecuencia se llama a ambas FM, aunque hay diferencias reales entre las dos. La modulación angular tiene varias ventajas sobre la modulación de amplitud, como la reducción de ruido, mejor fidelidad del sistema y uso más eficiente de la potencia. Sin embargo, también la modulación angular tiene varias desventajas en comparación con la AM, que incluyen la necesidad de mayor ancho de banda y el uso de circuitos más complicados tanto en los transmisores como en los receptores.

La modulación angular se introdujo por primera vez en 1931, como alternativa a la modulación de amplitud. Se sugería que una onda con modulación angular es menos susceptible al ruido que una onda de AM y, en consecuencia, podría mejorar el desempeño de las radiocomunicaciones. El mayor E. H. Armstrong (quien también desarrolló el receptor superheterodino) desarrolló el primer sistema de radio de FM funcional en 1936, y en julio de 1939 comenzaron las primeras emisiones con programa regular, en Alpine, New Jersey. Hoy, la modulación angular se usa en forma extensa para radioemisión comercial, televisión y transmisión de sonido, radioteléfonos, radios celulares y sistemas de comunicaciones por microondas y satelitales.

Los objetivos de este capítulo son presentar al lector los conceptos básicos de modulación de frecuencia y fase, y la forma en que se relacionan entre sí, mostrar algunos de los circuitos frecuentes para producir ondas con modulación angular y comparar el funcionamiento de la modulación angular con la modulación de amplitud.

MODULACIÓN ANGULAR

La *modulación angular* se produce siempre que se varía el ángulo de fase, θ, de una onda senoidal, con respecto al tiempo. Una onda con modulación angular se describe matemáticamente como sigue

$$m(t) = V_c \cos[\omega_c t + \theta(t)] \tag{6-1}$$

en donde $m(t)$ = onda con modulación angular
V_c = amplitud máxima de portadora (volts)
ω_c = frecuencia de la portadora en radianes, es decir, velocidad angular, $2\pi f_c$, en radianes por segundo
$\theta(t)$ = desviación instantánea de fase (radianes)

Con la modulación angular es necesario que $\theta(t)$ sea una función predeterminada de la señal moduladora. Por consiguiente, si $v_m(t)$ es la señal moduladora, la modulación angular se expresa como sigue

$$\theta(t) = F[v_m(t)] \tag{6-2}$$

en donde $v_m(t) = V_m \operatorname{sen}(\omega_m t)$
ω_m = velocidad angular de la señal moduladora ($2\pi f_m$ radianes por segundo)
f_m = frecuencia de la señal moduladora (hertz)
V_m = amplitud máxima de la señal moduladora (volts)

La diferencia entre modulación de frecuencia y modulación de fase es, en esencia, cuál propiedad de la portadora se hace variar en forma directa con la señal moduladora: la frecuencia o la fase, y cuál se hace variar en forma indirecta. Siempre que se varía la frecuencia de una portadora también varía la fase, y viceversa. Por consiguiente, la FM y la PM deben estar presentes al mismo tiempo siempre que se hace cualquier forma de modulación angular. Si la frecuencia de la portadora se hace variar directamente de acuerdo con la señal moduladora se obtiene la FM. Si se varía la fase de la portadora en forma directa por la señal moduladora, resulta la PM. Por consiguiente, la FM directa es PM indirecta, y la PM directa es FM indirecta. Se pueden definir las modulaciones de frecuencia y fase como sigue

Modulación directa de frecuencia (FM): Variar la frecuencia de una portadora de amplitud constante en proporción directa a la amplitud de la señal moduladora, con una rapidez igual a la frecuencia de la señal moduladora.

Modulación directa de fase (PM): Variar la fase de una portadora de amplitud constante en proporción directa a la amplitud de la señal moduladora, con una rapidez igual a la frecuencia de la señal moduladora.

La fig. 6-1 muestra una señal con modulación angular, $[m(t)]$, en el dominio de la frecuencia. Se ve allí cómo cambia la frecuencia f_c de la portadora cuando actúa sobre ella una señal moduladora $[v_m(t)]$. La magnitud y la dirección del desplazamiento de frecuencia, Δf, es proporcional a la amplitud y la polaridad de la señal moduladora, V_m, y la rapidez con la que se suceden los cambios de frecuencia es igual a la frecuencia f_m de la señal moduladora. Para este ejemplo,

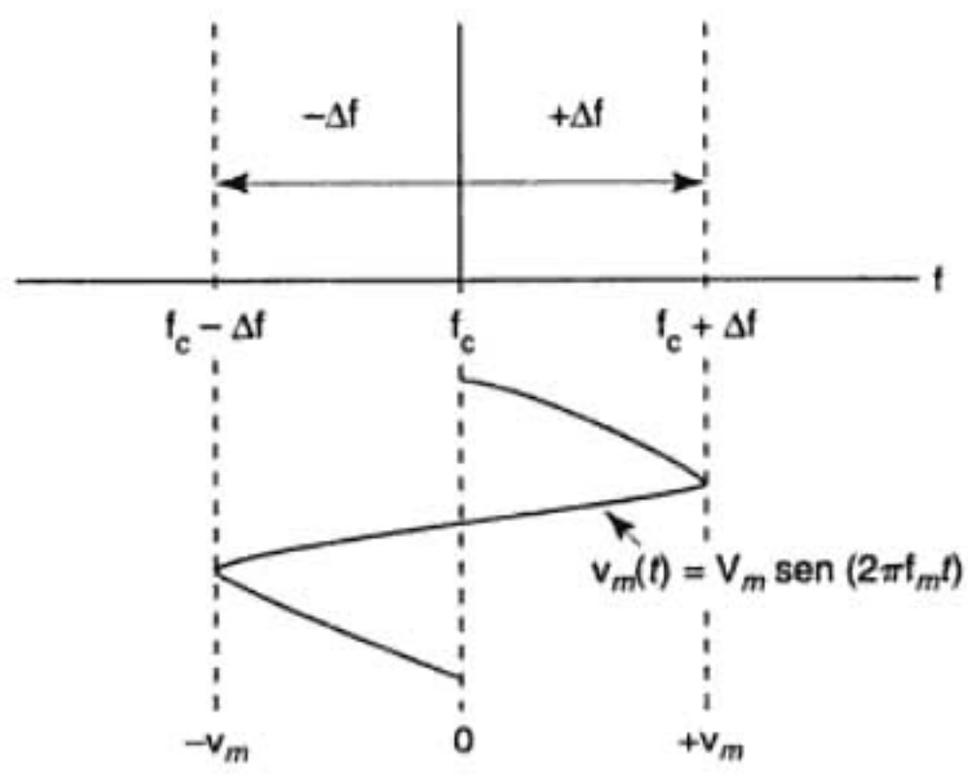


FIGURA 6-1 Onda con modulación angular en el dominio de la frecuencia

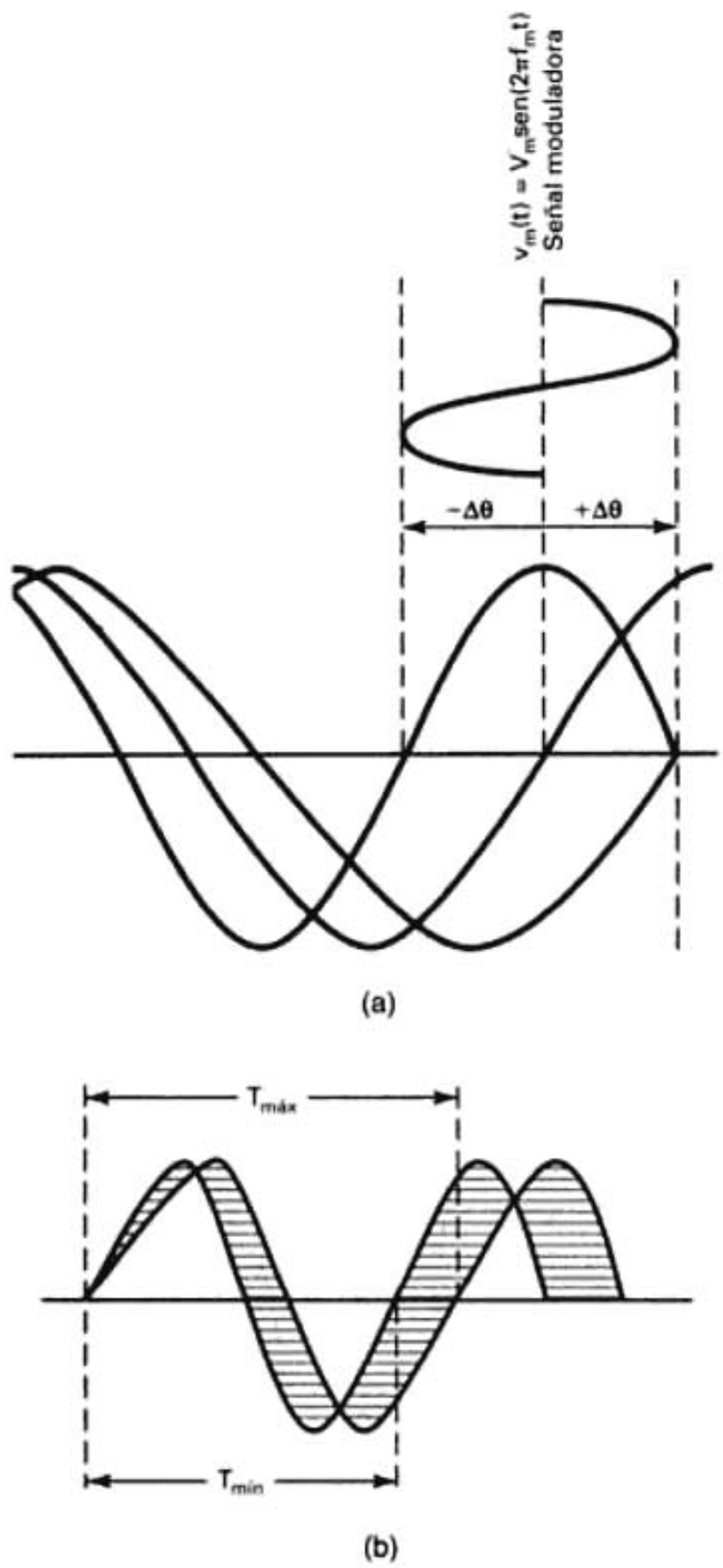


FIGURA 6-2 Modulación de ángulo en el dominio del tiempo: (a) cambio de fase a través del tiempo; (b) cambio de frecuencia a través del tiempo

una señal moduladora positiva produce un aumento de frecuencia, y una señal moduladora negativa produce una disminución de la frecuencia, aunque se podría tener la relación inversa, según el tipo de circuito modulador que se use.

La fig. 6-2a muestra la forma de onda, en el dominio de tiempo, de una portadora senoidal en la que hay modulación angular. Como se ve allí, la fase θ de la portadora cambia en forma proporcional a la amplitud, $V_m(t)$, de la señal moduladora. El desplazamiento angular relativo de la fase de la portadora, en radianes, con respecto a la fase de referencia, se llama *desviación de fase*, $\Delta\theta$. El cambio de fase de la portadora produce un cambio correspondiente de frecuencia. El desplazamiento relativo de la frecuencia de la portadora, en hertz, con respecto a su valor no modulado, se llama *desviación de frecuencia*, Δf. La magnitud de la desviación de frecuencia y fase es proporcional a la amplitud V_m de la señal moduladora, y la rapidez con que se efectúan los cambios es igual a la frecuencia f_m de la señal moduladora.

La fig. 6-2b muestra una portadora senoidal en la que la frecuencia moduladora f se cambia (*se desvía*) durante un periodo. La porción gruesa de la forma de onda corresponde al cambio

de periodo ΔT, pico a pico de la portadora. El periodo mínimo, $T_{mín}$, corresponde a la frecuencia máxima, $f_{máx}$, y el periodo máximo, $T_{máx}$, corresponde a la frecuencia mínima, $f_{mín}$. La desviación de frecuencia de pico a pico se determina en forma sencilla, midiendo la diferencia entre las frecuencias máxima y mínima: $\Delta f_{p\text{-}p} = 1/T_{mín} - 1/T_{máx}$.

Siempre que cambia el periodo T de una portadora senoidal, también cambian su frecuencia y su fase, y si los cambios son continuos ya no existe una sola frecuencia. Se demostrará que la forma de onda resultante de la modulación angular abarca la frecuencia de la portadora original no modulada (que con frecuencia se llama *frecuencia de portadora en reposo*) y una cantidad infinita de pares de frecuencias laterales, desplazadas a cada lado de la portadora, un múltiplo entero de la frecuencia de la señal moduladora.

Análisis matemático

La diferencia entre FM y PM se comprende con más facilidad si se definen los cuatro términos siguientes teniendo como referencia a la ecuación 6-1: desviación instantánea de fase, fase instantánea, desviación instantánea de frecuencia y frecuencia instantánea.

1. *Desviación instantánea de fase.* Es el cambio instantáneo de fase de la portadora, en determinado momento, e indica cuánto está cambiando la fase de la portadora con respecto a su fase de referencia. La desviación instantánea de fase se describe matemáticamente como sigue

$$\text{desviación instantánea de fase} = \theta(t) \quad \text{rad} \tag{6-3}$$

2. *Fase instantánea.* Es la fase precisa de la portadora en un momento dado, y se describe matemáticamente como sigue

$$\text{fase instantánea} = \omega_c t + \theta(t) \quad \text{rad} \tag{6-4}$$

donde $\omega_c t$ = fase de referencia de la portadora (radianes)
$= [2\pi(\text{rad/ciclo})]\,[f_c(\text{ciclos/s})][t\ (s)] = 2\pi f_c t(\text{rad})$
f_c = frecuencia de portadora (hertz)
$\theta(t)$ = desviación instantánea de fase (radianes)

3. *Desviación instantánea de frecuencia.* Es el cambio instantáneo en la frecuencia de la portadora, y se define como la primera derivada de la desviación instantánea de fase con respecto al tiempo. Por consiguiente, la desviación instantánea de fase es la primera integral de la desviación instantánea de frecuencia. En términos de la ecuación 6-3, la desviación instantánea de frecuencia se expresa como sigue

$$\text{desviación instantánea de frecuencia} = \theta'(t)\ \text{rad/s} \tag{6-5}$$

o bien

$$= \frac{\theta'(t)\ \text{rad/s}}{2\pi\ \text{rad/ciclo}} = \frac{\text{ciclos}}{\text{s}} = \text{Hz}$$

Se usa la prima (′) para indicar la primera derivada con respecto al tiempo.

4. *Frecuencia instantánea.* Es la frecuencia precisa de la portadora en determinado momento, y se define como la primera derivada de la fase instantánea respecto al tiempo. En términos de la ecuación 6-4, se expresa como sigue

$$\text{frecuencia instantánea} = \omega_i(t) = \frac{d}{dt}[\omega_c t + \theta(t)] \tag{6-6a}$$

$$= \omega_c + \theta'(t) \quad \text{rad/s} \tag{6-6b}$$

Al sustituir ω_c por $2\pi f_c$ se obtienen

frecuencia instantánea $= f_i(t)$

y $$\omega_i(t) = \left(2\pi \frac{\text{rad}}{\text{ciclo}}\right)\left(f_c \frac{\text{ciclos}}{\text{s}}\right) + \theta'(t) = 2\pi f_c + \theta'(t) \quad \text{rad/s}$$

o sea $$fi(t) = \frac{2\pi f_c + \theta(t)\ \text{rad/s}}{2\pi\ \text{rad/ciclo}} = f_c + \frac{\theta'(t)\ \text{ciclos}}{2\pi \quad \text{s}} = f_c + \frac{\theta'(t)}{2\pi} \quad \text{Hz} \tag{6-6c}$$

Sensibilidad a la desviación

Se puede definir la modulación de fase como modulación angular en la que la desviación instantánea de fase, $\theta(t)$, es proporcional a la amplitud del voltaje de la señal moduladora, y la desviación instantánea de frecuencia es proporcional a la pendiente, es decir, a la primera derivada, de la señal moduladora. En forma parecida, la modulación de frecuencia es la modulación angular en la que la desviación instantánea de frecuencia, $\theta'(t)$, es proporcional a la amplitud de la señal moduladora, y la desviación instantánea de fase es proporcional a la integral del voltaje de la señal moduladora.

Para una señal moduladora $v_m(t)$, la modulación de fase y la de frecuencia son

$$\text{modulación de fase} = \theta(t) = Kv_m(t) \quad \text{rad} \tag{6-7}$$

$$\text{modulación de frecuencia} = \theta'(t) = K_1 v_m(t) \quad \text{rad/s} \tag{6-8}$$

siendo K y K_1 constantes, y son las *sensibilidades a la desviación* de los moduladores de fase y de frecuencia, respectivamente. Las sensibilidades a la desviación son las funciones de transferencia de salida en función de la entrada de los moduladores, que producen la relación entre qué parámetro de salida cambia con respecto a los cambios especificados de la señal de entrada. Para un modulador de frecuencia, los cambios serían en la señal de salida con respecto a cambios en la amplitud del voltaje de entrada. Para un modulador de fase, los cambios serían en la fase de la frecuencia de salida con respecto a los cambios de amplitud del voltaje de entrada.

La sensibilidad a la desviación de un modulador de fase es

$$K = \frac{\text{rad}}{\text{V}}\left(\frac{\Delta\theta}{\Delta\text{V}}\right)$$

y para un modulador de frecuencia es

$$K_1 = \frac{\text{rad/s}}{\text{V}} \text{ o } \frac{\text{rad}}{\text{V} - \text{s}}\left(\frac{\Delta\omega}{\Delta\text{V}}\right)$$

La modulación de fase es la primera integral de la modulación de frecuencia. Por consiguiente, de acuerdo con las ecuaciones 6-7 y 6-8,

$$\begin{aligned} \text{modulación de fase} = \theta(t) &= \int \theta'(t)\, dt \\ &= \int K_1 v_m(t)\, dt \\ &= K_1 \int v_m(t)\, dt \end{aligned} \tag{6-9}$$

Por consiguiente, sustituyendo una señal moduladora $v_m(t) = V_m \cos(\omega_m t)$ en la ecuación 6-1 se obtiene

Para modulación de fase,
$$\begin{aligned} m(t) &= V_c \cos[\omega_c t + \theta(t)] \\ &= V_c \cos[\omega_c t + KV_m \cos(\omega_m t)] \end{aligned} \tag{6-10}$$

Para modulación de frecuencia,
$$\begin{aligned} m(t) &= V_c \cos[\omega_c t + \int \theta'(t)] \\ &= V_c \cos[\omega_c t + \int K_1 v_m(t)\, dt] \\ &= V_c \cos[\omega_c t + K_1 \int V_m \cos(\omega_m t)\, dt] \\ &= V_c \cos\left[\omega_c t + \frac{K_1 V_m}{\omega_m} \operatorname{sen}(\omega_m t)\right] \end{aligned} \tag{6-11}$$

Las ecuaciones anteriores se resumen en la tabla 6-1. También se presentan las ecuaciones de las ondas de FM y PM que resultan cuando la señal moduladora es una senoidal de frecuencia única.

TABLA 6-1 Ecuaciones para portadoras de fase y frecuencia modulada

Tipo de modulación	Señal moduladora	Onda con modulación angular, $m(t)$
(a) Fase	$v_m(t)$	$V_c \cos[\omega_c t + Kv_m(t)]$
(b) Frecuencia	$v_m(t)$	$V_c \cos[\omega_c t + K_1 \int v_m(t)\, dt]$
(c) Fase	$V_m \cos(\omega_m t)$	$V_c \cos[\omega_c t + KV_m \cos(\omega_m t)]$
(d) Frecuencia	$V_m \cos(\omega_m t)$	$V_c \cos\left[\omega_c t + \frac{K_1 V_m}{\omega_m} \operatorname{sen}(\omega_m t)\right]$

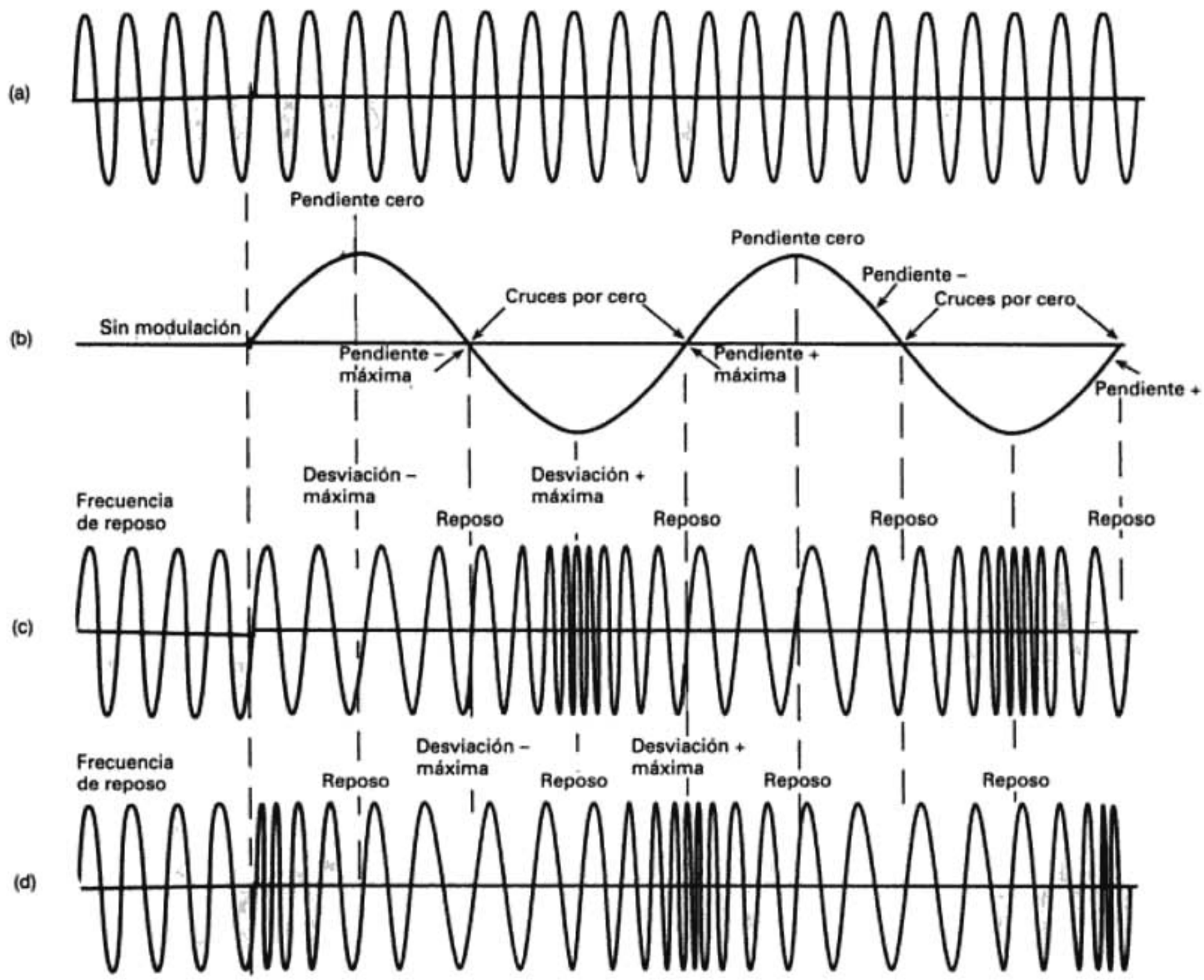


FIGURA 6-3 Modulación de fase y de frecuencia de una onda portadora senoidal por una señal de onda senoidal: (a) portadora no modulada; (b) señal moduladora; (c) onda con frecuencia modulada; (d) onda con fase modulada

Formas de onda de FM y PM

La fig. 6-3 ilustra la modulación de frecuencia y de fase de una portadora senoidal, mediante una señal moduladora de frecuencia única. Se ve que las formas de onda de FM y de PM son idénticas, a excepción de su relación temporal (fase). Así, es imposible distinguir una forma de onda de FM de una de PM, sin conocer las características dinámicas de la señal moduladora. Con FM, la desviación máxima de frecuencia (cambio de frecuencia de la portadora) se efectúa durante

los picos máximos positivos y negativos de la señal moduladora; es decir, la desviación de frecuencia es proporcional a la amplitud de la señal moduladora. Con la PM, la desviación máxima de frecuencia se efectúa durante los cruces de la señal moduladora por cero; es decir, la desviación de frecuencia es proporcional a la pendiente de la primera derivada de la señal moduladora. Tanto para la modulación de fase como la de frecuencia, la rapidez con que cambia la frecuencia es igual a la frecuencia de la señal moduladora.

De igual modo, no se puede saber si la ecuación 6-1 representa una onda de FM o de PM. Podría ser cualquiera de las dos. Sin embargo, el conocimiento de la señal moduladora permite hacer la identificación correcta. Si $\theta(t) = Kv_m(t)$, es modulación de fase, y si $\theta'(t) = K_1 v_m(t)$, es modulación de frecuencia. En otras palabras, si la frecuencia instantánea es directamente proporcional a la amplitud de la señal moduladora, es modulación de frecuencia, y si la fase instantánea es directamente proporcional a la amplitud de la frecuencia moduladora, es modulación de fase.

Desviación de fase e índice de modulación

Al comparar las ecuaciones (c) y (d) de la portadora con modulación angular, en la tabla 6-1, se ve que la ecuación de una portadora que se va a modular en fase o en frecuencia mediante una señal moduladora de frecuencia única, se puede escribir en forma general modificando como sigue la ecuación 6-1

$$m(t) = V_c \cos[\omega_c t + m \cos(\omega_m t)] \tag{6-12}$$

en la que $m \cos(w_m t)$ es la desviación instantánea de fase, $\theta(t)$. Cuando la señal moduladora es una senoidal de frecuencia única, se ve, en la ecuación 6-12, que el ángulo de fase de la portadora varía en forma senoidal sencilla respecto a su valor no modulado.

En la ecuación 6-12, m representa la *desviación máxima de fase*, en radianes, de una portadora con fase modulada. Esta desviación máxima de fase se llama *índice de modulación*. Una diferencia primaria entre modulación de frecuencia y de fase es la forma en que se define el índice de modulación. Para la PM, el índice de modulación es proporcional a la amplitud de la señal moduladora e independiente de su frecuencia. El índice de modulación de una portadora con fase modulada se define como sigue

$$m = KV_m \quad \text{(radianes)} \tag{6-13}$$

en donde m = índice de modulación y desviación máxima de fase ($\Delta\theta$, radianes)
K = sensibilidad a la desviación (radianes por volt)
V_m = amplitud máxima de la señal moduladora (volts)

y entonces, $m = K\left(\dfrac{\text{radianes}}{\text{volt}}\right)V_m \text{ (volts)} = \text{radianes}$

Por consiguiente, para PM, se puede escribir como sigue la ecuación 6-1

$$m(t) = V_c \cos[\omega_c t + KV_m \cos(\omega_m t)] \tag{6-14a}$$

o bien

$$m(t) = V_c \cos[\omega_c t + \Delta\theta \cos(\omega_m t)] \tag{6-14b}$$

o también

$$m(t) = V_c \cos[\omega_c t + m \cos(\omega_m t)] \tag{6-14c}$$

Para una portadora con frecuencia modulada, el índice de modulación es directamente proporcional a la amplitud de la señal moduladora, e inversamente proporcional a la frecuencia de la señal moduladora. Así, el índice de modulación para FM es

$$m = \frac{K_1 V_m}{\omega_m} \text{ (adimensional)} \tag{6-15}$$

donde m = índice de modulación (adimensional)
K_1 = sensibilidad a la desviación (radianes por segundo por volt, o radianes por volt)

V_m = amplitud máxima de la señal moduladora (volts)
ω_m = frecuencia en radianes (radianes por segundo)

por consiguiente, $$m = \frac{K_1\left(\frac{\text{radianes}}{\text{volt - s}}\right)V_m(\text{volt})}{\omega_m\,(\text{radianes/s})} = (\text{adimensional})$$

Se puede ver que, según la ecuación 6-15, el índice de modulación es una relación adimensional para la FM, y sólo se usa para describir la profundidad de modulación lograda en una señal moduladora con determinada amplitud máxima y frecuencia en radianes.

La sensibilidad a la desviación se puede expresar en hertz por volt, si se escribe la ecuación 6-15 en una forma más práctica, como sigue

$$m = \frac{K_1 V_m}{f_m}\ (\text{adimensional}) \tag{6-16}$$

en donde m = índice de modulación (adimensional)
K_1 = sensibilidad a la desviación (ciclos por segundo por volt o hertz por volt)
V_m = amplitud máxima de la señal moduladora (volts)
f_m = frecuencia de la señal moduladora (hertz)

por lo anterior, $$m = \frac{K_1\left(\frac{\text{hertz}}{\text{volt}}\right)V_m\,(\text{volt})}{f_m\,(\text{hertz})} = (\text{adimensional})$$

Desviación de frecuencia

La *desviación de frecuencia* es el cambio de frecuencia que sucede en la portadora cuando sobre ella actúa la frecuencia de la señal moduladora. Se expresa normalmente como un desplazamiento máximo de frecuencia, Δf, en hertz. La desviación de frecuencia pico a pico, $2\,\Delta f$, se llama a veces *variación de frecuencia*.

Para una FM, la sensibilidad a la desviación se suele expresar en hertz por volt. Por consiguiente, la desviación máxima de frecuencia no es más que el producto de la sensibilidad a la desviación y el voltaje máximo de la señal moduladora, y se expresa como sigue

$$\Delta f = K_1 V_m\ (\text{Hz}) \tag{6-17}$$

Se puede sustituir la ecuación 6-17 en la 6-16, para expresar como sigue el índice de modulación en FM

$$m = \frac{\Delta f\,(\text{Hz})}{f_m\,(\text{Hz})}\ (\text{adimensional}) \tag{6-18}$$

Por consiguiente, para la FM, la ecuación 6-1 se puede escribir como sigue

$$m(t) = V_c\cos\left[\omega_c t + \frac{K_1 V_m}{f_m}\,\text{sen}(\omega_m t)\right] \tag{6-19a}$$

o bien $$m(t) = V_c\cos\left[\omega_c t + \frac{\Delta f}{f_m}\,\text{sen}(\omega_m t)\right] \tag{6-19b}$$

o también $$m(t) = V_c\cos\left[\omega_c t + m\,\text{sen}(\omega_m t)\right] \tag{6-19c}$$

Se puede ver, al examinar las ecuaciones 6-15 y 6-16, que los índices de modulación para FM y PM se relacionan en forma distinta con la señal moduladora. En la PM, tanto el índice de modulación como la desviación máxima de fase son directamente proporcionales a la amplitud de la señal moduladora, y no son afectados por su frecuencia. Sin embargo, en la FM el índice de modulación y la desviación máxima de fase son directamente proporcionales a la

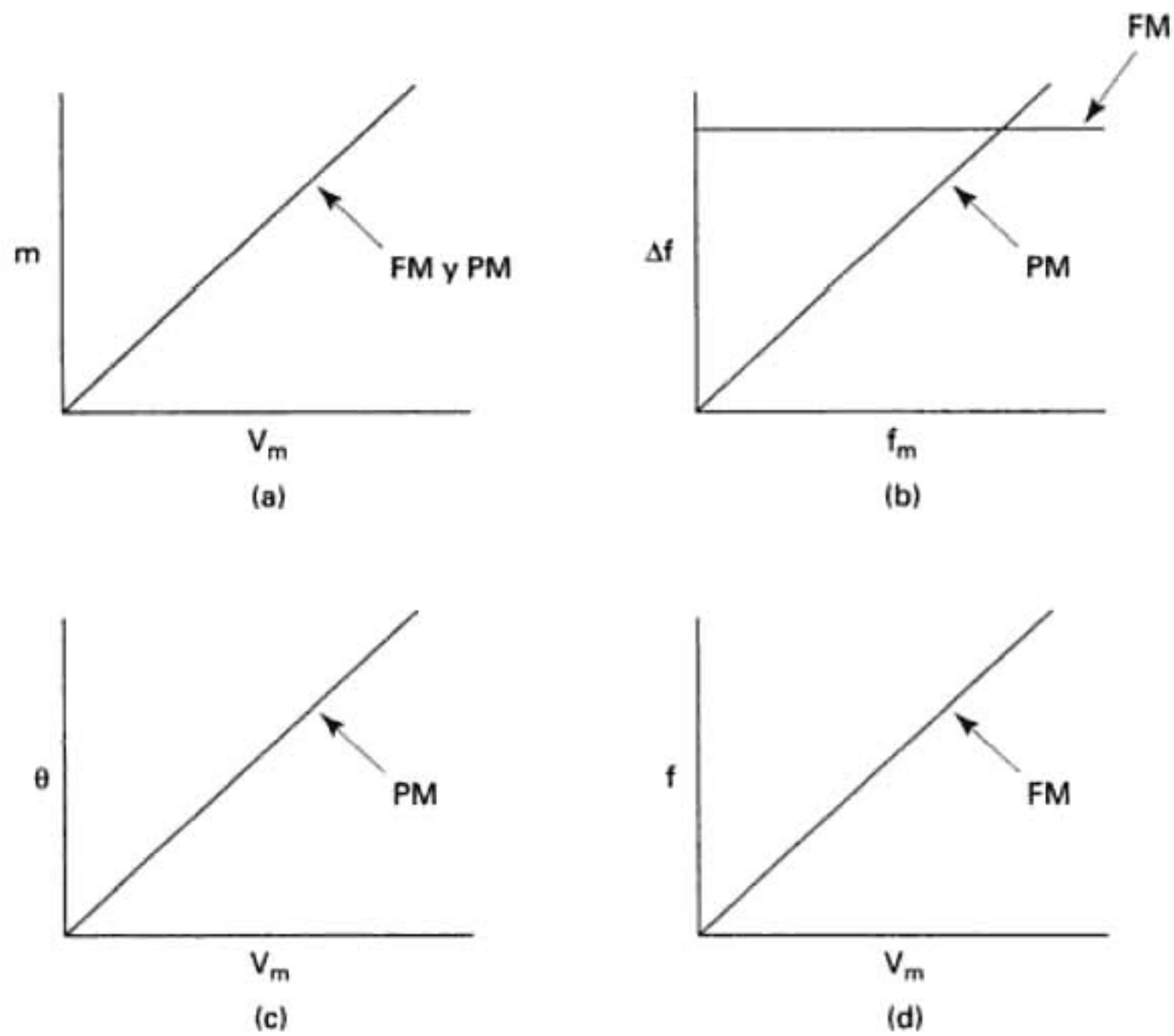


FIGURA 6-4 Relación entre índice de modulación, desviación de frecuencia y desviación de fase, con respecto a la amplitud y la frecuencia de la señal moduladora: (a) índice de modulación en función de amplitud; (b) desviación de frecuencia en función de la frecuencia de modulación; (c) desviación de fase en función de la amplitud; (d) desviación de frecuencia en función de la amplitud

amplitud de la señal moduladora, y el índice de modulación es inversamente proporcional a su frecuencia. La fig. 6-4 muestra en forma gráfica la relación entre el índice de modulación y la desviación máxima de fase para PM y para FM con respecto a la amplitud y a la frecuencia de la señal moduladora.

Las relaciones matemáticas anteriores se resumen en la tabla 6-2.

Ejemplo 6-1

(a) Determinar Δf, la desviación máxima de frecuencia, y m, el índice de modulación, para un modulador de FM con sensibilidad a la desviación $K_1 = 5$ kHz/V y una señal moduladora $v_m(t) = 2\cos(2\pi 2000t)$.

(b) Determinar la desviación máxima de fase, m, para un modulador PM con sensibilidad a la desviación $K = 2.5$ rad/V, y una señal moduladora $v_m(t) = 2\cos(2\pi 2000t)$.

Solución (a) La desviación máxima de frecuencia no es más que el producto de la sensibilidad a la desviación por la amplitud máxima de la señal moduladora, es decir

$$\Delta f = \frac{5\text{ kHz}}{\text{V}} \times 2\text{ V} = 10\text{ kHz}$$

El índice de modulación se calcula sustituyendo en la ecuación 6-18

$$m = \frac{10\text{ kHz}}{2\text{ kHz}} = 5$$

(b) El desplazamiento máximo de fase para una onda de fase modulada, es el índice de modulación, y se calcula sustituyendo en la ecuación 6-13

$$m = \frac{2.5\text{ rad}}{\text{V}} \times 2\text{ V} = 5\text{ rad}$$

TABLA 6-2 Resumen de modulación angular

	FM	PM
Onda modulada	$m(t) = V_c \cos\left[\omega_c t + \frac{K_1 V_m}{f_m} \text{sen}(\omega_m t)\right]$	$m(t) = V_c \cos[\omega_c t + KV_m \cos(\omega_m t)]$
o	$m(t) = V_c \cos[\omega_c t + m \,\text{sen}(\omega_m t)]$	$m(t) = V_c \cos[\omega_c t + m \cos(\omega_m t)]$
o	$m(t) = V_c \cos\left[\omega_c t + \frac{\Delta f}{f_m} \text{sen}(\omega_m t)\right]$	$m(t) = V_c \cos[\omega_c t + \Delta\theta \cos(\omega_m t)]$
Sensibilidad a la desviación	K_1 (Hz/V)	K (rad/V)
Desviación	$\Delta f = K_1 V_m$ (Hz)	$\Delta\theta = KV_m$ (rad)
Índice de modulación	$m = \frac{K_1 V_m}{f_m}$ (adimensional)	$m = KV_m$ (rad)
o	$m = \frac{\Delta f}{f_m}$ (adimensional)	$m = \Delta\theta$ (rad)
Señal moduladora	$v_m(t) = V_m \,\text{sen}(\omega_m t)$	$v_m(t) = V_m \cos(\omega_m t)$
Frecuencia moduladora	$\omega_m = 2\pi f_m$ rad/s	$\omega_m = 2\pi f_m$ rad/s
o	$\omega_m/2\pi = f_m$ (Hz)	$\omega_m/2\pi = f_m$ (Hz)
Señal portadora	$V_c \cos(\omega_c t)$	$V_c \cos(\omega_c t)$
Frecuencia de portadora	$\omega_c = 2\pi f_c$ (rad/s)	$\omega_c = 2\pi f_c$ (rad/s)
o	$\omega_c/2\pi = f_c$ (Hz)	$\omega_c/2\pi = f_c$ (Hz)

En el ejemplo 6-1, el índice de modulación para la portadora de frecuencia modulada fue igual al índice de modulación de la portadora con fase modulada (5). Si cambia la amplitud de la señal moduladora, cambiará en forma proporcional el índice de modulación, tanto para la onda de frecuencia modulada como la de fase modulada. Si embargo, si cambia la frecuencia de la señal moduladora, el índice de modulación en la onda de frecuencia modulada cambiará en forma inversamente proporcional, mientras que su índice de modulación no se altera. Por lo anterior y bajo condiciones idénticas, la FM y la PM son indistinguibles con una señal moduladora de una sola frecuencia; sin embargo, cuando cambia la frecuencia de la señal moduladora, el índice de modulación PM permanece constante, mientras que el índice de modulación de FM aumenta cuando disminuye la frecuencia de la señal moduladora, y viceversa.

Porcentaje de modulación. El porcentaje de modulación, o modulación porcentual, para una onda con modulación angular, se calcula en forma distinta que en la onda de amplitud modulada. Con la modulación angular, el porcentaje de modulación es tan sólo la relación de desviación de frecuencia producida realmente, entre la desviación máxima de frecuencia permitida, expresada en forma porcentual. La ecuación correspondiente es

$$\% \text{ de modulación} = \frac{\Delta f_{(\text{real})}}{\Delta f_{(\text{máx})}} \times 100 \tag{6-20}$$

Por ejemplo, en Estados Unidos la Comisión Federal de Comunicaciones (FCC) limita la desviación de frecuencia, para los transmisores de banda comercial de FM, a ±75 kHz. Si una señal moduladora produce una desviación de frecuencia de ±50 kHz, el porcentaje de modulación es

$$\% \text{ de modulación} = \frac{50 \text{ kHz}}{75 \text{ kHz}} \times 100 = 67\%$$

Moduladores y demoduladores de fase y de frecuencia

Un *modulador de fase* es un circuito en el que se hace variar la portadora de tal modo que su fase instantánea es proporcional a la señal moduladora. La portadora no modulada es una senoidal

de una sola frecuencia, que se suele llamar frecuencia *en reposo*. Un *modulador de frecuencia* (que va veces se llama *desviador de frecuencia*) es un circuito en el que se varía la portadora de tal manera que su fase instantánea es proporcional a la integral de la señal moduladora. Por consiguiente, con un modulador de frecuencia, si la señal moduladora $v(t)$ se deriva antes de aplicarla al modulador, la desviación instantánea de fase es proporcional a la integral de $v(t)$ o, en otras palabras, proporcional a $v(t)$, porque $\int v'(t) = v(t)$. Dicho de otra forma, un modulador de FM que está precedido por un diferenciador produce una onda de salida en la que la desviación de fase es proporcional a la señal moduladora y equivale, en consecuencia, a un modulador de fase. Son posibles algunas otras equivalencias interesantes. Por ejemplo, un demodulador de frecuencia seguido de un integrador equivale a un demodulador de fase. Las cuatro equivalencias de uso común son:

1. Modulador de PM = diferenciador seguido por un modulador de FM.
2. Demodulador de PM = demodulador de FM seguido por un integrador.
3. Modulador de FM = integrador seguido por un modulador de PM.
4. Demodulador de FM = demodulador de PM seguido de un diferenciador.

Análisis en frecuencia de las ondas con modulación angular

Con la modulación angular, los componentes de frecuencia de la onda modulada tienen una relación mucho más compleja con los componentes de frecuencia de la señal moduladora que en el caso de la modulación de amplitud. En un modulador de frecuencia o de fase, una señal moduladora de una sola frecuencia produce una cantidad infinita de pares de frecuencias laterales y, por lo mismo, tiene un ancho de banda infinito. Cada frecuencia lateral está desplazada respecto a la portadora por un múltiplo entero de la frecuencia de la señal moduladora. Sin embargo, en general la mayoría de las frecuencias laterales tienen una amplitud insignificante, y se pueden ignorar.

Modulación mediante una senoidal de frecuencia única. El análisis de frecuencia de una onda con modulación angular por una senoidal de una sola frecuencia produce una desviación máxima de fase de m radianes, donde m es el índice de modulación. De nuevo, según la ecuación 6-12, y para una frecuencia moduladora igual a ω_m, la onda $m(t)$ se puede escribir como sigue

$$m(t) = V_c \cos[\omega_c t + m \cos(\omega_m t)]$$

De acuerdo con la ecuación 6-12, no son obvios los componentes individuales de frecuencia que forman la onda modulada. Sin embargo, existen *identidades de funciones de Bessel* que se aplican en forma directa. Una de esas identidades es

$$\cos(\alpha + m \cos \beta) = \sum_{n=-\infty}^{\infty} J_n(m) \cos\left(\alpha + n\beta + \frac{n\pi}{2}\right) \tag{6-21}$$

La función $J_n(m)$ es la función de Bessel de primera clase de n-ésimo orden, con argumento m. Si se aplica la ecuación 6-21 a la ecuación 6-13, se podrá escribir $m(t)$ en la forma

$$m(t) = V_c \sum_{n=-\infty}^{\infty} J_n(m) \cos\left(\omega_c t + n\omega_m t + \frac{n\pi}{2}\right) \tag{6-22}$$

Esta ecuación se desarrolla en los cuatro primeros términos, como sigue

$$\begin{aligned} m(t) = V_c\Big\{ & J_0(m) \cos \omega_c t + J_1(m) \cos\left[(\omega_c + \omega_m)t + \frac{\pi}{2}\right] \\ & - J_1(m) \cos\left[(\omega_c - \omega_m)t - \frac{\pi}{2}\right] + J_2(m) \cos[(\omega + 2\omega_m)t)] \\ & + J_2(m) \cos[(\omega_c - 2\omega_m)t] + \ldots J_n(m) \ldots \end{aligned} \tag{6-23}$$

en la que $m(t)$ = onda con modulación angular
m = índice de modulación
V_c = amplitud máxima de la portadora no modulada
$J_0(m)$ = componente de la portadora
$J_1(m)$ = primer conjunto de frecuencias laterales, desplazadas ω_m respecto a la portadora
$J_2(m)$ = segundo conjunto de frecuencias laterales, desplazadas $2\omega_m$ respecto a la portadora
$J_n(m)$ = n-ésimo conjunto de frecuencias laterales, desplazadas $n\omega_m$ respecto a la portadora

Las ecuaciones 6-22 y 6-23 muestran que en la modulación angular, una señal moduladora de una sola frecuencia produce una cantidad infinita de conjuntos de frecuencias laterales, cada uno desplazado de la señal portadora en un múltiplo entero de la frecuencia de la señal moduladora. Un conjunto de banda lateral incluye una frecuencia lateral superior e inferior ($f_c \pm f_m, f_c \pm 2f_m, f_c \pm nf_m$, etcétera). Los conjuntos sucesivos de bandas laterales se llaman bandas laterales de primer orden, bandas latarales de segundo orden, y así sucesivamente; sus magnitudes están determinadas por los coeficientes $J_1(m)$, $J_2(m)$, etcétera, respectivamente.

Para calcular la amplitud J_n de las frecuencias laterales, se puede convertir la ecuación 6-23 en

$$J_n(m) = \left(\frac{m}{2}\right)^n \left[\frac{1}{n} - \frac{(m/2)^2}{1!(n+1)!} + \frac{(m/2)^4}{2!(n+2)!} - \frac{(m/2)^6}{3!(n+1)!} + \ldots\right] \qquad (6\text{-}24)$$

donde ! = factorial ($1 \times 2 \times 3 \times 4$, etc.)
n = J, o número de la frecuencia lateral
m = índice de modulación

La tabla 6-3 contiene las funciones de Bessel de primera clase para diversos valores del índice de modulación. Se verá que un índice de modulación de 0 (sin modulación) produce cero frecuencias laterales, y mientras mayor sea el índice de modulación, se producirán más conjuntos de frecuencias laterales. Los valores indicados de J_n son en relación con la amplitud de la portadora no modulada. Por ejemplo, $J_2 = 0.35$ indica que la amplitud del segundo conjunto de frecuencias laterales es igual a 35% de la amplitud de la portadora no modulada ($0.35\ V_c$). Se puede ver que la amplitud de las frecuencias laterales se vuelve insignificante con rapidez, a medida que el índice de modulación baja de la unidad. Para valores mayores de m, el valor de $J_n(m)$ comienza a decrecer con rapidez tan pronto como $n = m$. Al aumentar el índice de modulación respecto a cero, la magnitud de $J_0(m)$ de portadora decrece. Cuando m es igual a aproximadamente 2.4, entonces $J_0(m) = 0$ y el componente de la portadora es cero (esto se llama el *primer cero de portadora*). Esta propiedad se usa con frecuencia para determinar el índice de modulación, o establecer la sensibilidad a la desviación de un modulador de FM. La portadora vuelve a aparecer cuando m aumenta respecto a 2.4. Cuando m llega aproximadamente a 5.4, la componente de la portadora vuelve a desaparecer (esto se llama *segundo cero de portadora*). Los aumentos posteriores del índice de modulación producen más ceros de portadora, a intervalos periódicos.

La fig. 6-5 muestra las curvas de las amplitudes relativas de la portadora y de varios conjuntos de frecuencias laterales, para valores de m hasta 10. Se puede ver que las amplitudes de la portadora y de las frecuencias laterales varían con una rapidez periódica que se asemeja a la de una senoidal amortiguada. Los valores negativos para $J(m)$ indican simplemente la fase relativa de este conjunto de frecuencias laterales.

En la tabla 6-3 sólo se mencionan las frecuencias laterales significativas. Se considera que una frecuencia lateral es no significativa, a menos que su amplitud sea igual o mayor a 1% de la amplitud de la portadora no modulada ($J_n \geq 0.01$). Se ve en esa tabla que cuando aumenta m, aumenta la cantidad de frecuencias laterales importantes o significativas. Por consiguiente, el ancho de banda de una onda con modulación angular es función del índice de modulación.

Ejemplo 6-2

Para un modulador FM con índice de modulación $m = 1$, señal moduladora $v_m(t) = V_m \operatorname{sen}(2\pi 1000t)$, y portadora no modulada $v_c(t) = 10 \operatorname{sen}(2\pi 500kt)$, determinar:
(a) Cantidad de conjuntos de frecuencias laterales significativas.
(b) Sus amplitudes.

TABLA 6-3 Funciones de Bessel de primera clase, $J_n(m)$

Índice de modulación	Portadora	Pares laterales de frecuencia													
m	J_0	J_1	J_2	J_3	J_4	J_5	J_6	J_7	J_8	J_9	J_{10}	J_{11}	J_{12}	J_{13}	J_{14}
0.00	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—
0.25	0.98	0.12	—	—	—	—	—	—	—	—	—	—	—	—	—
0.5	0.94	0.24	0.03	—	—	—	—	—	—	—	—	—	—	—	—
1.0	0.77	0.44	0.11	0.02	—	—	—	—	—	—	—	—	—	—	—
1.5	0.51	0.56	0.23	0.06	0.01	—	—	—	—	—	—	—	—	—	—
2.0	0.22	0.58	0.35	0.13	0.03	—	—	—	—	—	—	—	—	—	—
2.4	0	0.52	0.43	0.20	0.06	0.02	—	—	—	—	—	—	—	—	—
2.5	−0.05	0.50	0.45	0.22	0.07	0.02	0.01	—	—	—	—	—	—	—	—
3.0	−0.26	0.34	0.49	0.31	0.13	0.04	0.01	—	—	—	—	—	—	—	—
4.0	−0.40	−0.07	0.36	0.43	0.28	0.13	0.05	0.02	—	—	—	—	—	—	—
5.0	−0.18	−0.33	0.05	0.36	0.39	0.26	0.13	0.05	0.02	—	—	—	—	—	—
5.45	0	−0.34	−0.12	0.26	0.40	0.32	0.19	0.09	0.03	0.01	—	—	—	—	—
6.0	0.15	−0.28	−0.24	0.11	0.36	0.36	0.25	0.13	0.06	0.02	—	—	—	—	—
7.0	0.30	0.00	−0.30	−0.17	0.16	0.35	0.34	0.23	0.13	0.06	0.02	—	—	—	—
8.0	0.17	0.23	−0.11	−0.29	−0.10	0.19	0.34	0.32	0.22	0.13	0.06	0.03	—	—	—
8.65	0	0.27	0.06	−0.24	−0.23	0.03	0.26	0.34	0.28	0.18	0.10	0.05	0.02	—	—
9.0	−0.09	0.25	0.14	−0.18	−0.27	−0.06	0.20	0.33	0.31	0.21	0.12	0.06	0.03	0.01	—
10.0	−0.25	0.05	0.25	0.06	−0.22	−0.23	−0.01	0.22	0.32	0.29	0.21	0.12	0.06	0.03	0.01

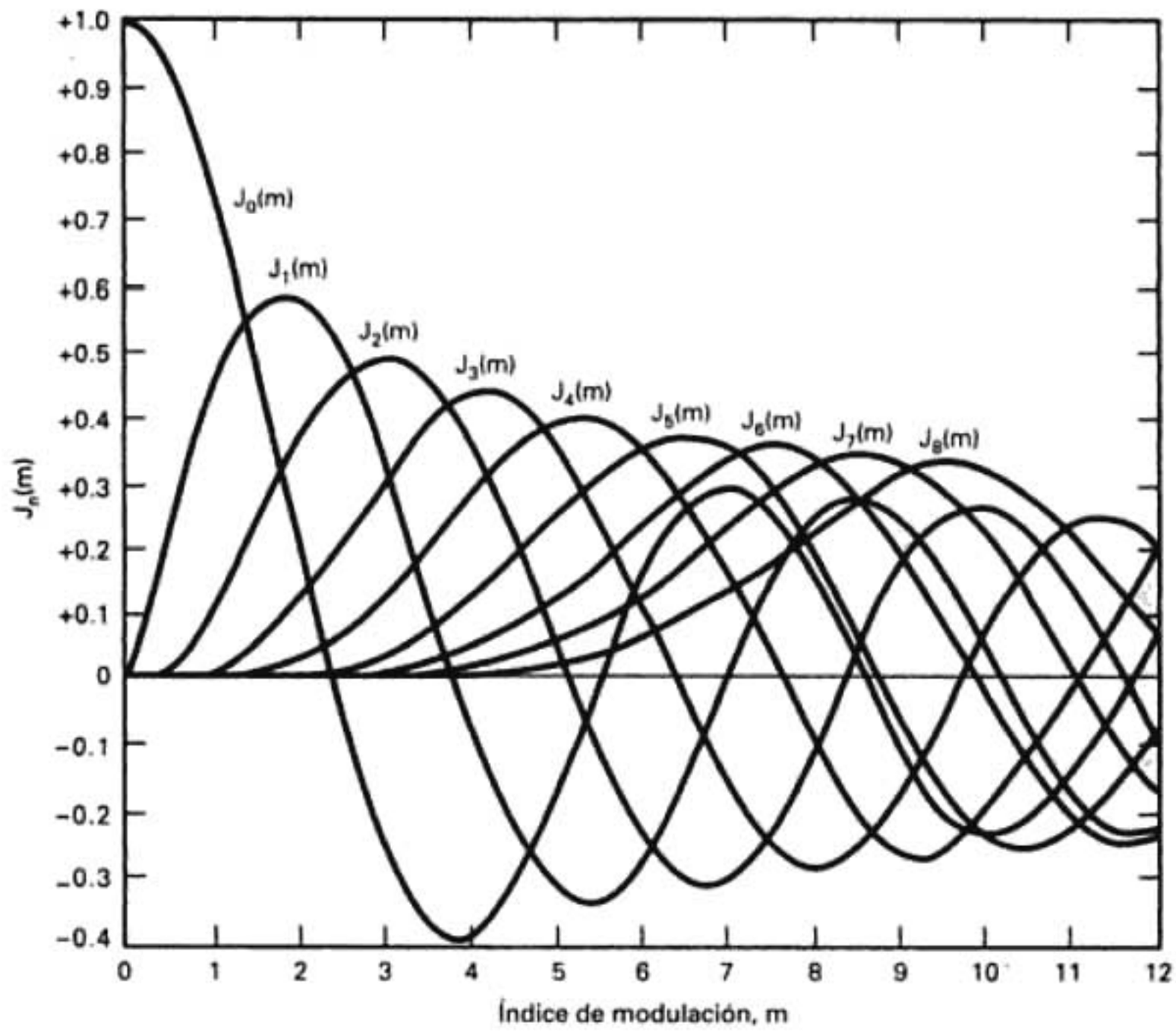


FIGURA 6-5 $J_n(m)$ en función de m

A continuación

(c) Trazar el espectro de frecuencias, mostrando sus amplitudes relativas.

Solución (a) De la tabla 6-3, un índice de modulación de 1 produce un componente reducido de portadora y tres conjuntos de frecuencias laterales significativas.

(b) Las amplitudes relativas de la portadora y de las frecuencias laterales son

$$J_0 = 0.77(10) = 7.7 \text{ V}$$
$$J_1 = 0.44(10) = 4.4 \text{ V}$$
$$J_2 = 0.11(10) = 1.1 \text{ V}$$
$$J_3 = 0.02(10) = 0.2 \text{ V}$$

(c) El espectro de frecuencias se ve en la fig. 6-6.

Si el modulador utilizado en el ejemplo 6.2 se reemplaza con un modulador PM, con las mismas frecuencias de portadora y moduladora, una desviación de fase máxima de 1 rad produce exactamente el mismo espectro de frecuencias.

Requisitos de ancho de banda para ondas con modulación angular

J. R. Carson demostró, en 1922, que para determinada frecuencia de señal moduladora, una onda de frecuencia modulada no puede caber en una banda más angosta que una onda de amplitud modulada. De acuerdo a la explicación anterior, y al ejemplo 6-2, se puede ver que el ancho de banda de una onda de modulación angular es una función de la frecuencia de la señal moduladora y del índice de modulación. En la modulación de ángulo se producen varios conjuntos de bandas laterales y, en consecuencia, el ancho de banda puede ser bastante mayor que el de una onda de amplitud modulada con la misma señal moduladora. La forma de onda de salida del modulador en el ejemplo 6-2 requiere 6 kHz de ancho de banda para pasar la portadora y todas las

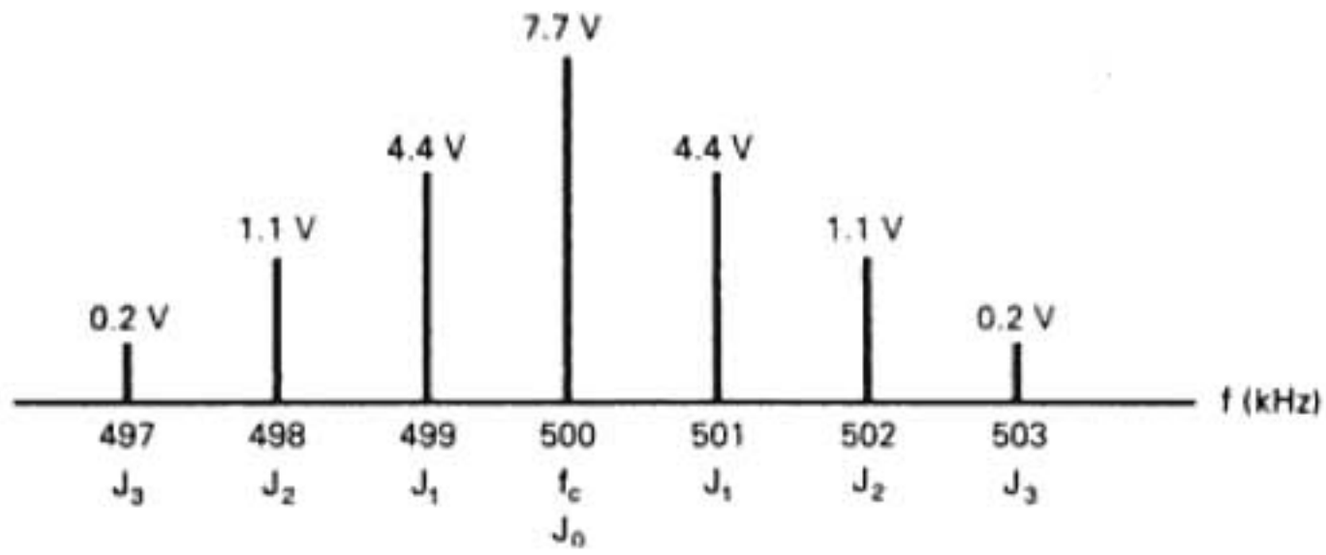


FIGURA 6-6 Espectro de frecuencias para el ejemplo 6-2

frecuencias laterales significativas. Un modulador convencional de AM de doble banda lateral sólo necesitaría 2 kHz de ancho de banda, y un sistema de banda lateral única sólo necesitaría 1 kHz.

En general, las formas de onda con modulación angular se clasifican como de *índice bajo, medio* o *alto*. Para el caso de bajo índice, el índice de modulación es menor que 1, y el caso de índice alto se da cuando el índice de modulación es mayor que 10. Los índices de modulación mayores que 1 y menores que 10 se clasifican como índice medio. Se ve, en la tabla 6-3, que con modulación angular de bajo índice, la mayor parte de la información de la señal está contenida en el primer conjunto de bandas laterales, y que el ancho mínimo de banda requerido es aproximadamente igual a dos veces la frecuencia máxima de la señal moduladora. Por esta razón, a los sistemas de FM de bajo índice se les llama de *FM de banda angosta*. Para una señal de índice alto se puede usar un método llamado *cuasi estacionario* para determinar el ancho de banda. En ese método se supone que la señal moduladora cambia con mucha lentitud. Por ejemplo, para un modulador de FM con sensibilidad a la desviación $K_1 = 2$ kHz/V y una señal moduladora de 1 V_p, la desviación máxima de frecuencia es $\Delta f = 2000$ Hz. Si la frecuencia de la señal moduladora es muy baja, el ancho de banda queda determinado por la desviación de frecuencia pico a pico. Por consiguiente, para grandes índices de modulación, el ancho mínimo de banda requerido para propagar una onda de frecuencia modulada es aproximadamente igual a la desviación de frecuencia pico a pico: $2\,\Delta f$.

Así, para una modulación de bajo índice, el espectro de frecuencias se asemeja al de la AM de banda lateral doble, y el ancho mínimo de banda es aproximadamente igual a

$$B = 2f_m \quad \text{Hz} \tag{6-25}$$

y para modulación de índice alto, el ancho mínimo de banda aproximado es

$$B = 2\Delta f \quad \text{Hz} \tag{6-26}$$

El ancho de banda real, necesario para pasar todas las bandas laterales significativas para una onda con modulación angular es igual a dos por el producto de la máxima frecuencia de señal moduladora, por la cantidad de bandas laterales significativas determinadas con la tabla de funciones Bessel. La ecuación para determinar el ancho mínimo de banda de una onda con modulación angular, usando la tabla de funciones de Bessel, es

$$B = 2(n \times f_m) \quad \text{Hz} \tag{6-27}$$

en donde n = cantidad de bandas laterales significativas

f_m = frecuencia de la señal moduladora (hertz)

En un memorando no publicado con fecha 28 de agosto de 1939, Carson estableció una regla general para estimar el ancho de banda de todos los sistemas con modulación angular, independiente del índice de modulación. A esto se llama *regla de Carson*. Enunciada en forma sencilla, esta regla establece el ancho de banda necesario para transmitir una onda con modulación angular, como igual a dos veces la suma de la desviación máxima de frecuencia por la frecuencia máxima de señal moduladora. Es decir,

$$B = 2(\Delta f + f_m) \quad \text{Hz} \tag{6-28}$$

donde Δf = desviación máxima de frecuencia (Hz)
f_m = frecuencia de la señal moduladora (Hz)

Para bajos índices de modulación, f_m es mucho mayor que Δf, y la ecuación 6-28 se reduce a la ecuación 6-25. Para grandes índices de modulación, Δf es mucho mayor que f_m, y la ecuación 6-28 se reduce a la ecuación 6-26.

La regla de Carson es una aproximación que da como resultado anchos de banda un poco menores que los determinados con la tabla de funciones de Bessel y la ecuación 6-27. La regla de Carson define un ancho de banda que abarca aproximadamente 98% de la potencia total en la onda modulada. El ancho real de banda necesario es una función de la forma de onda de la señal moduladora y de la calidad de transmisión deseada.

Ejemplo 6-3

Para un modulador de FM con desviación máxima de frecuencia Δf = 10 kHz, una frecuencia de señal moduladora f_m = 10 kHz, V_c = 10 V y una portadora de 500 kHz, determinar:

(a) El ancho de banda mínimo y real mediante la tabla de funciones de Bessel.
(b) El ancho mínimo aproximado de banda, con la regla de Carson.
A continuación:
(c) Graficar el espectro de frecuencias de salida con la aproximación de Bessel.

Solución (a) Al sustituir en la ecuación 6-18 se obtiene

$$m = \frac{10 \text{ kHz}}{10 \text{ kHz}} = 1$$

Según la tabla 6-3, un índice de modulación de 1 produce tres conjuntos de bandas laterales significativas. Se sustituye en la ecuación 6-27, y el ancho de banda es

$$B = 2(3 \times 10 \text{ kHz}) = 60 \text{ kHz}$$

(b) De acuerdo con la ecuación 6-28, el ancho mínimo de banda es

$$B = 2(10 \text{ kHz} + 10 \text{ kHz}) = 40 \text{ kHz}$$

(c) El espectro de frecuencias de salida determinado con la aproximación de Bessel se muestra en la fig. 6-7

Se puede ver que, según el ejemplo 6-3, hay una diferencia apreciable en los anchos de banda mínimos determinados con la regla de Carson y con la tabla de funciones de Bessel. El ancho de banda según la regla de Carson es menor que el ancho de banda mínimo y real requerido para pasar todos los conjuntos de bandas laterales significativas definidos por la tabla de funciones de Bessel. Por consiguiente, un sistema que se diseñara con la regla de Carson tendría banda más angosta y, por consiguiente, menor desempeño que uno diseñado con la tabla de funciones de Bessel. Para índices de modulación mayores que 5, la regla de Carson es una buena aproximación al ancho de banda real requerido.

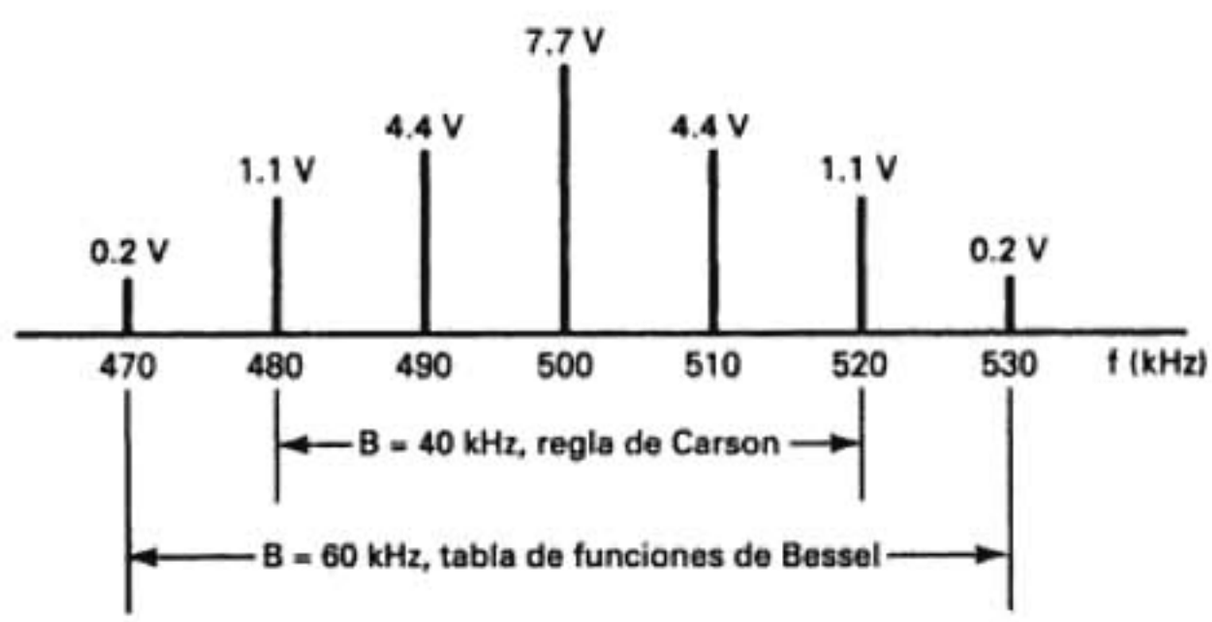


FIGURA 6-7 Espectro de frecuencias para el ejemplo 6-3

Relación de desviación. Para determinado sistema de FM, el ancho mínimo de banda es máximo cuando se obtiene la desviación máxima de frecuencia con la frecuencia máxima de la señal moduladora; es decir, la frecuencia moduladora máxima se presenta con la máxima amplitud permitida. Por definición, la *relación de desviación* (DR, de *deviation ratio*) es el índice de modulación *en el peor de los casos*, y es igual a la desviación máxima de frecuencia máxima dividida entre la frecuencia máxima de señal moduladora. El índice de modulación en el peor de los casos produce el espectro más amplio de frecuencias de salida. Es, por lo anterior,

$$\mathrm{DR} = \frac{\Delta f_{(\text{máx})}}{f_{m(máx)}} \tag{6-29}$$

donde DR = relación de desviación (adimensional)
$\Delta f_{(\text{máx})}$ = desviación máxima de frecuencia máxima (hertz)
$f_{m(\text{máx})}$ = frecuencia máxima de la señal moduladora (hertz)

Por ejemplo, para la porción de audio de una estación comercial de TV, la FCC establece que la desviación máxima de frecuencia sea 50 kHz, y la frecuencia máxima de la señal moduladora es 15 kHz. Por consiguiente, la relación de desviación para una estación comercial de TV es

$$\mathrm{DR} = \frac{50\ \text{kHz}}{15\ \text{kHz}} = 3.33$$

Esto no quiere decir que siempre que haya un índice de modulación de 3.33, también se obtenga al mismo tiempo el máximo ancho de banda. Quiere decir que siempre que haya un índice de modulación de 3.33 para una frecuencia máxima de señal moduladora, se presenta el ancho de banda máximo.

Ejemplo 6-4

(a) Determinar la relación de desviación y el ancho de banda para el índice de modulación en el peor de los casos (ancho de banda máximo) en un transmisor en la banda comercial de FM con una desviación máxima de frecuencia de 75 kHz y una frecuencia máxima de señal moduladora de 15 kHz.

(b) Calcular la relación de desviación y el ancho de banda máximo para un índice igual de modulación, sólo con la mitad de desviación de frecuencia máxima y la frecuencia de la señal moduladora.

Solución (a) La relación de desviación se calcula sustituyendo en la ecuación 6-29

$$\mathrm{DR} = \frac{75\ \text{kHz}}{15\ \text{kHz}} = 5$$

Según la tabla 6-3, un índice de modulación de 5 produce 8 bandas laterales significativas. Se sustituye en la ecuación 6-27, para obtener

$$B = 2(8 \times 15{,}000) = 240\ \text{kHz}$$

(b) Para una desviación de frecuencia de 37.5 kHz y una desviación de frecuencia de la señal moduladora f_m = 7.5 kHz, el índice de modulación es

$$m = \frac{37.5\ \text{kHz}}{7.5\ \text{kHz}} = 5$$

y el ancho de banda es $B = 2(8 \times 7500) = 120\ \text{kHz}$

En el ejemplo anterior se ve que, aunque se alcanzó el mismo índice de modulación de 5 con dos frecuencias y amplitudes distintas de señal moduladora, se produjeron dos bandas laterales distintas. Hay una cantidad infinita de combinaciones de frecuencia de señal moduladora y desviación de frecuencia que producen un índice de modulación igual a 5. Sin embargo, el caso obtenido a partir de la frecuencia máxima de señal moduladora y la desviación máxima de frecuencia producirá siempre el ancho de banda máximo.

A primera vista, parecerá que un índice de modulación mayor con una frecuencia menor de señal moduladora generaría un ancho de banda mayor, porque se producen más conjuntos de bandas laterales; pero recuérdese que las bandas laterales estarán más cercanas entre sí. Por ejemplo, una señal moduladora de 1 kHz que produzca 10 kHz de desviación de frecuencia tiene un índice de modulación $m = 10$ y produce 14 conjuntos distintos de bandas laterales. Sin embargo,

las bandas laterales sólo están desplazadas 1 kHz entre sí y, en consecuencia, el ancho de banda total es 28,000 Hz [2(14 × 1000)].

Con la regla de Carson, las mismas condiciones producen el ancho de banda más amplio (el peor de los casos). Para la desviación máxima de frecuencia y la frecuencia máxima de señal moduladora, el ancho de banda máximo usando la regla de Carson para el ejemplo 6-4a es

$$\begin{aligned} B &= 2[\Delta f_{(\text{máx})} + f_{m(\text{máx})}] \\ &= 2(75\text{ kHz} + 15\text{ kHz}) \\ &= 180\text{ kHz} \end{aligned}$$

BANDA COMERCIAL DE EMISIÓN EN FM

La FCC ha asignado una banda de frecuencias de 20 MHz al servicio de emisiones de FM, que va de los 88 a los 108 MHz. Esta banda de 20 MHz se divide en canales de 100 y 200 kHz de ancho que comienzan en 88.1 MHz; es decir, 88.3 MHz, 88.5 MHz y así sucesivamente. Para obtener una música de alta calidad y confiable, la desviación máxima de frecuencia permitida es 75 kHz con una frecuencia máxima de señal moduladora de 15 kHz.

De acuerdo con la ecuación 6-27, el índice de modulación en el peor de los casos, es decir, la relación de desviación, para un canal comercial es 75 kHz/15 kHz = 5. De acuerdo con la tabla de funciones de Bessel, se producen ocho pares de frecuencias laterales significativas cuando el índice de modulación es 5. Por consiguiente, de acuerdo con la ecuación 6-27, el ancho de banda mínimo y necesario para pasar todas las frecuencias laterales significativas es $B = 2(8 \times 15\text{ Hz}) = 240$ kHz, que es 40 kHz mayor que el ancho de banda asignado por la FCC. En esencia, esto quiere decir que se permite que las frecuencias laterales máximas de un canal entren a canales adyacentes, produciendo una interferencia llamada *interferencia por canal adyacente*. En general eso no constituye un problema, porque históricamente la FCC sólo ha asignado uno de cada dos canales, es decir, un canal sí y uno no, en un área geográfica dada. Por consiguiente, hay casi siempre una banda de protección de 200 kHz a cada lado de cada canal asignado. Además, el séptimo y octavo conjuntos de frecuencias laterales tienen poca potencia, y también es muy improbable obtener alguna vez la máxima desviación de frecuencia a la frecuencia máxima de señal moduladora. Es irónico que, si se usa la aproximación de Carson, el ancho de banda para los canales comerciales es 2(75 kHz + 15 kHz) = 180 kHz, bastante bien dentro de los límites de banda asignados por la FCC.

REPRESENTACIÓN FASORIAL DE LA ONDA CON MODULACIÓN ANGULAR

Como en la modulación de amplitud, se puede representar fasorialmente una onda con modulación de ángulo. El diagrama fasorial para una onda de modulación angular de bajo índice, con señal moduladora de una sola frecuencia, se muestra en la fig. 6-8. Para este caso especial de $m < 1$, sólo se considera el primer conjunto de pares de bandas laterales, y el diagrama fasorial se parece mucho al de una onda de AM, excepto que una de las frecuencias laterales tiene inversión de fase. El vector resultante tiene una amplitud cercana a la unidad, siempre, y una desviación máxima de fase de m radianes. Es importante observar que si se incluyeran las frecuencias laterales debidas a los términos de orden superior, el vector no tendría variaciones de amplitud. La línea interrumpida de la fig. 6-8e es el lugar geométrico de la resultante, formada por la portadora y el primer conjunto de frecuencias laterales.

La fig. 6-9 muestra el diagrama fasorial de una onda con modulación angular, de alto índice, con cinco conjuntos de frecuencias laterales; para simplificar sólo se muestran los vectores de los dos primeros conjuntos. El vector resultante es la suma del componente de portadora y de los componentes de las frecuencias laterales significativas, con sus magnitudes ajustadas de acuerdo con la tabla de funciones de Bessel. Cada frecuencia lateral se desplaza otros 90° respecto a la frecuencia lateral precedente. El lugar geométrico de la aproximación resultante para cinco componentes es curvo y sigue de cerca al lugar geométricop de la señal. Por definición, el lugar geométrico es un segmento del círculo cuyo radio es igual a la amplitud de la portadora no modulada. Se debe notar que la amplitud de la señal resultante y, en consecuencia, la potencia de la señal, permanecen constantes.

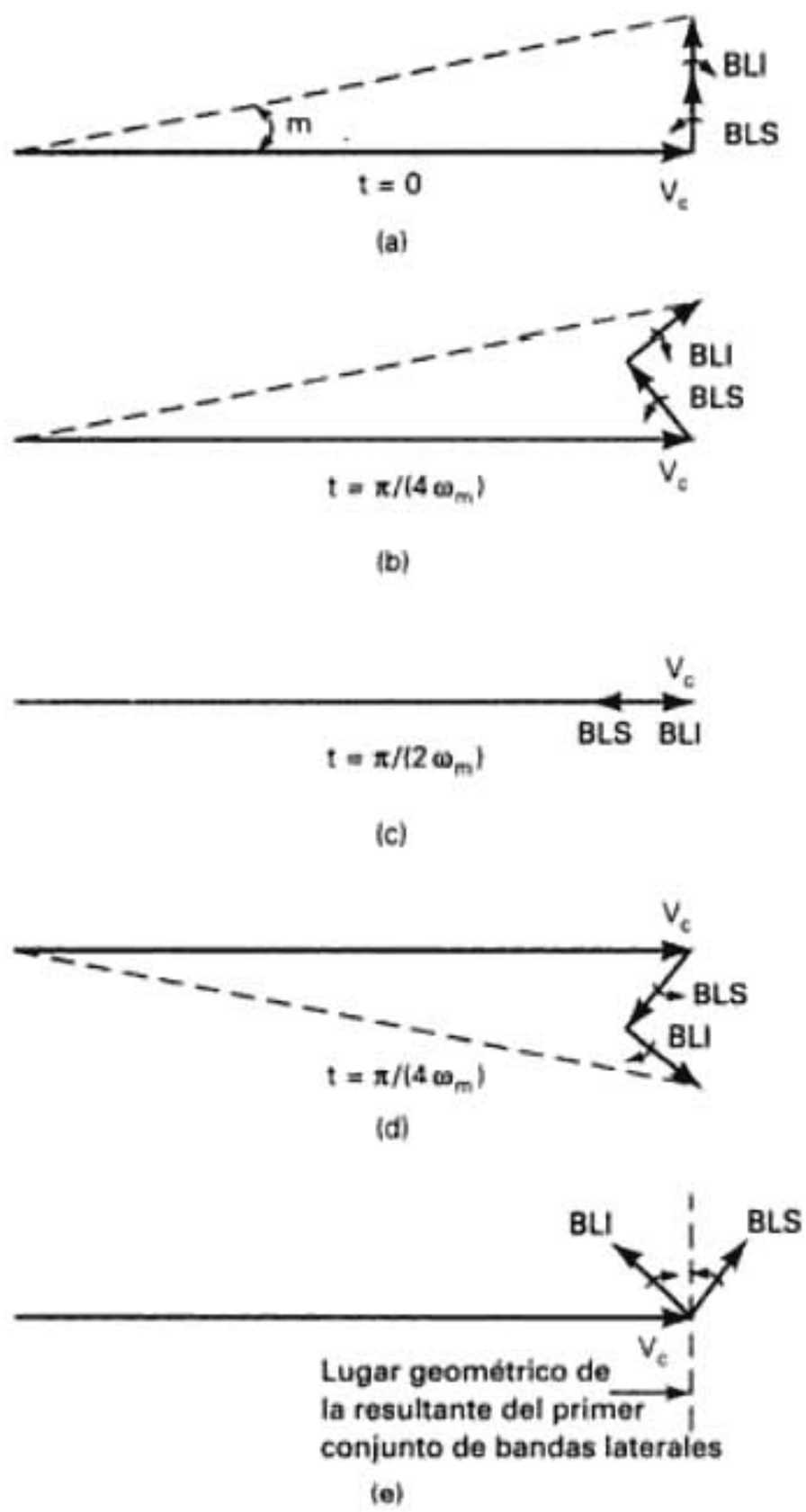


FIGURA 6-8 Representación fasorial de la modulación angular, bajo índice de modulación

POTENCIA PROMEDIO DE UNA ONDA DE MODULACIÓN ANGULAR

Una de las diferencias más importantes entre modulación de ángulo y modulación de amplitud es la distribución de la potencia en la onda modulada. A diferencia de la AM, la potencia total en la onda con modulación angular es igual a la potencia de la portadora no modulada; es decir, las bandas laterales no agregan potencia a la señal modulada compuesta. Por consiguiente, con la modulación de ángulo, la potencia que había originalmente en la portadora no modulada se redistribuye entre la portadora y sus bandas laterales. La potencia promedio de una onda con modulación angular es independiente de la señal moduladora, del índice de modulación y de la desviación de frecuencia. Es igual a la potencia promedio de la portadora no modulada, e independiente de la profundidad de modulación. La ecuación de la potencia promedio en la portadora no modulada es

$$P_c = \frac{V_c^2}{2R} \text{ W} \tag{6-30}$$

en donde P_c = potencia de la portadora (watts)
V_c = voltaje máximo de portadora no modulada (volts)
R = resistencia de carga (ohms)

La potencia instantánea total en una portadora con modulación de ángulo es

$$P_t = \frac{m(t)^2}{R} \text{ W} \tag{6-31a}$$

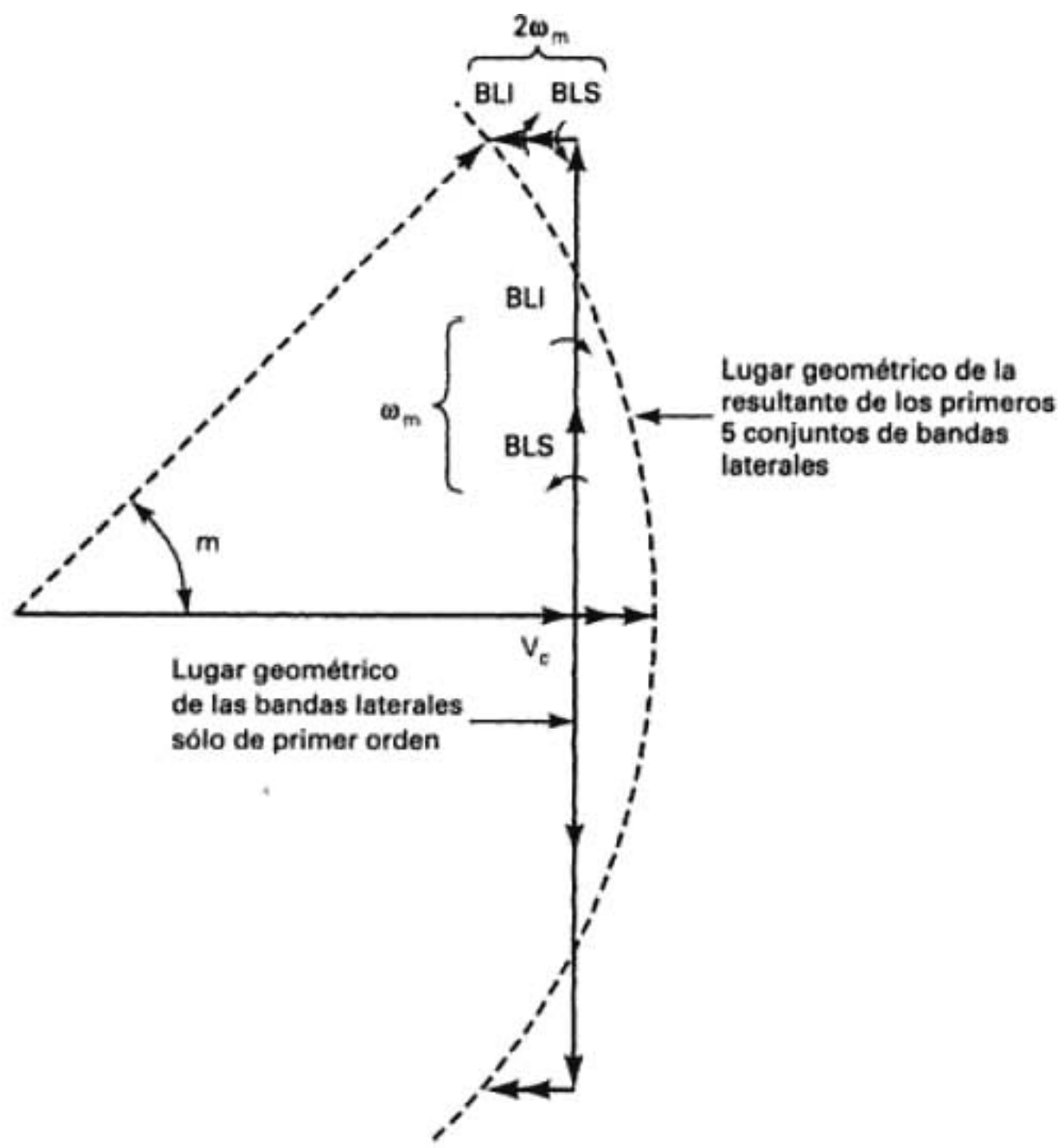


FIGURA 6-9 Representación fasorial de la modulación angular, alto índice de modulación

Se sustituye $m(t)$

$$P_t = \frac{V_c^2}{R}\cos^2[\omega_c t + \theta(t)] \tag{6-31b}$$

y al desarrollar se obtiene

$$= \frac{V_c^2}{R}\left\{\frac{1}{2} + \frac{1}{2}\cos[2\omega_c t + 2\theta(t)]\right\} \tag{6-31c}$$

En la ecuación 6-31c, el segundo término consiste en una cantidad infinita de componentes de frecuencias laterales, respecto a una frecuencia $2\omega_c$ igual al doble de la frecuencia de la portadora. En consecuencia, el valor promedio del segundo término es cero, y la potencia promedio de la onda modulada se reduce a

$$P_t = \frac{V_c^2}{2R}\ \text{W} \tag{6-32}$$

Nótese que las ecuaciones 6-30 y 6-32 son idénticas, por lo que la potencia promedio de la portadora modulada debe ser igual a la potencia promedio de la portadora no modulada. La potencia de la portadora modulada es la suma de las potencias de la portadora y de los componentes de frecuencias de banda lateral. Por consiguiente, la potencia total de la onda modulada es

$$P_t = P_0 + P_1 + P_2 + P_3 + P_n \tag{6-33a}$$

$$P_t = \frac{V_c^2}{2R} + \frac{2(V_1)^2}{2R} + \frac{2(V_2)^2}{2R} + \frac{2(V_3)^2}{2R} + \frac{2(V_n)^2}{2R} \tag{6-33b}$$

siendo P_t = potencia total (watts)
P_0 = potencia de la portadora modulada (watts)
P_1 = potencia en el primer conjunto de bandas laterales (watts)
P_2 = potencia en el segundo conjunto de bandas laterales (watts)
P_3 = potencia en el tercer conjunto de bandas laterales (watts)
P_n = potencia en el n-ésimo conjunto de bandas laterales (watts)

Ejemplo 6-5

(a) Determinar la potencia de la portadora no modulada para el modulador de FM y las condiciones que aparecen en el ejemplo 6-2. (Suponer una resistencia de carga $R_L = 50\ \Omega$.)
(b) Calcular la potencia total en la onda con modulación angular.

Solución (a) Se sustituye la ecuación 6-30, para obtener

$$P_c = \frac{10^2}{2(50)} = 1\ \text{W}$$

(b) Se sustituyen valores en la ecuación 6-33a para obtener

$$P_t = \frac{7.7^2}{2(50)} + \frac{2(4.4)^2}{2(50)} + \frac{2(1.1)^2}{2(50)} + \frac{2(0.2)^2}{2(50)}$$
$$= 0.5929 + 0.3872 + 0.0242 + 0.0008 = 1.0051\ \text{W}$$

Los resultados de a) y b) no son exactamente iguales, porque los valores de la tabla de funciones de Bessel se han redondeado. Sin embargo, esos resultados son lo bastante cercanos para ilustrar que es igual la potencia en la onda modulada y en la portadora no modulada.

RUIDO Y MODULACIÓN ANGULAR

Cuando se agrega ruido térmico con densidad espectral constante a una señal de FM, produce una desviación indeseada de la frecuencia de la portadora. La magnitud de esta desviación no deseada depende de la amplitud relativa del ruido con respecto a la portadora. Cuando se demodula esta desviación indeseada de portadora, se transforma en ruido, si sus componentes de frecuencia quedan dentro del espectro de frecuencias de información. La forma espectral del ruido demodulado depende de si se usa demodulador de PM o de FM. El voltaje de ruido en la salida de un demodulador de PM es constante respecto a la frecuencia, mientras que en la salida de un demodulador de FM aumenta en forma lineal con la frecuencia. A esto se le suele llamar *triángulo de ruido* de FM, y se ilustra en la fig. 6-10. Se ve que el voltaje de ruido demodulado es inherentemente mayor para mayores frecuencias de señal moduladora.

Modulación de fase debida a una señal de interferencia

La fig. 6-11 muestra la modulación de fase causada por una señal de ruido de una sola frecuencia. El componente de ruido V_n está separado del componente V_c de la señal, por la frecuencia

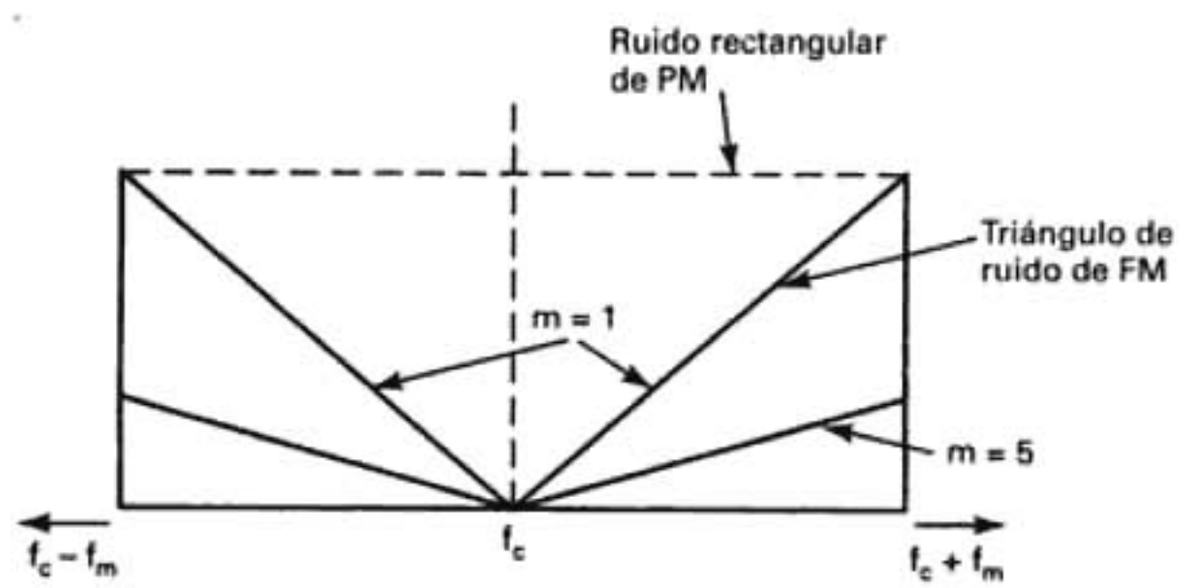


FIGURA 6-10 Triángulo de ruido de FM

f_n. Esto se ve en la fig. 6-11b. Si se supone que $V_c > V_n$, la desviación máxima de fase debida a la interferencia de una senoidal de una sola frecuencia se presenta cuando los voltajes de señal y de ruido están en cuadratura, y para ángulos pequeños se aproxima mediante

$$\Delta\theta_{\text{máx}} \simeq \frac{V_n}{V_c} \text{ rad} \tag{6-34}$$

La fig. 6-11c muestra el efecto de *limitar* la amplitud de la señal compuesta de ruido de FM. (La limitación se usa normalmente en receptores con modulación de ángulo, y se explica en el capítulo 7.) Se puede ver que la señal de ruido, de frecuencia única, se ha transpuesto en dos pares de bandas laterales de ruido, cada una con amplitud $V_n/2$. Estas bandas laterales son coherentes, y en consecuencia la desviación máxima de fase sigue siendo V_n/V_c radianes. Sin embargo, se han eliminado las variaciones indeseadas de amplitud, lo cual reduce la potencia total, pero no reduce la interferencia en la señal demodulada, debida a la desviación indeseada de fase.

Modulación de frecuencia debida a una señal de interferencia

De acuerdo con la ecuación 6-5, la desviación instantánea de frecuencia, $\Delta f(t)$, es la primera derivada de la desviación instantánea de fase, $\theta(t)$. Cuando el componente de la portadora es mucho mayor que el voltaje de ruido de interferencia, la desviación aproximada de fase es

$$\theta(t) = \frac{V_n}{V_c} \text{sen}(\omega_n t + \theta_n) \text{ rad} \tag{6-35}$$

y sacando la primera derivada se obtiene

$$\Delta\omega(t) = \frac{V_n}{V_c} \omega_n \cos(\omega_n t + \theta_n) \text{ rad/s} \tag{6-36}$$

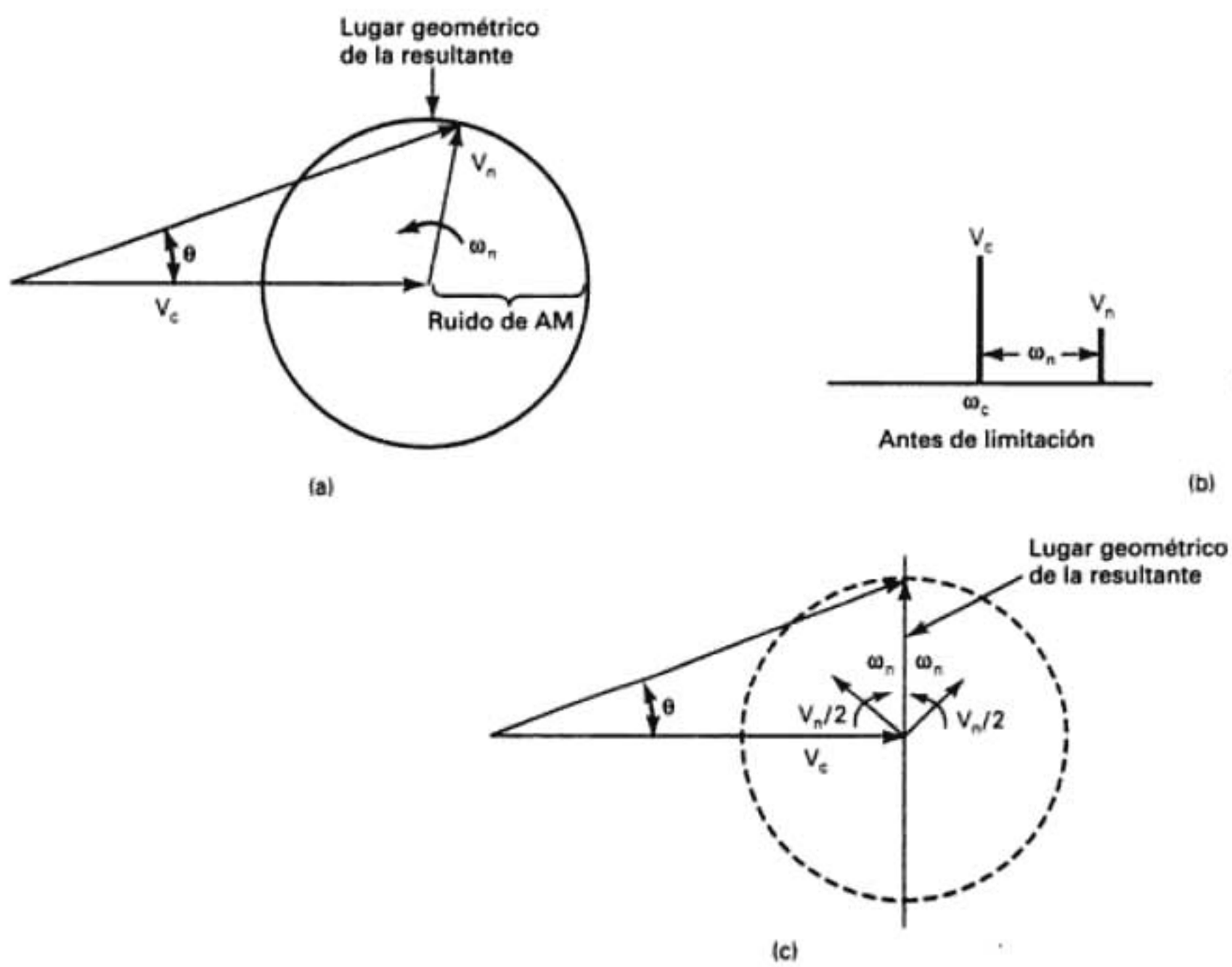


FIGURA 6-11 Interferencia senoidal de ruido: (a) antes de limitación; (b) espectro de frecuencias; (c) después de limitación

Entonces, la desviación máxima de frecuencia es

$$\Delta\omega_{\text{máx}} = \frac{V_n}{V_c}\,\omega_n \text{ rad / s} \tag{6-37}$$

$$\Delta f_{\text{máx}} = \frac{V_n}{V_c} f_n \text{ Hz} \tag{6-38}$$

Al reordenar la ecuación 6-18 se puede ver que la desviación máxima de frecuencia Δf es una función de la frecuencia de la señal moduladora y del índice de modulación. Así, para una frecuencia moduladora de ruido f_n, la desviación máxima de frecuencia es

$$\Delta f_{\text{máx}} = m f_n \quad \text{Hz} \tag{6-39}$$

en donde m es igual al índice de modulación ($m << 1$).

Se ve que, de acuerdo con la ecuación 6-39, mientras más lejos se desplaza la frecuencia de ruido respecto a la de la portadora, la desviación de frecuencia es mayor. Por consiguiente, las frecuencias de ruido que producen componentes en el extremo alto del espectro de frecuencias de señal moduladora producen más desviación de frecuencia, para la misma desviación de fase, que las frecuencias que están en el extremo bajo. Los demoduladores de FM generan un voltaje de salida que es proporcional a la desviación de frecuencia, y es igual a la diferencia entre la frecuencia de la portadora y la frecuencia de la señal de interferencia. Por lo anterior, los componentes de ruido de alta frecuencia producen más ruido demodulado que los componentes de baja frecuencia.

La relación de señal a ruido en la salida de un demodulador de FM, debida a desviación no deseada de frecuencia por una senoidal de interferencia, es la relación de la desviación máxima de frecuencia debida a la señal de información, entre la desviación máxima de frecuencia debida a la señal de interferencia.

$$\frac{S}{N} = \frac{\Delta f_{\text{debida a la señal}}}{\Delta f_{\text{debida al ruido}}} \tag{6-40}$$

Ejemplo 6-6

Determinar lo siguiente, para una portadora con modulación angular, $V_c = 6\cos(2\pi 110 \text{ MHz } t)$ con desviación de frecuencia de 75 kHz, debida a la señal de información, y una señal de interferencia con frecuencia única, $V_n = 0.3\cos(2\pi 109.985 \text{ MHz } t)$.

(a) Frecuencia de la señal demodulada de interferencia.

(b) Desviaciones máximas de fase y de frecuencia, debidas a la señal de interferencia.

(c) Relación de voltajes de señal a ruido en la salida del demodulador.

Solución (a) La frecuencia de la interferencia de ruido es igual a la diferencia entre la frecuencia de la portadora y la de la señal de interferencia (de frecuencia única).

$$f_c - f_n = 110 \text{ MHz} - 109.985 \text{ MHz} = 15 \text{ kHz}$$

(b) Al sustituir en la ecuación 6-34 se obtiene

$$\Delta\theta_{\text{máx}} = \frac{0.3}{6} = 0.05 \text{ rad}$$

Y al sustituir en la ecuación 6-38 se obtiene

$$\Delta f_{\text{máx}} = \frac{0.3 \times 15 \text{ kHz}}{6} = 750 \text{ Hz}$$

(c) La relación S/N de voltajes, debida al tono de interferencia, es la relación de la amplitud de la portadora entre la amplitud de la señal de interferencia, o sea

$$\frac{6}{0.3} = 20$$

La relación S/N de voltajes después de la modulación se calcula con la ecuación 6-40

$$\frac{S}{N} = \frac{75 \text{ kHz}}{750 \text{ Hz}} = 100$$

Por lo anterior, hay un mejoramiento de señal a ruido de 100/20 = 5, es decir, 20 log 5 = 14 dB.

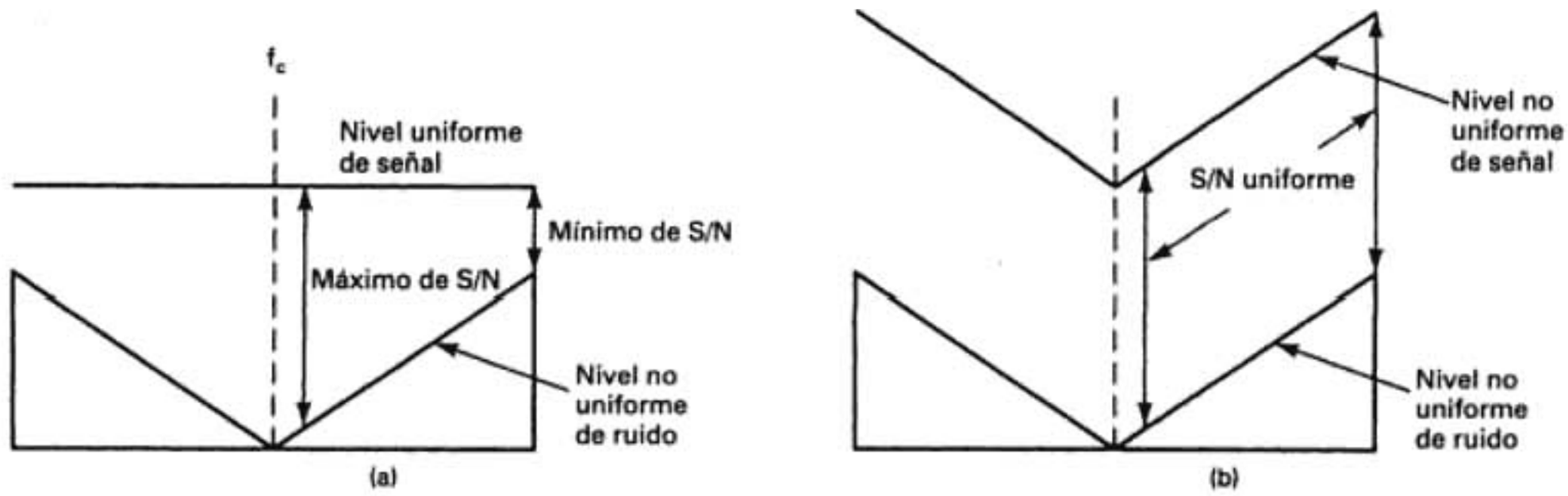


FIGURA 6-12 Relación señal a ruido en FM: (a) sin preénfasis; (b) con preénfasis

PREÉNFASIS Y DEÉNFASIS

El triángulo de ruido de la fig. 6-10 muestra que, con FM, hay una distribución no uniforme de ruido. El ruido en las frecuencias más altas de señal moduladora tiene, en forma inherente, mayor amplitud que en las frecuencias más bajas. Por lo anterior, para señales de información con nivel uniforme de señal, se produce una relación no uniforme de señal a ruido, y las frecuencias más altas de señal moduladora tienen menor relación de señal a ruido que las frecuencias más bajas. Esto se ve en la fig. 6-12a, donde la relación S/N es menor en los extremos de alta frecuencia del triángulo. Para compensar esto, las señales moduladoras se *enfatizan* o refuerzan en amplitud, en el transmisor, antes de hacer la modulación. Para compensar este refuerzo, se atenúan, o desenfatizan las señales de alta frecuencia en el receptor, después de hacer la demodulación. La deénfasis es el recíproco del preénfasis y, por consiguiente, una red de deénfasis restaura las características originales de amplitud en función de frecuencia, a las señales de información. En esencia, la red de preénfasis permite que las señales moduladoras de alta frecuencia modulen a la portadora a un grado mayor y así causen más desviación de frecuencia que la que producirían sus amplitudes originales. Las señales de alta frecuencia se propagan a través del sistema a un nivel elevado (mayor desviación de frecuencia), se demodulan y a continuación se restauran a sus proporciones originales de amplitud. La fig. 16-12b muestra los efectos del preénfasis y la deénfasis sobre la relación de señal a ruido. Allí se ve que producen una relación más uniforme de señal a ruido en todo el espectro de frecuencias de la señal moduladora.

Una red de preénfasis es un filtro pasaaltas, es decir, un diferenciador, y una red de deénfasis es un filtro pasabajas, o sea un integrador. La fig. 6-13a muestra el diagrama de una red activa de preénfasis, y una pasiva de deénfasis. Sus curvas correspondiente de respuesta se ven en la fig. 6-13b. Una red de preénfasis proporciona un aumento constante de amplitud de la señal moduladora con un aumento de frecuencia. En la FM, más o menos se logra una mejora de 12 dB en cuanto a desempeño contra ruido, cuando se usa preénfasis y deénfasis. La frecuencia de corte, que es donde comienzan el preénfasis y la deénfasis, se determina con la constante de tiempo RC o L/R de la red. Esta frecuencia de corte se presenta en la frecuencia donde X_C o X_L es igual a R. La definición matemática de la frecuencia de quiebre es

$$f_b = \frac{1}{2\pi RC} \tag{6-41a}$$

$$f_b = \frac{1}{2\pi L/R} \tag{6-41b}$$

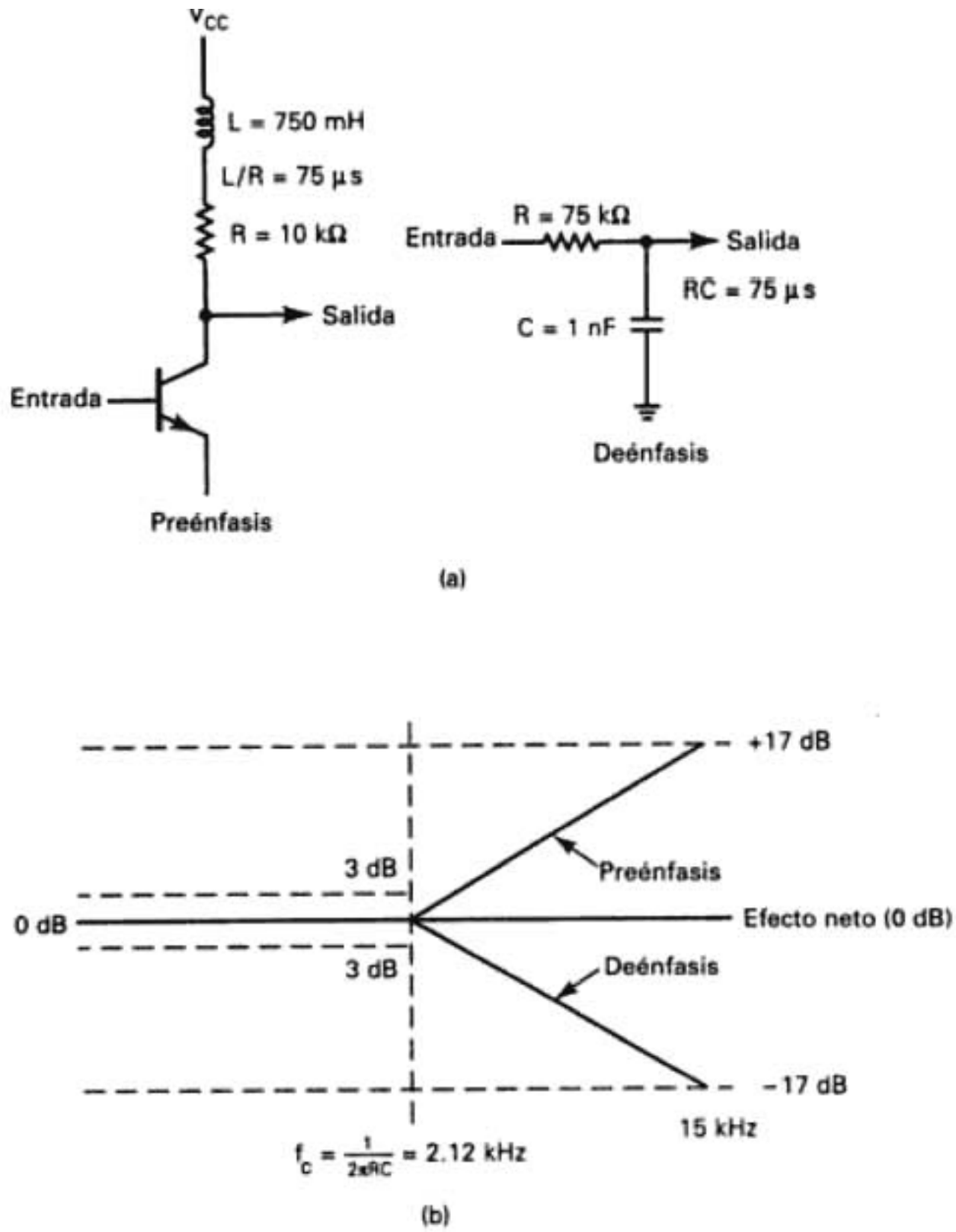


FIGURA 6-13 Preénfasis y deénfasis: (a) diagramas eléctricos; (b) curvas de atenuación

Las redes de la fig. 6-13 son para la banda comercial de FM, que usa una constante de tiempo de 75 μs. Por consiguiente, la frecuencia aproximada de corte es

$$f_b = \frac{1}{2\pi 75\ \mu s} = 2.12\ \text{kHz}$$

La transmisión en FM de la parte de audio en la emisión comercial de TV usa una constante de tiempo de 50 μs.

Como se ve en la fig. 6-13, se usa una red activa de preénfasis, y no pasiva, porque una red pasiva de preénfasis produce pérdidas en todas las frecuencias, y se introduce más pérdida en las frecuencias menores de la señal moduladora. El resultado de usar una red pasiva sería una disminución de la relación de señal a ruido en las frecuencias menores de la señal moduladora, más que un aumento de esa relación en las frecuencias mayores de la misma.

De acuerdo con la anterior explicación y la fig. 6-13, se puede ver que la amplitud de salida de una red de preénfasis aumenta con la frecuencia, cuando ésta es mayor que la frecuencia de corte. Regresando a la ecuación 6-16, se puede ver que si los cambios de frecuencia f_m de la señal moduladora producen cambios correspondientes en su amplitud V_m, el índice de modulación m permanece constante al cambiar la frecuencia. Esto es, naturalmente, una característica de la modulación de fase. En consecuencia, con moduladores de banda comercial, las frecuencias menores que 2112 Hz producen modulación de frecuencia, y las frecuencias mayores

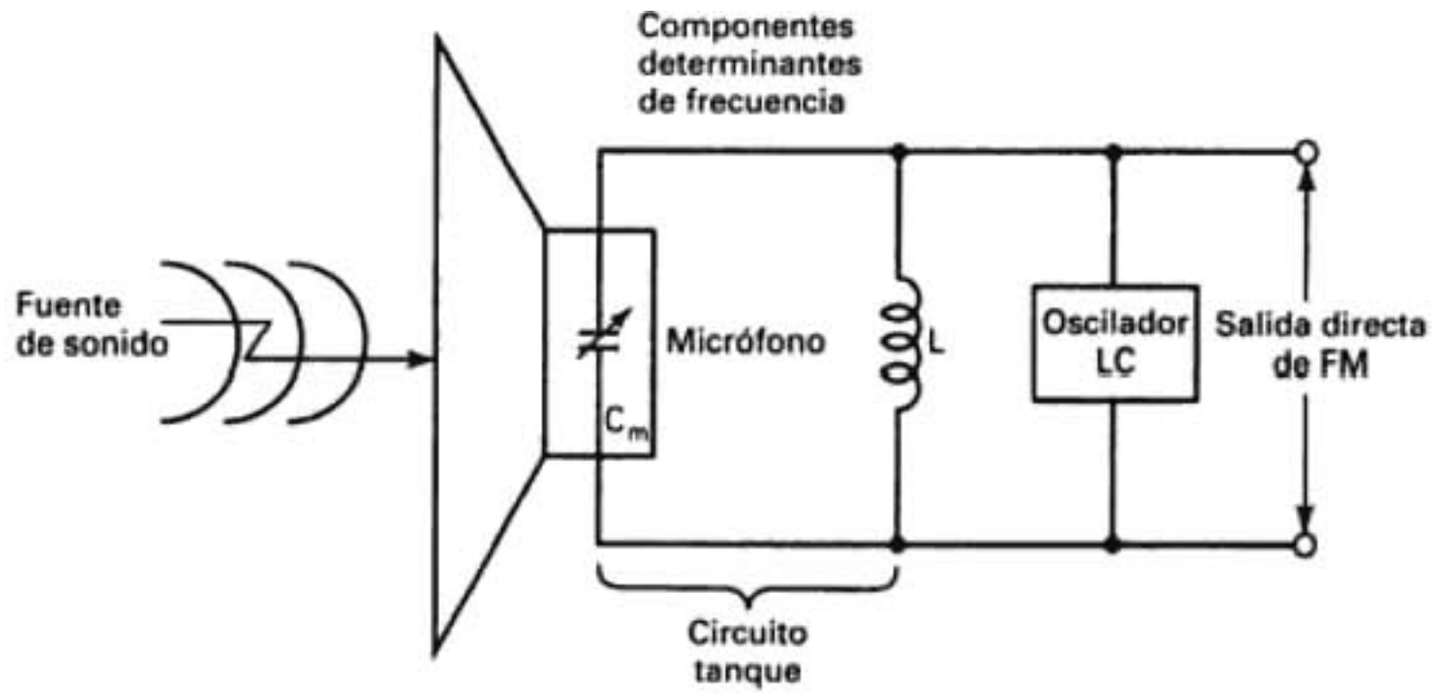


FIGURA 6-14 Modulador directo de FM sencillo

que 2112 Hz producen modulación de fase. Sin embargo, la conversión de FM a PM no es la función de una red de preénfasis, sino más bien una consecuencia.

MODULADORES DE FRECUENCIA Y FASE

En esencia, la diferencia principal entre los moduladores de frecuencia y los de fase estriba en si cambia la frecuencia o la fase de la portadora, en forma directa, debido a la señal moduladora, y de cuál propiedad cambia en forma indirecta. Cuando la señal de información modifica la frecuencia del oscilador de portadora, resulta la FM directa (PM indirecta). Cuando la señal de información modula la fase de la señal portadora, resulta la PM directa (FM indirecta).

La desventaja primaria de la FM directa es que se deben sintonizar osciladores LC, relativamente inestables, para producir la frecuencia de portadora, lo cual excluye el uso de osciladores de cristal. Por lo anterior, la FM directa requiere la adición de algún tipo de circuito de control automático de frecuencia para mantener la frecuencia de la portadora dentro de los estrictos límites de estabilidad de frecuencia de la FCC. La ventaja obvia de la FM directa es que se obtienen con facilidad desviaciones relativamente altas de frecuencia e índices de modulación, debido a que los osciladores son inherentemente inestables.

La principal ventaja de la PM directa es que el oscilador de portadora está separado del circuito real de modulación y, en consecuencia, puede ser una fuente extremadamente estable, como por ejemplo un oscilador de cristal. La desventaja obvia de la PM directa es que los osciladores de cristal son inherentemente estables, y es más difícil lograr con ellos grandes desviaciones de fase e índices de modulación.

Moduladores directos de FM

La modulación angular de *FM directa* es aquella en la que se varía (se desvía) la frecuencia de la moduladora, en forma directa, mediante la señal moduladora. Con la FM directa, la desviación instantánea de frecuencia es directamente proporcional a la amplitud de la señal moduladora. La fig. 6-14 muestra el diagrama de un generador sencillo (aunque muy impráctico) de FM directa. El circuito tanque (L y C_m), es la sección determinante de frecuencia, para un oscilador LC normal. El micrófono de capacitor es un transductor que convierte la energía acústica en energía mecánica, que se usa para variar la distancia entre las placas de C_m y, en consecuencia, cambia su capacitancia. Al variar C_m, varía la frecuencia de resonancia. Así, la frecuencia de salida del oscilador varía en forma directa con la fuente externa de sonido. Es FM directa porque cambia la frecuencia del oscilador en forma directa mediante la señal moduladora, y la magnitud del cambio de frecuencia es proporcional a la amplitud del voltaje de la señal moduladora. Hay tres métodos comunes para producir modulación directa de frecuencia: con moduladores de diodo varactor, con moduladores de reactancia de FM y con moduladores de FM directa en circuito integrado lineal.

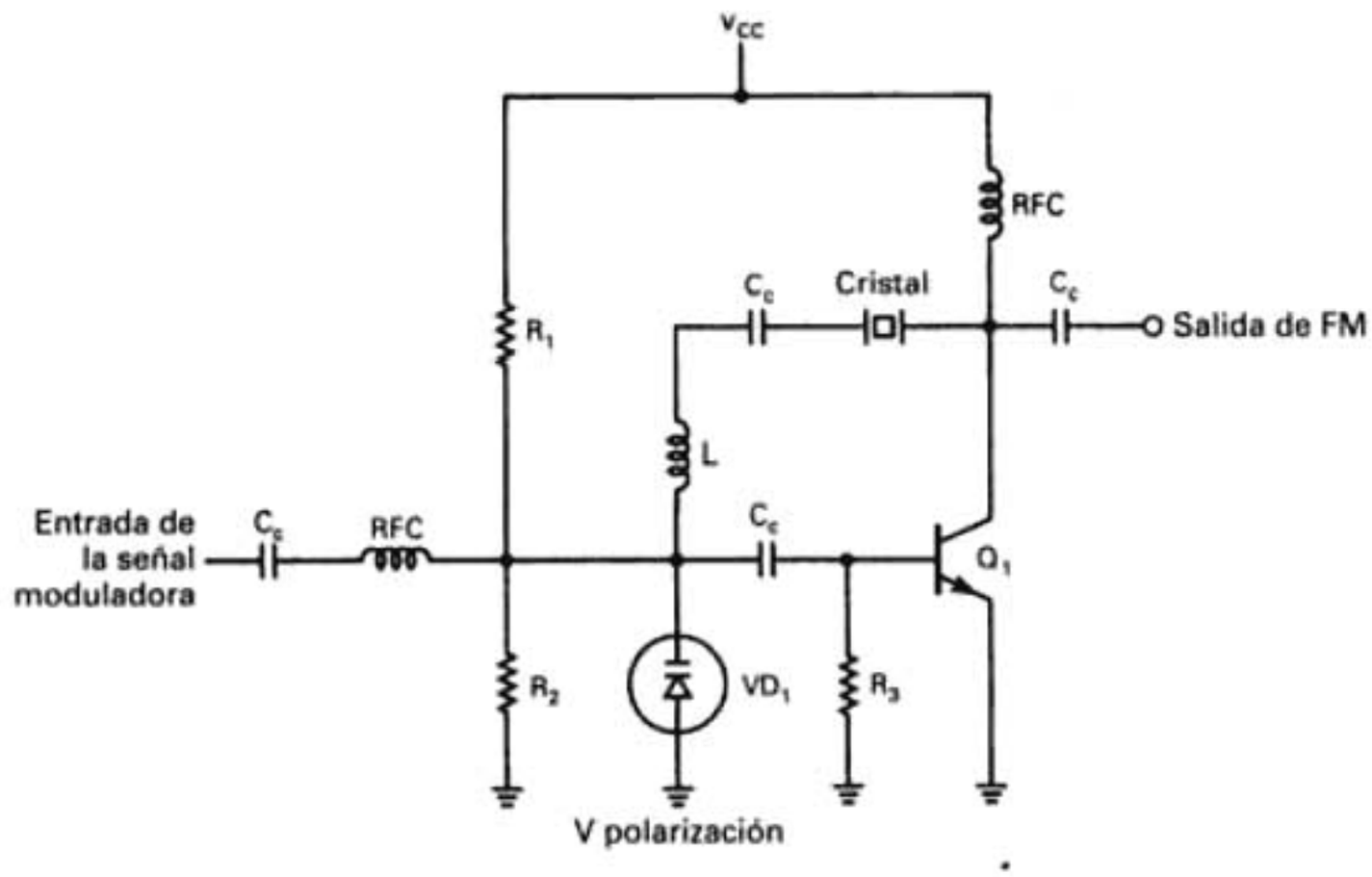


FIGURA 6-15 Modulador directo de FM con varactor

Moduladores con diodo varactor. La fig. 6-15 muestra el diagrama de un generador más práctico de FM que usa un diodo varactor, para desviar la frecuencia de un oscilador de cristal. R_1 y R_2 desarrollan un voltaje de cd que polariza inversamente al diodo varactor VD_1, y determina la frecuencia del oscilador en reposo. El voltaje de la señal moduladora externa se suma y se resta a la polarización de cd, con lo que cambia la capacitancia del diodo y, por consiguiente, la frecuencia de oscilación (véase una descripción más detallada de los diodos varactores en el capítulo 2). Las alternancias positivas de la señal moduladora aumentan la polarización inversa en VD_1, con lo que disminuye su capacitancia y aumenta la frecuencia de oscilación. Al revés, las alternancias negativas de la señal moduladora disminuyen la frecuencia de oscilación. Los moduladores de diodo varactor son muy populares, porque son sencillos de usar, son confiables y tienen la estabilidad de un oscilador de cristal. Sin embargo, ya que se usa un cristal, la desviación máxima de frecuencia se limita a valores relativamente bajos. En consecuencia, se usan principalmente en aplicaciones de bajo índice, como por ejemplo, en radiomóvil en dos sentidos.

La fig. 6-16 muestra el diagrama simplificado de un generador de FM con oscilador controlado por voltaje (VCO, de *voltage-controlled oscillator*). De nuevo se usa un diodo varactor para transformar los cambios de amplitud de la señal moduladora en cambios de frecuencia. La frecuencia central del oscilador se calcula como sigue

$$f_c = \frac{1}{2\pi\sqrt{LC}} \text{ Hz} \tag{6-42}$$

en donde f_c = frecuencia de la portadora en reposo (hertz)
L = inductancia del devanado primario de T_1 (henries)
C = capacitancia del diodo varactor (faradios)

Con una señal moduladora aplicada, la frecuencia es

$$f = \frac{1}{2\pi\sqrt{L(C + \Delta C)}} \text{ Hz} \tag{6-43}$$

donde f es la nueva frecuencia de oscilación y ΔC es el cambio de capacitancia del diodo varactor debido a la señal moduladora. Ese cambio de frecuencia es

$$\Delta f = |f_c - f| \tag{6-44}$$

donde Δf = desviación máxima de frecuencia (hertz)

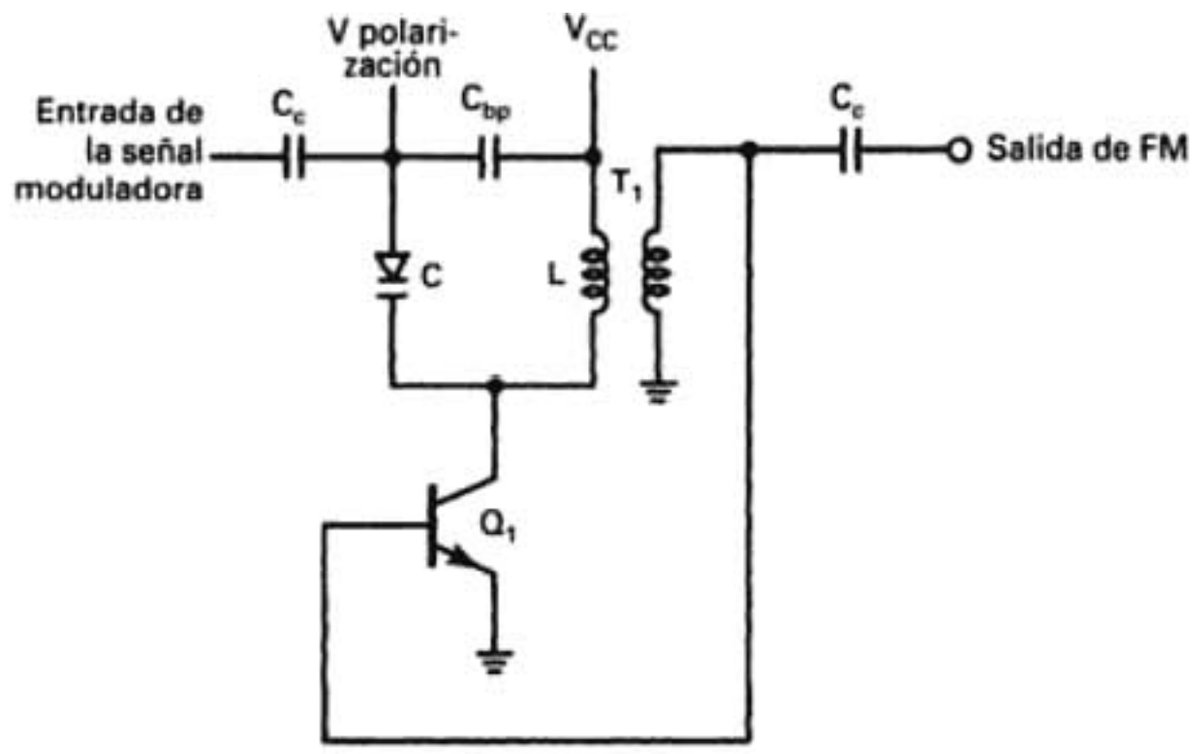


FIGURA 6-16 Modulador directo de FM con VCO y diodo varactor

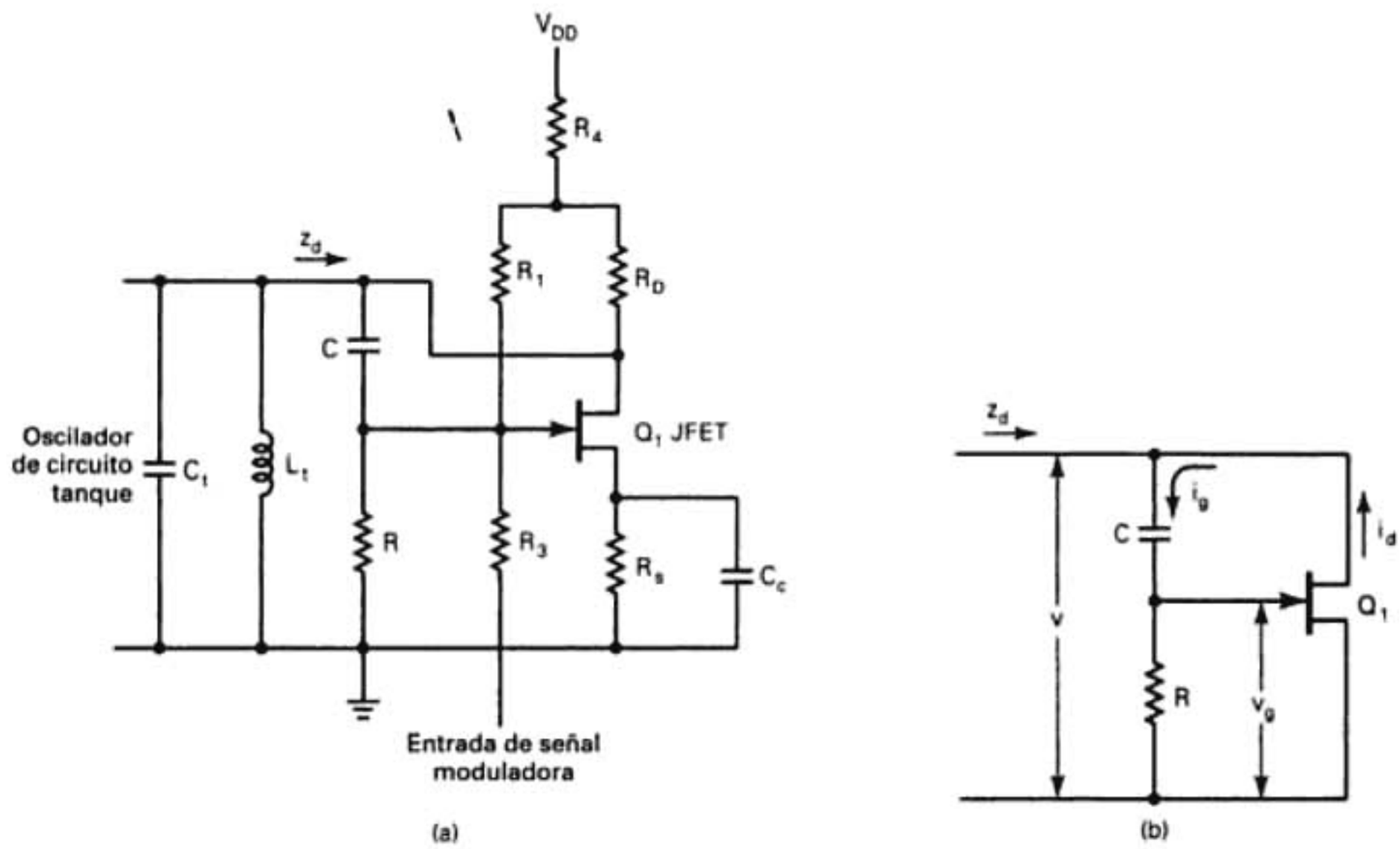


FIGURA 6-17 Modulador de reactancia y JFET: (a) diagrama eléctrico; (b) circuito equivalente de ca

Moduladores de FM de reactancia. La fig. 6-17a muestra el diagrama de un modulador de reactancia que usa un JFET (transistor de unión de efecto de campo, *junction field-effect transistor*) como dispositivo activo. Esta configuración de circuito se llama modulador de reactancia, porque el JFET se ve como una carga de reactancia variable desde el circuito tanque *LC*. La señal moduladora hace variar la reactancia de Q_1, lo que causa un cambio correspondiente en la frecuencia de resonancia del circuito tanque oscilador.

La fig. 6-17b muestra el circuito equivalente en ca: R_1, R_3, R_4 y R_D proporcionan la polarización cd de Q_1. La resistencia R_S se conecta en paralelo con C_c y en consecuencia se omite del circuito equivalente de ca. El funcionamiento del circuito es como sigue. Suponiendo un JFET ideal (corriente de compuerta $i_g = 0$),

$$v_g = i_g R$$

donde

$$i_g = \frac{v}{R - jX_C}$$

Por consiguiente, $$v_g = \frac{v}{R - jX_C} \times R$$

y la corriente de drenaje del JFET es

$$i_d = g_m v_g = g_m \left(\frac{v}{R - jX_C}\right) \times R$$

donde g_m es la transconductancia del JFET, y la impedancia entre el drenaje y tierra es

$$z_d = \frac{v}{i_d}$$

Sustituyendo y reordenando se llega a

$$z_d = \frac{R - jX_C}{g_m R} = \frac{1}{g_m}\left(1 - \frac{jX_C}{R}\right)$$

Suponiendo que $R <<< X_C$, $$z_d = -j\frac{X_C}{g_m R} = \frac{-j}{2\pi f_m g_m RC}$$

La expresión *anterior* equivale a una impedancia variable, y es inversamente proporcional a la resistencia R, a la velocidad angular de la señal moduladora, $2\pi f_m$, y a la transconductancia g_m de Q_1 que varía con el voltaje de compuerta a frente. Cuando se aplica una señal a la parte inferior de R_3, el voltaje de compuerta a frente también varía, provocando un cambio proporcional en g_m. Como resultado, la impedancia z_d del circuito equivalente es una función de la señal moduladora. En consecuencia, la frecuencia de resonancia del oscilador de circuito tanque es una función de la amplitud de la señal moduladora y la rapidez con que cambia es igual a f_m. Si se intercambian R y C se hace que la reactancia variable sea inductiva y no capacitiva, pero no afecta la forma de onda de salida de FM. La desviación máxima de frecuencia que se obtiene con un modulador de reactancia es más o menos 5 kHz.

Moduladores directos de FM en circuito integrado lineal. Los *osciladores controlados por voltaje en circuito integrado* lineal y los *generadores de función* pueden generar una forma de onda de salida de FM directa que es relativamente estable, exacta y directamente proporcional a la señal moduladora de entrada. La desventaja primaria de usar VCO y generadores de función en LIC (circuito integrado lineal, *linear integrated circuit*) para modulación directa de FM es su baja potencia de salida y la necesidad de otros componentes externos más para que funcionen, como por ejemplo capacitores de temporización y resistores para determinación de frecuencia, así como filtros para las fuentes de energía.

La fig. 6-18 muestra el diagrama simplificado de bloques de un generador de funciones monolítico, en circuito integrado lineal, que se puede usar para generación directa de FM. La frecuencia central del VCO se determina con el resistor y el capacitor externos (R y C). La señal moduladora de entrada, $v_m(t) = V_m \operatorname{sen}(2\pi f_m t)$, se aplica en forma directa a la entrada del oscilador controlado por voltaje, donde desvía la frecuencia en reposo f_c de la portadora y produce una señal de salida de FM. La desviación máxima de frecuencia se determina con el producto

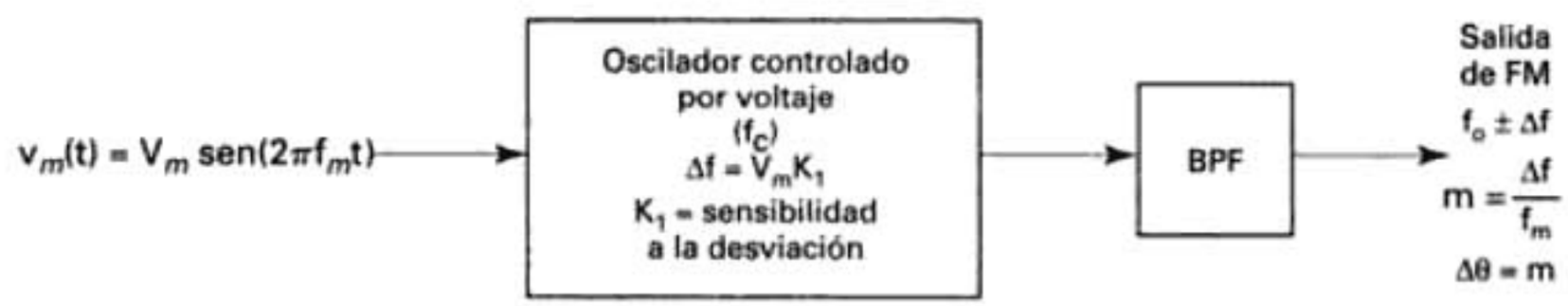


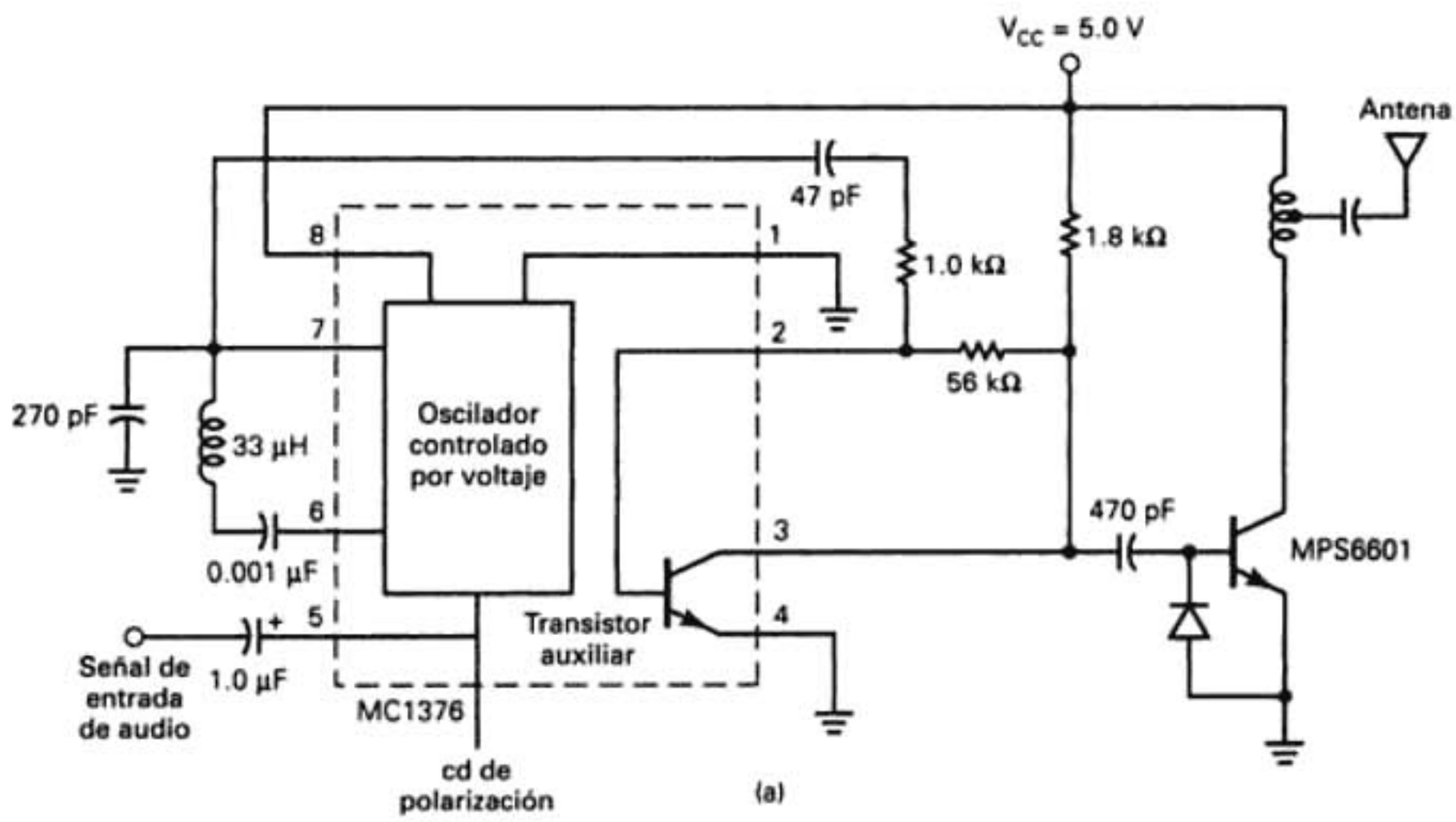
FIGURA 6-18 Modulador directo de FM en circuito integrado lineal; diagrama simplificado de bloques

de la amplitud máxima V_m de la señal moduladora, por K_1, la sensibilidad del VCO a la desviación. La salida del modulador es

$$FM_{(sal)} = f_c + \Delta f$$

donde f_c = frecuencia en reposo de la portadora (frecuencia natural del VCO − $1/RC$ hertz)
Δf = desviación máxima de frecuencia ($V_m K_1$ hertz)
K_1 = sensibilidad a la desviación, (hertz por volt)

La fig. 6-19 muestra el diagrama del transmisor monolítico de FM Motorola MC1376. Es un modulador completo de FM en un circuito integrado DIP con 8 terminales. Puede funcionar con frecuencias de portadora de 1.4 a 14 MHz, y es para usarse en la producción directa de ondas de FM para aplicaciones de baja potencia, como por ejemplo teléfonos inalámbricos. Cuando se conecta el transistor auxiliar a un voltaje de suministro de 12 V, se pueden alcanzar



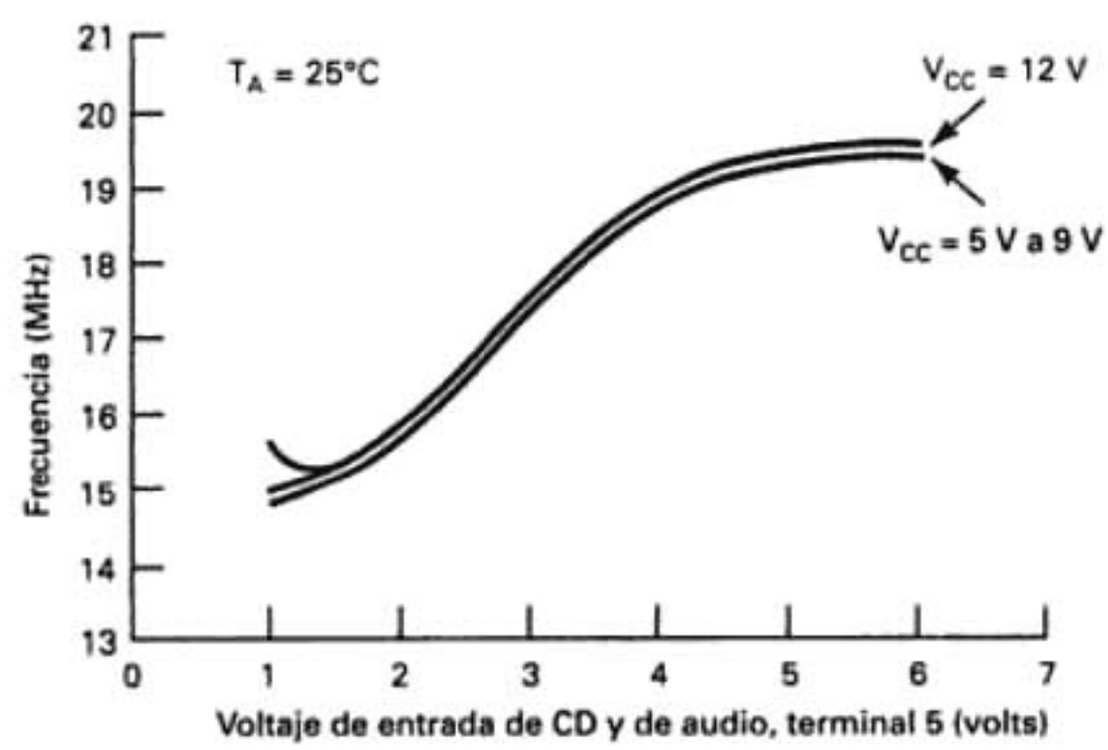


FIGURA 6-19 Transmisor de FM en circuito integrado lineal MC1376: (a) diagrama esquemático; (b) Curva de respuesta de salida de VCO en función de frecuencia de entrada

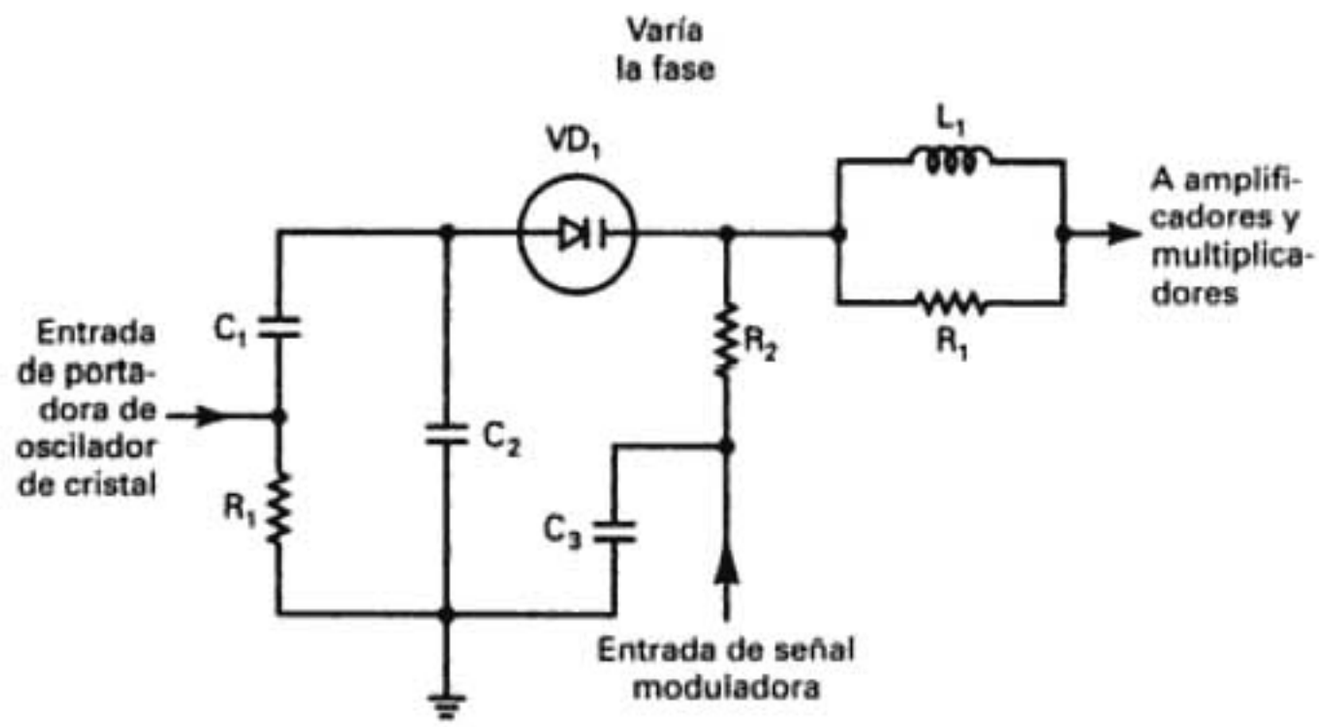


FIGURA 6-20 Esquema de un modulador directo de PM

potencias de salida hasta de 600 mW. La fig. 6-19b muestra la curva de frecuencia de salida en función del voltaje de entrada para el VCO interno. Como allí se ve, la curva es bastante lineal entre 2 y 4 V, y puede producir una desviación máxima de frecuencia casi de 150 kHz.

Moduladores directos de PM

Modulador directo de PM con diodo varactor. La modulación directa de PM (es decir, *indirecta de FM*) es una modulación angular en la que se desvía la frecuencia de la portadora en forma indirecta, mediante la señal moduladora. La PM directa se logra cambiando en forma directa la fase de la portadora y, en consecuencia, es una forma de modulación directa de fase. La fase instantánea de la portadora es directamente proporcional a la señal moduladora.

La fig. 6-20 muestra el diagrama de un modulador directo de PM. Comprende un diodo varactor VD_1 en serie con la red inductiva (la bobina sintonizable L_1 y el resistor R_1). La red serie-paralelo combinada aparece como un circuito resonante en serie a la frecuencia de salida del oscilador de cristal. Se aplica una señal moduladora a VD_1, que cambia su capacitancia y, en consecuencia, varía el ángulo de fase que ve la portadora, y eso resulta en un desplazamiento correspondiente de fase en la portadora. El desplazamiento de fase es directamente proporcional a la amplitud de la señal moduladora. Una ventaja de la FM indirecta es que se usa un oscilador aislado de cristal como fuente de la señal portadora. En consecuencia, los transmisores de FM indirecta tienen frecuencia más estable que sus contrapartes directos. Una de sus desventajas es que las características de capacitancia en función de voltaje, de un diodo varactor, son no lineales. De hecho, se parecen mucho a una función de raíz cuadrada. En consecuencia, para minimizar la distorsión en la forma de onda modulada, se debe mantener bastante pequeña la amplitud de la señal moduladora, y eso limita la desviación de fase a valores pequeños, y los usos a aplicaciones de bajo índice y banda angosta.

Modulador directo de PM con transistor. La fig. 6-21 muestra el diagrama de un modulador directo de PM con transistor. El circuito es un amplificador normal de clase A, de emisor común con dos entradas externas: una entrada de señal moduladora $v_m(t)$, y una entrada de portadora externa $v_c(t)$. Las condiciones de operación en reposo hacen que el transistor funcione como un resistor del punto x a tierra. La resistencia R_t del emisor al colector del transmisor es una parte del desplazador de fase formado por C_1 en serie con R_t y el resistor R_E del emisor. La salida se toma a través de la combinación de R_t y R_E en serie. Si el circuito se diseña de tal manera que, a la frecuencia de entrada de la portadora f_c, la suma de R_t y R_E sea igual a la reactancia capacitiva de C_1 (es decir, X_{C1}), la señal de entrada portadora se desplaza 45°.

Cuando se aplica una señal moduladora, su voltaje se suma y se resta a la polarización de la base en cd y produce los cambios correspondientes en la corriente del colector. Estos cambios en la corriente del colector hacen cambiar dinámicamente la resistencia del emisor al colector

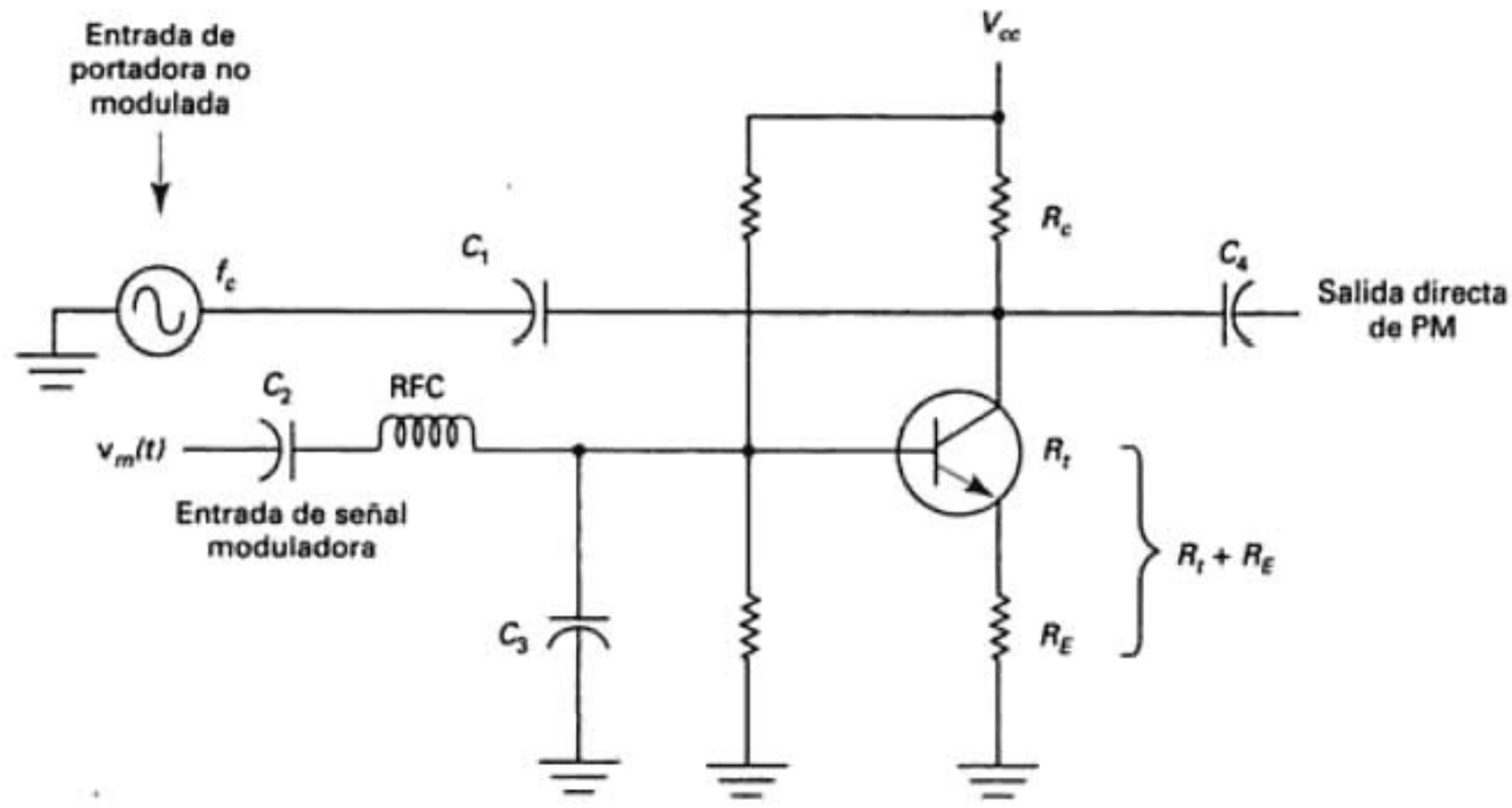


FIGURA 6-21 Modulador directo de PM con transistor

del transistor, y producen cambios en el desplazamiento de fase que sufre la portadora al pasar por la red de desplazamiento de fase. El desplazamiento de fase es proporcional a la amplitud de la señal moduladora y se presenta con una rapidez igual a la frecuencia de la señal moduladora. Mientras mayor sea la amplitud de entrada de la señal moduladora, el cambio de resistencia del emisor al colector será mayor, y el desplazamiento de fase será mayor.

Los desplazadores de fase con transistor pueden producir desplazamientos máximos de fase hasta de 0.375 radianes, es decir, un índice de modulación de 0.375. Con una frecuencia de señal moduladora de 15 kHz, un índice de modulación de 0.375 corresponde a un desplazamiento indirecto de frecuencia de $15{,}000 \times 0.375 = 5625$ Hz.

CONVERSIÓN ELEVADORA DE FRECUENCIA

Con los moduladores de FM y de PM, la frecuencia de portadora en la salida del modulador es, en general, algo mayor que la frecuencia deseada de transmisión. Por lo anterior, en los transmisores de FM y de PM suele ser necesario elevar la frecuencia de la portadora modulada, después de haber hecho la modulación. Hay dos métodos básicos de hacer la conversión elevadora de frecuencia. Uno es usar un proceso de heterodinado, y el segundo es mediante multiplicación de frecuencia. La fig. 6-22 muestra esos dos métodos.

Método de heterodinado para elevar la frecuencia

La fig. 6-22a muestra el método de heterodinado para elevar la frecuencia. En él, se aplica una portadora de frecuencia relativamente baja y modulación angular, junto con sus frecuencias laterales, a una entrada de un modulador balanceado. La segunda entrada es una señal portadora no modulada de RF, de alta frecuencia. En el modulador balanceado las dos entradas se mezclan en forma no lineal, y producen frecuencias de suma y de diferencia en su salida. También se producen frecuencias de suma y de diferencia entre las frecuencias laterales de la señal moduladora y de la portadora de RF. El filtro pasabanda (BPF) se sintoniza a la frecuencia de suma, con una banda de paso con ancho suficiente para pasar la portadora más las frecuencias laterales superior e inferior. Así, el BPF pasa la suma de las portadoras modulada y no modulada, mientras que se bloquean las frecuencias de diferencia. La salida del filtro pasabanda es

$$f_{c(\text{sal})} = f_{c(\text{ent})} + f_{\text{RF}}$$

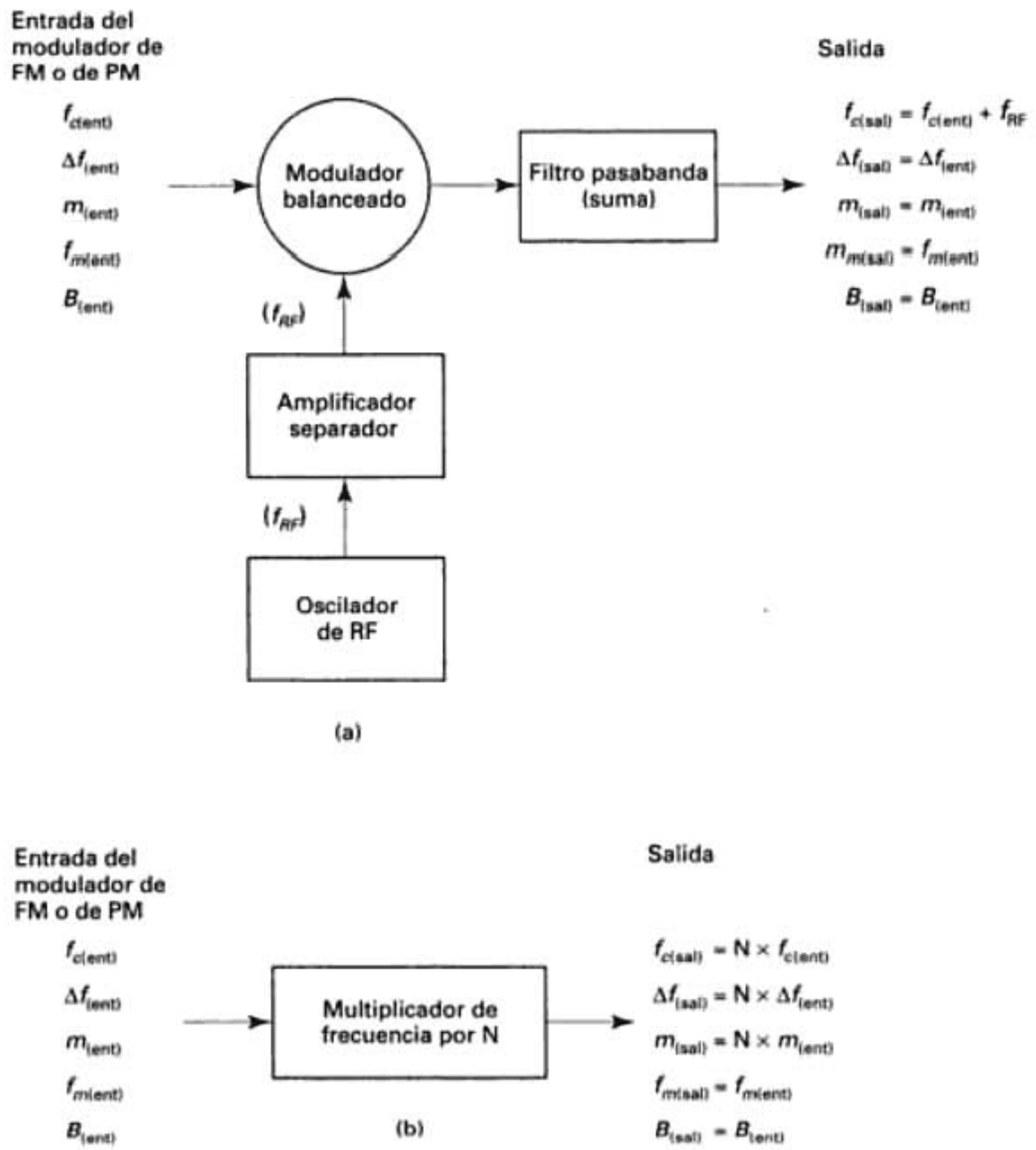


FIGURA 6-22 Conversión elevadora de frecuencia: (a) método de heterodinado; (b) método de multiplicación

en donde $f_{c(sal)}$ = señal modulada con frecuencia aumentada
$f_{c(ent)}$ = señal modulada de entrada
f_{RF} = portadora de RF

Como las frecuencias de la portadora modulada no se alteran por el proceso de heterodinado, tampoco se afecta la desviación de frecuencia. Así, la salida del filtro BPF contiene la desviación original de frecuencia (tanto su magnitud Δf como la rapidez de cambio f_m). El índice de modulación, la desviación de fase y el ancho de banda tampoco se afectan por el proceso de heterodinado. En consecuencia

$$\Delta f_{(sal)} = \Delta f_{(ent)}$$
$$m_{(sal)} = m_{(ent)}$$
$$\Delta\theta_{(sal)} = \Delta\theta_{(ent)}$$
$$B_{(sal)} = B_{(ent)}$$
$$f_{m(sal)} = f_{m(ent)}$$

Si el filtro pasabanda que está en la salida del modulador balanceado de la fig. 6-22a se sintonizara con la frecuencia de diferencia, sucedería la conversión reductora de frecuencias.

Como en la conversión elevadora, sólo se afecta la frecuencia de la portadora. No cambiarían el índice de modulación, la desviación de frecuencia, la desviación de fase, el ancho de banda ni la rapidez de cambio.

En esencia, con el método de heterodinado por aumento de frecuencia, una portadora modulada de baja frecuencia puede elevarse o disminuirse a un lugar distinto del espectro de frecuencias sin cambiar sus propiedades de modulación.

Método de multiplicación para aumento de frecuencia

La fig. 6-22b muestra el método de multiplicación para elevar la frecuencia. En él, se pueden aumentar las propiedades de modulación de una portadora a la vez que se eleva su frecuencia. Con el método de multiplicación, la frecuencia de la portadora modulada se multiplica por un factor N en el multiplicador. Además, también se multiplica la desviación de frecuencia. Sin embargo, la tasa de desviación no es afectada por el proceso de multiplicación. Por lo anterior, la frecuencia de la portadora a la salida, la desviación de frecuencia, el índice de modulación y la desviación de fase en la salida del multiplicador de frecuencia son

$$f_{c(\text{sal})} = \text{N}f_{c(\text{ent})}$$
$$\Delta f_{(\text{sal})} = \text{N}\Delta f_{(\text{ent})}$$
$$m_{(\text{sal})} = \text{N}m_{(\text{ent})}$$
$$\Delta\theta_{(\text{sal})} = \text{N}\Delta\theta_{(\text{ent})}$$
$$f_{m(\text{sal})} = f_{m(\text{ent})}$$

donde N es el factor de multiplicación.

Como la desviación de frecuencia y el índice de modulación se multiplican en esta clase de multiplicadores, la cantidad de frecuencias laterales también aumenta, y se produce un aumento correspondiente de ancho de banda. Para índices de modulación mayores que 10, es decir, con modulación de alto índice, se puede aplicar la regla de Carson y el ancho de banda aproximado a la salida del multiplicador es

$$\begin{aligned} B(\text{sal}) &= \text{N}(2\Delta f) \\ &= \text{N}B_{(\text{ent})} \end{aligned}$$

Sin embargo, es importante notar que la desviación de frecuencia se presenta en la frecuencia de la señal moduladora, f_m, que permanece inalterada por el proceso de multiplicación. En consecuencia, la separación entre las frecuencias laterales adyacentes permanece sin cambio, es decir, $\pm f_m$, $\pm 2f_m$, $\pm 3f_m$, etcétera.

En las siguientes secciones de este libro se verá que a veces conviene usar el método de heterodinado para elevar la frecuencia en los transmisores de FM y PM, a veces conviene usar el método de multiplicación y a veces se usan las dos técnicas en el mismo transmisor.

Ejemplo 6-7

Para la frecuencia de portadora y las propiedades de modulación mencionadas, determinar la frecuencia de portadora y propiedades de modulación a la salida de:

(a) Un modulador balanceado con un filtro pasabanda sintonizado a la frecuencia de suma y una RF de entrada de 99.5 MHz.

(b) Un multiplicador de frecuencia por 10.

$$\Delta f = 3 \text{ kHz},\ f_m = 10 \text{ kHz},\ m = 0.3, \text{ y } f_c = 500 \text{ kHz}$$

Solución (a) La frecuencia de salida de portadora no es más que la suma de la portadora modulada de entrada y la portadora de RF

$$\begin{aligned} f_{c(\text{sal})} &= 0.5 \text{ MHz} + 99.5 \text{ MHz} \\ &= 100 \text{ MHz} \end{aligned}$$

Las propiedades de modulación en la salida son idénticas a las de la entrada

$$\Delta f(\text{ent}) = 3 \text{ kHz},\ f_{m(\text{ent})} = 10 \text{ kHz},\ m = 0.3, \text{ y } f_{c(\text{ent})} = 500 \text{ kHz}$$

(b) La frecuencia de salida de portadora, la desviación de frecuencia, el índice de modulación y la frecuencia moduladora son simplemente

$$f_{c(\text{sal})} = 10(500\text{ kHz}) = 5\text{ MHz}$$
$$\Delta f_{(\text{sal})} = 10(3\text{ kHz}) = 30\text{ kHz}$$
$$m_{(\text{sal})} = 10(0.3) = 3$$
$$f_{m(\text{sal})} = 10\text{ kHz}$$

TRANSMISORES DIRECTOS DE FM

Los transmisores directos de FM producen una forma de onda de salida en la que la desviación de frecuencia es directamente proporcional a la señal moduladora. En consecuencia, el oscilador de portadora se debe desviar en forma directa. Así, para los sistemas de FM de índice intermedio y alto, el oscilador no puede ser de cristal, porque la frecuencia a la que oscila un cristal no se puede variar mucho. Como resultado, la estabilidad de los osciladores en los transmisores directos de FM no cumplen, a menudo, con las especificaciones de la FCC. Para resolver este problema se usa el *control automático de frecuencia* (AFC, de *automatic frequency control*). Un circuito de AFC compara la frecuencia del oscilador de la portadora (no de cristal) con un oscilador de cristal de referencia, y produce un voltaje de corrección proporcional a la diferencia entre las dos frecuencias. El voltaje de corrección se retroalimenta al oscilador de la portadora para compensar, en forma automática, cualquier error que se pueda presentar.

Transmisor directo de FM de Crosby

La fig. 6-23 muestra el diagrama de bloques de un transmisor de banda comercial. Esta configuración en especial se llama *transmisor directo de FM de Crosby*, e incluye un *lazo de AFC*. El modulador de frecuencia puede ser uno de reactancia o un oscilador controlado por voltaje. La frecuencia de reposo de la portadora es la frecuencia no modulada de salida del oscilador maestro, f_c. Para el transmisor de la fig. 6-23, la frecuencia central del oscilador maestro $f_c = 5.1$ MHz, que se multiplica por 18 en tres pasos ($3 \times 2 \times 3$) para producir una frecuencia final de portadora de transmisión $f_t = 91.8$ MHz. Aquí se deben hacer notar tres aspectos de la conversión de frecuencia. Primero, que cuando la frecuencia de una portadora de frecuencia modulada se multiplica, también se multiplican sus desviaciones de frecuencia y de fase. Segundo, que la rapidez con la que se desvía la portadora (es decir, la frecuencia f_m de la señal

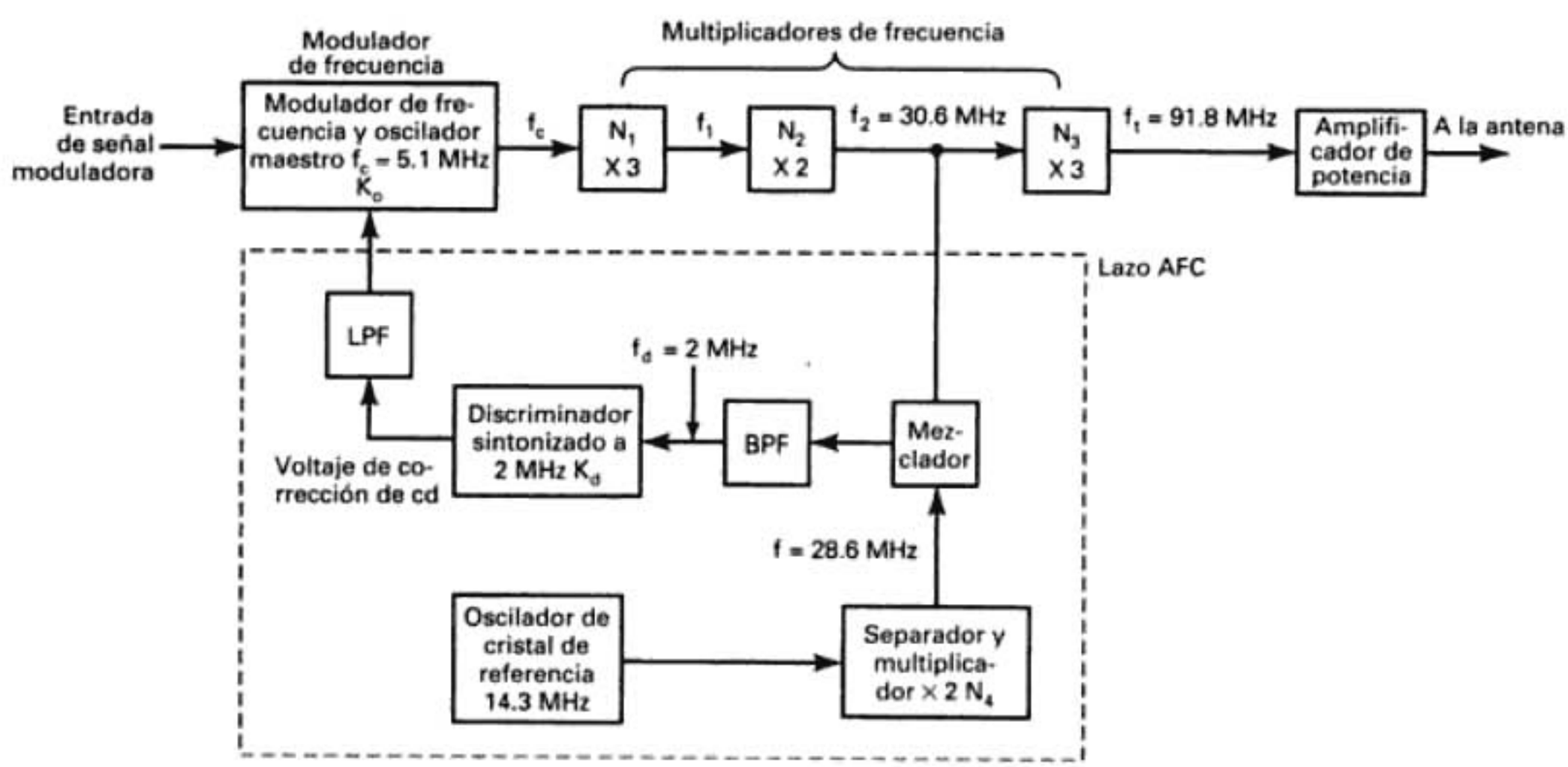


FIGURA 6-23 Transmisor directo de FM de Crosby

moduladora) no se afecta en el proceso de multiplicación. Por consiguiente, también se multiplica el índice de modulación. Tercero, que cuando se heterodina una portadora con otra frecuencia en un mezclador no lineal, la portadora puede tener una conversión elevadora o reductora, dependiendo de la banda de paso del filtro de salida. Sin embargo, el proceso de heterodinado no afecta a la desviación de frecuencia, desviación de fase ni a la rapidez de cambio. Por consiguiente, para el transmisor de la fig. 6-23, las desviaciones de frecuencia y de fase en la salida del modulador también están multiplicadas por 18. Para lograr la desviación máxima de frecuencia permitida para las estaciones de emisión en FM en la antena (75 kHz), la desviación en la salida del modulador debe ser

$$\Delta f = \frac{75 \text{ kHz}}{18} = 4166.7 \text{ Hz}$$

y el índice de modulación debe ser

$$m = \frac{4166.7 \text{ Hz}}{f_m}$$

Para la frecuencia máxima permitida de señal moduladora, $f_m = 15$ kHz,

$$m = \frac{4166.7 \text{ Hz}}{15{,}000 \text{ Hz}} = 0.2778$$

Así, el índice de modulación en la antena es

$$m = 0.2778(18) = 5$$

que es la relación de desviación para los transmisores comerciales de FM con una señal moduladora de 15 kHz.

Lazo AFC. El propósito del *lazo de AFC* es lograr una estabilidad de la frecuencia de portadora de transmisión parecida a la de un oscilador de cristal, sin usar el cristal. Con el AFC, la señal de portadora se mezcla con la señal de salida de un oscilador de referencia de cristal en un dispositivo no lineal, se reduce su frecuencia y se retroalimenta a la entrada de un *discriminador de frecuencias*. Éste es un dispositivo selectivo de frecuencia cuyo voltaje de salida es proporcional a la diferencia entre la frecuencia de entrada y su frecuencia de resonancia. Su funcionamiento se explica en el capítulo 7. Para el transmisor de la fig. 6-23, la salida del doblador $f_2 = 30.6$ MHz, que se mezcla con una frecuencia de referencia controlada por cristal, $f_r = 28.6$ MHz, para producir una frecuencia de diferencia $f_d = 2$ MHz. El discriminador es un circuito sintonizado de Q relativamente alta (de banda angosta) que reacciona sólo a frecuencias cercanas a su frecuencia central, en este caso 2 MHz. Por consiguiente, el discriminador responde a cambios de frecuencia bajos y de largo plazo en la frecuencia central de portadora, debidos a errores en la frecuencia del oscilador maestro, y porque el filtrado pasabajas no responde a la desviación producida por la señal moduladora. Si el discriminador respondiera a la desviación de frecuencia, el lazo de retroalimentación anularía la desviación y así eliminaría la modulación de la onda de FM; a esto se le llama *borrado*. El voltaje de corrección se suma a la señal moduladora para ajustar en forma automática la frecuencia central del oscilador maestro, y así compensar el desplazamiento de baja frecuencia.

Ejemplo 6-8

Usar el modelo de transmisor de la fig. 6-23 para contestar lo siguiente. Para una multiplicación total de frecuencia de 20 y una frecuencia de portadora de transmisión $f_t = 88.8$ MHz, determinar:

(a) Frecuencia central del oscilador maestro.

(b) Desviación de frecuencia a la salida del modulador, para una desviación de frecuencia de 75 kHz en la antena.

(c) La relación de desviación en la salida del modulador para una frecuencia máxima de señal moduladora $f_m = 15$ kHz.

(d) Relación de desviación en la antena.

Solución (a) $$f_c = \frac{f_t}{N_1N_2N_3} = \frac{88.8\text{ MHz}}{20} = 4.43\text{ MHz}$$

(b) $$\Delta f = \frac{\Delta f_t}{N_1N_2N_3} = \frac{75\text{ kHz}}{20} = 3750\text{ Hz}$$

(c) $$\text{DR} = \frac{\Delta f_{(\text{máx})}}{f_{m(\text{máx})}} = \frac{3750\text{ Hz}}{15\text{ kHz}} = 0.25$$

(d) $$\text{DR} = 0.25 \times 20 = 5$$

Control automático de frecuencia. Como el transmisor de Crosby usa un VCO o un oscilador de reactancia o un oscilador de circuito integrado lineal para generar la frecuencia de la portadora, es más susceptible a errores de frecuencia debidos a cambios de temperatura, fluctuaciones en la fuente de energía, etc., en comparación con un oscilador de cristal. Como se dijo en el capítulo 2, la estabilidad de un oscilador se suele especificar en partes por millón (ppm) por grado Celsius. Por ejemplo, para el transmisor de la fig. 6-23, una estabilidad de oscilador de ±40 ppm podría producir ±204 Hz (5.1 MHz × ±40 Hz/millón) de error de frecuencia por grado Celsius a la salida del oscilador maestro. Eso correspondería a un error de ±3672 Hz en la antena (18 × ±204), que se sale mucho del máximo de ±2 kHz establecido por la FCC para emisiones comerciales de FM. Aunque un circuito AFC no elimina en su totalidad el error de frecuencia, lo puede reducir bastante. Si se supone un oscilador de referencia de cristal, estable, y un discriminador perfectamente sintonizado, el error de frecuencia a la salida del segundo multiplicador sin retroalimentación, es decir, con lazo abierto, es

$$\text{error con lazo abierto} = df_{o1} = N_1N_2df_c \tag{6-45}$$

donde d representa el error. El error con lazo cerrado es

$$\text{error con lazo cerrado} = df_{c1} = df_{o1} - N_1N_2k_dk_odf_{c1} \tag{6-46}$$

Por consiguiente, $$df_{c1} + N_1N_2k_dk_odf_{c1} = df_{o1}$$

y $$df_{c1}(1 + N_1N_2k_dk_o) = df_{o1}$$

Así, $$df_{c1} = \frac{df_{o1}}{1 + N_1N_2k_dk_o} \tag{6-47}$$

donde k_d = función de transferencia del discriminador (volts por hertz)
k_o = función de transferencia del oscilador maestro (hertz por volt)

Se puede ver en esta ecuación que el error de frecuencia a la salida del segundo multiplicador y, en consecuencia, en la entrada del discriminador, se reduce por un factor de $1 + N_1N_2k_dk_o$ cuando se cierra el lazo de AFC. El error de frecuencia de la portadora se multiplica por la ganancia del lazo AFC y se retroalimenta al oscilador maestro, como voltaje de corrección. El error total de frecuencia no se puede anular, porque entonces no habría voltaje de error en la salida del discriminador, para retroalimentar al oscilador maestro. Además, las ecuaciones 6-45, 6-46 y 6-47 se dedujeron suponiendo que el discriminador y el oscilador de referencia de cristal son perfectamente estables, lo que naturalmente no es cierto.

Ejemplo 6-9

Usar el diagrama de bloques del transmisor y los valores de la fig. 6-23, para contestar lo siguiente: Determinar la reducción de error de frecuencia en la antena, para un transmisor sin AFC, comparado con uno con AFC. Usar estabilidad de VCO de +200 ppm, k_o = 10 kHz/V y k_d = 2 V/kHz.

Solución Estando abierto el lazo de retroalimentación, la frecuencia de salida del oscilador maestro es

$$f_c = 5.1\text{ MHz} + (200\text{ ppm} \times 5.1\text{ MHz}) = 5{,}101{,}020\text{ Hz}$$

y la frecuencia en la salida del segundo multiplicador es

$$f_2 = N_1N_2f_c = (5{,}101{,}020)(6) = 30{,}606{,}120\text{ Hz}$$

Así, el error de frecuencia es

$$df_2 = 30{,}606{,}120 - 30{,}600{,}000 = 6120 \text{ Hz}$$

Por consiguiente, la frecuencia de transmisión en la antena es

$$f_t = 30{,}606{,}120(3) = 91.81836 \text{ MHz}$$

18.36 kHz arriba de la frecuencia asignada, y bastante fuera de límites.

Con el lazo de retroalimentación cerrado, el error de frecuencia en la salida del segundo multiplicador se reduce en un factor de $1 + N_1N_2k_dk_o$, es decir,

$$1 + \frac{(2)(3)(10 \text{ kHz})}{\text{V}} \frac{2\text{V}}{\text{kHz}} = 121$$

Por consiguiente,

$$df_2 = \frac{6120}{121} = 51 \text{ Hz}$$

Así,

$$f_2 = 30{,}600{,}051 \text{ Hz}$$

y la frecuencia de transmisión en la antena es

$$f_t = 30{,}600{,}051 \times 3 = 91{,}800{,}153 \text{ Hz}$$

El error de frecuencia en la antena se redujo de 18,360 a 153 Hz, que ahora está dentro de los requisitos de ±2 kHz por parte de la FCC.

En la descripción anterior y en el ejemplo 6-9 se supuso un oscilador de cristal de referencia con estabilidad perfecta, y un discriminador con sintonía perfecta. En realidad tanto el discriminador como el oscilador de referencia son susceptibles al error, y el peor de los casos es cuando se desplazan en la misma dirección que el oscilador maestro. Las características de error de un discriminador normal son del orden de ±100 ppm. Quizá se pueda ver ahora por qué la frecuencia de salida del segundo multiplicador se mezcló y redujo a una frecuencia relativamente baja, antes de alimentarla al discriminador. Para un discriminador sintonizado a 2 MHz con una estabilidad de ±200 ppm, su error máximo es

$$df_d = \pm 100 \text{ ppm} \times 2 \text{ MHz} = \pm 200 \text{ Hz}$$

Si se alimentara la señal de 30.6 MHz en forma directa al discriminador, el error máximo hubiera sido

$$df_d = \pm 100 \text{ ppm} \times 30.6 \text{ MHz} = \pm 3060 \text{ Hz}$$

El error de frecuencia debido a la inestabilidad del discriminador se multiplica por la ganancia del AFC de lazo abierto. En consecuencia, el cambio en la frecuencia de salida del segundo multiplicador, debido al error del discriminador, es

$$df_2 = df_dN_1N_2k_dk_o \tag{6-48}$$

De igual forma, el oscilador de referencia de cristal se puede desplazar y contribuir también al error total de frecuencia en la salida del segundo multiplicador. El error debido a la inestabilidad del cristal se multiplica por 2 antes de entrar al mezclador no lineal; entonces

$$df_2 = N_4df_oN_1N_2k_dk_o \tag{6-49}$$

y el error máximo de frecuencia con lazo abierto a la salida del segundo multiplicador es

$$df_{(\text{total})} = N_1N_2(df_c + k_ok_df_d + k_ok_dN_4df_o) \tag{6-50}$$

Transmisor directo de FM con lazo de fase cerrada

La fig. 6-24 muestra un transmisor de FM de *banda amplia* que usa un lazo de fase cerrada para lograr una estabilidad como de cristal con un oscilador maestro de VCO y, al mismo tiempo, generar una señal de salida de FM de banda ancha e índice alto. La frecuencia de salida del VCO se divide entre N y se retroalimenta al comparador de fase PLL (PLL = lazo de fase cerrada), donde se compara con una frecuencia estable de referencia de un cristal. El comparador de fases genera un voltaje de corrección que es proporcional a la diferencia entre las dos frecuencias. Este voltaje de corrección se suma a la señal moduladora y se aplica a la entrada del VCO. El

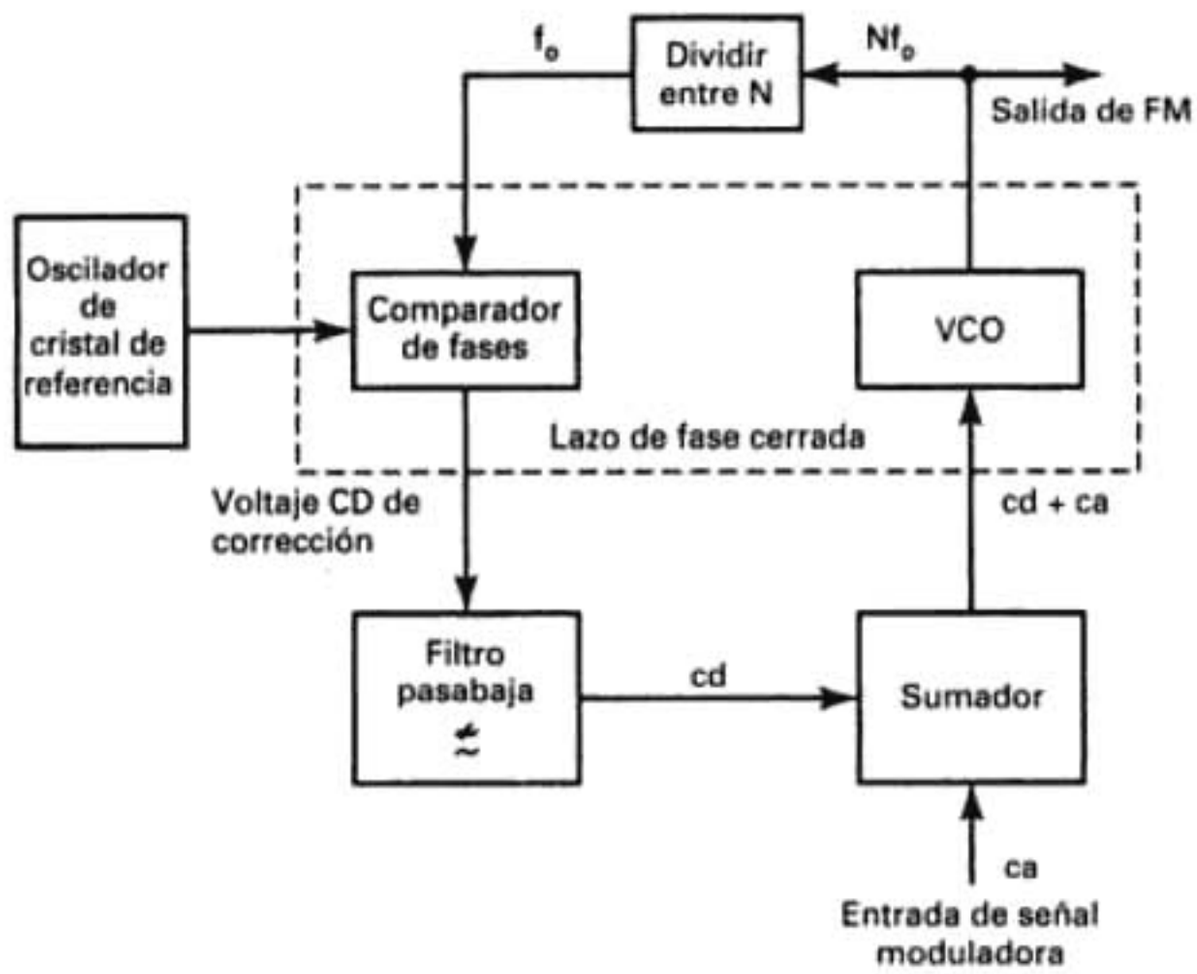


FIGURA 6-24 Transmisor de FM con lazo de fase cerrada

voltaje de corrección ajusta la frecuencia central a su valor correcto. De nuevo, el filtro pasabajas evita que los cambios en la frecuencia de salida del VCO, debidos a la señal moduladora, se conviertan en un voltaje, se retroalimenten al VCO y borren la modulación. El filtro pasabajas también evita que se cierre el lazo a una frecuencia lateral.

PM a partir de FM. Un modulador de FM precedido por un diferenciador genera una forma de onda PM. Si a los transmisores de las figuras 6-23 y 6-24 les precede una red de preénfasis, que es un diferenciador (filtro pasaaltas), se presenta un caso interesante. Para una constante de tiempo de 75 μs, se enfatiza la amplitud de frecuencias mayores que 2.12 kHz, por diferenciación. Así, para frecuencias moduladoras inferiores a 2.12 kHz, la forma de onda de salida es proporcional a la señal moduladora, y para frecuencias mayores que 2.12 kHz, la forma de onda de salida es proporcional a la derivada de la señal de entrada. En otras palabras, hay modulación de frecuencias menores que 2.12 kHz, y modulación de fase para frecuencias mayores que 2.12 kHz. Como la ganancia de un diferenciador aumenta cuando la frecuencia es mayor que la de corte (2.12 kHz), y como la desviación de frecuencia es proporcional a la amplitud de la señal moduladora, la desviación de frecuencia también aumenta cuando las frecuencias son mayores que 2.12 kHz. Se puede ver en la ecuación 6-18 que si Δf y f_m aumentan en forma proporcional, el índice de modulación permanece constante, lo que es característico de la modulación de fase.

TRANSMISORES INDIRECTOS DE FM

Los transmisores indirectos de FM producen una forma de onda de salida en la que la desviación de fase es proporcional a la señal moduladora. En consecuencia, el oscilador de portadora no se desvía en forma directa. Así, el oscilador de portadora puede ser un cristal, porque el oscilador mismo no es el modulador. Como resultado, la estabilidad de los osciladores con transmisores indirectos de FM puede cumplir con las especificaciones de la FCC sin usar un circuito AFC.

Transmisor indirecto de FM de Armstrong

Con la FM indirecta, la señal moduladora desvía en forma directa la fase de la portadora, y ésta, a su vez, cambia la frecuencia en forma indirecta. La fig. 6-25 muestra el diagrama de bloques de

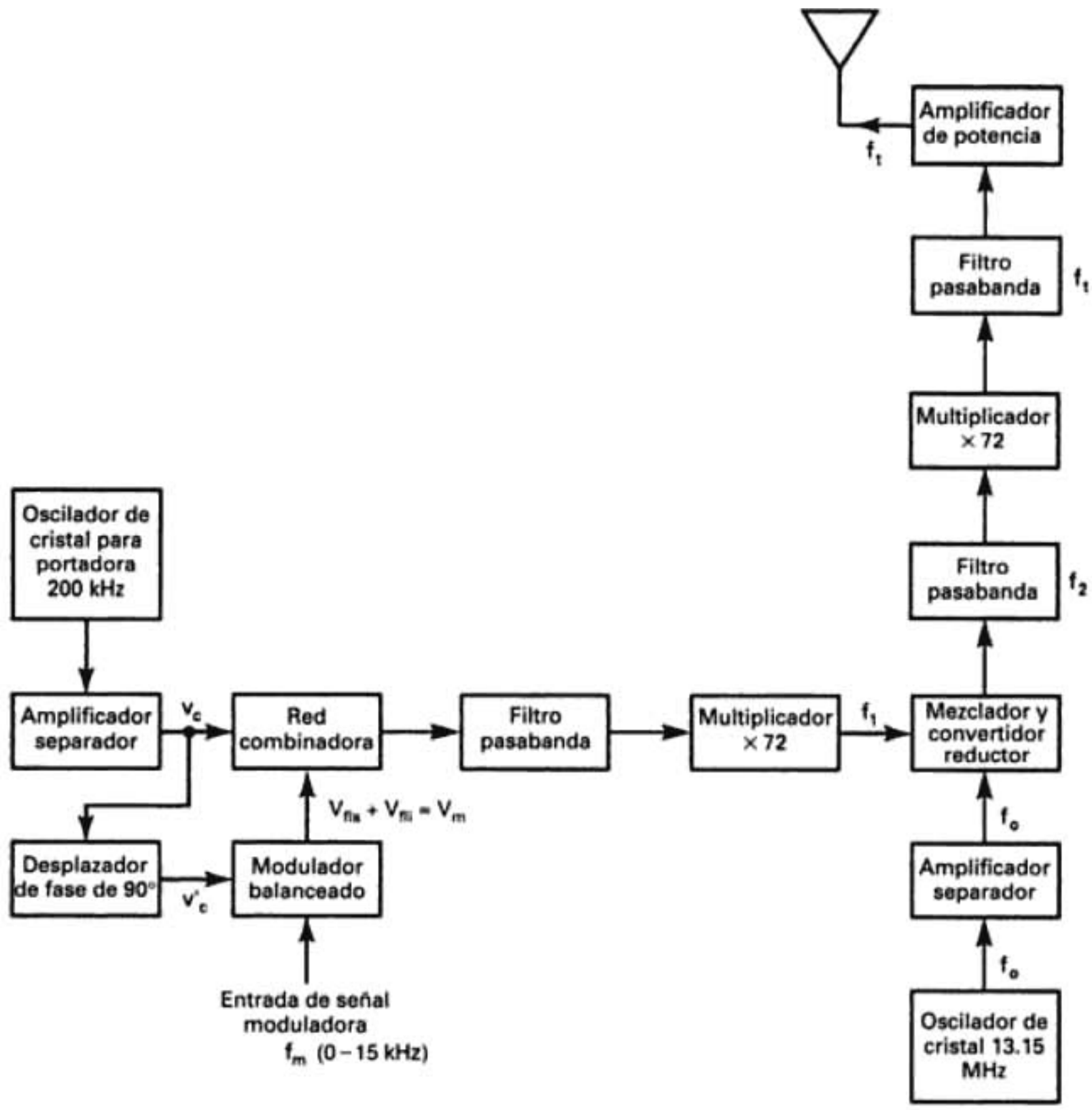


FIGURA 6-25 Transmisor indirecto de FM de Armstrong

un *transmisor indirecto de FM de Armstrong*. La fuente de la portadora es un cristal y en consecuencia se pueden llenar los requisitos de estabilidad de frecuencia de portadora establecidos por la FCC, sin usar un lazo AFC.

En un transmisor de Armstrong, una subportadora de frecuencia f_c relativamente baja se desplaza 90° en fase (f'_c) y se alimenta a un modulador balanceado, donde se mezcla con la señal moduladora de entrada, f_m. La salida del modulador balanceado es una onda de doble banda lateral y portadora suprimida, que se combina con la portadora original en una red combinadora y se produce una forma de onda de bajo índice y fase modulada. La fig. 6-26a muestra el fasor de la portadora original, V_c, y la fig. 6-26b muestra los fasores de los componentes de frecuencia lateral de la onda portadora suprimida, V_{fls} y V_{fli}. Como el voltaje de la portadora suprimida V'_c está desfasado 90° respecto a V_c, se combinan las bandas laterales superior e inferior y producen un componente V_m que está siempre en cuadratura (en ángulo recto) con V_c. Las figs. 6-26c a f, muestran la suma fasorial progresiva de V_c, V_{fls} y V_{fli}. Se puede ver que la salida de la red combinadora es una señal cuya fase se varía con una rapidez igual a f_m, y cuya magnitud es directamente proporcional a la magnitud de V_m. En la fig. 6-26 se puede ver que la desviación máxima de fase, o índice de modulación, se puede calcular como sigue:

$$\theta = m = \arctan \frac{V_m}{V_c} \tag{6-51a}$$

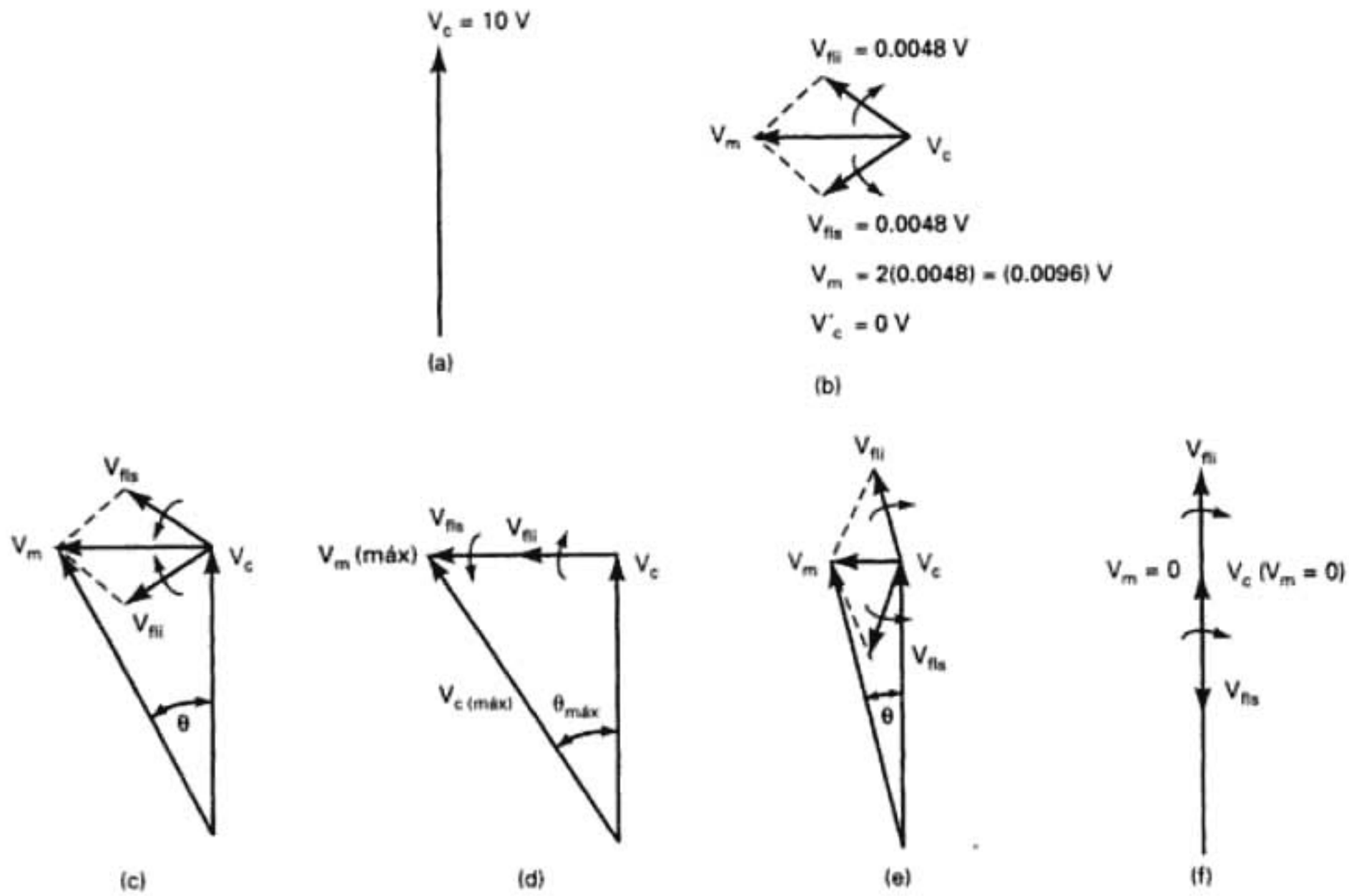


FIGURA 6-26 Suma fasorial de V_c, V_{fls} y V_{fli}: (a) fasor de portadora; (b) fasores de banda lateral; (c) a (f) suma fasorial progresiva. La parte (d) muestra el desplazamiento máximo de fase

Para ángulos muy pequeños, la tangente de un ángulo es aproximadamente igual al ángulo mismo y, por consiguiente,

$$\theta = m = \frac{V_m}{V_c} \tag{6-51b}$$

Ejemplo 6-10

Para el transmisor de Armstrong de la fig. 6-25 y la portadora de fase desplazada V'_c, y los componentes de frecuencia lateral superior V_{fls} y frecuencia lateral inferior V_{fli}, que se ven en la fig. 6-26, determinar:

(a) El desplazamiento máximo de fase, tanto en radianes como en grados.

(b) La desviación de frecuencia cuando la frecuencia de la señal moduladora f_m = 15 kHz.

Solución (a) La amplitud máxima del componente modulador es

$$V_m = V_{fls} + V_{fli}$$
$$= 0.0048 + 0.0048 = 0.0096$$

La desviación máxima de fase es el índice de modulación, y se calcula sustituyendo en la ecuación 6-51a

$$\theta = m = \arctan \frac{0.0096}{10} = 0.055°$$
$$= 0.055° \times \frac{\pi \text{ rad}}{180°} = 0.00096 \text{ rad}$$

(b) Se reordena la ecuación 6-18 para obtener

$$\Delta f = mf_m = (0.00096)(15 \text{ kHz}) = 14.4 \text{ Hz}$$

Se ve, en los diagramas fasoriales de la fig. 6-26, que se varía la amplitud de la portadora, y eso produce modulación no deseada de amplitud en la forma de onda de salida, y que $V_{c(máx)}$ se presenta cuando V_{fls} y V_{fli} están en fase entre sí y con V_c. La desviación máxima de

fase que se puede producir con este tipo de modulador es, más o menos, 16.7 milirradianes. Así, según la ecuación 6-18 y con frecuencia máxima de señal moduladora de $f_{m(\text{máx})} = 15$ kHz, la desviación máxima posible de frecuencia es

$$\Delta f_{\text{máx}} = (0.00167)(15{,}000) = 25 \text{ Hz}$$

Es evidente, de acuerdo con lo anterior, que el índice de modulación en la salida de la red combinadora es insuficiente para producir un espectro amplio de frecuencias de FM y que, por consiguiente, se debe multiplicar bastante antes de transmitir. Para el transmisor de la fig. 6-25, una subportadora de fase modulada de 200 kHz con desviación máxima de frecuencia $m = 0.00096$ rad sólo produce una desviación de frecuencia de 14.4 Hz a la salida de la red combinadora. Para lograr una desviación de frecuencia de 75 kHz en la antena, se debe multiplicar la frecuencia por 5208, aproximadamente. Sin embargo, así se obtendría una frecuencia de portadora de transmisión en la antena igual a

$$f_t = 5208 \times 200 \text{ kHz} = 1041.6 \text{ MHz}$$

muy afuera de los límites de frecuencia para la banda comercial de FM. Se ve que la multiplicación es inadecuada en sí. En consecuencia, es necesaria una combinación de multiplicación y mezclado, para desarrollar la frecuencia deseada de portadora de transmisión con una desviación de frecuencia de portadora igual a 75 kHz. La forma de onda en la salida de la red combinadora se multiplica por 72 y produce la señal siguiente

$$f_1 = 72 \times 200 \text{ kHz} = 14.4 \text{ MHz}$$
$$m = 72 \times 0.00096 = 0.06912 \text{ rad}$$
$$\Delta f = 72 \times 14.4 \text{ Hz} = 1036.8 \text{ Hz}$$

La salida del primer multiplicador se mezcla con una frecuencia f_o controlada por cristal de 13.15 MHz, para producir una señal de diferencia f_2 con las siguientes características

$$f_2 = 14.4 \text{ MHz} - 13.15 \text{ MHz} = 1.25 \text{ MHz (reducida)}$$
$$m = 0.6912 \text{ rad (no cambia)}$$
$$\Delta f = 1036.8 \text{ Hz (no cambia)}$$

Nótese que sólo se afecta la frecuencia de portadora con el proceso de heterodinado. La salida del mezclador se vuelve a multiplicar por 72 para producir una señal de transmisión con las siguientes características

$$f_t = 1.25 \text{ MHz} \times 72 = 90 \text{ MHz}$$
$$m = 0.06912 \times 72 = 4.98 \text{ rad}$$
$$\Delta f = 1036.8 \times 72 = 74{,}650 \text{ Hz}$$

En el ejemplo anterior, con el uso de los procesos de multiplicación y de heterodinado, se aumentó la portadora en un factor de 450; al mismo tiempo, se aumentaron la desviación de frecuencia y el índice de modulación en un factor de 5184.

Con el transmisor de Armstrong, la fase de la portadora se modula en forma directa en la red combinadora mediante suma, y se produce modulación indirecta de frecuencia. La magnitud de la desviación de fase es directamente proporcional a la amplitud de la señal moduladora, pero es independiente de su frecuencia. Por consiguiente, el índice de modulación permanece constante para todas las frecuencias de señal moduladora de determinada amplitud. Por ejemplo, para el transmisor de la fig. 6-26, si se mantiene constante la amplitud de la señal moduladora y se disminuye al mismo tiempo su frecuencia a 5 kHz, el índice de modulación permanece en 5, mientras que la desviación de frecuencia se reduce a $\Delta f = 5 \times 5000 = 25{,}000$ Hz.

FM a partir de PM. Un modulador de PM precedido por un integrador produce una forma de onda de FM. Si el transmisor de PM de la fig. 6-25 está precedido por un filtro de paso bajo (que es un integrador), se obtiene la FM. El filtro pasabajas simplemente es un filtro $1/f$, que se suele llamar *red de predistorsión*, o *de corrección de frecuencia*.

Comparación de FM y PM

Desde un punto de vista puramente teórico, la diferencia entre la FM y la PM es bastante sencilla: el índice de modulación para FM se define en forma distinta que para PM. En la PM, el índice de modulación es directamente proporcional a la amplitud de la señal moduladora, y es independiente de su frecuencia. En la FM, el índice de modulación es directamente proporcional a la amplitud de la señal moduladora y es inversamente proporcional a su frecuencia.

Si se considera la FM como una forma de modulación de fase, mientras mayor sea la desviación de frecuencia, mayor será la desviación de fase. Por consiguiente, la última depende, al menos hasta cierto grado, de la amplitud de la señal de la moduladora, del mismo modo que en la PM. En ésta, el índice de modulación es proporcional sólo a la amplitud del voltaje de la señal moduladora, mientras que en la FM también el índice de modulación es inversamente proporcional a la frecuencia de la señal moduladora. Si se reciben transmisiones de FM con un receptor de PM, las frecuencias bajas tendrían bastante más desviación de fase que la que les comunicaría un modulador de PM. Como el voltaje de salida de un demodulador de PM es proporcional a la desviación de fase, la señal aparece demasiado reforzada en los graves. También, y éste es un caso más práctico, la PM demodulada por un receptor de FM produce una señal de información en la que se refuerzan las señales moduladoras de mayor frecuencia.

MODULACIÓN ANGULAR EN COMPARACIÓN CON MODULACIÓN DE AMPLITUD

Ventajas de la modulación angular

La modulación angular tiene varias ventajas inherentes sobre la modulación de amplitud.

Inmunidad al ruido. Es probable que la ventaja más importante de la transmisión por modulación angular (FM y PM) respecto a la de modulación de amplitud, es la inmunidad al ruido. La mayor parte del ruido, incluyendo el ruido causado por el hombre, produce variaciones no deseadas de amplitud en la onda modulada (es decir, ruido de AM). Los receptores de FM y de PM tienen limitadores que eliminan la mayor parte del ruido de AM en la señal recibida, antes de que se lleve a cabo el proceso final de demodulación; este proceso no se puede usar en los receptores de AM, porque la información también está contenida en variaciones de amplitud, y la eliminación de ruido también eliminaría la información.

Desempeño ante el ruido y mejoramiento de la relación señal a ruido. Con el uso de limitadores, los demoduladores de FM y PM pueden, en realidad, reducir el nivel de ruido y mejorar la relación de señal a ruido durante el proceso de demodulación. Este tema se explica con más detalle en el capítulo 7. A esto se le llama *umbral de FM*. Con la AM, una vez que el ruido contaminó la señal, no se puede eliminar.

Efecto de captura. En la FM y la PM hay un fenómeno llamado *efecto de captura* que permite a un receptor diferenciar entre dos señales recibidas con la misma frecuencia. Siempre que una de ellas tenga cuando menos el doble de amplitud que la otra, el receptor capturará la señal más intensa y eliminará la más débil. En la modulación de amplitud, si se reciben dos o más señales con la misma frecuencia, ambas se demodularán y producirán señales de audio. Una podrá tener mayor amplitud que la otra, pero ambas se podrán oír.

Uso de energía y eficiencia energética. En la transmisión de AM, en especial con DSBFC, la mayor parte de la potencia transmitida está en la portadora, mientras que la información está contenida en las bandas laterales, de mucho menor potencia. Con la modulación angular, la potencia total permanece constante independientemente de si hay modulación. En la AM, la potencia de portadora permanece constante con la modulación, y la potencia de banda lateral tan sólo se suma a la potencia de portadora. En la modulación angular, se toma la potencia de la portadora, con la modulación, y se redistribuye en las bandas laterales, por lo que, se podría decir, que la modulación angular pone la mayor potencia en la información.

Desventajas de la modulación angular

También la modulación angular tiene varias desventajas inherentes, respecto a la modulación de amplitud.

Ancho de banda. La modulación angular de alta calidad produce muchas frecuencias laterales y, en consecuencia, necesita un ancho de banda mucho mayor que el necesario para transmitir AM. La FM de banda angosta usa un índice de modulación bajo y, en consecuencia, sólo produce un conjunto de bandas laterales. Sin embargo, esas bandas laterales contienen un porcentaje todavía más desproporcionado de la potencia total que un sistema comparable de AM. Para tener una transmisión de alta calidad, la FM y la PM requieren mucho más ancho de banda que la AM. Cada estación, en la banda comercial de radio de AM tiene asignados 10 kHz de ancho de banda, mientras que en la banda comercial de FM se asignan 200 kHz a cada estación.

Complejidad y costo del circuito. Los moduladores, demoduladores, transmisores y receptores de PM y de FM tienen un diseño y fabricación más complejos que sus contrapartes de AM. Hace algún tiempo, más complejidad equivalía a más costo. Sin embargo, hoy, con el advenimiento de CI poco costosos con integración en gran escala, el costo de fabricar circuitos de FM y PM es comparable a sus contrapartes de AM.

PREGUNTAS

6-1. Defina *modulación angular.*

6-2. Defina *FM directa* y *FM indirecta.*

6-3. Defina *PM directa* y *PM indirecta.*

6-4. Defina *desviación de frecuencia* y *desviación de fase.*

6-5. Defina *fase instantánea, desviación instantánea de fase, frecuencia instantánea* y *desviación instantánea de frecuencia.*

6-6. Defina *sensibilidad a la desviación* para un modulador de frecuencia y para un modulador de fase.

6-7. Describa la relación entre la frecuencia instantánea de portadora y la señal moduladora para FM.

6-8. Describa la relación entre la fase instantánea de la portadora y la señal moduladora para PM.

6-9. Describa la relación entre la desviación de frecuencia y la amplitud y la frecuencia de la señal moduladora.

6-10. Defina la *variación de la portadora.*

6-11. Defina el *índice de modulación* para FM y para PM.

6-12. Describa la relación entre el índice de modulación y la señal moduladora para FM, y para PM.

6-13. Defina el porcentaje de modulación para señales con modulación de ángulo.

6-14. Describa la diferencia entre un modulador directo de frecuencia y un modulador directo de fase.

6-15. ¿Cómo se puede convertir un modulador de frecuencia en un modulador de fase? ¿Y un modulador de fase en un modulador de frecuencia?

6-16. ¿Cuántos conjuntos de bandas laterales se producen cuando una frecuencia de portadora se modula con una frecuencia única de entrada?

6-17. ¿Cuáles son los requisitos para que una frecuencia lateral se considere significativa?

6-18. Defina el índice de modulación *bajo, intermedio* y *alto.*

6-19. Describa la importancia de la *tabla de funciones de Bessel.*

6-20. Enuncie la *regla general de Carson* para determinar el ancho de banda de una onda con modulación angular.

6-21. Defina la *relación de desviación.*

6-22. Describa la relación entre la potencia en la portadora no modulada y en la onda modulada, para la FM.

6-23. Describa la importancia del *triángulo de ruido* de FM.

6-24. ¿Qué efecto tiene la *limitación* en la forma de onda compuesta de FM?

6-25. Defina lo que son *preénfasis* y *deénfasis.*

6-26. Describa una red de preénfasis y una red de deénfasis.

6-27. Describa el funcionamiento básico de un generador de FM con diodo varactor.

6-28. Describa el funcionamiento básico de un modulador de FM de reactancia.

6-29. Describa el funcionamiento básico de un modulador de FM en circuito integrado lineal.

6-30. Trace el diagrama de bloques de un transmisor directo de FM de Crosby, y describa su funcionamiento.

6-31. ¿Cuál es el objeto de un lazo AFC? ¿Por qué el receptor de Crosby requiere uno?

6-32. Trace el diagrama de bloques de un transmisor de FM con lazo de fase cerrada y describa su funcionamiento.

6-33. Trace el diagrama de bloques de un transmisor indirecto de FM de Armstrong, y describa su funcionamiento.

6-34. Compare la FM con la PM.

PROBLEMAS

6-1. Si un modulador de frecuencia produce 5 kHz de desviación de frecuencia para una señal moduladora de 10 V, determine la sensibilidad a la desviación. ¿Cuánta desviación de frecuencia produce una señal moduladora de 2 V?

6-2. Si un modulador de fase produce desviación de fase de 2 rad con una señal moduladora de 5 V, calcule la sensibilidad a la desviación. ¿Cuánta desviación de fase produciría una señal moduladora de 2 V?

6-3. Calcule: a) la desviación máxima de frecuencia, b) la variación de portadora y c) el índice de modulación de un modulador de FM con sensibilidad a la desviación $K_1 = 4$ kHz/V y una señal moduladora $v_m(t) = 10\ \text{sen}(2\pi 2000t)$. ¿Cuál es la desviación máxima de frecuencia producida, si la señal moduladora tuviera el doble de amplitud?

6-4. Calcule la desviación máxima de fase de un modulador de PM con una sensibilidad a la desviación $K = 1.5$ rad/V y una señal moduladora $v_m(t) = 2\ \text{sen}(2\pi 2000t)$. ¿Cuánta desviación de fase produce una señal moduladora con el doble de amplitud?

6-5. Calcule la modulación porcentual para una estación emisora de TV con desviación máxima de frecuencia $\Delta f = 50$ kHz, cuando la señal moduladora produce 40 kHz de desviación de frecuencia en la antena. ¿Cuánta desviación se requiere para llegar a 100% de modulación de la portadora?

6-6. Con la tabla de funciones de Bessel, calcule la cantidad de conjuntos de bandas laterales producidos con los siguientes índices de modulación: 0.5, 1.0, 2.0, 5.0 y 10.0.

6-7. Para un modulador de FM con índice de modulación $m = 2$, señal moduladora $v_m(t) = V_m\ \text{sen}(2\pi 2000t)$ y portadora no modulada $v_c(t) = 8\ \text{sen}(2\pi 800kt)$, determine:

(a) La cantidad de conjuntos de bandas laterales significativas.

(b) Sus amplitudes.

(c) Trace el espectro de frecuencias, indicando las amplitudes relativas de las frecuencias laterales.

(d) El ancho de banda.

(e) El ancho de banda si la amplitud de la señal moduladora aumenta en un factor de 2.5.

6-8. Para un transmisor de FM con variación de portadora de 60 kHz, determine la desviación de frecuencia. Si la amplitud de la señal moduladora disminuye en un factor de 2, determine la nueva desviación de frecuencia.

6-9. Para determinada señal de entrada, un transmisor comercial de FM tiene desviación de frecuencia $\Delta f = 20$ kHz. Calcule la desviación de frecuencia si aumenta en un factor de 2.5 la amplitud de la señal moduladora.

6-10. Un transmisor de FM tiene frecuencia en reposo $f_c = 96$ MHz, y sensibilidad a la desviación $K_1 = 4$ kHz/V. Calcule la desviación de frecuencia para una señal moduladora $v_m(t) = 8\ \text{sen}(2\pi 2000t)$. Calcule el índice de modulación.

6-11. Calcule la relación de desviación y el ancho de banda en el peor de los casos, para una señal de FM con desviación máxima de frecuencia $\Delta f = 25$ kHz y una frecuencia máxima de señal moduladora $f_{m(máx)} = 12.5$ kHz.

6-12. Para un modulador de FM con desviación de frecuencia de 40 kHz y frecuencia de señal moduladora $f_m = 10$ kHz, determine el ancho de banda mediante la tabla de funciones de Bessel y la regla de Carson.

6-13. Para un modulador de FM con amplitud de portadora no modulada $V_c = 20$ V, índice de modulación $m = 1$ y resistencia de carga $R_L = 10$ W, calcule la potencia en la portadora modulada y en cada frecuencia lateral, y trace el espectro de potencia para la onda modulada.

6-14. Para una portadora con modulación de ángulo $v_c(t) = 2\cos(2\pi 200 \text{ MHz } t)$ con 50 kHz de desviación de frecuencia debida a la señal moduladora, y una señal de interferencia de frecuencia única $V_n(t) = 0.5\cos(2\pi 200.01 \text{ MHz } t)$, determine:

(a) La frecuencia de la señal de interferencia demodulada.
(b) La desviación máxima de fase y de frecuencia debida a la señal de interferencia.
(c) La relación de señal a ruido en la salida del demodulador.

6-15. Calcule la desviación máxima total de fase producida por ruido aleatorio en una banda de 5 kHz con un voltaje máximo $V_n = 0.08$ V y una portadora $v_c(t) = 1.5 \text{ sen}(2\pi 40 \text{ MHz } t)$.

6-16. Calcule, para un transmisor directo de FM de Crosby, parecido al de la fig. 6-23, y con los parámetros indicados:

(a) La desviación de frecuencia en la salida del VCO y del amplificador de potencia.
(b) El índice de modulación en los mismos dos puntos.
(c) El ancho de banda en la salida del amplificador de potencia.

$N_1 = \times 3$
$N_2 = \times 3$
$N_3 = \times 2$
Frecuencia del oscilador de cristal de referencia = 13 MHz
Multiplicador de referencia = $\times 3$
Sensibilidad del VCO a la desviación $K_1 = 450$ Hz/V
Señal moduladora $v_m(t) = 3 \text{ sen}(2\pi 5 \times 10^3\, t)$
Frecuencia en reposo del VCO $f_c = 4.5$ MHz
Frecuencia de resonancia del discriminador $f_d = 1.5$ MHz

6-17. Para un transmisor indirecto de FM de Armstrong parecido al de la fig. 6-25, con los parámetros indicados, determine:

(a) El índice de modulación a la salida de la red combinadora y del amplificador de potencia.
(b) La desviación de frecuencia en esos mismos puntos.
(c) La frecuencia de la portadora de transmisión.

Oscilador de cristal de portadora = 210 kHz
Oscilador de cristal de referencia = 10.2 MHz
Voltaje de banda lateral $V_m = 0.018$ V
Voltaje de entrada de portadora al combinador $V_c = 5$ V
Primer multiplicador = $\times 40$
Segundo multiplicador = $\times 50$
Frecuencia de la señal moduladora $f_m = 2$ kHz

6-18. Si un modulador de frecuencia produce 4 kHz de desviación de frecuencia para una señal moduladora de 10 V_p, calcule la sensibilidad a la desviación.

6-19. Si un modulador de fase produce 1.5 rad de desviación de fase para una señal moduladora de 5 V_p, determine la sensibilidad a la desviación.

6-20. Determine: a) la desviación máxima de frecuencia, b) la variación de portadora y c) el índice de modulación para un modulador de FM con sensibilidad a la desviación $K_1 = 3$ kHz/V y una señal moduladora $v_m = 6 \text{ sen}(2\pi 2000t)$.

6-21. Calcule la desviación máxima de frecuencia para un modulador de PM con sensibilidad a la desviación $K = 2$ rad/V y una señal moduladora $v_m = 4 \text{ sen}(2\pi 1000t)$.

6-22. Calcule el porcentaje de modulación para una estación emisora de TV con desviación máxima de frecuencia $\Delta f = 50$ kHz, cuando la señal moduladora produce 30 kHz de desviación de frecuencia.

6-23. Con la tabla de funciones de Bessel determine la cantidad de frecuencias laterales producidas para los siguientes índices de modulación: 0.25, 0.5, 1.0, 2.0, 5.0 y 10.

6-24. Para un modulador de FM con índice de modulación $m = 5$, señal moduladora $v_m = 2 \text{ sen}(2\pi 5kt)$ y frecuencia de portadora no modulada $f_c = 400$ kHz, determine:

(a) La cantidad de conjuntos de bandas laterales significativas.
(b) Las amplitudes de bandas laterales.
A continuación, (c) Trace el espectro de frecuencias de salida.

6-25. Para un transmisor de FM con variación de portadora de 80 kHz, determine la desviación de frecuencia. Si la amplitud de la señal moduladora disminuye en un factor de 4, determine la nueva desviación de frecuencia.

6-26. Para cierta señal de entrada, un transmisor comercial de FM tiene una desviación de frecuencia $\Delta f = 40$ kHz. Determine la desviación de frecuencia si aumenta la amplitud de la señal moduladora en un factor de 4.3.

6-27. Un transmisor de FM tiene una frecuencia en reposo $f_c = 94$ MHz y una sensibilidad a la desviación $K_1 = 5$ kHz/V. Calcule la desviación de frecuencia para una señal moduladora $v_m(t) = 4\ V_p$.

6-28. Determine la relación de desviación y el ancho de banda en el peor de los casos para un sistema de FM con desviación máxima de frecuencia de 40 kHz y frecuencia máxima de señal moduladora $f_m = 10$ kHz.

6-29. Para un modulador de FM con 50 kHz de desviación de frecuencia y una frecuencia de señal moduladora $f_m = 8$ kHz, determine el ancho de banda, usando la tabla de funciones de Bessel y la regla de Carson.

6-30. Para un modulador de FM con voltaje de portadora no modulada $v_c = 12\ V_p$, índice de modulación $m = 1$ y resistencia de carga $R_L = 12\ \Omega$, determine la potencia en la portadora modulada y cada frecuencia lateral significativa, y trace el espectro de potencia para la onda de salida modulada.

6-31. Para una portadora con modulación angular, $v_c = 4\cos(2\pi 300\text{ MHz } t)$, con 75 kHz de desviación de frecuencia debida a la señal moduladora, y una señal de interferencia con frecuencia única $v_n = 0.2\cos(2\pi 300.015\text{ MHz } t)$, determine:

(a) La frecuencia de la señal demodulada de interferencia.
(b) La desviación máxima y rms de fase y de frecuencia debida a la señal de interferencia.
(c) La relación S/N a la salida del demodulador de FM.

6-32. Calcule la desviación total de fase rms producida por ruido aleatorio en una banda de 10 kHz con un voltaje máximo $V_n = 0.04$ V y una portadora con voltaje máximo $V_c = 4.5\ V_p$.

6-33. Para un transmisor directo de FM de Crosby, semejante al de la fig. 6-23, con las características mencionadas abajo, determine:

(a) Desviación de frecuencia en la salida del VCO y del amplificador de potencia.
(b) El índice de modulación en la salida del VCO y del amplificador de potencia.
(c) El ancho de banda en la salida del amplificador de potencia.

$N_1 = \times 3$
$N_2 = \times 3$
$N_3 = \times 2$
Frecuencia de referencia del oscilador de cristal = 13 MHz
Multiplicador de referencia = $\times 3$
Sensibilidad del VCO a la desviación $k_1 = 250$ Hz/V
Amplitud máxima de la señal moduladora $v_m = 4\ V_p$
Frecuencia de la señal moduladora $f_m = 10$ kHz
Frecuencia de reposo del VCO $f_c = 4.3$ MHz
Frecuencia de resonancia del discriminador $f_d = 1.5$ MHz

6-34. Para un transmisor indirecto de FM de Armstrong, semejante al de la fig. 6-25 y con los parámetros mencionados abajo, determine:

(a) Índice de modulación en la salida de la red combinadora y del amplificador de potencia.
(b) Desviación de frecuencia en esos dos puntos.
(c) Frecuencia de transmisión de la portadora.

Oscilador de cristal de portadora = 220 kHz
Oscilador de cristal de referencia = 10.8 MHz
Voltaje de banda lateral $V_m = 0.012\ V_p$

C A P Í

Imagen protegida por derechos de autor

Receptores por modulación de ángulo, FM estéreo y radio de FM en dos sentidos

INTRODUCCIÓN

Los receptores para señales con modulación angular son muy parecidos a los que se usan para recepción de AM convencional o de banda lateral única, a excepción del método que se usa para extraer la información de audio de la forma de onda compuesta de FI. En los receptores de FM, el voltaje a la salida del detector de audio es directamente proporcional a la desviación de frecuencia en su entrada. Con los receptores de PM, el voltaje a la salida del detector de audio es directamente proporcional a la desviación de fase en su entrada. Como se presentan la modulación de frecuencia y de fase al mismo tiempo con cualquiera de los sistemas de modulación angular, las señales de FM se pueden demodular con los receptores de PM, y viceversa. En consecuencia, los circuitos para demodular las señales de PM y de FM se describen a la vez bajo el título de "Receptores de FM".

Con la AM convencional, la señal moduladora se imprime en la portadora en forma de variaciones de amplitud. Sin embargo, la introducción de ruido al sistema produce también cambios en la amplitud de la envolvente. En consecuencia, el ruido no se puede quitar de la forma de onda compuesta, sin quitar también una porción de la señal de información. Con la modulación angular, la información se imprime sobre la portadora en forma de variaciones de frecuencia o de fase. Por consiguiente, en los receptores de modulación angular, se pueden quitar las variaciones de amplitud causadas por el ruido, de la forma de onda compuesta, tan sólo *limitando* (*recortando*) los máximos de la envolvente, antes de la detección. Con la modulación angular, se logra una mejoría en la relación de señal a ruido durante el proceso de demodulación; así, mediante la limitación, se puede mejorar el desempeño del sistema en presencia de ruido. En esencia, esta es la mayor ventaja de la modulación angular sobre la AM convencional.

Los objetivos de este capítulo son presentar al lector las configuraciones básicas de receptores y circuitos que se usan para recibir y demodular señales de FM y de PM, y describir su funcionamiento y en qué difieren de los receptores y circuitos de AM convencional o de banda lateral única. Además, se describirán varios sistemas de comunicaciones de FM que incluyen la FM estéreo y las radiocomunicaciones de FM en dos sentidos.

RECEPTORES DE FM

Los receptores de FM, como sus contrapartes de AM, son superheterodinos. La fig. 7-1 muestra el diagrama de bloques simplificado de un receptor superheterodino de FM de doble conversión. Allí se ve que el receptor de FM es parecido a los receptores de AM que se describieron en el capítulo 4. Las secciones de preselector, amplificador de RF, el primero y segundo mezclador, el amplificador de FI y de detector de un receptor de FM efectúan funciones casi idénticas a las que hacían en los receptores de AM: el preselector rechaza la frecuencia imagen, el amplificador de RF establece la relación de señal a ruido y la cifra de ruido, la sección de mezclador/convertidor hace conversión descendente de RF a FI, los amplificadores de FI proporcionan la mayor parte de la ganancia y selectividad del receptor, y el detector quita la información de la onda modulada. A excepción de el AGC demorado, para evitar la saturación del mezclador cuando se reciben señales intensas de RF, el AGC que se usa en los receptores de AM no se usa en los de FM, porque con la transmisión de FM no hay información en la amplitud de la señal recibida. Debido a las características inherentes de supresión de ruido de los receptores de FM, es común que tampoco se requieran amplificadores de RF en los receptores de FM.

En los receptores de FM es deseable tener una señal de FI de amplitud constante en el demodulador. En consecuencia, estos receptores tienen, en general, mucho más ganancia de FI que los receptores de AM. De hecho, con los receptores de FM se prefiere que esté saturado el amplificador final de FI. Las armónicas producidas al saturar el amplificador final de FI son suficientemente altas como para que se reduzcan en forma apreciable con los filtros pasabanda que sólo tienen el ancho de banda mínimo necesario para preservar las señales de información. El amplificador final de FI se diseña especialmente para tener características ideales de saturación, y se llama *limitador*, o a veces *limitador pasabanda* si se filtra la salida.

Las secciones de preselector, amplificadores de RF, el mezclador/convertidor y la sección de FI de un receptor de FM funcionan esencialmente igual que lo hacían en los receptores de AM; sin embargo, la etapa de detector de audio que se usa en los receptores de FM es muy distinta. El detector de envolvente (de picos), común a los receptores de AM, se sustituye por una *red limitadora, discriminadora de frecuencia* y *deénfasis* en los receptores de FM. El discriminador de frecuencia extrae la información de la onda modulada, mientras que el circuito limitador y el de deénfasis contribuyen a mejorar la relación de señal a ruido que se alcanza en la etapa de demodulador de audio en los receptores de FM.

Para los receptores de FM en la banda comercial, la primera FI es una frecuencia relativamente alta, por ejemplo de 10.7 MHz, para tener buen rechazo de la frecuencia imagen, y la segunda FI es una frecuencia relativamente baja (muchas veces de 455 kHz) que permite tener, a los amplificadores de FI, una ganancia relativamente alta, sin ser susceptibles a la oscilación. Con una primera FI de 10.7 MHz, la frecuencia imagen hasta para la estación de FM con frecuencia mínima (88.1 MHz) es 109.5 MHz, muy afuera de la banda de emisión de FM.

DEMODULADORES DE FM

Los demoduladores de FM son circuitos dependientes de la frecuencia, diseñados para producir un voltaje de salida que sea proporcional a la frecuencia instantánea en su entrada. La función general de transferencia para un demodulador de FM es no lineal, pero cuando se trabaja en su intervalo lineal, es

$$K_d = \frac{V\,(\text{volts})}{f\,(\text{Hz})} \tag{7-1}$$

donde K_d es igual a la función de transferencia.

La salida de un demodulador de FM se expresa como sigue

$$v_{\text{sal}}\,(\text{t}) = K_d \Delta f \tag{7-2}$$

en donde $v_{\text{sal}}\,(t)$ = señal demodulada de salida (volts)
K_d = función de transferencia del demodulador (volts por hertz)
Δf = diferencia entre la frecuencia de entrada y la frecuencia central del demodulador (hertz)

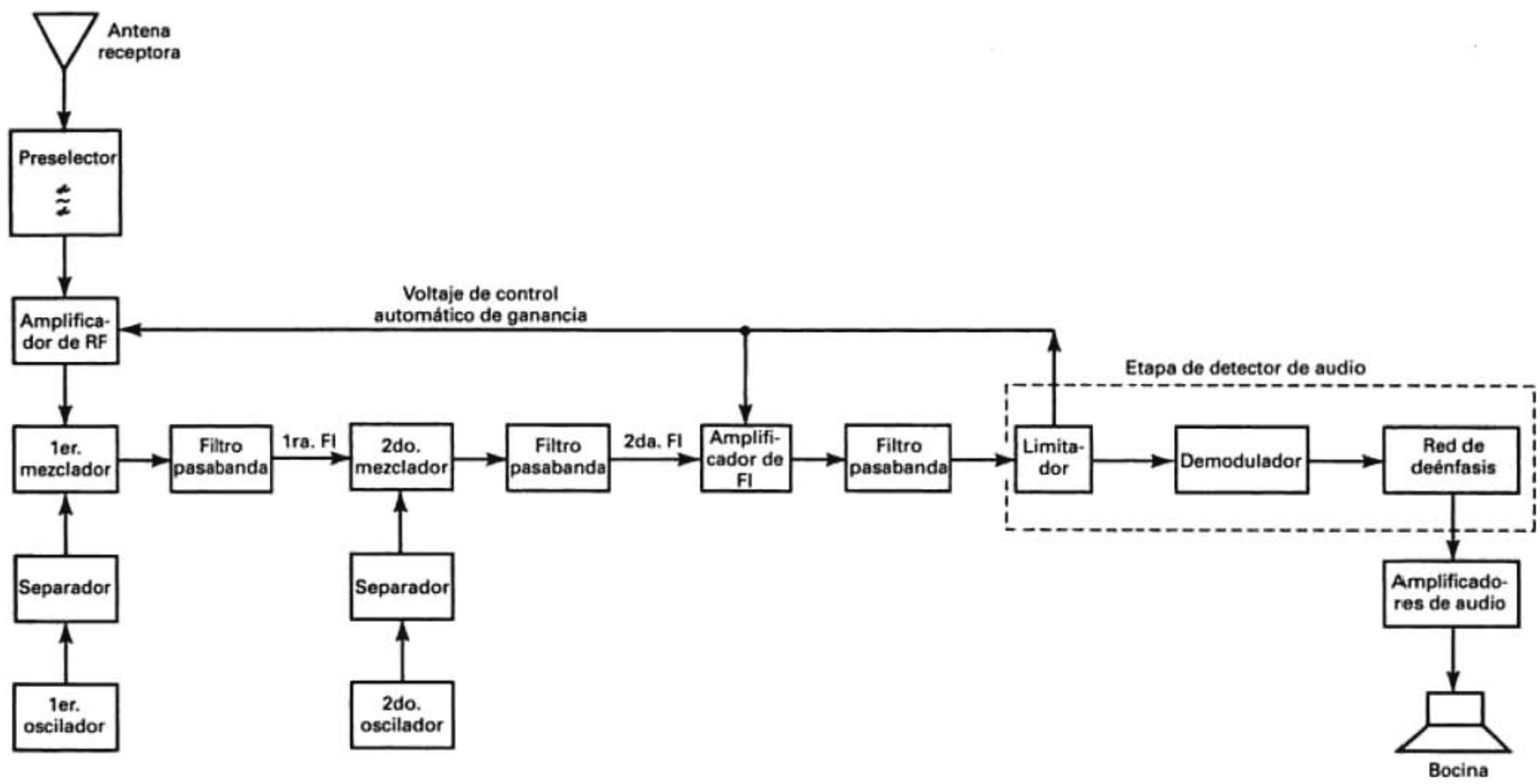


FIGURA 7-1 Diagrama de bloques de un receptor de FM con doble conversión

Ejemplo 7-1

Calcular el voltaje máximo de salida para un circuito demodulador de FM con función de transferencia $K_d = 0.2$ V/kHz y señal de entrada de FM con desviación máxima de frecuencia de 20 kHz.

Solución Se sustituye en la ecuación 7-2 para obtener el voltaje máximo de salida:

$$v_{sal}(t) = \frac{0.2\ \text{V}}{\text{kHz}} \times 20\ \text{kHz}$$
$$= 4\ \text{V}_p$$

Hay varios circuitos que se usan para demodular las señales de FM. Los más comunes son el *detector de pendiente*, el *discriminador de Foster-Seeley*, el *detector de relación*, el *demodulador PLL* y el *detector de cuadratura*. El detector de pendiente, el discriminador de Foster-Seeley y el detector de relación son formas de *discriminadores de frecuencia con circuito sintonizado*.

Discriminadores de frecuencia con circuito sintonizado

Estos discriminadores convierten la FM en AM y a continuación demodulan la envolvente de AM con detectores convencionales de picos. También, la mayor parte de los discriminadores de frecuencia requieren un inversor de fase de 180°, un circuito sumador y uno o más circuitos dependientes de frecuencia.

Detector por pendiente. La fig. 7-2a muestra el esquema de un *detector por pendiente no balanceado*, que es la forma más sencilla de discriminador de frecuencia con circuito sintonizado. El detector por pendiente no balanceado tiene las características no lineales de voltaje en función de frecuencia y, en consecuencia, casi nunca se usa. Sin embargo, su funcionamiento es básico para todos los discriminadores de frecuencia con circuito sintonizado.

En la fig. 7-2a, el circuito sintonizado (L_a y C_a) produce un voltaje de salida que es proporcional a la frecuencia de entrada. El voltaje máximo de salida se presenta en la frecuencia de resonancia f_c del circuito tanque, y su salida disminuye en forma proporcional a la desviación de la frecuencia respecto a f_o. El circuito está diseñado de modo que la frecuencia intermedia f_c central queda en el centro de la parte más lineal de la curva de voltaje en función de frecuencia, como se ve en la fig. 7-2b. Cuando la frecuencia intermedia se desvía arriba de f_c, el voltaje de salida aumenta; cuando la frecuencia intermedia se desvía abajo de f_c, el voltaje de salida disminuye. En consecuencia, el circuito sintonizado convierte variaciones de frecuencia en variaciones de amplitud (conversión de FM a AM). Los elementos D_i, C_i y R_i forman un detector sencillo de picos que convierte las variaciones de amplitud en un voltaje de salida que varía con una rapidez igual a la de los cambios de frecuencia, y cuya amplitud es proporcional a la magnitud de los cambios de frecuencia.

Detector por pendiente balanceado. La fig. 7-3a muestra el diagrama de un *detector por pendiente balanceado*. Un detector por pendiente balanceado es un discriminador de frecuencias

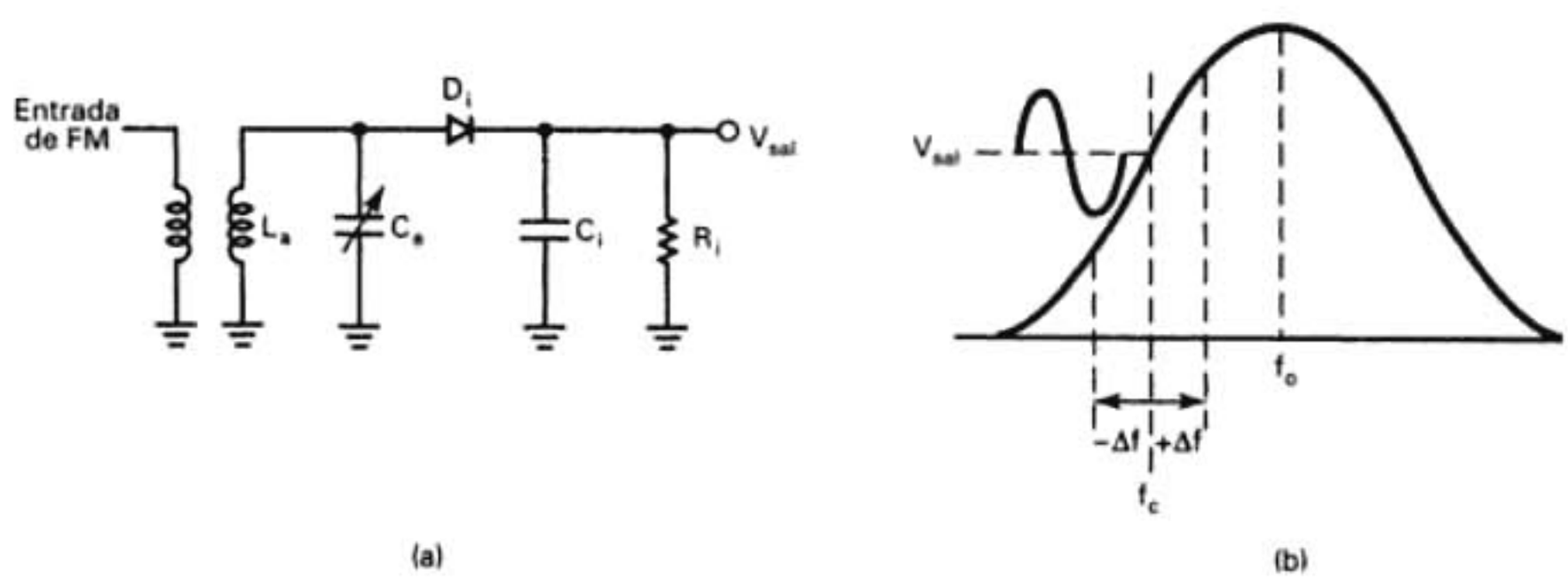


FIGURA 7-2 Detector de pendiente: (a) diagrama; (b) curva de voltaje en función de frecuencia

de circuito sintonizado, y un detector por pendiente balanceado sólo consiste en dos detectores por pendiente no balanceados conectados en paralelo y alimentados con un desfase de 180°. La inversión de fase se logra con la toma central del devanado secundario del transformador T_1. En la fig. 7-3a, los circuitos sintonizados L_a, C_a y L_b, C_b hacen la conversión de FM a AM, y los detectores balanceados de pico D_1, C_1, R_1 y D_2, C_2 y R_2 obtienen la información de la envolvente de AM. El circuito sintonizado superior L_a y C_a se sintoniza a la frecuencia f_a que es mayor que la frecuencia intermedia f_o central, la cantidad $1.33 \times \Delta f$ (para la banda comercial de FM, esto es aproximadamente 1.33×75 kHz = 100 kHz). El circuito sintonizado inferior L_b y C_b se sintoniza a una frecuencia f_b que es menor que la frecuencia intermedia central en una cantidad igual.

El funcionamiento del circuito es bastante sencillo. El voltaje de salida de cada circuito sintonizado es proporcional a la frecuencia de entrada, y cada salida se rectifica con su respectivo detector de picos. Por consiguiente, mientras más cercana es la frecuencia de entrada a la frecuencia de resonancia del circuito tanque, el voltaje de salida del circuito tanque es mayor. La frecuencia central FI cae exactamente a la mitad entre las frecuencias de resonancia de los dos circuitos sintonizados. Así, en la frecuencia intermedia central, los voltajes de salida de los dos circuitos sintonizados tienen amplitud igual, pero polaridad opuesta. En consecuencia, el voltaje de salida rectificado a través de R_1 y R_2, sumado, produce un voltaje diferencial de salida $V_{sal} = 0$ V. Cuando la FI se desvía arriba de la resonancia, el circuito sintonizado superior produce un voltaje mayor de salida que el circuito tanque inferior, y V_{sal} se hace positivo. Cuando la FI se desvía por abajo de la resonancia, el voltaje de salida del circuito tanque inferior es mayor que el voltaje de salida del superior, y V_{sal} se vuelve negativo. En la fig. 7-3b se muestra una curva de respuesta de salida en función de la frecuencia.

Aunque el detector por pendiente es, probablemente, el detector más sencillo de FM, tiene varias desventajas inherentes, entre las que están la mala linealidad, la dificultad de sintonía y la falta de partes limitadoras. Como no hay limitación, un detector de pendiente produce un voltaje de salida que es proporcional a las variaciones tanto de amplitud como de frecuencia de la señal de entrada, y en consecuencia, debe estar precedido por una etapa limitadora. Un detector balanceado por pendiente se alinea inyectando una frecuencia igual a la FI central y sintonizando C_a y C_b para 0 V en la salida. Entonces, las frecuencias iguales a f_a y f_b se inyectan alternadamente, mientras que C_a y C_b se sintonizan para los voltajes máximos e iguales, con polaridades opuestas.

Discriminador de Foster-Seeley. Un *discriminador de Foster-Seeley*, que también se llama *discriminador de desplazamiento de fase* es un discriminador de frecuencia con circuito sintonizado cuyo funcionamiento es muy parecido al del detector por pendiente balanceado. El esquema de un discriminador de Foster-Seeley se ve en la fig. 7-4a. Los valores de capacitancia de C_c, C_1 y C_2 se escogen de tal modo que son un cortocircuito para la FI central. Así, el lado derecho de L_3 está al potencial de tierra de ca, y la señal de FI, V_{ent}, se alimenta en forma directa (en

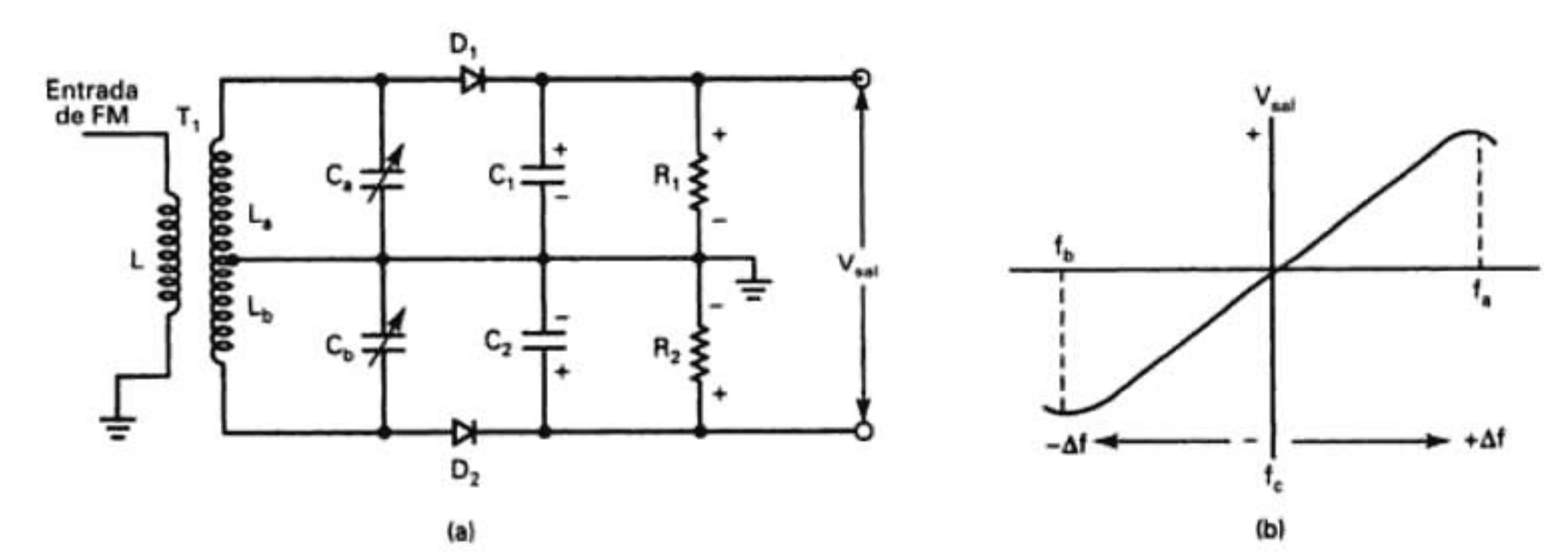


FIGURA 7-3 Detector balanceado por pendiente: (a) diagrama eléctrico; (b) voltaje en función de la frecuencia

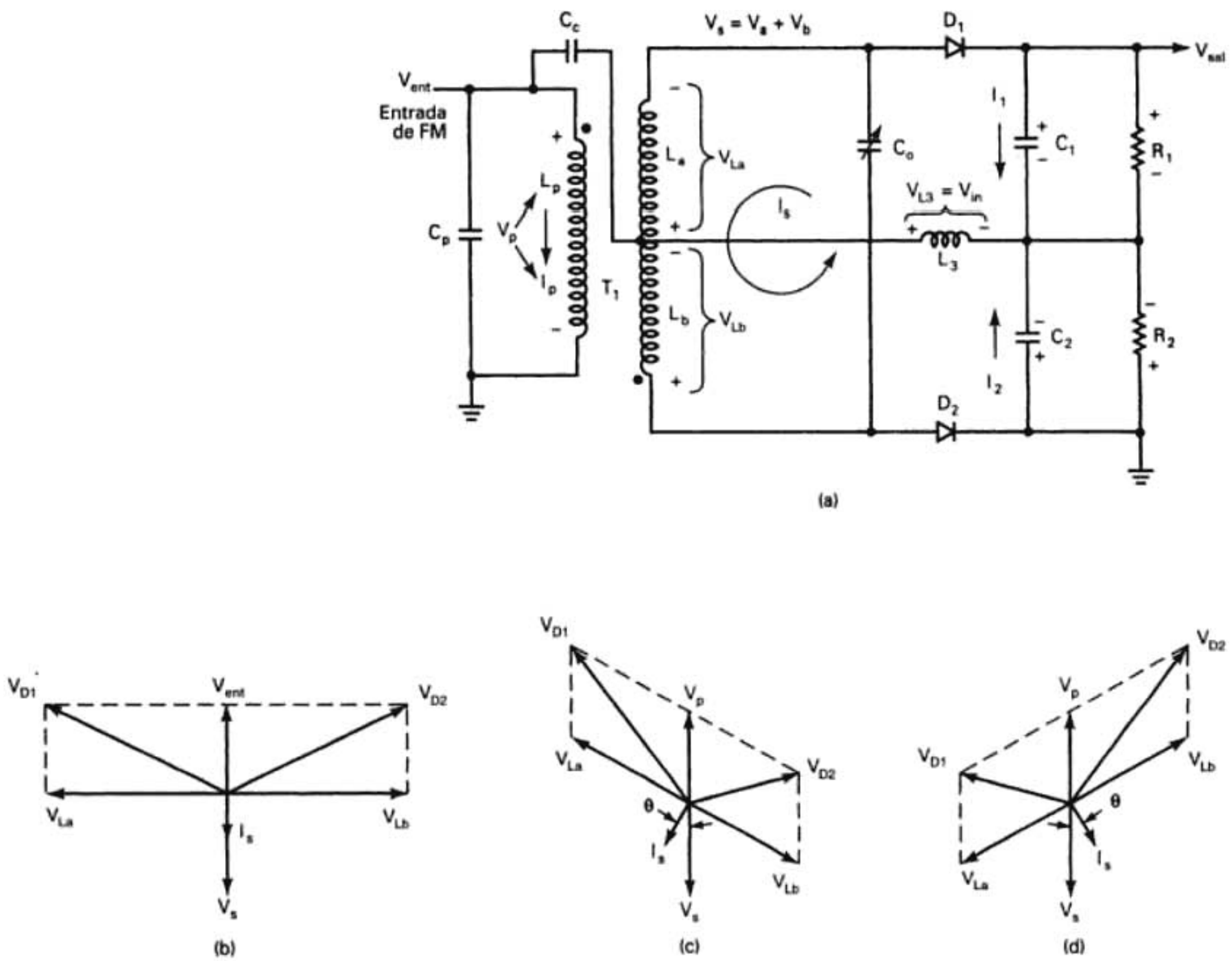


FIGURA 7-4 Discriminador de Foster-Seeley: (a) diagrama eléctrico; (b) diagrama vectorial, $f_{ent} = f_o$; (c) diagrama vectorial, $f_{ent} > f_o$; (d) diagrama vectorial, $f_{ent} < f_o$

fase) a través de L_3 (V_{L3}). La FI entrante se invierte 180° con el transformador T_1 y se divide por igual entre L_a y L_b. En la frecuencia de resonancia (la FI central) del circuito tanque secundario, la corriente I_s del secundario está en fase con el voltaje total secundario, V_s, y desfasada 180° respecto a V_{L3}. También, debido al acoplamiento flojo, el primario de T_1 funciona como un inductor, y la corriente I_p del primario está desfasada 90° con respecto a V_{ent} y, debido a que la inducción magnética depende de la corriente del primario, el voltaje inducido en el secundario está desfasada 90° con respecto a V_{ent} (V_{L3}). Por consiguiente, V_{La} y V_{Lb} están desfasadas 180° entre sí, y en cuadratura, o desfasadas 90° respecto a V_{L3}. El voltaje a través del diodo superior V_{D1} es la suma vectorial de V_{L3} y V_{La}, y el voltaje a través del diodo inferior, V_{D2}, es la suma vectorial de V_{L3} y V_{Lb}. En la fig. 7-4b se muestran los diagramas vectoriales correspondientes. De acuerdo con la figura, los voltajes a través de D_1 y D_2 son iguales. Así, en la resonancia, I_1 e I_2 son iguales, y C_1 y C_2 cargan a voltajes de igual magnitud, pero polaridades opuestas. En consecuencia, $V_{sal} = V_{C1} - V_{C2} = 0$ V. Cuando la FI sube de la resonancia ($X_L > X_C$), la impedancia del circuito tanque secundario se vuelve inductiva, y la corriente del secundario se retrasa un ángulo θ respecto al voltaje del secundario; el ángulo es proporcional a la magnitud de la desviación de frecuencia. El diagrama fasorial correspondiente se ve en la fig. 7-4c. Allí se ve que la suma vectorial del voltaje a través de D_1 es mayor que la suma vectorial de los voltajes a través de D_2. En consecuencia, C_1 carga mientras que C_2 descarga, y V_{sal} se hace positivo. Cuando la FI baja de la resonancia ($X_L < X_C$), la corriente del secundario se adelanta al voltaje del secundario un ángulo θ, que de nuevo es proporcional a la

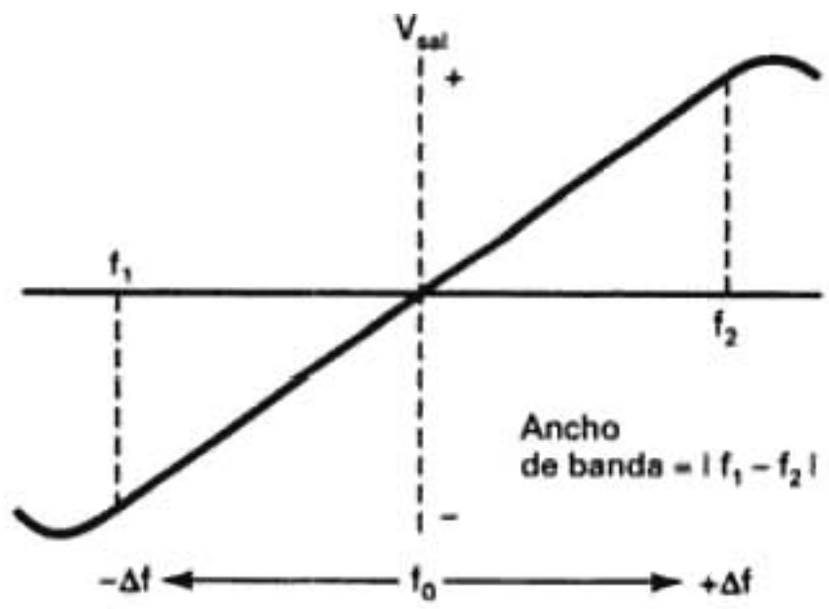


FIGURA 7-5 Curva de respuesta de voltaje en función de la frecuencia de un discriminador

magnitud del cambio de frecuencia. Los fasores correspondientes se ven en la fig. 7-4d. Se aprecia que la suma vectorial de los voltajes a través de D_1 es ahora menor que la suma vectorial de los voltajes a través de D_2. En consecuencia, C_1 descarga mientras que C_2 carga, y V_{sal} se vuelve negativo. Un discriminador de Foster-Seeley se sintoniza inyectando una frecuencia igual a la FI central y sintonizando a C_o para 0 V en la salida.

La descripción anterior y la fig. 7-4 indican que el voltaje de salida de un discriminador de Foster-Seeley es directamente proporcional a la magnitud y dirección de la desviación de frecuencia. La fig. 7-5 muestra una curva característica de respuesta de voltaje en función de la frecuencia para un discriminador de Foster-Seeley. Por razones obvias se le llama con frecuencia *curva S*. Se ve que la curva de voltaje de salida en función de desviación de frecuencia es más lineal que la de un detector por pendiente, y como sólo hay un circuito tanque, es más fácil de sintonizar. Para una demodulación sin distorsión se debe restringir la desviación de frecuencia a la parte lineal de la curva de respuesta en frecuencia del circuito secundario sintonizado. Como en el caso del detector por pendiente, un discriminador de Foster-Seeley responde a variaciones de amplitud y también de frecuencia y, en consecuencia, debe estar precedido por un circuito limitador separado.

Detector de relación. El *detector de relación* tiene una gran ventaja sobre el detector por pendiente y el discriminador de Foster-Seeley para demodulación de FM: es relativamente inmune a variaciones de amplitud en su señal de entrada. La fig. 7-6a muestra el diagrama de un detector de relación. Al igual que el discriminador de Foster-Seeley, un detector de relación tiene un solo circuito sintonizado en el secundario del transformador. Así, su funcionamiento se parece al del discriminador de Foster-Seeley. De hecho, los vectores de voltaje para D_1 y D_2 son idénticos con los del circuito discriminador de Foster-Seeley de la fig. 7-4. Sin embargo, en el detector de relación se invierte un diodo (D_2) y la corriente I_d puede pasar en torno al lazo más externo del circuito. Así, después de varios ciclos de la señal de entrada, el capacitor *en paralelo* C_s se carga a más o menos al voltaje máximo a través del devanado secundario de T_1. La reactancia de C_s es baja, y R_s tan sólo es un camino de cd para la corriente del diodo. Por consiguiente, la constante de tiempo para R_s y C_s es suficientemente larga como para que los cambios rápidos de amplitud de la señal de entrada, debidos al ruido térmico u otras señales de interferencia, pasen a tierra y no tengan efecto sobre el voltaje promedio a través de C_s. Entonces, C_1 y C_2 se cargan y descargan en forma proporcional a los cambios de frecuencia en la señal de entrada, y son relativamente inmunes a variaciones de amplitud. También, el voltaje de salida de un detector de relación se toma con respecto a tierra, y para las polaridades del diodo que se ven en la fig. 7-6a, el voltaje promedio de salida es positivo. En la resonancia, el voltaje de salida se divide por igual entre C_1 y C_2, y se redistribuye de acuerdo a como se desvía la frecuencia de entrada sobre o abajo de la resonancia. Por lo anterior, los cambios de V_{sal} se deben a la relación cambiante de voltajes a través de C_1 y C_2, mientras que el voltaje total está fijado por C_s.

La fig. 7-6b muestra la curva de la respuesta en frecuencia de la salida para el detector de relación de la fig. 7-6a. Se ve que en la resonancia, V_{sal} no es igual a 0 V, sino más bien a la mitad del voltaje a través de los devanados secundarios de T_1. Como un detector de relación es relativamente inmune a variaciones de amplitud, se selecciona con frecuencia como discriminador.

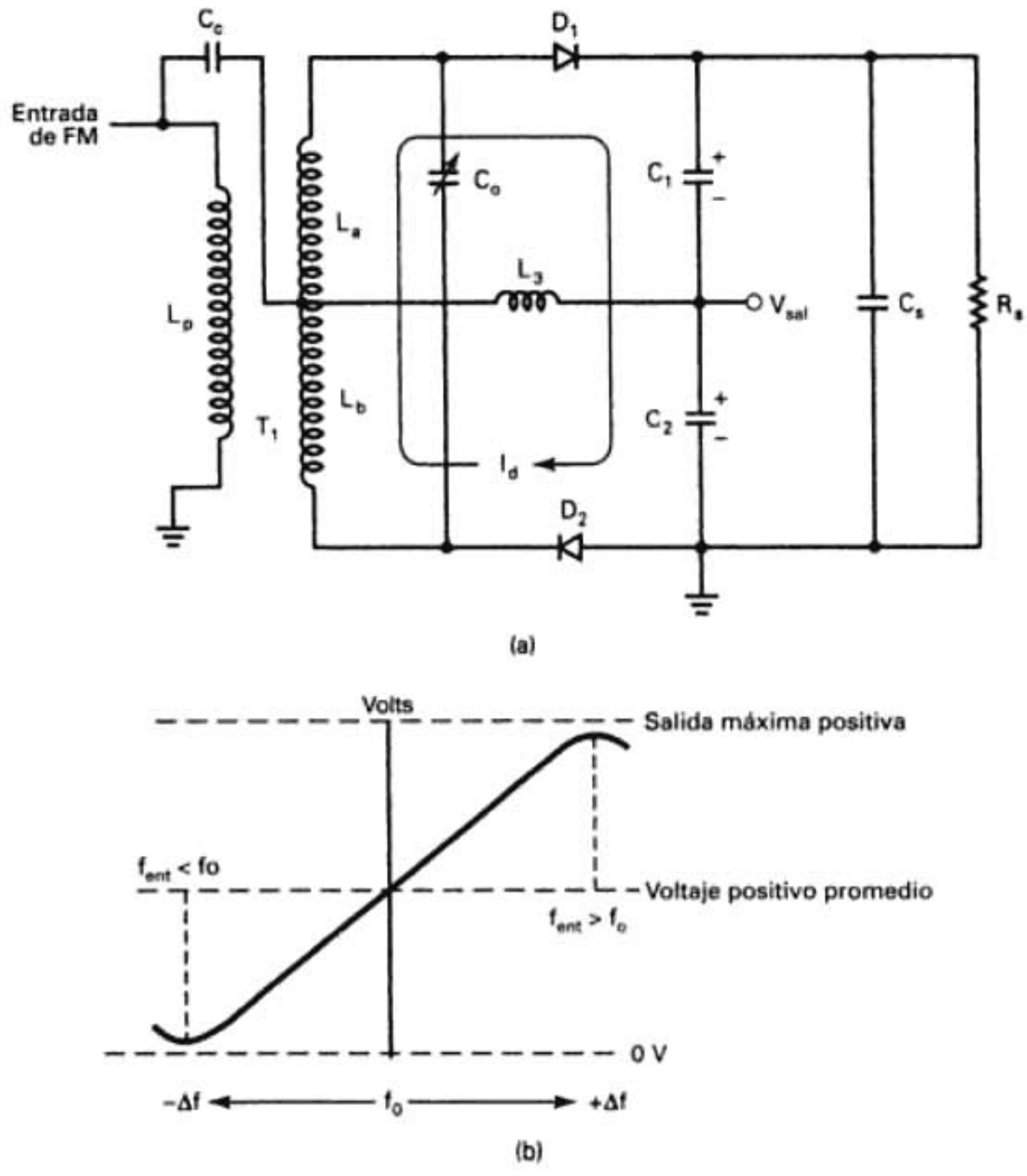


FIGURA 7-6 Detector de relación: (a) diagrama; (b) curva de respuesta de voltaje en función de la frecuencia

Sin embargo, un discriminador produce una curva de respuesta más lineal, de voltaje en función de frecuencia.

DEMODULADOR DE FM CON LAZO DE FASE CERRADA

Desde que se desarrollaron los circuitos integrados lineales en gran escala, se puede lograr con bastante facilidad la demodulación de FM con un lazo de fase cerrada (PLL, de *phase-locked loop*). Aunque el funcionamiento de un PLL es muy complicado, es probable que el funcionamiento de un *demodulador de FM con PLL* sea más sencillo y fácil de comprender. Un demodulador de frecuencia con PLL no requiere circuitos sintonizados, y compensa en forma automática los cambios de frecuencia de portadora debidos a inestabilidad del oscilador del transmisor. La fig. 7-7a muestra el diagrama de bloques simplificado de un demodulador de FM con PLL.

En el capítulo 2 se presentó una explicación detallada del funcionamiento del lazo de fase cerrada (PLL). Se demostró que después de haber sucedido la fijación de frecuencia, el VCO rastrea los cambios de frecuencia en la señal de entrada, manteniendo un error de fase en la entrada del comparador de fases. Por consiguiente, si la entrada al PLL es una señal desviada de FM, y la frecuencia natural del VCO es igual a la FI central, el voltaje de corrección que se produce en la salida del comparador de fases, y se retroalimenta a la entrada del VCO, es proporcional a la desviación de frecuencia y es, por consiguiente, la señal de información demodulada. Si la amplitud de la FI se limita lo suficiente antes de llegar al PLL, y se compensa bien el lazo, la

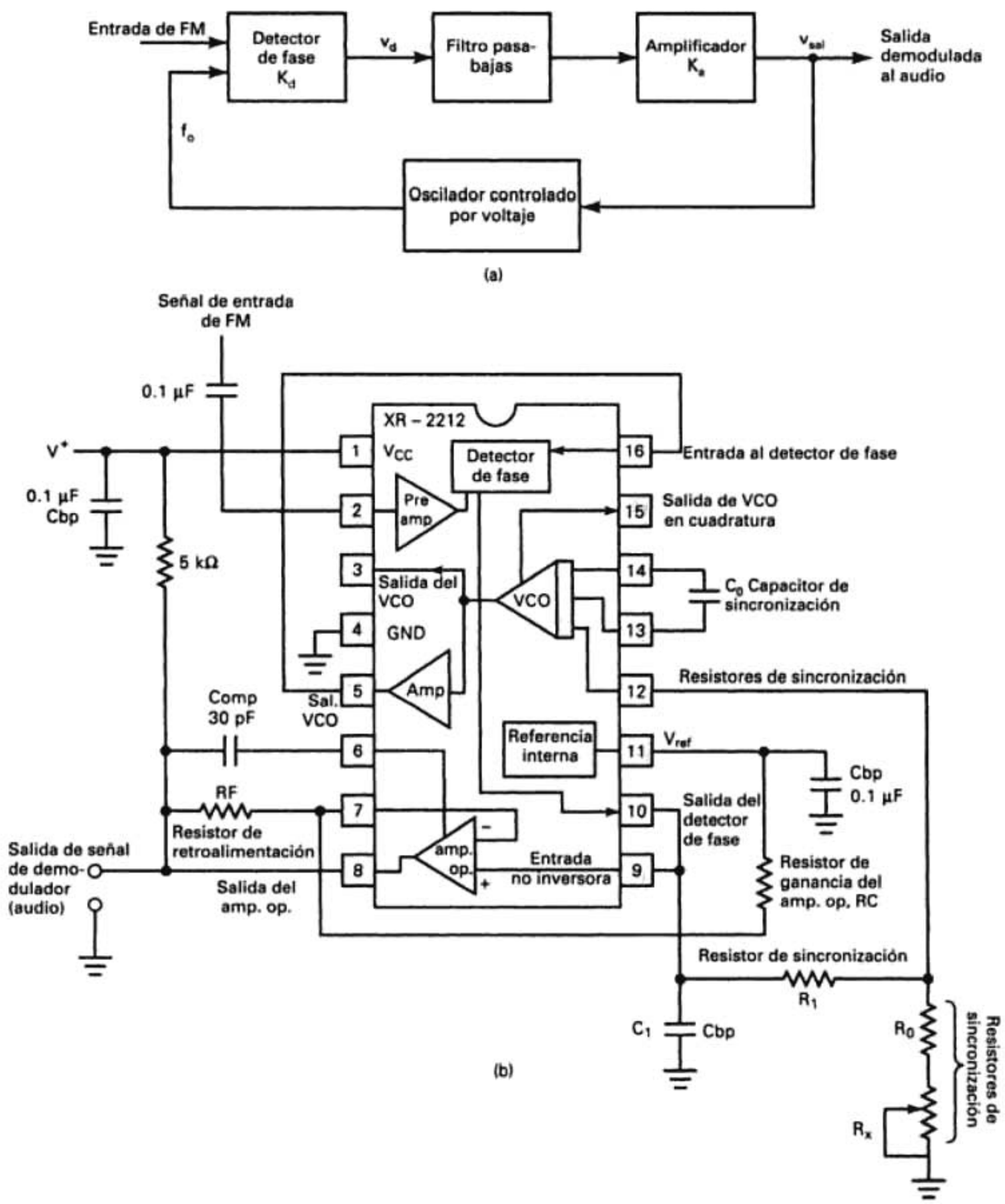


FIGURA 7-7 (a) Diagrama de bloques de un demodulador de FM con PLL; (b) demodulador PLL FM con PLL XR-2212 en circuito integrado

ganancia del lazo de PLL es constante e igual a K_v. Por consiguiente, la señal demodulada se puede tomar en forma directa de la salida del separador interno, y se describe con la ecuación

$$V_{sal} = \Delta f K_d K_a \tag{7-3}$$

La fig. 7-7b muestra el esquema de un demodulador de FM con el circuito XR-2212. R_0 y C_0 son elementos de ajuste, para establecer la frecuencia propia, o de oscilación libre. El resistor R_x es para sintonía fina, y R_F y R_C establecen la ganancia de voltaje del amplificador operacional interno, K_a. La respuesta del lazo cerrado de PLL a la frecuencia se debe compensar, para

permitir una demodulación no atenuada de todo el ancho de banda de la señal de información. El amplificador operacional separador del PLL proporciona ganancia de voltaje y estabilidad de corriente de salida.

DEMODULADOR DE FM POR CUADRATURA

Un *demodulador de FM por cuadratura* (llamado a veces *detector por coincidencia*) extrae la señal de información original de la forma de onda compuesta, de FI, multiplicando dos señales en cuadratura, es decir, desfasadas 90°. Un detector por cuadratura usa un desplazador de fase de 90°, un solo circuito sintonizado y un detector de producto, para demodular señales de FM. El desplazador de fase de 90° produce una señal que está en cuadratura con las señales recibidas de FI. El circuito sintonizado convierte variaciones de frecuencia en variaciones de fase, y el detector de producto multiplica las señales recibidas de FI por la señal de FI con fase desplazada.

La fig. 7-8 muestra un diagrama simplificado de un detector de FM por cuadratura. El capacitor C_i es de alta reactancia que, cuando se conecta en serie con el circuito tanque R_o, L_o y C_o, produce un desplazamiento de fase de 90° en la FI central. El circuito tanque se sintoniza a la FI central, y produce un desplazamiento adicional de fase θ, que es proporcional a la desviación de frecuencia. La señal de entrada de FI, v_i, se multiplica por la señal en cuadratura v_o, en el detector de producto, y produce una señal de salida que es proporcional a la desviación de frecuencia. En la frecuencia de resonancia, la impedancia del circuito tanque es resistiva. Sin embargo, las variaciones de frecuencia en la señal de FI producen un desplazamiento adicional de fase, positivo o negativo. Así, el voltaje de salida del detector de producto es proporcional a la diferencia de fases entre las dos señales de entrada y se expresa como sigue

$$v_{sal} = v_i v_o = [V_i \operatorname{sen}(\omega_i t + \theta)]\,[V_o \cos(\omega_o t)] \tag{7-4}$$

Se sustituye en esta ecuación la identidad trigonométrica del producto de un seno por un coseno de igual frecuencia, y el resultado es

$$v_{sal} = \frac{V_i V_o}{2}\,[\operatorname{sen}(2\omega_i t + \theta) + \operatorname{sen}(\theta)] \tag{7-5}$$

La segunda armónica ($2\omega_i$) se filtra y se elimina, y queda

$$v_{sal} = \frac{V_i V_o}{2} \operatorname{sen}(\theta) \tag{7-6}$$

en donde $\theta = \tan^{-1} \rho Q$
$\rho = 2\pi f/f_o$, la desviación fraccionaria de frecuencia
Q = factor de calidad del circuito tanque

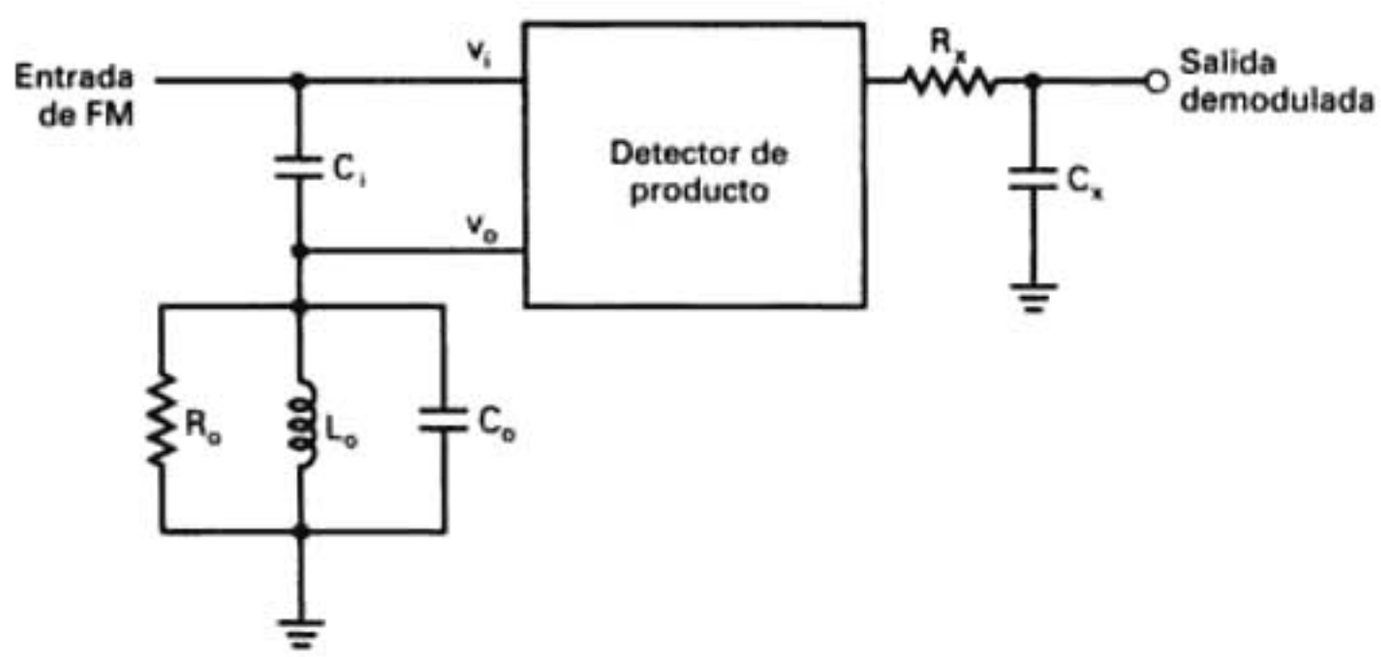


FIGURA 7-8 Demodulador de FM por cuadratura

SUPRESIÓN DE RUIDO EN FM

Es probable que la ventaja más importante de la modulación de frecuencia, sobre la modulación de amplitud, es la capacidad que tienen los receptores de FM para suprimir el ruido. Ya que la mayor parte del ruido aparece en forma de variaciones de amplitud en la onda modulada, los demoduladores de AM no pueden eliminar el ruido sin eliminar también algo de la información. Esto se debe a que también la información está contenida en las variaciones de amplitud. Sin embargo, en la FM, la información está contenida en variaciones de frecuencia, lo cual permite eliminar las variaciones indeseadas de amplitud, con circuitos especiales llamados *limitadores*.

Limitadores de amplitud y umbral de FM

La vasta mayoría de los sistemas de radiocomunicaciones terrestres de FM usan la demodulación convencional no coherente, porque la mayoría de los discriminadores normales de frecuencia aplican la detección de envolvente para sacar la información de la onda de FM. Desafortunadamente, los detectores de envolvente, incluidos los detectores de relación, demodulan variaciones incidentales de amplitud, al igual que variaciones de frecuencia. El ruido de transmisión y de interferencia se suman a la señal y producen variaciones indeseadas de amplitud. También y en general, la modulación de frecuencia se acompaña por pequeñas cantidades de modulación residual de amplitud. El ruido es más notable en los máximos de la forma de onda de FM, y relativamente insignificante durante los cruces con cero. Un limitador es un circuito que produce una salida de amplitud constante para todas las señales de entrada mayores que un nivel mínimo predeterminado, que se suele llamar nivel *umbral*, *de silenciamiento* o *de captura*. Se requieren los limitadores en la mayoría de los receptores de FM, porque muchos de los demoduladores que se describieron antes en este capítulo demodulan variaciones de amplitud y también de frecuencia. Con los limitadores de amplitud, la relación de señal a ruido en la salida del demodulador (post detección) se puede mejorar hasta en 20 dB o más, respecto a la relación de señal a ruido en la entrada (predetección).

En esencia, un limitador de amplitud es un amplificador adicional de FI que está sobreexcitado. La limitación comienza cuando la señal de FI es suficientemente grande para activar al amplificador hasta la saturación y al corte, en forma alternativa. En las figs. 7-9a y 7-9b se ven las formas de onda de entrada y salida de un limitador típico. La fig. 7-9b muestra que para señales de FI por abajo del umbral, el ruido de AM no se reduce, y arriba del umbral hay una gran reducción del nivel de ruido de AM. El objeto del limitador es eliminar todas las variaciones de amplitud de la señal de FI.

La fig. 7-10a muestra la salida del limitador cuando el ruido es mayor que la señal, es decir, el ruido ha capturado al limitador. Los anchos irregulares de los picos son causados por pulsos de ruido que saturan al limitador. La fig. 7-10b muestra la salida del limitador cuando la señal es suficientemente mayor que el ruido (la señal ha capturado al limitador). Los picos de la señal han conducido al limitador tan adentro de la saturación, que el ruido más débil se elimina en forma total. El mejoramiento de la relación S/N se llama *umbral de FM, silenciamiento de FM* o *efecto de captura de FM*. Se deben satisfacer tres criterios para poder obtener el umbral de FM:

1. La relación de señal a ruido en predetección debe ser 10 dB o mayor.
2. La señal de FI debe amplificarse lo suficiente como para sobreexcitar al limitador.
3. La señal debe tener un índice de modulación igual o mayor que la unidad ($m \geq 1$).

La fig. 7-11 muestra curvas características umbral de FM para señales de índice bajo ($m = 1$) e intermedio ($m = 4$). El voltaje de salida de un detector de FM es proporcional a m^2. Así, al subir m al doble, la relación S/N aumenta en un factor de 4 (6 dB). La relación de silenciamiento para $m = 1$ es una relación S/N = 13 dB de entrada, y para $m = 4$, de 22 dB. Para relaciones de S/N abajo del umbral, se dice que el receptor está capturado por el ruido, y con relaciones de S/N arriba del umbral, el receptor está capturado por la señal. La fig. 7-11 muestra que las señales de FI en la entrada del limitador, con 13 dB o más de S/N, tienen un mejoramiento de S/N. El silenciamiento en la FM comienza con una relación S/N en la entrada de 10 dB,

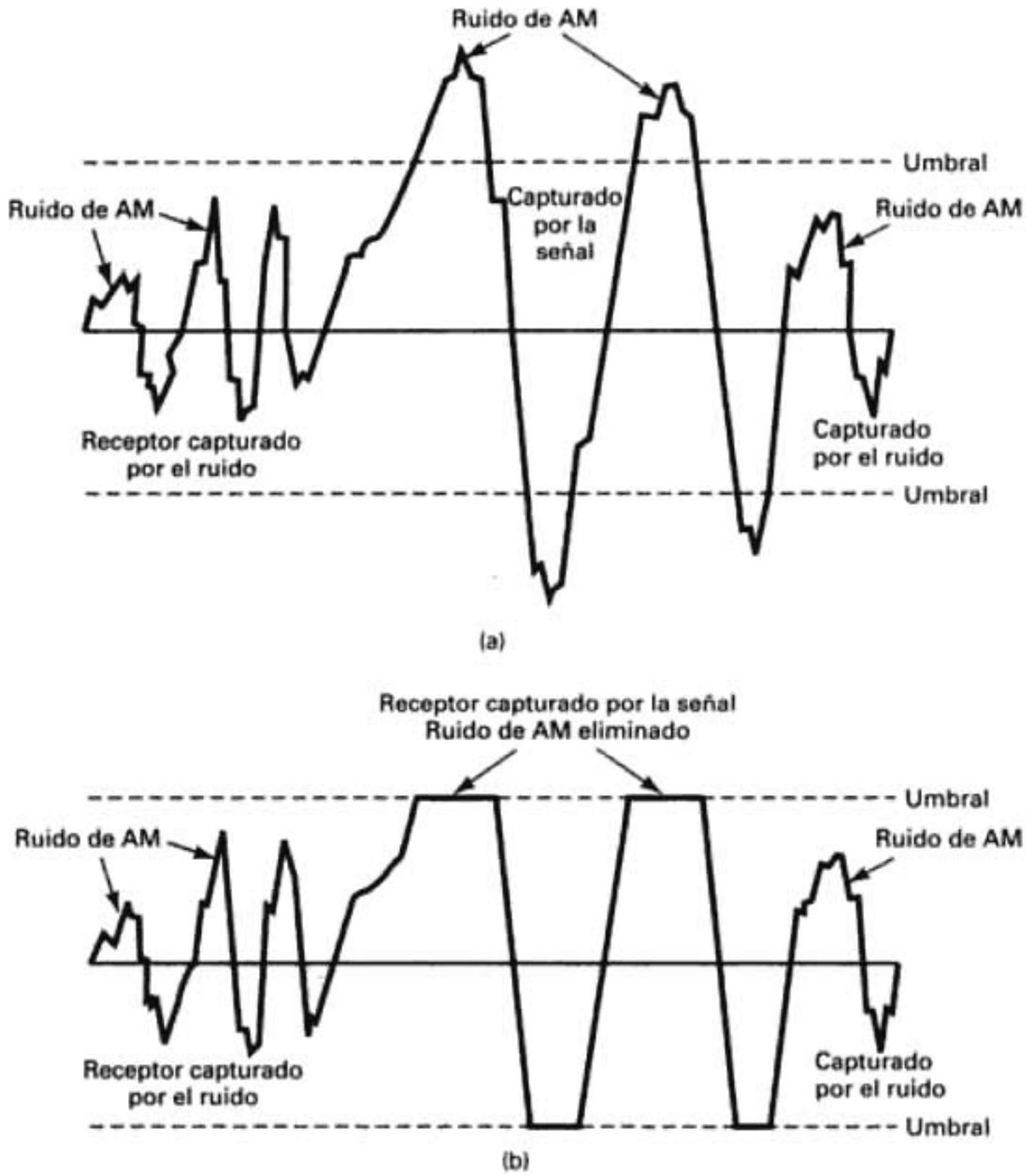


FIGURA 7-9 Formas de onda de entrada y salida del limitador de amplitud: (a) forma de onda de entrada; (b) forma de onda de salida

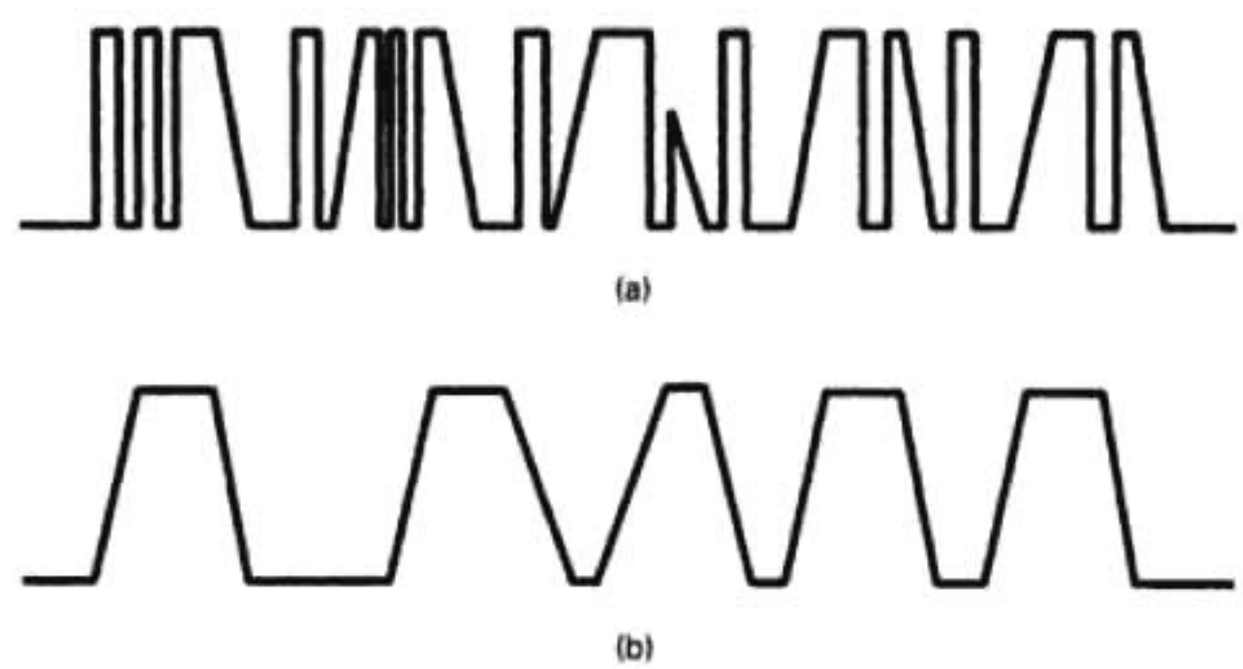


FIGURA 7-10 Salida del limitador: (a) capturado por el ruido; (b) capturado por la señal

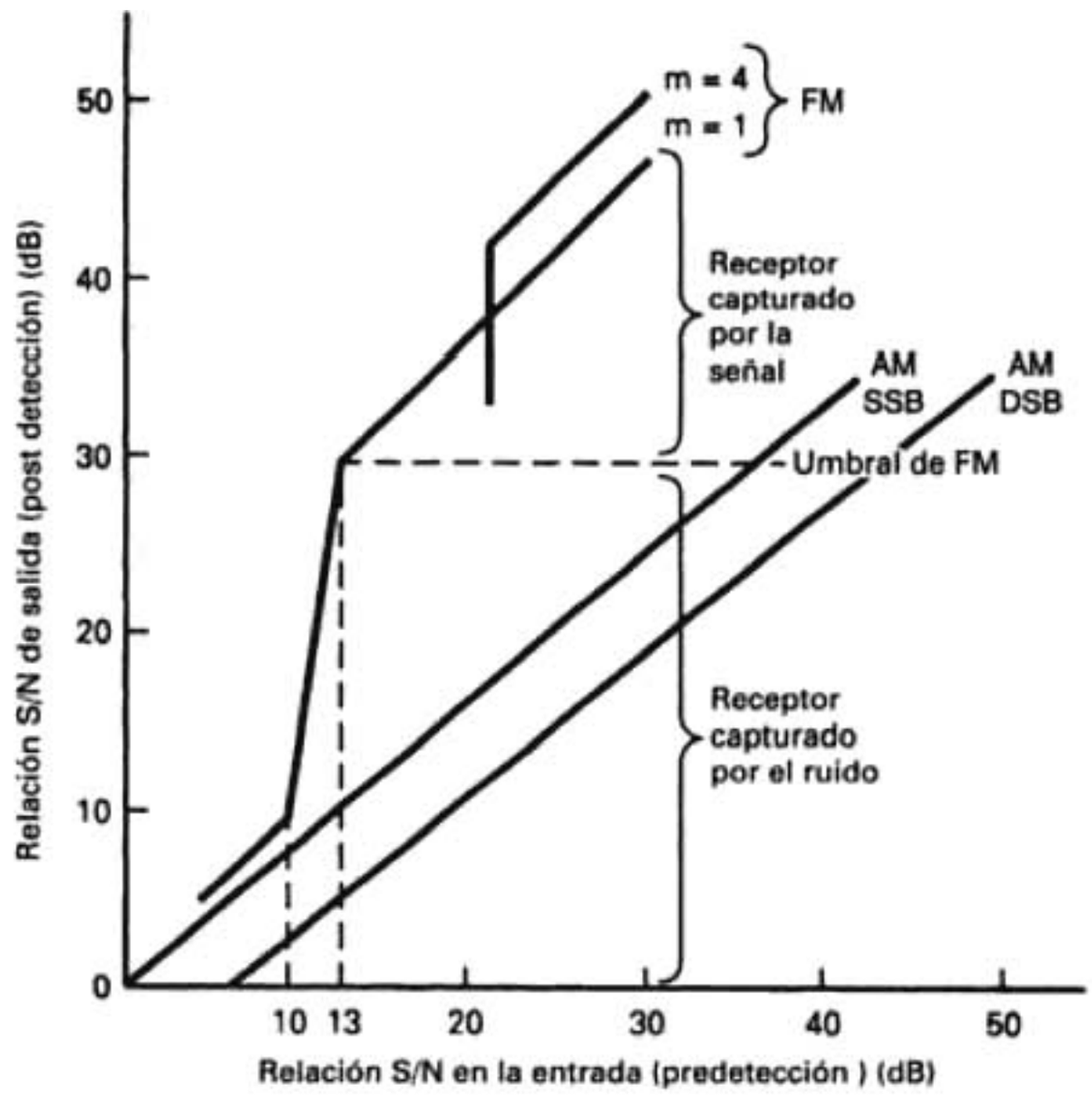


FIGURA 7-11 Umbral de FM

pero no produce la mejoría completa de 17 dB sino hasta que la relación de señal a ruido en la entrada llega a 13 dB.

Como se ve en la fig. 7-11, no hay mejoría en la relación de señal a ruido en la transmisión de AM por doble banda lateral o banda lateral única. En la AM, las relaciones de señal a ruido en predetección y post detección son iguales, en esencia.

Circuitos limitadores

La fig. 7-12a muestra el diagrama de un circuito limitador de una etapa con filtro incorporado de salida. A esta configuración se le suele llamar *limitador/amplificador pasabanda* (BPL, de *bandpass limiter/amplifier*). Un BPL es un amplificador de FI clase A polarizado y sintonizado, y para que haya limitación y silenciamiento de FM, requiere que la señal de entrada de FI sea suficiente para llevarlo tanto a la saturación como al corte. El circuito tanque de salida se sintoniza con la FI central. El filtrado elimina la distorsión por armónicas e intermodulación que hay en los pulsos rectangulares, debida a la *limitación dura*. El efecto del filtrado se ve en la fig. 7-13. Si se quitara todo el resistor R_2, el amplificador estaría polarizado para operación en clase C, que también es adecuada para esta clase de circuitos, pero requiere más filtrado. La fig. 7-12b muestra la acción limitadora del circuito de la fig. 7-12a. Para señales pequeñas, menores que el voltaje de umbral, no hay limitación. Cuando V_{ent} llega a V_{umbral}, comienza la limitación y para amplitudes mayores que $V_{máx}$, en realidad hay una disminución de V_{sal} al aumentar V_{ent}. Esto se debe a que con altos valores de excitación en la entrada, los pulsos de corriente del colector son lo bastante angostos como para desarrollar en realidad menos potencia en el circuito tanque. El problema de sobreexcitar el limitador se puede rectificar incorporando un control automático de ganancia (CAG) en el circuito.

Efecto de captura de FM

La capacidad inherente que tiene la FM para disminuir los efectos de las señales de interferencia se llama *efecto de captura*. A diferencia de los receptores de AM, los de FM tienen la capacidad de diferenciar entre dos señales recibidas con la misma frecuencia. Así, si se reciben dos estaciones en forma simultánea, con la misma o casi la misma frecuencia, el receptor se amarra a la estación más intensa y suprime la más débil. La supresión de la señal más débil se logra en

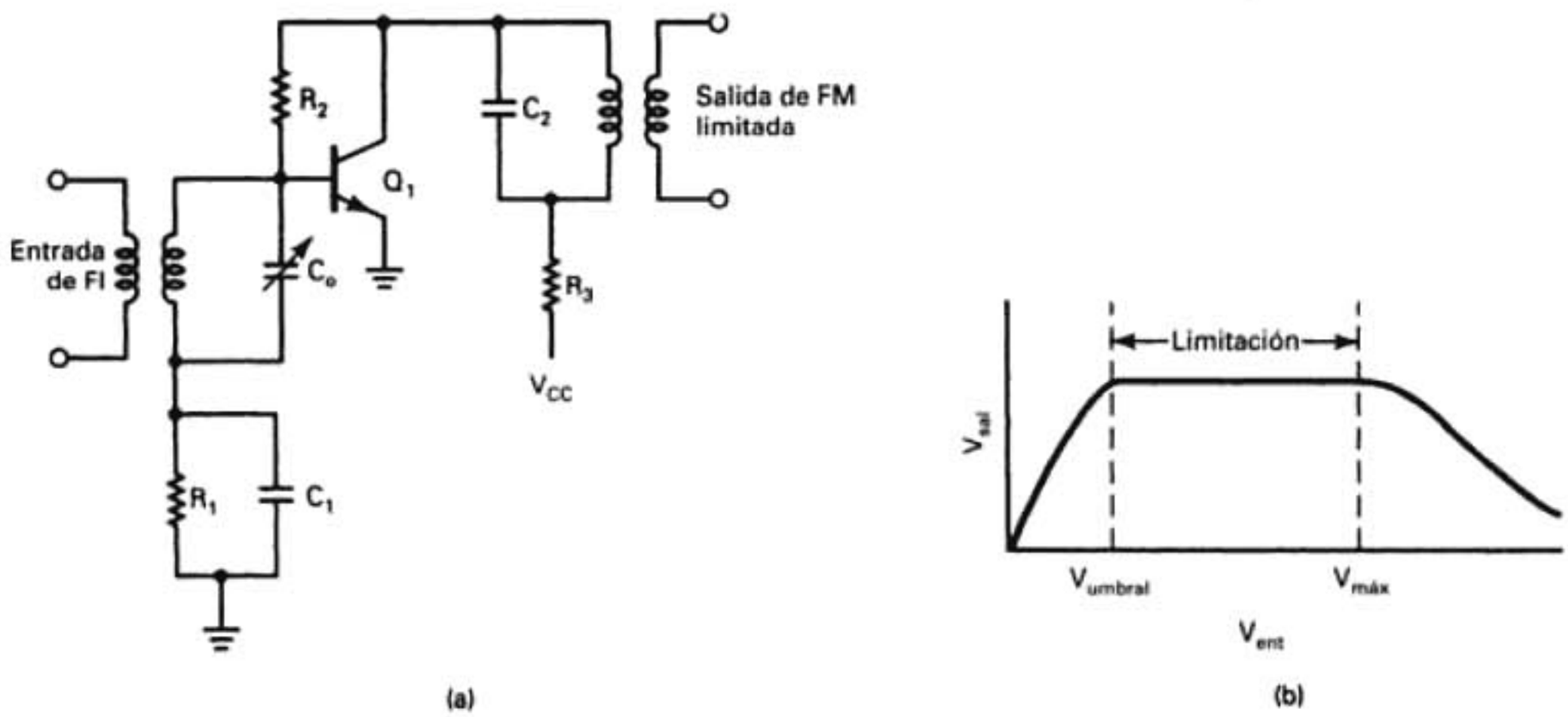


FIGURA 7-12 Limitador sintonizado de una etapa: (a) diagrama; (b) acción del limitador

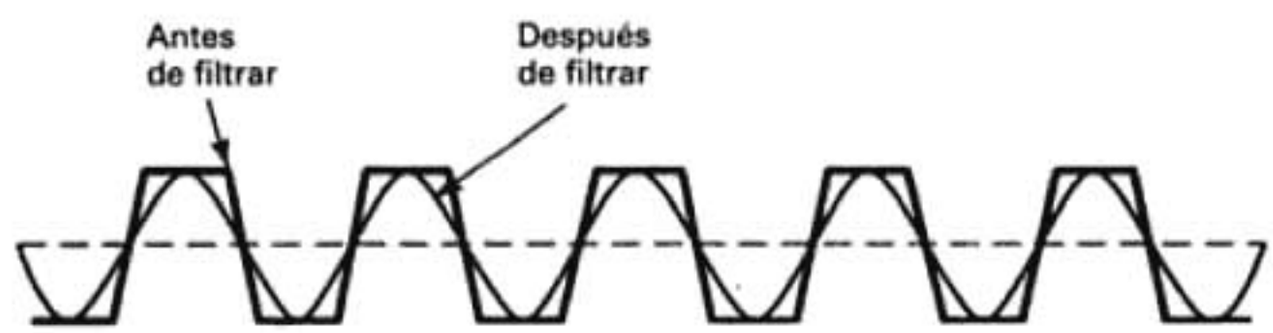


FIGURA 7-13 Salida filtrada del limitador

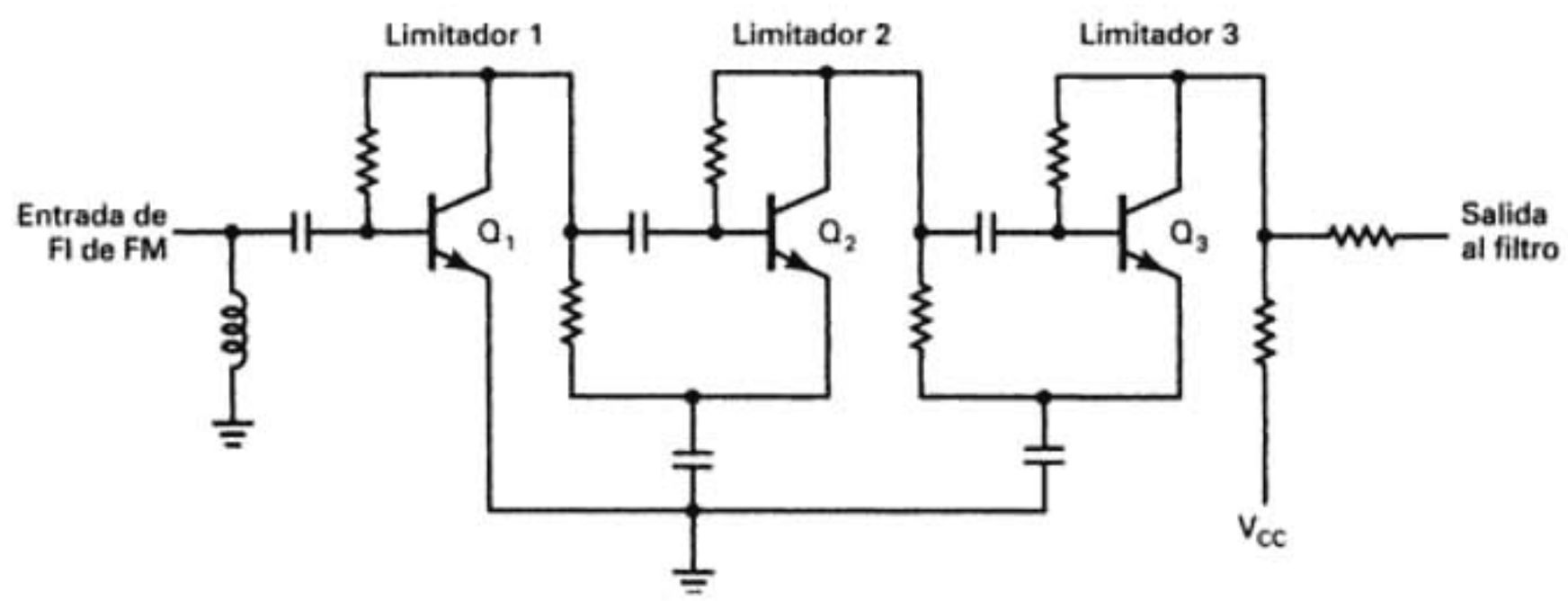


FIGURA 7-14 Limitador de tres etapas en cascada

los limitadores de amplitud, de la misma manera que se suprime el ruido de AM. Si se reciben dos estaciones con más o menos el mismo nivel de señal, el receptor no puede diferenciar lo suficiente entre ellas, y se puede cambiar de una a otra y viceversa. La *relación de captura* de un receptor de FM es la diferencia mínima, en dB, de intensidades de dos señales recibidas, necesaria para que el efecto de captura suprima a la señal más débil. Las relaciones de captura de 1 dB son características en los receptores de FM de alta calidad.

Cuando se usan dos etapas limitadoras se tiene *doble limitación*; con tres etapas, *triple limitación*, etcétera. La fig. 7-14 muestra un *limitador en cascada* de tres etapas, sin filtro incorporado. Un circuito limitador de esta clase debe estar seguido por un filtro de cerámica o de

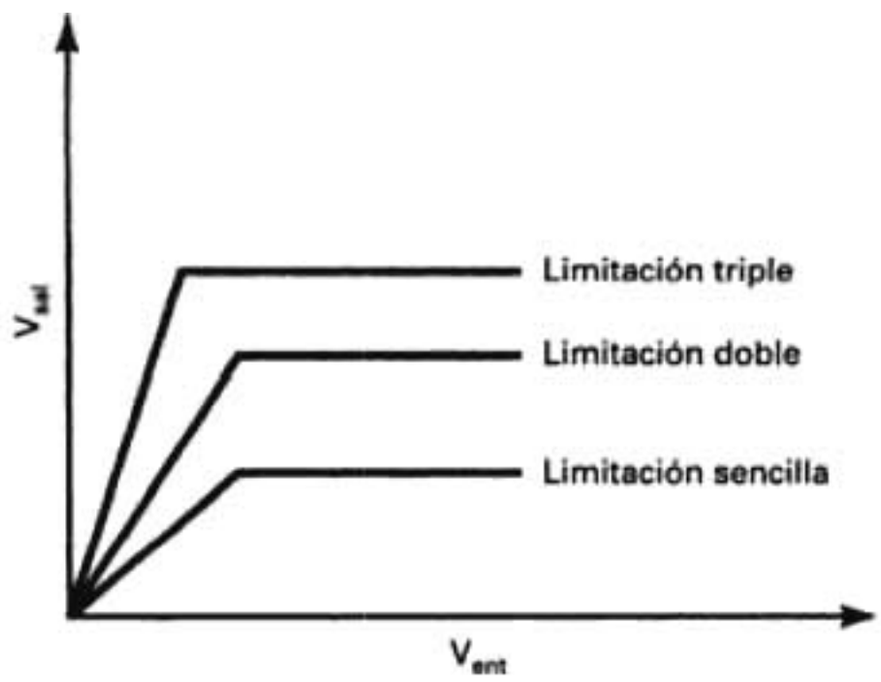


FIGURA 7-15 Curvas de respuesta del limitador

cristal, para eliminar la distorsión no lineal. El limitador que se ve tiene tres etapas limitadoras acopladas *RC*, que están conectadas en serie respecto a la cd para reducir el drenaje de corriente. Los amplificadores en cascada combinan varias de las ventajas de los amplificadores de emisor común y de compuerta común. Al conectar en cascada los amplificadores también disminuye el nivel de umbral y, en consecuencia, se mejoran las posibilidades de silenciamiento de la etapa. En la fig. 7-15 se muestran los efectos de doble y triple limitación. Como los receptores de FM tienen la ganancia suficiente para saturar los limitadores dentro de márgenes relativamente amplios de nivel de señal de RF de entrada, el CAG suele ser innecesario. De hecho, en realidad y con frecuencia el CAG degrada el desempeño de un receptor de FM.

Ejemplo 7-2

Calcular la potencia mínima de portadora en recepción, necesaria para alcanzar una relación de señal a ruido de 37 dB en post detección, para un receptor de FM con ancho de banda $B = 200$ kHz, cifra de potencia de ruido NF = 8 dB y temperatura de ruido de entrada $T = 100$ K. Usar el diagrama de bloques del receptor de la fig. 7-1 como modelo, y la curva de limitación de picos de FM de la figura 7-11, para $m = 1$.

Solución Se ve en la fig. 7-11 que es evidente una mejoría de señal a ruido de 17 dB en el detector, suponiendo que los limitadores estén saturados, y que la relación de señal a ruido en la entrada es mayor que 13 dB. En consecuencia, para lograr una relación de señal a ruido de 37 dB en post detección, la correspondiente en predetección debe ser, como mínimo

$$37 \text{ dB} - 17 \text{ dB} = 20 \text{ dB}$$

Por consiguiente, para una cifra de ruido general del receptor de 8 dB, la relación de S/N en la entrada del mismo debe ser, al menos

$$20 \text{ dB} + 8 \text{ dB} = 28 \text{ dB}$$

La potencia de ruido en la entrada del receptor es

$$N_{(\text{dBm})} = 10 \log \frac{KTB}{0.001} = 10 \log \frac{(1.38 \times 10^{-23})(100)(200{,}000)}{0.001} = -125.6 \text{ dBm}$$

En consecuencia, la potencia mínima de señal en el receptor, para tener una relación de S/N de 28 dB es

$$S = -125.6 \text{ dBm} + 28 \text{ dB} = -97.6 \text{ dBm}$$

COMPARACIÓN DE MODULACIÓN DE FRECUENCIA Y MODULACIÓN DE FASE

Aunque en muchos aspectos se parecen la modulación de frecuencia y la modulación de fase, sí hay diferencias entre ellas y, en consecuencia, hay ventajas y desventajas en ambas formas de modulación angular. Alguna vez se prefirió la FM para aplicaciones en gran escala, como emisión comercial, porque la PM requiere una demodulación coherente, casi siempre mediante un lazo de fase cerrada. Por otra parte, en la FM se puede demodular usando demoduladores no coherentes. Sin embargo, hoy es probable que los PLL sean menos costosos que sus contrapartes no coherentes, principalmente porque se consiguen como circuitos integrados y no requieren transformadores ni circuitos tanque *LC*.

En la PM, el índice de modulación es independiente de la frecuencia de la señal moduladora. Así, la PM ofrece mejor desempeño de señal a ruido que la FM, y la PM no requiere una red de preénfasis. Otra ventaja importante de la PM es que la modulación de fase se hace en un circuito separado del oscilador de portadora. En consecuencia, se pueden usar osciladores de cristal muy estables como fuente de portadora. Con la FM, la señal moduladora se aplica en forma directa al oscilador de portadora, por lo que no se pueden usar osciladores de cristal para producir la señal de portadora. Por lo anterior, los moduladores de FM requieren circuitos de control automático de frecuencia para lograr la estabilidad de frecuencia que dicta la FCC.

Una ventaja notable de la FM sobre la PM es que los VCO que se usan en la FM se pueden modular en forma directa, y producir salidas con grandes desviaciones de frecuencia y altos índices de modulación. En general, los moduladores de PM requieren multiplicadores de frecuencia para aumentar el índice de modulación y la desviación de frecuencia hasta niveles útiles.

RECEPTORES DE FM EN CIRCUITO INTEGRADO LINEAL

En los últimos años varios fabricantes de circuitos integrados, como Signetics, RCA y Motorola, han desarrollado circuitos integrales monolíticos de baja potencia que hacen virtualmente todas las funciones del receptor, para sistemas de comunicaciones de AM y de FM a la vez. Estos circuitos integrados tienen las ventajas de ser confiables, predecibles, miniaturizados y de fácil diseño. El desarrollo de esos circuitos integrados es una de las razones principales del abrumador crecimiento de las radiocomunicaciones de FM en dos sentidos y las celulares, en los últimos años.

Sistema de FI para FM en circuito integrado y baja potencia

El CI NE/SA614A es un sistema monolítico mejorado, de FI para FM de baja potencia, fabricado por Signetics Corporation. Es un dispositivo de alta ganancia y alta frecuencia que tiene bajo consumo de energía (3.3 mA de consumo normal de corriente) y una sensibilidad excelente de entrada (1.5 μV entre terminales) a 455 kHz. El NE/SA614A tiene un *indicador de intensidad de señal recibida* (RSSI, *received signal-strength indicator*) incorporado, compensado térmicamente y con salida logarítmica e intervalo dinámico mayor que 90 dB. Tiene dos salidas de audio, una enmudecida y la otra no. Requiere pocos componentes externos para funcionar, y cumple con las especificaciones de radio celular. Se puede usar para las siguientes aplicaciones:

1. Radio celular de FM.
2. Receptores de comunicaciones de FM de alto rendimiento.
3. Amplificación y detección de frecuencia intermedia hasta de 25 MHz.
4. Medidor de intensidad de señal de RF.
5. Aplicaciones de analizador de espectro.
6. Circuitos de instrumentación.
7. Receptores y transmisores de datos.

En la fig. 7-16 se ve el diagrama de bloques del NE/SA614A. Incluye dos amplificadores limitadores de frecuencia intermedia, un detector de cuadratura de FM, un circuito de silenciamiento de audio, un indicador logarítmico de intensidad de señal recibida (RSSI) y un regulador de voltaje. Es un sistema de procesamiento de señal de FI, adecuado para frecuencias hasta de 21.4 MHz.

Amplificadores de FI. La fig. 7-17 muestra el circuito equivalente del NE/SA614A. La sección de amplificación de FI consiste en dos etapas de amplificación con limitación logarítmica. La primera está formada por dos amplificadores diferenciales con 39 dB de ganancia, y ancho de banda a pequeña señal de ca, de 41 MHz cuando se activa con una fuente de 50 Ω. La salida del primer limitador es un seguidor de emisor, de baja impedancia, con resistencia en serie equivalente de 1 kΩ. La segunda etapa limitadora consiste en tres amplificadores diferenciales con una ganancia total de 62 dB y un ancho de banda a pequeña señal de ca de 28 MHz. Las salidas del amplificador diferencial final pasan por un separador hacia el detector interno de

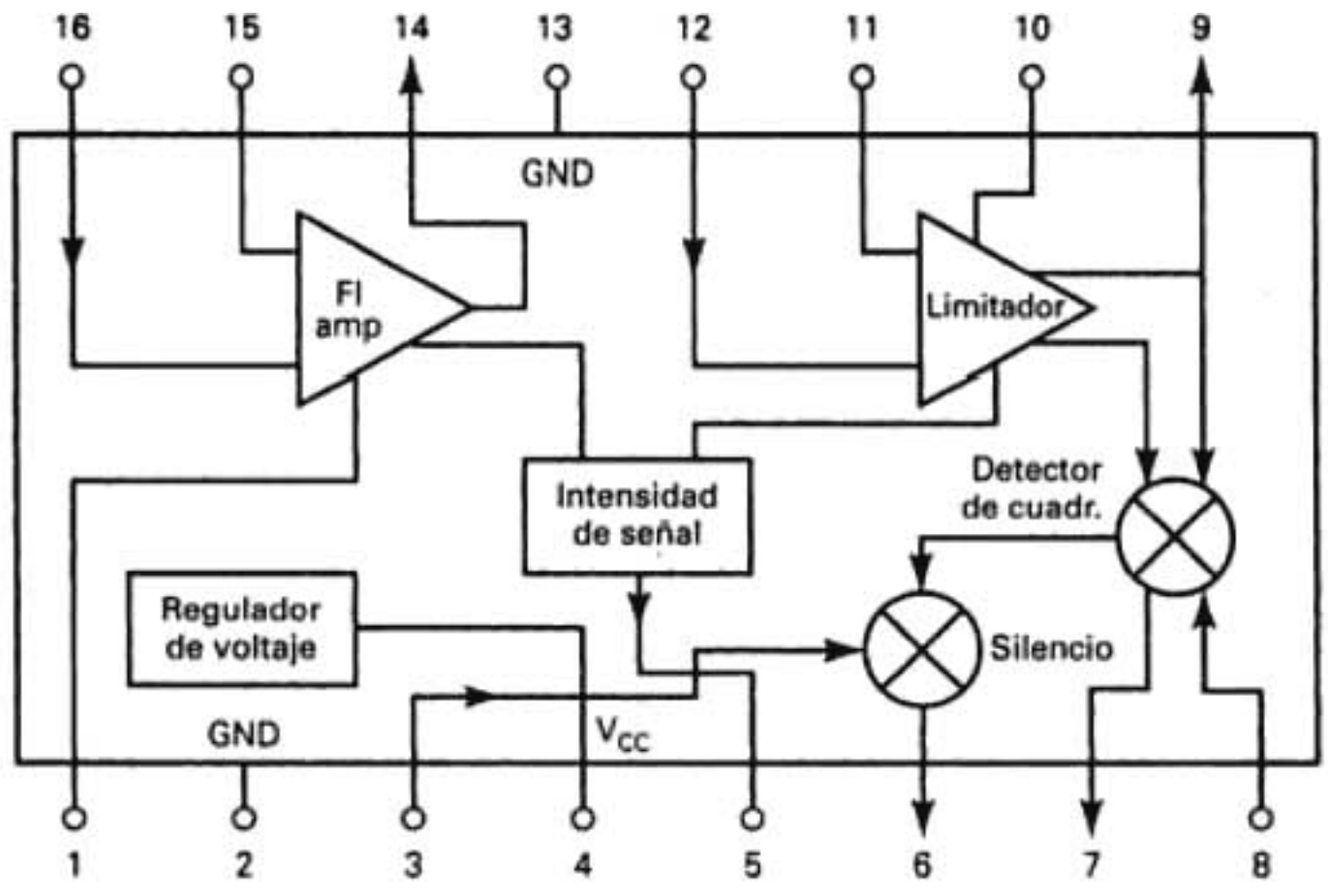


FIGURA 7-16 Diagrama de bloques del sistema de FI para FM en circuito integrado y baja potencia Signetics NE/SA614A

cuadratura. Una está disponible para excitar un capacitor externo de cuadratura y tanque de cuadratura *L/C*. Las dos etapas limitadoras están polarizadas con cd con retroalimentación. Las salidas del amplificador diferencial final de cada etapa se retroalimentan, después de pasar por los separadores, a la entrada de esa etapa a través de un resistor de 42 kΩ. Debido a la ganancia tan alta, al gran ancho de banda y a la alta impedancia de entrada de los limitadores, la etapa de limitación es potencialmente inestable a frecuencias intermedias mayores que 455 kHz. Se puede mejorar la estabilidad reduciendo la ganancia. Esto se logra agregando atenuadores entre las etapas de amplificación. Los amplificadores de FI también tienen bajo desplazamiento de fase (normalmente de unos pocos grados dentro de un amplio margen de frecuencias de entrada).

Detector de cuadratura. La fig. 7-18 muestra el diagrama de bloques del circuito equivalente del detector de cuadratura en el NE/SA614A. Un detector de cuadratura es una celda multiplicadora parecida a una etapa mezcladora, pero en lugar de mezclar dos frecuencias distintas, mezcla dos señales con la misma frecuencia, pero con distintas fases. Una señal de amplitud constante (amplitud limitada) se aplica a la parte inferior del multiplicador. La misma señal (en modo no diferencial) se aplica a un capacitor externo conectado con la terminal 9. Hay un desplazamiento de fase de 90° a través de las placas del capacitor. La señal con fase desplazada se aplica a la conexión superior del multiplicador, en la terminal 8. Un tanque de cuadratura, que es una red *LC* paralela, permite un desplazamiento de fase selectivo, dependiente de la frecuencia, en la señal de FI. El detector de cuadratura produce una señal de salida cuya amplitud es proporcional a la magnitud de desviación de frecuencia de la señal de FM de entrada.

Salidas de audio. El NE/SA614A tiene dos salidas de audio. Ambas son convertidores PNP de corriente a voltaje, con cargas internas nominales de 55 kΩ. La salida no silenciada siempre está activa, para permitir el uso de tonos de señal, como por ejemplo, para radio celular. La otra salida se puede silenciar con atenuación normal de 70 dB. Las dos salidas tienen una diferencia interna de fase de 180° y, en consecuencia, se pueden aplicar a las entradas diferenciales de un amplificador operacional o comparador. Una vez establecido el umbral de la frecuencia de referencia, las dos amplitudes de salida se desplazarán en direcciones opuestas cuando se mueva la frecuencia de entrada.

RRSI. El RSSI o indicador de intensidad de señal recibida tiene una salida monótona y logarítmica dentro de un intervalo de 90 dB. La relación de señal a ruido en la salida se deriva de la suma de corrientes en la etapa de los amplificadores limitadores. En esencia es independiente de la FI. Así, las señales no filtradas en la entrada del limitador, como por ejemplo productos

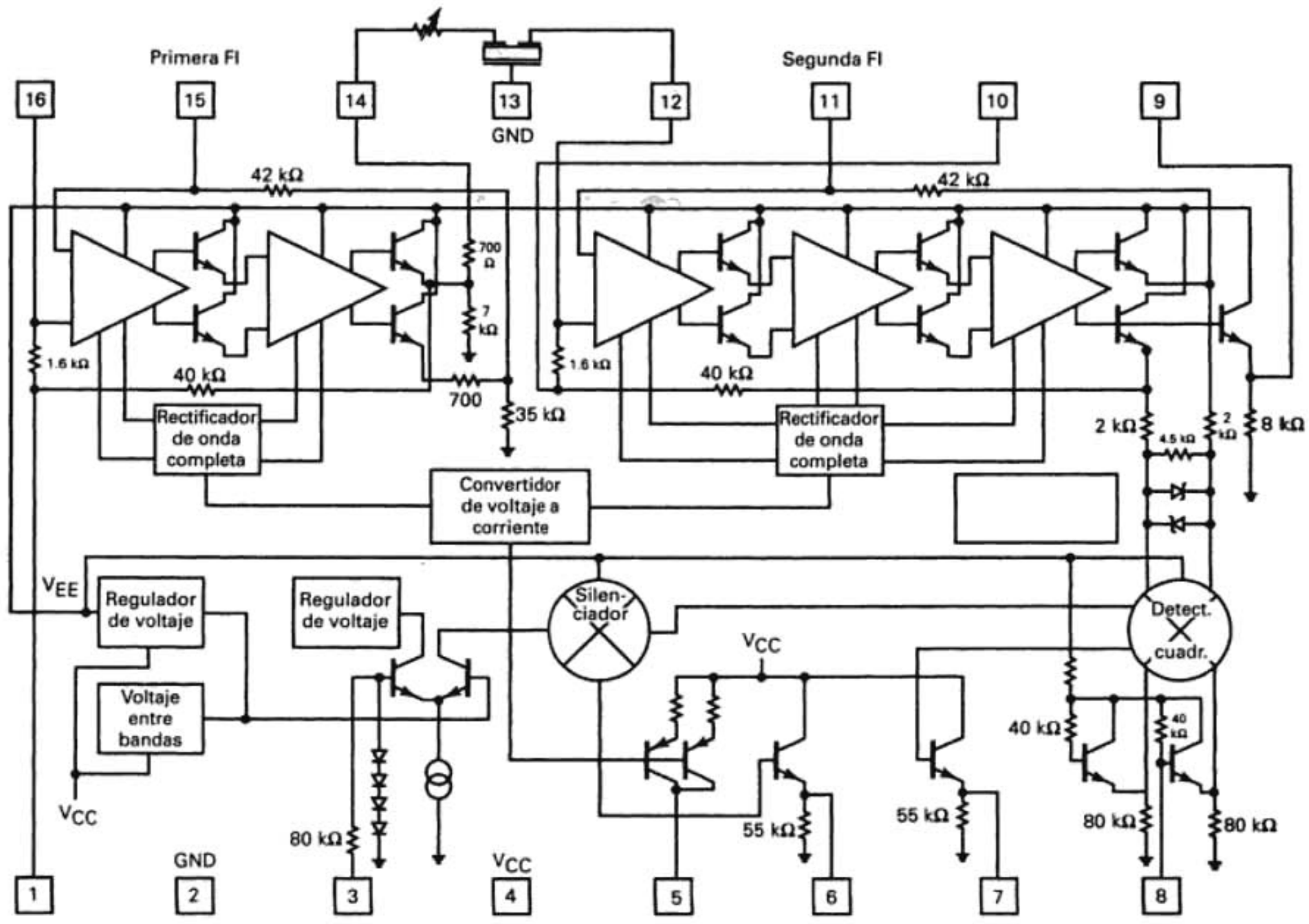


FIGURA 7-17 Circuito equivalente del sistema de FI para FM en circuito integrado y baja potencia Signetics NE/SA614A

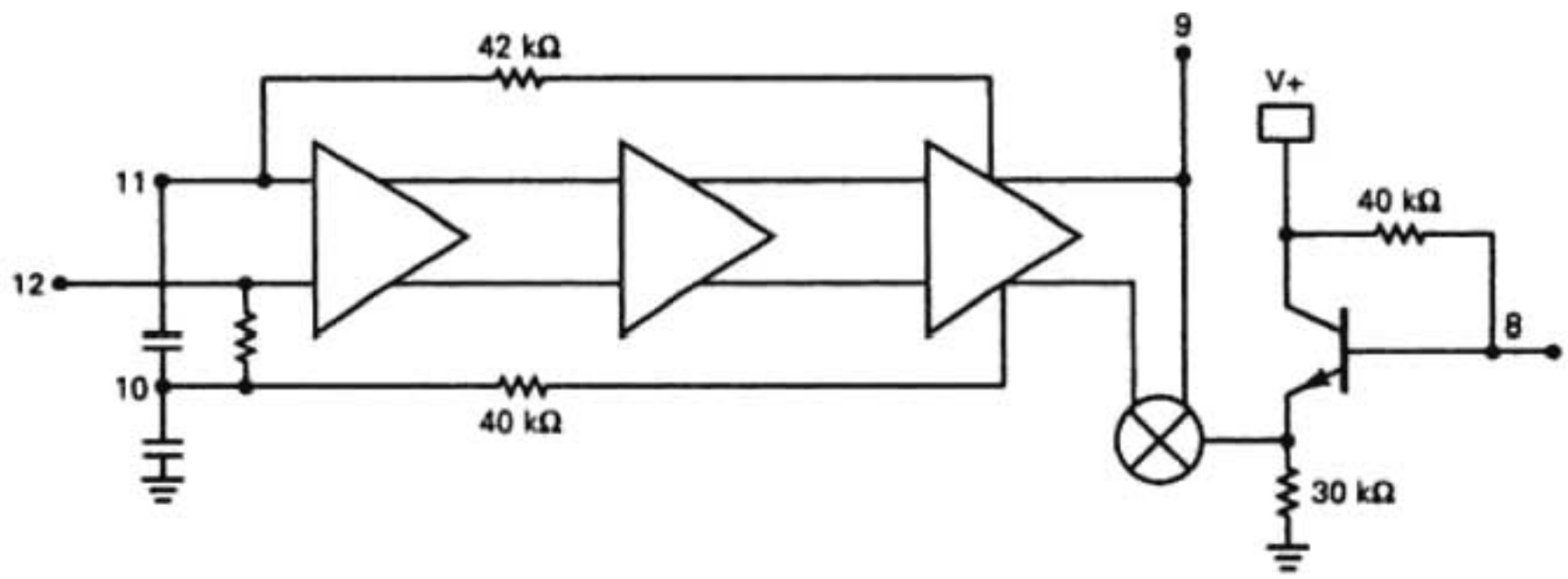


FIGURA 7-18 Diagrama de bloques de un detector de cuadratura

espurios o señales regeneradas, se manifestarán como salida de RSSI. A bajas frecuencias el RSSI es un excelente voltímetro logarítmico para ca. La salida del RSSI es un convertidor de corriente a voltaje parecido a las salidas de audio.

Sistema de FI para FM de alto rendimiento, bajo voltaje y con mezclador

El NE/SA616 es un sistema monolítico de bajo voltaje, de FI para FM de alto rendimiento parecido al NE/SA614A, excepto por la adición de un circuito mezclador/oscilador. El NE/SA616 trabaja con frecuencias hasta de 150 MHz y con 2.7 V cd. Tiene bajo consumo de energía, una ganancia de conversión de potencia en mezclador de 17 dB a 45 MHz, 102 dB de ganancia de amplificador/limitador de FI y un ancho de banda a pequeña señal de ca de 2 MHz en el amplificador/limitador de FI. Se puede usar en las siguientes aplicaciones:

1. Radio celular FM portátil.
2. Teléfonos inalámbricos.
3. Sistemas de comunicaciones inalámbricas.
4. Medidor de intensidad de señal de RF.
5. Análisis de espectro.
6. Circuitos de instrumentación.
7. Transceptores de datos.
8. Amplificadores logarítmicos.
9. Receptores VHF de una conversión.

El diagrama de bloques del NE/SA616 se ve en la fig. 7-19. Es similar al NE/SA614A pero tiene un mezclador y una etapa de oscilador local. La etapa de entrada es un mezclador de celda de Gilbert con un oscilador. Las características normales del mezclador incluyen una cifra de ruido de 6.2 dB, ganancia de conversión de 17 dB y una entrada para intersección de tercer orden de −9 dBm. El oscilador trabajará por arriba de 200 MHz en una configuración de circuito tanque LC. La impedancia de salida del mezclador es un resistor de 1.5 kΩ, que permite la conexión directa con un filtro de cerámica de 455 kHz. El amplificador de FI tiene 43 dB de ganancia y 5.5 MHz de ancho de banda. El limitador de FI tiene 60 dB de ganancia y 4.5 MHz de ancho de banda. El detector de cuadratura también usa una celda de Gilbert. Una conexión de la celda es activada internamente por la señal de FI y la otra salida de la FI se acopla en ca a una red sintonizada de cuadratura, donde sufre un desplazamiento de fase de 90° antes de retroalimentarse a la otra conexión de la celda de Gilbert. La salida del demodulador en el detector de cuadratura excita un amplificador operacional interno. Este último se puede configurar como separador de ganancia unitaria, o para ganancia, filtrado y compensación de temperatura de segundo orden en forma simultánea, si es necesario.

Sistema de radio de FM con un solo CI

El TDA7000 es un sistema de radio de FM en circuito integrado, fabricado por Signetics Corporation para radios portátiles de FM. En esencia es un radiorreceptor completo sobre un solo

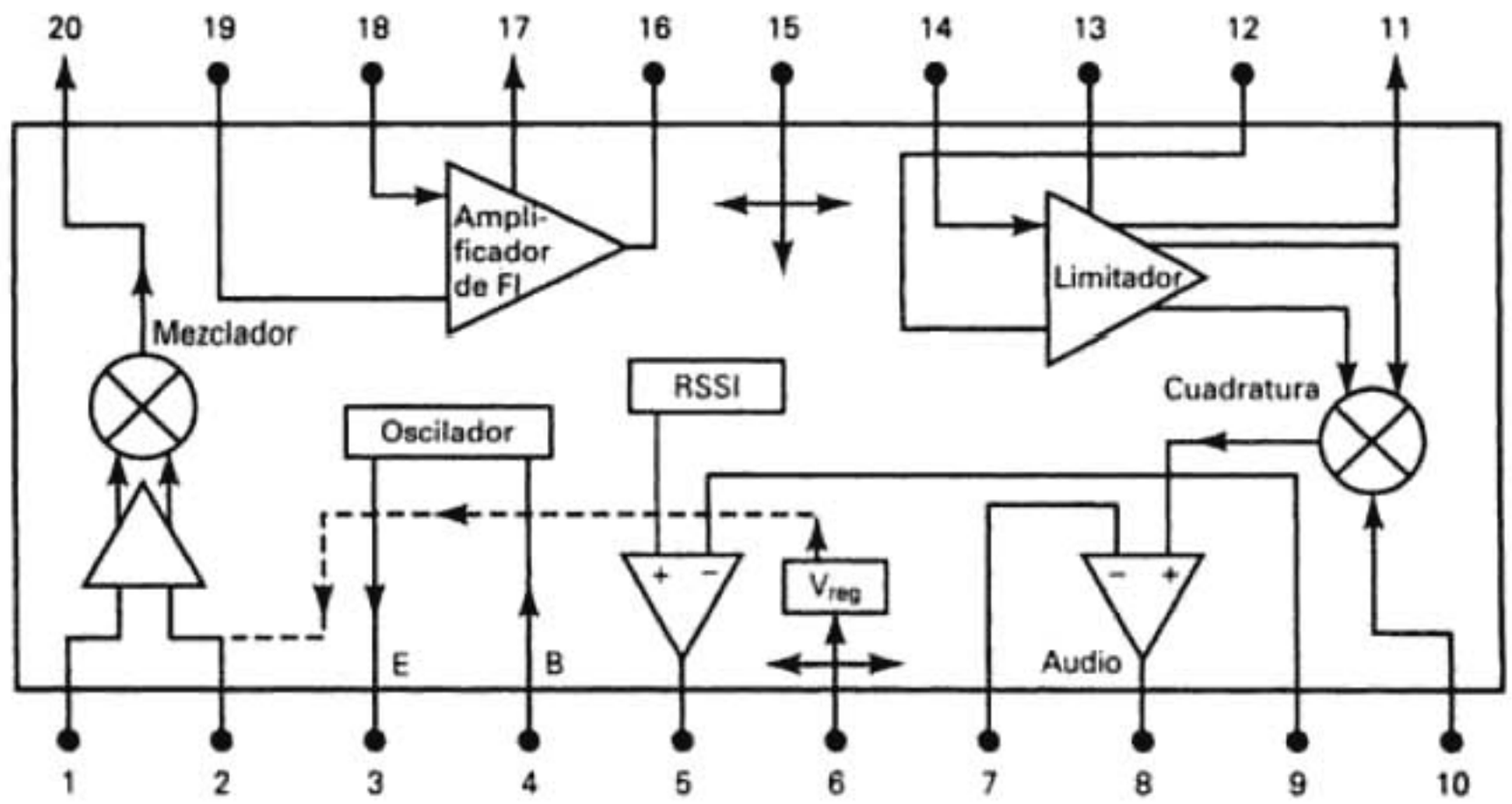


FIGURA 7-19 Diagrama de bloques del sistema monolítico de FI para FM Signetics NE/SA616

circuito integrado. El TDA7000 tiene pequeño tamaño, no tiene bobinas de FI, se ensambla con facilidad y tiene bajo consumo de energía. Aparte del CI sólo hay un circuito tanque *LC* sintonizable para el oscilador local, unos pocos capacitores de cerámica, poco costosos, y un resistor. Al usarlo, un radio completo de FM puede tener el tamaño suficientemente pequeño como para caber dentro de una calculadora, un encendedor para cigarrillos, un llavero o hasta un reloj plano. También se puede usar en equipos como teléfonos inalámbricos, modelos a control remoto, sistemas de localización o como canal de sonido de un receptor de TV.

El diagrama de bloques del TDA7000 está en la fig. 7-20. Comprende los siguientes bloques funcionales: etapa de entrada de RF, mezclador, oscilador local, limitador de amplitud de FI, demodulador de fase, detector de silenciamiento e interruptor de silenciamiento. El CI tiene un sistema FLL (lazo de frecuencia cerrada) interno con una frecuencia intermedia de 70 MHz. El FLL es para reducir la distorsión armónica total (THD) comprimiendo la variación (desviación) de la FI. Esto se logra usando una salida de audio del demodulador de FM para desplazar la frecuencia del oscilador local en sentido contrario al de la desviación de la FI. La idea es comprimir 75 kHz de desviación de frecuencia hasta unos 15 kHz. Esto limita la distorsión armónica total a 0.7% con desviación de ± 22.5 kHz, y a 2.3% con desviación de ± 75 kHz. La selectividad en FI se obtiene con filtros activos *RC* de Sallen-Key. La única función que necesita ajuste es el circuito de resonancia del oscilador.

EMISIÓN ESTEREOFÓNICA EN FM

Hasta 1961, todas las transmisiones comerciales de FM eran *monofónicas*. Es decir, un solo canal de audio de 50 Hz a 15 kHz formaba todo el espectro de frecuencias de información de voz y música. Este único canal de audio modulaba a una portadora de alta frecuencia, y era transmitido a través de un canal de comunicaciones de FM con ancho de banda de 200 kHz. Con la transmisión monofónica cada conjunto de bocinas en el receptor reproduce exactamente la misma información. Es posible separar las frecuencias de información con bocinas especiales como *bocinas de graves* para bajas frecuencias y *bocinas de altos* para altas frecuencias. Sin embargo, es imposible separar *espacialmente* el sonido monofónico. Toda la señal de información suena como si viniera de la misma dirección, es decir, de una *fuente puntual*, sin que el sonido tenga direccionalidad. En 1961, la FCC autorizó la transmisión *estereofónica* en banda comercial de FM. En la transmisión estereofónica, la señal de información se divide espacialmente en dos canales de audio, de 50 Hz a 15 kHz: uno derecho y uno izquierdo. La música que se originó en el lado izquierdo sólo se reproduce en la bocina izquierda, y la que se originó en el lado derecho

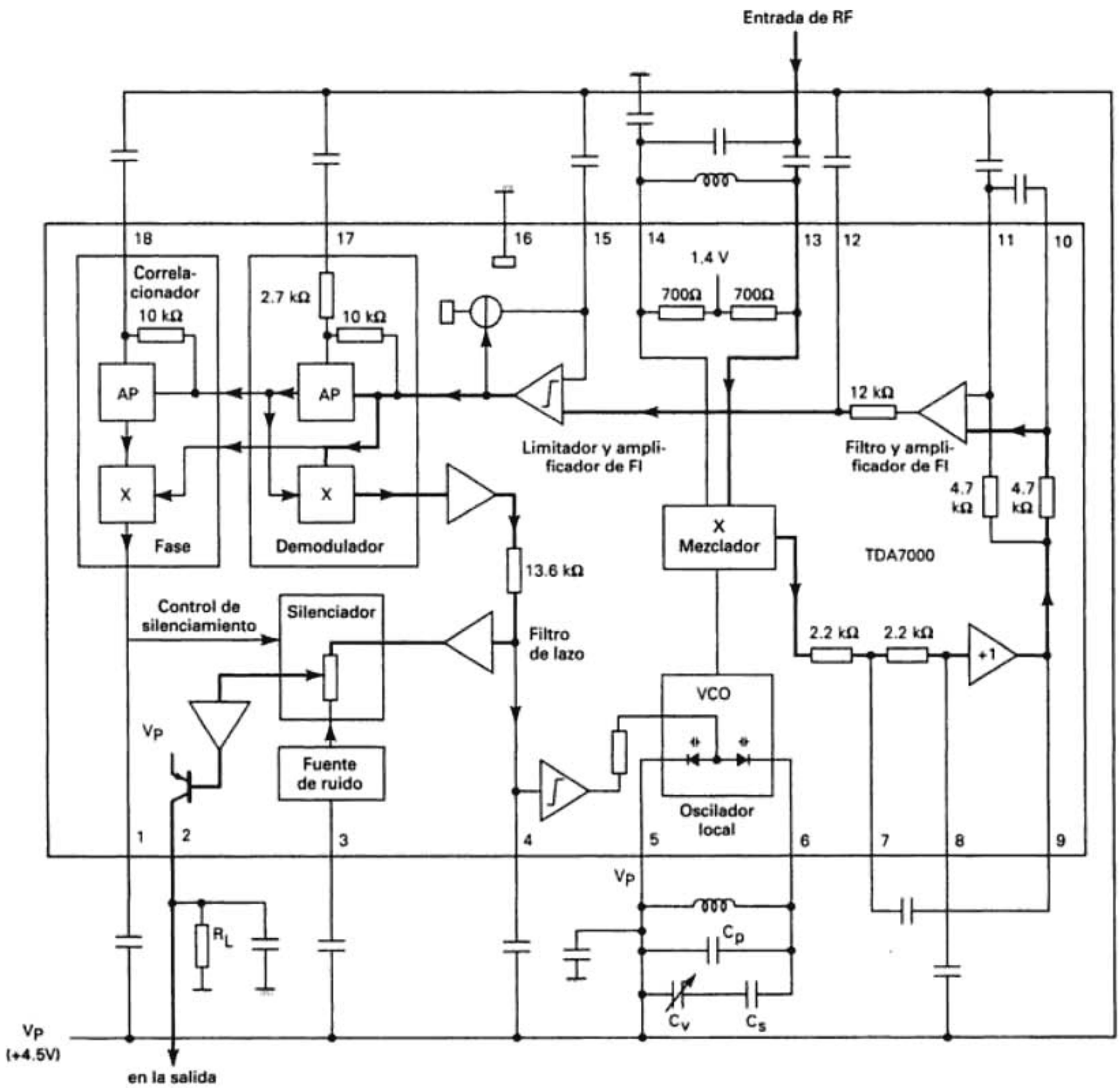


FIGURA 7-20 Diagrama de bloques del radio de FM de circuito integrado Signetics TDA7000

sólo se reproduce en la bocina derecha. Así, con la transmisión estereofónica es posible reproducir música con una direccionalidad y dimensión espacial exclusivas, que antes sólo era posible con la fuente viva, es decir, con una fuente *extendida*. También, en la transmisión estereofónica es posible separar música o sonido por *calidad tonal*, por ejemplo, percusiones, cuerdas, vientos, etc.

Uno de los cuidados principales de la FCC antes de autorizar la transmisión estereofónica era su compatibilidad con los receptores monofónicos. La transmisión estereofónica no debería afectar la recepción monofónica. También, los receptores monofónicos deben poder recibir transmisiones estereofónicas como monoaurales, sin degradación perceptible de la calidad del programa. Además, los receptores estereofónicos deben recibir la programación estereofónica con separación casi perfecta, de 40 dB o más, entre sus canales izquierdo y derecho.

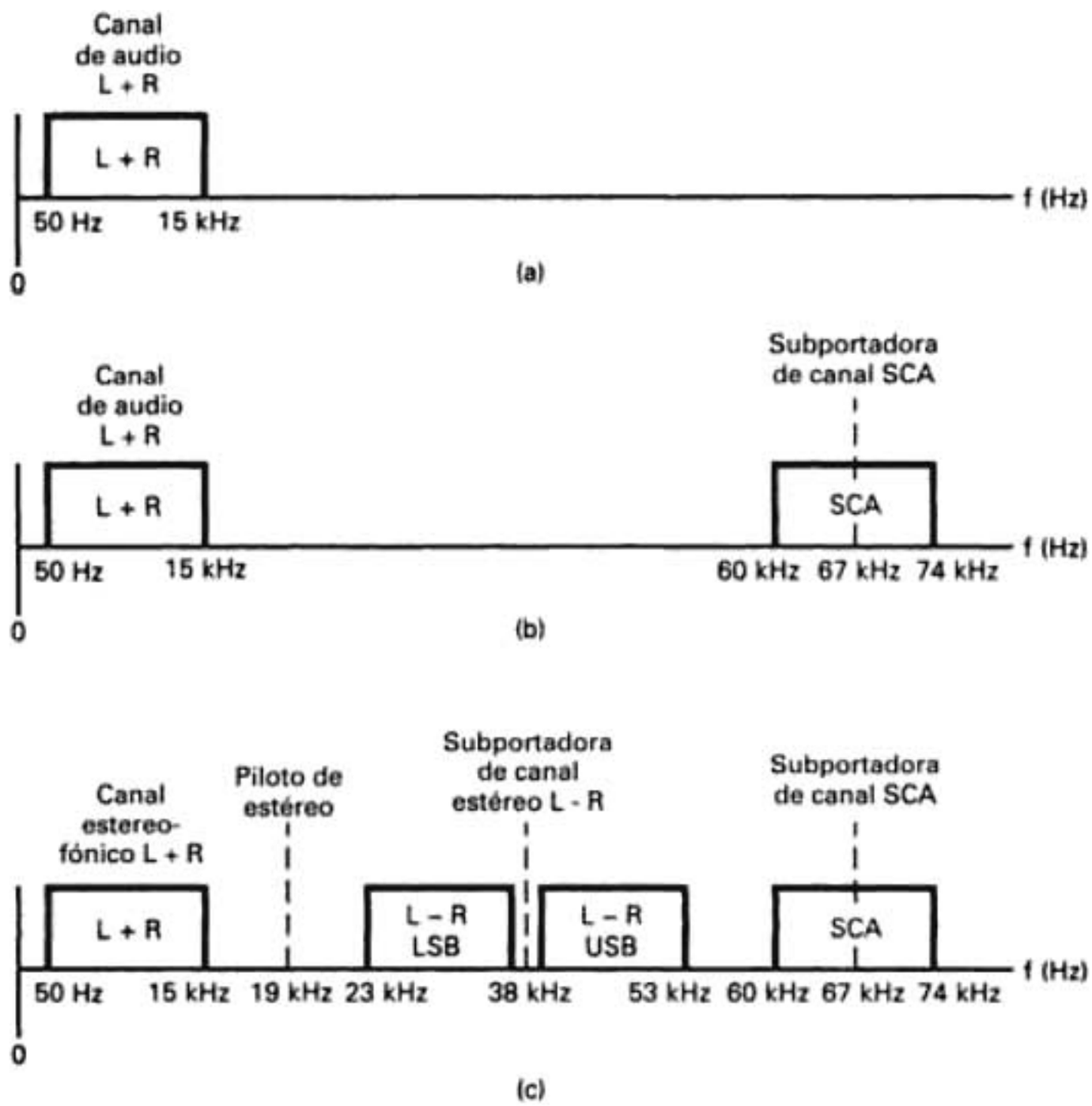


FIGURA 7-21 Espectro de banda base de FM: (a) antes de 1955; (b) antes de 1961; (c) a partir de 1961

El espectro de audio original de FM se ve en la fig. 7-21a. El canal de audio iba de 50 Hz a 15 kHz. En 1955, la FCC aprobó transmisión por subportadora con la Autorización de Comunicaciones Subsidiarias (SCA, *Subsidiary Communications Authorization*). Se usa la SCA para transmitir música ininterrumpida a suscriptores privados, como por ejemplo tiendas de departamentos, restaurantes y consultorios médicos que cuenten con los receptores SCA especiales. Es la música que a veces se le llama "música ambiental". Al principio, la subportadora SCA iba de 25 kHz a 75 kHz, pero se ha normalizado en 67 kHz. La subportadora y sus bandas laterales asociadas son parte de la señal total que modula a la portadora principal. En el receptor, la subportadora se demodula junto con el canal primario, pero no se puede oír debido a su alta frecuencia. El proceso de poner dos o más canales independientes uno junto a otro en el dominio de la frecuencia (apilar los canales) para modular entonces una sola portadora de alta frecuencia con la señal combinada se llama *multiplexado por división de frecuencia* (FDM, por *frequency division multiplexing*). En la emisión estereofónica de FM se multiplexan tres canales de voz o de música por división de frecuencia sobre una sola portadora de FM. La fig. 7-21b muestra el espectro total de frecuencias de la banda base para transmisión de FM antes de 1961. La banda base compuesta abarca el espectro total de frecuencias de la señal moduladora. El canal primario de audio quedó en 50 Hz a 15 kHz, mientras que un canal SCA adicional tiene su frecuencia trasladada a la banda de 50 kHz a 74 kHz. La subportadora SCA puede ser una transmisión de AM de banda lateral única o doble, o de FM con una frecuencia máxima moduladora de 7 kHz. Sin embargo, la modulación SCA de la portadora principal es FM de banda angosta y bajo índice y, en consecuencia, es una transmisión con mucho menor calidad que el canal primario de FM. La desviación total de frecuencia quedó en 75 kHz, con 90% (67.5 kHz) reservados al canal primario y 10% (7.5 kHz) reservados al SCA.

La fig. 7-21c muestra el espectro de frecuencias de la banda base de FM, como es a partir de 1961. Comprende el canal estereofónico de 50 Hz a 15 kHz, más un canal estereofónico

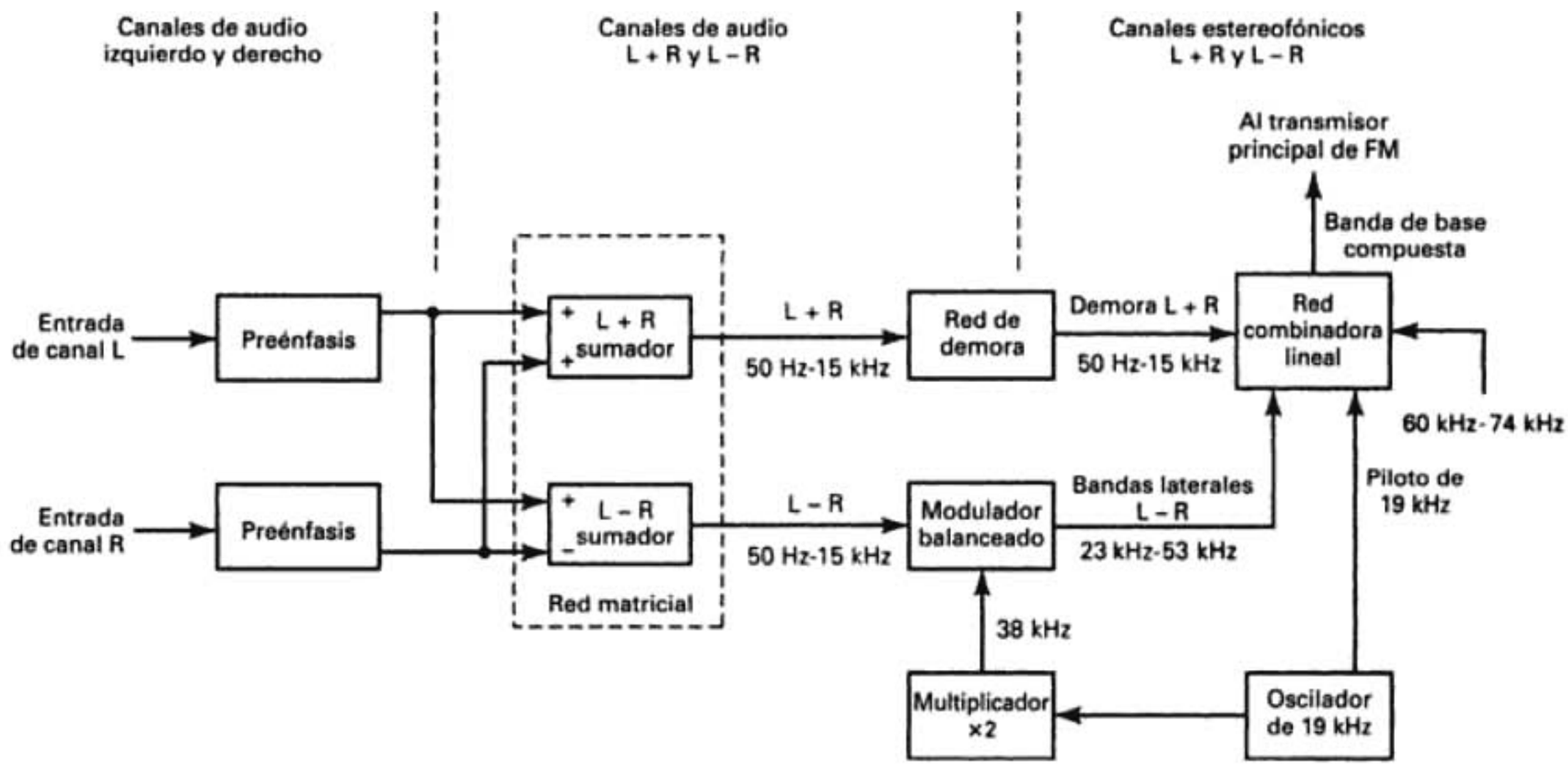


FIGURA 7-22 Transmisor estereofónico de FM con multiplexado por división de frecuencia

adicional multiplexado por división de frecuencia en una señal compuesta en banda base y un piloto de 19 kHz. Los tres canales son: 1) el izquierdo (L) más el derecho (R) de audio (el canal estereofónico de audio L + R), 2) el izquierdo más el derecho invertido, de audio (el canal estereofónico L − R) y 3) la subportadora SCA y sus bandas laterales asociadas. El canal estereofónico L + R ocupa la banda de 0 a 15 kHz (en esencia, la información de audio L y R combinadas). La amplitud del canal de audio L − R modula una subportadora de 38 kHz y produce el canal estereofónico L − R, que es una señal de doble banda lateral con portadora suprimida, que ocupa la banda de 23 kHz a 53 kHz, que sólo se usa para transmitir FM estereofónica. Las transmisiones SCA ocupan el espectro de frecuencias de 60 a 74 kHz. La información contenida en los canales estereofónicos L + R y L − R es idéntica, excepto su fase. Con este esquema, los receptores monoaurales pueden demodular el espectro total de la banda base, pero sólo amplifica y alimenta a sus bocinas el canal de audio de 50 Hz a 15 kHz L + R. Por lo tanto, cada bocina reproduce el espectro completo del sonido original. Los receptores estereofónicos deben proporcionar demodulación adicional, del canal estereofónico L − R de 23 kHz a 53 kHz, separar los canales izquierdo y derecho de audio, para entonces alimentarlos a sus bocinas respectivas. De nuevo, la subportadora SCA es demodulada en todos los receptores de FM, aunque sólo los que tienen el equipo especial SCA siguen demodulando la subportadora en frecuencias de audio.

En la transmisión estereofónica, la desviación máxima de frecuencia sigue siendo 75 kHz; se reservan 7.5 kHz (10%) para transmisión SCA y otros 7.5 kHz (10%) para una piloto estereofónica de 19 kHz. Esto deja 60 kHz de desviación de frecuencia para la transmisión estereofónica real de los canales L + R y L − R. Sin embargo, esos canales no necesariamente están limitados a una desviación de frecuencia de 30 kHz cada uno. Se aplica una técnica muy sencilla, pero exclusiva, para entrelazar los dos canales en forma tal que a veces el canal estereofónico L + R o el L − R puedan desviar, por sí mismos, 60 kHz a la portadora principal. Sin embargo, la desviación total nunca será mayor de 60 kHz. Esta técnica de entrelazamiento se explicará más adelante en esta sección.

Transmisión estereofónica de FM

La fig. 7-22 muestra el diagrama simplificado de bloques de un transmisor estereofónico de FM. Los canales L y R de audio se combinan en una red matricial para producir los canales de audio

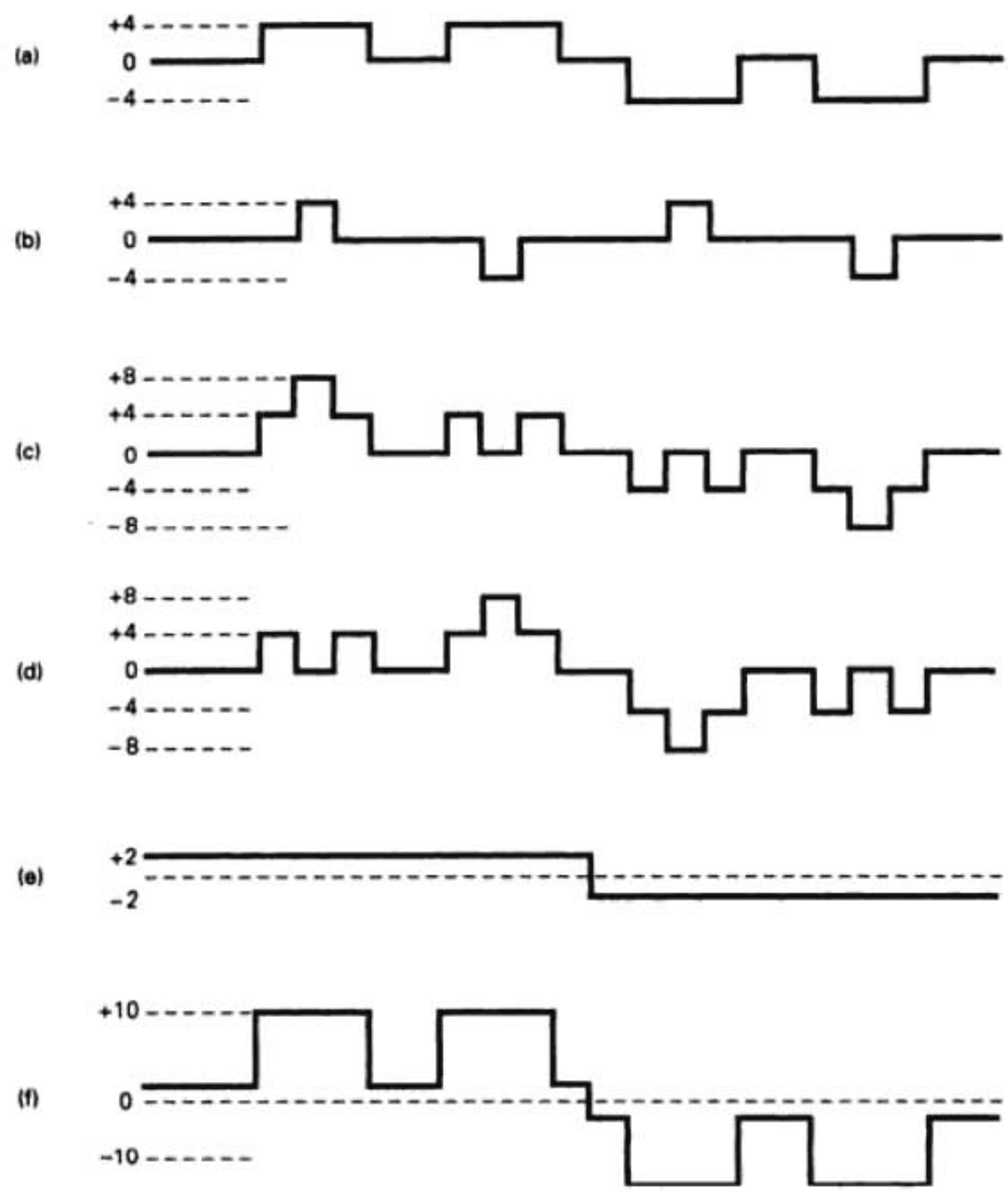


FIGURA 7-23 Desarrollo de la señal estereofónica compuesta para tener señales L y R de igual amplitud: (a) señal de audio; (b) señal de audio R; (c) canal estereofónico L + R; (d) canal estereofónico L − R; (e) piloto SCA + 19 kHz; (f) forma de onda en banda base compuesta

L + R y L − R. El canal de audio L − R modula una subportadora de 38 kHz y produce un canal estereofónico L − R de 23 a 53 kHz. Como hay un retardo de tiempo introducido en la trayectoria de la señal L − R, al propagarse por el modulador balanceado, el canal estereofónico L + R debe retardarse un poco en forma artificial para mantener la integridad de fase con el canal L − R cuando se demodulen. También, para fines de demodulación, se transmite una piloto de 19 kHz, y no una subportadora de 38 kHz, porque es bastante más difícil recuperar la de 38 kHz en el receptor. La banda base de la señal compuesta se alimenta al transmisor de FM, donde modula a la portadora principal.

Entrelazamiento de canales L + R y L − R. La fig. 7-23 muestra el desarrollo de la señal estereofónica compuesta para señales de canal de audio L y R de igual amplitud. Para fines de ilustración se muestran ondas rectangulares. La tabla 7-1 es un resumen tabular de los voltajes de señal individual y total para la fig. 7-23. Nótese que el canal de audio L − R no aparece en la forma de onda compuesta. Ese canal modula a la subportadora de 38 kHz para formar las bandas laterales estereofónicas L − R, que son parte del espectro compuesto.

Para el modulador de FM de este ejemplo se supone que 10 V de señal en banda base produce 75 kHz de desviación de frecuencia de la portadora principal, y las polaridades de las

TABLA 7-1 Voltajes compuestos de FM

L	R	L + R	L − R	SCA y piloto	Total
0	0	0	0	2	2
4	0	4	4	2	10
0	4	4	−4	2	2
4	4	8	0	2	10
4	−4	0	8	2	10
−4	4	0	−8	−2	−10
−4	−4	−8	0	−2	−10

piloto SCA y de 19 kHz que se ven son para la desviación máxima de frecuencia. Los canales de audio L y R se limitan, cada uno, a un valor máximo de 4 V; 1 V es para la SCA y 1 V es para la piloto estereofónica de 19 kHz. Por consiguiente, quedan 8 V para los canales estereofónicos L + R y L − R. La fig. 7-23 muestra los canales L, R, L + R y L − R, la piloto SCA y la de 19 kHz, y la forma de onda estereofónica compuesta. Se puede ver que los canales estereofónicos L + R y L − R se entrelazan y nunca producen más de 8 V de amplitud total y, en consecuencia, nunca producen una desviación de frecuencia mayor que 60 kHz. La banda base total compuesta nunca es mayor que 10 V (desviación de 75 kHz).

La fig. 7-24 muestra el desarrollo de la forma de onda estereofónica compuesta, para valores desiguales de las señales L y R. Se puede ver de nuevo que la forma de onda estereofónica compuesta nunca es mayor de 10 V, o 75 kHz de desviación de frecuencia. Para el primer conjunto de formas de onda parece que la suma de las ondas L + R y L − R se anula por completo; en realidad no es verdad; sólo parece así porque en este ejemplo se usaron formas rectangulares de onda.

Recepción estereofónica de FM

Los receptores estereofónicos de FM son idénticos a los normales de FM hasta la salida de la etapa de detección de audio. La salida del discriminador es el espectro total de la banda base que se mostró en la fig. 7-21c.

La fig. 7-25 muestra el diagrama simplificado de bloques de un receptor de FM que tiene salidas de audio tanto monoaural como estereofónica. En la sección monoaural del procesador de señales tan sólo se filtra, amplifica y alimenta a las bocinas L y R el canal estereofónico L + R, que contiene toda la información original de los canales de audio L y R. En la sección estereofónica del procesador de señales, la señal de banda base se alimenta a un demodulador estereofónico, donde se separan los canales de audio L y R, y se alimentan a sus bocinas respectivas. Los canales estereofónicos L + R y L − R, y la piloto de 19 kHz, se separan, con filtros, de la señal de banda base compuesta. La piloto de 19 kHz se filtra con un filtro pasabanda de alta Q, se multiplica por 2, se amplifica y a continuación se alimenta al demodulador L − R. El canal estereofónico L + R se filtra y se elimina con un filtro pasabajas con frecuencia de corte superior a 15 kHz. La señal L − R de doble banda lateral se separa con un filtro pasabanda sintonizado aproximadamente, y a continuación se mezcla con la portadora recuperada de 38 kHz en un modulador balanceado, para producir la información de audio L − R. La red matricial combina a las señales L + R y L − R de tal modo que se separan las señales de información de audio L y R, y alimentan a sus redes respectivas de deénfasis y a sus bocinas.

La fig. 7-26 muestra el diagrama de bloques de un decodificador matricial estereofónico. El canal de audio L − R se suma en forma directa al canal de audio L + R. La salida del sumador es

$$\begin{array}{r} L + R \\ \underline{+ (L - R)} \\ 2L \end{array}$$

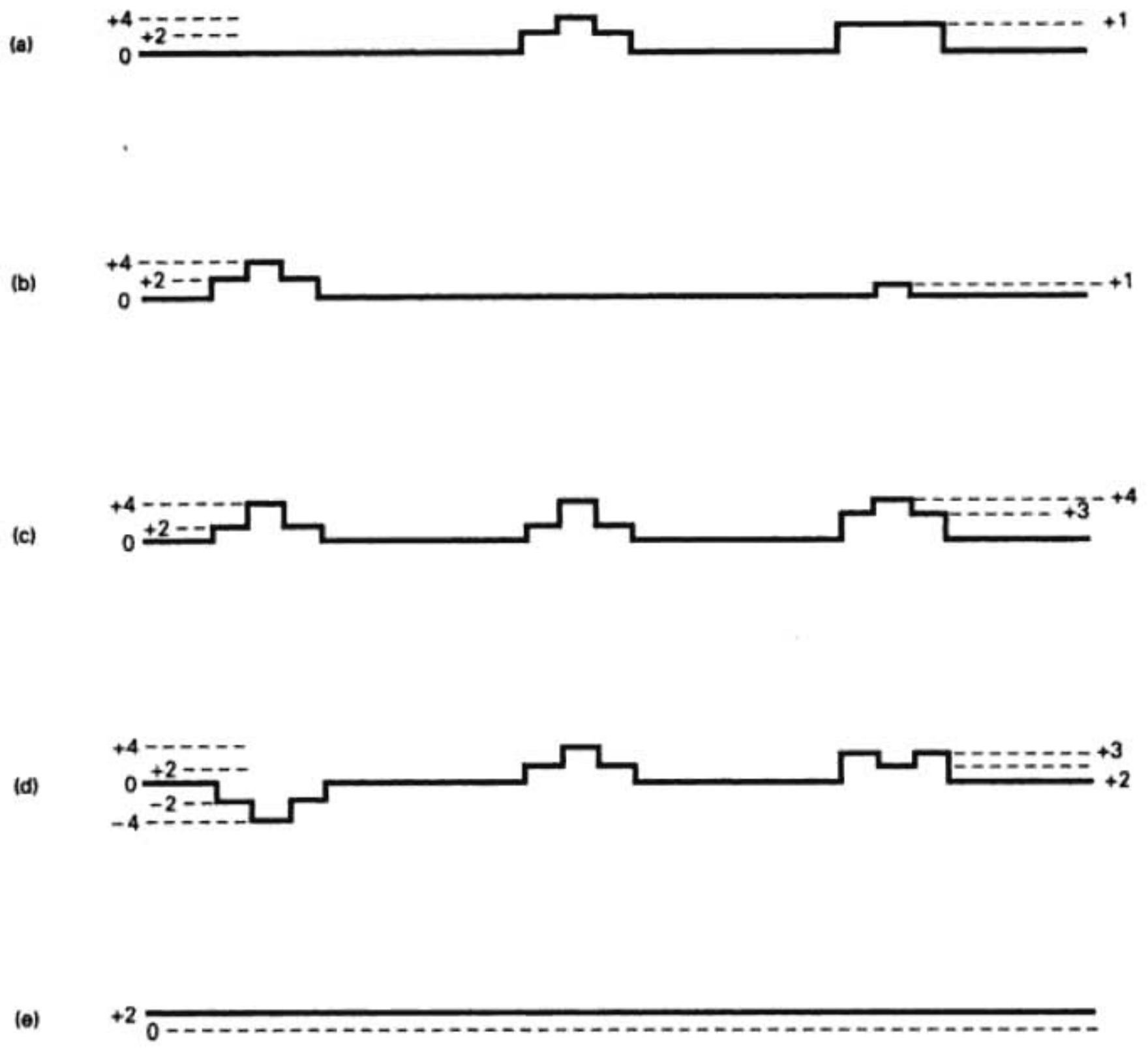


FIGURA 7-24 Desarrollo de la señal estereofónica compuesta para señales L y R de amplitud desigual: (a) señal de audio L; (b) señal de audio R; (c) canal estereofónico L + R; (d) canal estereofónico L − R; (e) piloto SCA + 19 kHz; (f) forma de onda en banda base compuesta

El canal de audio L − R se invierte y después se suma al canal de audio L + R. La salida del sumador es el canal estereofónico

$$\begin{array}{r} L + R \\ -(L - R) \\ \hline 2R \end{array}$$

que contiene toda la información original de los canales L y R de audio, sólo se filtra, amplifica y a continuación se alimenta a las bocinas L y R. En la sección estereofónica del procesador de señal, la señal de banda base se alimenta a un demodulador estereofónico en el que los canales de audio L y R se separan y se alimentan a sus bocinas respectivas. Los canales estereofónicos L + R y L − R, y la piloto de 19 kHz se separan con filtros de la señal compuesta de banda

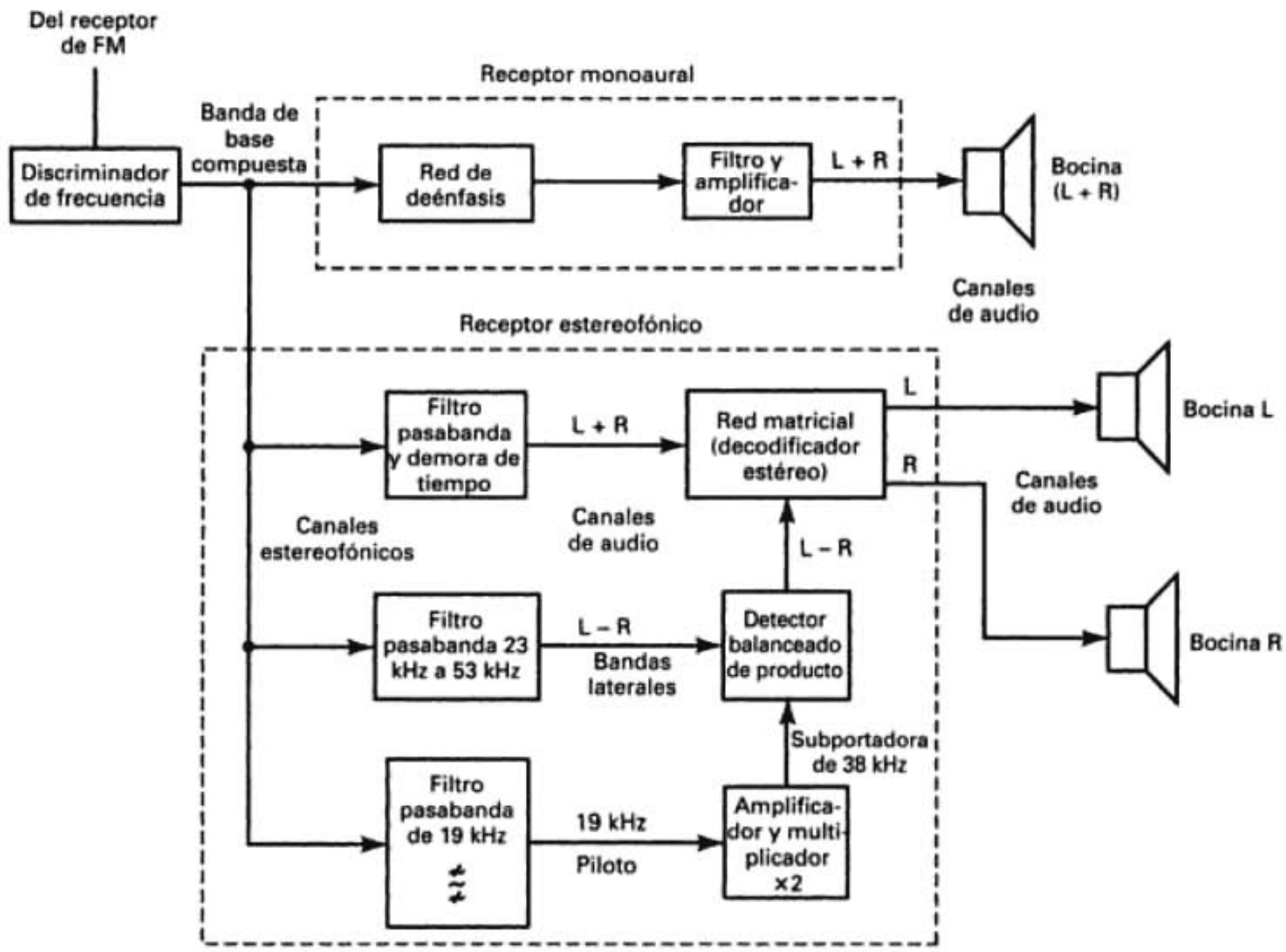


FIGURA 7-25 Receptor FM estereofónico y monoaural

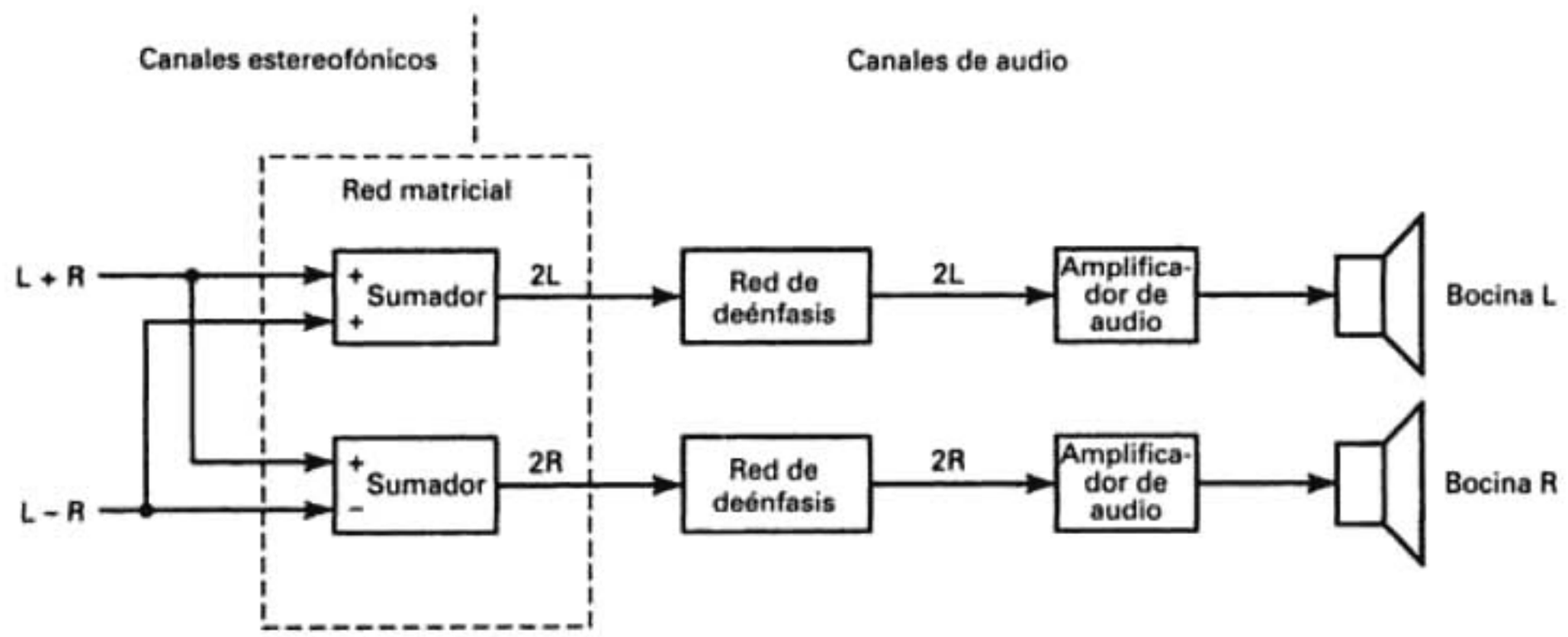


FIGURA 7-26 Decodificador estereofónico de red matricial

base. La piloto de 19 kHz se filtra con un filtro pasabanda de alta Q, se multiplica por 2, se amplifica y se alimenta al demodulador L − R. El canal estereofónico L + R se filtra y se elimina con un filtro pasabajas con frecuencia superior de corte de 15 kHz. La señal L − R de doble banda lateral se separa con un filtro pasabanda de sintonía amplia, y a continuación se mezcla con la portadora recuperada de 38 kHz en un modulador balanceado, para producir la informa-

DEMODULADOR ESTEREOFÓNICO

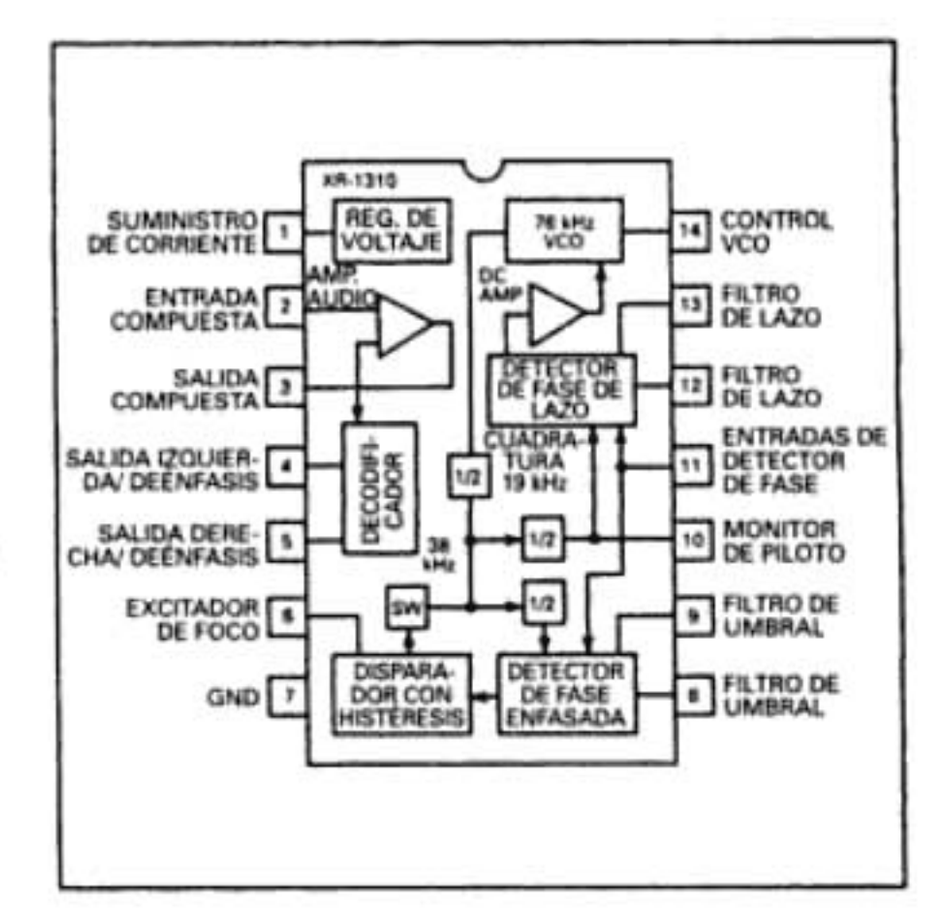
DESCRIPCIÓN GENERAL

El XR-1310 es un demodulador estereofónico FM exclusivo, que emplea técnicas de amarre de fase para extraer los canales derecho e izquierdo de audio de la señal compuesta. Con un lazo de fase cerrada para regenerar la subportadora de 38 kHz, no requiere tanques L-C externos para sintonizar. El alineamiento se logra con un solo potenciómetro

PROPIEDADES

No requiere inductores
Poca cantidad de partes externas
Sintonía sencilla y no crítica, mediante ajuste con un solo potenciómetro
Conmutador interno estéreo/monoaural con capacidad de activación de un foco de 100 mA
Amplio intervalo dinámico: 600 mV (RMS) de señal de entrada máxima compuesta
Amplio intervalo de voltaje de suministro: de 8 a 14 volts
Excelente separación de canales
Baja distorsión
Excelente rechazo de SCA

DIAGRAMA DE BLOQUES FUNCIONALES Marzo de 1982



INFORMACIÓN PARA PEDIDOS

Número de parte	Empaque	Temperatura de funcionamiento
XR-1310CP	Plástico	−40° C a +85° C

APLICACIONES

Demodulación de FM estereofónica

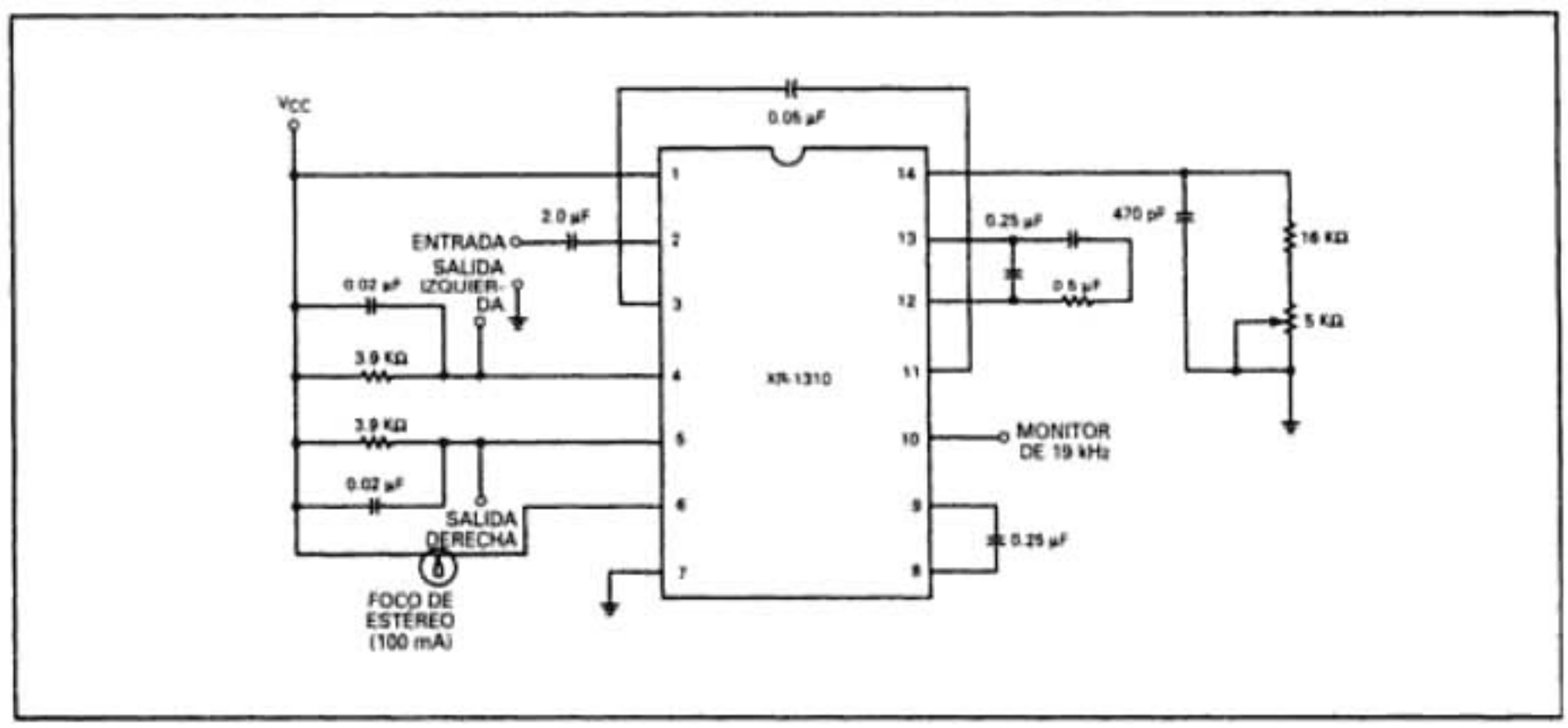


FIGURA 7-27 Demodulador estereofónico XR-1310 (continúa)

ción de audio L − R. La red matricial combina las señales L + R y L − R de tal forma que se separan las señales de información de audio L y R, que se alimentan a sus respectivas redes de deénfasis y a sus respectivas bocinas.

Demodulador estereofónico en circuito en gran escala de integración. La fig. 7-27 muestra la traducción de la hoja de especificaciones para el demodulador/decodificador estereofónico XR-1310. Es un demodulador estereofónico de FM que usa técnicas de lazo de fase cerrada para obtener los canales derecho e izquierdo de audio a partir de la señal estereofónica compuesta. Usa un lazo de fase cerrada para fijarse a la piloto de 19 kHz y regenerar la portadora de 38 kHz. Este circuito XR-1310 no requiere circuito tanque *LC* para sintonizarse, y el alinea-

CARACTERÍSTICAS ELÉCTRICAS

Condiciones de prueba: A menos que se diga otra cosa, V_{CC}* = + 12 Vcd, TA = +25° C, 560 mV(RMS) (2.8 Vp-p) señal estándar multiplexada compuesta sólo con canal L o R modulado a 1.0 kHz con 100 mV (RMS) (10% nivel de la piloto), con el circuito de la figura 1

PARÁMETROS	MÍN	NORMAL	MÁX	UNIDAD
Señal de entrada máxima compuesta estándar (0.5% THD)	2.8			V (p-p)
Señal máxima de entrada monoaural (1.0% THD)	2.8			V (p-p)
Impedancia de entrada		50		kΩ
Separación de canal estereofónico (50 Hz-15 kHz)	30	40		dB
Voltaje de salida de audio (canal deseado)		485		mV (rms)
Balance de canal monoaural (tono piloto "apagado")			1.5	dB
Distorsión armónica total		0.3		%
Rechazo de frecuencia ultrasónica 19 kHz 38 kHz	50	34.4 45		dB
Rechazo de SCA inherente (f = 67 kHz; nota de pulso de 9.0 kHz medida con modulación de 1.0 kHz "apagada")		80		dB
Nivel de cambio a estéreo (Entrada de 19 kHz para foco "encendido") Histéresis	13	 6	20	mV (rms) dB
Intervalo de captura (error permisible de sintonía del oscilador interno, valores de referencia del circuito de la figura 1)		±3.5		%
Voltaje de suministro de operación (cargas reducidas a 2.7 kΩ para funcionamiento con 8 V)	8.0		14	V (dc)
Corriente de drenado (foco "apagado")		13		mA (dc)

* Los símbolos se apegan al Boletín No. 1 de JEDEC Engineering, cuando sea el caso.

CAPACIDADES MÁXIMAS ABSOLUTAS

(TA = +25° C a menos que se indique otra cosa)

Voltaje mínimo de suministro	14 V
Corriente de foco (capacidad nominal, foco de 12 V)	75 mA
Disipación de potencia (limitación de paquete)	625 mW
Disminuir arriba de TA = +25° C	5.0 mW/°C
Intervalo de temperatura (ambiente) de operación	−40 a +85° C
Intervalo de temperatura de almacenamiento	−65 a +150° C

FIGURA 7-27 (*Continuación*) Demodulador estereofónico XR-1310

miento se logra con un solo potenciómetro. Tiene una sintonía sencilla y no crítica, excelente separación de canales, baja distorsión y amplio margen dinámico.

RADIOCOMUNICACIONES EN FM DE DOS SENTIDOS

La radiocomunicación en FM de dos sentidos se usa mucho en comunicaciones móviles de *seguridad pública*, como por ejemplo de policía y de bomberos, y en los servicios médicos de emergencia. La FCC asignó tres bandas primarias de frecuencia para radiocomunicaciones en FM de dos sentidos: 132 a 174 MHz, 450 a 470 MHz y 806 a 947 MHz. La desviación máxima de frecuencias para los transmisores de FM en dos sentidos suele ser de 5 kHz, y la frecuencia máxima de señal moduladora es 3 kHz. Esos valores producen una relación de desviación de 1.67, y un ancho de banda máximo de Bessel aproximado de 24 kHz. Sin embargo, la distancia entre canales asignada por la FCC es 30 kHz. La radio FM de dos sentidos es semidúplex, que soporta comunicaciones en dos sentidos, pero no en forma simultánea. Sólo una parte puede transmitir a un tiempo. Las transmisiones se inician cerrando un interruptor *para hablar* (PTT, de *push-to-talk*), que enciende el transmisor y apaga el receptor. Durante las condiciones sin uso,

el transmisor de apaga y el receptor se enciende, para permitir vigilar el canal de radio y comprobar si hay transmisiones desde otras estaciones.

Perspectiva histórica

La radio móvil ya se usaba en 1921, cuando el Departamento de Policía de Detroit usó un sistema que funcionaba a una frecuencia cercana a 2 MHz. En 1940, la FCC puso nuevas frecuencias a a disposición de la radio móvil en la banda de frecuencias de 30 a 40 MHz. Sin embargo, no fue hasta que se desarrollaron las técnicas de modulación de frecuencia, para mejorar la recepción en presencia de ruido eléctrico y desvanecimiento de señal, que la radio móvil fue de utilidad. El primer sistema telefónico móvil comercial en Estados Unidos se estableció en 1946, en St. Louis Missouri, cuando la FCC asignó seis canales de 60 kHz a la telefonía móvil en el intervalo de frecuencias de 150 MHz. En 1947, un sistema público de telefonía móvil se estableció a lo largo de la carretera de la ciudad de Nueva York a Boston, que trabajaba en el intervalo de frecuencias de 35 a 40 MHz. En 1949, la FCC autorizó 6 canales móviles adicionales a *portadoras comunes de radio*, que se definieron como empresas que no proporcionan un servicio telefónico público alámbrico, pero que sí se interconectan con la red telefónica pública y proporcionan servicio telefónico *sin conductores*. Después, la FCC aumentó la cantidad de canales de 6 a 11, reduciendo el ancho de banda a 30 kHz y colocando los nuevos canales entre los anteriores. En 1950, la FCC agregó 12 nuevos canales en la banda de 450 MHz.

Hasta 1964, los sistemas telefónicos móviles sólo funcionaban en el *modo manual*: un operador especial de teléfono móvil manejaba cada llamada hacia y desde cada *unidad móvil*. En 1964 se pusieron en servicio *sistemas de selección automática de canal* para los sistemas telefónicos móviles. Así se eliminó la necesidad del funcionamiento de oprimir para hablar, y se permitió que los clientes *marcaran directamente* sus llamadas sin ayuda de un operador. El *establecimiento automático de llamada* se extendió a la banda de 450 MHz en 1969, y los *sistemas mejorados de teléfono móvil* (IMTS, de *improved mobile telephone systems*) se volvieron la norma de servicio telefónico móvil en Estados Unidos. Hoy hay más de 200,000 suscriptores en el *servicio telefónico móvil* (MTS, *mobile telephone service)* en todo ese país. El MTS usa canales de radio FM para establecer enlaces de comunicación entre teléfonos móviles y los transceptores de la *estación base* central, enlazados a la central telefónica local a través de cables conductores telefónicos normales. La mayoría de los sistemas MTS dan servicio en un área aproximada de 40 millas de diámetro, y cada canal funciona en forma parecida a una *línea compartida*. Cada canal se puede asignar a varios suscriptores, pero sólo uno lo puede usar cada vez. Si está ocupado el canal preasignado, el abonado debe esperar hasta que esté libre, antes de hacer o recibir una llamada.

La creciente demanda en el saturado espectro de frecuencias de la telefonía móvil impulsó a la FCC a emitir el "Expediente 18262", que solicitó un medio de proporcionar mayor eficiencia en el espectro de frecuencias. En 1971, la AT&T presentó una propuesta sobre la factibilidad técnica de proporcionar el uso eficiente del espectro de frecuencias de telefonía móvil. El informe de la AT&T, titulado *Servicio de teléfono móvil de alta capacidad*, describió las bases de la radio celular.

En abril de 1981, la FCC aprobó un esquema de autorización para los mercados de *radio celular*. Cada mercado da servicio a un *área de servicio*, definida de acuerdo con las Áreas estadísticas metropolitanas normales (SMSA, por *Standard Metropolitan Statistical Areas*) de la Oficina del Censo. A principios de 1982, la FCC aprobó un plan definitivo para aceptar aplicaciones de permisos celulares a partir de junio de 1982, y una ronda final de aplicaciones en marzo de 1983. Las batallas legales que siguieron, para licencias celulares entre AT&T, MCI, GTE y muchos otros proveedores salen del alcance de este libro.

Radiotransmisor de FM en dos sentidos

El diagrama simplificado de bloques de un radiotransmisor indirecto de *circuito integrado modular* para FM se ve en la fig. 7-28. Se usa en general la FM indirecta, porque los transmisores directos de FM no tienen la estabilidad de frecuencia necesaria para cumplir con las normas de la FCC sin usar lazos de control automático de frecuencia (AFC). El transmisor que se ve es una unidad de cuatro canales, que funciona en la banda de frecuencias de 150 kHz a

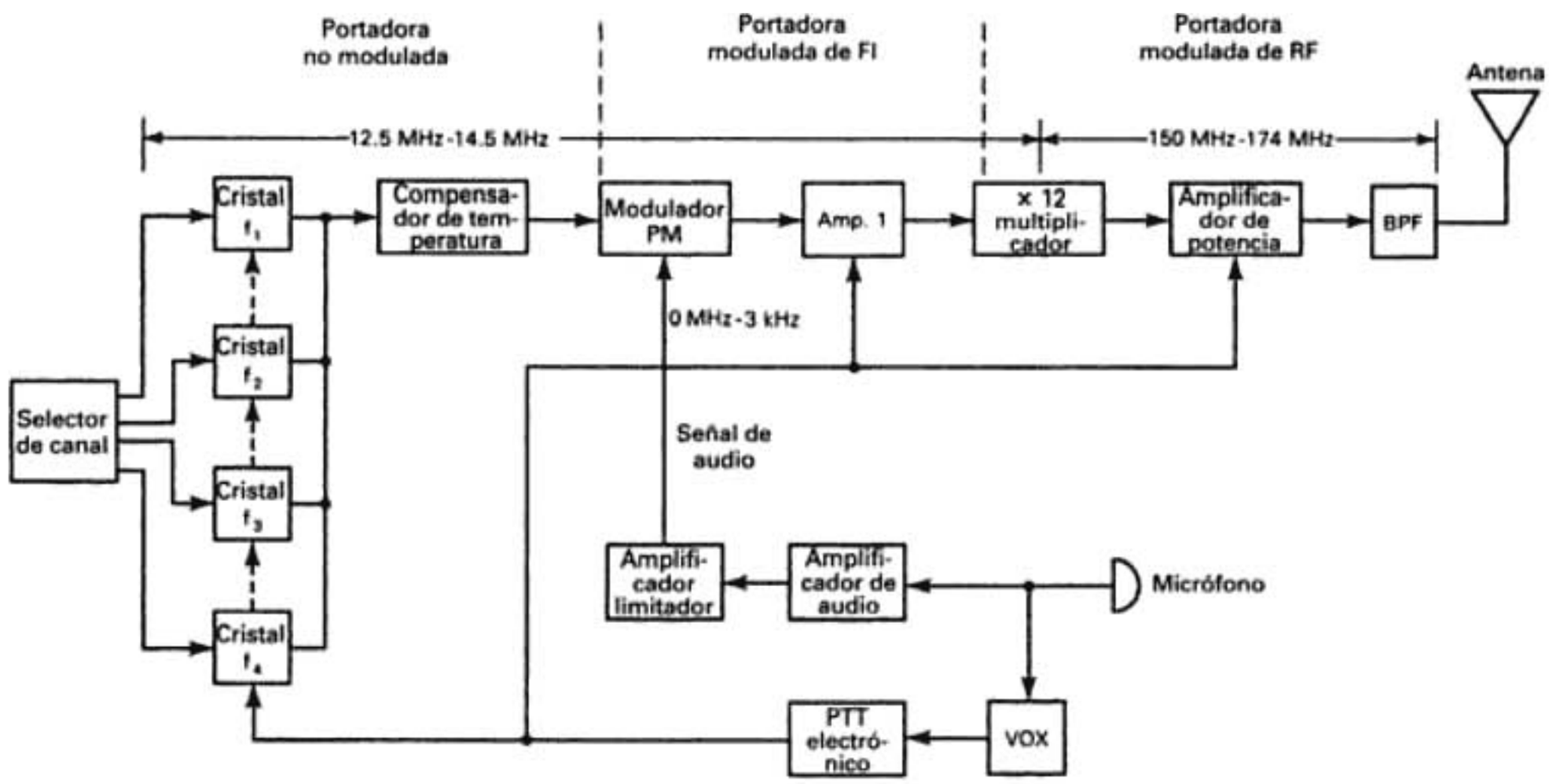


FIGURA 7-28 Diagrama de bloques de un transmisor de FM en dos sentidos

174 MHz. El selector de canal aplica energía a uno de cuatro módulos de oscilador de cristal, que funciona a una frecuencia entre 12.5 MHz y 14.5 MHz, dependiendo de la frecuencia de la portadora final de transmisión. La frecuencia del oscilador está compensada mediante el módulo de compensación, para asegurar una estabilidad de ±0.0002%. El modulador de fase usa un diodo varactor, al que modula la señal de audio a la salida del limitador de audio. La amplitud de la señal de audio se limita para asegurar que el transmisor no se desvíe en exceso. La portadora modulada de FI se amplifica y después se multiplica por 12, para producir la RF deseada en la transmisión. La señal de RF se vuelve a amplificar y se filtra antes de su transmisión. Se usa un *botón para hablar* (PTT) *electrónico* y no un interruptor mecánico para reducir el ruido de estática asociado con el *rebote de contactos* en los interruptores mecánicos. Al manipular el PTT se aplica corriente cd al módulo oscilador seleccionado y a los amplificadores de potencia de RF.

La fig. 7-29 muestra el diagrama de un módulo típico de PTT electrónico. Al oprimir el interruptor PTT se aterriza la base de Q_1, haciéndolo conducir y apagando a Q_2. Estando desactivado Q_2, se aplica V_{CC} al transmisor y se quita del receptor. Cuando se suelta el PTT, Q_1 se apaga y quita V_{CC} del transmisor, encendiendo Q_2 y aplicando V_{CC} al receptor.

Los transmisores equipados con VOX (*transmisor operado con la voz*) se cambian en forma automática cada vez que el operador habla frente al micrófono, independientemente de si está oprimido el PTT. Estos transmisores requieren un micrófono externo. En la fig. 7-30 se muestra el esquema de un módulo VOX normal. La corriente de la señal de audio, en la banda de 400 a 600 Hz se filtra y amplifica con Q_1, Q_2 y Q_3. La salida de Q_3 se rectifica y se usa para activar Q_4, que pone una tierra en el circuito de PTT, activando el transmisor y desactivando el receptor. Cuando no hay señal de audio a la entrada, Q_4 está apagado y se abre la terminal del PTT, desactivando el transmisor y activando el receptor.

Radiorreceptor de FM en dos sentidos

En la fig. 7-31 se muestra el diagrama de bloques de un radiorreceptor típico de FM en dos sentidos. El receptor que se ve es uno de circuito integrado de cuatro canales, con cuatro módulos separados de oscilador de cristal. Cuando el receptor está encendido, se activa uno de los cuatro módulos de oscilador, dependiendo de la posición del selector de canales. La frecuencia del oscilador tiene compensación por temperatura, y se multiplica por 9. La salida del multiplicador

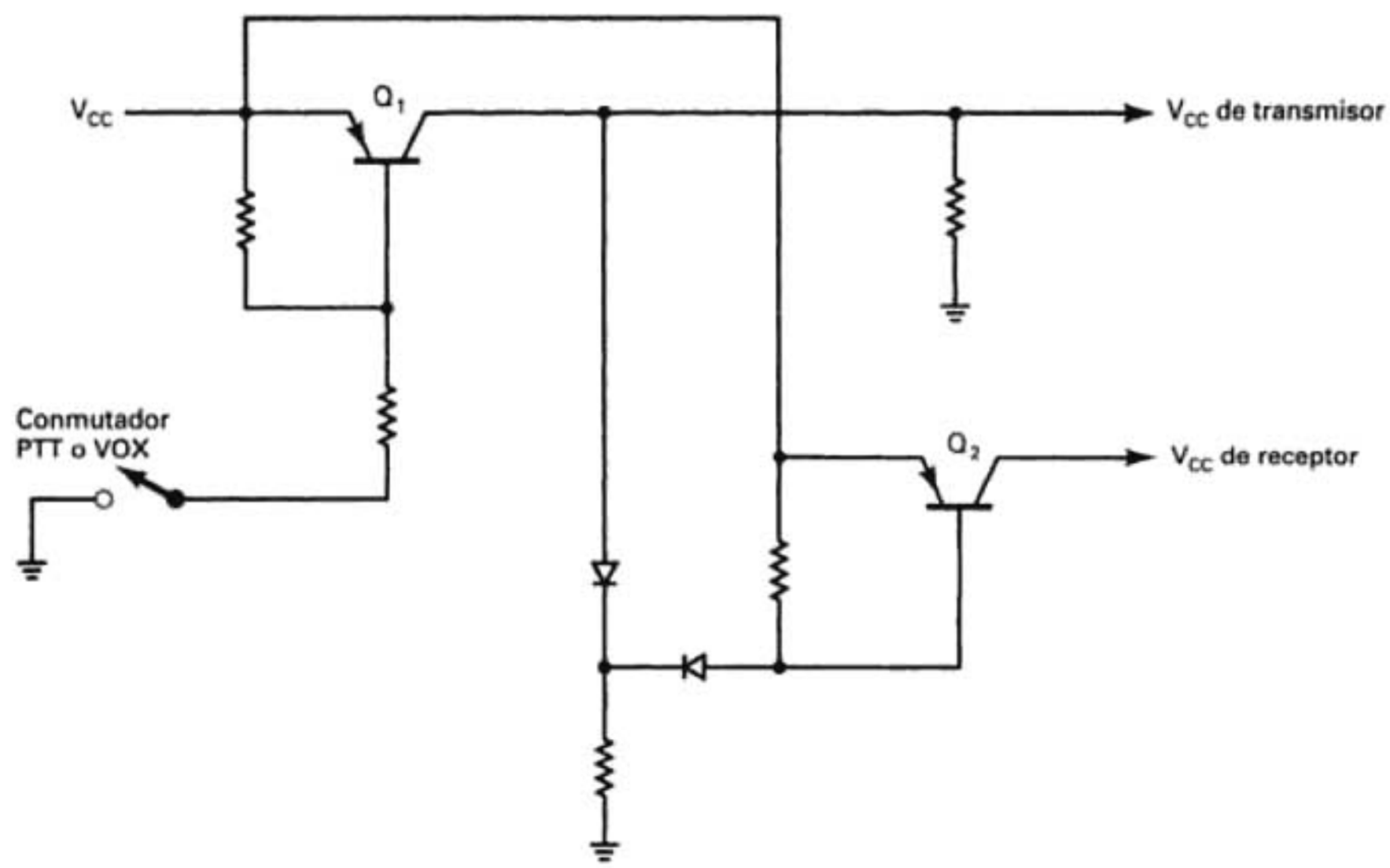


FIGURA 7-29 Diagrama de un PTT (botón para hablar) electrónico

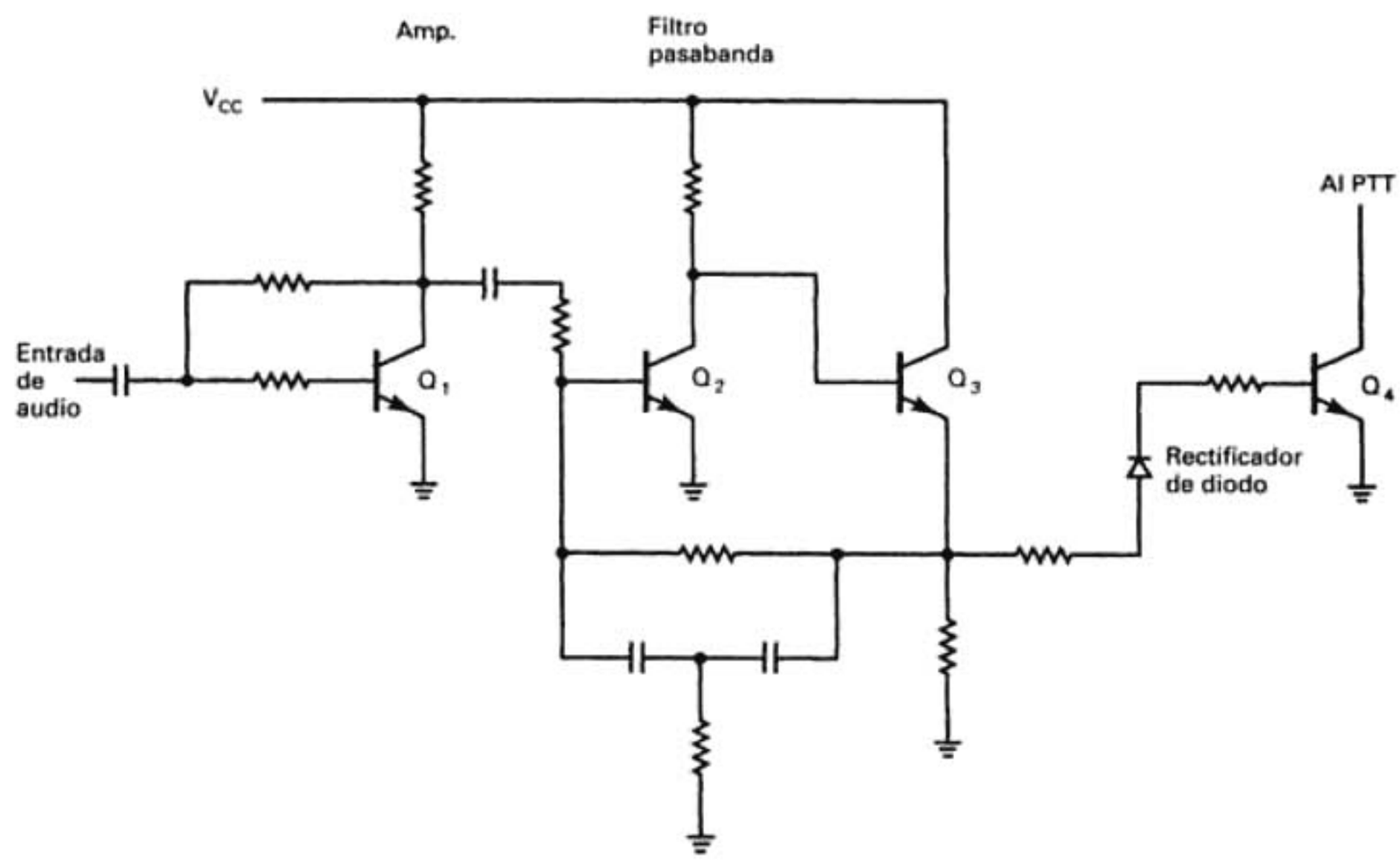


FIGURA 7-30 Diagrama de un VOX (interruptor activado por la voz)

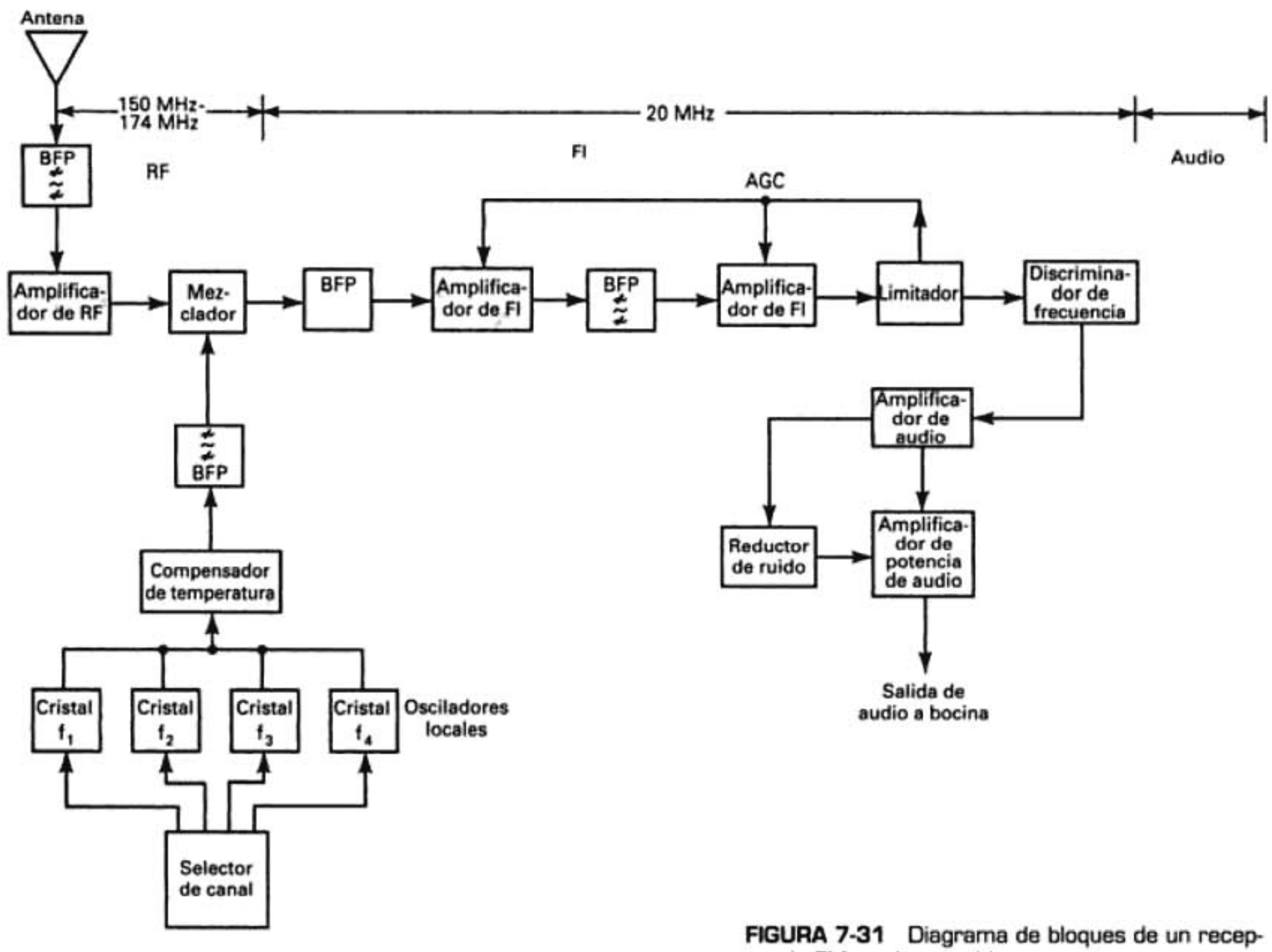


FIGURA 7-31 Diagrama de bloques de un receptor de FM en dos sentidos

se aplica al mezclador, donde heterodina la señal de RF de llegada, para producir una frecuencia intermedia de 20 MHz. Este receptor usa inyección lateral inferior, y la frecuencia del oscilador se determina como sigue

$$\text{frecuencia del cristal} = \frac{\text{frecuencia de RF} - 20\ \text{MHz}}{9}$$

La señal de FI se filtra, amplifica, limita y después se aplica al discriminador de frecuencia para su demodulación. El voltaje de salida del discriminador se amplifica y a continuación se aplica a la bocina. Un circuito típico de amplificador y reducción de ruido se ve en la fig. 7-32. El circuito de reducción se manipula con el ruido fuera de banda en la salida del amplificador de audio. Cuando no hay señal de recepción de RF, el control automático de ganancia (CAG) hace que aumente la ganancia de los amplificadores de FI hasta el máximo, con lo que aumenta el ruido del receptor en la banda de 3 a 5 kHz. Siempre que hay demasiado ruido, el amplificador de audio se apaga y el receptor se silencia. El filtro pasabanda de entrada pasa la señal de ruido de 3 a 5 kHz, que se amplifica y rectifica. El voltaje rectificado de salida determina la condición de apagado o encendido del interruptor Q_3 del reductor de ruido. Cuando Q_3 está encendido, se aplica V_{CC} al amplificador de audio. Cuando Q_3 está apagado, V_{CC} se quita del amplificador de audio y silencia al receptor. R_x es un ajuste de sensibilidad de reducción de ruido.

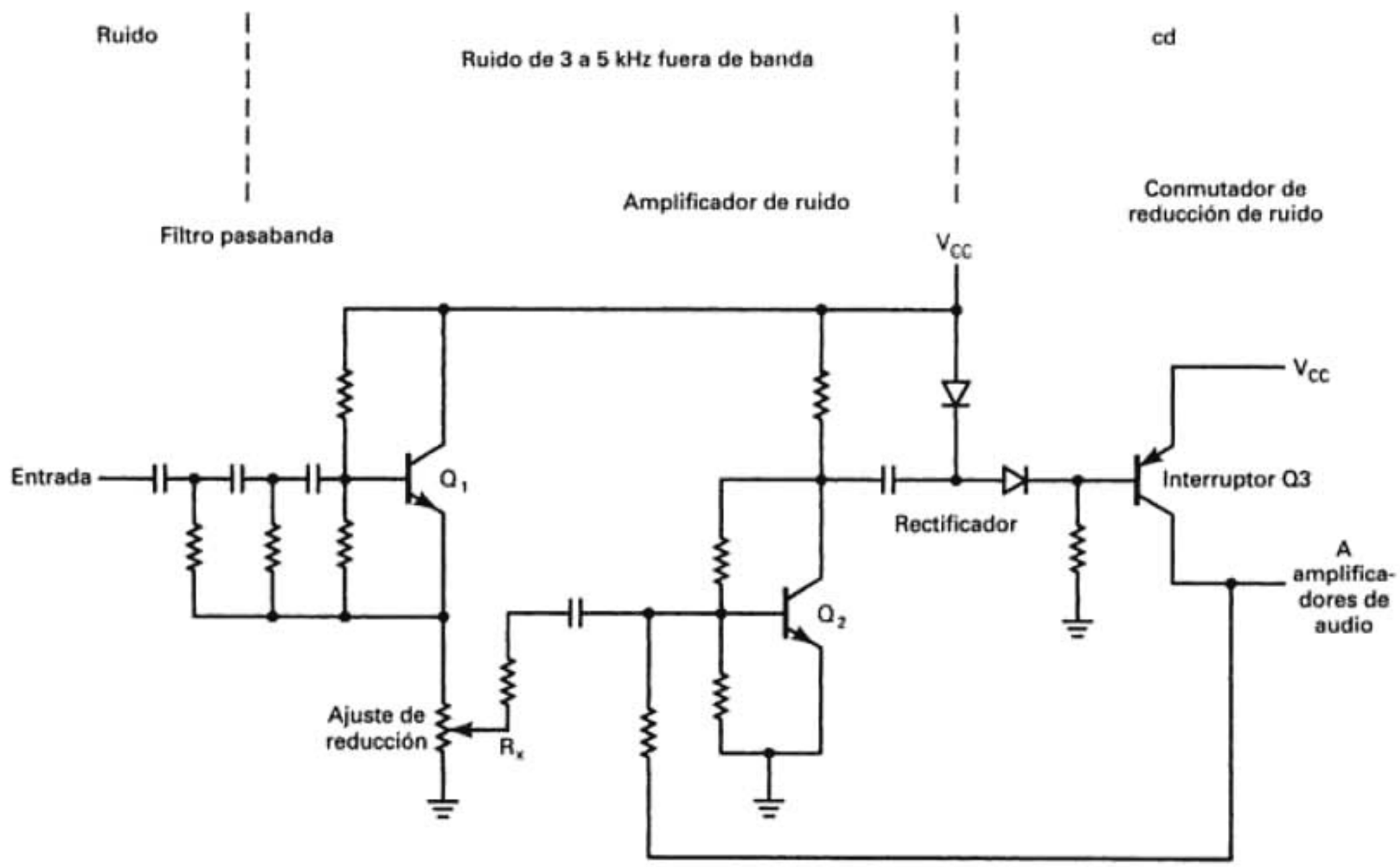


FIGURA 7-32 Circuito reductor de ruido

PREGUNTAS

7-1. Describa las diferencias básicas entre los receptores de AM y FM.

7-2. Trace el diagrama de un *detector por pendiente* no balanceado y describa su funcionamiento.

7-3. Trace el diagrama de un *detector por pendiente* balanceado y describa su funcionamiento.

7-4. Trace el diagrama de un *discriminador de Foster-Seeley* y describa su funcionamiento.

7-5. Trace el diagrama de un *detector de relación* y describa su funcionamiento.

7-6. Describa el funcionamiento de un demodulador PLL FM.

7-7. Trace el esquema de un *demodulador de FM de cuadratura* y describa su funcionamiento.

7-8. Compare las ventajas y desventajas de los circuitos demoduladores de FM que se mencionaron en las preguntas 7-1 a 7-7.

7-9. ¿Cuál es el objeto de un *limitador* en un receptor de FM?

7-10. Describa el *umbral de FM*.

7-11. Describa el funcionamiento de un *transmisor estereofónico* de FM, y de un *receptor estereofónico* de FM.

7-12. Trace el diagrama de bloques de un radiotransmisor de FM en dos sentidos, y explique su funcionamiento.

7-13. Trace el diagrama de bloques de un radiorreceptor de FM en dos sentidos y explique su funcionamiento.

7-14. Describa el funcionamiento de un *circuito de botón para hablar* electrónico.

7-15. Describa el funcionamiento de un *circuito VOX*.

7-16. Explique en forma breve cómo se produce una señal de *FM estereofónica* compuesta.

7-17. ¿Qué significa el término *entrelazamiento* de señales L y R en la transmisión estereofónica?

7-18. ¿Cuál es el objetivo de la *piloto* de 19 kHz en la emisión FM?

7-19. ¿Cuál es la diferencia entre *radio móvil* y *teléfono móvil*?

PROBLEMAS

7-1. Determine la relación mínima S/N en la entrada, que requiere un receptor con mejoramiento de FM de 15 dB, cifra de ruido NF = 4 dB y una S/N = 33 dB deseada en post detección.

7-2. Para un receptor de FM con ancho de banda de 100 kHz, cifra de ruido NF = 6 dB y temperatura de ruido en la entrada $T = 200°$ C, calcule la potencia mínima de portadora de recepción para lograr una S/N = 40 dB en postdetección. Use el diagrama de bloques del receptor de la fig. 7-1 como modelo y la curva de umbral de FM de la fig. 7-11.

7-3. Para un receptor de FM sintonizado a 92.75 MHz que usa inyección lateral superior y una primera FI de 10.7 MHz, determine la frecuencia imagen y la frecuencia del oscilador local.

7-4. Determine V_{sal} para un receptor de FM con desviación de frecuencia de entrada $\Delta f = 40$ kHz y relación de transferencia $K = 0.01$ V/kHz.

7-5. En el detector balanceado por pendiente de la fig. 7-3a, la frecuencia central $f_c = 20.4$ MHz y la desviación máxima de frecuencia es $\Delta f = 50$ kHz. Calcule las frecuencias de corte superior e inferior para el circuito sintonizado.

7-6. En el discriminador de Foster-Seeley de la fig. 7-4, $V_{C1} = 1.2$ V y $V_{C2} = 0.8$ V; determine V_{sal}.

7-7. Para el detector de relación de la fig. 7-6, $V_{C1} = 1.2$ V y $V_{C2} = 0.8$ V; determine V_{sal}.

7-8. Determine la S/N en postdetección para un modulador de FM, con factor de mejoramiento de FM de 23 dB y S/N = 26 dB en la entrada.

7-9. Con la fig. 7-11, determine el factor aproximado de mejoramiento de FM para una S/N = 10.5 dB en la entrada y $m = 1$.

7-10. Calcule la relación mínima de S/N en la entrada, que requiere un receptor con 15 dB de mejoramiento de FM, cifra de ruido NF = 6 dB y una relación deseada de señal a ruido = 38 dB en post detección.

7-11. Un receptor de FM tiene 200 kHz de ancho de banda, cifra de ruido NF = 8 dB y temperatura de ruido en la entrada $T = 100°$ C. Determine la potencia mínima de portadora en la entrada para lograr una S/N = 40 dB en postdetección. Use el diagrama de bloques del receptor de la fig. 7-1 como modelo de receptor, y la curva de umbral de FM de la fig. 7-11.

7-12. Para un receptor de FM sintonizado a 94.5 MHz que usa inyección lateral superior y una primera FI de 10.7 MHz, determine la frecuencia imagen y la frecuencia del oscilador local.

7-13. Determine V_{sal} para un receptor de FM con desviación de frecuencia en la entrada $\Delta f = 50$ kHz y una relación de transferencia $K = 0.02$ V/kHz.

7-14. Calcule las frecuencias de corte superior e inferior del circuito de detector balanceado por pendiente de la fig. 7.3a; tiene frecuencia central $f_c = 10.7$ MHz y desviación máxima de frecuencia en la entrada $\Delta f = 40$ kHz.

7-15. Para el discriminador de Foster-Seeley de la fig. 7-4, $V_{C1} = 1.6$ V y $V_{C2} = 0.4$ V. Calcule V_{sal}.

7-16. Para el detector de relación de la fig. 7-6, $V_{C1} = 1.6$ V. Determine V_{sal}.

7-17. Calcule la S/N de postdetección para un demodulador de FM con factor de mejoramiento de FM igual a 18 dB y una señal a ruido (predetección) en la entrada $S_i/N_i = 32$ dB.

7-18. Con la fig. 7-11 determine el factor aproximado de mejoramiento con una S/N = 11 dB en la entrada y $m = 1$.

CAPÍT

Imagen protegida por derechos de autor

Líneas de transmisión

INTRODUCCIÓN

Una *línea de transmisión* es un *sistema de conductores metálicos* para transferir energía eléctrica de un punto a otro. En forma más específica, una línea de transmisión consiste en dos o más conductores separados por un aislador, como por ejemplo un par de alambres o un sistema de pares de alambres. Una línea de transmisión puede tener desde unas pocas pulgadas hasta varios miles de millas de longitud. Se pueden usar las líneas de transmisión para propagar cd o ca de baja frecuencia, como la corriente eléctrica de 60 ciclos y las señales de audio; también se pueden usar para propagar frecuencias muy altas, como las señales de frecuencia intermedia y de radiofrecuencia. Cuando propagan señales de baja frecuencia, el comportamiento de una línea de transmisión es bastante sencillo y muy predecible; sin embargo, cuando propagan señales de alta frecuencia se complican las características de las líneas de transmisión, y su comportamiento es algo especial para un estudiante de circuitos y sistemas de elementos concentrados y constantes.

ONDAS ELECTROMAGNÉTICAS TRANSVERSALES

La propagación de la energía eléctrica por una línea de transmisión se hace en forma de *ondas electromagnéticas transversales* (EMT). Una onda es un *movimiento oscilatorio*. La vibración de una partícula excita vibraciones semejantes en las partículas vecinas. Una onda EMT se propaga principalmente en el no conductor (dieléctrico) que separa los dos conductores de una línea de transmisión. En consecuencia, la onda viaja, o se propaga, a través de un medio. Para una onda transversal, la dirección del desplazamiento es perpendicular a la dirección de propagación. Una onda superficial de agua es una onda longitudinal. Una onda en la que el desplazamiento tiene la dirección de propagación se llama *onda longitudinal*. Las ondas sonoras son longitudinales. Una onda electromagnética (EM) se produce por la aceleración de una carga eléctrica. En un conductor, la corriente y el voltaje siempre se acompañan por un campo eléctrico E y un campo magnético H en la región vecina del espacio. La fig. 8-1a muestra las relaciones espaciales entre los campos E y H de una onda electromagnética. En la fig. 8-1b se ven los cortes transversales de los campos E y H que rodean a una línea de dos alambres paralelos

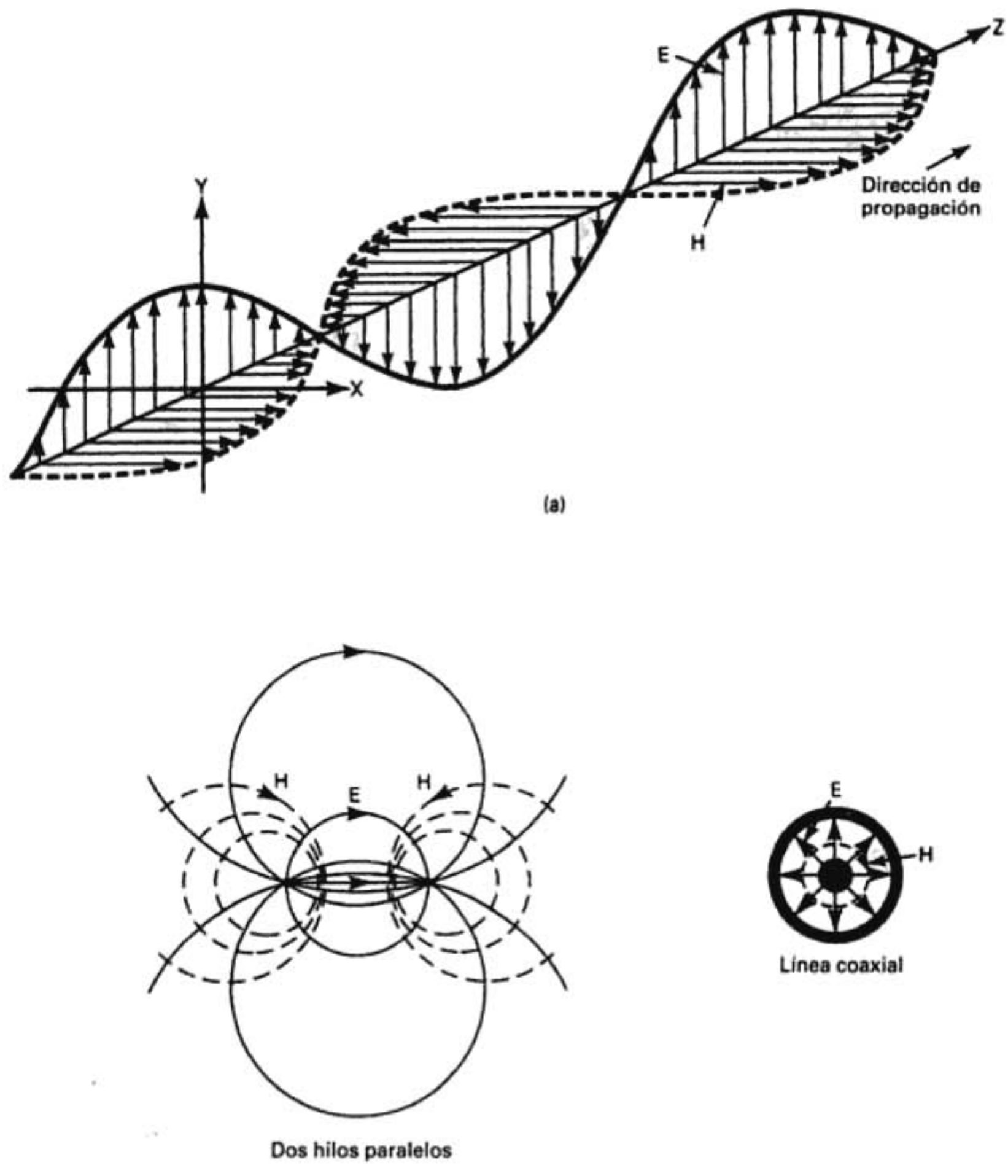


FIGURA 8-1 Vistas: (a) en perspectiva; (b) transversal indicando el desplazamiento relativo de los campos *E* y *H* en una línea de transmisión

y una coaxial. Se puede ver que los campos *E* y *H* son perpendiculares (en ángulo de 90°) entre sí en todos los puntos. Esto se llama *cuadratura espacial*. Las ondas electromagnéticas que viajan por una línea de transmisión, de la fuente a la carga, se llaman *ondas incidentes*, y las que regresan de la carga a la fuente se llaman *ondas reflejadas*.

Características de las ondas electromagnéticas

Velocidad de la onda. Las ondas viajan a diversas velocidades que dependen del tipo de onda y de las características del medio de propagación. Las ondas sonoras viajan aproximadamente a 335 m/s en la atmósfera normal. Las ondas electromagnéticas viajan mucho más rápido. En el espacio libre (es decir, en el vacío), las EMT viajan a la velocidad de la luz, $c = 299{,}793{,}000$ m/s, que se redondea a 3×10^8 m/s. Sin embargo, en el aire de la atmósfera terrestre, viajan un poco más despacio, y por una línea de transmisión viajan con bastante menor velocidad.

Frecuencia y longitud de onda. Las oscilaciones de una onda electromagnética son periódicas y repetitivas. Por consiguiente, se caracterizan por su frecuencia. La rapidez con que se repite la onda periódica es su frecuencia. La distancia de un ciclo en el espacio se llama *longitud de onda*, y se calcula con la siguiente ecuación fundamental

$$\text{distancia} = \text{velocidad} \times \text{tiempo} \tag{8-1}$$

Si se sustituye el tiempo de un ciclo (un *periodo*) en la ecuación 8-1 se obtiene la longitud de un ciclo, que se llama longitud de onda, y cuyo símbolo es la letra griega lambda minúscula (λ).

$$\begin{aligned}\lambda &= \text{velocidad} \times \text{periodo}\\ &= v \times T\end{aligned}$$

Y como $T = 1/f$,

$$\lambda = \frac{v}{f} \tag{8-2}$$

En la propagación por el espacio vacío $v = c$, y por consiguiente la longitud de un ciclo es

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8 \text{ m/s}}{f \text{ ciclos/s}} = \frac{\text{metros}}{\text{ciclo}} \tag{8-3a}$$

Para calcular la longitud de onda en pulgadas o en pies, la ecuación 8-3a se replantea como sigue

$$\lambda = \frac{11.8 \times 10^9 \text{ pulg/s}}{f \text{ ciclos/s}} \quad \text{(pulgadas)} \tag{8-3b}$$

$$\lambda = \frac{9.83 \times 10^8 \text{ pies/s}}{f \text{ ciclos/s}} \quad \text{(pies)} \tag{8-3c}$$

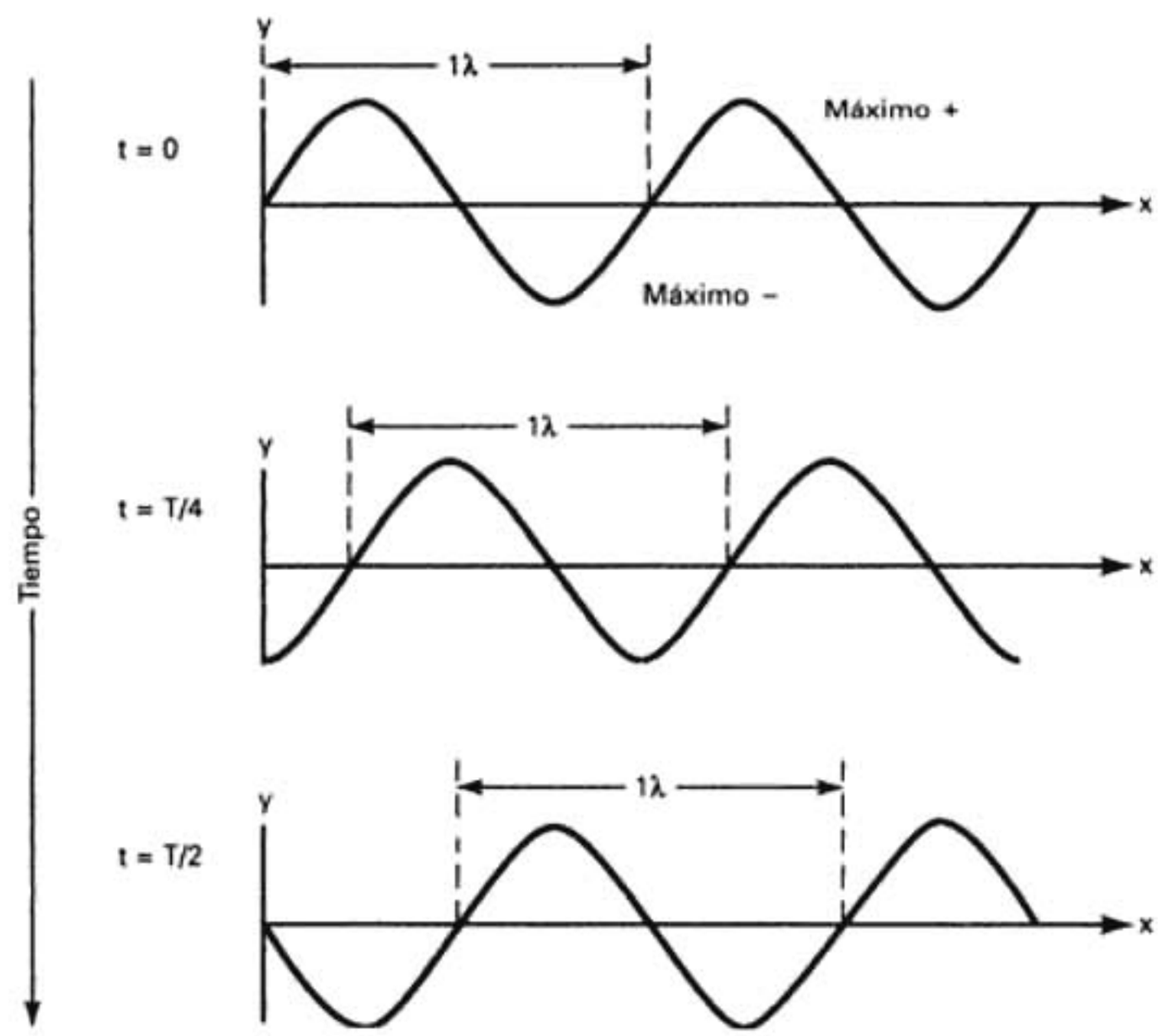


FIGURA 8-2 Desplazamiento y velocidad de una onda transversal al propagarse por una línea de transmisión

La fig. 8-2 es una gráfica del desplazamiento y la velocidad de una onda transversal cuando se propaga en una línea de transmisión, de una fuente a una carga. El eje horizontal (X) es la distancia, y el eje vertical (y) es el desplazamiento. Una longitud de onda es la distancia que ocupa un ciclo de la onda. Se ve que la onda se mueve hacia la derecha, o se propaga por la línea al paso del tiempo. Si se instala un voltímetro en cualquier punto estacionario de la línea, el voltaje medido fluctuará desde cero hasta un máximo positivo, regresará a cero, pasará por un máximo negativo, regresará a cero y a continuación se repetirá el ciclo.

TIPOS DE LÍNEAS DE TRANSMISIÓN

En general, las líneas de transmisión se pueden clasificar en *balanceadas* y *desbalanceadas*. En las líneas balanceadas de dos alambres ambos conductores llevan corriente; uno lleva la señal y el otro es el regreso. Este tipo de transmisión se llama transmisión *diferencial* o *balanceada* de señal. La señal que se propaga por el alambre se mide como diferencia de potencial entre los dos conductores. La fig. 8-3 muestra un sistema de transmisión balanceado. Ambos conductores de una línea balanceada conducen corriente de señal, y las corrientes tienen igual magnitud con respecto a la masa o tierra eléctrica, pero viajan en direcciones opuestas. Las corrientes que fluyen en direcciones opuestas en un par balanceado de alambres se llaman *corrientes de circuito metálico*. Las corrientes que tienen las mismas direcciones se llaman *corrientes longitudinales*. Un par balanceado de alambres tiene la ventaja de que la mayor parte del ruido de interferencia (que a veces se llama *voltaje de modo común*) se induce por igual en ambos conductores, y produce corrientes longitudinales que se anulan en la carga. La anulación de las señales de modo común se llama rechazo de modo común (CMR, de *common-mode rejection*). Son comunes las relaciones de rechazo de modo común (CMRR, de *common-mode rejection ratio*) de 40 a 70 dB.

Todo par de alambres puede trabajar en el modo balanceado, siempre que ninguno de ellos esté al potencial de tierra. Aquí se incluye el cable coaxial que tiene dos conductores centrales y un blindaje. En general, el blindaje se conecta a tierra para evitar que la interferencia estática penetre a los conductores centrales.

La fig. 8-4 muestra el resultado de las corrientes metálicas y longitudinales en una línea de transmisión. Se ve que las corrientes longitudinales, que se producen con frecuencia debido a la interferencia de estática, se anulan en la carga.

En una línea de transmisión desbalanceada, un alambre está al potencial de tierra, mientras que el otro tiene el potencial de una señal. A este tipo de transmisión se le llama transmisión de señal *desbalanceada* o *asimétrica*. En la transmisión desbalanceada, el alambre de tierra puede ser también la referencia para otros conductores portadores de señal. Si ése es el caso, el alambre de tierra debe ir donde vaya cualquiera de los conductores de señal. A veces esto origina problemas, porque un tramo de alambre tiene resistencia, inductancia y capacitancia y, en consecuencia, puede existir una pequeña diferencia de potencial entre dos puntos cualesquiera

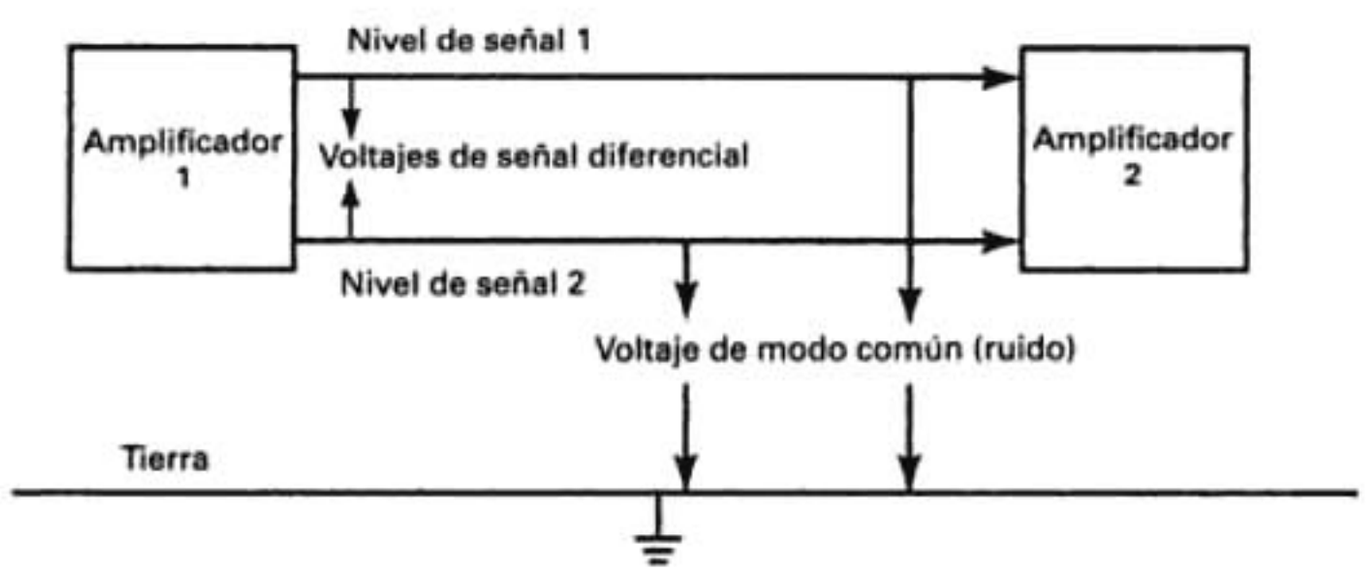


FIGURA 8-3 Sistema de transmisión diferencial o balanceado

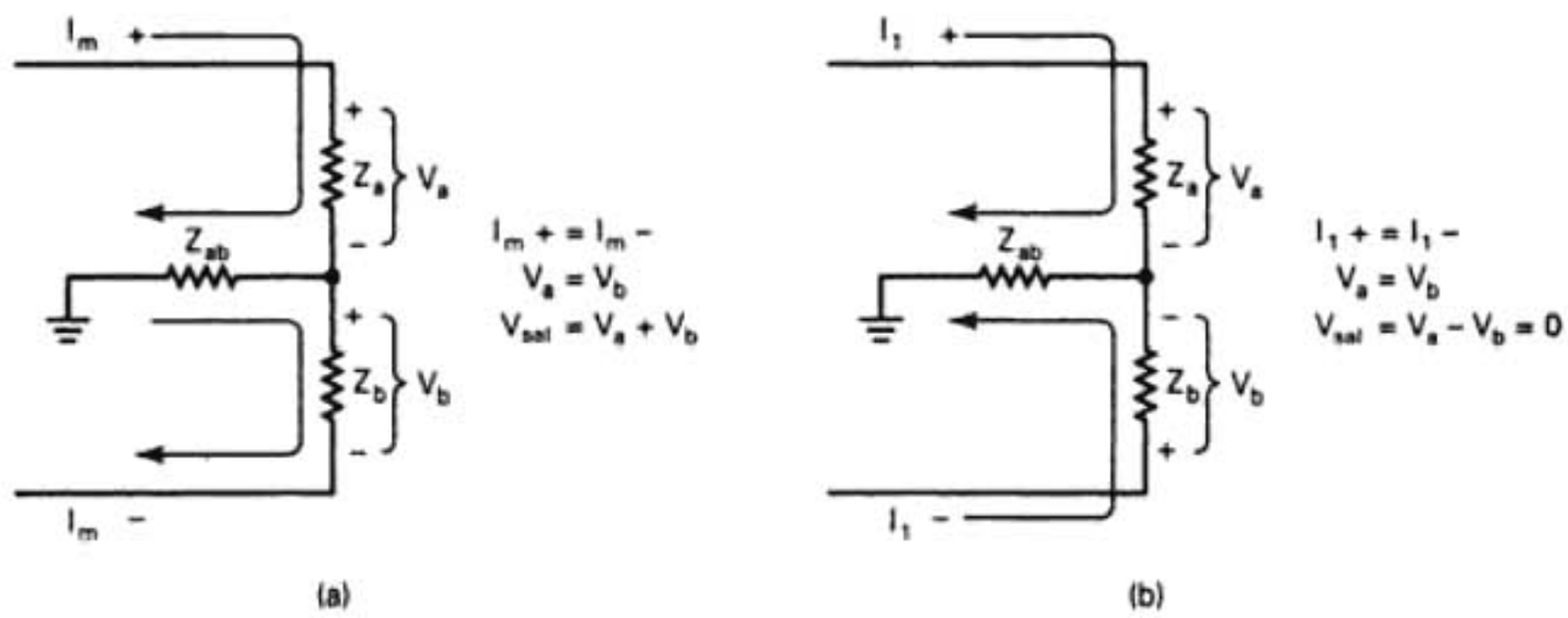


FIGURA 8-4 Resultados de corrientes metálicas y longitudinales en una línea de transmisión balanceada: (a) corrientes metálicas debidas a voltajes de señal; (b) corrientes longitudinales debidas a voltajes de ruido

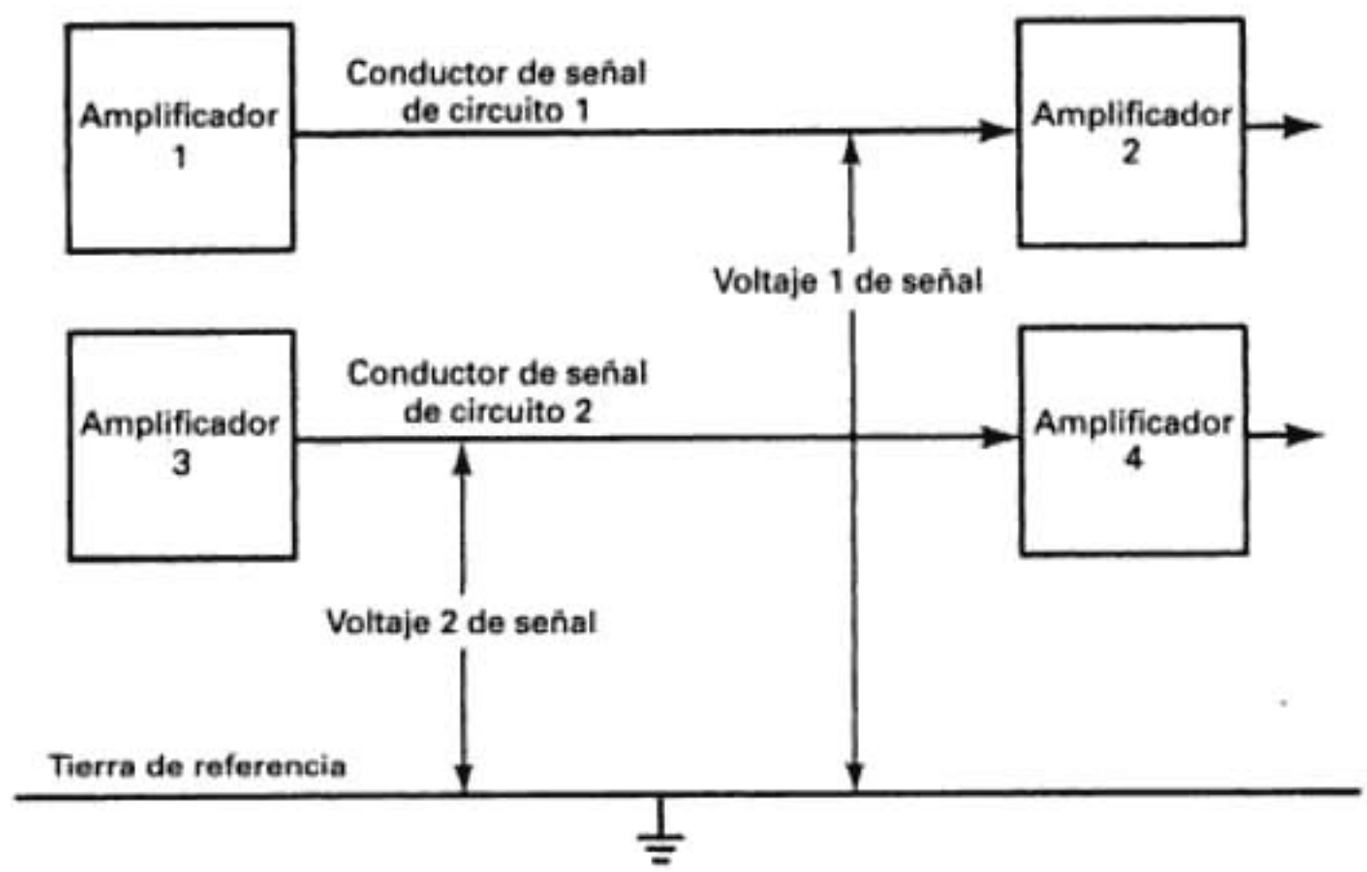


FIGURA 8-5 Sistema de transmisión asimétrico o desbalanceado

en el conductor de tierra. En consecuencia, ese conductor no es un punto de referencia perfecto, y puede tener ruido inducido en él. Un cable coaxial normal de dos conductores es una línea desbalanceada. El segundo conductor es el blindaje, que casi siempre se conecta a tierra.

La fig. 8-5 muestra dos sistemas desbalanceados de transmisión. La diferencia de potencial en cada alambre de señal se mide entre él y la tierra. Las líneas de transmisión balanceadas se pueden conectar a líneas desbalanceadas, y viceversa, con transformadores especiales llamados *balunes*.

Balunes

Un dispositivo que se usa para conectar una línea de transmisión balanceada con una carga desbalanceada se llama *balún* (*balanceado a desbalanceado*, de *balanced to unbalanced*). También, lo que es más común, una línea de transmisión desbalanceada, como un cable coaxial, se puede conectar con una carga balanceada, como una antena, mediante un transformador especial con desbalanceado primario y devanado secundario con toma central. El conductor externo (*blindaje*) de una línea de transmisión desbalanceada se suele conectar a tierra. A frecuencias relativamente bajas se puede usar un transformador ordinario para aislar la tierra de la carga, como se ve en la fig. 8-6a. El balún debe tener un blindaje electrostático conectado a tierra física, para reducir al mínimo los efectos de las capacitancias parásitas.

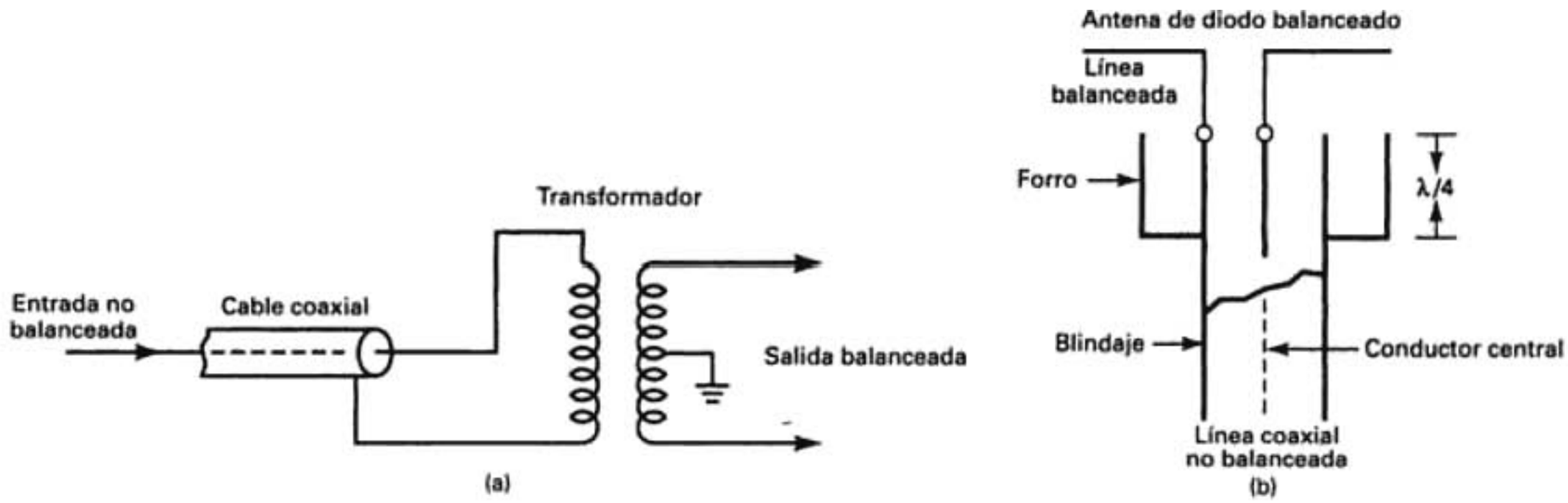


FIGURA 8-6 Balunes: (a) balún de transformador; (b) balún de bazuca

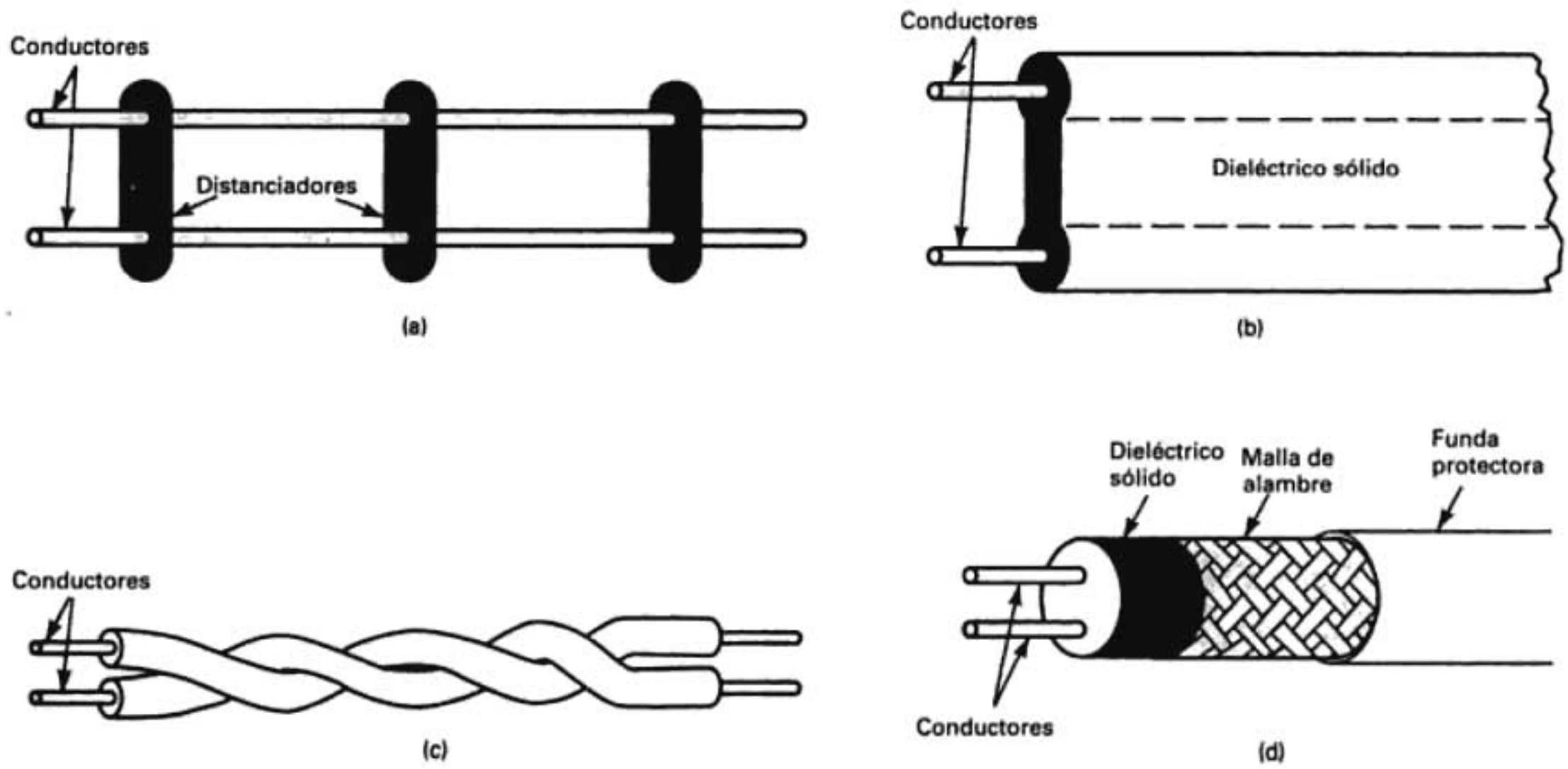


FIGURA 8-7 Líneas de transmisión: (a) alambres desnudos; (b) conductores gemelos; (c) par trenzado; (d) par blindado

Cuando las frecuencias son relativamente altas se usan balunes de varios tipos para líneas de transmisión. El más común es el balún de *banda angosta*, que a veces se llama *choke*, *forro* o balún *bazuca*, y se ve en la fig. 8-6b. Un choke de cuarto de onda se instala en torno al conductor externo de un cable coaxial y se conecta con él. Así, la impedancia que se ve hacia la línea de transmisión se forma por el choke y el conductor externo, y es igual a infinito, es decir, el conductor externo ya no tiene impedancia cero a tierra. Por lo anterior, un alambre del par balanceado se puede conectar con el choke sin poner en corto la señal. El segundo conductor se conecta al conductor interno del cable coaxial.

Líneas de transmisión de conductores paralelos

Línea de transmisión de alambre desnudo. Una *línea de transmisión de alambre desnudo* es un *conductor de dos alambres paralelos;* se ve en la fig. 8-7a. Consiste simplemente en dos alambres paralelos a corta distancia y separados por aire. Se colocan espaciadores no conductores a intervalos periódicos, para sostenerlos y mantener constante la distancia entre ellos. La distancia entre los dos conductores en general es entre 2 y 6 pulgadas. El dieléctrico no es

más que el aire entre y en torno a los dos conductores en los que se propaga la EMT. La única ventaja real de este tipo de línea de transmisión es su construcción sencilla. Como no tiene blindaje, las pérdidas por radiación son altas y es susceptible de captar ruido. Son las principales desventajas de una línea de transmisión de cable desnudo. Por consiguiente, estas líneas se trabajan normalmente en el modo balanceado.

Conductores gemelos. Los *conductores gemelos* son otra forma de línea de transmisión de dos alambres paralelos, y se ve en la fig. 8-7b. A los conductores gemelos también se les llama con frecuencia *cable de cinta*. Los conductores gemelos son, en esencia, lo mismo que la línea de transmisión de conductores desnudos, pero los distanciadores entre los dos conductores se reemplazan con un dieléctrico macizo continuo. Así se asegura la distancia uniforme a lo largo de todo el cable, lo cual es una buena característica, por razones que se explicarán después en el capítulo. En forma normal, la distancia entre los dos conductores es 5/16 de pulgada para el cable de transmisión de TV. Los materiales dieléctricos más frecuentes son el teflón y el polietileno.

Cable de par trenzado. Un *cable de par trenzado* se forma torciendo entre sí dos conductores aislados. Con frecuencia, los pares se trenzan en *unidades* y las unidades se llevan en *núcleos* que a su vez se cubren con varios tipos de *forros*, dependiendo de la aplicación. Los pares vecinos se trenzan con distintos *pasos* (longitud de torcimiento) para reducir la interferencia debida a la inducción mutua entre los pares. Las *constantes primarias* del cable de par trenzado son sus parámetros eléctricos: resistencia, inductancia, capacitancia y conductancia, que están sujetas a variaciones de acuerdo con el ambiente físico, como temperatura, humedad y esfuerzos mecánicos, y dependen de las diferencias de manufactura. En la fig. 8-7c se muestra un cable de par trenzado.

Par de cable blindado. Para reducir las pérdidas por radiación y la interferencia, con frecuencia las líneas de transmisión se encierran en una *malla* de alambre metálica y conductora. La malla se conecta a tierra y funciona como blindaje. También, la malla evita que se irradien señales fuera de ella, y evita que la interferencia electromagnética llegue a los conductores de señal. En la fig. 8-7d se ve un par de cable blindado. Está formado por dos alambres conductores paralelos separados por un material dieléctrico macizo. Toda la estructura se encierra en un tubo de conductor integrado por una malla, y después se cubre con una capa protectora de plástico.

Líneas de transmisión concéntricas o coaxiales

Las líneas de transmisión de conductores paralelos son adecuadas para aplicaciones en baja frecuencia. Sin embargo, con las altas frecuencias aumentan demasiado sus pérdidas por radiación y en dieléctrico, así como su susceptibilidad a la interferencia externa. Por lo anterior, se usan mucho los *conductores coaxiales* en aplicaciones de alta frecuencia, para reducir las pérdidas y para aislar las trayectorias de transmisión. El cable coaxial básico consiste en un conductor central rodeado por un *conductor externo concéntrico*, a distancia uniforme del centro. A frecuencias de trabajo relativamente altas, el conductor externo coaxial proporciona un excelente blindaje contra la interferencia externa. Sin embargo, no es económico usar un blindaje con frecuencias relativamente bajas. También, casi siempre el conductor externo de un cable coaxial se conecta a tierra, y eso limita su empleo a aplicaciones desbalanceadas o asimétricas.

En esencia hay dos tipos de cables coaxiales: líneas *rígidas llenas de aire* o *líneas flexibles macizas*. La fig. 8-8a muestra una línea coaxial rígida de aire. Se ve que el conductor central está coaxialmente rodeado por un conductor externo tubular, y que el material aislador es aire. El conductor externo está aislado físicamente, y separado del conductor central por un distanciador, que puede ser de vidrio *pyrex*, poliestireno o algún otro material no conductor. La fig. 8-8b representa un cable coaxial flexible y macizo. El conductor externo es una malla de alambre flexible, y es coaxial respecto al conductor central. El material aislante es polietileno macizo no conductor, que proporciona tanto soporte como aislamiento eléctrico entre los conductores interno y externo. El conductor interno es un alambre flexible de cobre, que puede ser macizo o hueco.

Es relativamente costoso fabricar los cables coaxiales rígidos de aire, y para minimizar las pérdidas, el aislador de aire debe estar relativamente libre de humedad. Los cables coaxiales

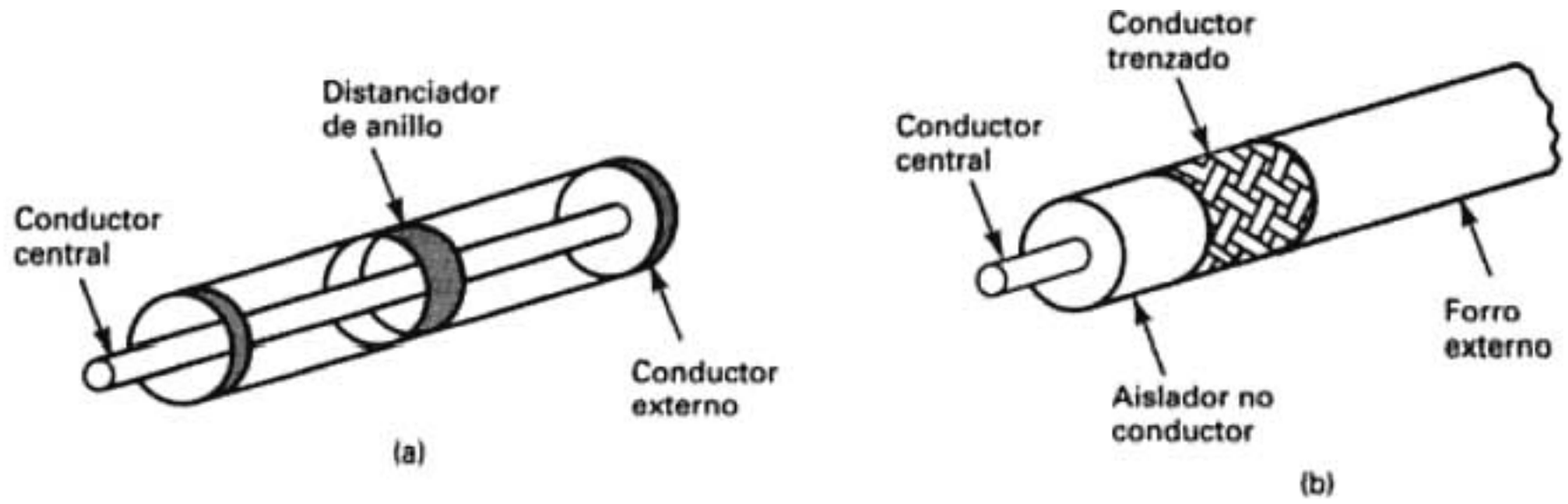


FIGURA 8-8 Líneas de transmisión concéntricas o coaxiales: (a) rígida llena de aire; (b) línea maciza flexible

macizos tienen menos pérdidas y son más fáciles de fabricar, instalar y mantener. Los dos tipos de cable coaxial son relativamente inmunes a la radiación externa, irradian poco ellos mismos, y pueden funcionar a mayores frecuencias que sus contrapartes de conductores paralelos. Las desventajas básicas de las líneas coaxiales de transmisión son su alto costo y que se deben usar en el modo desbalanceado.

CIRCUITO EQUIVALENTE DE UNA LÍNEA DE TRANSMISIÓN

Líneas uniformemente distribuidas

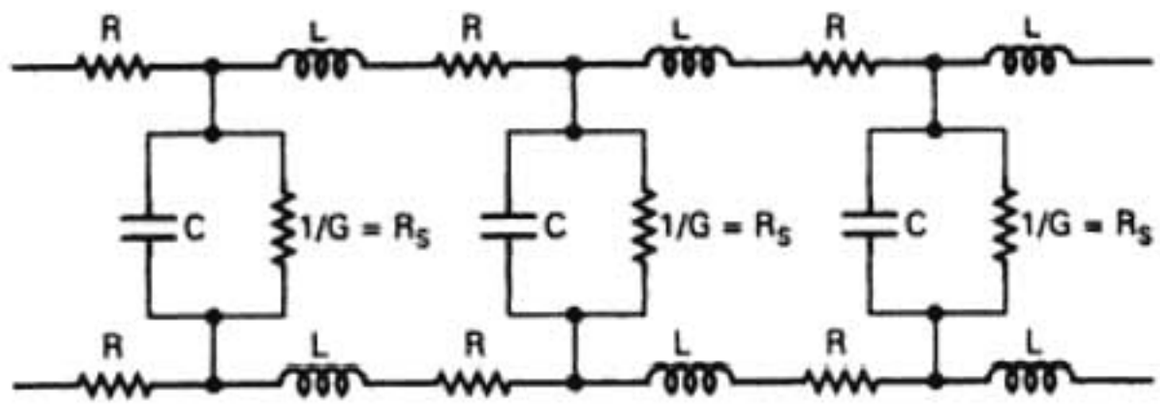
Las características de una línea de transmisión están determinadas por sus propiedades eléctricas, como por ejemplo la conductividad de los alambres y la constante dieléctrica del aislamiento, y de sus propiedades físicas, como diámetro del alambre y distancia entre conductores. Estas propiedades, a su vez, determinan las constantes eléctricas primarias: resistencia de cd en serie (R), inductancia en serie (L), capacitancia en paralelo (C) y conductancia en paralelo (G). A lo largo de la línea hay resistencia e inductancia, mientras que entre los dos conductores se desarrollan capacitancia y conductancia. Las constantes primarias se distribuyen uniformemente en toda la longitud de la línea y, en consecuencia, se les llama *parámetros distribuidos*. Para simplificar el análisis, los parámetros distribuidos se *agrupan* entre sí por unidad de longitud, para formar un modelo eléctrico artificial de la línea. Por ejemplo, la resistencia en serie se especifica en general en ohms por unidad de longitud (por ejemplo, ohms por metro).

La fig. 8-9 muestra el circuito eléctrico equivalente de una línea de transmisión metálica de dos conductores, donde se muestra la colocación relativa de los diversos parámetros agrupados. La conductancia entre los dos alambres se muestra en su forma recíproca, y se menciona como resistencia de fugas en paralelo, R_s.

Características de transmisión

Las características de transmisión de una línea se llaman *constantes secundarias*, y se calculan a partir de las cuatro constantes primarias. Las constantes secundarias son la impedancia característica y la constante de propagación.

Impedancia característica. Para que haya una transferencia máxima de energía de la fuente a la carga, es decir, que no haya energía reflejada, una línea de transmisión debe terminar en una carga puramente resistiva, igual a la *impedancia característica* de ella. La impedancia característica, Z_o, de una línea de transmisión es una cantidad compleja que se expresa en ohms, y que en el caso ideal es independiente de la longitud de la línea y que no se puede medir. Esa impedancia característica, se define como la impedancia vista hacia una línea de longitud infinita, o la impedancia vista hacia una línea de longitud finita que termina en una carga puramente resistiva igual a la impedancia característica de la línea. Una línea de transmisión almacena energía en su inductancia y capacitancia distribuidas. Si la línea es infinitamente larga, puede almacenar energía en forma indefinida; la energía procede de la fuente, entra a la línea y nada regresa. Por lo mismo, la línea funciona como un resistor que disipa toda la



C = capacitancia – dos conductores separados por un aislador
R = resistencia – oposición al flujo de corriente
L = inductancia propia
1/G = resistencia de pérdida del dieléctrico
R_s = resistencia de pérdida en paralelo

FIGURA 8-9 Línea de transmisión de dos hilos paralelos; circuito eléctrico equivalente

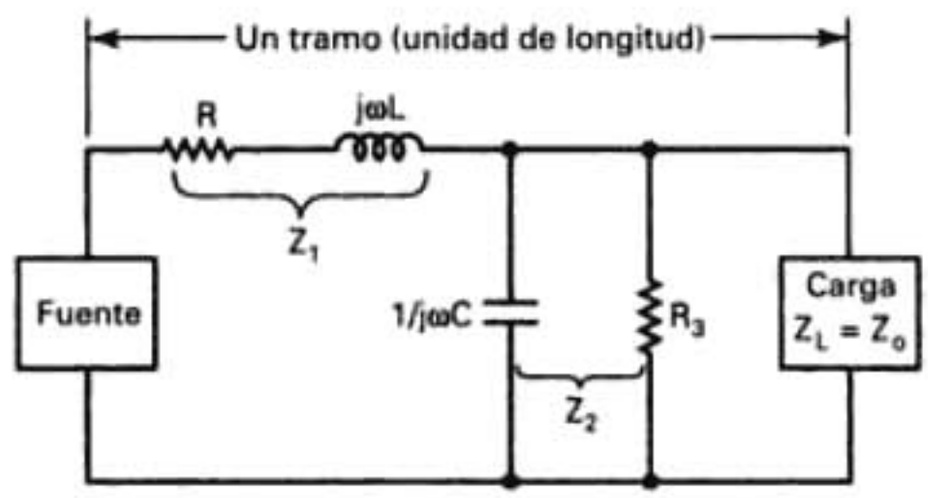


FIGURA 8-10 Circuito equivalente de un solo tramo de línea de transmisión terminada en una carga igual a Z_o

energía. Se puede simular una línea infinita si una línea finita termina en una carga puramente resistiva igual a Z_o; toda la energía que entra de la fuente a la línea se disipa en la carga (esto supone que la línea es totalmente sin pérdidas).

La fig. 8-10 muestra una sola sección de una línea de transmisión terminada en una carga Z_L que es igual a Z_o. La impedancia que se ve hacia una línea de n de dichas secciones se calcula con la siguiente ecuación

$$Z_o^2 = Z_1 Z_2 + \frac{Z_L^2}{n} \tag{8-4}$$

siendo n la cantidad de secciones. Para una cantidad infinita de secciones, Z_L^2/n tiende a 0 si

$$\lim \frac{Z_L^2}{n}\Big|_{n\to\infty} = 0$$

Entonces,
$$Z_o = \sqrt{Z_1 Z_2}$$

donde
$$Z_1 = R + j\omega L$$

$$Y_2 = \frac{1}{Z_2} = \frac{1}{R_s} + \frac{1}{1/j\omega C}$$
$$= G + j\omega C$$

$$Z_2 = \frac{1}{G + j\omega C}$$

Por consiguiente,
$$Z_o = \sqrt{(R + j\omega L)\frac{1}{G + j\omega C}}$$

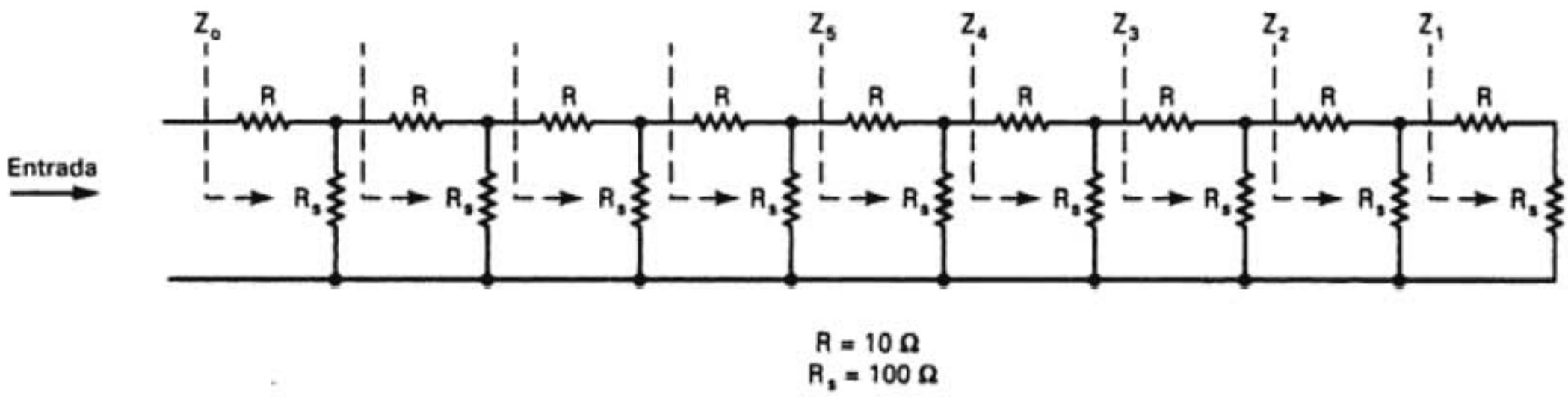


FIGURA 8-11 Impedancia característica de una línea de transmisión de una cantidad infinita de tramos, o una terminada en una carga igual a Z_o

o sea
$$Z_o = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \tag{8-5}$$

Para frecuencias extremadamente bajas, las resistencias dominan, y la ecuación 8-5 se simplifica a

$$Z_o = \sqrt{\frac{R}{G}} \tag{8-6}$$

Cuando las frecuencias son extremadamente altas, dominan la inductancia y la capacitancia, y la ecuación 8-5 se simplifica a

$$Z_o = \sqrt{\frac{j\omega L}{j\omega C}} = \sqrt{\frac{L}{C}} \tag{8-7}$$

Se puede ver en esta ecuación que para altas frecuencias, la impedancia característica de una línea de transmisión tiende a una constante, es independiente de la frecuencia y la longitud, y sólo se determina por la inductancia y la capacitancia distribuidas. También se puede ver que el ángulo de fase es 0°. Por consiguiente, Z_0 se ve como puramente resistiva, y toda la energía incidente la absorbe la línea.

Desde un punto de vista puramente resistivo se puede apreciar con facilidad que la impedancia, viendo hacia una línea de transmisión formada por una cantidad infinita de secciones, tiende a la impedancia característica. Esto se ve en la fig. 8-11. De nuevo y para simplificar, sólo se consideran la resistencia en serie R y la resistencia en paralelo R_s. La impedancia vista hacia la última sección de la línea no es más que la suma de R y R_s. La ecuación correspondiente es en este caso

$$Z_1 = R + R_s = 10 + 100 = 110$$

Si se agrega una segunda sección, Z_2 será

$$Z_2 = R + \frac{R_s Z_1}{R_s + Z_1} = 10 + \frac{100 \times 110}{100 + 110} = 10 + 52.38 = 62.38$$

y para una tercera sección, Z_3 será

$$Z_3 = R + \frac{R_s Z_2}{R_s + Z_2}$$

$$= 10 + \frac{100 \times 62.38}{100 + 62.38} = 10 + 38.42 = 48.32$$

La Z_4 con una cuarta sección es

$$Z_4 = 10 + \frac{100 \times 48.32}{100 + 48.32} = 10 + 32.62 = 42.62$$

Se puede apreciar que después de cada sección individual, la impedancia total vista hacia la línea disminuye respecto a su valor anterior; sin embargo, cada vez es menor la magnitud de la disminución. Si se continuara el proceso anterior, la impedancia que se ve hacia la línea disminuiría en forma asintótica hacia 37 Ω, que es la impedancia característica de la línea.

Si la línea de transmisión de la fig. 8-11 se terminara en una resistencia de carga $Z_L = 37\ \Omega$, la impedancia que se ve hacia cualquier cantidad de secciones sería igual a 37 Ω, la impedancia característica. Para una sola sección de línea, Z_o es

$$Z_o = Z_1 = R + \frac{R_s \times Z_L}{R_s + Z_L} = 10 + \frac{100 \times 37}{100 + 37} = 10 + \frac{3700}{137} = 37\ \Omega$$

Al agregar una segunda sección, Z_2 es

$$Z_o = Z_2 = R + \frac{R_s \times Z_1}{R_s + Z_1} = 10 + \frac{100 \times 37}{100 + 37} = 10 + \frac{3700}{137} = 37\ \Omega$$

Por consiguiente, si esta línea terminara en una resistencia de carga $Z_L = 37\ \Omega$, $Z_o = 37\ \Omega$, sin importar cuántas secciones se incluyan.

También se puede determinar la impedancia característica de una línea de transmisión con la ley de Ohm. Cuando una fuente se conecta a una línea de longitud infinita y se aplica un voltaje, pasa una corriente. Aun cuando la carga esté abierta, el circuito se completa a través de las constantes distribuidas de la línea. La impedancia característica no es más que la relación del voltaje de la fuente (E_o) a la corriente de la línea (I_o). La ecuación es

$$Z_o = \frac{E_o}{I_o} \tag{8-8}$$

La impedancia característica de una línea de dos conductores paralelos con dieléctrico de aire se puede determinar a partir de sus dimensiones físicas (véase la fig. 8-12a), con la fórmula

$$Z_o = 276 \log \frac{D}{r} \tag{8-9}$$

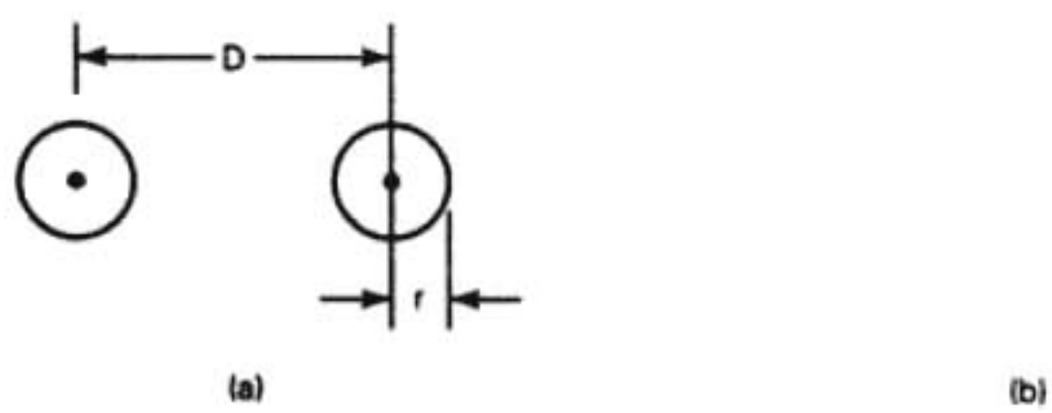


FIGURA 8-12 Dimensiones físicas de las líneas de transmisión: (a) línea de transmisión de dos alambres paralelos; (b) línea de transmisión de cable coaxial

en donde D = distancia entre los centros de los dos conductores (pulg)
r = radio del conductor (pulg)

y $D >> r$.

Ejemplo 8-1

Calcular la impedancia característica de una línea de transmisión de dos conductores paralelos en aire, con una relación $D/r = 12.22$.

Solución Al sustituir en la ecuación 8-9 se obtiene

$$Z_o = 276 \log 12.22 = 300\ \Omega$$

La impedancia característica de un cable concéntrico coaxial también se puede calcular a partir de sus dimensiones físicas (véase la fig. 8-12b) con la fórmula

$$Z_o = \frac{138}{\sqrt{\epsilon_r}} \log \frac{D}{d} \qquad (8\text{-}10)$$

en la que D = diámetro interno del conductor externo (pulg)
d = diámetro exterior del conductor interno (pulg)
ϵ_r = constante dieléctrica (permeabilidad relativa) del material aislante (sin unidades)

Ejemplo 8-2

Determinar la impedancia caracterítica de un cable coaxial RG-59A cuyas especificaciones son las siguientes: $L = 0.118\ \mu$H/pie, $C = 21$ pF/pie, $d = 0.025$ pulg, $D = 0.15$ pulg y $\epsilon = 2.23$.

Solución Se sustituyen valores en la ecuación 8-7 y el resultado es

$$Z_o = \sqrt{\frac{L}{C}} = \sqrt{\frac{0.118 \times 10^{-6}\text{H/pie}}{21 \times 10^{-12}\text{pF/pie}}} = 75\ \Omega$$

Al sustituir en la ecuación 8-10 se obtiene

$$Z_o = \frac{138}{\sqrt{2.23}} \log \frac{0.15\ \text{pulg.}}{0.25\ \text{pulg.}} = 71.9\ \Omega$$

Se pueden resumir como sigue las líneas de transmisión:

1. La impedancia de entrada de una línea infinitamente larga a radiofrecuencias es resistiva, igual a Z_o.
2. Las ondas electromagnéticas recorren la línea sin reflexiones. A esa línea se le llama *no resonante*.
3. La relación de voltaje a corriente en cualquier punto de la línea es igual a Z_o.
4. El voltaje y la corriente incidentes en cualquier punto de la línea están en fase.
5. Las pérdidas en una línea no resonante son mínimas, por unidad de longitud.
6. Toda línea de transmisión que esté terminada en una carga puramente resistiva igual a Z_o funciona como si fuera una línea infinita.
   (a) $Z_i = Z_o$.
   (b) No hay ondas reflejadas.
   (c) V e I están en fase.
   (d) Hay un máximo de transferencia de energía de la fuente a la carga.

Constante de propagación. La *constante de propagación*, que a veces se le llama *coeficiente de propagación* se usa para expresar la atenuación (pérdida de señal) y el desplazamiento de fase por unidad de longitud de una línea de transmisión. Cuando una onda se propaga por una línea de transmisión disminuye su amplitud con la distancia recorrida. Se usa la constante de propagación para determinar la reducción de voltaje o de corriente con la distancia, cuando una onda EMT se propaga por una línea de transmisión. Cuando la línea es de longitud infinita, toda la potencia incidente se disipa en la resistencia del conductor al avanzar la onda por la línea. Por

consiguiente, con una línea de longitud infinita, o una que se vea infinitamente larga, como puede ser una línea finita terminada en una carga equilibrada ($Z_o = Z_L$), no regresa o se refleja energía alguna hacia la fuente. La ecuación de la constante de propagación es

$$\gamma = \alpha + j\beta \tag{8-11a}$$

siendo γ = constante de propagación
α = coeficiente de atenuación (nepers por unidad de longitud)
β = coeficiente de desplazamiento de fase (radianes por unidad de longitud)

La constante de propagación es una cantidad compleja y se define como sigue

$$\gamma = \sqrt{(R + j\omega L)(G + j\omega C)} \tag{8-11b}$$

Como en cada distancia igual a la longitud de onda se produce un desplazamiento de fase de 2π,

$$\beta = \frac{2\pi}{\lambda} \tag{8-12}$$

A frecuencias intermedias y de radio, $\omega L > R$ y $\omega C > G$, entonces

$$\alpha = \frac{R}{2Z_o} + \frac{GZ_o}{2} \tag{8-13}$$

y

$$\beta = \omega\sqrt{LC} \tag{8-14}$$

La distribución de corriente y voltaje a lo largo de una línea de transmisión que termina en una carga igual a su impedancia característica (una línea equilibrada) se calculan con las siguientes fórmulas

$$I = I_s e^{-l\gamma} \tag{8-15}$$

$$V = V_s e^{-l\gamma} \tag{8-16}$$

en las que I_s = corriente en el extremo de la línea que da a la fuente (amp)
V_s = voltaje en el extremo de la línea que da a la fuente (volts)
γ = constante de propagación
l = distancia de la fuente hasta donde se determina la corriente o el voltaje

Para una carga equilibrada $Z_L = Z_o$, y para determinada longitud del cable l, la pérdida de voltaje o corriente de señal es la parte real de γl, y el desplazamiento de fase es la parte imaginaria.

PROPAGACIÓN DE ONDAS EN LÍNEA DE TRANSMISIÓN

Las ondas electromagnéticas viajan a la velocidad de la luz cuando se propagan en el vacío, y casi a la velocidad de la luz cuando lo hacen a través del aire. Sin embargo, en las líneas metálicas de transmisión, donde el conductor suele ser cobre, y en los materiales dieléctricos, la velocidad varía mucho de acuerdo con el tipo de cable, y una onda electromagnética viaja con mucha mayor lentitud.

Factor de velocidad

El *factor de velocidad* (llamado a veces *constante de velocidad*) se define como la relación de la velocidad real de propagación a través de determinado medio, entre la velocidad de propagación a través del espacio vacío. La definición matemática del factor de velocidad es

$$V_f = \frac{V_p}{c} \tag{8-17}$$

donde V_f = factor de velocidad (adimensional)
V_p = velocidad real de propagación (metros por segundo)
c = velocidad de propagación a través del espacio vacío ($c = 3 \times 10^8$ m/s)

TABLA 8-1 Factores de velocidad

Material	Factor de velocidad
Aire	0.95–0.975
Hule	0.56–0.65
Polietileno	0.66
Teflón	0.70
Espuma de teflón	0.82
Espigas de teflón	0.81
Espiral de teflón	0.81

TABLA 8-2 Constantes dieléctricas

Material	Constante dieléctrica relativa (ϵ_r)
Vacío	1.0
Aire	1.0006
Teflón	2.1
Polietileno (PE)	2.27
Poliestireno	2.5
Papel parafinado	2.5
Hule	3.0
Cloruro de polivinilo (PVC)	3.3
Mica	5.0
Vidrio	7.5

y
$$V_f \times c = V_p$$

La velocidad a la que viaja una onda electromagnética por una línea de transmisión depende de la constante dieléctrica del material aislante que separa a los dos conductores. El factor de velocidad se calcula en forma muy aproximada con la fórmula

$$V_f = \frac{1}{\sqrt{\epsilon_r}} \tag{8-18}$$

en la que ϵ_r es la constante dieléctrica del material dado (la permitividad del material en relación con la permitividad en el vacío; es la misma relación ϵ/ϵ_o).

La constante dieléctrica es tan sólo la *permitividad relativa* de un material. La constante dieléctrica del aire es 1.0006. Sin embargo, la constante dieléctrica de los materiales que se usan con frecuencia en las líneas de transmisión va de 1.2 a 2.8, que dan lugar a factores de velocidad de 0.6 a 0.9. En la tabla 8-1 se muestran los factores de velocidad para varias configuraciones comunes en líneas de transmisión, y las constantes dieléctricas de varios materiales aislantes se ven en la tabla 8-2.

La constante dieléctrica depende del material que se use. Los inductores almacenan energía magnética, y los capacitores almacenan energía eléctrica. Se necesita un tiempo finito para que un inductor o un capacitor tome o ceda energía. Por consiguiente, la velocidad con la que se propaga una onda electromagnética por una línea de transmisión varía de acuerdo con la inductancia y la capacitancia. Se puede demostrar que ese tiempo de carga es $T = \sqrt{LC}$. Así, la inductancia, capacitancia y velocidad de propagación se relacionan mediante la fórmula

$$\text{velocidad} \times \text{tiempo} = \text{distancia}$$

Por consiguiente,
$$V_p = \frac{\text{distancia}}{\text{tiempo}} = \frac{D}{T} \tag{8-19}$$

Al sustituir el tiempo se obtiene $$V_p = \frac{D}{\sqrt{LC}} \tag{8-20}$$

Si la distancia se normaliza a 1 m, la velocidad de propagación de una línea sin pérdidas es

$$V_p = \frac{1 \text{ m}}{\sqrt{LC}} = \frac{1}{\sqrt{LC}} \text{ m/s} \tag{8-21}$$

donde V_p = velocidad de propagación (metros por segundo)
$\sqrt{LC}$ = segundos

Ejemplo 8-3

Para un tramo de cable coaxial RG8A/U con capacitancia distribuida $C = 96.6$ pF/m, inductancia distribuida $L = 241.56$ nH/m y constante dieléctrica relativa $\epsilon_r = 2.3$, determinar la velocidad de propagación y el factor de velocidad.

Solución De acuerdo con la ecuación 8-21,

$$V_p = \frac{1}{\sqrt{96.6 \times 10^{-12} \times 241.56 \times 10^{-9}}} = 2.07 \times 10^8 \text{ m/s}$$

De la ecuación 8-17,

$$V_f = \frac{2.07 \times 10^8 \text{ m/s}}{3 \times 10^8 \text{ m/s}} = 0.69$$

Sustituyendo en la ecuación 8-18,

$$V_f = \frac{1}{\sqrt{2.3}} = 0.66$$

Como la longitud de onda es directamente proporcional a la velocidad, y la velocidad de propagación de una onda EMT varía en forma directa con la constante dieléctrica, la longitud de una onda EMT también varía con la constante dieléctrica. Así, para medios de transmisión distintos del espacio vacío, se puede replantear como sigue la ecuación 8-3a

$$\lambda = \frac{V_p}{f} = \frac{cV_f}{f} = \frac{c}{f\sqrt{\epsilon_r}} \tag{8-22}$$

Longitud eléctrica de una línea de transmisión

La longitud de una línea de transmisión, en relación con la longitud de la onda que se propaga por ella, es una consideración importante al analizar el comportamiento de la línea. A bajas frecuencias (grandes longitudes de onda), el voltaje a lo largo de la línea permanece relativamente constante. Sin embargo, con las altas frecuencias pueden estar varias longitudes de onda de la señal en la línea, al mismo tiempo. Así, puede variar bastante el voltaje a lo largo de la línea. En consecuencia, a menudo la longitud de la línea de transmisión se especifica en longitudes de onda, y no en dimensiones lineales. Los fenómenos de línea de transmisión se aplican a líneas largas. En general, se define como larga una línea de transmisión si es mayor que un dieciseisavo de longitud de onda; si no es así, se considera como línea corta. Un tramo determinado de línea de transmisión podrá ser corto en cierta frecuencia, y largo en otra. Por ejemplo, un tramo de 10 m de línea de transmisión, a 1000 Hz, es corto ($\lambda = 300{,}000$ m; 10 m es sólo una pequeña fracción de una longitud de onda). Sin embargo, la misma línea es larga a 6 GHz ($\lambda = 5$ cm; la línea tiene 200 longitudes de onda). Después en este capítulo, en el capítulo 9 y en el apéndice A se verá que la longitud eléctrica se usa mucho en los cálculos de línea de transmisión y diseño de antenas.

Líneas de retardo

En la sección anterior se indicó que la velocidad de propagación de una onda electromagnética depende del medio por donde viaja. Esta velocidad en el espacio libre, es decir, en el vacío, es la velocidad de la luz, 3×10^8 m/s, y es un poco menor a través de la atmósfera terrestre (es decir, el aire). La velocidad de propagación a través de una línea metálica de transmisión depende

de las constantes eléctricas del cable, de la inductancia y la capacitancia. Es algo menor que la velocidad de propagación a través del espacio libre y también de la atmósfera terrestre.

Las líneas de retardo son líneas de transmisión diseñadas en forma intencional para introducir un retardo de tiempo en la trayectoria de una onda electromagnética. La cantidad de retardo es función de la inductancia y la capacitancia de la línea de transmisión. La inductancia se opone a cambios de corriente, al igual que los tiempos de carga y descarga de la capacitancia. El retardo se calcula como sigue

$$t_d = LC \text{ (segundos)} \tag{8-23a}$$

donde t_d = retardo (segundos)
L = inductancia (henrys)
C = capacitancia (farads)

Si la inductancia y la capacitancia se expresan por unidad de longitud de línea de transmisión, por ejemplo por metro o por pie, el retardo también será por unidad de longitud; por ejemplo, 1.5 ns/metro.

La demora introducida por un tramo de cable coaxial se calcula con la siguiente fórmula

$$t_d = 1.016\,\epsilon \tag{8-23b}$$

en la que ϵ es la constante dieléctrica del cable.

PÉRDIDAS EN LÍNEAS DE TRANSMISIÓN

Para fines de análisis, las líneas de transmisión se consideran, con frecuencia, sin pérdidas. Sin embargo, en realidad hay varias formas en las que se pierde la energía en una línea de transmisión. Están las pérdidas en el conductor, pérdidas por radiación, pérdidas por calentamiento del dieléctrico, pérdidas por acoplamiento y efecto de corona.

Pérdidas en el conductor

Como la corriente pasa por una línea de transmisión, y ésta tiene una resistencia finita, hay una pérdida inherente e inevitable de potencia. A veces a esto se le llama *pérdida en el conductor* o *pérdida por calentamiento del conductor*, y es tan sólo una pérdida de la forma I^2R. Como la resistencia está distribuida en una línea de transmisión, la pérdida en el conductor es directamente proporcional a la longitud de la línea. También, ya que la disipación de potencia es directamente proporcional al cuadrado de la corriente, la pérdida en el conductor es inversamente proporcional a la impedancia característica. Para reducir las pérdidas en el conductor no hay más que acortar la línea de transmisión o usar un alambre de mayor diámetro (téngase en cuenta que al cambiar el diámetro del alambre también cambia la impedancia característica y, en consecuencia, la corriente).

La pérdida en el conductor depende algo de la frecuencia, debido a una acción llamada *efecto de superficie*. Cuando pasa la corriente por un alambre redondo aislado, el flujo magnético asociado con ella tiene la forma de círculos concéntricos. Esto se ve en la fig. 8-13. Se puede demostrar que la densidad de flujo cerca del centro del conductor es mayor que cerca de la superficie. Entonces, las líneas de flujo cercanas al centro del conductor encierran la corriente y reducen la movilidad de los electrones encerrados. Es una forma de autoinductancia, y hace que la inductancia cercana al centro del conductor sea mayor que en la superficie. Así, en las radiofrecuencias, la mayor parte de la corriente pasa por la superficie y no cerca del centro del conductor. Esto equivale a reducir el área transversal del conductor, y a aumentar la oposición al flujo de corriente (es decir, a aumentar la resistencia). La oposición adicional tiene ángulo de fase igual a 0°, y en consecuencia es una resistencia, y no una reactancia. Así, la resistencia en ca del conductor es proporcional a la raíz cuadrada de la frecuencia. La relación de la resistencia en ca entre la resistencia en cd de un conductor se llama relación de resistencias. Arriba de más o menos 100 MHz, se puede eliminar por completo el centro de un conductor, sin tener absolutamente efecto alguno sobre la pérdida en el conductor o la propagación de la onda electromagnética.

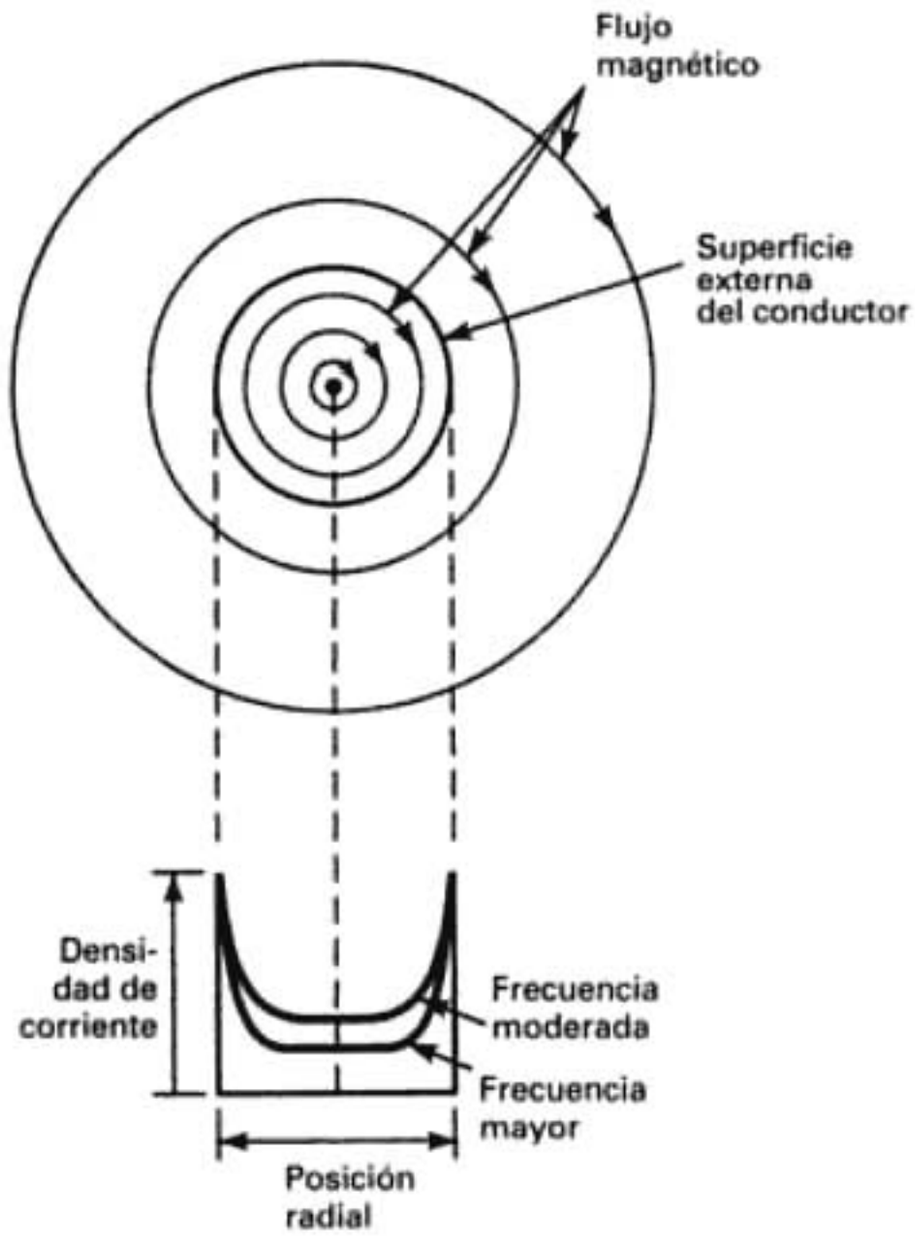


FIGURA 8-13 Conductor redondo aislado mostrando las líneas de flujo magnético, distribución de corrientes y el efecto de superficie

Las pérdidas en el conductor, en las líneas de transmisión, varía desde una fracción de decibel por 100 m en los cables coaxiales rígidos con dieléctrico de aire, hasta 200 dB por 100 m en una línea flexible de dieléctrico rígido.

Tanto las pérdidas I^2R como las del dieléctrico son proporcionales a la longitud, con frecuencia se agrupan y se expresan en decibeles de pérdida por unidad de longitud, es decir, dB/m.

Pérdida por calentamiento del dieléctrico

Una diferencia de potencial entre los dos conductores de una línea de transmisión causa el *calentamiento del dieléctrico*. El calor es una forma de energía que se debe tener en cuenta cuando se propaga energía por la línea. Para las líneas con dieléctrico de aire, la pérdida por calentamiento es despreciable. Sin embargo, con las líneas rígidas el calentamiento del dieléctrico aumenta con la frecuencia.

Pérdida por radiación

Si la separación entre los conductores de una línea de transmisión es una fracción apreciable de una longitud de onda, los campos electrostático y electromagnético que rodean al conductor hacen que la línea funcione como si fuera una antena, y transfiera energía a cualquier material conductor cercano. La cantidad de energía irradiada depende del material dieléctrico, la distancia entre conductores y la longitud de la línea. Las *pérdidas por radiación* se reducen blindando el cable en forma adecuada. Así, los cables coaxiales tienen menores pérdidas por radiación que las líneas de dos alambres paralelos. La pérdida por radiación también es proporcional a la frecuencia.

Pérdida por acoplamiento

La *pérdida por acoplamiento* se presenta siempre que se hace una conexión con o de una línea de transmisión, o cuando se conectan dos tramos separados de línea de transmisión. Las conexiones mecánicas son discontinuidades, es decir, lugares donde se unen materiales distintos. Las discontinuidades se tienden a calentar, irradian energía y disipan potencia.

Efecto corona (o efecto de arco voltaico)

El *arco voltaico* es una descarga luminosa que se produce entre dos conductores de una línea de transmisión, cuando la diferencia de potencial entre ellos es mayor que el voltaje de *ruptura* del

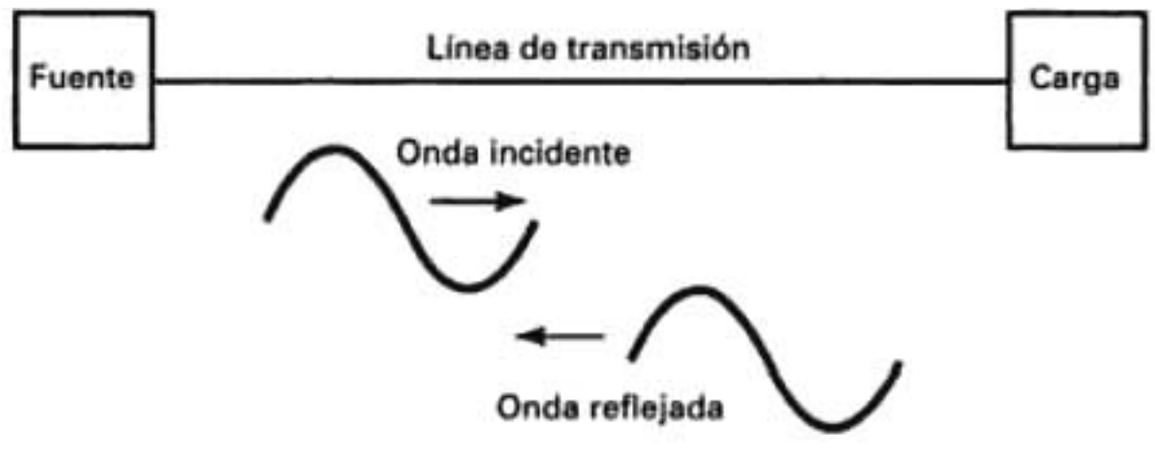


FIGURA 8-14 Fuente, carga, línea de transmisión y sus correspondientes ondas incidente y reflejada

dieléctrico aislante. En general, una vez que se produce el efecto de arco voltaico o efecto corona, la línea de transmisión se destruye.

ONDAS INCIDENTES Y REFLEJADAS

Una línea de transmisión ordinaria es bidireccional: la energía se puede propagar por igual en ambas direcciones. El voltaje que se propaga desde la fuente hacia la carga se llama *voltaje incidente*, y el que se propaga de la carga hacia la fuente se llama *voltaje reflejado*. De igual manera hay corrientes incidente y reflejada. En consecuencia, la potencia incidente se propaga hacia la carga, y la potencia reflejada se propaga hacia la fuente. El voltaje y la corriente incidentes siempre están en fase cuando la impedancia es de carácter resistivo. En una línea infinitamente larga, toda la potencia incidente se almacena en ella, y no hay potencia reflejada. También, si la línea se termina en una carga puramente resistiva igual a la impedancia característica de la línea, la carga absorbe toda la potencia incidente (suponiendo que la línea no tenga pérdidas). Para contar con una definición más práctica, la potencia reflejada es la parte de la potencia incidente que no absorbió la carga. En consecuencia, la potencia reflejada nunca puede ser mayor que la potencia incidente.

Líneas de transmisión resonantes y no resonantes

Una línea de transmisión sin potencia reflejada se llama línea *plana* o *no resonante*. Una línea de transmisión es no resonante si su longitud es infinita o si termina en una carga resistiva igual al valor óhmico de su impedancia característica. En una línea plana, el voltaje y la corriente son constantes en toda su longitud, si se supone que no tiene pérdidas. Cuando la carga no es igual a la impedancia característica de la línea, algo de la potencia se refleja y va hacia la fuente. Si la carga es un circuito cerrado o abierto, toda la potencia incidente se refleja hacia la fuente. Si se sustituyera la fuente por una abertura y un corto y la línea fuera sin pérdidas, la energía presente en ella se reflejaría de un lado a otro (oscilaría) entre los extremos de la fuente y la carga, de modo parecido a cuando la energía va y viene entre el capacitor y el inductor de un circuito tanque *LC*. A esto se le llama línea de transmisión *resonante*. En una línea resonante, la energía es transferida en forma alternativa entre los campos magnéticos y eléctricos de la inductancia y la capacitancia distribuida de la línea. La fig. 8-14 muestra una fuente, una línea de transmisión y una carga, con sus respectivas ondas incidente y reflejada.

Coeficiente de reflexión

El *coeficiente de reflexión* es una cantidad vectorial que representa la relación del voltaje reflejado entre el voltaje incidente, o la corriente reflejada entre la corriente incidente. La definición matemática del coeficiente de reflexión, representado por la gamma mayúscula Γ, es

$$\Gamma = \frac{E_r}{E_i} \text{ o } \frac{I_r}{I_i} \tag{8-24}$$

siendo Γ = coeficiente de reflexión (adimensional)
E_i = voltaje incidente (volts)
E_r = voltaje reflejado (volts)
I_i = corriente incidente (amps)
I_r = corriente reflejada (amps)

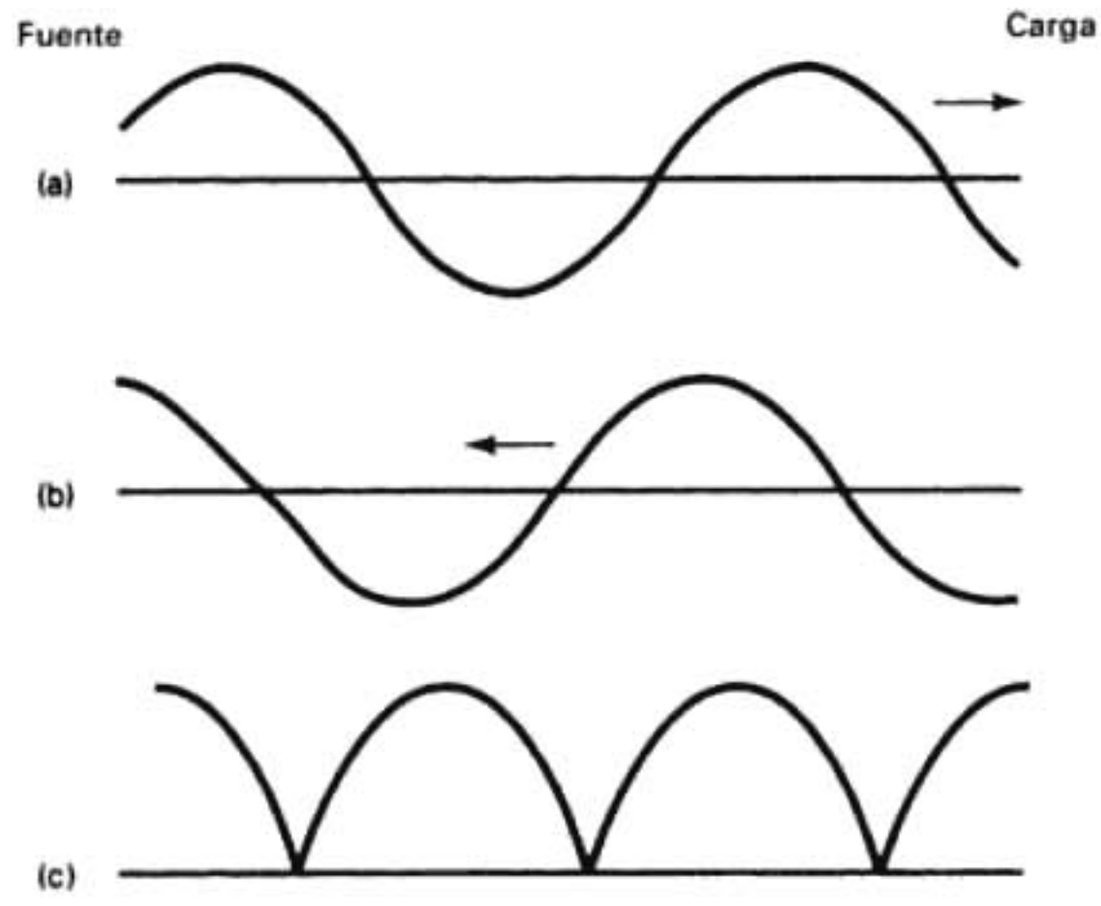


FIGURA 8-15 Formación de una onda estacionaria en una línea de transmisión: (a) onda incidente; (b) onda reflejada; (c) onda estacionaria

En la ecuación 8-24 se ve que el valor máximo, y el peor de los casos para $\Gamma = 1$, cuando $E_r = E_i$, y el valor mínimo, y de condición ideal, es cuando $\Gamma = 0$ ($E_r = 0$).

ONDAS ESTACIONARIAS

Cuando $Z_o = Z_L$, toda la potencia incidente es absorbida por la carga. A esto se le llama *línea compensada*. Cuando $Z_o \neq Z_L$, algo de la potencia incidente queda absorbida en la carga y algo se regresa (se refleja) a la fuente. Es lo que se llama *línea no compensada* o *línea descompensada*. En una línea no compensada hay dos ondas electromagnéticas que viajan en direcciones opuestas, al mismo tiempo (de hecho, a esas ondas se les llama *ondas viajeras*). Las dos ondas viajeras establecen un patrón de interferencia llamada *onda estacionaria*. Esto se ve en la fig. 8-15. Cuando las ondas incidente y reflejada pasan una por la otra, se producen patrones estacionarios de voltaje y corriente en la línea. Esas ondas se llaman estacionarias porque parecen permanecer en un lugar fijo de la línea, y sólo varían de amplitud. La onda estacionaria tiene mínimos (nodos) separados por una mitad de longitud de onda de las ondas viajeras, y tiene máximos (antinodos) también separados por una mitad de longitud de onda.

Relación de onda estacionaria

La *relación de onda estacionaria* (SWR, de *standing-wave ratio*) se define como la relación del voltaje máximo al voltaje mínimo, o de la corriente máxima entre la corriente mínima de una onda estacionaria en una línea de transmisión. A ello también se le llama *relación de voltajes de onda estacionaria* (VSWR, por *voltage standing-wave ratio*). En esencia, la SWR es una medida de la falta de compensación entre la impedancia de carga y la impedancia característica de la línea de transmisión. La ecuación correspondiente es

$$\text{SWR} = \frac{V_{\text{máx}}}{V_{\text{mín}}} \text{ (adimensional)} \tag{8-25}$$

Los máximos de voltaje ($V_{\text{máx}}$) se presentan cuando las ondas incidente y reflejada están en fase (es decir, sus máximos pasan por el mismo punto de la línea, con la misma polaridad), y los mínimos de voltaje ($V_{\text{mín}}$) se presentan cuando las ondas incidente y reflejada están desfasadas 180°. Las descripciones matemáticas de $V_{\text{máx}}$ y $V_{\text{mín}}$ son

$$V_{\text{máx}} = E_i + E_r \tag{8-26a}$$

$$V_{\text{mín}} = E_i - E_r \tag{8-26b}$$

Por lo anterior, la ecuación 8-25 se puede escribir como sigue

$$\text{SWR} = \frac{V_{\text{máx}}}{V_{\text{mín}}} = \frac{E_i + E_r}{E_i - E_r} \tag{8-27}$$

En esta ecuación se puede ver que cuando las ondas incidente y reflejada tienen igual amplitud (descompensación total), SWR = infinito. Es el peor de los casos. También, en la ecuación 8-27 se puede ver que cuando no hay onda reflejada ($E_r = 0$), SWR = E_i/E_i, es decir 1. Este caso se presenta cuando $Z_o = Z_L$ y es el caso ideal.

También se puede expresar la relación de onda estacionaria en función de Γ. Si se reordena la ecuación 8-25 se llega a

$$\Gamma E_i = E_r$$

Se sustituye en la ecuación 8-27, y se obtiene

$$\text{SWR} = \frac{E_i + E_i\Gamma}{E_i - E_i\Gamma}$$

Se saca E_i como factor común

$$\text{SWR} = \frac{E_i(1 + \Gamma)}{E_i(1 - \Gamma)} = \frac{1 + \Gamma}{1 - \Gamma} \tag{8-28}$$

Se desarrolla por producto cruzado como sigue

$$\text{SWR}(1 - \Gamma) = 1 + \Gamma$$

$$\text{SWR} - \text{SWR}\Gamma = 1 + \Gamma$$

$$\text{SWR} = 1 + \Gamma + (\text{SWR})\Gamma$$

$$\text{SWR} - 1 = \Gamma(1 + \text{SWR}) \tag{8-29}$$

$$\Gamma = \frac{\text{SWR} - 1}{\text{SWR} + 1} \tag{8-30}$$

Ejemplo 8-4

Para una línea de transmisión con voltaje incidente $E_i = 5$ V y voltaje reflejado $E_r = 3$ V, determinar:

(a) El coeficiente de reflexión.

(b) La relación de onda estacionaria, SWR.

Solución (a) Se sustituyen valores en la ecuación 8-24:

$$\Gamma = \frac{E_r}{E_i} = \frac{3}{5} = 0.6$$

(b) Se sustituye en la ecuación 8-27

$$\text{SWR} = \frac{E_i + E_r}{E_i - E_r} = \frac{5 + 3}{5 - 3} = \frac{8}{2} = 4$$

Se sustituye en la ecuación 8-30, para obtener

$$\Gamma = \frac{4 - 1}{4 + 1} = \frac{3}{5} = 0.6$$

Cuando la carga es puramente resistiva, también se puede expresar la SWR como una relación de la impedancia característica entre la impedancia de carga, o viceversa. La ecuación es

$$\text{SWR} = \frac{Z_o}{Z_L} \text{ o } \frac{Z_L}{Z_o} \quad \text{(la que produzca una SWR mayor que 1)} \tag{8-31}$$

En la ecuación 8-31 se escogen el numerador y el denominador de tal modo que la SWR siempre sea un número mayor que 1, para evitar confusiones y apegarse a la convención establecida en la ecuación 8-27. En la ecuación 8-31 se ve que una resistencia de carga $Z_L = 2Z_o$ produce la misma SWR que una resistencia de carga $Z_L = Z_o/2$; el grado de desajuste es el mismo.

Las desventajas de no tener una línea de transmisión equilibrada (plana) se pueden resumir como sigue:

1. A la carga no le llega el cien por ciento de la potencia incidente desde la fuente.
2. El dieléctrico que separa los dos conductores puede fallar y causar un efecto corona como resultado de la alta relación de voltajes de onda estacionaria.
3. Las reflexiones y las reflexiones repetidas causan más pérdida de potencia.
4. Las reflexiones causan imágenes fantasma.
5. Los desajustes causan interferencia de ruido.

Aunque es muy improbable que una línea de transmisión termine en una carga que sea un circuito abierto o cerrado, se examinan esas condiciones, porque ilustran los peores casos posibles que se producirían, y que producen ondas estacionarias que son características de condiciones menos severas.

Ondas estacionarias en una línea abierta

Cuando las ondas incidentes de voltaje y corriente llegan a una terminación abierta, nada de energía se absorbe; se refleja por todo el trayecto hacia la fuente. La onda de voltaje incidente se refleja en la forma exacta como si fuera a continuar por una línea infinitamente larga. Sin embargo, la corriente incidente se refleja invertida 180° respecto a como hubiera continuado si la línea no estuviera abierta. Al pasar las ondas incidente y reflejada, se producen en la línea ondas estacionarias. La fig. 8-16 muestra las ondas estacionarias de voltaje y corriente en una línea de transmisión terminada en circuito abierto. Se ve que la onda estacionaria de voltaje tiene un valor máximo en el extremo abierto, y uno mínimo a un cuarto de longitud de onda de la abertura. La onda estacionaria de corriente tiene un valor mínimo en el extremo abierto, y un valor máximo a un cuarto de longitud de onda de la abertura. Es razonable que el máximo voltaje esté en la abertura, y que allí haya corriente mínima.

Se pueden resumir como sigue las características de una línea de transmisión terminada en una abertura:

1. La onda incidente de voltaje se refleja y regresa exactamente como si hubiera de continuar, es decir, sin inversión de fase.
2. La onda incidente de corriente se refleja y regresa a 180° respecto a como hubiera continuado.
3. La suma de las formas de onda de corriente incidente y reflejada es mínima en la abertura.
4. La suma de las formas de onda de voltaje incidente y reflejado es máxima en la abertura.

También se ve en la fig. 8-16 que las ondas estacionarias de voltaje y corriente se repiten cada mitad de longitud de onda. La impedancia en el extremo abierto es $Z = V_{máx}/I_{mín}$ y es máxima allí. La impedancia a un cuarto de longitud de onda de la abertura es $Z = V_{mín}/I_{máx}$ y es mínima allí. En consecuencia, a un cuarto de longitud de onda se presenta una inversión de impedancia, y hay otras inversiones de impedancia a cada cuarto de longitud de onda.

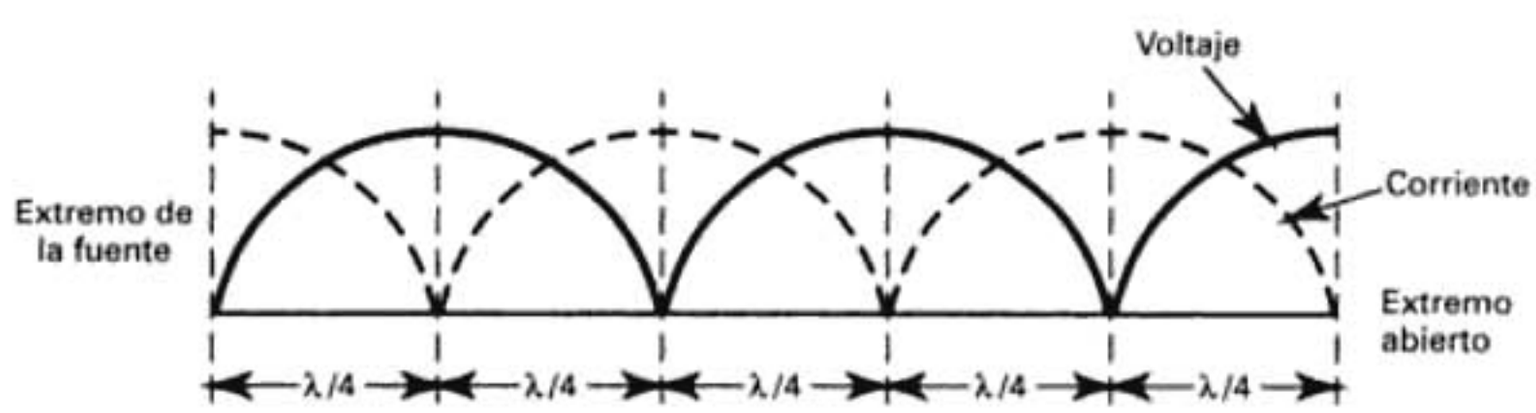


FIGURA 8-16 Ondas estacionarias de voltaje y de corriente en una línea de transmisión que termina en un circuito abierto

La fig. 8-16 muestra el desarrollo de una onda estacionaria de voltaje en una línea de transmisión que se termina en un circuito abierto. La fig. 8-17 muestra una onda incidente propagándose por una línea de transmisión hacia la carga. Esa onda viaja aproximadamente a la velocidad de la luz; sin embargo, para ilustrarla, se ha congelado a intervalos de un octavo de longitud de onda. Se puede ver en la fig. 8-17a que la onda incidente no ha llegado al circuito abierto. La fig. 8-17 muestra la onda una unidad de tiempo después (para este ejemplo, la onda recorre un octavo de longitud de onda por unidad de tiempo). Se ve que la onda se ha movido y está un

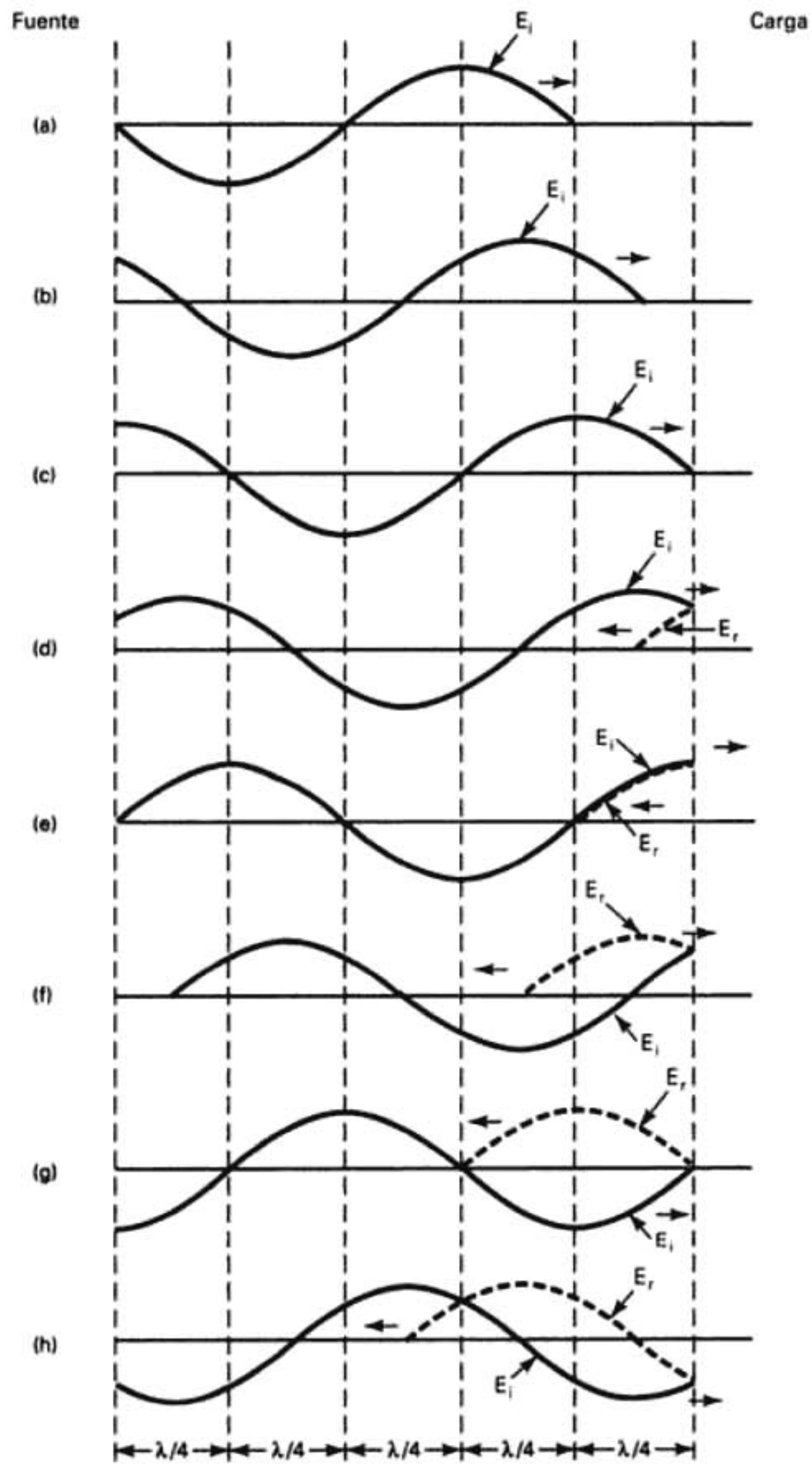


FIGURA 8-17 Ondas incidentes y reflejadas en una línea de transmisión terminada en circuito abierto (*continúa*)

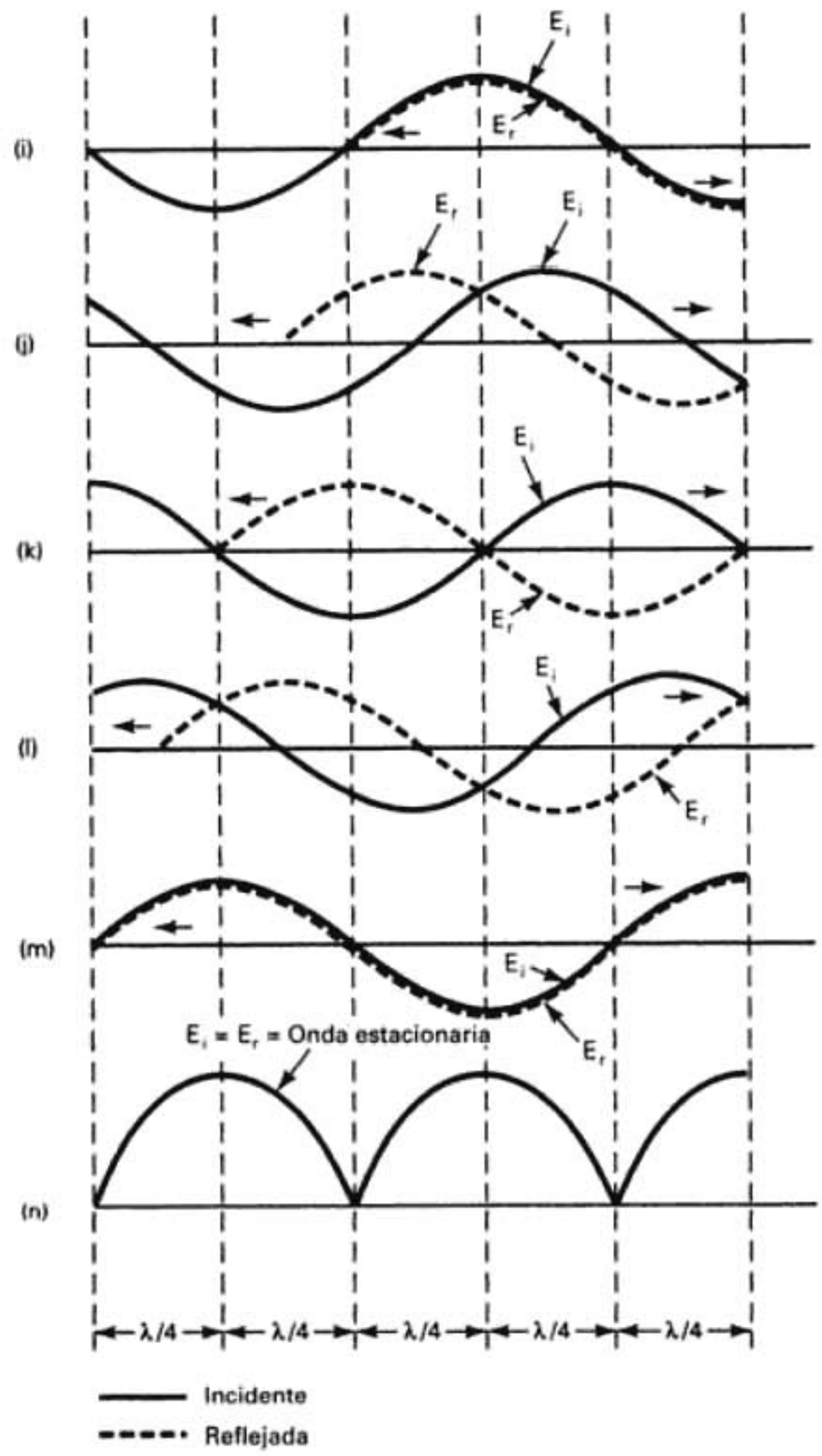


FIGURA 8-17 (Continuación) Ondas incidentes y reflejadas en una línea de transmisión terminada en circuito abierto

cuarto de longitud de onda más cercana a la abertura. La fig. 8-17c muestra la onda en el momento de arribar a la abertura. Hasta ahora, no ha habido onda reflejada y, en consecuencia, no ha habido onda estacionaria. La fig. 8-17d muestra las ondas incidente y reflejada una unidad de tiempo después de que la onda incidente llegó a la abertura; la onda reflejada se propaga alejándose de la abertura. Las figs. 8-17e, f y g muestran las ondas incidente y reflejada durante las siguientes unidades de tiempo. Se ve que en la fig. 8-17e las ondas incidente y reflejada tienen sus valores máximos positivos al mismo tiempo, y así producen un máximo de voltaje en la abertura. También se ve que a un cuarto de longitud de onda de la abertura, la suma de las ondas incidente y reflejada (es decir, la onda estacionaria) siempre es igual a 0 V (un mínimo). Las figuras 8-17h a m, muestran la propagación de las ondas incidente y reflejada hasta que la onda reflejada llega a la fuente; la figura 8-17n muestra la onda estacionaria que resulta. Se ve que la onda estacionaria queda en el mismo lugar: los nodos y antinodos de voltaje permanecen en los mismos puntos. Sin embargo, la amplitud de los antinodos varía desde un máximo posible

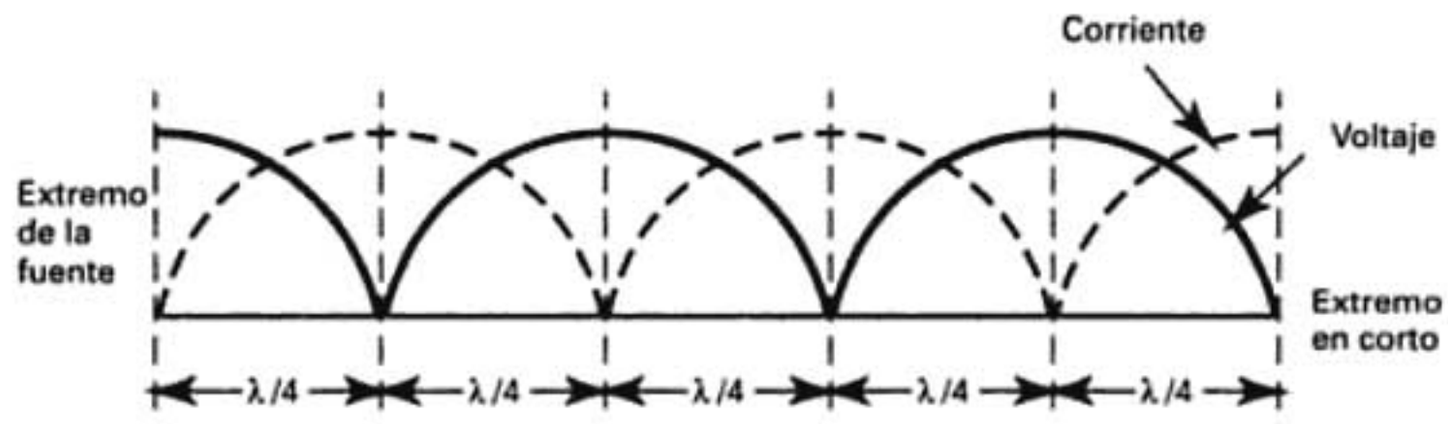


FIGURA 8-18 Ondas estacionarias de voltaje y de corriente en una línea de transmisión que termina en cortocircuito

hasta cero, y después se repite. Para una carga abierta, todo el voltaje incidente se refleja; $E_r = E_i$, y, en consecuencia, $V_{máx} = E_i + E_r$ o $2E_i$. Se puede hacer una demostración parecida para una onda estacionaria de corriente (sin embargo, recuérdese que la corriente se refleja y regresa con una inversión de fase de 180°).

Ondas estacionarias en una línea en corto

Como en el caso de la línea abierta, la carga no absorbe potencia alguna cuando la línea de transmisión se termina en un cortocircuito. Sin embargo, si la línea está en corto, el voltaje y la corriente incidentes se reflejan y regresan en la forma contraria. La onda de voltaje se refleja con su fase invertida 180° respecto a la que tendría si continuara por una línea infinitamente larga, y la onda de corriente se refleja exactamente de la misma manera que si no hubiera corto.

La fig. 8-18 representa las ondas estacionarias de corriente y de voltaje en una línea de transmisión que termina en un cortocircuito. Se ve que la onda estacionaria de voltaje tiene un valor mínimo en el extremo puesto en corto, y un valor máximo a un cuarto de longitud de onda del corto. La onda estacionaria de corriente tiene un valor máximo en el corto y un valor mínimo un cuarto de longitud de onda más atrás. Las ondas estacionarias de voltaje y de corriente se repiten cada cuarto de longitud de onda. Así, hay una inversión de impedancia cada intervalo de cuarto de longitud de onda. La impedancia en el corto es $Z = V_{mín}/I_{máx}$ = mínima, y un cuarto de longitud de onda atrás es $Z = V_{máx}/I_{mín}$ = máxima. De nuevo, es razonable que un mínimo de voltaje se presente en un corto, y que allí haya corriente máxima.

Las características de una línea de transmisión terminada en corto se pueden resumir como sigue:

1. La onda estacionaria de voltaje se refleja y se invierte 180° respecto a la forma en que continuaría.
2. La onda estacionaria de corriente se refleja del mismo modo que si hubiera continuado.
3. La suma de las formas de onda incidente y reflejada de corriente es máxima en el corto.
4. La suma de las formas de onda incidente y reflejada de voltaje es cero en el corto.

Para una línea de transmisión terminada en un circuito cerrado o abierto, el coeficiente de reflexión es 1 (el peor de los casos) y la SWR (relación de onda estacionaria) es infinita (también el peor de los casos).

IMPEDANCIA DE ENTRADA DE UNA LÍNEA DE TRANSMISIÓN

En la sección anterior se demostró que cuando se termina una línea de transmisión en un cortocircuito o en un circuito abierto, hay una *inversión de impedancia* cada cuarto de onda. Para una línea sin pérdidas, la impedancia varía de infinito a cero. Sin embargo, en un caso más práctico donde sí hay pérdidas de energía, la amplitud de la onda reflejada siempre es menor que la de la onda incidente, excepto en la terminación. Por consiguiente, la impedancia varía desde cierto valor máximo hasta cierto valor mínimo, o viceversa, dependiendo de si la línea termina en corto o en circuito abierto. La impedancia de entrada de una línea sin pérdida, vista hacia una línea de

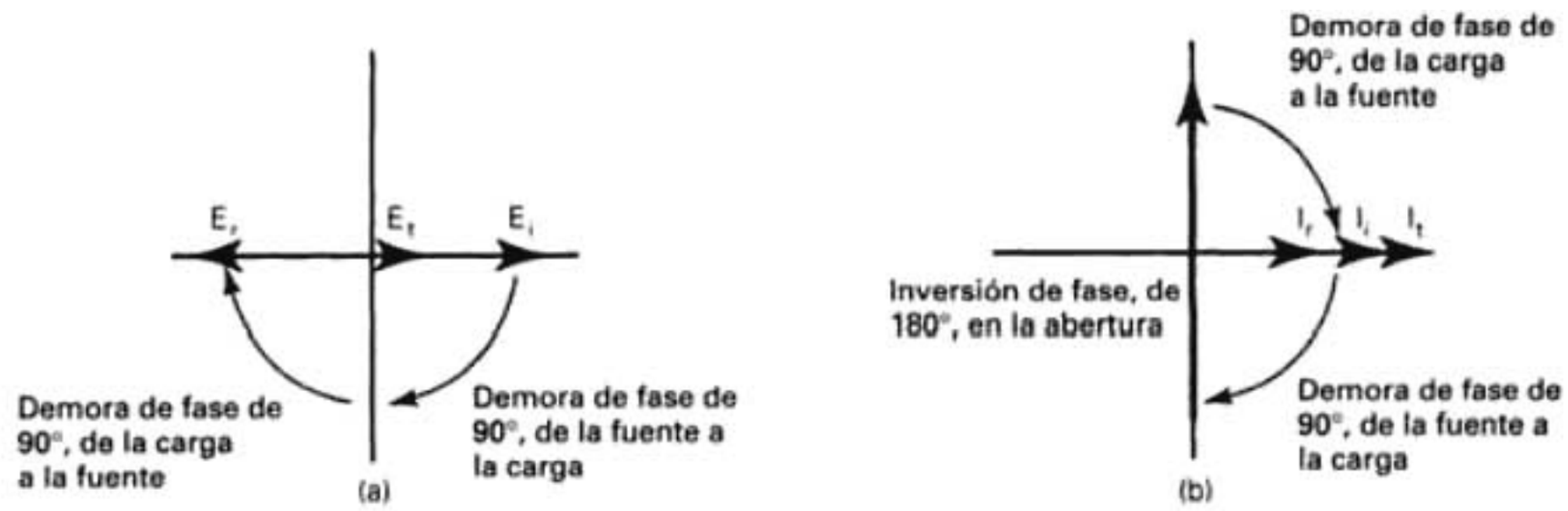


FIGURA 8-19 Relaciones de fase para corriente y voltaje en una línea de un cuarto de onda terminada en circuito abierto: (a) relaciones de voltaje a fase; (b) relaciones de corriente a fase

transmisión que termina en un corto o en circuito abierto, puede ser resistiva, inductiva o capacitiva, dependiendo de la distancia a la terminación.

Análisis fasorial de la impedancia de entrada: línea abierta

Se usan en general los diagramas fasoriales para analizar la impedancia de entrada de una línea de transmisión, por ser relativamente sencillos y dar una representación gráfica de las relaciones de fase de corriente y voltaje. Estas relaciones se refieren a variaciones en el tiempo. Las figs. 8-16, 8-17 y 8-18 muestran ondas estacionarias de voltaje y corriente, graficadas en función de la distancia y, en consecuencia, no indican las relaciones reales entre fases. Las siguientes secciones usan diagramas fasoriales para analizar la impedancia de entrada de varias configuraciones de línea de transmisión.

Línea de transmisión de un cuarto de longitud de onda. La fig. 8-19a muestra el diagrama fasorial del voltaje, y la fig. 8-19b el diagrama fasorial de la corriente en la entrada de un tramo de un cuarto de onda, de una línea de transmisión terminada en un circuito abierto. Las formas de onda de la corriente y del voltaje incidentes son I_i y V_i, respectivamente, en el extremo de la entrada (de la fuente) de la línea en determinado instante. Todo voltaje reflejado, E_r, en la entrada de la línea ha viajado la mitad de una longitud de onda, de la fuente a la abertura y de regreso y en consecuencia está 180° retrasado al voltaje incidente. Por consiguiente, el voltaje total E_t en el extremo de entrada es la suma de E_i y E_r. Es decir, $E_t = E_i + E_r \angle -180°$, y, suponiendo una pérdida pequeña en la línea, $E_t = E_i - E_r$. La corriente reflejada se demora 90° al propagarse de la fuente a la carga, y otros 90° en el regreso de la carga a la fuente. También, la corriente reflejada sufre una inversión de 180° en el extremo abierto. La corriente reflejada se ha demorado en realidad 360°. Por consiguiente, cuando la corriente reflejada llega al extremo de la fuente, está en fase con la corriente incidente y la corriente total es $I_t = I_i + I_r$. Se puede ver, al examinar la fig. 8-19, que E_t e I_t están en fase. Por consiguiente, la impedancia de entrada viendo hacia una línea de transmisión de un cuarto de longitud de onda de largo terminada en un circuito abierto es $Z_{ent} = E_t \angle 0° / I_t \angle 0° = Z_{ent} \angle 0°$. Z_{ent} tiene ángulo de fase 0° y en consecuencia es resistiva y es mínima. Así, una línea de transmisión terminada en un circuito abierto equivale a un circuito LC resonante en serie.

La fig. 8-20 muestra varios fasores de voltaje para las ondas incidente y reflejada en una línea de transmisión terminada en circuito abierto, y la forma en que produceh una onda de voltaje estacionaria.

Línea de transmisión de longitud menor que un cuarto de longitud de onda. La fig. 8-21a muestra el diagrama fasorial de voltaje, y la fig. 8-21b muestra el diagrama fasorial de corriente de una línea de transmisión más corta que un cuarto de longitud de onda ($\lambda/4$), terminada en circuito abierto. De nuevo, la corriente incidente I_i y el voltaje incidente E_i están en fase. La onda reflejada de voltaje se demora 45° al ir de la fuente a la carga (distancia que equivale a un octavo de longitud de onda) y otros 45° al regresar de la carga a la fuente (otro octavo de longitud de onda más). Por consiguiente, cuando la onda reflejada llega al extremo de la fuente, está retrasada 90° respecto a la onda incidente. El voltaje total en el extremo de la fuente es la

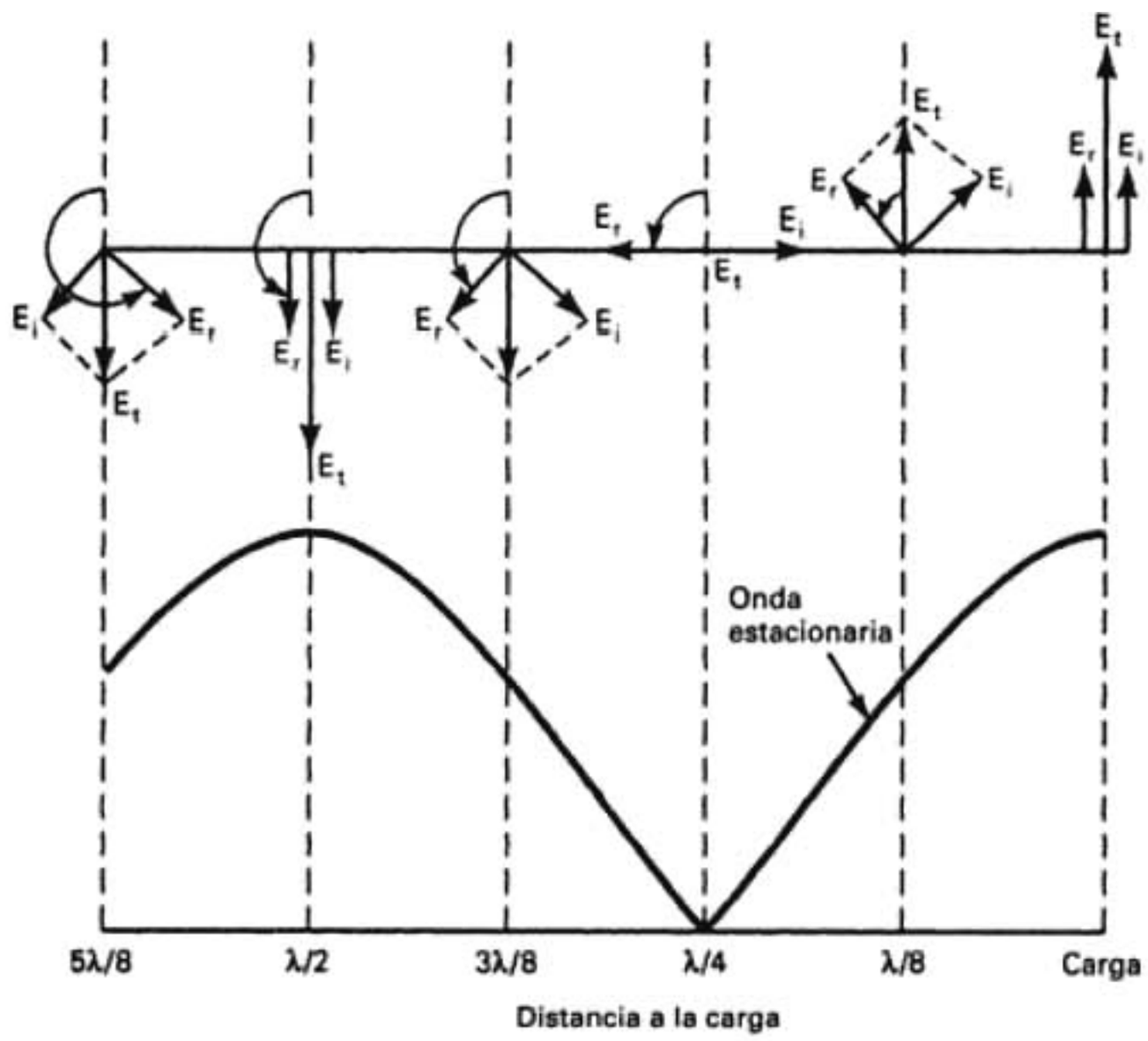


FIGURA 8-20 Suma vectorial de las ondas incidente y reflejada para producir una onda estacionaria

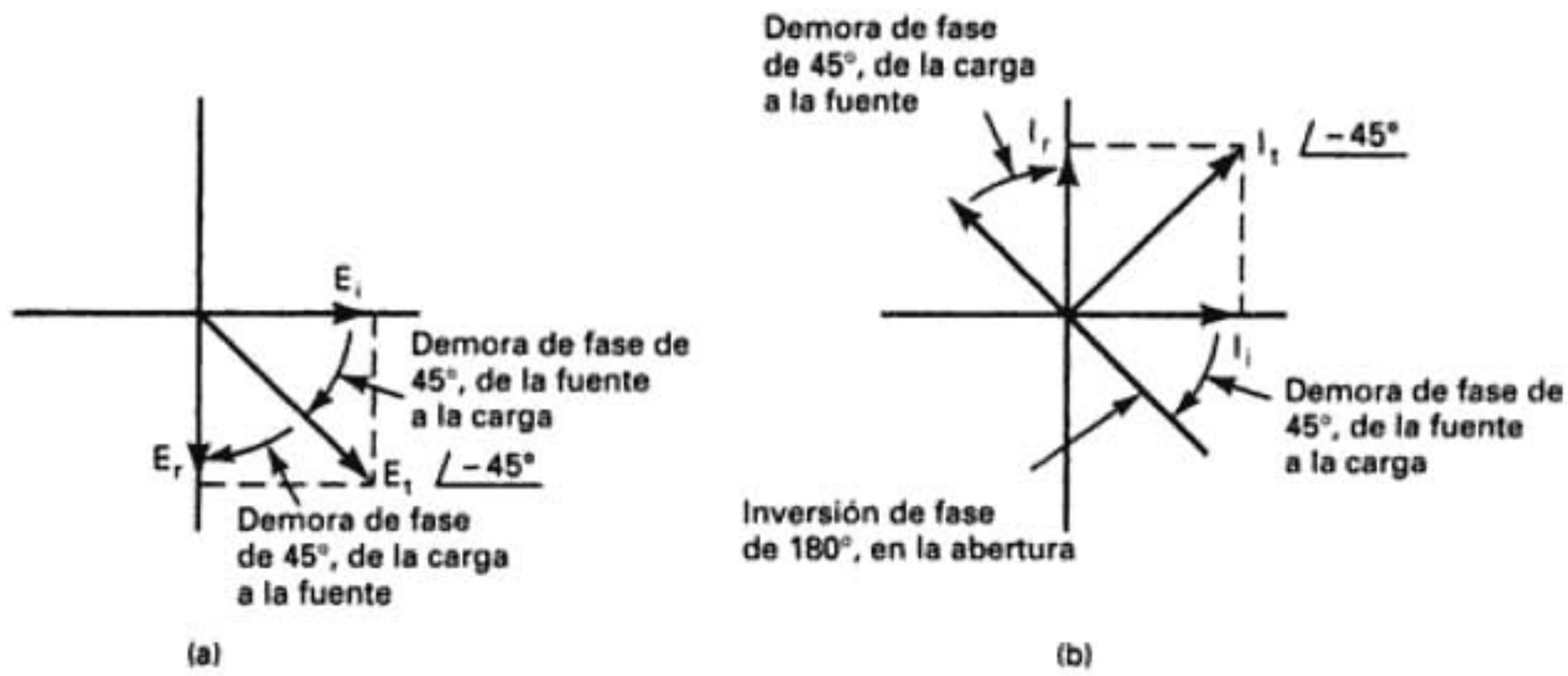


FIGURA 8-21 Relaciones de fase de voltaje y corriente para una línea de transmisión de menos de un cuarto de longitud de onda terminada en circuito abierto: (a) relaciones de voltaje a fase; (b) relaciones de corriente a fase

suma vectorial de las ondas incidente y reflejada. Así, $E_t = \sqrt{E_i^2 + E_r^2} = E_t \angle{-45°}$. La onda reflejada de corriente se demora 45° al ir de la fuente a la carga y otros 45° al regresar de la carga a la fuente (una distancia total de un cuarto de longitud de onda). Además, la onda reflejada de corriente ha sufrido una inversión de fase de 180° en el extremo abierto antes de ser reflejada. Así, la onda reflejada de corriente tiene una demora total de 270°. Por lo anterior, la onda reflejada de hecho antecede 90° a la onda incidente. La corriente total en el extremo de la fuente es la suma vectorial de las ondas incidente y reflejada. Es decir, $I_t = \sqrt{I_i^2 + I_r^2} = I_t \angle{+45°}$. Al examinar la fig. 8-21 se puede ver que E_t está retrasado 90° respecto a I_t. Entonces, $Z_{ent} = E_t \angle{-45°} / I_t \angle{+45°} = Z_{ent} \angle{-90°}$. La impedancia Z_{ent} tiene ángulo de fase de −90° y por consiguiente es capacitiva. Toda línea de transmisión más corta que un cuarto de longitud de onda que esté terminada en un circuito abierto equivale a un capacitor. La cantidad de capacitancia depende de la longitud eléctrica exacta de la línea.

Línea de transmisión mayor que un cuarto de longitud de onda. La fig. 8-22a muestra el diagrama fasorial de voltaje y la fig. 8-22b el diagrama fasorial de corriente para una línea de transmisión más larga que un cuarto de longitud de onda, terminada en circuito abierto. Para este ejemplo se usa una línea de transmisión de tres octavos de longitud de onda. El voltaje reflejado está demorado tres cuartos de longitud de onda, o 270°. Por consiguiente, de hecho el voltaje se adelanta 90° al voltaje incidente. Entonces, el voltaje total es $E_t = \sqrt{E_i^2 + E_r^2}\,\underline{/+45^\circ} = E_t\,\underline{/-45^\circ}$. La onda reflejada de corriente está demorada 270° y ha sufrido una inversión de fase de 180° en el extremo abierto. Por ello, de hecho la corriente reflejada está retrasada 90° respecto a la corriente incidente. Entonces, la corriente total es $I_t = \sqrt{I_i^2 + I_r^2} = I_t\,\underline{/-45^\circ}$. Así, $Z_{ent} = E_t\,\underline{/+45^\circ}\,/I_t\,\underline{/-45^\circ} = Z_{ent}\,\underline{/+90^\circ}$. La impedancia Z_{ent} tiene un ángulo de fase de +90° y en consecuencia es inductiva. La magnitud de la impedancia de entrada es igual a la impedancia característica en los puntos a un octavo de longitud de onda. Una línea de transmisión entre un cuarto y media longitud de onda que termina en un circuito abierto equivale a un inductor. La cantidad de inductancia depende de la longitud eléctrica exacta de la línea.

Línea de transmisión abierta como elemento de circuito. Es obvio, de acuerdo con la descripción anterior y las figuras 8-19 a 8-22, que una línea de transmisión abierta se puede comportar como un resistor, un inductor o un capacitor, dependiendo de su longitud eléctrica. Como las pautas de onda estacionaria en una línea abierta se repiten cada intervalo de media longitud de onda, la impedancia de entrada también se repite. La fig. 8-23 muestra las variaciones en la impedancia de entrada para una línea de transmisión abierta de varias longitudes eléctricas. Se ve que la impedancia de una línea abierta es resistiva y máxima en el extremo abierto y a cada intervalo sucesivo de media longitud de onda, y resistiva y mínima a un cuarto de longitud de onda del extremo abierto y a cada intervalo sucesivo de media longitud de onda. Para longitudes eléctricas menores que un cuarto de longitud de onda, la impedancia de entrada es capacitiva y disminuye con la longitud. Para longitudes eléctricas entre un cuarto y media longitud de onda, la impedancia de entrada es inductiva y aumenta con la longitud. Los patrones de capacitancia y de inductancia también se repiten cada media longitud de onda.

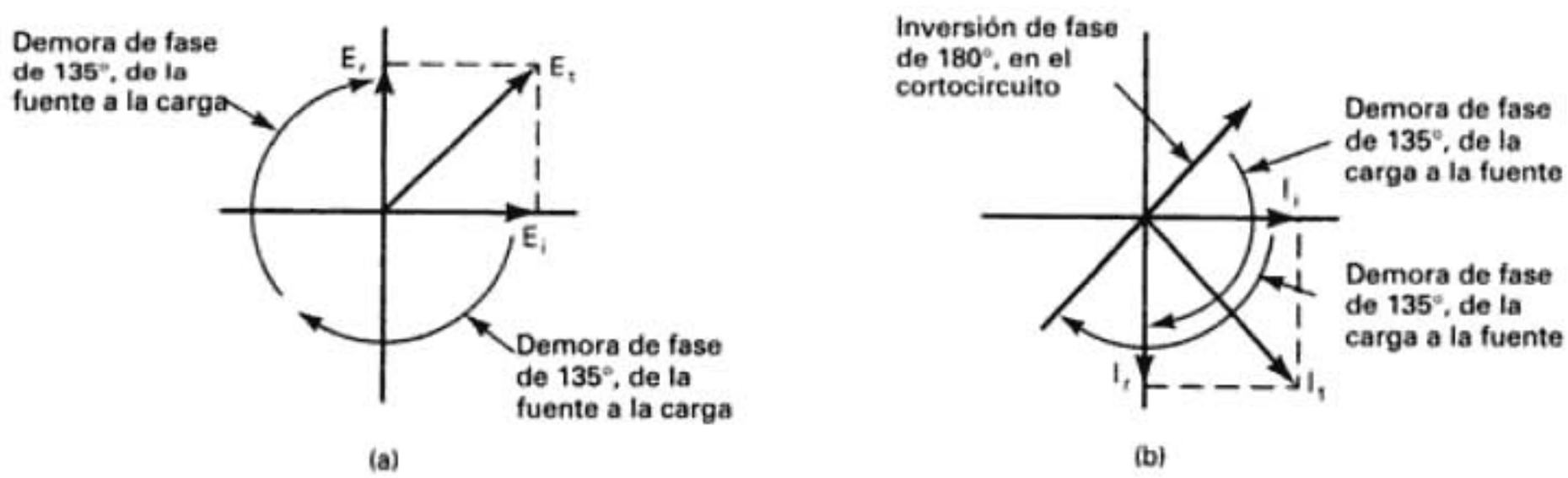


FIGURA 8-22 Relaciones de voltaje y de corriente a fase para una línea de transmisión de más de un cuarto de longitud de onda, terminada en circuito abierto: (a) relaciones de voltaje a fase; (b) relaciones de corriente a fase

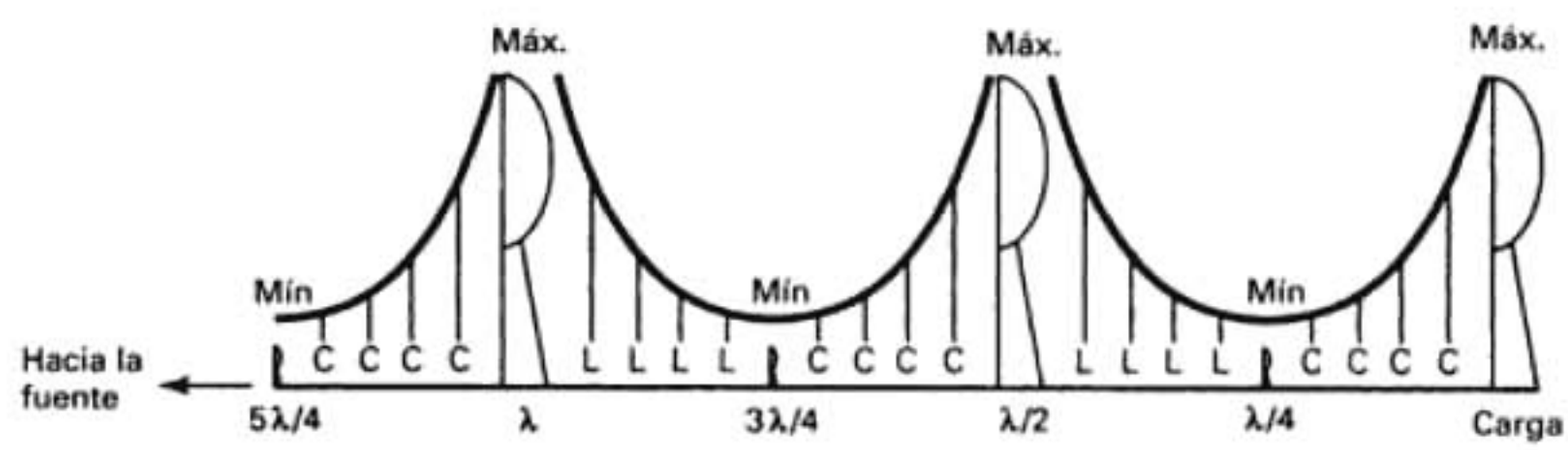


FIGURA 8-23 Variaciones de impedancia de entrada para una línea de transmisión en circuito abierto

Análisis fasorial de la impedancia de entrada: línea en cortocircuito

Las siguientes explicaciones usarán diagramas fasoriales para el análisis de líneas de transmisión puestas en corto de la misma forma que con las líneas abiertas. La diferencia es que en las líneas de transmisión en corto, la forma de onda de voltaje se regresa con una inversión de fase de 180°, mientras que la onda de corriente se refleja como si no hubiera corto.

Línea de transmisión de un cuarto de longitud de onda. Los diagramas fasoriales de voltaje y corriente, para una línea de transmisión de un cuarto de longitud de onda terminada en un cortocircuito son idénticos a los de la fig. 8-19, pero invertidas. Los voltajes incidente y reflejado están en fase y en consecuencia $E_t = E_i + E_r$ y es máximo. Las corrientes incidente y reflejada están desfasadas 180° y, en consecuencia, $I_t = I_i - I_r$ y es mínima. Entonces $Z_{ent} = E_i \angle 0° / I_t \angle 0° = Z_{ent} \angle 0°$ y es máxima. La impedancia Z_{ent} tiene ángulo de fase de 0°, es resistiva y es máxima. Por consiguiente, una línea de transmisión terminada en un cortocircuito equivale a un circuito *LC* en paralelo.

Línea de transmisión menor que un cuarto de longitud de onda. Los diagramas fasoriales de voltaje y corriente para una línea de transmisión menor que un cuarto de longitud de onda y terminadas en cortocircuito son idénticos a los de la fig. 8-21, pero invertidos. El voltaje se invierte 180° en el corto y la corriente se refleja con la misma fase que tendría si hubiera continuado. Por consiguiente, el voltaje total en el extremo de la fuente de la línea se adelanta 90° a la corriente y la línea se ve inductiva.

Línea de transmisión mayor que un cuarto de longitud de onda. Los diagramas fasoriales de voltaje y corriente para una línea de transmisión mayor que un cuarto de longitud de onda, terminada en cortocircuito, son idénticos a los de la fig. 8-22, pero invertidos. El voltaje total en el extremo de la línea que da a la fuente se retrasa 90° a la corriente, y la línea se ve capacitiva.

Línea de transmisión en cortocircuito como elemento de circuito. De acuerdo con la descripción anterior es obvio que una línea de transmisión puesta en corto se puede comportar como si fuera un resistor, un inductor o un capacitor, dependiendo de su longitud eléctrica. En una línea de transmisión en corto, las ondas estacionarias se repiten cada media longitud de onda, y en consecuencia también se repite la impedancia de entrada. La fig. 8-24 muestra las variaciones de impedancia de entrada de una línea de transmisión en corto, para diversas longitudes eléctricas. Se ve que una línea en corto tiene impedancia resistiva y mínima en el corto y en cada intervalo sucesivo de media longitud de onda, y resistiva y máxima a un cuarto de longitud de onda del corto, y en cada intervalo sucesivo de media longitud de onda. Para longitudes eléctricas menores que un cuarto de longitud de onda, la impedancia de entrada es inductiva y aumenta con la longitud. Para longitudes eléctricas de un cuarto a media longitud de onda, la impedancia de entrada es capacitiva y disminuye con la longitud. Los comportamientos de inductancia y capacitancia también se repiten cada intervalo de media longitud de onda.

Resumen de impedancia de entrada a una línea de transmisión. La fig. 8-25 es un resumen de las configuraciones de línea de transmisión que se describieron en las secciones anteriores, sus características de impedancia de entrada y sus circuitos *LC* equivalentes. Se ve que los tramos en corto y abiertos de las líneas de transmisión se pueden comportar como resistores, inductores o capacitores, dependiendo de su longitud eléctrica.

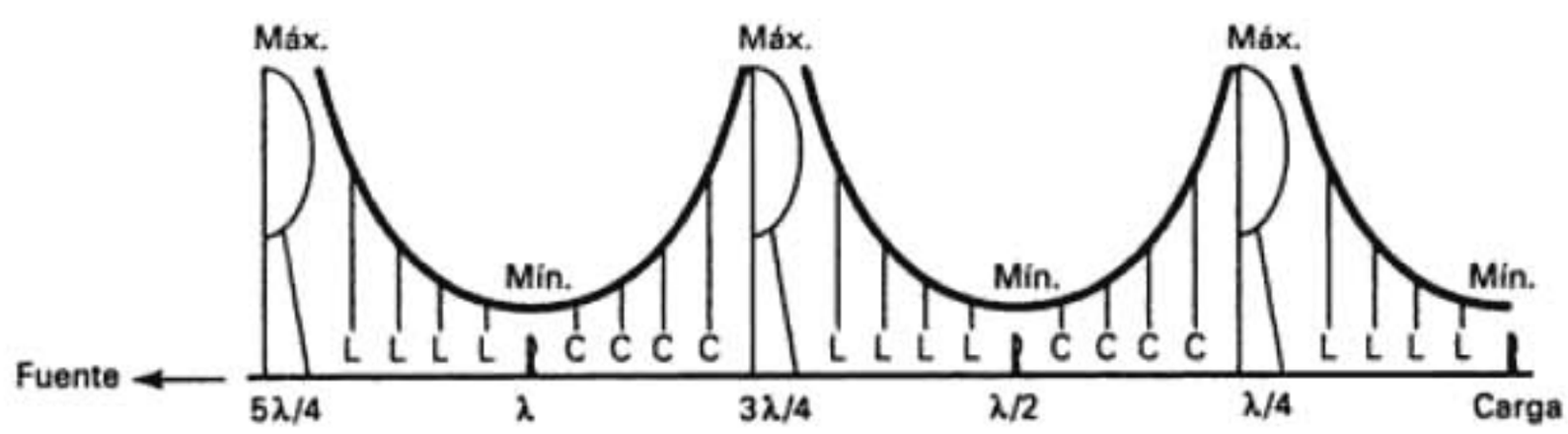


FIGURA 8-24 Variaciones de impedancia de entrada para una línea de transmisión en cortocircuito

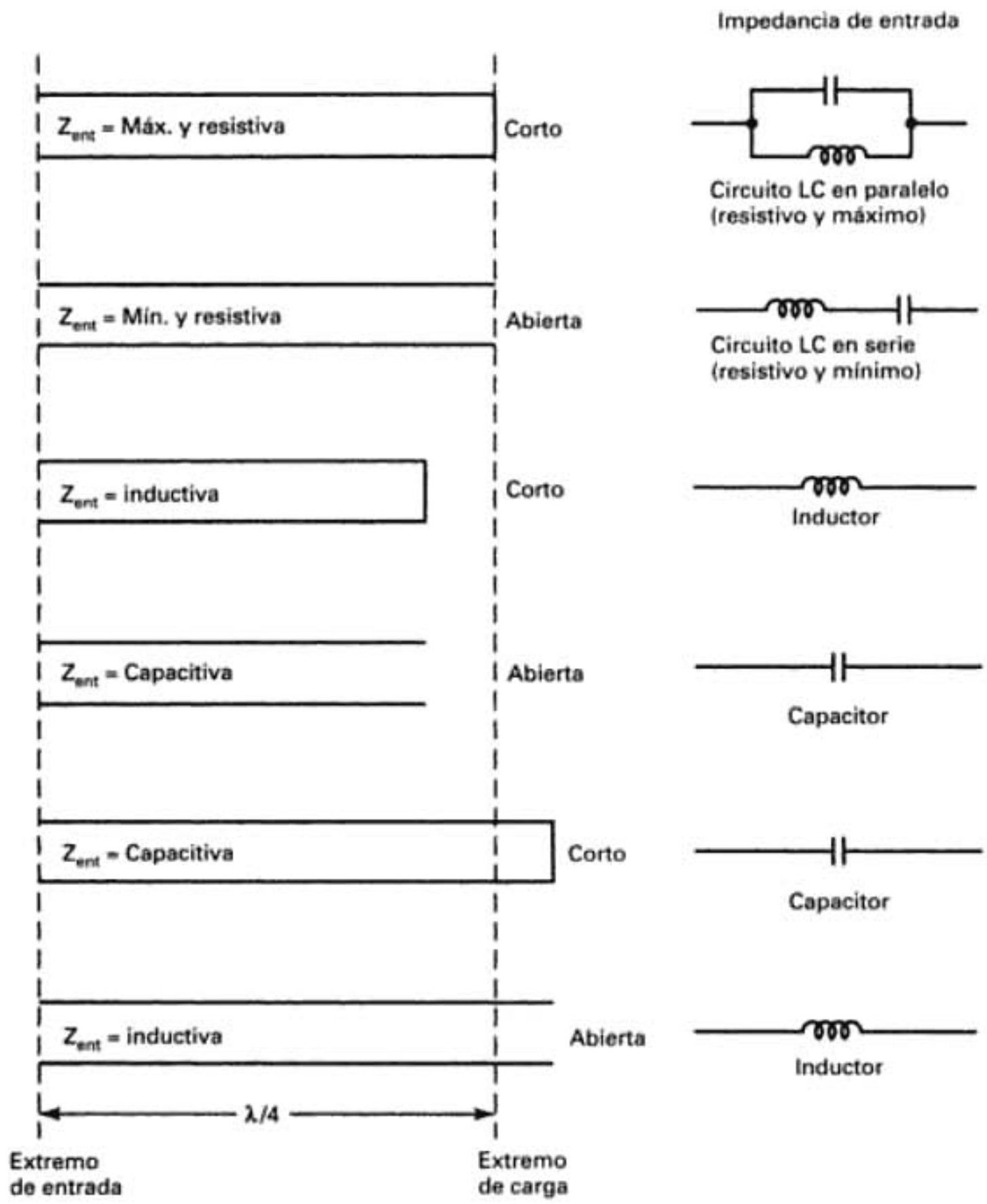


FIGURA 8-25 Resumen de impedancias en líneas de transmisión

Acoplamiento de impedancias en líneas de transmisión

La potencia se transfiere a una carga cuando no hay ondas reflejadas, esto es, cuando la carga es puramente resistiva e igual a Z_o. Siempre que la impedancia característica de una línea de transmisión con su carga no estén balanceadas (sean iguales), habrá ondas estacionarias en la línea, y no se transfiere la potencia máxima a la carga. Las ondas estacionarias causan pérdida de potencia, falla del dieléctrico, ruido, radiación y *señales fantasma*. Así, de ser posible, una línea de transmisión se debe adaptar a su carga. Para adaptar una línea de transmisión a una carga que tenga una impedancia distinta de Z_o se usan dos técnicas comunes. Son la compensación con transformador de un cuarto de longitud de onda mediante línea de acoplamiento.

Adaptación con transformador de un cuarto de longitud de onda. Los *transformadores de un cuarto de longitud de onda* se usan para compensar las líneas de transmisión con cargas puramente resistivas cuya resistencia no sea igual a la impedancia característica de la línea. Téngase en cuenta que un transformador de un cuarto de longitud de onda en realidad no es un transformador, sino más bien una sección de línea de transmisión de un cuarto de longitud de onda que funciona como si fuera un transformador. La impedancia de entrada a una línea de transmisión varía desde un valor máximo hasta uno mínimo o viceversa, cada cuarto de longitud de onda. Así, una línea de transmisión de un cuarto de longitud de onda de largo funciona como un *transformador elevador* o un *transformador reductor*, dependiendo de si Z_L es mayor

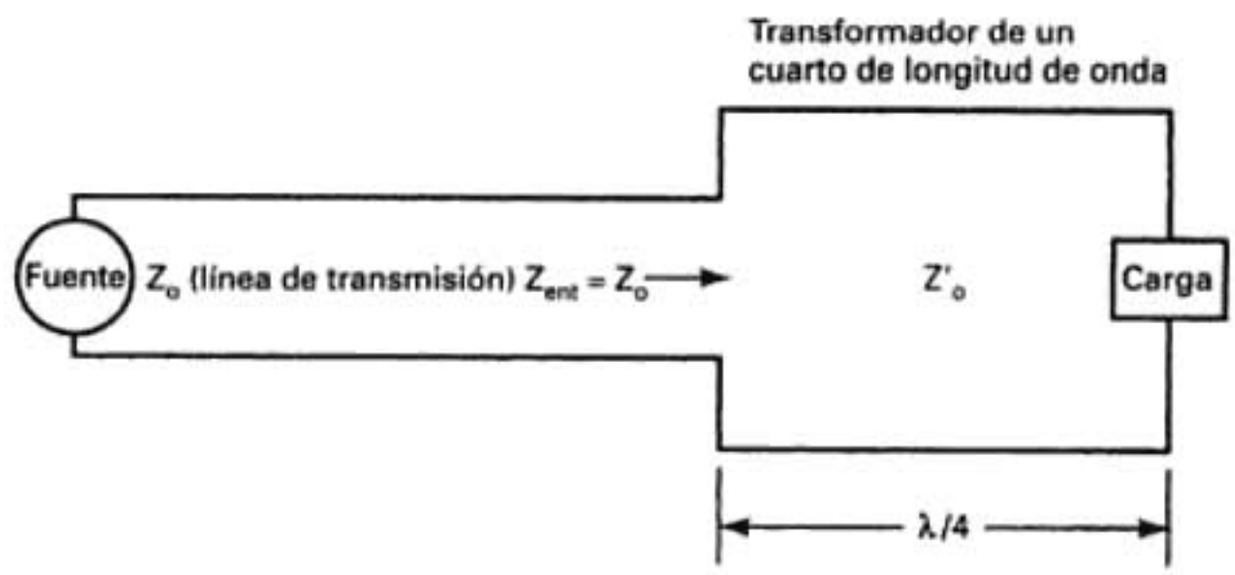


FIGURA 8-26 Transformador de un cuarto de longitud de onda

o menor que Z_o. Un transformador de un cuarto de longitud de onda no es un dispositivo de acoplamiento de impedancia de banda ancha; es de cuarto de longitud de onda en una sola frecuencia. Las transformaciones de impedancia para una línea de transmisión de un cuarto de longitud de onda son las siguientes:

1. $R_L = Z_o$: la línea de un cuarto de longitud de onda funciona como un transformador con relación de vueltas 1:1.
2. $R_L > Z_o$: la línea de un cuarto de longitud de onda funciona como un transformador reductor.
3. $R_L < Z_o$: la línea de un cuarto de longitud de onda funciona como un transformador elevador.

Como en los transformadores, uno de un cuarto de longitud de onda se pone entre una línea de transmisión y su carga. El transformador de un cuarto de longitud de onda no es más que un tramo de línea de transmisión de un cuarto de longitud de onda de largo. La fig. 8-26 muestra cómo se usa ese transformador para adaptar una línea de transmisión a una carga puramente resistiva. La impedancia característica de la parte de un cuarto de longitud de onda se calcula con la fórmula

$$Z'_0 = \sqrt{Z_o Z_L} \tag{8-32}$$

siendo Z'_0 = impedancia característica de un transformador de un cuarto de longitud de onda
Z_o = impedancia característica de la línea de transmisión que se está compensando
Z_L = impedancia de la carga

Ejemplo 8-5

Calcular la longitud física y la impedancia característica de un transformador de un cuarto de longitud de onda que se debe usar para compensar un tramo de cable RG-8A/U ($Z_o = 50\ \Omega$) a una carga resistiva de 150 Ω. La frecuencia de funcionamiento es 150 MHz y el factor de velocidad es $V_f = 1$.

Solución La longitud física del transformador depende de la longitud de onda de la señal. Al sustituir en la ecuación 8-2 se obtiene

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8 \text{ m/s}}{150 \text{ MHz}} = 2 \text{ m}$$

$$\frac{\lambda}{4} = \frac{2 \text{ m}}{4} = 0.5 \text{ m}$$

La impedancia característica del transformador de 0.5 m se determina con la ecuación 8-32

$$Z'_o = \sqrt{Z_o Z_L} = \sqrt{(50)(150)} = 86.6\ \Omega$$

Adaptación con línea de acoplamiento. Cuando una carga es puramente inductiva o puramente capacitiva no absorbe energía. El coeficiente de reflexión es 1, y la relación de onda estacionaria (SWR) es infinita. Cuando la carga es una impedancia compleja, y este suele ser el caso, es necesario eliminar la componente reactiva para adaptar la línea de transmisión a la carga. Para este objeto se utilizan líneas de *acoplamiento*. Una línea de acoplamiento a una línea

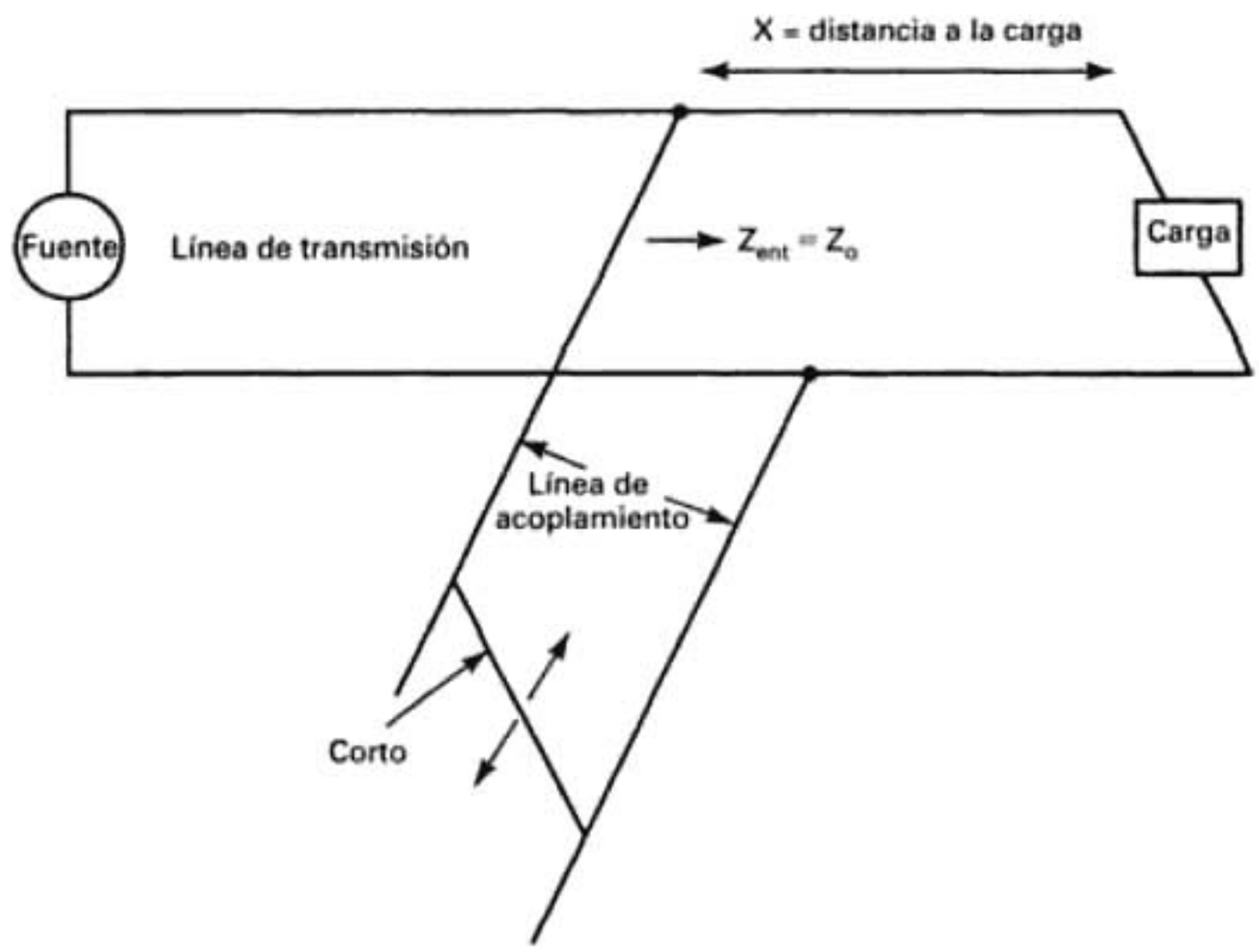


FIGURA 8-27 Acoplamiento de impedancia por línea de acoplamiento en corto

de transmisión no es más que un tramo adicional de línea de transmisión que se conecta entre los hilos de la línea primaria, tan cerca como sea posible de la carga. La susceptancia de la línea se usa para sintonizar la susceptancia de la carga. Para la *adaptación con línea de acoplamiento* se colocan líneas en corto o abiertas. Sin embargo, se prefieren las líneas en corto, porque las líneas abiertas tienen la tendencia a irradiar, en especial a frecuencias más elevadas.

La fig. 8-27 muestra cómo se conecta una línea en corto para anular la susceptancia de la carga y balancear la resistencia de carga con la impedancia característica de la línea de transmisión. Se ha demostrado cómo un tramo de línea de transmisión puesto en corto puede verse resistivo, inductivo o capacitivo, dependiendo de su longitud eléctrica. Una línea de transmisión de media longitud de onda o menor se puede usar para sintonizar y eliminar el componente resistivo de una carga.

El proceso de adaptar una carga a una línea de transmisión con una línea de acoplamiento en corto es el siguiente:

1. Ubicar un punto tan cerca de la carga como sea posible, donde el componente conductivo de la admitancia de entrada sea igual a la admitancia característica de la línea de transmisión.

$$Y_{ent} = G - jB, \text{ donde } G = \frac{1}{Z_o}$$

2. Fijar la línea puesta en corto al punto de la línea de transmisión determinado en el paso 1.
3. Dependiendo si en el punto localizado en el paso 1 el componente reactivo es inductivo o capacitivo, ajustar la longitud de la línea.

$$Y_{ent} = G_o - jB + jB_{acoplamiento}$$
$$= G_o$$

si $$B = B_{acoplamiento}$$

Véase el apéndice A, donde se explica con más detalle la adaptación con línea de acoplamiento usando el diagrama de Smith.

REFLECTOMETRÍA EN EL DOMINIO DEL TIEMPO

Los cables metálicos, como todos los componentes de un sistema electrónico de comunicaciones, pueden causar problemas que reduzcan su capacidad esperada. A menudo, los problemas

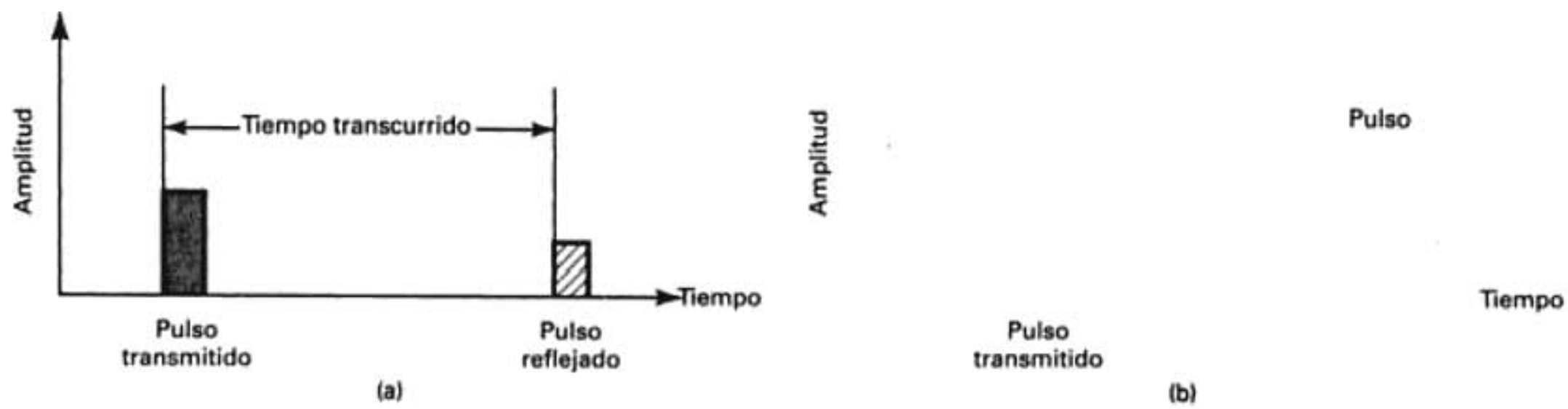


FIGURA 8-28 Reflectometría en el dominio del tiempo: (a) tiempo transcurrido; (b) pulso transmitido demasiado largo

de cables crean casos únicos porque a veces los cables recorren grandes distancias, que pueden ser de cientos de pies o más. Los problemas con cable se suelen atribuir a erosión química en los puntos de interconexión y a falla mecánica. Cuando se presenta un problema en un cable puede ser muy tardada y, en consecuencia, muy costosa la determinación de la clase y ubicación exacta del problema.

Una técnica que se usa para ubicar un defecto en un cable metálico se llama *reflectometría en el dominio del tiempo* (TDR, por *time-domain reflectometry*). Con ella, se pueden localizar los defectos de las líneas de transmisión con precisión de algunos pies en distancias de 10 millas. La TDR usa la bien demostrada teoría de que los defectos en las líneas de transmisión, como cortos o aberturas, hacen que una parte de la señal incidente regrese a la fuente. La cantidad de retorno de la señal transmitida depende de la clase y la magnitud del defecto. El punto en la línea donde está el defecto representa una discontinuidad para la señal. Esa discontinuidad hace que una parte de la señal transmitida se refleje, en vez de continuar por el cable. Si no regresa energía alguna, es decir, si la línea de transmisión y la carga están perfectamente balanceadas, la línea tiene longitud infinita o termina en una carga resistiva con impedancia igual a la impedancia característica de la línea. La TDR funciona en forma similar al *radar*. Un pulso de corta duración con corto tiempo de subida se propaga por un cable; a continuación se mide el tiempo en que regresa una parte de la señal a la fuente. Esta señal de regreso se llama *eco*, a veces. Al conocer la velocidad de propagación por el cable, se puede calcular la distancia exacta entre el defecto y la fuente, con la siguiente ecuación:

$$d = \frac{v \times t}{2} \tag{8-33}$$

siendo d = distancia a la discontinuidad (metros)
v = velocidad real (metros por segundo)
$v = k \times c$ (metros por segundo)
k = factor de velocidad (v/c) (adimensional)
c = velocidad de la luz en el vacío (3×10^8 metros por segundo)
t = tiempo transcurrido (segundos)

El tiempo transcurrido se mide desde la parte delantera del pulso transmitido hasta la recepción de la señal reflejada, como se indica en la fig. 8-28a. Es importante que el pulso transmitido sea tan corto como sea posible. De no ser así, cuando el defecto está cerca de la fuente, la señal reflejada podría regresar mientras todavía se estuviera transmitiendo el pulso (fig. 8-28b), dificultando la detección. Para las señales que viajan a la velocidad de la luz (c), la velocidad de propagación es 3×10^8 m/s, que aproximadamente es 1 ns/pie. En consecuencia, un pulso de varios microsegundos de ancho limitaría la utilidad de la TDR sólo para defectos de cable que estuvieran a varios miles de pies o más de distancia. La producción de pulsos extremadamente angostos fue uno de los factores limitantes en el desarrollo de la TDR para localizar fallas o cortos en cables.

Ejemplo 8-6

Un pulso se transmite por un cable, y la velocidad de propagación es 0.8 c. La señal reflejada se recibe 1 μs después ¿A qué distancia está el defecto en el cable?

Solución Se sustituyen cantidades en la ecuación 8-33

$$d = \frac{(0.8\,c) \times 1\ \mu\text{s}}{2}$$

$$= \frac{0.8 \times (3 \times 10^8\ \text{m/s}) \times 1 \times 10^{-6}\ \text{s}}{2} = 120\ \text{m}$$

Ejemplo 8-7

Se localiza un defecto en una línea de transmisión, usando TDR, a 3000 m de la fuente. La velocidad de propagación es 0.9 c. Calcular el tiempo transcurrido desde el inicio del pulso hasta la recepción del eco.

Solución Se reacomoda la ecuación 8-33 y se sustituyen valores como sigue

$$t = \frac{2d}{v} = \frac{2d}{k \times c}$$

$$= \frac{2(3000\ \text{m})}{0.9(3 \times 10^8\ \text{m/s})} = 22.22\ \mu\text{s}$$

LÍNEAS DE TRANSMISIÓN DE MICROCINTA Y DE CINTA

A frecuencias menores que unos 300 MHz, las características de las líneas de transmisión abiertas y en corto, como las que se describieron antes en este capítulo tienen poca importancia. Así, a bajas frecuencias, las líneas de transmisión normales serían demasiado largas para tener aplicación práctica como componentes reactivos o circuitos sintonizados. Sin embargo, para aplicaciones de alta frecuencia, de 300 a 3000 MHz, se han desarrollado líneas especiales de transmisión hechas con patrones de cobre en una tarjeta de *circuito impreso* (PC, por *printed circuit, o PCB, por printed circuit board*), llamadas *microcintas* y *línea de cinta*, para interconectar componentes en las tarjetas PC. También, cuando la distancia entre los extremos de fuente y de carga en una línea de transmisión es de unas pocas pulgadas o menor, es impráctico usar líneas de transmisión de cable coaxial normal, simplemente porque los conectores, las terminaciones y los cables mismos son demasiado grandes.

Tanto las microcintas como las líneas de cinta usan las pistas (o trazas) de cobre sobre la misma tarjeta de circuito impreso. Las trazas se pueden grabar con los mismos procesos que se usan en la fabricación de circuitos impresos, y, por consiguiente, ya no requieren más procesos de manufactura. Si las líneas se graban sólo en la superficie de la tarjeta PC, se llaman líneas de microcinta. Cuando se graban en la capa intermedia de una tarjeta PC de varias capas se llaman líneas de cinta. Las microcintas y las líneas de cinta se pueden usar como líneas de transmisión, inductores, capacitores, circuitos sintonizados, filtros, desplazadores de fase y dispositivos de acoplamiento de impedancia.

Microcinta

Una microcinta no es más que un conductor plano separado de un plano de tierra con un material dieléctrico aislante. En la fig. 8-29a se ve una línea de microcinta de una pista. El plano de tierra sirve como punto común del circuito, y debe ser cuando menos 10 veces más ancho que el conductor superior, y debe conectarse a tierra. En general, la microcinta es de un cuarto o media longitud de onda, a la frecuencia de operación, y equivale a una línea de transmisión desbalanceada. Se prefieren las líneas en corto sobre las líneas abiertas, porque éstas tienen mayor tendencia a irradiar. La fig. 8-29b muestra una línea de transmisión balanceada de dos hilos.

Como en cualquier línea de transmisión, la impedancia característica de una línea de microcinta depende de sus características físicas. Así, cualquier impedancia característica de 50 a 200 ohms puede obtenerse en las líneas de microcinta sólo con cambiar sus dimensiones. Lo

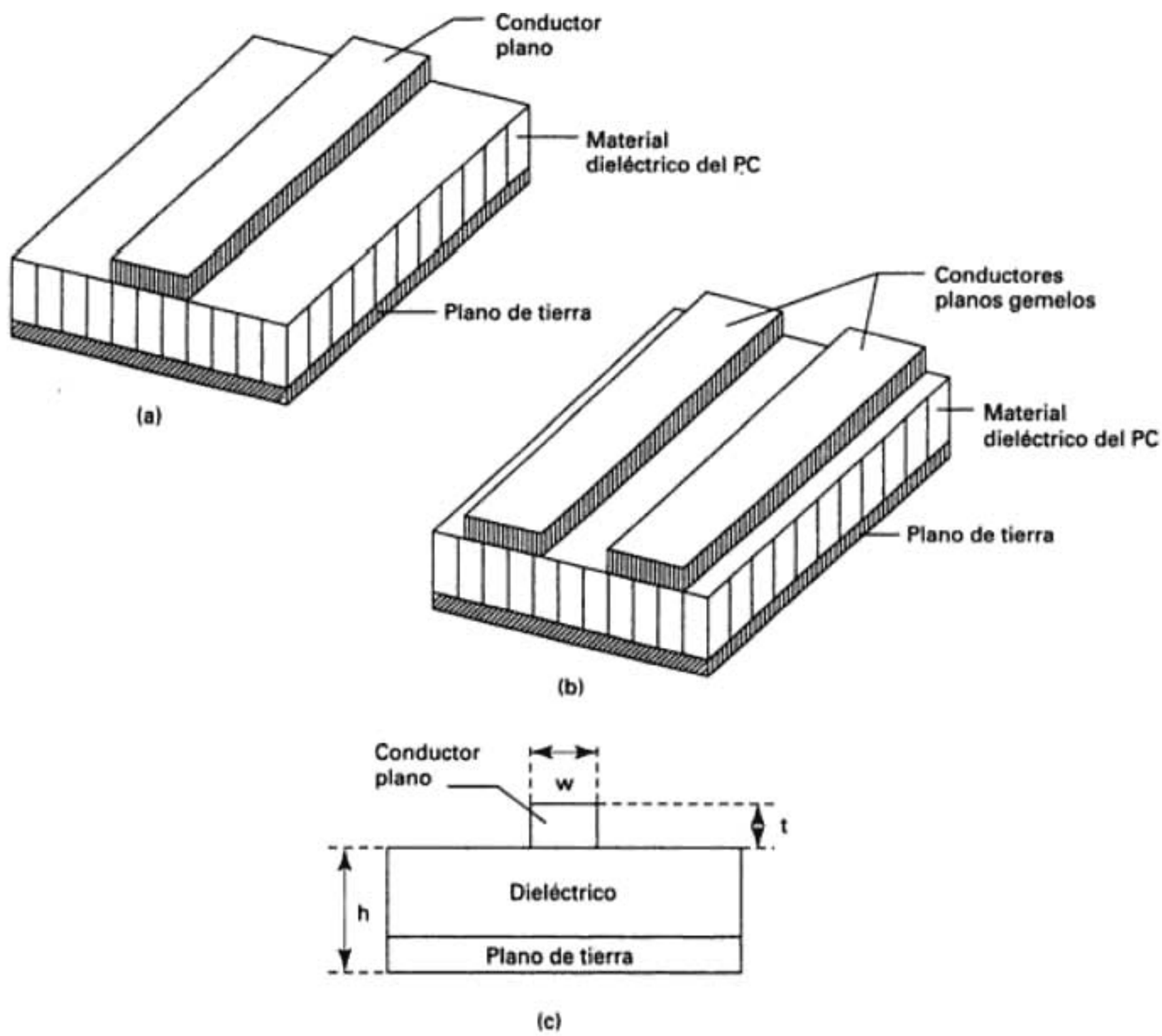


FIGURA 8-29 Línea de transmisión de microcinta: (a) desbalanceada; (b) balanceada; (c) dimensiones

mismo sucede con la cinta. Desafortunadamente, toda configuración de microcinta tiene su propia y exclusiva fórmula. La ecuación para calcular la impedancia característica de una línea desbalanceada de microcinta, como la de la fig. 8-29c, es

$$Z_o = \frac{87}{\sqrt{\epsilon + 1.41}} \ln\left(\frac{5.98h}{0.8w + t}\right) \tag{8-34}$$

en donde Z_o = impedancia característica (ohms)
ϵ = constante dieléctrica (para plástico reforzado con fibra de vidrio FR-4, $\epsilon = 4.5$ y para el teflón $\epsilon = 3$)
w = ancho de la pista de cobre*
t = espesor de la pista de cobre*
h = distancia entre la pista de cobre y el plano de tierra, es decir, el espesor del dieléctrico*

Línea de cinta

La línea de cinta no es más que un conductor plano emparedado entre dos planos de tierra, como se ve en la fig. 8-30. Aunque es más difícil de fabricar que la microcinta, es menos propensa a

* Las dimensiones de w, t y h pueden expresarse en cualquier unidad de longitud (pulgadas, milímetros, etc.), siempre y cuando se usen las mismas unidades.

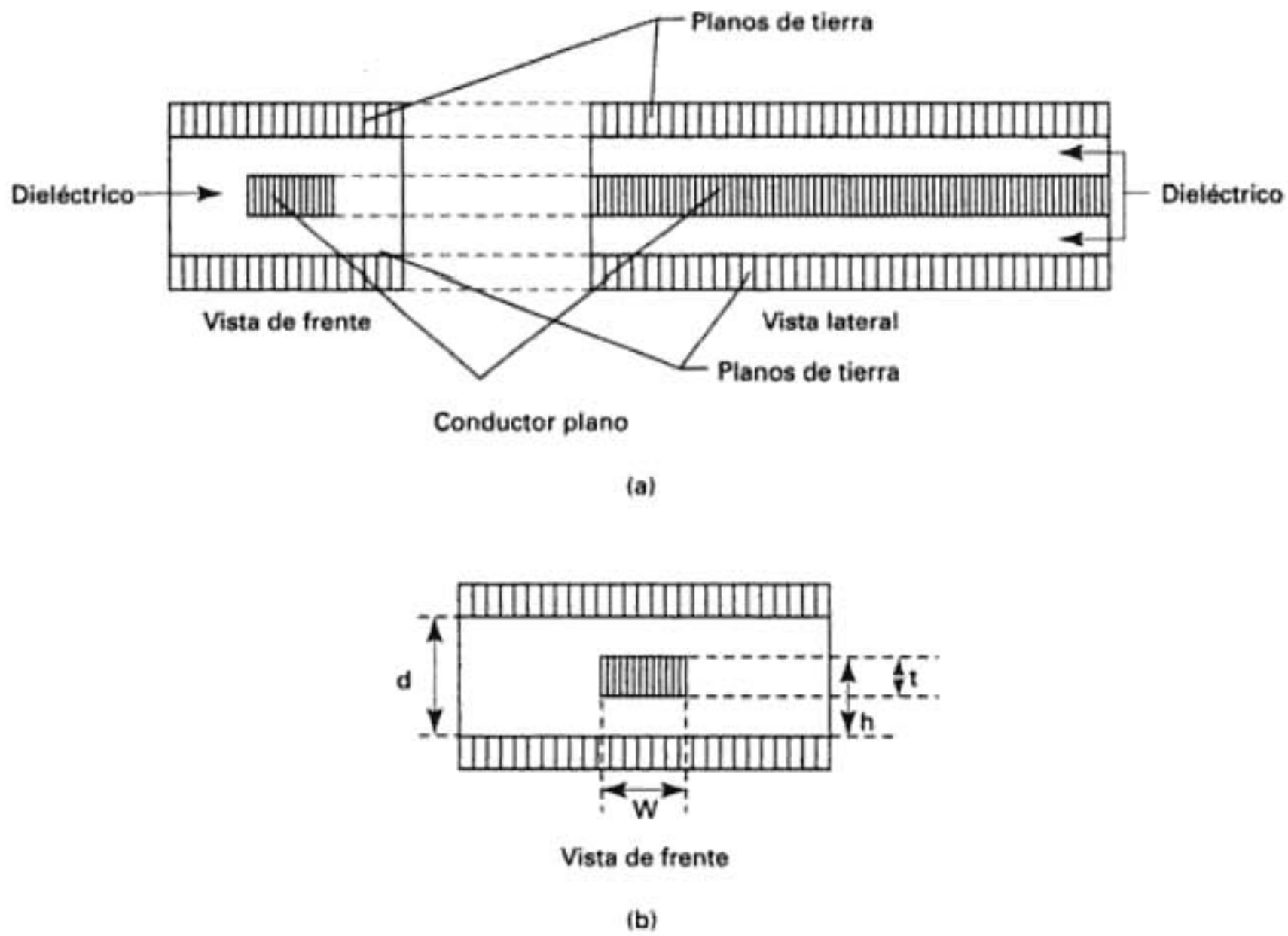


FIGURA 8-30 Línea de transmisión de cinta: (a) vistas de frente y lateral; (b) dimensiones

irradiar y así sus pérdidas son menores que las de la microcinta. De nuevo, la longitud de una línea de cinta puede ser de un cuarto o media longitud de onda, y se usan con más frecuencia las líneas en corto que las abiertas. La impedancia característica de una línea de cinta, configurada como se ve en la fig. 8-30 es

$$Z_o = \frac{60}{\epsilon} \ln\left(\frac{4d}{0.67\pi w(0.8 + t/h)}\right) \tag{8-35}$$

siendo Z_o = impedancia característica (ohms)
ϵ = constante dieléctrica (para plástico reforzado con fibra de vidrio FR-4, $\epsilon = 4.5$ y para el teflón $\epsilon = 3$)
d = espesor del dieléctrico*
w = ancho de la traza conductora de cobre*
t = espesor de la traza conductora de cobre*
h = distancia entre la traza de cobre y el plano de tierra*

PREGUNTAS

8-1. Defina *línea de transmisión*.

8-2. Describa una onda electromagnética transversal.

8-3. Defina *velocidad de onda*.

8-4. Defina *frecuencia* y *longitud de onda* para una onda electromagnética transversal.

8-5. Describa lo que son líneas de transmisión balanceadas y desbalanceadas.

8-6. Describa una línea de transmisión de conductores desnudos.

* Las dimensiones de d, w, t y h pueden expresarse en cualquier unidad de longitud (pulgadas, milímetros, etc.), siempre y cuando se usen las mismas unidades.

8-7. Describa una línea de transmisión de conductores gemelos.

8-8. ¿Qué es una línea de transmisión de par trenzado?

8-9. ¿Qué es una línea de transmisión de cable blindado?

8-10. Describa una línea de transmisión concéntrica.

8-11. Describa las propiedades eléctricas y físicas de una línea de transmisión.

8-12. Mencione y describa las cuatro constantes primarias de una línea de transmisión.

8-13. Defina la *impedancia característica* de una línea de transmisión.

8-14. ¿Qué propiedades de una línea de transmisión determinan su impedancia característica?

8-15. Defina la *constante de propagación* de una línea de transmisión.

8-16. Defina el *factor de velocidad* de una línea de transmisión.

8-17. ¿Qué propiedades de una línea de transmisión determinan su factor de velocidad?

8-18. ¿Qué propiedades de una línea de transmisión determinan su constante dieléctrica?

8-19. Defina lo que es la *longitud eléctrica* de una línea de transmisión.

8-20. Mencione y describa cinco clases de pérdidas en línea de transmisión.

8-21. Describa lo que es una onda incidente y una reflejada.

8-22. Describa lo que es una línea de transmisión resonante y una no resonante.

8-23. Defina el *coeficiente de reflexión*.

8-24. Describa qué son ondas estacionarias y qué es la relación de onda estacionaria.

8-25. Describa las ondas estacionarias que hay en una línea de transmisión abierta.

8-26. Describa las ondas estacionarias que hay en una línea de transmisión en cortocircuito.

8-27. Defina lo que es la *impedancia de entrada* en una línea de transmisión.

8-28. Describa el comportamiento de una línea de transmisión que termina en un cortocircuito, y es más larga que un cuarto de longitud de onda. Haga lo mismo con una línea más corta que un cuarto de longitud de onda.

8-29. Describa el comportamiento de una línea de transmisión que termina en circuito abierto y que es más larga que un cuarto de longitud de onda. Haga lo mismo con una línea más corta que un cuarto de longitud de onda.

8-30. Describa el comportamiento de una línea de transmisión abierta como elemento de circuito.

8-31. Describa el comportamiento de una línea de transmisión en corto como elemento de circuito.

8-32. Describa las características de impedancia de entrada de una línea de transmisión de un cuarto de longitud de onda.

8-33. Describa las características de impedancia de entrada de una línea de transmisión más corta que un cuarto de longitud de onda; haga lo mismo con una línea más larga que un cuarto de longitud de onda.

8-34. Describa la adaptación con un transformador de un cuarto de longitud de onda.

8-35. Describa como se hace la adaptación con línea de acoplamiento.

8-36. Describa la reflectometría en el dominio del tiempo.

PROBLEMAS

8-1. Calcule las longitudes de onda para ondas electromagnéticas en el espacio libre, que tengan las siguientes frecuencias: 1 kHz, 100 kHz, 1 MH y 1 GHz.

8-2. Calcule las frecuencias de ondas electromagnéticas en el espacio libre que tengan las siguientes longitudes de onda: 1 cm, 1 m, 10 m, 100 m y 1000 m.

8-3. Determine la impedancia característica de una línea de transmisión con dieléctrico de aire y relación $D/r = 8.8$.

8-4. Calcule la impedancia característica de una línea de transmisión concéntrica y llena de aire, con relación $D/d = 4$.

8-5. Calcule la impedancia característica de un cable coaxial con inductancia $L = 0.2\ \mu$H/pies y capacitancia $C = 16$ pF/pies.

8-6. Para determinado tramo de cable coaxial con capacitancia distribuida $C = 48.3$ pF/m e inductancia distribuida $L = 241.56$ nH/m, calcule el factor de velocidad y la velocidad de propagación.

8-7. Calcule el coeficiente de reflexión en una línea de transmisión con voltaje incidente $E_i = 0.2$ V y voltaje reflejado $E_r = 0.01$ V.

8-8. Determine la relación de onda estacionaria (SWR) para la línea de transmisión del problema 8-7.

8-9. Calcule la SWR para una línea de transmisión con amplitud máxima de onda estacionaria de voltaje $V_{máx} = 6$ V y amplitud mínima de onda estacionaria de voltaje $V_{mín} = 0.5$.

8-10. Calcule la SWR para una línea de transmisión de 50 Ω que termina en una resistencia de carga $Z_L = 75\ \Omega$.

8-11. Calcule la SWR para una línea de transmisión de 75 Ω que termina en una resistencia de carga $Z_L = 50\ \Omega$.

8-12. Calcule la impedancia característica de un transformador de un cuarto de longitud de onda con el que se adapta un tramo de línea de transmisión de 75 Ω a una carga resistiva de 100 Ω.

8-13. Al usar la TDR, se transmite un pulso por un cable, y la velocidad de propagación es 0.7 c. La señal reflejada se recibe 1.2 μs después. ¿A qué distancia está el defecto del cable?

8-14. Usando la TDR se ubicó un defecto en una línea de transmisión a 2500 m de la fuente. Para una velocidad de propagación de 0.95 c, calcule el tiempo transcurrido desde el inicio del pulso hasta la recepción del eco.

8-15. Al usar la TDR se ubicó un defecto en la línea de transmisión a 100 m de la fuente. Si el tiempo transcurrido desde el inicio del pulso hasta la recepción del eco es 833 ns, calcule el factor de velocidad.

8-16. Calcule las longitudes de las ondas electromagnéticas con las siguientes frecuencias: 5 kHz, 50 kHz, 500 kHz y 5 MHz.

8-17. Calcule las frecuencias de ondas electromagnéticas con las siguientes longitudes de onda: 5 cm, 50 cm, 5 m y 50 m.

8-18. Calcule la impedancia característica de una línea de transmisión de dieléctrico de aire, con relación $D/r = 6.8$.

8-19. Calcule la impedancia característica de una línea de transmisión concéntrica, llena de aire, con relación $D/d = 6$.

8-20. Calcule la impedancia característica de un cable coaxial con inductancia distribuida $L = 0.15$ μH/pie y capacitancia $C = 20$ pF/pie.

8-21. Para determinada longitud de cable coaxial, cuya capacitancia distribuida es $C = 24.15$ pF/m e inductancia distribuida es $L = 483.12$ nH/m, calcule el factor de velocidad y la velocidad de propagación.

8-22. Calcule el coeficiente de reflexión para una línea de transmisión con voltaje de incidencia $E_i = 0.4$ V y voltaje reflejado $E_r = 0.002$ V.

8-23. Calcule la relación de onda estacionaria para la línea de transmisión del problema 8-22.

8-24. Calcule la SWR para una línea de transmisión cuya amplitud máxima de onda estacionaria de voltaje es $V_{máx} = 0.8$ V y amplitud mínima de onda estacionaria de voltaje es $V_{mín} = 0.8$ V.

8-25. Calcule la SWR para una línea de transmisión de 50 Ω que termina en una resistencia de carga $Z_L = 60\ \Omega$.

8-26. Calcule la SWR para una línea de transmisión de 60 Ω que termina en una resistencia de carga $Z_L = 50\ \Omega$.

8-27. Calcule la impedancia característica de un transformador de cuarto de onda con el que se adapta un tramo de línea de transmisión de 50 Ω a una carga resistiva de 60 Ω.

CAPÍT

Imagen protegida por derechos de autor

Propagación de las ondas electromagnéticas

INTRODUCCIÓN

En el capítulo 8 se describieron las ondas electromagnéticas transversales (TEM, por *transverse electromagnetic*) y también se describió cómo los conductores metálicos se pueden usar como medio de transmisión para transferir las ondas TEM de uno a otro punto. Sin embargo, en los sistemas electrónicos de comunicaciones con frecuencia es impráctico o imposible interconectar dos equipos con una instalación física, como un hilo o cable metálico. Esto es válido en especial cuando los equipos están separados por grandes trechos de agua, escarpadas montañas o inhóspitos desiertos, o al comunicarse con transpondedores satelitales, en órbita a 22,000 millas sobre la Tierra. También, cuando los transmisores o receptores son móviles, como por ejemplo, en las radiocomunicaciones en dos sentidos y en los teléfonos móviles, es imposible tener conexiones con instalaciones metálicas. En consecuencia, se suele usar el espacio libre o la atmósfera terrestre como medio de transmisión.

La propagación de las ondas electromagnéticas por el *espacio libre* se suele llamar *propagación de radiofrecuencia* (RF), o simplemente *radio propagación*. Aunque el espacio libre implica al vacío, con frecuencia la propagación por la atmósfera terrestre se llama propagación por el espacio libre, y se puede considerar casi siempre así. La diferencia principal es que la atmósfera de la Tierra introduce pérdidas de la señal que no se encuentran en el vacío. Las TEM se propagan a través de cualquier material dieléctrico, incluyendo el aire. Sin embargo, no se propagan bien a través de conductores con pérdidas, como por ejemplo agua de mar, porque los campos eléctricos hacen que fluyan corrientes en el material, que disipan con rapidez la energía de las ondas.

Las ondas de radio son ondas electromagnéticas y, como la luz, se propagan a través del espacio libre en línea recta y con velocidad de 300,000,000 metros por segundo. Otras formas de ondas electromagnéticas son los rayos infrarrojos, los ultravioleta, los rayos X y los rayos gamma. Para propagar las ondas de radio por la atmósfera terrestre se necesita que la energía se irradie de la fuente. A continuación, la energía se debe capturar en el lado de recepción. La irradiación y la captura de energía son funciones de las antenas, y se explicarán en el capítulo 10; las propiedades de las ondas electromagnéticas se explicaron en el capítulo 8. El objetivo de este capítulo es describir la naturaleza, comportamiento y propiedades ópticas de las ondas de radio que se propagan a través de la atmósfera terrestre.

POLARIZACIÓN ELECTROMAGNÉTICA

Como se explicó en el capítulo 8, una onda electromagnética contiene un campo eléctrico y uno magnético, que forman 90° entre sí. La *polarización* de una onda electromagnética plana no es más que la orientación del vector de campo eléctrico con respecto a la superficie de la Tierra, es decir, respecto al horizonte. Si la polarización permanece constante se llama *polarización lineal*. La *polarización horizontal* y la *polarización vertical* son dos formas de polarización lineal. Si el campo eléctrico se propaga en dirección paralela a la superficie terrestre, se dice que la onda está polarizada horizontalmente. Si el campo eléctrico se propaga en dirección perpendicular a la superficie terrestre, se dice que la onda está polarizada verticalmente. Si el vector de polarización gira 360° a medida que la onda recorre una longitud de onda por el espacio, y la intensidad de campo es igual en todos los ángulos de polarización, se dice que la onda tiene *polarización circular*. Cuando la intensidad de campo varía con cambios en la polarización, se dice que es una *polarización elíptica*. Una onda rotatoria puede girar en cualquier dirección. Si el vector gira en dirección de las manecillas del reloj (_), es derecho, y si gira en dirección contraria (_), se considera izquierdo.

RAYOS Y FRENTES DE ONDA

Las ondas electromagnéticas son invisibles y, en consecuencia, se deben analizar con métodos indirectos, mediante esquemas. Los conceptos de *rayos* y de *frentes de onda* son auxiliares para ilustrar los efectos de la propagación de ondas electromagnéticas a través del espacio vacío. Un rayo es una línea trazada a lo largo de la dirección de propagación de una onda electromagnética. Los rayos se usan para mostrar la dirección relativa de la propagación de la onda electromagnética; sin embargo, no necesariamente representan la propagación de una sola onda electromagnética. En la fig. 9-1 se muestran varios rayos (R_a, R_b, R_c, etc.). Un frente de onda representa una superficie de ondas electromagnéticas de fase constante. Se forma un frente de onda cuando se unen puntos de igual fase en rayos que se propagan desde la misma fuente. La fig. 9-1 muestra un frente de onda con una superficie que es perpendicular a la dirección de propagación: el rectángulo *ABCD*. Cuando una superficie es plana, su frente de onda es perpendicular a la dirección de propagación. Cuanto más cerca está de su fuente, el frente de onda se vuelve más complicado.

La mayoría de los frentes de onda son más complicados que los de una simple onda plana. La fig. 9-2 muestra una fuente puntual, varios rayos que se propagan desde ella, y el frente de onda correspondiente. Una *fuente puntual* es un solo lugar desde el cual se propagan rayos por igual en todas direcciones: es una *fuente isotrópica*. El frente de onda generado por una fuente puntual sólo es una esfera con radio *R*, y su centro está en el punto de origen de las ondas. En el espacio libre, y a una distancia suficiente de la fuente, los rayos dentro de una superficie pequeña de un frente de onda esférico son casi paralelos. Por consiguiente, a mayor distancia de la fuente, la propagación de la onda se parece más a la de un frente de onda plano.

RADIACIÓN ELECTROMAGNÉTICA

Densidad de potencia e intensidad de campo

Las ondas electromagnéticas representan el flujo de energía en la dirección de propagación. La rapidez con que la energía pasa a través de una superficie dada en el espacio libre se llama *den-*

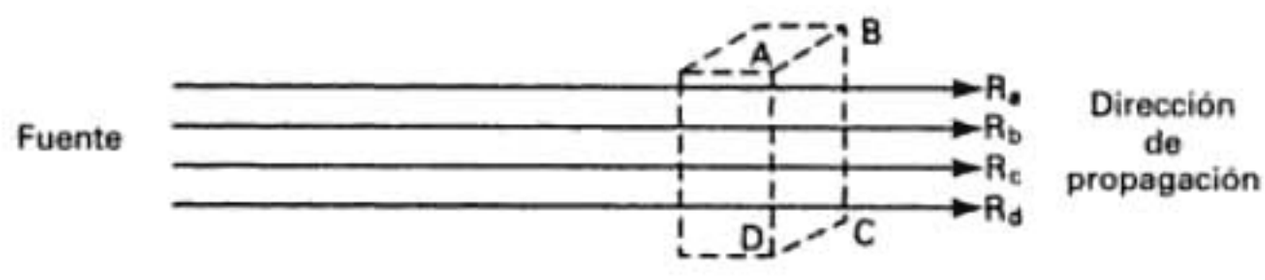


FIGURA 9-1 Onda plana

sidad de potencia. Por consiguiente, la densidad de potencia es la energía por unidad de tiempo y por unidad de área, y se suele expresar en watts por metro cuadrado. La *intensidad de campo* es la intensidad de los campos eléctrico y magnético de una onda electromagnética que se propaga por el espacio libre. La intensidad del campo eléctrico se suele expresar en volts por metro, y la del campo magnético en amperes por metro (A/m). La densidad de potencia es:

$$\mathcal{P} = \mathcal{E}\mathcal{H} \quad \text{W/m}^2 \tag{9-1}$$

siendo $\mathcal{P}$ = densidad de potencia (watts por metro cuadrado)
$\mathcal{E}$ = intensidad rms del campo eléctrico (volts por metro)
$\mathcal{H}$ = intensidad rms del campo magnético (amperes por metro)

Impedancia característica del espacio libre

Las intensidades del campo eléctrico y magnético de una onda electromagnética en el espacio libre se relacionan a través de la impedancia característica (resistencia) del espacio vacío. La impedancia característica de un medio de transmisión sin pérdidas es igual a la raíz cuadrada de la relación de su permeabilidad magnética entre su permitividad eléctrica. La ecuación de la impedancia característica del espacio libre, Z_s, es

$$Z_s = \sqrt{\frac{\mu_0}{\epsilon_0}} \tag{9-2}$$

en la que Z_s = impedancia característica del espacio libre (ohms)
μ_0 = permeabilidad magnética del espacio libre, 1.26×10^{-6} H/m
ϵ_0 = permitividad eléctrica del espacio libre, 8.85×10^{-12} F/m

Al sustituir valores en la ecuación 9-2, se obtiene

$$Z_s = \sqrt{\frac{1.26 \times 10^{-6}}{8.85 \times 10^{-12}}} = 377\ \Omega$$

Por consiguiente, al aplicar la ley de Ohm se obtienen

$$\mathcal{P} = \frac{\mathcal{E}^2}{377} = 377\mathcal{H}^2 \quad \text{W/m}^2 \tag{9-3}$$

$$\mathcal{H} = \frac{\mathcal{E}}{377} \quad \text{A/m} \tag{9-4}$$

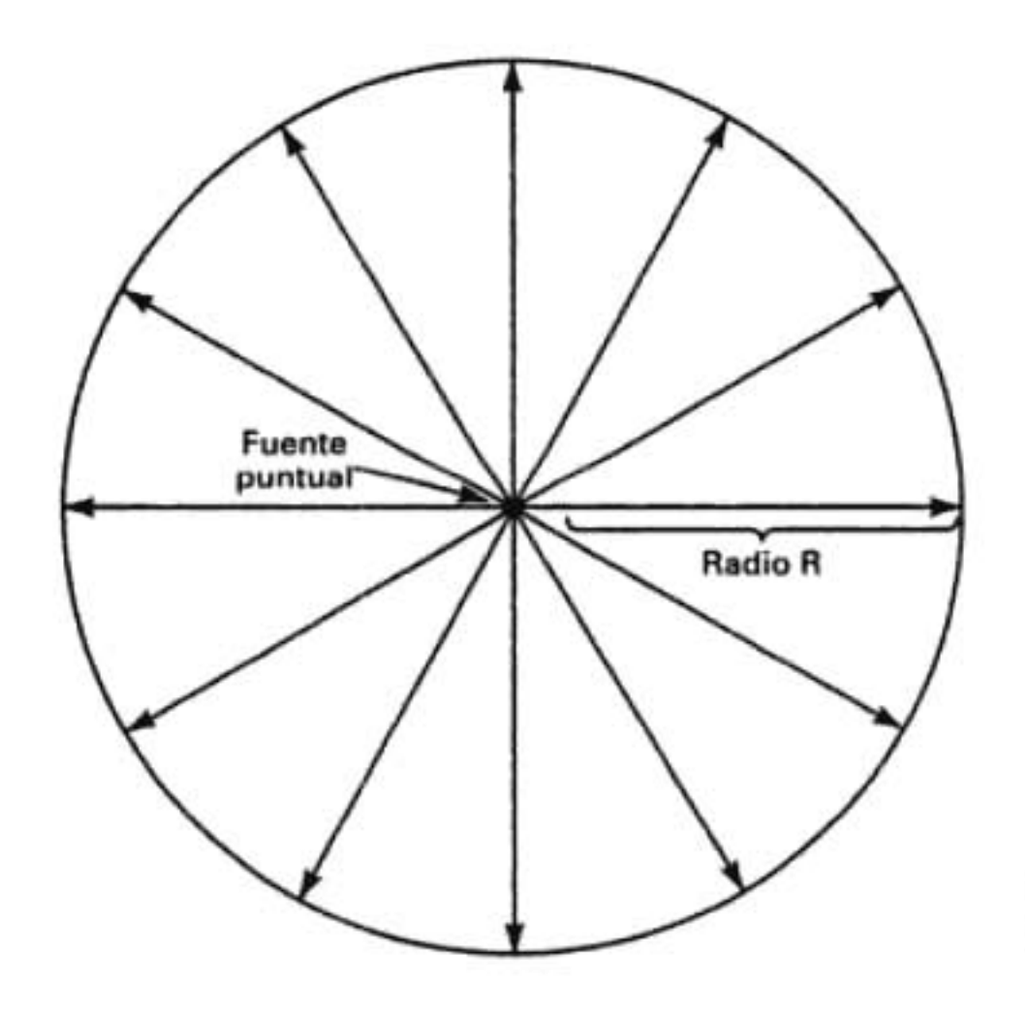


FIGURA 9-2 Frente de onda producido por una fuente puntual

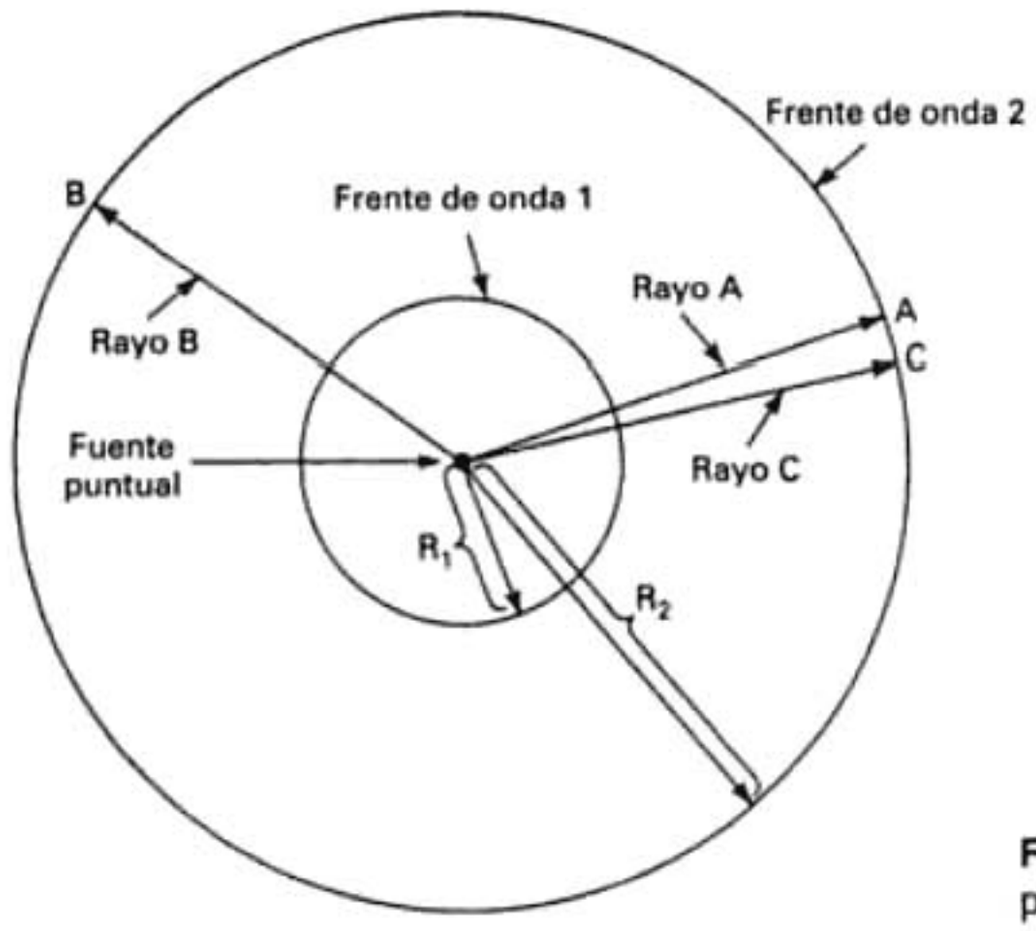


FIGURA 9-3 Frente de onda esférico producido por una fuente isotrópica

FRENTE DE ONDA ESFÉRICO Y LA LEY DEL CUADRADO INVERSO

Frente de onda esférico

La fig. 9-3 muestra una fuente puntual que irradia potencia a una tasa constante y uniformemente en todas direcciones. Esa fuente se llama *radiador isotrópico*. No existe un radiador realmente isotrópico; sin embargo, se puede aproximar a una *antena omnidireccional*. Un radiador isotrópico produce un frente de onda esférico cuyo radio es R. Todos los puntos que están a la distancia R de la fuente están en la superficie de una esfera, y tienen igual densidad de potencia. Por ejemplo, en la fig. 9-3, los puntos A y B están a igual distancia de la fuente. En consecuencia, las densidades de potencia en ellos son iguales. En cualquier momento, la potencia irradiada, P_r watts, está uniformemente distribuida sobre la superficie total de la esfera (se supone que el medio de transmisión no tiene pérdidas). Así, la densidad de potencia en cualquier punto de la esfera es la potencia total irradiada dividida entre el área total de la esfera. La ecuación que expresa la densidad de potencia en cualquier punto de la superficie de un frente de onda esférico es

$$\mathcal{P} = \frac{P_{rad}}{4\pi R^2} \tag{9-5}$$

en donde P_{rad} = potencia total irradiada (watts)
R = radio de la esfera, que es igual a la distancia de cualquier punto de la superficie de la esfera a la fuente
$4\pi R^2$ = área de la esfera

y para una distancia R_a metros de la fuente, la densidad de potencia es

$$\mathcal{P}_a = \frac{P_{rad}}{4\pi R_a^2}$$

Se igualan las ecuaciones 9-3 y 9-5 para obtener

$$\frac{P_{rad}}{4\pi R^2} = \frac{\mathcal{E}^2}{377}$$

Por consiguiente,

$$\mathcal{E}^2 = \frac{377 P_{rad}}{4\pi R^2} \quad \text{y} \quad \mathcal{E} = \frac{\sqrt{30 P_{rad}}}{R} \tag{9-6}$$

Ley del cuadrado inverso

Se ve, en la ecuación 9-5, que cuanto más lejos va el frente de onda respecto a la fuente, la densidad de potencia es más pequeña: R_a y R_c se alejan entre sí. La potencia total distribuida en la

superficie de la esfera queda igual. Sin embargo, como el área de la esfera aumenta en proporción directa a la distancia a la fuente elevada al cuadrado, es decir, al radio de la esfera al cuadrado, la densidad de potencia es inversamente proporcional al cuadrado de la distancia de la fuente. A esta relación se le llama *ley del cuadrado inverso*. Entonces, la densidad de potencia en cualquier punto de la superficie de la esfera exterior es

$$\mathcal{P}_2 = \frac{P_{rad}}{4\pi R_2^2}$$

y la densidad de potencia en cualquier punto de la esfera interior es

$$\mathcal{P}_1 = \frac{P_{rad}}{4\pi R_1^2}$$

Por consiguiente,

$$\frac{\mathcal{P}_2}{\mathcal{P}_1} = \frac{P_{rad}/4\pi R_2^2}{P_{rad}/4\pi R_1^2} = \frac{R_1^2}{R_2^2} = \left(\frac{R_1}{R_2}\right)^2 \tag{9-7}$$

Según esta ecuación, a medida que se duplica la distancia a la fuente, la densidad de potencia decrece en un factor de 2^2, o 4. Cuando se dedujo la ley del cuadrado inverso de la radiación, ecuación 9-7, se supuso que la fuente irradia isotrópicamente, aunque eso no es necesario; sin embargo, sí es necesario que la velocidad de propagación en todas direcciones sea uniforme. A un medio de propagación con estas propiedades se le llama *medio isotrópico*.

Ejemplo 9-1

Determinar, para una antena isotrópica que irradia 100 W de potencia:
(a) Densidad de potencia a 1000 m de la fuente.
(b) Densidad de potencia a 2000 m de la fuente.

Solución (a) Se sustituye en la ecuación 9-5, para obtener

$$\mathcal{P}_1 = \frac{100}{4\pi 1000^2} = 7.96\ \mu\text{W/m}^2$$

(b) De nuevo se sustituyen valores en la ecuación 9-5

$$\mathcal{P}_2 = \frac{100}{4\pi 2000^2} = 1.99\ \mu\text{W/m}^2$$

o bien, si se sustituyen en la ecuación 9-7, se obtendrá

$$\frac{\mathcal{P}_2}{\mathcal{P}_1} = \frac{1000^2}{2000^2} = 0.25$$

o sea que $\mathcal{P}_2 = 7.96\ \mu\text{W/m}^2\ (0.25) = 1.99\ \mu\text{W/m}^2$

ATENUACIÓN Y ABSORCIÓN DE ONDAS

El espacio libre es el vacío, por lo que no hay pérdida de energía al propagarse una onda por él. Sin embargo, cuando las ondas se propagan por el espacio vacío, se dispersan y resulta una reducción de la densidad de potencia. A esto se le llama *atenuación*, y se presenta tanto en el espacio libre como en la atmósfera terrestre. Ya que la atmósfera terrestre no es un vacío, contiene partículas que pueden absorber energía electromagnética. A este tipo de reducción de potencia se le llama *pérdida por absorción*, y no se presenta en ondas que viajan fuera de nuestra atmósfera.

Atenuación

La ley del cuadrado inverso de la radiación describe en forma matemática la reducción de densidad de potencia con la distancia a la fuente. A medida que se aleja un frente de onda de la fuente, el campo electromagnético continuo que irradia la fuente se dispersa. Esto es, las ondas se alejan cada vez más entre sí y, en consecuencia, la cantidad de ondas por unidad de área es menor. No se pierde o disipa nada de la potencia irradiada, porque el frente de onda se aleja de la fuente; la on-

da sólo se extiende, o se dispersa, sobre un área mayor y disminuye la densidad de potencia. La reducción de densidad de potencia con la distancia equivale a una pérdida de potencia, y se suele llamar *atenuación de la onda*. Como la atenuación se debe a la dispersión esférica de la onda, a veces se le llama *atenuación espacial* de la onda. La atenuación de la onda se expresa en general en función del logaritmo común de la relación de densidades de potencia (pérdida en dB). La definición matemática de γ_a es

$$\gamma_a = 10 \log \frac{\mathcal{P}_1}{\mathcal{P}_2} \tag{9-8}$$

La relación de densidad de potencia debida a la ley del cuadrado inverso supone que la propagación es en el espacio libre: en el vacío o casi en el vacío, y se llama atenuación de onda. La reducción de la densidad de potencia debida a propagación en espacio no libre se llama *absorción*.

Absorción

La atmósfera terrestre no es un vacío. Más bien está formada por átomos y moléculas de diversas sustancias gaseosas, líquidas y sólidas. Algunos de esos materiales pueden absorber las ondas electromagnéticas. Cuando una onda electromagnética se propaga a través de la atmósfera terrestre, se transfiere energía de la onda a los átomos y moléculas atmosféricos. La absorción de onda por la atmósfera es análoga a una pérdida de potencia I^2R. Una vez absorbida, la energía se pierde para siempre, y causa una atenuación en las intensidades de voltaje y campo magnético, y una reducción correspondiente de densidad de potencia.

La absorción de las radiofrecuencias en una atmósfera normal depende de su frecuencia, y es relativamente insignificante a menos de unos 10 GHz. La fig. 9-4 muestra la absorción atmosférica, en decibeles por kilómetro, debida al oxígeno y al vapor de agua, para radiofrecuencias mayores de 10 GHz. Se aprecia que ciertas frecuencias se afectan más o menos por la absorción, y se producen picos y valles en las curvas. La atenuación de ondas debida a la absorción no depende de la distancia a la fuente de radiación, sino más bien a la distancia total que la onda se propaga a través de la atmósfera. En otras palabras, para un *medio homogéneo*, cuyas propiedades son uniformes en todo él, la absorción sufrida durante el primer kilómetro de propagación es igual que la del último kilómetro. También, las condiciones atmosféricas anormales, como por ejemplo lluvias intensas o neblina densa, absorben más energía que una atmósfera normal. La absorción atmosférica se representa por η y, para una onda que se propaga de R_1 a R_2, es $\gamma(R_2 - R_1)$, siendo γ el coeficiente de absorción. Así, la atenuación de onda depende de la relación R_2/R_1, y la absorción de onda depende de la distancia entre R_1 y R_2. En el

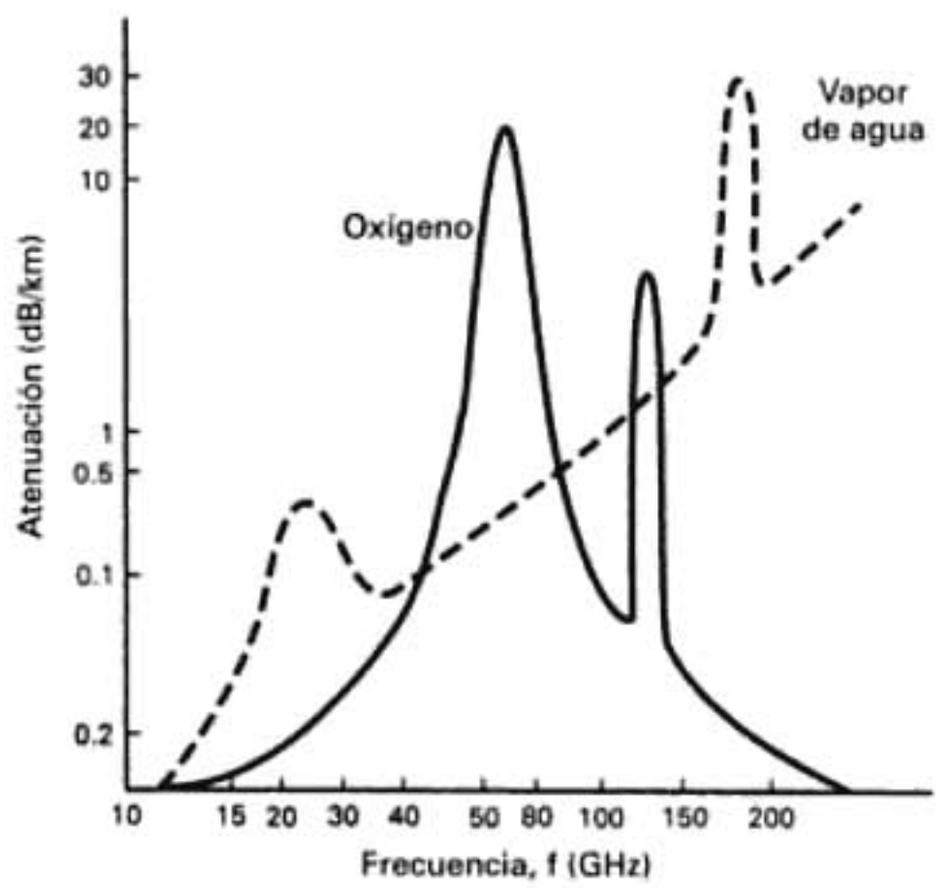


FIGURA 9-4 Absorción atmosférica de las ondas electromagnéticas

caso más real, es decir, en un *medio no homogéneo*, el coeficiente de absorción varía mucho de acuerdo con el lugar y por lo mismo origina difíciles problemas para los ingenieros de sistemas de radio.

PROPIEDADES ÓPTICAS DE LAS ONDAS DE RADIO

En la atmósfera terrestre, la propagación de frentes de ondas y rayos puede diferir del comportamiento en el espacio libre, debido a efectos *ópticos*, como *refracción, reflexión, difracción* e *interferencia*. En una terminología muy coloquial, se puede imaginar la refracción como la *flexión*, la reflexión como *rebote*, la difracción como *dispersión* y la interferencia como *choques*. Se dice que la refracción, la reflexión, la difracción y la interferencia son propiedades ópticas porque se observaron primero en la ciencia de la óptica, que estudia el comportamiento de las ondas luminosas. Como las ondas luminosas son ondas electromagnéticas de alta frecuencia, parece razonable que las propiedades ópticas también se apliquen a la propagación de las ondas de radio. Aunque se pueden analizar por completo los principios ópticos aplicando las ecuaciones de Maxwell, lo cual es complicado por necesidad. Para la mayoría de las aplicaciones, se pueden sustituir las ecuaciones de Maxwell por el *trazo geométrico de rayos*.

Refracción

La *refracción* electromagnética es el cambio de dirección de un rayo al pasar en dirección oblicua de un medio a otro con distinta velocidad de propagación. La velocidad a la que se propaga una onda electromagnética es inversamente proporcional a la densidad del medio en el que lo hace. Por consiguiente, hay refracción siempre que una onda de radio pasa de un medio a otro de distinta densidad. La fig. 9-5 muestra la refracción de un frente de onda en una frontera *plana* entre dos medios con distintas densidades. Para este ejemplo, el medio 1 es menos denso que el medio 2, por lo que $v_1 > v_2$. Se puede ver que el rayo *A* entra al medio más denso antes que el rayo *B*. Así, el rayo *B* se propaga con más rapidez que el rayo *A*, y viaja la distancia *B-B′* durante el mismo tiempo que el rayo *A* recorre la distancia *A-A′*. Por consiguiente, el frente de onda *A′B′* se *inclina* o se dobla hacia abajo. Como un rayo se define como perpendicular al frente de onda en todos los puntos de éste, los rayos de la fig. 9-5 cambiaron de dirección en la interfase entre los dos medios. Siempre que un rayo pasa de un medio menos denso a uno más denso, se dobla hacia la *normal*. La normal no es más que una línea imaginaria, trazada perpendicular a la interfase en el punto de incidencia. Al revés, siempre que un rayo pasa de un medio más

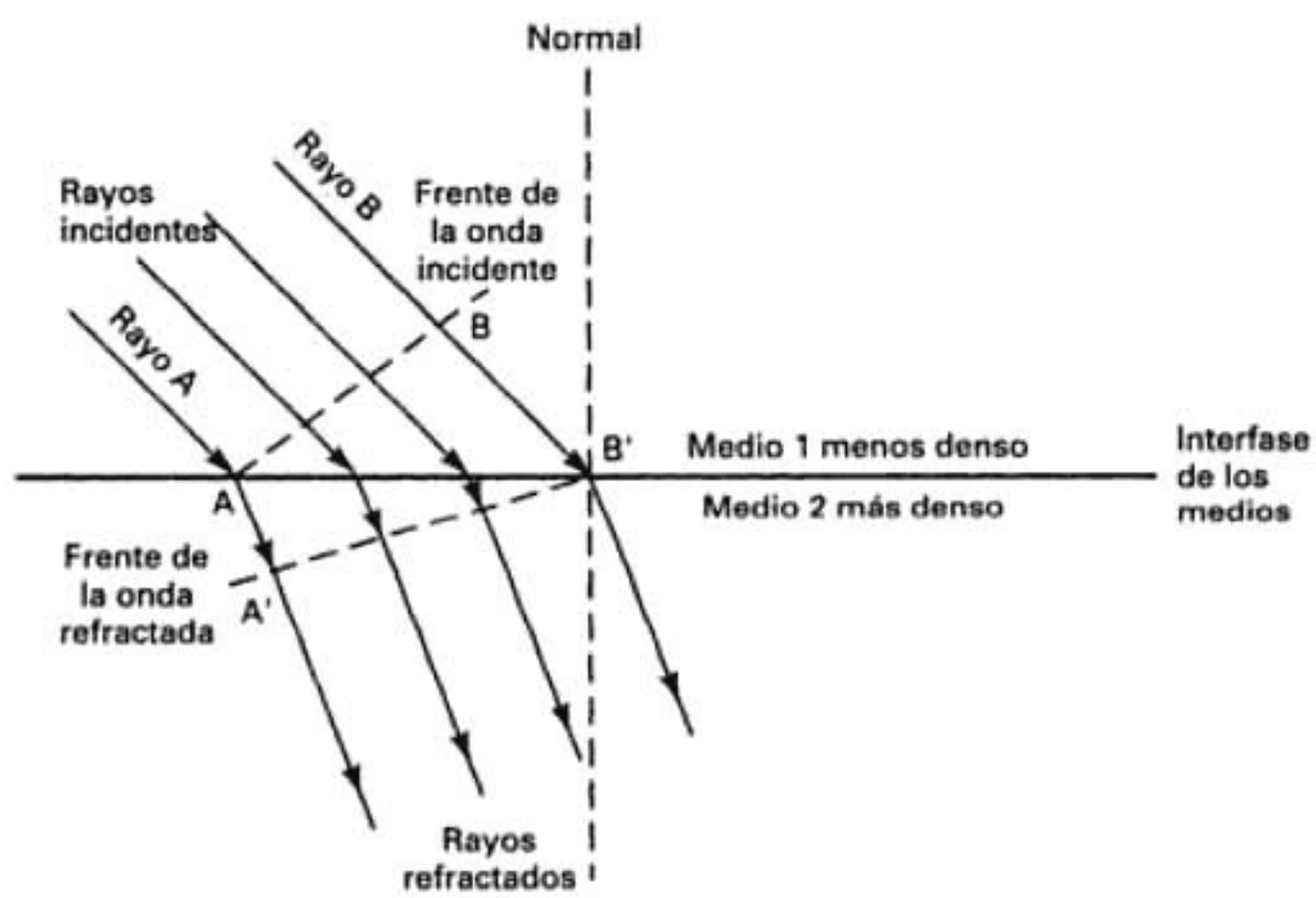


FIGURA 9-5 Refracción en una frontera plana entre dos medios

denso a uno menos denso se dobla alejándose de la normal. El *ángulo de incidencia* es el que forman la onda incidente y la normal, y el *ángulo de refracción* es el que forman la onda refractada y la normal.

El grado de flexión o refracción que hay en la interfase entre dos materiales de distintas densidades es bastante predecible, y depende del *índice de refracción* de cada material. El índice de refracción no es más que la relación de la velocidad de propagación de la luz en el espacio vacío entre la velocidad de propagación de la luz en determinado material, es decir,

$$n = \frac{c}{v} \tag{9-9}$$

en la que n = índice de refracción (adimensional)
c = velocidad de la luz en el espacio libre (3×10^8 m/s)
v = velocidad de la luz en determinado material (metros por segundo)

El índice de refracción también es una función de la frecuencia. Sin embargo, en la mayoría de las aplicaciones la variación es insignificante y, en consecuencia, se omitirá de esta descripción. La forma en que reacciona una onda electromagnética cuando llega a la interfase entre dos materiales transmisores que tienen distintos índices de refracción se describe con la *ley de Snell*. La ley de Snell establece que

$$n_1 \operatorname{sen} \theta_1 = n_2 \operatorname{sen} \theta_2 \tag{9-10}$$

y

$$\frac{\operatorname{sen} \theta_1}{\operatorname{sen} \theta_2} = \frac{n_2}{n_1}$$

en donde n_1 = índice de refracción del material 1
n_2 = índice de refracción del material 2
θ_1 = ángulo de incidencia (grados)
θ_2 = ángulo de refracción (grados)

y como el índice de refracción de un material es igual a la raíz cuadrada de su constante dieléctrica,

$$\frac{\operatorname{sen} \theta_1}{\operatorname{sen} \theta_2} = \sqrt{\frac{\epsilon_{r2}}{\epsilon_{r1}}} \tag{9-11}$$

donde ϵ_{r1} = constante dieléctrica del medio 1
ϵ_{r2} = constante dieléctrica del medio 2

También se presenta la refracción cuando un frente de onda se propaga en un medio que tiene un *gradiente de densidad*, perpendicular a la dirección de propagación, es decir, paralelo al frente de onda. La fig. 9-6 representa la refracción de un frente de onda en un medio de transmisión que tiene una variación gradual en su índice de refracción. El medio es más denso en la parte inferior, y menos denso en la parte superior. Entonces, los rayos que viajan cerca de la parte superior lo hacen con mayor rapidez que los que están cerca de la parte inferior y, en consecuencia, el frente de onda se inclina hacia abajo. La desviación se hace en forma gradual a medida que avanza el frente de onda, como se ve en la figura.

Reflexión

Reflejar quiere decir regresar, y la *reflexión* es el acto de reflejar. La reflexión electromagnética se presenta cuando una onda incidente choca con una frontera entre dos medios, y algo o toda la potencia incidente no entra al segundo material. Las ondas que no penetran al segundo medio se reflejan. La fig. 9-7 muestra la reflexión de una onda electromagnética en un plano limítrofe entre dos medios. Como todas las ondas reflejadas permanecen en el medio 1, las velocidades de las ondas incidente y reflejada son iguales. En consecuencia, el *ángulo de reflexión* es igual al *ángulo de incidencia*, $\theta_i = \theta_r$. Sin embargo, la intensidad del campo del voltaje reflejado es

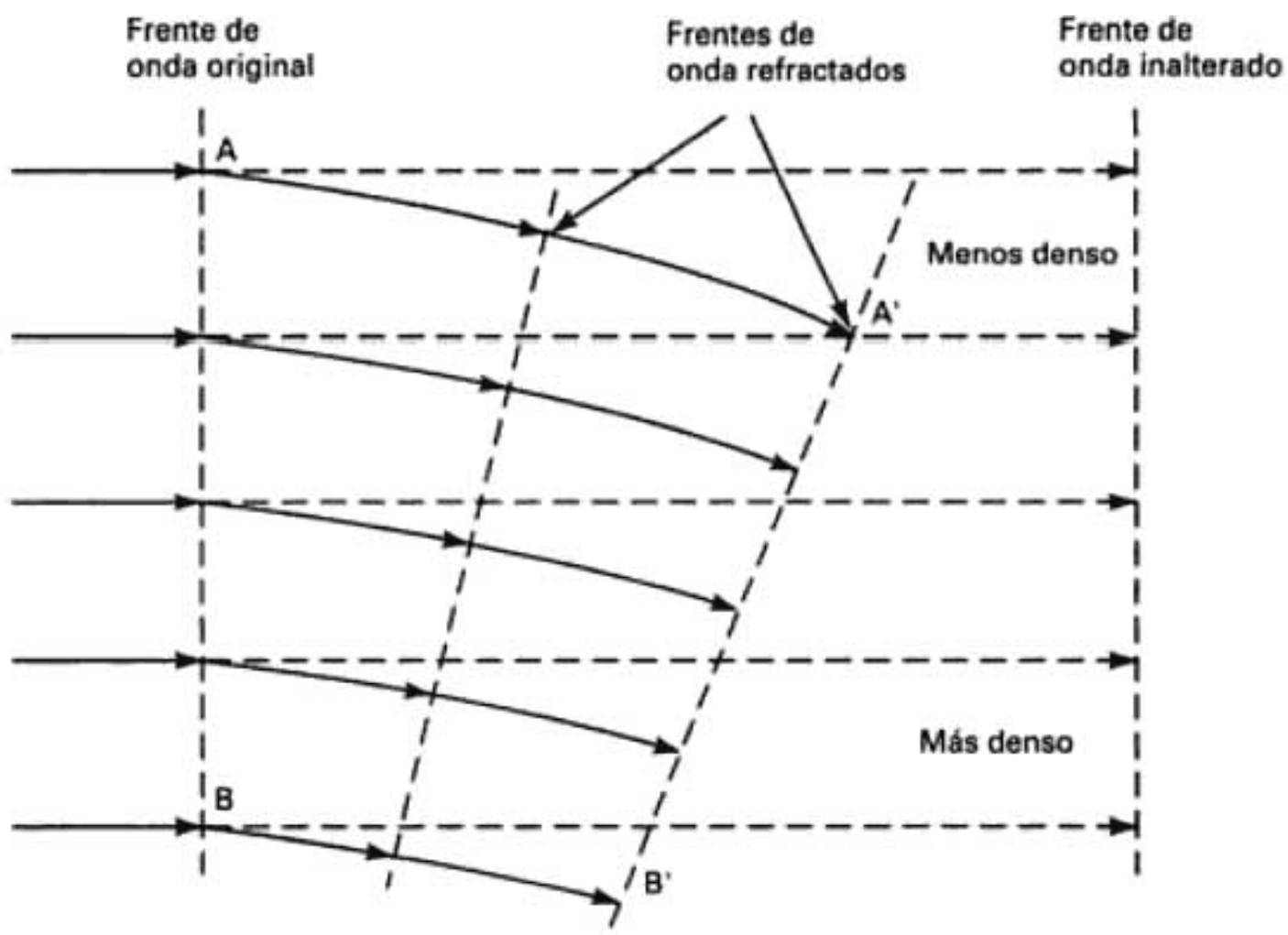


FIGURA 9-6 Refracción de un frente de onda en un medio con gradiente

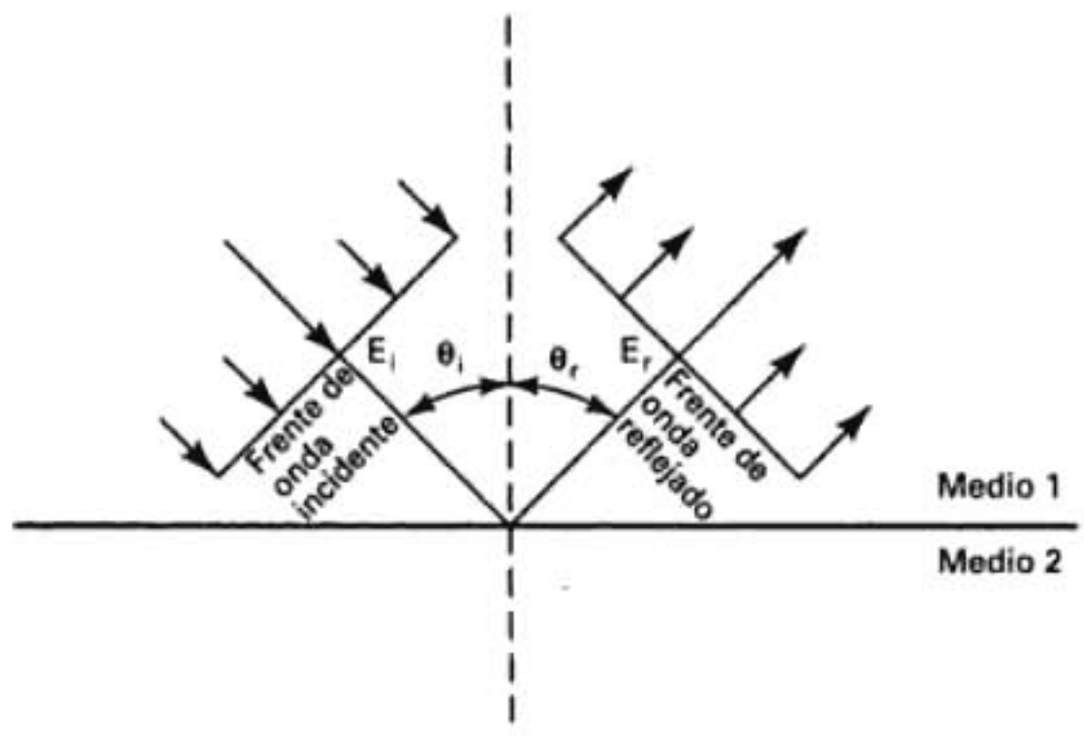


FIGURA 9-7 Reflexión electromagnética en una frontera plana entre dos medios

menor que la del voltaje incidente. La relación de las intensidades de voltaje reflejado a incidente se llama *coeficiente de reflexión*, Γ. Para un conductor perfecto, $\Gamma = 1$. Se usa Γ para indicar tanto la amplitud relativa de los campos incidente y reflejado, como el desplazamiento de fase que hay en el punto de reflexión. La ecuación de este coeficiente es

$$\Gamma = \frac{E_r e^{j\theta_r}}{E_i e^{j\theta_i}} = \frac{E_r}{E_i} = e^{j(\theta_r - \theta_i)} \tag{9-12}$$

en la que Γ = coeficiente de reflexión (adimensional)
E_i = intensidad de voltaje incidente (volts)
E_r = intensidad de voltaje reflejado (volts)
θ_i = fase incidente (grados)
θ_r = fase reflejada (grados)

La relación de las densidades de potencia reflejada a incidente es Γ. La parte de la potencia incidente total que no es reflejada se llama *coeficiente de transmisión de potencia, T*, o simplemente el *coeficiente de transmisión*. Para un conductor perfecto, $T = 0$. La *ley de la conservación de la energía* establece que, para una superficie reflectora perfecta, la potencia total reflejada debe ser igual a la potencia total incidente y, en consecuencia,

$$T + |\Gamma|^2 = 1 \tag{9-13}$$

Para los conductores imperfectos, tanto $|\Gamma|^2$ como T son funciones del ángulo de incidencia, la polarización del campo eléctrico y las constantes dieléctricas de los dos materiales. Si el medio 2 no es conductor perfecto, algunas de las ondas incidentes penetran en él y se absorben. Las ondas absorbidas establecen corrientes a través de la resistencia del material, y la energía se convierte en calor. La fracción de la potencia que penetra al medio 2 se llama *coeficiente de absorción*.

Cuando la superficie reflectora no es plana, sino curva, la curvatura de la onda reflejada es distinta de la de la onda incidente. Cuando el frente de la onda incidente es curvo, y la superficie reflectora es plana, la curvatura del frente de la onda reflejada es igual a la del frente de la onda incidente.

También se produce reflexión cuando la superficie reflectora es *irregular* o *áspera*; sin embargo, esa superficie puede destruir la forma del frente de onda. Cuando un frente de onda incidente choca con una superficie irregular, se dispersa al azar en muchas direcciones. A esa condición se le llama *reflexión difusa*, mientras que a la reflexión de una superficie perfectamente lisa se llama *reflexión especular* (como de espejo). Las superficies que caen entre lisas e irregulares se llaman *superficies semiásperas*. Esas superficies causan una mezcla de reflexión difusa y especular. Una superficie semiáspera no destruye en su totalidad la forma del frente de onda reflejado. Sin embargo, hay una reducción de la potencia total. El *criterio de Rayleigh* establece que una superficie semiáspera refleja como si fuera una superficie lisa siempre que el coseno del ángulo de incidencia sea mayor que $\lambda/8d$, donde d es la profundidad de la irregularidad de la superficie, y λ es la longitud de la onda incidente. La reflexión en una superficie semiáspera se ilustra en la fig. 9-8. La ecuación del criterio de Rayleigh es

$$\cos \theta_i > \frac{\lambda}{8d} \tag{9-14}$$

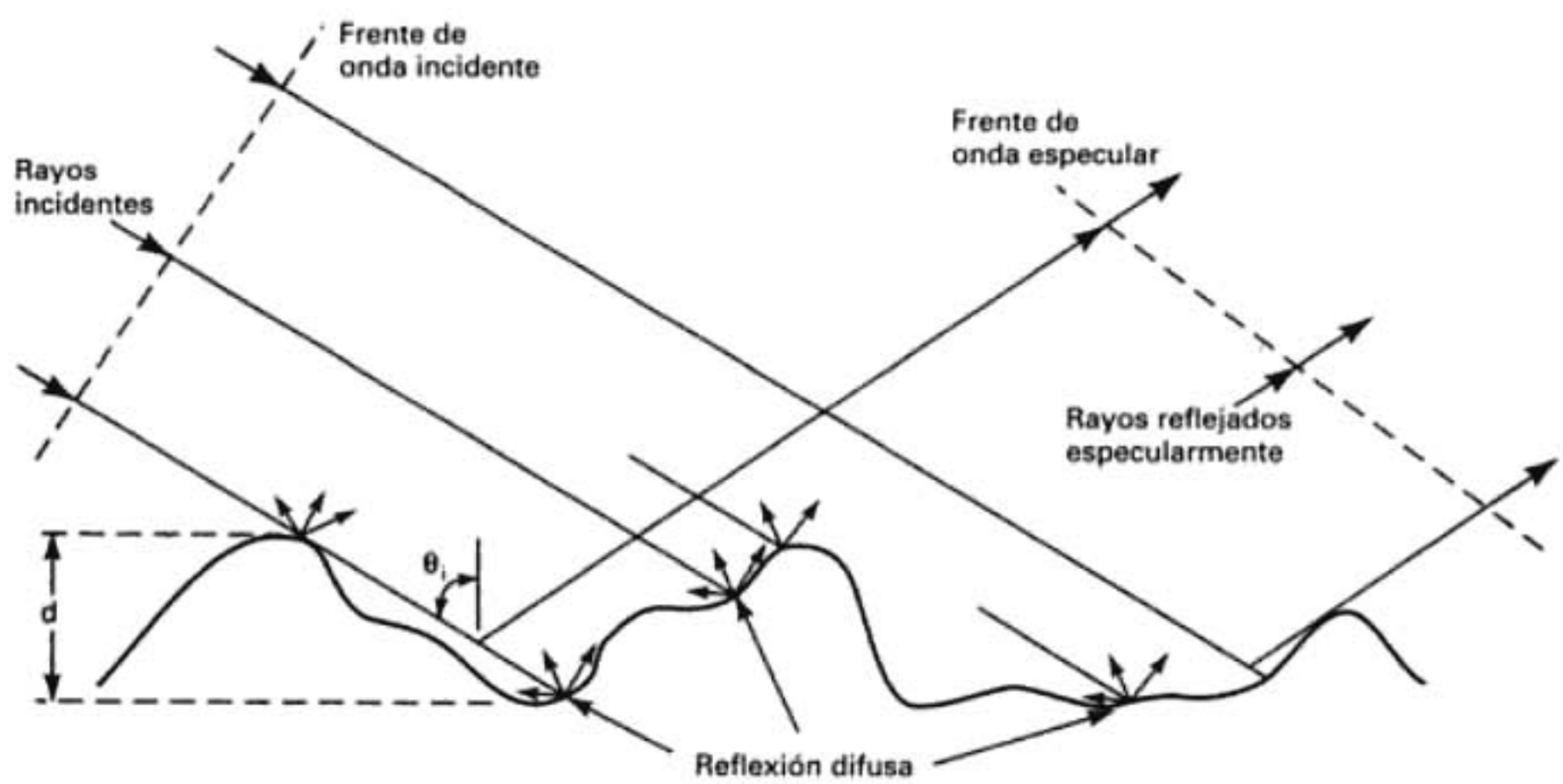


FIGURA 9-8 Reflexión en una superficie semiáspera

Difracción

Se define a la *difracción* como la modulación o redistribución de la energía dentro de un frente de onda, al pasar cerca de la orilla de un objeto *opaco*. La difracción es el fenómeno que permite que las ondas luminosas o de radio se propaguen en torno a esquinas. En la descripción anterior de la refracción y la reflexión se supuso que las dimensiones de las superficies refractora y reflectora eran grandes con respecto a una longitud de onda de la señal. Sin embargo, cuando un frente de onda pasa cerca de un obstáculo o discontinuidad cuyas dimensiones sean de tamaño comparable a una longitud de onda, no se puede usar el análisis geométrico simple para explicar los resultados, y es necesario recurrir al *principio de Huygens*, que se puede deducir de las ecuaciones de Maxwell.

El principio de Huygens establece que todo punto sobre determinado frente de onda esférico se puede considerar como una fuente puntual secundaria de ondas electromagnéticas, desde la cual se irradian y se alejan otras ondas secundarias. El principio de Huygens se ilustra en la fig. 9-9. En la figura 9-9a se muestra la propagación normal de ondas considerando un plano infinito. Cada fuente puntual secundaria (p_1, p_2, etc.) irradia energía hacia afuera, en todas direcciones. Sin embargo, el frente de onda sigue en su dirección original, y no se reparte, porque la anulación de ondas secundarias se hace en todas direcciones, excepto en la de avance. Por lo anterior, el frente de onda permanece plano.

Cuando se considera un frente de onda plano y finito, como en la fig. 9-9b, es incompleta la anulación en direcciones aleatorias. En consecuencia, el frente de onda se reparte hacia afuera, o se *dispersa*. A este efecto de dispersión se le llama *difracción*. La fig. 9-9c muestra la difracción en torno a la orilla de un obstáculo. Se ve que la anulación de ondulaciones sólo es parcial. Se lleva a cabo en torno a la orilla del obstáculo, lo que permite que las ondas secundarias se "escurran" en torno a las aristas de un obstáculo, hacia lo que se llama la *zona de sombra*. Este fenómeno se puede observar cuando se abre la puerta de un cuarto oscuro. Los rayos de luz se difractan en torno a la orilla de la puerta, e iluminan lo que hay detrás de ella.

Interferencia

Interferir quiere decir estar en oposición, y la *interferencia* es el acto de interferir. La interferencia de ondas de radio se produce siempre que se combinan dos o más ondas electromagnéticas de tal manera que se degrada el funcionamiento del sistema. La refracción, la reflexión y la difracción pertenecen a la óptica geométrica, y eso quiere decir que su comportamiento se analiza principalmente en función de rayos y de frentes de onda. Por otro lado, la interferencia está sujeta al principio de la *superposición lineal* de las ondas electromagnéticas, y se presenta siempre que dos o más ondas ocupan el mismo punto del espacio en forma simultánea. El principio de la superposición lineal establece que la intensidad total de voltaje en un punto dado en el espacio es la suma de los vectores de onda individuales. Ciertos tipos de medios de propagación tienen propiedades no lineales; sin embargo, en un medio ordinario, como la atmósfera terrestre, es válida la superposición lineal.

La fig. 9-10 muestra la suma lineal de dos vectores de voltaje instantáneo, cuyos ángulos de fase difieren en el ángulo θ. Se aprecia que el voltaje total no es tan sólo la suma de las dos magnitudes vectoriales, sino más bien la suma fasorial. En la propagación por el espacio libre, puede existir una diferencia de fases sólo porque difieran las *polarizaciones electromagnéticas* de las dos ondas. Según los ángulos de fase de los dos vectores, puede suceder una suma o una resta. Esto implica simplemente que el resultado puede ser mayor o menor que cualquiera de los vectores, porque las dos ondas electromagnéticas se pueden reforzar o se pueden anular.

La fig. 9-11 muestra la interferencia entre dos ondas electromagnéticas en el espacio libre. En el punto X las dos ondas ocupan el mismo lugar en el espacio. Sin embargo, la onda B ha recorrido una trayectoria diferente a la de la onda A y, en consecuencia, sus ángulos de fase relativos pueden ser distintos. Si la diferencia de distancias recorridas es un múltiplo entero impar de la mitad de la longitud de onda. se presenta la anulación total. Si la diferencia es un múltiplo entero y par de la mitad de la longitud de onda, tiene lugar un reforzamiento. Lo más probable es que la diferencia de distancias sea intermedia entre las dos, y se produce una anulación o reforzamiento parcial. Para frecuencias menores que VHF, las longitudes de onda relativamente grandes evitan que la interferencia sea un problema apreciable. Sin embargo, con UHF o más, la interferencia ondulatoria puede ser grave.

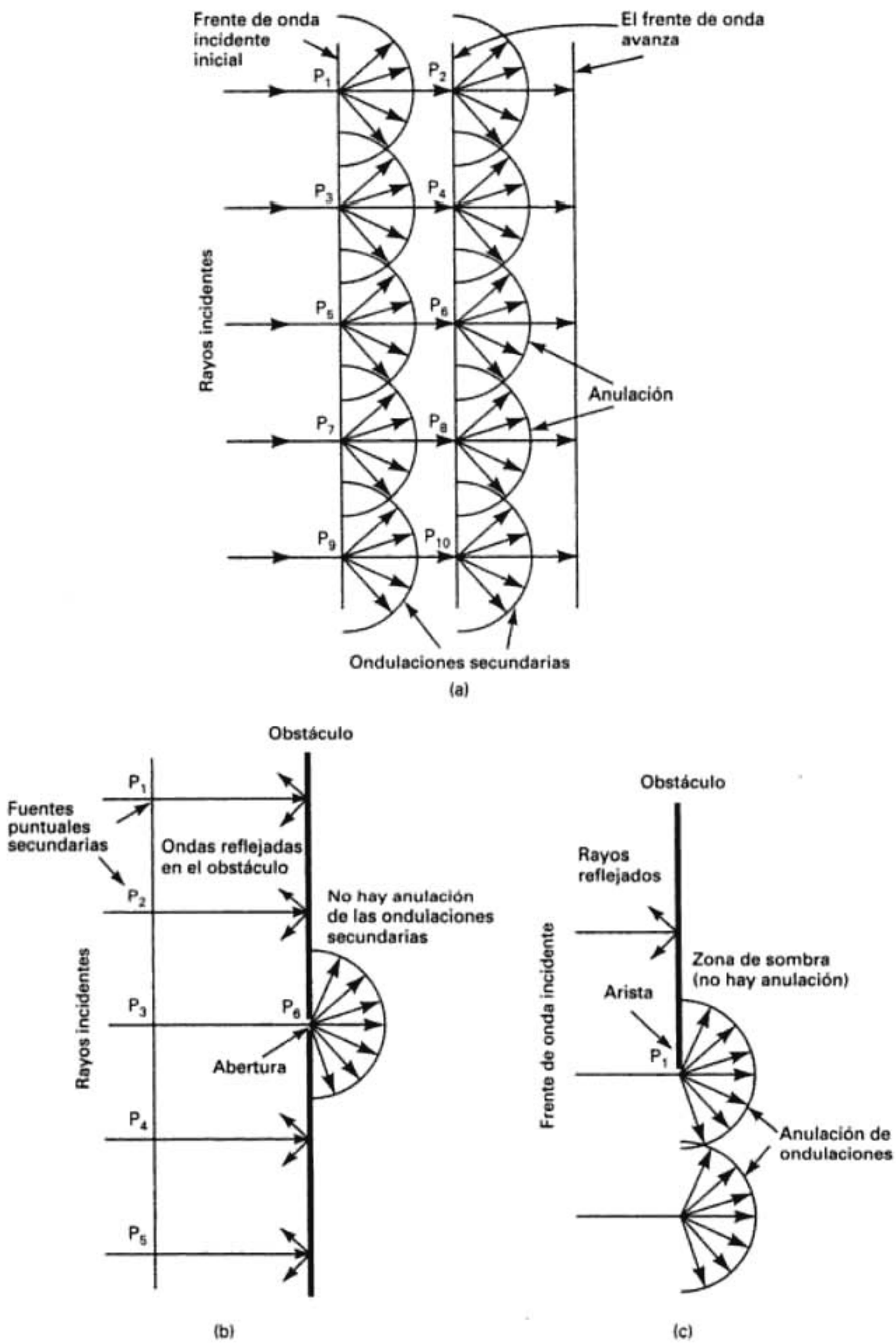


FIGURA 9-9 Difracción de las ondas electromagnéticas: (a) principio de Huygens para un frente de onda plano; (b) frente de una onda finita a través de una abertura; (c) frente de onda rodeando una arista

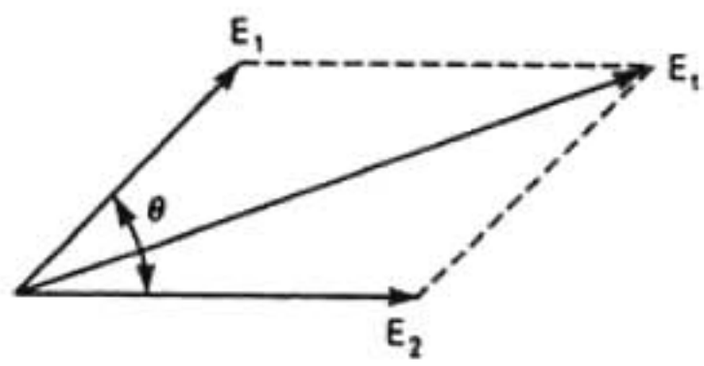


FIGURA 9-10 Suma lineal de dos vectores con distintos ángulos de fase

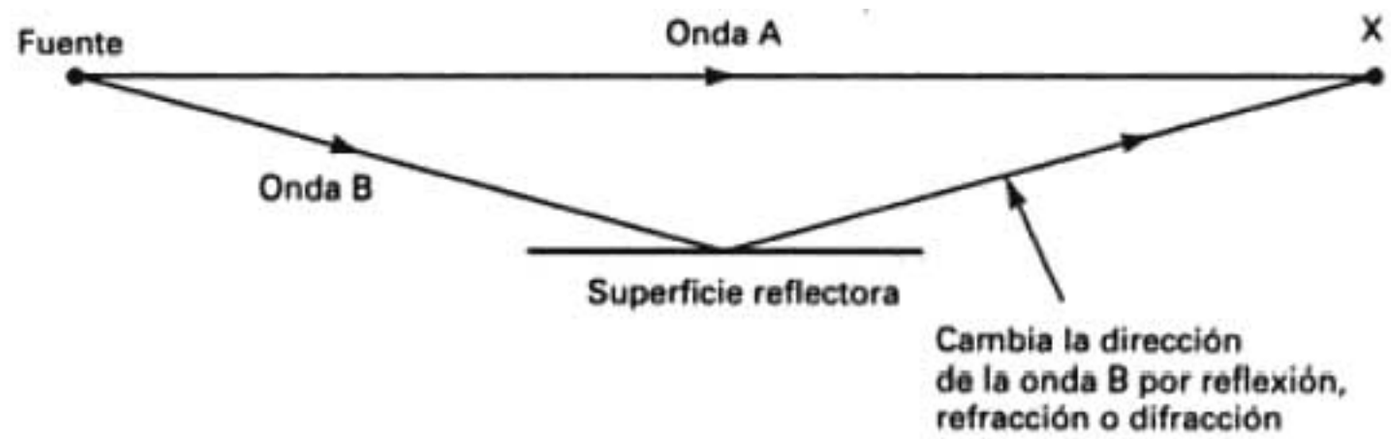


FIGURA 9-11 Interferencia de las ondas electromagnéticas

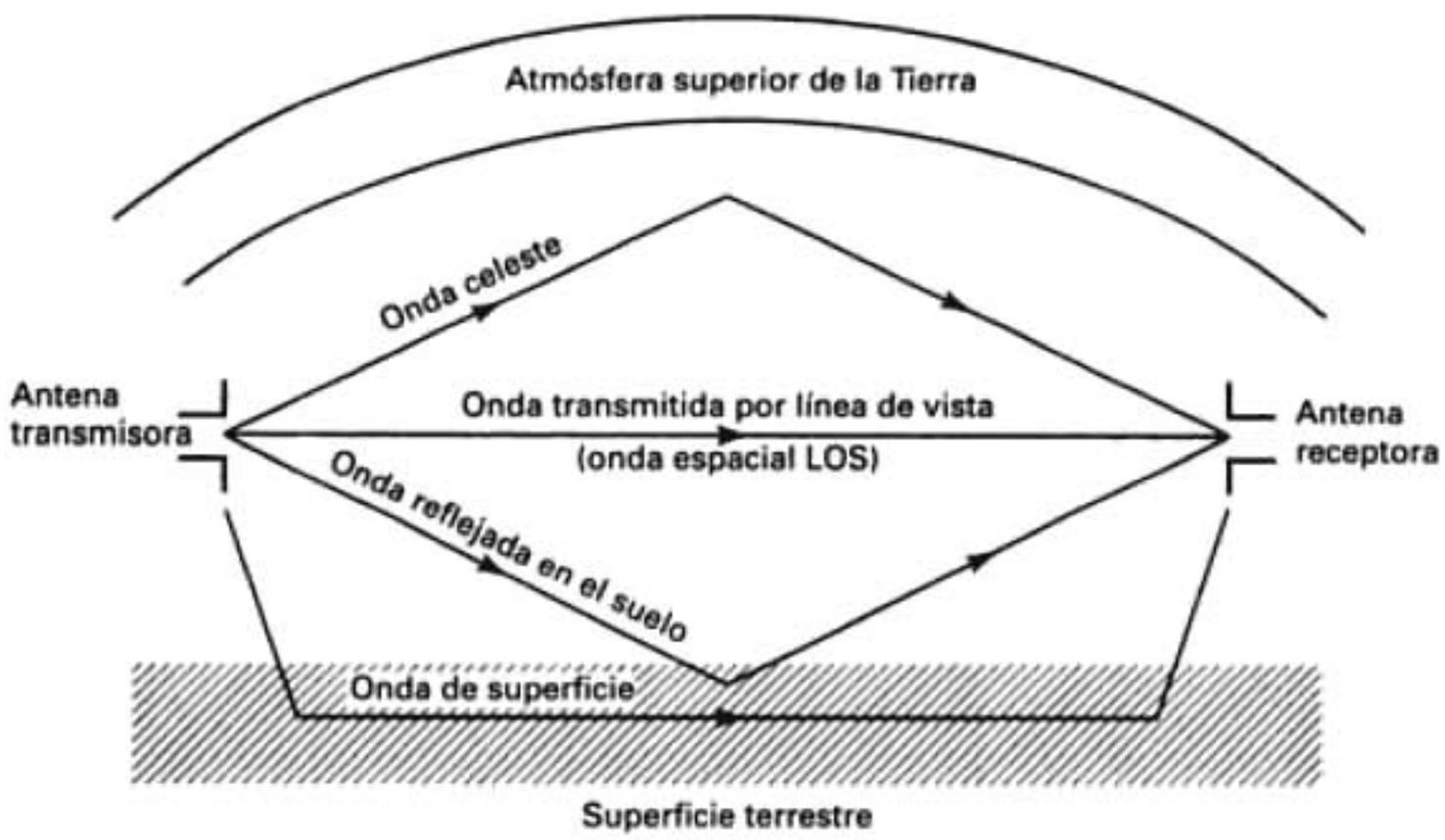


FIGURA 9-12 Modos normales de propagación de ondas

PROPAGACIÓN TERRESTRE DE LAS ONDAS ELECTROMAGNÉTICAS

Las ondas electromagnéticas de radio que viajan dentro de la atmósfera terrestre se llaman *ondas terrestres*, y las comunicaciones entre dos o más puntos de la Tierra se llaman *radiocomunicaciones terrestres*. Las ondas terrestres se ven influidas por la atmósfera y por la Tierra misma. En las radiocomunicaciones terrestres, las ondas se pueden propagar de varias formas, que dependen de la clase del sistema y del ambiente. Como se dijo antes, las ondas electromagnéticas también viajan en línea recta, excepto cuando la Tierra y su atmósfera alteran sus trayectorias. En esencia, hay tres formas de propagación de ondas electromagnéticas dentro de la atmósfera terrestre: onda terrestre, onda espacial (que comprende ondas directas y reflejadas en el suelo) y ondas celestes o ionosféricas.

La fig. 9-12 ilustra los modos normales de propagación entre dos antenas de radio. En todo sistema de radio existen los tres modos, sin embargo, algunos son despreciables en ciertos intervalos de frecuencia, o sobre determinada clase de terreno. A frecuencias menores que 1.5 MHz, las ondas terrestres tienen la mejor difusión, porque las pérdidas en el suelo aumentan con rapidez al aumentar la frecuencia. Las ondas celestes se usan para aplicaciones de alta frecuencia, y las ondas espaciales se usan para frecuencias muy elevadas.

Propagación de ondas terrestres

Una *onda terrestre* es una onda electromagnética que viaja por la superficie de la Tierra. Por eso a las ondas terrestres también se les llama *ondas superficiales*. Las ondas terrestres deben estar polarizadas verticalmente. Esto se debe a que el campo eléctrico, en una onda polarizada horizontalmente, sería paralelo a la superficie de la tierra, y esas ondas se pondrían en corto por la conductividad del suelo. Con las ondas terrestres, el campo eléctrico variable induce voltajes en

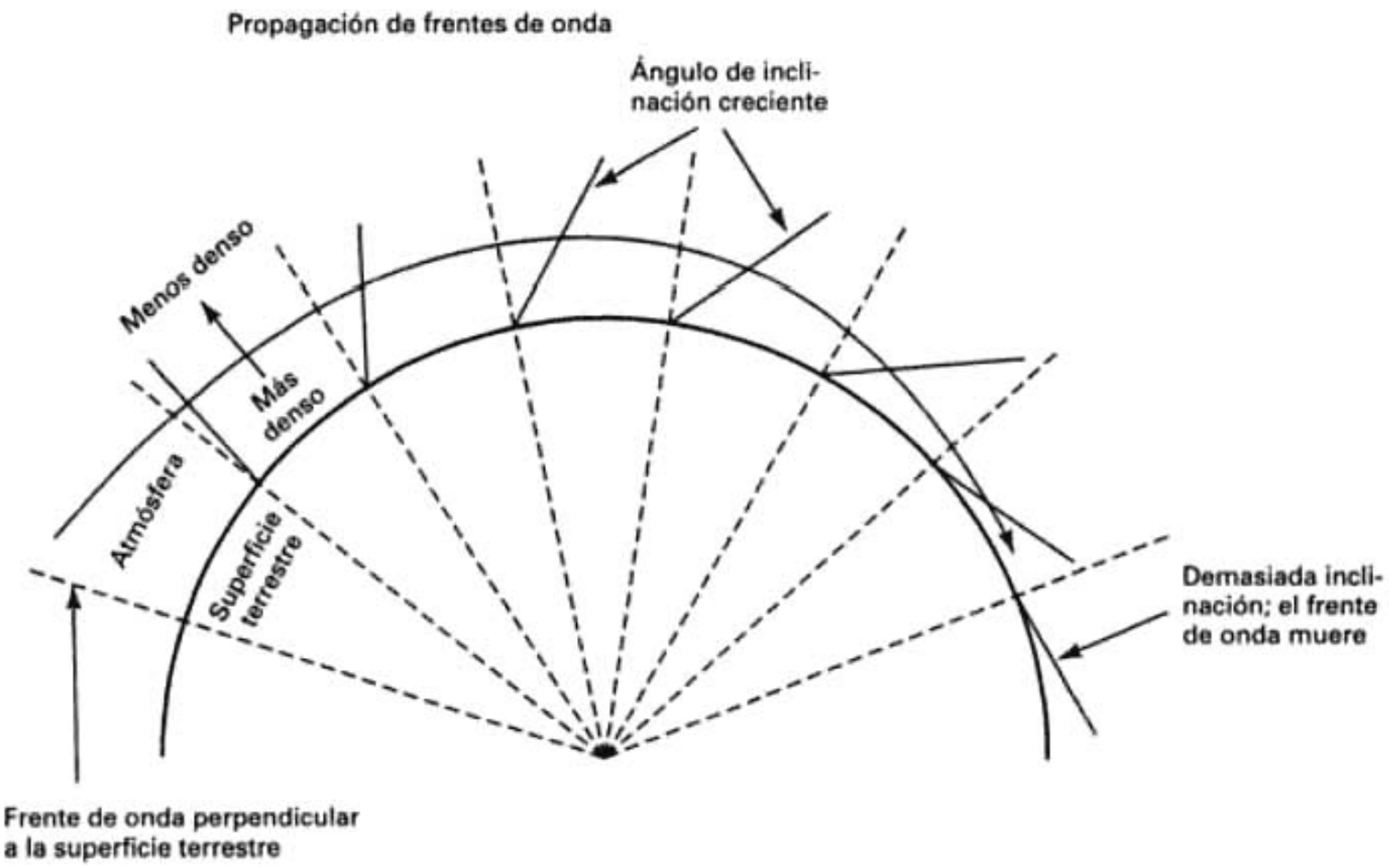


FIGURA 9-13 Propagación de ondas terrestres

la superficie terrestre, que hacen circular corrientes muy parecidas a las de una línea de transmisión. La superficie terrestre también tiene pérdidas por resistencia y por dieléctrico. Por consiguiente, las ondas terrestres se atenúan a medida que se propagan. Se propagan mejor sobre una superficie buena conductora, como por ejemplo, agua salada, y se propagan mal sobre superficies desérticas. Las pérdidas en las ondas terrestres aumentan rápidamente al aumentar la frecuencia. Por consiguiente, su propagación se limita en general a frecuencias menores que 2 MHz.

La fig. 9-13 representa la propagación de las ondas terrestres. La atmósfera terrestre tiene un *gradiente de densidad*, es decir, la densidad disminuye en forma gradual conforme aumenta la distancia a la superficie de la Tierra; esto hace que el frente de onda se incline en forma progresiva hacia adelante. Así, la onda terrestre se propaga en torno a la Tierra y queda cerca de su superficie; si se transmite la potencia suficiente, el frente de onda se podría propagar más allá del horizonte, o hasta por toda la circunferencia de la Tierra. Sin embargo, se debe tener cuidado al seleccionar la frecuencia y el terreno sobre el que se va a propagar la onda terrestre, para asegurarse de que no se incline demasiado y se "caiga", de plano sobre el suelo, cesando de propagarse.

La propagación por ondas terrestres se usa normalmente en comunicaciones entre barcos y de barco a tierra, para la radionavegación y para comunicaciones marítimas móviles. Las ondas terrestres tienen frecuencias tan bajas como 15 kHz.

Las desventajas de la propagación de ondas terrestres son las siguientes:

1. Requieren una potencia de transmisión relativamente alta.
2. Se limitan a frecuencias muy bajas, bajas e intermedias (VLF, LF y MF) y requieren grandes antenas. La razón de esto se explica en el capítulo 11.
3. Las pérdidas en el terreno varían mucho de acuerdo con el material superficial y su composición.

Las ventajas de la propagación de ondas terrestres son las siguientes:

1. Con la potencia suficiente de transmisión, se pueden usar las ondas terrestres para comunicarse entre dos lugares cualesquiera en el mundo.
2. Las ondas terrestres se afectan poco por las condiciones variables de la atmósfera.

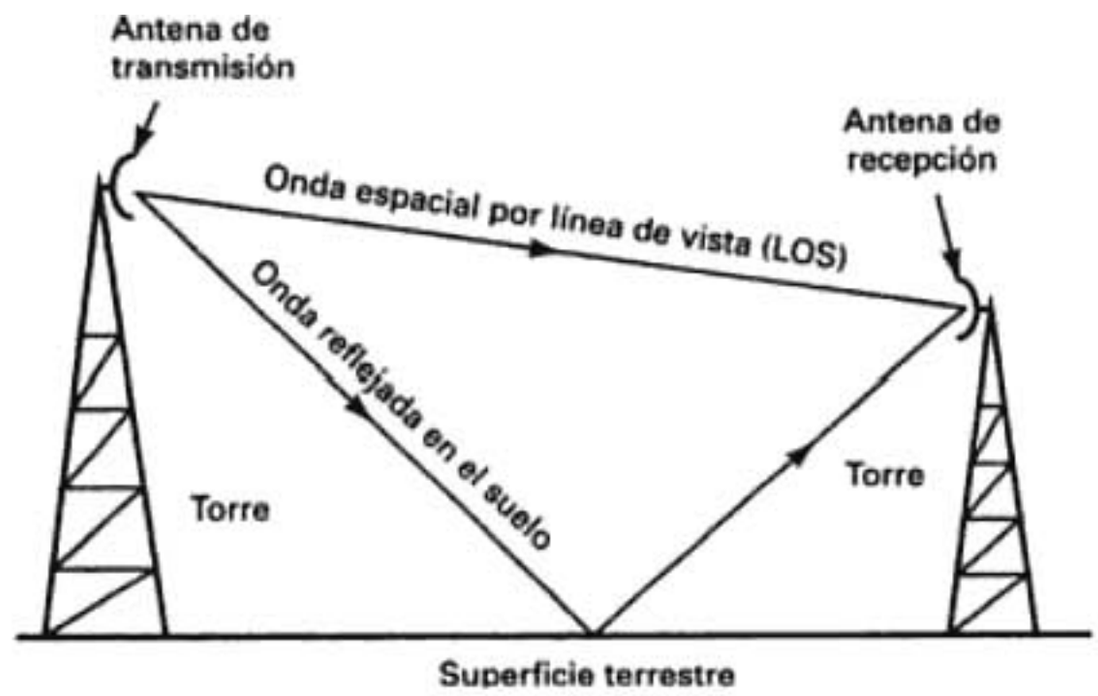


FIGURA 9-14 Propagación de ondas espaciales

Propagación de las ondas espaciales

La propagación de la energía electromagnética en forma de ondas espaciales incluye la energía irradiada que viaja en los kilómetros inferiores de la atmósfera terrestre. Las ondas espaciales incluyen ondas directas y las reflejadas en el suelo (véase la fig. 9-14). Las *ondas directas* viajan esencialmente en línea recta entre las antenas de transmisión y recepción. La propagación de ondas espaciales directas se llama *transmisión por línea de vista* (LOS, por *line-of-sight*). Por consiguiente, la propagación directa de ondas espaciales está limitada por la curvatura de la Tierra. Las ondas reflejadas en el suelo son las que refleja la superficie terrestre cuando se propagan entre las antenas emisora y receptora.

En la fig. 9-14 se ve la propagación de una onda espacial entre dos antenas. Se aprecia que la intensidad de campo en la antena receptora depende de la distancia entre las dos antenas (atenuación y absorción) y de si las ondas directas y las reflejadas en el suelo están en fase (interferencia).

La curvatura de la Tierra presenta un horizonte en la propagación de las ondas espaciales, que se suele llamar el *horizonte de radio*. A causa de la refracción atmosférica, el horizonte de radio está más allá del *horizonte óptico* para la *atmósfera estándar* común. El horizonte de radio está, más o menos, a cuatro tercios del horizonte óptico. La refracción se debe a la troposfera, a cambios en su densidad, temperatura, contenido de vapor de agua y conductividad relativa. El horizonte de radio se puede alargar sólo con elevar las antenas de transmisión o recepción, o ambas, respecto a la superficie terrestre, con torres, o colocándolas sobre montañas o edificios altos.

La fig. 9-15 muestra el efecto de la altura de la antena sobre el horizonte de radio. El horizonte visual de radio para una sola antena es

$$d = \sqrt{2}\,h \tag{9-15}$$

siendo d = distancia al horizonte de radio (millas)
h = altura de la antena sobre el nivel del mar (millas)

Por consiguiente, para una antena de transmisión y una de recepción, la distancia máxima entre ellas es

$$d = d_t + d_r$$

$$d = \sqrt{2}h_t + \sqrt{2}h_r \tag{9-16}$$

donde d = distancia total (millas)
d_t = horizonte de radio de la antena transmisora (millas)
d_r = horizonte de radio de la antena receptora (millas)
h_t = altura de la antena transmisora (pies)
h_r = altura de la antena receptora (pies)

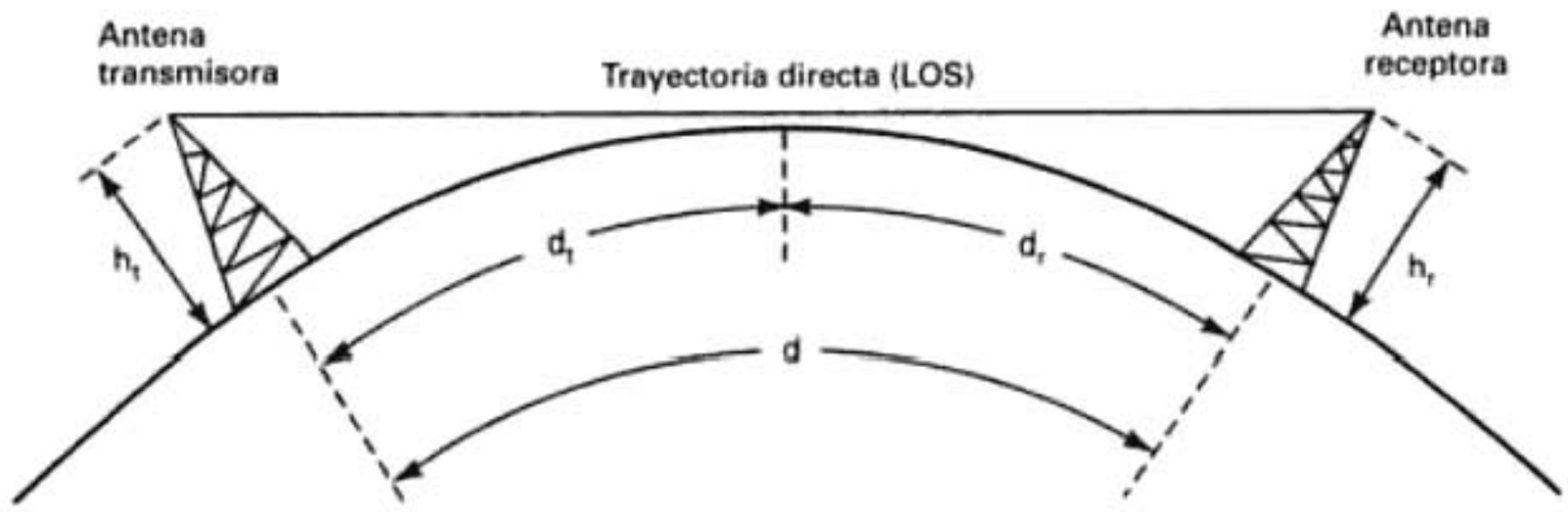


FIGURA 9-15 Ondas espaciales y horizonte de radio

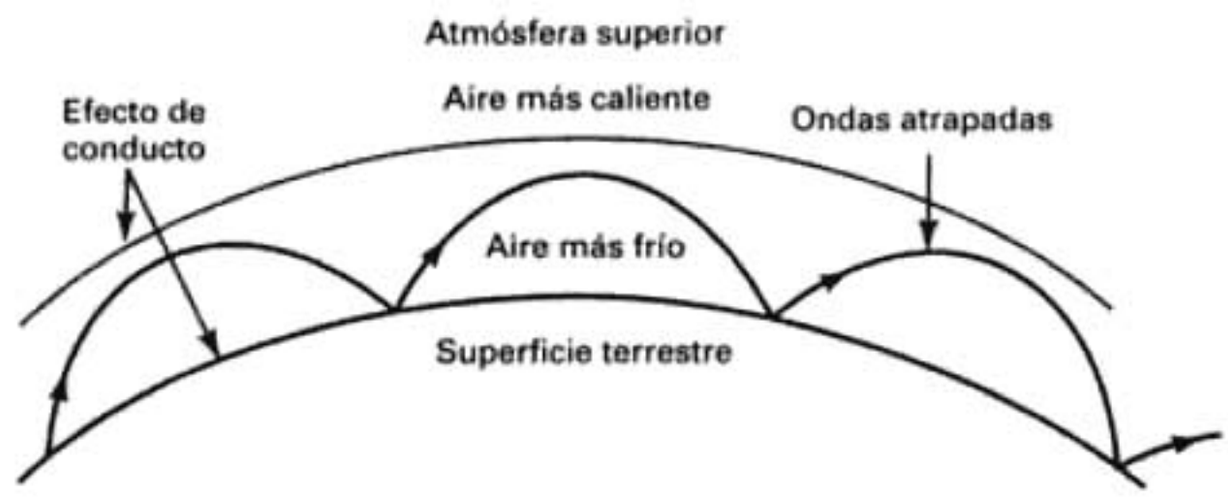


FIGURA 9-16 Propagación por conductos atmosféricos

La distancia máxima entre un transmisor y un receptor sobre terreno promedio se puede aproximar en unidades métricas como sigue

$$d(\text{máx}) = 17h_t + 17h_r \tag{9-17}$$

siendo $d(\text{máx})$ = distancia máxima entre el transmisor y el receptor (kilómetros)
h_t = altura de la antena transmisora (metros)
h_r = altura de la antena receptoras (metros)

De acuerdo con las ecuaciones 9-16 y 9-17, la distancia de propagación de las ondas espaciales se puede incrementar aumentando la altura de la antena transmisora o de la antena receptora, o de ambas.

Como las condiciones de la atmósfera terrestre inferior cambian, el grado de refracción puede variar a través del tiempo. Se tiene una condición especial, llamada *propagación por conductos* cuando la densidad de la atmósfera inferior es tal que las ondas electromagnéticas quedan atrapadas entre ella y la superficie terrestre. Las capas atmosféricas funcionan como un conducto, y una onda electromagnética se puede propagar grandes distancias siguiendo la curvatura de la Tierra dentro de este conducto. En la fig. 9-16 se representa la propagación por conducto.

Propagación por ondas celestes

Las ondas electromagnéticas que se dirigen sobre el nivel del horizonte se llaman *ondas celestes*. En el caso normal, las ondas celestes se irradian en una dirección que forma un ángulo relativamente grande con la Tierra. Se irradian hacia el cielo, donde son reflejadas o refractadas hacia la superficie terrestre por la ionosfera. Debido a lo anterior, a la propagación de las ondas celestes se le llama a veces propagación ionosférica. La ionosfera es la región del espacio que está entre 50 y 400 km (30 a 250 mi) sobre la superficie terrestre. Es la parte superior de la atmósfera terrestre. Por su situación, absorbe grandes cantidades de la energía solar radiante, que ioniza las moléculas de aire y forma electrones libres. Cuando una onda de radio atraviesa la

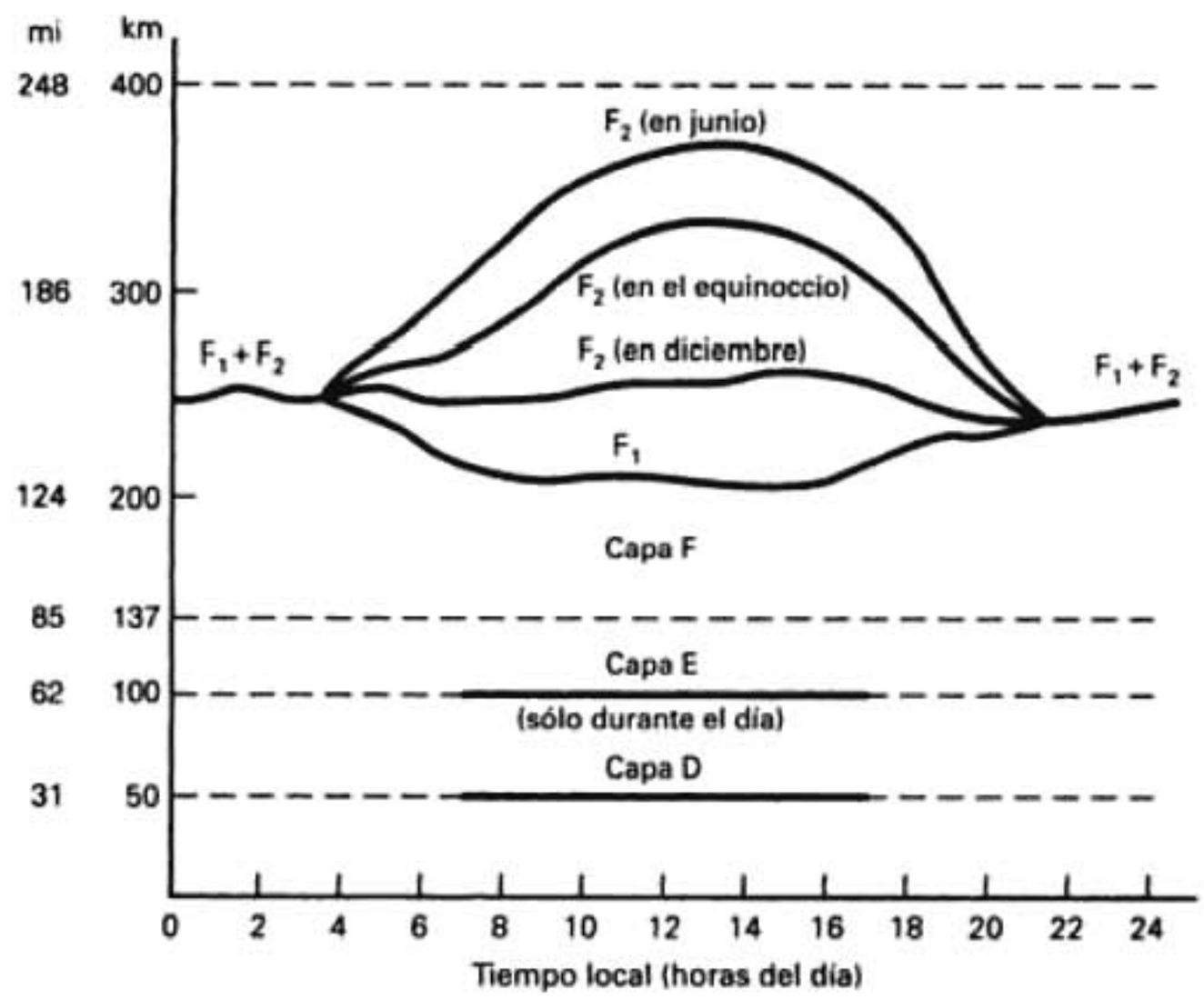


FIGURA 9-17 Capas ionosféricas

ionosfera, el campo eléctrico de la onda ejerce una fuerza sobre los electrones libres y los pone a vibrar. Los electrones en vibración disminuyen la corriente, lo que equivale a reducir la constante dieléctrica. Al reducir la constante dieléctrica aumenta la velocidad de propagación y las ondas electromagnéticas se desvían y se alejan de las regiones de alta densidad de electrones hacia regiones de baja densidad de electrones; es decir, aumenta la refracción. Al alejarse la onda de la Tierra aumenta la ionización; sin embargo, hay menos moléculas de aire que se ionizan. Por consiguiente, la atmósfera superior tiene mayor porcentaje de moléculas ionizadas que la atmósfera inferior. Mientras mayor sea la densidad de iones, la refracción es mayor. También, debido a la composición no uniforme de la ionosfera y a sus variaciones de temperatura y de densidad, está *estratificada*. En esencia son tres las capas que forman la ionosfera: las capas D, E y F, que se ven en la fig. 9-17. Las tres capas de ionosfera varían en localización y en *densidad de ionización* según la hora del día. También fluctúan en una forma cíclica durante el año, y también siguiendo el *ciclo de manchas solares* de 11 años. La ionosfera es más densa durante las horas de más luz solar: durante las horas del día y en el verano.

Capa D. La *capa D* es la más inferior de la ionosfera, entre 30 y 60 mi (50 a 100 km) sobre la superficie de la Tierra. Como es la capa más alejada del Sol, tiene poca ionización. En consecuencia, la capa D tiene muy poco efecto sobre la dirección de propagación de las ondas de radio. Sin embargo, los iones de la capa D pueden absorber cantidades apreciables de energía electromagnética. La cantidad de ionización en la capa D depende de la altura del Sol sobre el horizonte. En consecuencia, desaparece durante la noche. La capa D refleja las ondas de VLF y LF, y absorbe las ondas MF y HF (véase la tabla 1-1, con las regiones de frecuencias VLF, LF, MF y HF).

Capa E. La *capa E* se ubica entre las 60 y las 85 millas (100 a 140 km) sobre la superficie terrestre. A veces se le llama *capa Kennelly-Heaviside* en honor de los dos científicos que la descubrieron. Tiene su densidad máxima a unas 70 millas a mediodía, cuando el Sol está en su altura máxima. Como en la capa D, la capa E desaparece casi totalmente por la noche. La capa E favorece la propagación de ondas superficiales de MF y refleja algo las ondas HF durante el día. La parte superior de la capa E se considera a veces por separado, y se llama capa E esporádica, porque parece llegar e irse en forma impredecible. La capa E esporádica se produce

durante los *destellos solares* y durante la *actividad de manchas solares*. Es una capa delgada con una densidad de ionización muy alta. Cuando aparece, produce una mejoría inesperada en la radiotransmisión a grandes distancias.

Capa F. La *capa F* está formada en realidad por dos capas: la F_1 y la F_2. Durante el día, la capa F_1 está entre 85 y 155 mi (140 a 250 km) sobre la superficie terrestre. La capa F_2 está de 85 a 185 mi (140 a 300 km) sobre la superficie terrestre durante el invierno, y de 155 a 220 mi (250 a 350 km) en el verano. Durante la noche, las capas F_1 y F_2 se combinan y forman una sola capa. La capa F_1 absorbe y atenúa algunas ondas de HF, aunque la mayoría de las ondas atraviesan hasta la capa F_2, donde son reflejadas hacia la Tierra.

TÉRMINOS Y DEFINICIONES DE PROPAGACIÓN

Frecuencia crítica y ángulo crítico

Virtualmente, la ionosfera no afecta a las frecuencias mayores que las UHF, porque las ondas son extremadamente cortas. Las distancias entre iones son bastante mayores que las longitudes de onda de estas frecuencias, y en consecuencia, las ondas electromagnéticas pasan a través de ellos con pocos efectos notables. Por consiguiente es razonable que haya un límite superior de frecuencias que se propaguen como ondas celestes. La *frecuencia crítica* (f_c) se define como la máxima frecuencia que se puede propagar directo hacia arriba y es reflejada por la ionosfera hacia la Tierra. La frecuencia crítica depende de la densidad de ionización y, en consecuencia, varía con la hora del día y con la estación. Si disminuye el ángulo vertical de irradiación, las frecuencias iguales o mayores que la crítica se pueden reflejar hacia la superficie terrestre, porque recorren una distancia mayor en la ionosfera y por consiguiente tienen mayor tiempo de refractarse. Así, la frecuencia crítica sólo se usa como punto de referencia para fines de comparación. Sin embargo, cada frecuencia tiene un ángulo vertical máximo al cual se puede propagar y seguir reflejándose por la ionosfera. Ese ángulo se llama *ángulo crítico*. El ángulo crítico, θ_c, se ilustra en la fig. 9-18.

A veces se usa una técnica de medición llamada *sondeo ionosférico* para determinar la frecuencia crítica. Una señal se propaga directo hacia arriba, desde la superficie de la Tierra, y su frecuencia se aumenta en forma gradual. A las frecuencias inferiores, la señal se absorberá por completo en la atmósfera. Sin embargo, a medida que se aumenta la frecuencia, la señal, o una

FIGURA 9-18 Ángulo crítico

parte de ella, regresará a la Tierra. Pero en determinada frecuencia la señal pasará por la atmósfera terrestre, hacia el espacio exterior, y no regresará hacia la Tierra. La máxima frecuencia que regresa a la Tierra en dirección vertical es la frecuencia crítica.

Altura virtual

La *altura virtual* es la altura, sobre la superficie terrestre, desde la que parece reflejarse una onda refractada. La fig. 9-19 muestra una onda que se ha irradiado desde la superficie terrestre hacia la ionosfera. La onda irradiada se refracta y regresa a la tierra, describiendo la trayectoria B. La altura máxima real que alcanzó la onda es h_a. Sin embargo, la trayectoria A muestra la trayectoria proyectada que podría haber tomado la onda reflejada y ser regresada a la Tierra hacia el mismo lugar. La altura máxima que habría alcanzado esta onda reflejada hipotética es la altura virtual (h_v).

Máxima frecuencia útil

La *máxima frecuencia útil* (MUF, de *maximum usable frequency*) es la mayor frecuencia que se puede usar en propagación de ondas celestes entre dos puntos específicos de la superficie terrestre. Es razonable, entonces, que haya tantos valores posibles de MUF como puntos y frecuencias hay en la Tierra: una cantidad infinita. La MUF, como la frecuencia crítica, es una frecuencia límite para la propagación de las ondas celestes. Sin embargo, la MUF es para determinado ángulo de incidencia: el ángulo que forma la onda incidente con la normal. La ecuación correspondiente es

$$\text{MUF} = \frac{\text{frecuencia crítica}}{\cos\theta} \tag{9-18a}$$

$$= \text{frecuencia crítica} \times \sec\theta \tag{9-18b}$$

donde θ es el ángulo de incidencia.

La ecuación 9-18a se llama *ley de la secante*. Supone que la Tierra es plana y que la capa reflectora es plana, lo cual, naturalmente, no puede ser. Por esta razón la MUF sólo se usa para hacer cálculos preliminares.

Debido a la inestabilidad general de la ionosfera, la máxima frecuencia que se usa entre dos puntos se suele seleccionar menor que la MUF. Se ha demostrado que si se trabaja con una frecuencia igual al 85% de la MUF se obtienen comunicaciones más confiables. A veces, a esta frecuencia se le llama *frecuencia óptima de trabajo* (OWF, de *optimum working frequency*).

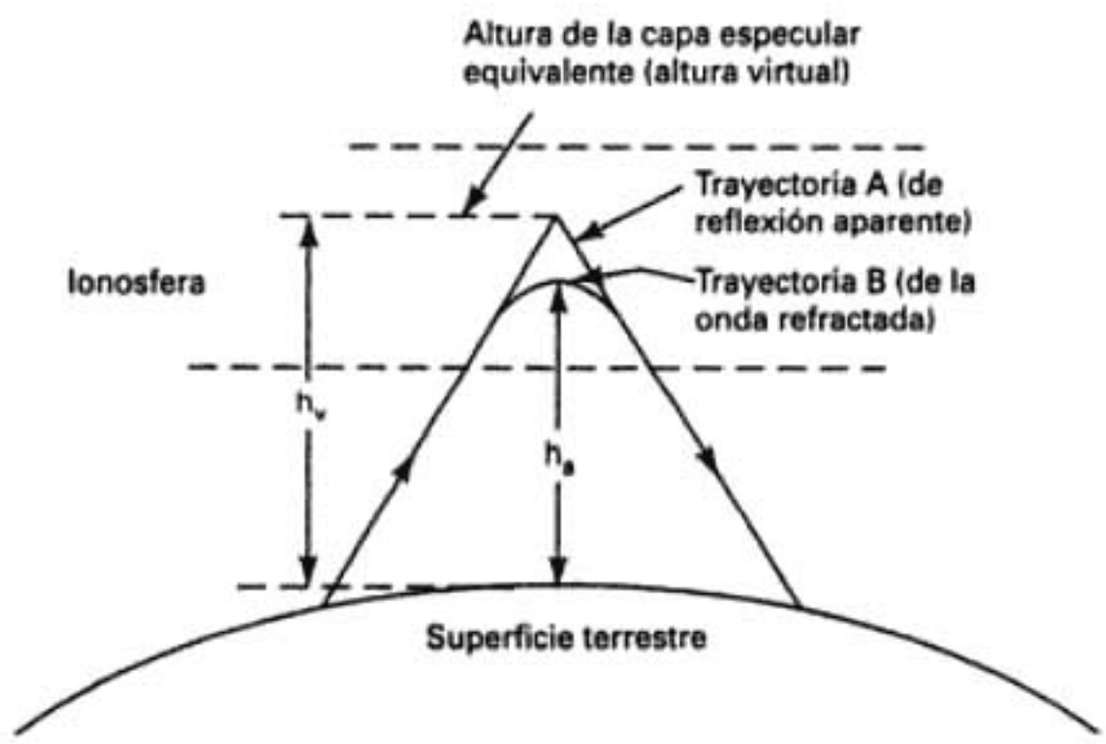


FIGURA 9-19 Altura virtual y altura real

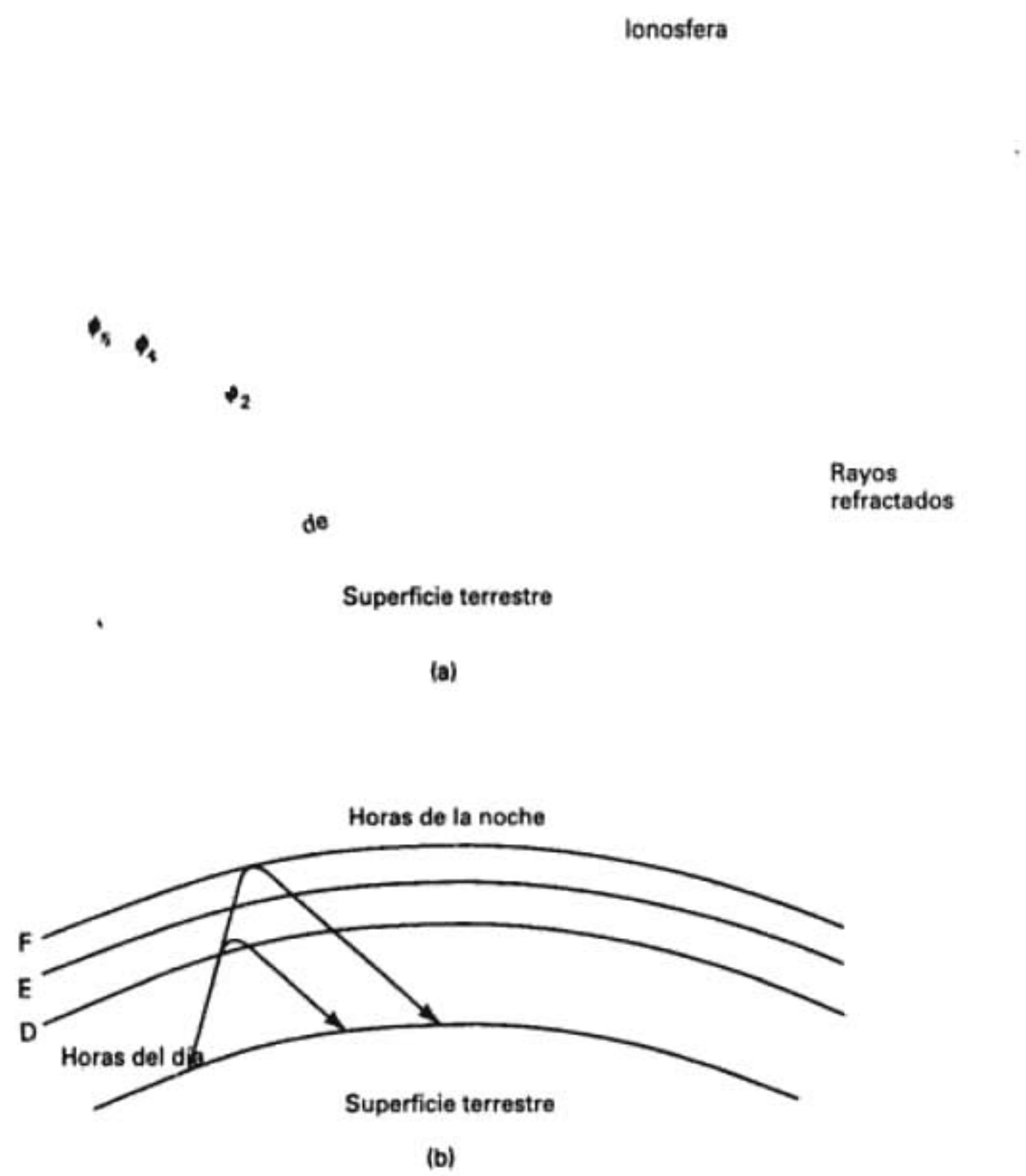


FIGURA 9-20 (a) Distancia de salto; (b) propagación durante el día y durante la noche

Distancia de salto

La *distancia de salto*, d_s, es la distancia mínima desde una antena de transmisión a la que regresará a la Tierra una onda celeste de determinada frecuencia (que debe ser menor que la MUF). La fig. 9-20a muestra varios rayos con distintos ángulos de elevación, irradiados desde el mismo punto sobre la Tierra. Se aprecia que el punto en el que regresa la onda a la Tierra se acerca al transmisor a medida que aumenta ϕ, el ángulo de elevación. Sin embargo, al final el ángulo de elevación es tan alto que la onda penetra y atraviesa la ionosfera, y escapa de la atmósfera terrestre en forma total.

La fig. 20-9b muestra el efecto de la desaparición de las capas D y E durante la noche, sobre la distancia de salto. En forma efectiva, el *cielo* que forma la ionosfera se eleva y permite a las ondas celestes viajar más arriba antes de ser refractadas hacia la Tierra. Este efecto explica por qué a veces las estaciones lejanas de radio se oyen durante la noche, pero no durante las horas del día.

PÉRDIDAS EN TRAYECTORIA POR EL ESPACIO LIBRE

La pérdida *en trayectoria por el espacio libre* se suele definir como la pérdida sufrida por una onda electromagnética al propagarse en línea recta por un vacío, sin absorción ni reflexión de energía en objetos cercanos. Es una definición mala y con frecuencia engañosa. La pérdida en trayectoria por el espacio libre es una cantidad técnica artificial que se originó debido a la manipulación de las ecuaciones de presupuesto de un enlace de comunicaciones, que deben tener

determinado formato en el que se incluye la ganancia de la antena transmisora, la pérdida en trayectoria por el espacio libre y el área efectiva de la antena receptora. En realidad no se pierde energía alguna; tan sólo se reparte al propagarse alejándose de la fuente, y se produce una menor densidad de potencia en determinado punto a determinada distancia de la fuente. En consecuencia, un término más adecuado para definir el fenómeno es *pérdida por dispersión*. La pérdida por dispersión se debe simplemente a la ley del cuadrado inverso. La ecuación que define a la pérdida en trayectoria por el espacio libre es

$$L_p = \left(\frac{4\pi D}{\lambda}\right)^2 = \left(\frac{4\pi Df}{c}\right)^2 \tag{9-19}$$

siendo L_p = pérdida en trayectoria por el espacio libre (adimensional)
D = distancia (kilómetros)
f = frecuencia (hertz)
λ = longitud de onda (metros)
c = velocidad de la luz en el espacio libre (3×10^8 metros por segundo)

Expresada en decibeles, esta ecuación es

$$L_{p(\mathrm{dB})} = 20 \log \frac{4\pi fD}{c} = 20 \log \frac{4\pi}{c} + 20 \log f + 20 \log D$$

Cuando la frecuencia se expresa en MHz y la distancia en km

$$L_{p(\mathrm{dB})} = 20 \log \frac{4\pi(10)^6(10)^3}{3 \times 10^8} + 20 \log f_{(\mathrm{MHz})} + 20 \log D_{(\mathrm{km})} \tag{9-20a}$$

$$= 32.4 + 20 \log f_{(\mathrm{MHz})} + 20 \log D_{(\mathrm{km})}$$

Cuando la frecuencia se expresa en GHz y la distancia en km

$$L_{p(\mathrm{dB})} = 92.4 + 20 \log f_{(\mathrm{GHz})} + 20 \log D_{(\mathrm{km})} \tag{9-20b}$$

Se pueden hacer conversiones parecidas cuando la distancia está en millas, la frecuencia en kHz, etc.

Ejemplo 9-2

Calcular la pérdida en trayectoria por el espacio libre para una frecuencia de portadora de 6 GHz y una distancia de 50 km.

Solución

$$L_p = 32.4 + 20 \log 6000 + 20 \log 50$$
$$= 32.4 + 75.6 + 34 = 142 \text{ dB}$$
$$L_p = 92.4 + 20 \log 6 + 20 \log 50$$
$$= 92.4 + 15.6 + 34 = 142 \text{ dB}$$

MARGEN DE DESVANECIMIENTO

Las radiocomunicaciones entre lugares remotos, sean de tierra a tierra o de tierra a satélite, requieren la propagación de señales electromagnéticas por el espacio libre. Al propagarse una onda electromagnética por la atmósfera terrestre, la señal puede tener pérdidas intermitentes de intensidad, además de la pérdida normal en la trayectoria. Esas pérdidas de pueden atribuir a diversos fenómenos, que incluyen efectos de corto y de largo plazo. Esta variación en la pérdida de la señal se llama *desvanecimiento* y se puede atribuir a perturbaciones meteorológicas como lluvia, nieve, granizo, etc.; a trayectorias múltiples de transmisión y a una superficie terrestre irregular. Para tener en cuenta el desvanecimiento temporal, se agrega una pérdida adicional de transmisión a la pérdida en trayectoria normal. A esta pérdida se le llama *margen de desvanecimiento*.

En esencia, el margen de desvanecimiento es un "factor espurio" que se incluye en la ecuación de ganancia del sistema para considerar las características no ideales y menos predecibles de la propagación de las ondas de radio, como por ejemplo la *propagación por trayectorias*

múltiples (*pérdida por trayectorias múltiples*) y la *sensibilidad del terreno*. Estas características causan condiciones atmosféricas temporales y anormales que alteran la pérdida por trayectoria en el espacio libre, y suelen ser perjudiciales para la eficiencia general del sistema. El margen de desvanecimiento también tiene en cuenta los objetivos de confiabilidad del sistema. Así, el margen de desvanecimiento se incluye en la ecuación de ganancia de un sistema como una pérdida.

Al resolver las ecuaciones de confiabilidad de Barnett-Vignant para una disponibilidad anual especificada en un sistema no protegido sin diversidad se obtiene la siguiente ecuación

$$F_m = \underbrace{30 \log D}_{\text{Efecto de trayectoria múltiple}} + \underbrace{10 \log (6ABf)}_{\text{Sensiblidad del terreno}} - \underbrace{10 \log (1 - R)}_{\text{Objetivos de confiabilidad}} - \underbrace{70}_{\text{constante}} \tag{9-21}$$

siendo F_m = margen de desvanecimiento (decibeles)
D = distancia (kilómetros)
f = frecuencia (gigahertz)
R = confiabilidad en tanto por uno (es decir, 99.99% = 0.9999 de confiabilidad)
$1 - R$ = objetivo de confiabilidad para una ruta de 400 km en un sentido
A = factor de rugosidad
= 4 sobre agua o sobre un terreno muy liso
= 1 sobre un terreno promedio
= 0.25 sobre un terreno muy áspero y montañoso
B = factor para convertir la peor probabilidad mensual en una probabilidad anual
= 1 para pasar una disponibilidad anual a la peor base mensual
= 0.5 para áreas calientes y húmedas
= 0.25 para áreas continentales promedio
= 0.125 para áreas muy secas o montañosas

Ejemplo 9-3

Calcular el margen de desvanecimiento para las siguientes condiciones: distancia entre sitios $D = 40$ km; frecuencia $f = 1.8$ GHz; terreno liso; clima húmedo y objetivo de confiabilidad 99.99%.

Solución Se sustituyen valores en la ecuación 9-21, para obtener

$$F_m = 30 \log 40 + 10 \log[(6)(4)(0.5)(1.8)] - 10 \log(1 - 0.9999) - 70$$
$$= 48.06 + 13.34 - (-40) - 70 = 31.4 \text{ dB}$$

PREGUNTAS

9-1. Describa un rayo electromagnético y un frente de onda.
9-2. Describa lo que es densidad de potencia e intensidad de voltaje.
9-3. Describa un frente de onda esférico.
9-4. Explique la ley del cuadrado inverso.
9-5. Describa la atenuación de ondas.
9-6. Describa la absorción de ondas.
9-7. Describa la refracción; explique la ley de Snell de la refracción.
9-8. Describa la reflexión.
9-9. Describa la difracción. Explique el principio de Huygens.
9-10. Describa la composición de un buen reflector.
9-11. Describa las condiciones atmosféricas que causan la refracción electromagnética.
9-12. Defina la *interferencia de ondas electromagnéticas*.
9-13. Describa la propagación de ondas terrestres. Haga una lista de sus ventajas y sus desventajas.

9-14. Describa la propagación de las ondas espaciales.

9-15. Explique por qué el horizonte de radio está a mayor distancia que el horizonte óptico.

9-16. Describa las diversas capas de la ionosfera.

9-17. Describa la propagación de ondas celestes.

9-18. Explique por qué las condiciones ionosféricas varían con la hora del día, el mes, etc.

9-19. Defina la *frecuencia crítica* y el *ángulo crítico.*

9-20. Describa lo que es altura virtual.

9-21. Defina lo que es *máxima frecuencia útil.*

9-22. Defina la *distancia de salto* y describa las razones por las que varía.

9-23. Describa lo que es pérdida en trayectoria.

9-24. Describa qué es margen de desvanecimiento.

9-25. Describa el desvanecimiento.

PROBLEMAS

9-1. Calcule la densidad de potencia cuando la potencia irradiada es 1000 W y la distancia a la antena isotrópica es 20 km.

9-2. Calcule la densidad de potencia para el problema 9-1, para un punto a 30 km de la antena.

9-3. Describa los efectos sobre la densidad de potencia, si se eleva al triple la distancia a la antena de transmisión.

9-4. Calcule el horizonte de radio para una antena de transmisión de 100 pies de alto, y una antena de recepción de 50 pies de alto, y para antenas respectivas de 100 m y de 50 m.

9-5. Calcule la máxima frecuencia útil para una frecuencia crítica de 10 MHz y un ángulo de incidencia de 45°.

9-6. Calcule la intensidad de campo eléctrico para el mismo punto en el problema 9-1.

9-7. Calcule la intensidad de campo eléctrico para el mismo punto en el problema 9-2.

9-8. Para una potencia irradiada P_{rad} = 10 kW, calcule la intensidad de voltaje a 20 km de la fuente.

9-9. Calcule el cambio de densidad de potencia cuando la distancia a la fuente aumenta en un factor de 4.

9-10. Si la distancia a la fuente se reduce a la mitad, ¿qué efecto tiene sobre la densidad de potencia?

9-11. La densidad de potencia en un punto es 0.001 μW, y en otro punto es 0.00001 μW para la misma señal. Calcule la atenuación en decibeles.

9-12. Para una relación dieléctrica $\sqrt{\epsilon_{r2}/\epsilon_{r1}} = 0.8$ y un ángulo de incidencia $\theta_i = 26°$, calcule el ángulo de refracción θ_r.

9-13. Calcule la distancia al horizonte de radio para una antena a 40 pies sobre el nivel del mar.

9-14. Calcule la distancia al horizonte de radio para una antena que está a 40 pies sobre la cumbre de una montaña de 4000 pies de altura.

9-15. Calcule la distancia máxima entre antenas idénticas equidistantes sobre el nivel del mar para el problema 9-13.

9-16. Calcule la densidad de potencia cuando la potencia irradiada es 1200 W, a 50 km de una antena isotrópica.

9-17. Calcule la densidad de potencia para el problema 9-16 y un punto a 100 km de la misma antena.

9-18. Describa los efectos sobre la densidad de potencia si la distancia a una antena de transmisión se reduce en un factor de 3.

9-19. Calcule el horizonte de radio para una antena de transmisión de 200 pies de alto, y una de recepción de 100 pies de alto. También para antenas de 200 m y de 100 m.

9-20. Calcule la máxima frecuencia útil cuando la frecuencia crítica es de 20 MHz y el ángulo de incidencia es 35°.

9-21. Calcule la intensidad de voltaje para el mismo punto en el problema 9-16.

9-22. Calcule la intensidad de voltaje para el mismo punto en el problema 9-17.

9-23. Calcule el cambio de densidad de potencia cuando la distancia a la fuente disminuye en un factor de 8.

9-24. Calcule el cambio de densidad de potencia cuando la distancia a la fuente aumenta en un factor de 8.

9-25. Si la distancia a la fuente se reduce a una cuarta parte, ¿qué efecto tiene sobre la densidad de potencia?

9-26. La densidad de potencia en un punto es 0.002 μW, y en otro punto, para la misma señal, es 0.00002 μW. Calcule la atenuación en dB.

9-27. Calcule el ángulo de refracción para una relación de dieléctrico de 0.4, y un ángulo de incidencia $\theta_i = 18°$.

9-28. Determine la distancia al horizonte de radio de una antena de 80 pies sobre el nivel del mar.

9-29. Determine la distancia al horizonte de radio de una antena de 80 pies sobre la cumbre de una montaña de 5000 pies.

9-30. Calcule la distancia máxima entre antenas idénticas, a la misma altura sobre el nivel del mar, que la del problema 9-29.

9-31. Calcule la pérdida en trayectoria para las siguientes frecuencias y distancias:

f (MHz)	D (km)
400	0.5
800	0.6
3000	10
5000	5
8000	20
18,000	15

9-32. Calcule el margen de desvanecimiento para una transmisión de microondas a 30 km. La RF es 10 GHz, el terreno es agua y el objetivo de confiabilidad es 99.995%.

CAPÍ

Imagen protegida por derechos de autor

Antenas y guías de onda

INTRODUCCIÓN

Una *antena* es un sistema conductor metálico capaz de *radiar* y capturar ondas electromagnéticas. Las antenas son para conectar las líneas de transmisión con el espacio libre, el espacio libre a líneas de transmisión, o ambas cosas. En esencia, una línea de transmisión acopla la energía de un transmisor o de un receptor con una antena, que a su vez acopla la energía con la atmósfera terrestre, y de la atmósfera terrestre a una línea de transmisión. En el extremo transmisor de un sistema de radiocomunicaciones con el espacio libre, una antena convierte la energía eléctrica que viaja por una línea de transmisión en ondas electromagnéticas que se emiten al espacio. En el extremo receptor, una antena convierte las ondas electromagnéticas en el espacio en energía eléctrica en una línea de transmisión.

Una *guía de ondas* es un tipo especial de línea de transmisión formado por un tubo metálico conductor, a través del cual se propaga energía electromagnética. Una (o *un*) guía de ondas se usa para interconectar en forma eficiente ondas electromagnéticas entre una antena y un transceptor.

FUNCIONAMIENTO BÁSICO DE LA ANTENA

El funcionamiento básico de una antena se comprende mejor con referencia en las ilustraciones de ondas estacionarias en una línea de transmisión, fig. 10-1a. La línea de transmisión termina en un circuito abierto, que representa una discontinuidad abrupta para la onda incidente de voltaje y tiene la forma de una inversión de fase. La inversión de fase hace que se irradie algo del voltaje incidente, sin reflejarse hacia la fuente. La energía *radiada* se propaga alejándose de la antena, en forma de ondas electromagnéticas transversales. La *eficiencia de irradiación* de una línea de transmisión abierta es extremadamente baja. Es la relación de la energía irradiada entre la energía reflejada. Para irradiar más energía tan sólo se apartan entre sí los conductores. A la antena obtenida así se le llama un *dipolo* (quiere decir dos polos), y se ve en la fig. 10-1b.

En la fig. 10-1c, los conductores se apartan en una línea recta a la distancia de un cuarto de onda. Esa antena se llama *antena de cuarto de onda* o *monopolo vertical* (a veces se le

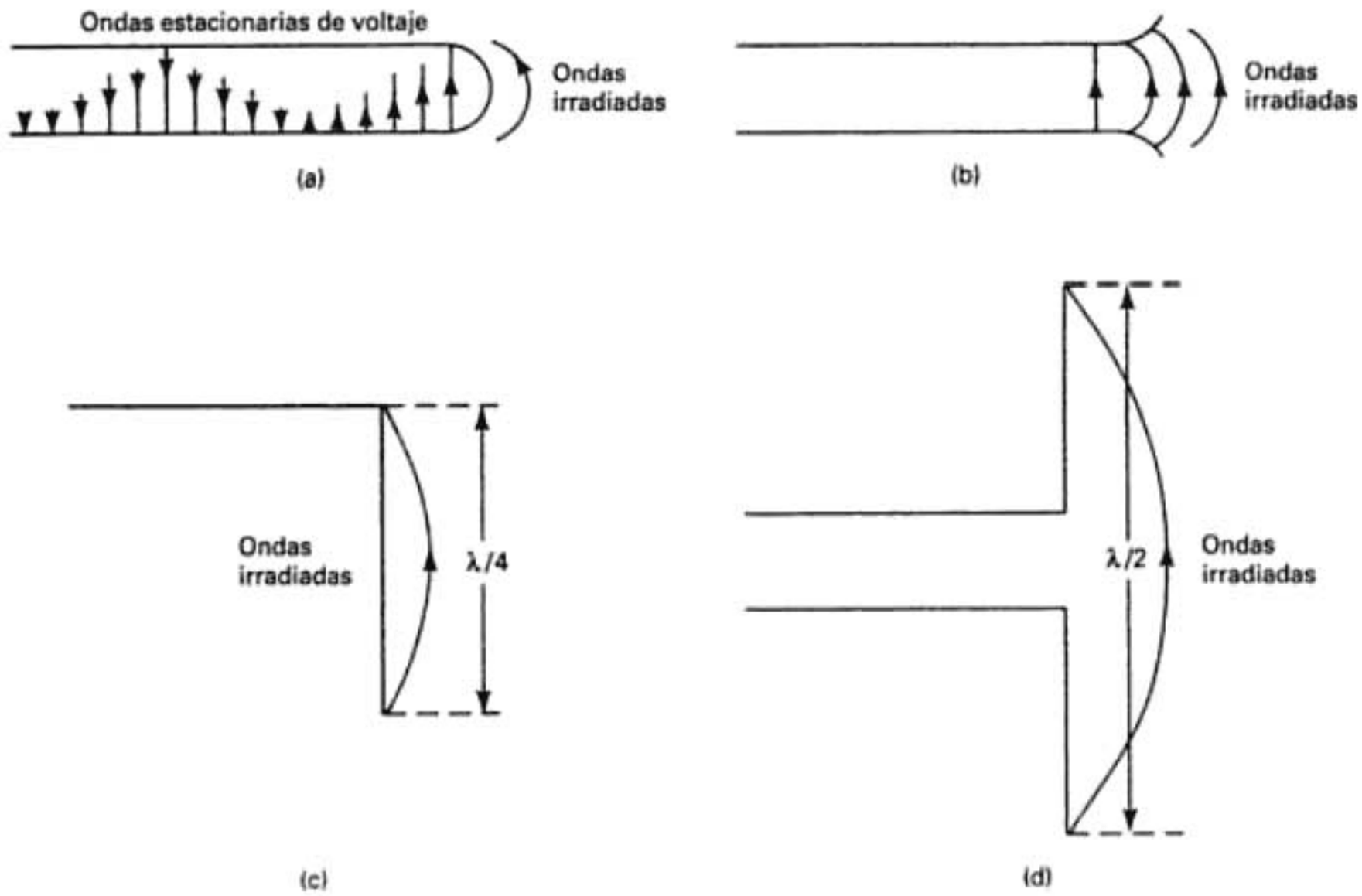


FIGURA 10-1 Radiación de una línea de transmisión: (a) radiación de línea de transmisión; (b) conductores divergentes; (c) antena de Marconi; (d) antena de Hertz

llama antena de Marconi). Un dipolo de media onda se llama *antena de Hertz*, y se ve en la fig. 10-1d.

Circuitos equivalentes a una antena

En las radiocomunicaciones, los transmisores están conectados a los receptores a través de líneas de transmisión, antenas y el espacio libre. Las ondas electromagnéticas se acoplan desde las antenas transmisoras a las receptoras, a través del espacio libre en una forma parecida a cuando la energía se acopla del primario al secundario de un transformador. Sin embargo, con las antenas el grado de acoplamiento es mucho menor que en un transformador, e interviene una onda electromagnética y no sólo una onda magnética. Un sistema de acoplamiento con antena se puede representar con una red de cuatro terminales como la de la fig. 10-2a. La energía electromagnética se debe transferir de la antena transmisora a un espacio libre y después desde el espacio libre a una antena receptora. La fig. 10-2b muestra el circuito equivalente de una antena transmisora, y la fig. 10-2c muestra el circuito equivalente para una antena receptora.

RECIPROCIDAD DE ANTENAS

Una antena básica es un *dispositivo recíproco pasivo*; pasivo porque en realidad no puede amplificar una señal, al menos no en el sentido verdadero de la palabra. Sin embargo, después, en este capítulo, se verá que una antena puede tener ganancia. Una antena es un dispositivo recíproco porque las características y el desempeño de transmisión y de recepción son idénticas, como la ganancia, directividad, frecuencia de operación, ancho de banda, resistencia de radiación, eficiencia, etc.

Las antenas de transmisión deben poder manejar potencias grandes y, en consecuencia, deben ser de materiales que soporten altos voltajes y grandes potencias, como por ejemplo, de tubo metálico. Por otra parte, las antenas de recepción producen voltajes y corrientes muy pequeños y se pueden hacer con alambre de diámetro pequeño. Sin embargo, en muchos sistemas de radio comunicación, se usa la misma antena para transmitir y para recibir. En estos casos, la

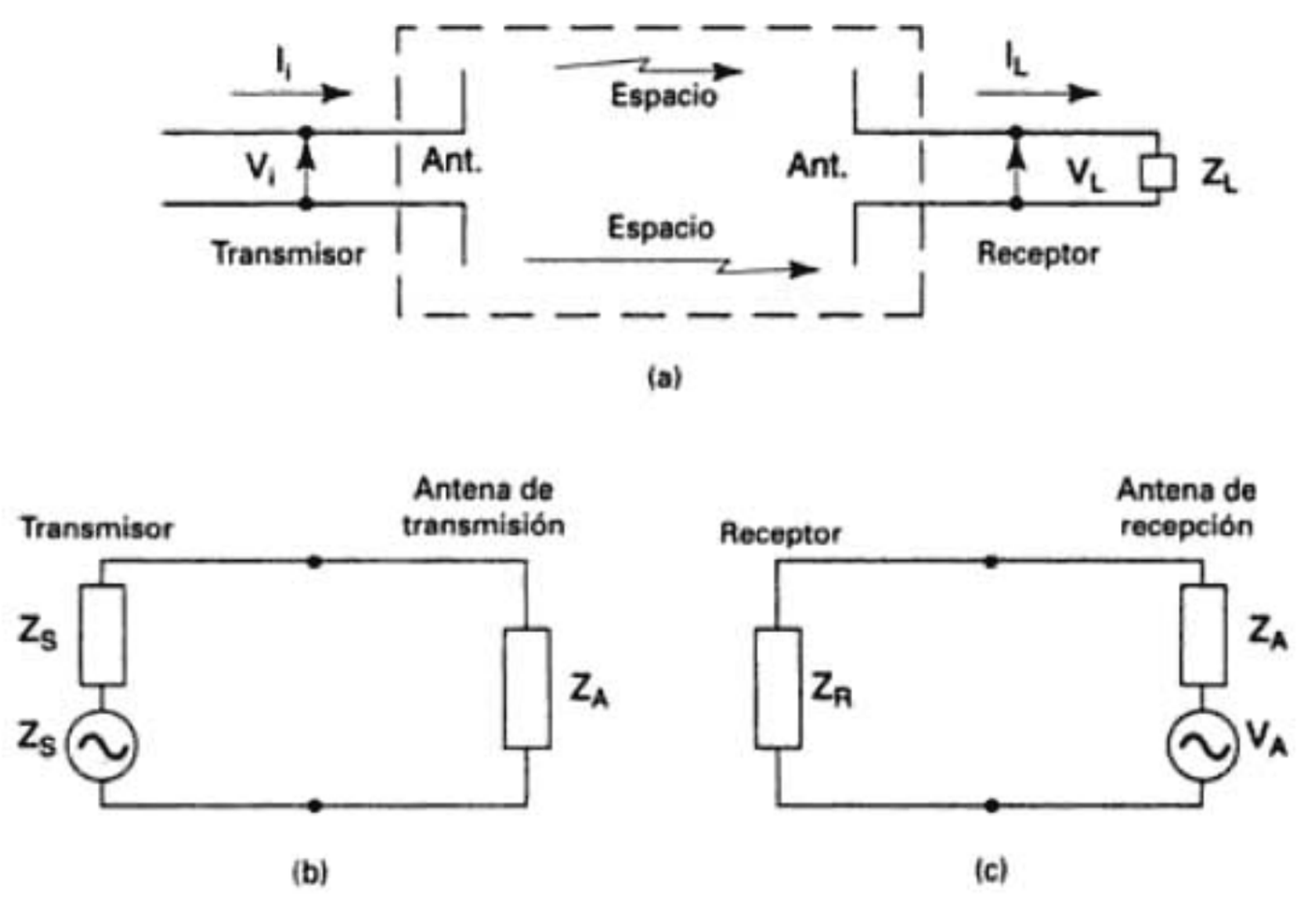


FIGURA 10-2 (a) Antena como red de cuatro terminales; (b) circuito equivalente a una antena de transmisión; (c) circuito equivalente a una antena de recepción

antena debe ser de materiales robustos. Si se usa una antena para transmitir y para recibir debe existir un medio para evitar que las señales de transmisión, que son de gran potencia, se acoplen hacia el receptor, que es relativamente sensible. Para dirigir las señales de transmisión y de recepción se puede usar un elemento especial llamado *diplexor*, así como para proporcionarles el aislamiento necesario.

Las antenas normales no tienen componentes activos (diodos, transistores, FET, etc.); por lo mismo son pasivas y recíprocas. En la práctica, una antena activa no existe. Lo que se suele llamar antena activa es en realidad la combinación de una antena pasiva y un amplificador de bajo ruido (LNA, de *low-noise amplifier*). Las antenas activas no son recíprocas, es decir, transmiten o reciben, pero no ambas cosas. Es importante hacer notar que las antenas activas y las pasivas introducen pérdidas de potencia, independientemente de si se usan para transmitir o para recibir señales. La ganancia de la antena es un término engañoso que se explicará con detalle después en este capítulo.

TERMINOLOGÍA Y DEFINICIONES PARA ANTENAS

Sistema de coordenadas de la antena

Las características direccionales de una onda electromagnética, irradiada o recibida por una antena, se describen en general en términos de coordenadas esféricas, como se ve en la fig. 10-3. Imagine el lector la antena colocada en el centro de la esfera, y la distancia a cualquier punto en la superficie de la esfera se puede definir con respecto a la antena, mediante el radio de la esfera d y los ángulos θ y Φ. El plano xy de la figura se llama plano ecuatorial, y cualquier plano que forma ángulo recto con él se llama plano meridiano.

Diagrama de radiación

Un *diagrama de radiación* es un diagrama o gráfica *polar* que representa intensidades de campo o densidades de potencia en diversas posiciones angulares en relación con una antena. Si la gráfica de radiación se traza en términos de intensidad del campo eléctrico ($\mathscr{E}$) o de densidad de potencia ($\mathscr{P}$) se llama gráfica de radiación *absoluta* (es decir, distancia variable y potencia fija). Si se grafica intensidad de campo o densidad de potencia con respecto al valor en algún punto de referencia, se llama gráfica de radiación *relativa* (es decir, potencia variable, distancia fija). La fig. 10-4a muestra una gráfica de radiación absoluta para una antena no especificada. La gráfica se traza en papel de coordenadas *polares*, y la línea gruesa representa puntos de igual

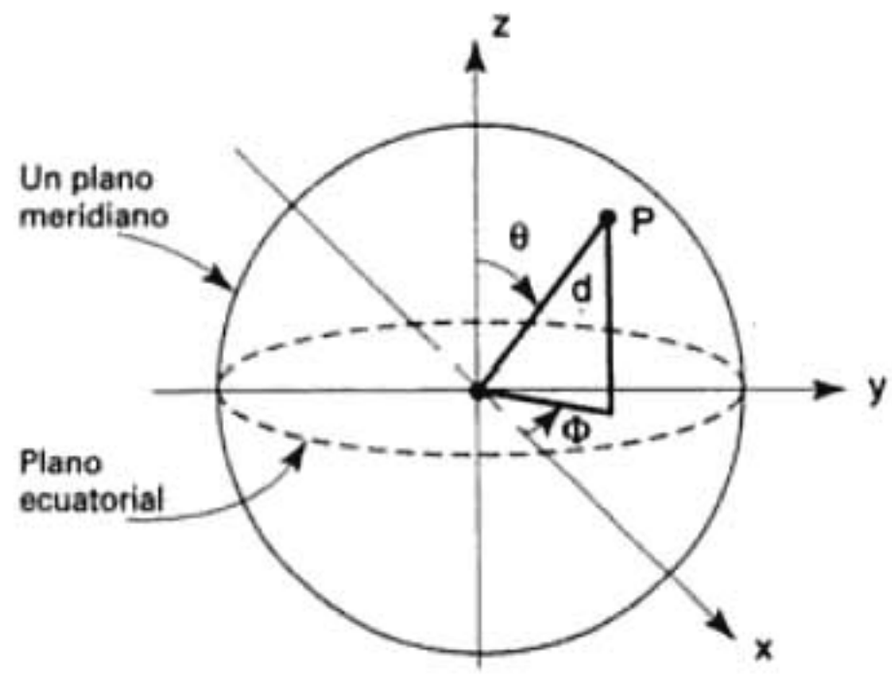


FIGURA 10-3 Coordenadas esféricas

densidad de potencia (10 μW/m^2). Los gradientes circulares indican la distancia en incrementos de 2 km. Se puede ver que la radiación máxima forma 90° con la referencia. La densidad de potencia a 10 km de la antena, en una dirección de 90° es 10 μW/m^2. En una dirección a 45°, el punto de igual densidad de potencia está a 5 km de la antena; a 180°, sólo a 4 km, y en dirección de –90°, esencialmente no hay radiación.

En la fig. 10-4a, el haz primario está en la dirección de 90°, y se llama *lóbulo mayor* o *principal*. Puede haber más de un lóbulo mayor. También hay un haz *secundario*, o lóbulo *menor* en dirección de –180°. En el caso normal, los lóbulos menores representan radiación o recepción no deseada. Como el lóbulo mayor se propaga y recibe la mayor cantidad de energía, ese lóbulo se llama lóbulo *frontal* (el frente de la antena). Los lóbulos adyacentes al frontal se llaman lóbulos *laterales* (el lóbulo menor de 180° es un lóbulo lateral) y los lóbulos cuya dirección es exactamente opuesta al lóbulo frontal son lóbulos *traseros* (en este diagrama no hay lóbulo trasero). La relación de la potencia de lóbulo frontal a la de lóbulo trasero se llama simplemente *eficiencia direccional*, y la relación de lóbulo frontal a un lóbulo lateral se llama *relación frontal a lateral*. La línea que bisecta al lóbulo mayor, o que apunta desde el centro de la antena con dirección de radiación máxima se llama *línea de tiro*, o a veces *punto de tiro*.

La fig. 10-4b muestra una gráfica de radiación relativa para una antena no especificada. La línea gruesa representa puntos a igual distancia de la antena (10 km), y los gradientes circulares indican densidad de potencia en divisiones de 1 –μW/m^2. Se ve que la radiación máxima (5 μW/m^2) está en la dirección de la referencia (0°), y que la antena irradia la menor potencia (1 μW/m^2) a 180° de la referencia. En consecuencia, la eficiencia direccional es 5:1 = 5. En general, la intensidad relativa de campo y la densidad de potencia se grafican en decibelios (dB), siendo dB = 20 log($\mathscr{E}/\mathscr{E}_{máx}$) o bien 10 log($\mathscr{P}/\mathscr{P}_{máx}$). La fig. 10-4c muestra una gráfica de radiación relativa con la densidad de potencia en decibelios. En direcciones a ±45° de la referencia, la densidad de potencia es –3 dB (mitad de potencia) en relación con la densidad de potencia en la dirección de radiación máxima (0°). La fig. 10-4d muestra la gráfica de radiación para una antena omnidireccional. Una antena omnidireccional, o *isotrópica*, irradia energía en todas direcciones por igual y, en consecuencia, la gráfica de radiación es un círculo (en realidad, una esfera). También, con una antena omnidireccional no hay lóbulos frontales, traseros o laterales, porque la radiación es igual en todas direcciones.

Las gráficas de radiación de la fig. 10-4 son bidimensionales. Sin embargo, la radiación de una antena es en realidad tridimensional. En consecuencia, las gráficas de radiación se toman tanto en el plano horizontal (vista superior) como los verticales (vista lateral). Para la antena omnidireccional de la fig. 10-4d, las gráficas de radiación en los planos horizontal y vertical son circulares e iguales, porque la gráfica real de radiación para un radiador isotrópico es una esfera.

Recuérdese que en el capítulo 9 se dijo que un radiador isotrópico verdadero irradia potencia a una tasa constante y uniformemente en todas direcciones. Una antena isotrópica ideal también irradia toda la potencia que se le suministra. No existen los radiadores isotrópicos y, sin embargo, sólo se usan en descripciones y comparaciones analíticas.

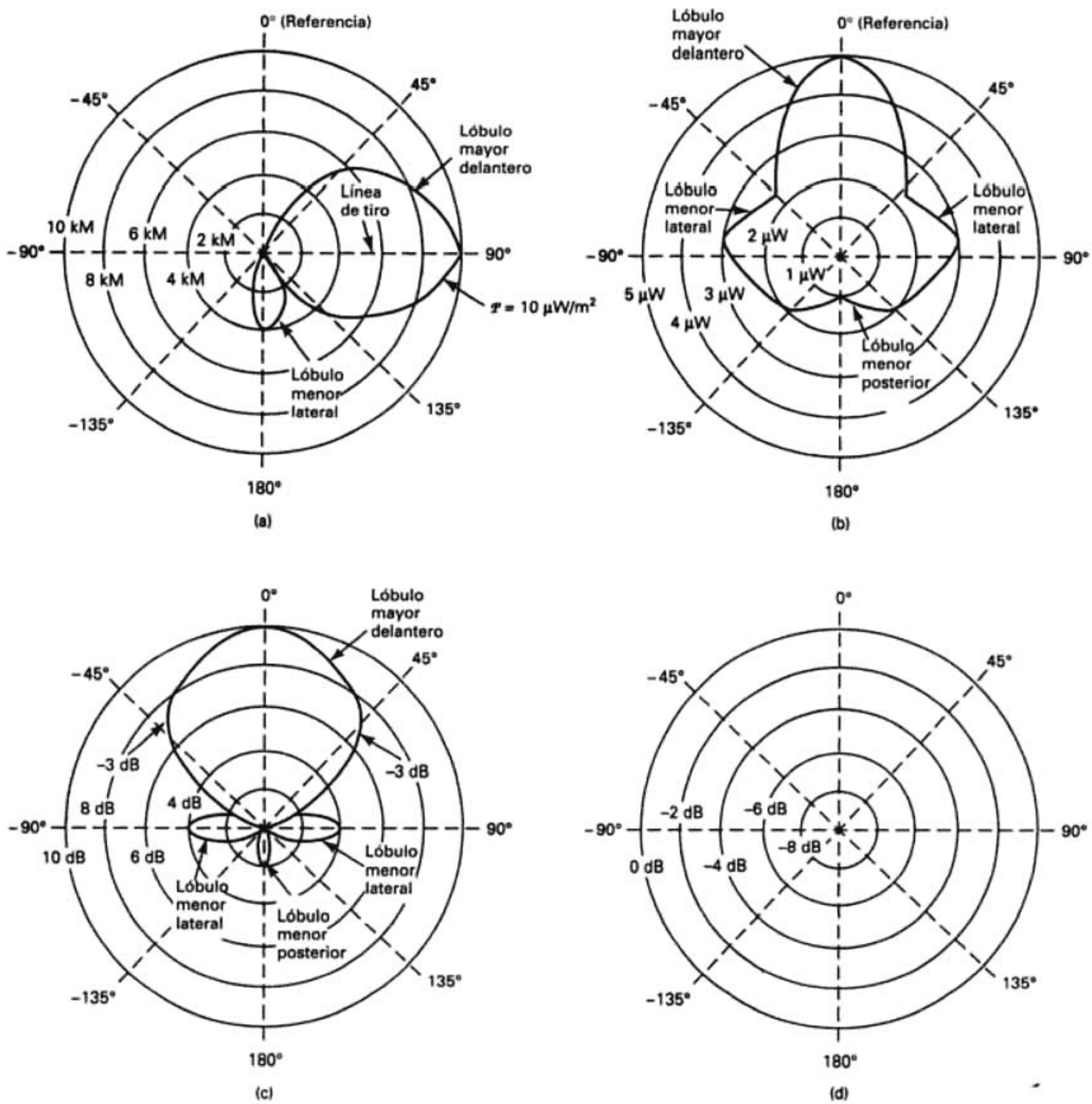


FIGURA 10-4 Gráficas de radiación: (a) gráfica de radiación absoluta (distribución fija); (b) gráfica de radiación relativa (distancia fija); (c) gráfica de radiación relativa (distancia fija) en decibelios; (d) gráfica de radiación relativa (distancia fija) en decibelios para una antena direccional (fuente puntual)

Campos cercano y lejano

El campo de radiación cercano a una antena no es igual que el que está a una gran distancia. El término *campo cercano* se refiere a la gráfica de radiación cerca de una antena, y el término *campo lejano* indica una gráfica de radiación a gran distancia. Durante medio ciclo, se irradia potencia desde una antena, donde algo de la potencia se almacena en forma temporal en el campo cercano. Durante el siguiente medio ciclo, la potencia en el campo cercano regresa a la antena. Esta acción se parece a la forma en la que un inductor almacena y libera energía. En consecuencia,

el campo cercano se llama a veces *campo de inducción*. La potencia que llega al campo lejano continúa irradiándose y alejándose, y nunca regresa a la antena. En consecuencia, a veces al campo lejano se le llama *campo de radiación*. La potencia irradiada suele ser la más importante de las dos y, en consecuencia, las gráficas de radiación son para el campo lejano. El campo cercano se define como la zona dentro de una distancia D^2/λ de la antena, siendo λ la longitud de onda y D el diámetro de la antena, en las mismas unidades.

Resistencia de radiación

No toda la potencia que se suministra a una antena se irradia. Algo de ella se convierte en calor y se disipa. La *resistencia de radiación* es un poco "irreal", porque no se puede medir en forma directa. La resistencia de radiación es una resistencia de antena a la corriente alterna, y es igual a la relación de la potencia irradiada por la antena entre el cuadrado de la corriente en su punto de alimentación. La ecuación que define la resistencia de radiación es

$$R_r = \frac{P_{rad}}{i^2} \tag{10-1}$$

en donde R_r = resistencia de radiación (ohms)
P_{rad} = potencia irradiada por la antena (watts)
i = corriente en el punto de alimentación de la antena (amperes)

La resistencia de radiación es aquella que si reemplazara a la antena disiparía exactamente la misma potencia que la que irradia la antena. La resistencia de radiación de una antena, tal como se define en la ecuación 10-1, es una cantidad ficticia en cierto sentido, porque se refiere a un punto arbitrario de la antena, que tendría distintos valores de corriente para distintos puntos de referencia. Sin embargo, se acostumbra referir la resistencia de radiación al punto de máxima corriente, o a veces a la corriente en el punto de alimentación, aunque en muchos casos los dos puntos son el mismo. Cuando se refiere al punto de corriente máxima, la resistencia de radiación se llama a veces *resistencia de radiación de lazo*, o *resistencia de radiación de antinodo*, porque un máximo de corriente también es un antinodo de corriente.

Es aparente que la resistencia de radiación es, a veces, un concepto nebuloso, porque no siempre se mide con facilidad. Es concepto útil sólo cuando se puede medir con facilidad, y no tiene significado en antenas en las que no hay un valor bien definido que se pueda tomar como referencia.

La *eficiencia de la antena* es la relación de la potencia irradiada por ella entre la suma de la potencia irradiada y la potencia disipada, o la relación de la potencia irradiada por la antena entre la potencia total de entrada. La ecuación de definición es

$$\eta = \frac{P_{\text{rad}}}{P_{\text{ent}}} \times 100 \tag{10-2a}$$

donde η = eficiencia de la antena (porcentaje)
P_{rad} = potencia irradiada (watts)
P_{ent} = potencia de entrada (watts)

o bien

$$\eta = \frac{P_{\text{rad}}}{P_{\text{rad}} + P_d} \times 100 \tag{10-2b}$$

donde P_{rad} = potencia irradiada por la antena (watts)
P_d = potencia disipada en la antena (watts)

La fig. 10-5 muestra un circuito eléctrico simplificado y equivalente a una antena. Algo de la potencia de entrada se disipa en la resistencia efectiva (del suelo, corona, dieléctricos imperfectos, corrientes parásitas, etc,), y el resto se irradia. La potencia total de la antena es la

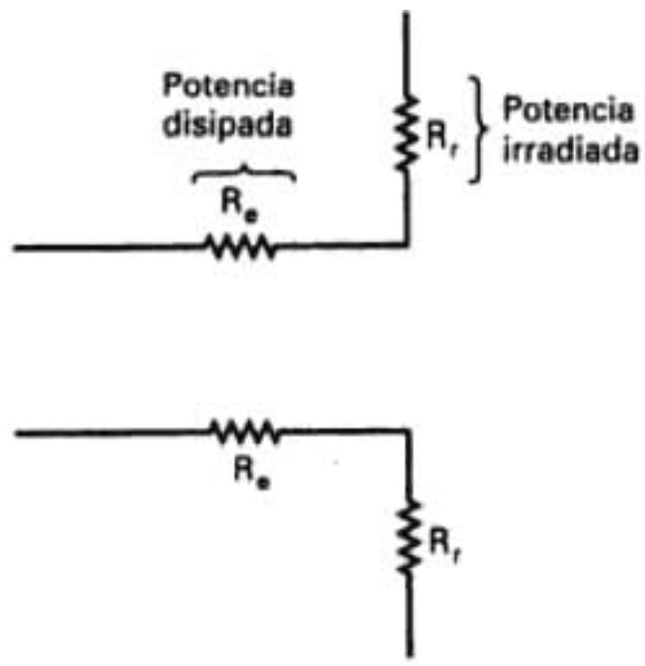


FIGURA 10-5 Circuito equivalente simplificado de una antena

suma de las potencias disipada e irradiada. En consecuencia, en términos de resistencia y de corriente, la eficiencia de la antena es

$$\eta = \frac{i^2 R_r}{i^2(R_r + R_e)} = \frac{R_r}{R_r + R_e} \tag{10-3}$$

en donde η = eficiencia de la antena
i = corriente de la antena (amperes)
R_r = resistencia de radiación (ohms)
R_e = resistencia efectiva de la antena (ohms)

Ganancia directiva y ganancia de potencia

Los términos *ganancia directiva* y *ganancia de potencia* se malentienden con frecuencia, y por lo tanto, se usan mal. La ganancia directiva es la relación de la densidad de potencia irradiada en una dirección particular entre la densidad de potencia irradiada al mismo punto por una antena de referencia, suponiendo que ambas antenas estén irradiando la misma cantidad de potencia. La gráfica de densidad de potencia de radiación para una antena en realidad es una gráfica de ganancia directiva, si se toma la referencia de densidad de potencia para una antena normal de referencia, que en general es una antena isotrópica. La ganancia directiva máxima se llama *directividad*. La ecuación de definición es

$$\mathcal{D} = \frac{\mathcal{P}}{\mathcal{P}_{\text{ref}}} \tag{10-4}$$

en donde $\mathcal{D}$ = ganancia directiva (adimensional)
$\mathcal{P}$ = densidad de potencia en un punto, con determinada antena (watts/m^2)
$\mathcal{P}_{\text{ref}}$ = densidad de potencia en el mismo punto, con una antena de referencia (watts/m^2)

La *ganancia de potencia* es lo mismo que la ganancia directiva, excepto que se usa la potencia total alimentada a la antena; es decir, se toma en cuenta la eficiencia de la antena. Se supone que la antena dada y la antena de referencia tienen la misma potencia de entrada, y que la antena de referencia no tiene pérdidas (η = 100%). La ecuación de la ganancia de potencia, A_p, es

$$A_p = \mathcal{D}\eta \tag{10-5}$$

Si la antena es sin pérdidas, irradia 100% de la potencia de entrada, y la ganancia de potencia es igual a la ganancia directiva. La ganancia de potencia de una antena también se expresa en decibelios en relación con una antena de referencia. En este caso, la ganancia de potencia es

$$A_{\text{p (dB)}} = 10 \log \frac{\mathcal{P}\eta}{\mathcal{P}_{\text{ref}}} \tag{10-6}$$

Para una referencia isotrópica, la ganancia de potencia en decibelios, de un dipolo de media onda, es 1.64 (2.15 dB), aproximadamente. Se acostumbra expresar la ganancia de potencia en decibelios, cuando se refiere a un dipolo de $\lambda/2$ (dBd). Sin embargo, si la referencia es un radiador isotrópico, se mencionan los decibelios como dBi, o dB/radiador isotrópico, y es 2.15 dB mayor que si se usara un dipolo de media onda como referencia. Es importante notar que la potencia irradiada de una antena nunca puede ser mayor que la potencia de entrada. Por consiguiente, en realidad la antena no amplifica la potencia de entrada. Una antena tan sólo concentra su potencia irradiada en determinada dirección. Entonces, los puntos en donde la potencia irradiada se concentra obtienen una ganancia aparente, en relación con la densidad de potencia en esos mismos puntos si se usara una antena isotrópica. Si la ganancia se obtiene en una dirección, debe haber una reducción (una pérdida) de densidad de potencia en otra dirección. La dirección a la que "apunta" una antena siempre es la de máxima radiación. Como una antena es un dispositivo recíproco, su gráfica de radiación también es su gráfica de recepción. Para alcanzar una potencia *capturada* máxima, una antena de recepción debe apuntar en la dirección desde donde se desea recibir. En consecuencia, las antenas de recepción tienen directividad y ganancia de potencia, exactamente como las de transmisión.

Potencia isotrópica efectiva irradiada

La *potencia isotrópica efectiva irradiada* (EIRP, por *effective isotropic radiated power*) se define como la potencia equivalente de transmisión y se expresa como sigue

$$\text{EIRP} = P_{\text{rad}} D_t \text{ (watts)} \tag{10-7a}$$

siendo P_{rad} = potencia total irradiada (watts)
D_t = ganancia directiva de la antena de transmisión (adimensional)

o bien

$$\text{EIRP}_{(\text{dBm})} = 10 \log \frac{P_{\text{rad}}}{0.001} + 10 \log D_t \tag{10-7b}$$

o también

$$\text{EIRP}_{(\text{dBW})} = 10 \log (P_{\text{rad}} D_t) \tag{10-7c}$$

La ecuación 10-7a se puede escribir en función de la potencia de entrada y la ganancia de potencia de la antena, como sigue

$$\text{EIRP} = P_{\text{ent}} A_t \tag{10-7d}$$

en donde P_{ent} = potencia total de entrada de la antena (watts)
A_t = ganancia de potencia de la antena de transmisión (adimensional)

o bien

$$\text{EIRP}_{(\text{dBm})} = 10 \log \left(\frac{P_{\text{ent}} A_t}{0.001} \right) \tag{10-7e}$$

$$\text{EIRP}_{(\text{dBW})} = 10 \log (P_{\text{ent}} A_t) \tag{10-7f}$$

La EIRP, o simplemente ERP (de *effective radiated power*, potencia efectiva radiada) es la potencia equivalente que tendría que irradiar una antena isotrópica para alcanzar la misma densidad de potencia en la dirección elegida y en determinado punto, que otra antena. Por ejemplo, si determinada antena de transmisión tiene ganancia de potencia de 10, la densidad de potencia a una distancia dada de la antena es 10 veces mayor que la que sería si la antena fuera un radiador isotrópico. Una antena isotrópica tendría que irradiar 10 veces más potencia para alcanzar la misma densidad de potencia. Por lo tanto, la antena dada irradia 10 veces más potencia que una antena isotrópica con la misma potencia y eficiencia.

Para calcular la densidad de potencia en determinado punto a una distancia R de la antena transmisora, se puede ampliar la ecuación 9-5 para incluir la ganancia de esa antena, y escribir como sigue

$$\mathscr{P} = \frac{P_{\text{ent}} A_t}{4\pi R^2} \tag{10-8a}$$

o en términos de ganancia directiva

$$\mathcal{P} = \frac{P_{rad} D_t}{4\pi R^2} \quad (10\text{-}8b)$$

siendo $\mathcal{P}$ = densidad de potencia (watts por metro cuadrado)
P_{ent} = potencia de entrada a la antena de transmisión (watts)
P_{rad} = potencia radiada por la antena de transmisión (watts)
A_t = ganancia de potencia de la antena de transmisión (adimensional)
D_t = ganancia directiva de potencia de la antena de transmisión (adimensional)
R = distancia a la antena de transmisión (metros)

Ejemplo 10-1

Para una antena transmisora con ganancia de potencia $A_t = 10$ y potencia de entrada $P_{ent} = 100$ W, calcular:

(a) EIRP en watts, dBm y dBW.
(b) Densidad de potencia en un punto a 10 km de la antena.
(c) Densidad de potencia si se hubiera usado una antena isotrópica, con la misma potencia y eficiencia.

Solución (a) Se sustituyen valores en las ecuaciones 10-7d, e y f, para obtener

$$\begin{aligned} \text{EIRP} &= (100\text{ W})(10) \\ &= 1000\text{ W} \\ \text{EIRP}_{(\text{dBm})} &= 10 \log \frac{1000}{0.001} \\ &= 60\text{ dBm} \\ \text{EIRP}_{(\text{dBW})} &= 10 \log 1000 \\ &= 30\text{ dBW} \end{aligned}$$

(b) Sustituyendo en la ecuación 10-8a

$$\begin{aligned} \mathcal{P} &= \frac{(100\text{ W})(10)}{4\pi(10{,}000\text{ m})^2} \\ &= 0.796\ \mu\text{W/m}^2 \end{aligned}$$

(c) De acuerdo con la ecuación 9-5

$$\begin{aligned} \mathcal{P} &= \frac{(100\ W)}{4\pi(10{,}000\text{ m})^2} \\ &= 0.0796\ \mu\text{W/m}^2 \end{aligned}$$

Se ve que en el ejemplo 10-1, la densidad de potencia en un punto a 10 km de la antena de transmisión es 10 veces mayor con esta antena que la que se hubiera obtenido con un radiador isotrópico. Para lograr la misma densidad de potencia, la antena isotrópica hubiera requerido una potencia de entrada 10 veces mayor, o 100 W. La antena de transmisión del ejemplo irradia, en forma efectiva, el equivalente de 1000 W.

Ejemplo 10-2

Para una antena de transmisión con resistencia de radiación $R_r = 72$ ohms, resistencia efectiva de antena $R_e = 8$ ohms, ganancia directiva $D = 20$ y potencia de entrada $P_{ent} = 100$ W, calcular:

(a) La eficiencia de la antena.
(b) La ganancia de la antena (absoluta y en dB).
(c) La potencia irradiada en watts, dBm y dBW.
(d) La EIRP en watts, dBM y dBW.

Solución (a) La eficiencia de la antena se calcula con la ecuación 10-3

$$\begin{aligned} \eta &= \frac{72}{72 + 8} \times 100 \\ &= 90\% \end{aligned}$$

(b) La ganancia de la antena no es más que el producto de su eficiencia por su ganancia directiva

$$\begin{aligned} A &= (0.9)(20) \\ &= 18 \end{aligned}$$

y

$$A(\text{dB}) = 10 \log 18 = 12.55 \text{ dB}$$

(c) La potencia irradiada se calcula con la ecuación 10-2a, reordenada

$$P_{\text{rad}} = \eta P_{\text{ent}} = (0.9)(100 \text{ W}) = 90 \text{ W}$$

$$P_{\text{rad(dBm)}} = 10 \log \frac{90}{0.001} = 49.54 \text{ dBm}$$

$$P_{\text{rad(dBW)}} = 10 \log 90 = 19.54 \text{ dBW}$$

(d) La EIRP se calcula sustituyendo en las ecuaciones 10-7d, e y f

$$\text{EIRP} = (100 \text{ W})(18) = 1800 \text{ W}$$

$$\text{EIRP}_{\text{(dBm)}} = 10 \log \frac{1800}{0.001} = 62.55 \text{ dBm}$$

$$\text{EIRP}_{\text{(dBW)}} = 10 \log 1800 = 32.55 \text{ dBW}$$

Densidad de potencia capturada

Las antenas son dispositivos recíprocos; así, tienen la misma resistencia de radiación, ganancia de potencia y directividad, cuando se usan para recibir las ondas electromagnéticas que cuando se usan para transmitirlas. En consecuencia, la densidad de potencia recibida o capturada por una antena es el producto de la densidad de potencia en el espacio que rodea a la antena de recepción, por su ganancia directiva. Así, se puede ampliar la ecuación 10-8a para incluir la ganancia de potencia de la antena receptora, escribiéndola en la siguiente forma

$$C = \frac{(P_{\text{ent}})(A_t)(A_r)}{4\pi R^2} \tag{10-9}$$

en la que C = densidad de potencia capturada (watts por metro cuadrado)
P_{ent} = potencia de entrada a la antena de transmisión (watts)
A_t = ganancia de potencia de la antena de transmisión (adimensional)
A_r = ganancia de potencia de la antena de recepción (adimensional)
R = distancia entre las antenas de recepción y transmisión (metros)

Densidad de potencia capturada; es la densidad de potencia en el espacio, en W/m^2, y es una cantidad algo engañosa. Lo que tiene más importancia es la potencia real, en watts, que produce una antena en sus terminales de salida; esto, naturalmente, depende de cuánta potencia captura la antena de recepción, así como de su eficiencia.

Área de captura y potencia capturada

Si bien existe una relación recíproca entre las propiedades de transmisión y de recepción de una antena, con frecuencia es más útil describir las propiedades de recepción en una forma un poco distinta. Mientras que la ganancia de potencia es el parámetro natural para describir la mayor densidad de potencia de una señal transmitida, por las propiedades direccionales de la antena transmisora, para describir las propiedades receptoras de una antena se usa una cantidad relacionada, que se llama *área de captura*.

El *área de captura* de una antena es un *área efectiva*, y se puede describir como sigue: una antena de transmisión irradia una onda electromagnética que tiene cierta densidad de potencia, en W/m^2, en el lugar de la antena de recepción. No es la potencia real recibida, sino más bien la cantidad de potencia que incide en, o pasa a través de cada área unitaria de una superficie imaginaria que es perpendicular a la dirección de propagación de las ondas electromagnéticas. Una antena receptora expuesta al campo electromagnético tendrá inducidos en ella un voltaje y

una corriente de radiofrecuencia, que producirán una potencia correspondiente de radiofrecuencia en las terminales de salida de la antena. En principio, la potencia (en watts) disponible en las terminales de salida de la antena, es la *potencia capturada*. La potencia capturada se puede entregar a una carga, que puede ser una línea de transmisión o los circuitos de entrada a un receptor. Para que aparezca la potencia capturada en las terminales de la antena, ésta debe haber capturado potencia de un espacio inmediato que la rodee.

La potencia capturada es directamente proporcional a la densidad de potencia recibida y al área de captura de la antena receptora. Como cabe esperar, el área física transversal de una antena, y su área efectiva de captura, no necesariamente son iguales. De hecho, algunas veces hay antenas con áreas transversales físicamente pequeñas que pueden tener áreas de captura bastante mayores que sus áreas físicas. En esos casos, es como si la antena se extendiera y capturara o absorbiera potencia de una zona mayor que su tamaño físico.

Hay una relación obvia entre el tamaño de una antena y su capacidad para capturar la energía electromagnética. Esto sugiere que también debe haber una relación entre la ganancia de una antena y su área transversal de recepción. La relación entre las dos cantidades se expresa como sigue

$$A_c = \frac{A_r \lambda^2}{4\pi} \tag{10-10}$$

siendo A_c = área efectiva de captura (metros cuadrados)
λ = longitud de onda de la señal recibida (metros)
A_r = ganancia de potencia de la antena receptora (adimensional)

Si se despeja la ganancia de la antena de la ecuación 10-10, se obtiene

$$A_r = \frac{A_c 4\pi}{\lambda^2} \tag{10-11}$$

La potencia capturada es tan sólo el producto de la densidad de potencia en la zona que rodea a la antena receptora, por el área de captura de esa antena. La potencia capturada se define como

$$P_{cap} = \mathcal{P} A_c \tag{10-12}$$

en donde P_{cap} = potencia capturada (watts)
A_c = área efectiva de captura (metros cuadrados)
$\mathcal{P}$ = densidad de potencia capturada (watts por metro cuadrado)

Se pueden sustituir las ecuaciones 10-8a y 10-10 en la ecuación 10-12, y se obtiene

$$P_{cap} = \frac{(P_{ent} A_t)(A_r \lambda^2)}{16\pi^2 R^2} \tag{10-13}$$

en donde P_{cap} = potencia capturada (watts)
λ = longitud de onda de la señal recibida (metros)
A_r = ganancia de potencia de la antena receptora (adimensional)
A_t = ganancia de potencia de la antena de transmisión (adimensional)
R = distancia entre las antenas de transmisión y de recepción (metros)
P_{ent} = potencia de entrada a la antena de transmisión (watts)

Ejemplo 10-3

Calcular, para una densidad de potencia recibida de 10 μW/m^2 y una antena receptora con área de captura de 0.2 m^2, lo siguiente:
(a) Potencia capturada, en watts.
(b) Potencia capturada, en dBm.

Solución (a) De acuerdo con la ecuación 10-12:

$$P_{cap} = (10\ \mu W/m^2)(0.2 m^2)$$
$$= 2\ \mu W$$

(b)
$$P_{cap}(dBm) = 10 \log \frac{2\ \mu W}{0.001\ W}$$
$$= -27\ dBm$$

Polarización de antena

La *polarización* de una antena no es más que la orientación del campo eléctrico que se irradia de ella. Una antena puede estar *polarizada linealmente* (en general, horizontal o verticalmente, suponiendo que los elementos de la antena están en un plano horizontal o en uno vertical), *elípticamente* o *circularmente*. Si una antena irradia una onda electromagnética verticalmente polarizada, se define a la antena como verticalmente polarizada (o polarizada verticalmente). Si una antena irradia una onda electromagnética horizontalmente polarizada, se dice que la antena está polarizada horizontalmente; si el campo eléctrico gira describiendo una elipse, está elípticamente polarizada; si el campo eléctrico gira en forma circular, está circularmente polarizada. La fig. 10-6 muestra las diversas polarizaciones que se acaban de describir.

Abertura del haz de la antena

La *abertura (angular) del haz* de una antena es la separación angular entre dos puntos de media potencia (−3 dB) en el lóbulo mayor de la gráfica de radiación de una antena, que se suele tomar en uno de los planos "principales". La abertura del haz de la antena cuya gráfica de radiación se ve en la fig. 10-7, es el ángulo definido por los puntos A, X y B (el ángulo θ). Los puntos A y B son los puntos de media potencia (la densidad de potencia en ellos es la mitad de la que hay,

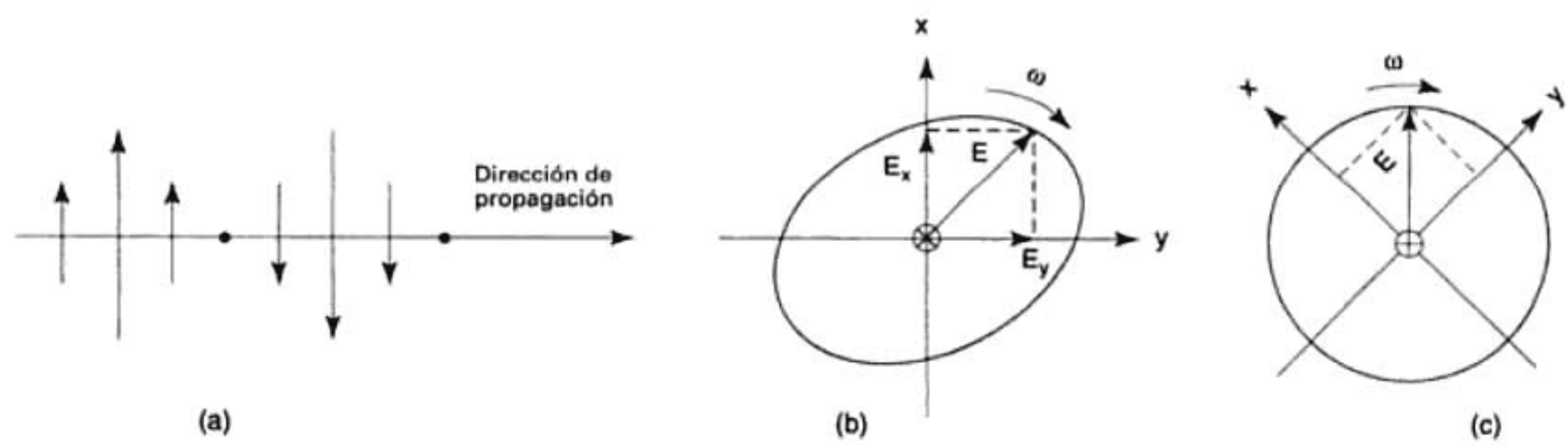


FIGURA 10-6 Polarizaciones de antena: (a) lineal; (b) elíptica; (c) circular

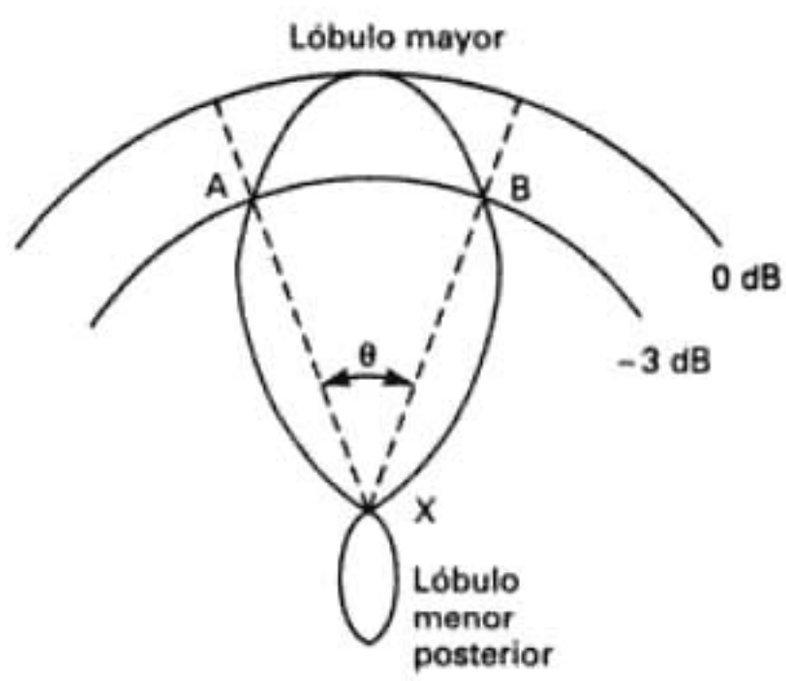


FIGURA 10-7 Abertura de haz de antena

a igual distancia de la antena, en la dirección de radiación máxima). La abertura del haz de una antena se llama a veces ancho de haz de −3 dB, o ancho de haz de media potencia, o *ancho de lóbulo*.

La ganancia de la antena es inversamente proporcional a la abertura del haz: mientras más grande es la ganancia de una antena, la abertura del haz es menor. Una antena omnidireccional (isotrópica) irradia por igual en todas direcciones. Por consiguiente, tienen ganancia de unidad y abertura de haz de 360°. Las antenas normales tienen abertura de haz de 30° a 60°, y no es raro que las antenas de microondas de alta ganancia tengan una abertura de haz hasta de 1°.

Ancho de banda de antena

El *ancho de banda* de una antena se define, en forma vaga, como el intervalo de frecuencias dentro del cual el funcionamiento de la antena es "satisfactorio". Se toma, en el caso normal, como la diferencia entre las frecuencias de media potencia (diferencia entre las frecuencias máxima y mínima de operación), pero a veces indica variaciones en la impedancia de entrada de la antena. El ancho de banda de una antena se expresa, también normalmente, como un porcentaje de la frecuencia óptima de operación de esa antena.

Ejemplo 10-4

Calcular el ancho de banda porcentual para una antena con frecuencia óptima de operación de 400 MHz, y frecuencias de −3 dB de 380 MHz y 420 MHz.

Solución

$$\text{Ancho de banda} = \frac{420 - 380}{400} \times 100$$
$$= 10\%$$

Impedancia de entrada a la antena

La radiación de una antena es un resultado directo del flujo de corriente de RF. La corriente va hacia la antena pasando por una línea de transmisión, que está conectada con un espacio pequeño entre los conductores que forman la antena. El punto de la antena donde se conecta la línea de transmisión se llama terminal de entrada de la antena, o simplemente *punto de alimentación*. El punto de alimentación presenta una carga de ca a la línea de transmisión, llamada *impedancia de entrada de* (o *a*) *la antena*. Si la impedancia de salida del transmisor y la impedancia de entrada de la antena son iguales a la impedancia característica de la línea de transmisión, no habrá ondas estacionarias en la línea y se transmitirá una potencia máxima a la antena, potencia que será irradiada.

La impedancia de entrada de una antena es sólo la relación del voltaje de entrada a la antena a la corriente de entrada a la misma. Es decir

$$Z_{\text{ent}} = \frac{E_i}{I_i} \tag{10-14}$$

en donde Z_{ent} = impedancia de entrada a la antena (ohms)
E_i = voltaje de entrada a la antena (volts)
I_i = corriente de entrada a la antena (amperes)

En general, la impedancia de entrada a la antena es compleja; sin embargo, si el punto de alimentación está en un máximo de corriente y no hay componente reactivo, la impedancia de entrada es igual a la suma de la resistencia de radiación más la resistencia efectiva.

ANTENAS BÁSICAS

Doblete elemental

El tipo más sencillo de antena es el *doblete elemental*. Es un dipolo eléctricamente corto, y con frecuencia se llama *dipolo corto*, *dipolo elemental* o *dipolo hertziano*. Eléctricamente corto quiere decir que es corto en comparación con media longitud de onda, pero no necesariamente onda de corriente uniforme; en general, cualquier dipolo menor que un décimo de longitud de

onda se considera eléctricamente corto. En realidad, no se puede obtener un doblete elemental; sin embargo, el concepto de dipolo corto es útil para comprender las antenas más prácticas.

Un doblete elemental tiene corriente uniforme en toda su longitud. Sin embargo, se supone que la corriente varía en forma senoidal en función del tiempo, y que en cualquier instante es

$$i(t) = I \operatorname{sen}(2\pi ft + \theta)$$

siendo $i(t)$ = corriente instantánea (amperes)
I = amplitud máxima de la corriente de RF (amperes)
f = frecuencia (hertz)
t = tiempo instantáneo (segundos)
θ = ángulo de fase (radianes)

Se puede demostrar, mediante las ecuaciones de Maxwell, que el campo lejano de radiación es

$$\mathscr{E} = \frac{60\pi Il \operatorname{sen} \phi}{\lambda R} \tag{10-15}$$

en donde $\mathscr{E}$ = intensidad del campo eléctrico (volts por metro)
I = corriente del dipolo (amperes rms)
l = longitud de uno a otro extremo del dipolo (metros)
R = distancia al dipolo (metros)
λ = longitud de onda (metros)
ϕ = ángulo que forma el eje de la antena con la dirección de radiación, que se indica en la fig. 10-8a

Al graficar la ecuación 10-15 se obtiene la figura de intensidad relativa del campo eléctrico para un dipolo elemental, que se ve en la fig. 10-8b. Se aprecia que la radiación es máxima en ángulo recto con el dipolo, y baja a cero en los extremos.

La gráfica de densidad de potencia relativa se puede deducir con la ecuación 10-10, sustituyendo $\mathscr{P} = \mathscr{E}^2/120\pi$. Entonces el resultado es

$$\mathscr{P} = \frac{30\pi I^2 l^2 \operatorname{sen}^2\phi}{\lambda^2 R^2} \tag{10-16}$$

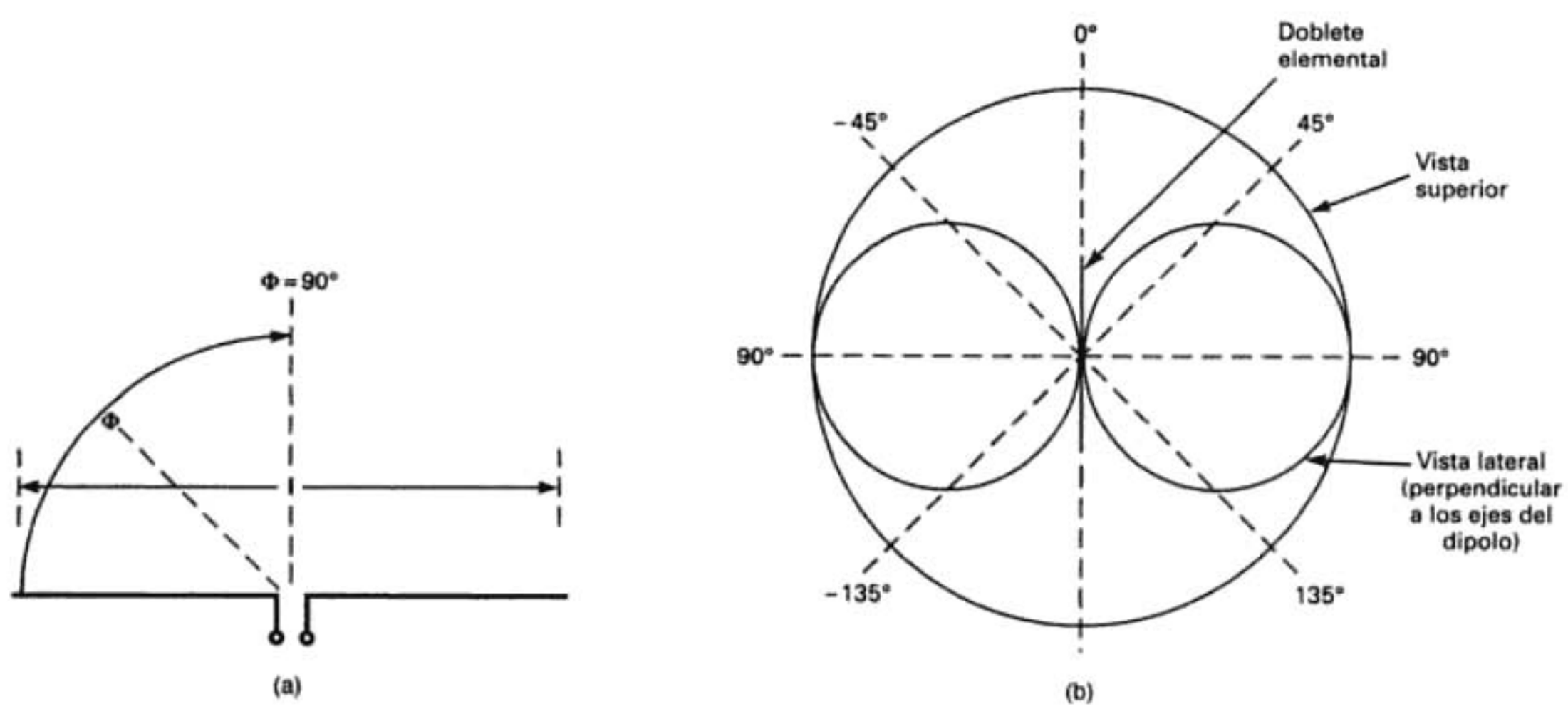


FIGURA 10-8 (a) Doblete elemental; (b) gráfica de radiación relativa

Dipolo de media onda

El dipolo lineal de media onda es una clase de antena entre las más usadas para frecuencias mayores de 2 MHz. A frecuencias menores que 2 MHz, la longitud física de una antena de media onda la hace prohibitiva. Al dipolo de media onda se le llama en general *antena de Hertz*, en honor de Heinrich Hertz, quien fue el primero en demostrar la existencia de las ondas electromagnéticas.

Una antena de Hertz es una antena *resonante*. Esto es, tiene un múltiplo de cuartos de longitud de onda de largo, y tiene circuito abierto en los extremos lejanos. A lo largo de una antena resonante se desarrollan ondas estacionarias de voltaje y corriente. La fig. 10-9 muestra las distribuciones idealizadas de corriente y voltaje a lo largo de un dipolo de media onda. Cada polo de la antena se ve como si fuera un tramo de línea de transmisión de cuarto de onda. Por lo mismo, hay un máximo de voltaje y un mínimo de corriente en los extremos, y un mínimo de voltaje y máximo de corriente en la parte media. En consecuencia, suponiendo que el punto de alimentación está en el centro de la antena, la impedancia de entrada es $E_{mín}/I_{mín}$, y tiene valor mínimo. La impedancia en los extremos de la antena es $E_{máx}/I_{máx}$, y tiene valor máximo. La fig. 10-10 muestra la curva de impedancia en un dipolo de media onda alimentado en el centro. La impedancia varía desde un valor máximo en los extremos, de aproximadamente 2500 Ω, hasta un mínimo en el punto de alimentación, de unos 73 Ω, de los cuales, entre 68 y 70 Ω es la resistencia de radiación.

Para una antena ideal, la eficiencia es 100%, la directividad es igual a la ganancia de potencia y la resistencia de radiación es igual a la impedancia de entrada (73), así que

$$D = A = \frac{120}{\text{resistencia de radiación}} \tag{10-17}$$

$$\frac{120}{73}$$

$$= 1.64$$

y

$$10 \log 1.64 = 2.16 \text{ dB}$$

Un radiador de alambre, como un dipolo de media onda, se puede concebir como una cantidad infinita de dobletes elementales lado a lado. En consecuencia, la pauta de radiación se puede obtener integrando la ecuación 10-15 sobre la longitud de la antena. La pauta de radiación de un dipolo de media onda en el espacio libre depende de si la antena se pone en dirección horizontal o vertical con respecto a la superficie terrestre. La fig. 10-11a muestra la pauta de radiación de un dipolo de media onda montado verticalmente, visto de canto. Nótese que dos lóbulos irradian en direcciones opuestas, en ángulo recto con la antena. También nótese que los lóbulos no son círculos. Los lóbulos circulares sólo se obtienen en el caso ideal cuando la corriente es constante en toda la longitud de la antena, lo cual es inalcanzable en una antena práctica. La fig. 10-11b muestra el corte transversal. Obsérvese que la pauta de radiación tiene

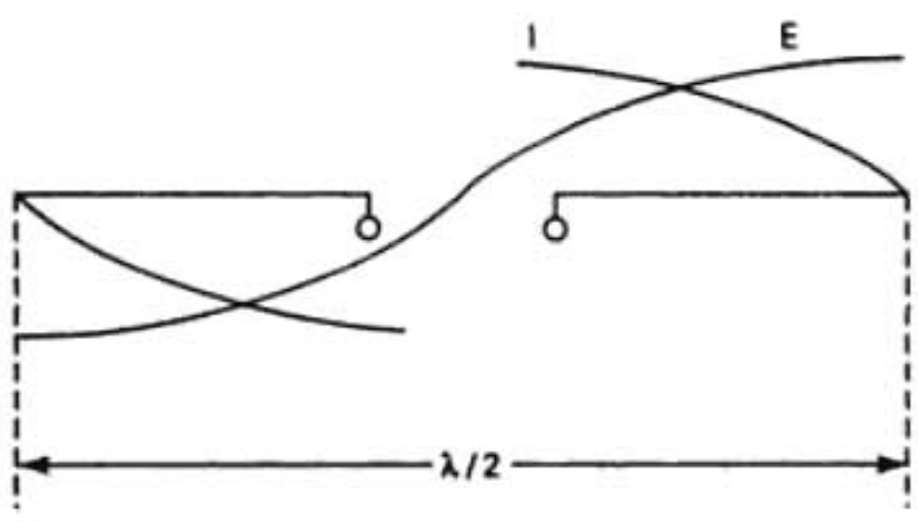


FIGURA 10-9 Distribuciones idealizadas de voltaje y de corriente a lo largo de un dipolo de media onda

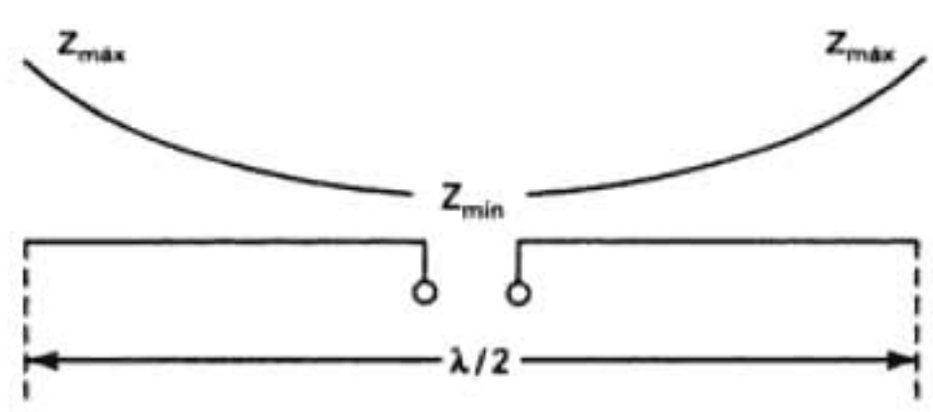


FIGURA 10-10 Curva de impedancia para un dipolo de media onda con alimentación central

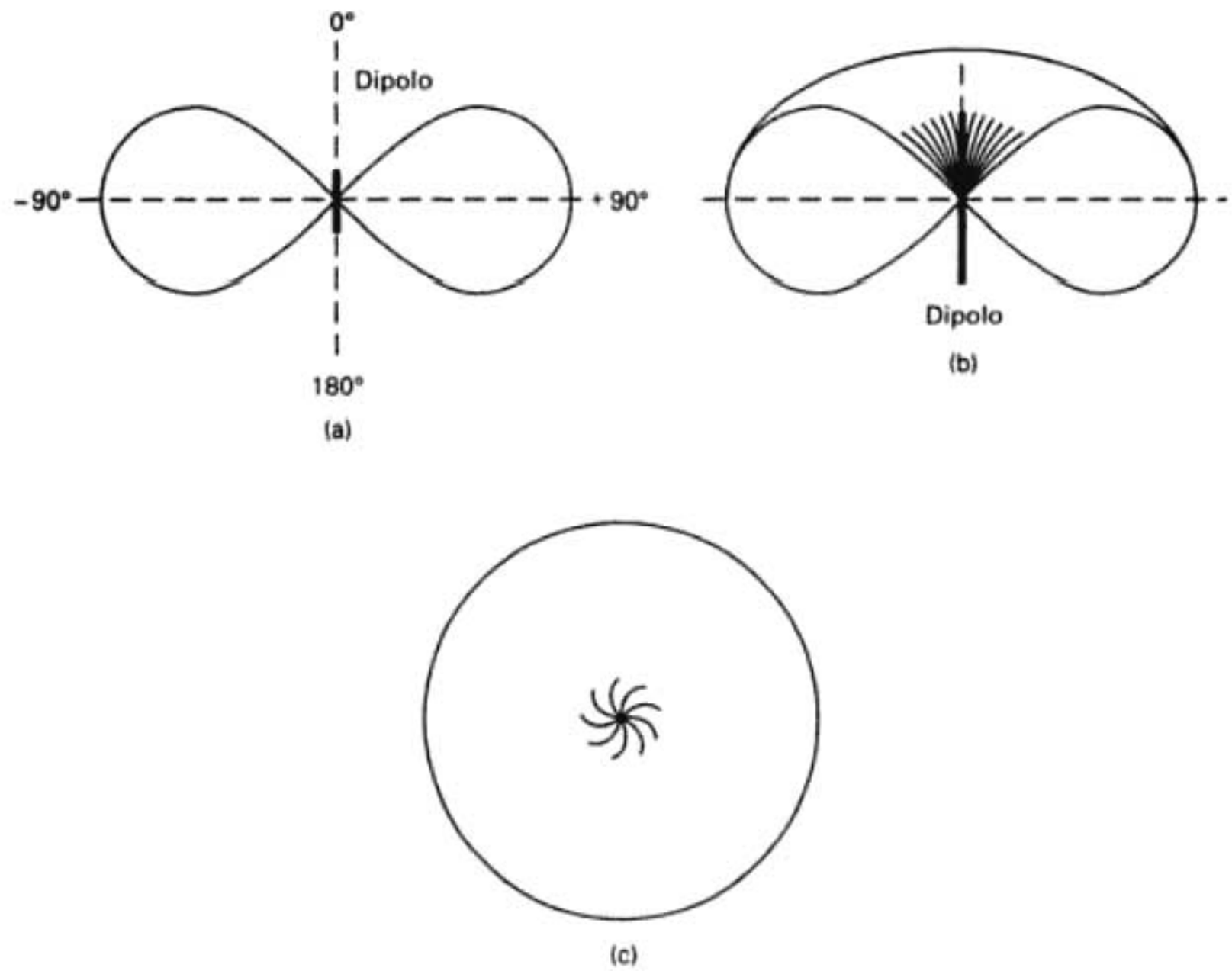


FIGURA 10-11 Gráficas de radiación de un dipolo de media onda: (a) vista vertical (lateral) de un dipolo montado verticalmente; (b) vista del corte transversal; (c) vista horizontal (superior)

forma de ocho, que se parece a la de una dona. La radiación máxima está en un plano paralelo a la superficie terrestre. Mientras mayor es el ángulo de elevación, la radiación es menor, y para 90° no hay radiación. La fig. 10-11c muestra la vista superior de la pauta de radiación para un dipolo de media onda montado en dirección vertical. La forma es circular, porque la radiación es uniforme en todas las direcciones perpendiculares a la antena.

Efectos del terreno sobre un dipolo de media onda. Las pautas de radiación de la fig. 10-11 son para condiciones en el espacio libre. En la atmósfera terrestre, la propagación ondulatoria está influida por la orientación de la antena, la absorción atmosférica y los efectos del suelo, como la reflexión. El efecto de la reflexión de un dipolo de media onda no aterrizado sobre el suelo se ve en la fig. 10-12. La antena se monta a una cantidad apreciable de longitudes de onda (altura h) sobre la superficie de la Tierra. La intensidad de campo en cualquier punto dado del espacio es la suma de las ondas directa y reflejada en el suelo. La onda reflejada en el suelo parece irradiarse de una imagen de la antena situada a la distancia h bajo el suelo. Esta antena aparente es una imagen especular de la antena real. La onda reflejada en el piso está invertida, o desfasada 180°, y recorre una distancia $2h$ sen θ mayor que la onda directa, para llegar al mismo punto P en el espacio. La pauta de radiación que resulta es una suma de las radiaciones de la antena real y la antena espejo, o especular. Nótese que es la técnica de trazado de rayos, clásica en óptica.

La fig. 10-13 muestra las gráficas de radiación vertical para un dipolo de media onda montado horizontalmente un cuarto de onda arriba del suelo. Para este caso, el lóbulo inferior desaparece por completo y la intensidad de campo directo hacia arriba se duplica. La fig. 10-13a muestra, en línea de puntos, la gráfica en espacio libre, y en línea llena, la distribución vertical en un plano que pasa por la antena. La fig. 10-13b muestra la distribución vertical en un plano en ángulo recto con la antena. La fig. 10-13c muestra la gráfica de radiación vertical para un dipolo horizontal a media onda arriba del suelo. Se ve que ahora la gráfica se descompone en dos lóbulos, y que la dirección de radiación máxima (vista de frente) está a 30° de la horizontal, en

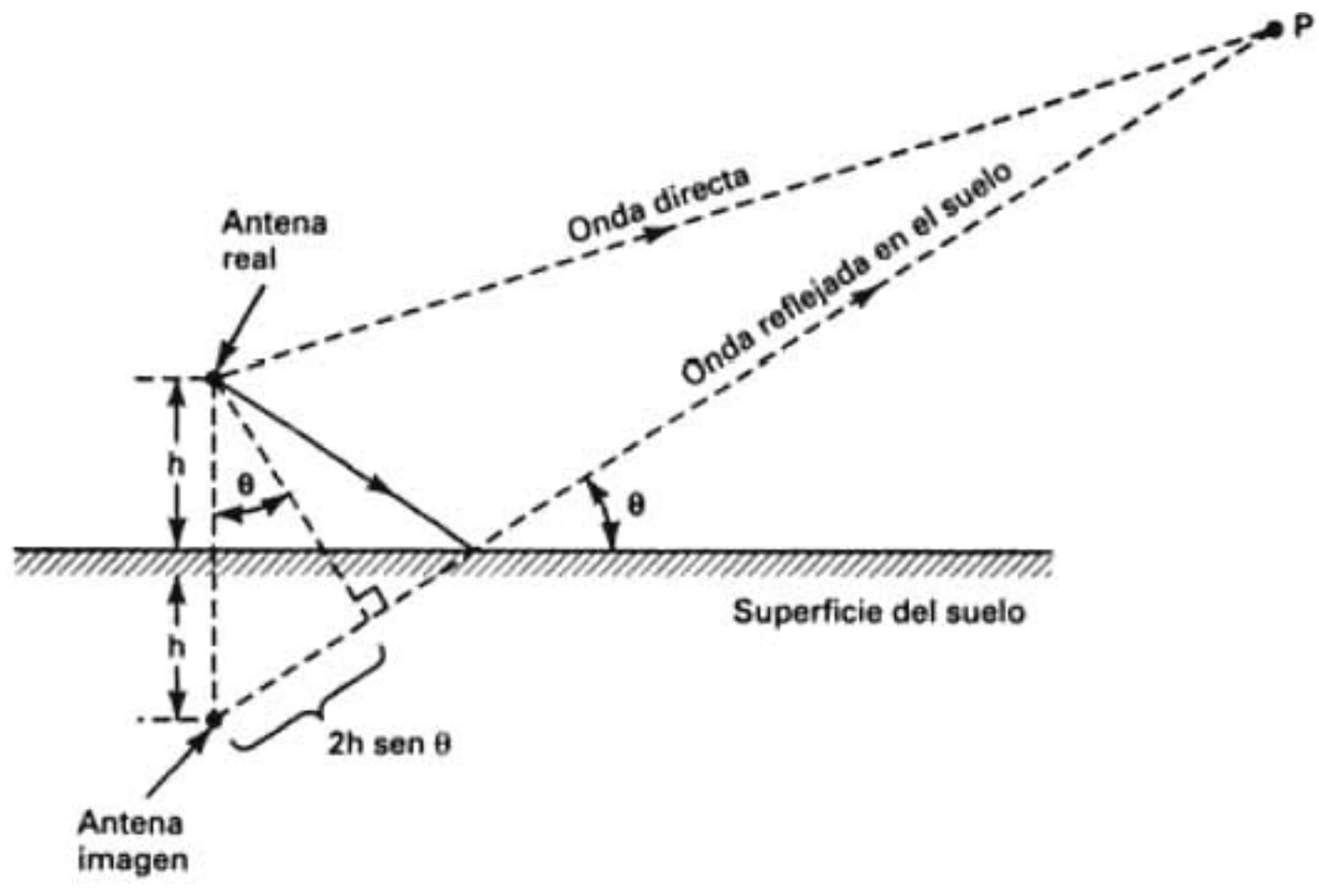


FIGURA 10-12 Efectos del suelo sobre un dipolo de media onda

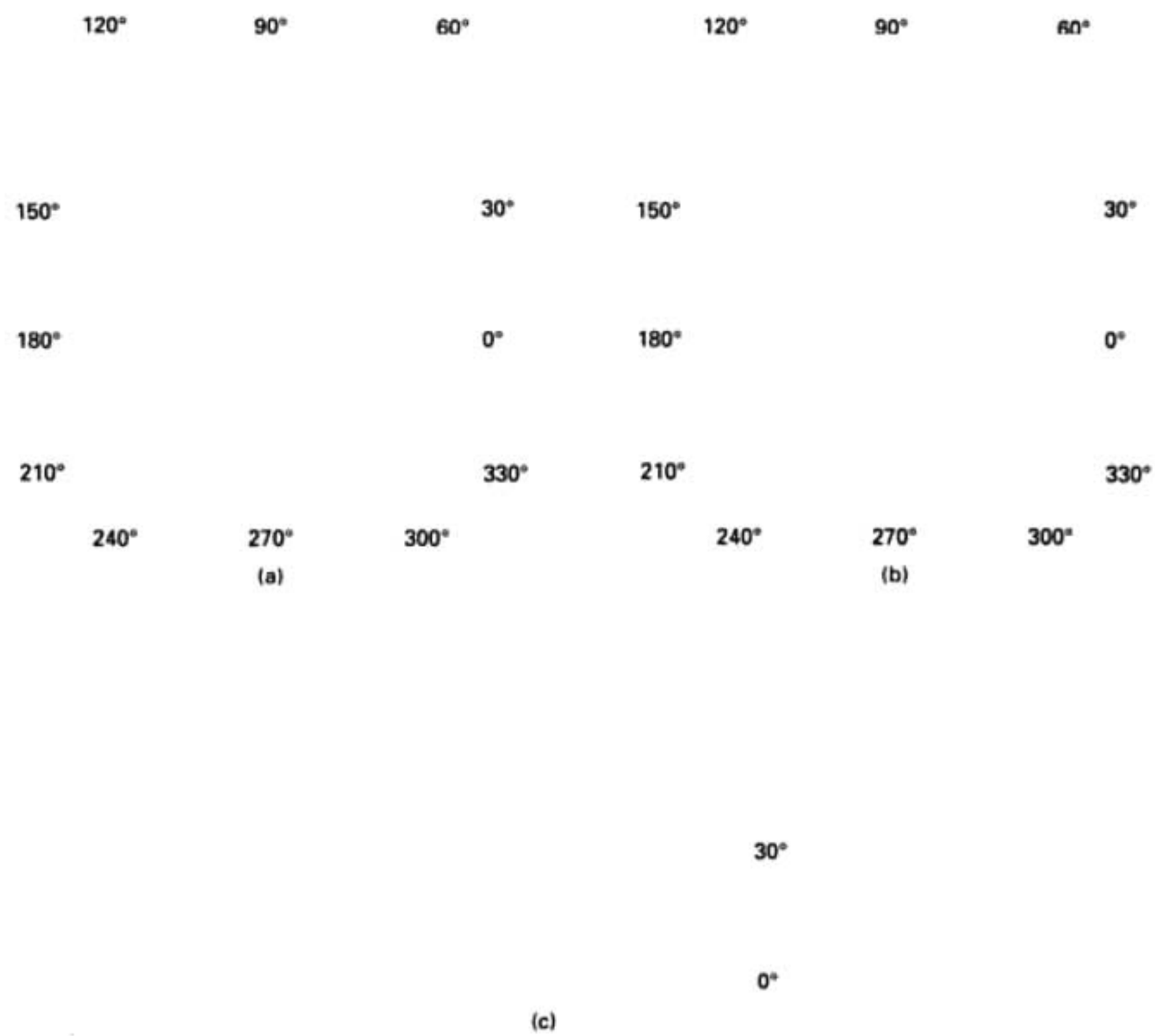


FIGURA 10-13 Gráfica de radiación vertical para un dipolo de media onda: (a) En un plano a través de la antena; (b) en un plano perpendicular a la antena; (c) dipolo horizontal a media longitud de onda sobre el suelo

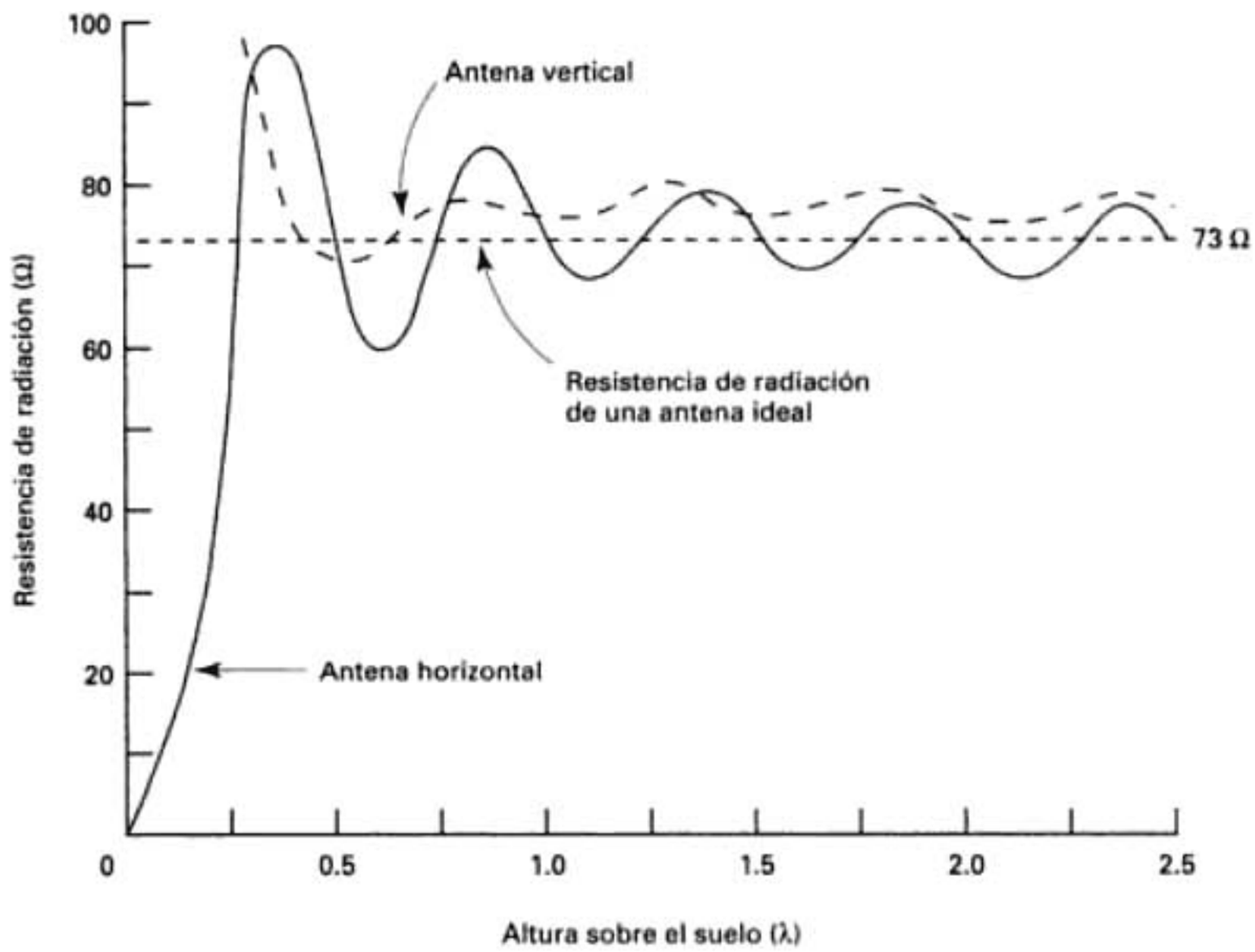


FIGURA 10-14 Resistencia de radiación en función de altura sobre el suelo

lugar de ser directo hacia arriba. No hay componente por el suelo, para la polarización horizontal, debido al corrimiento de fase del componente reflejado. Las ondas reflejadas en el piso tienen efectos parecidos en todas las antenas. La mejor forma de eliminar o reducir el efecto de las ondas reflejadas en el suelo es montar la antena con suficiente alejamiento del suelo para obtener condiciones de espacio libre. Sin embargo, en muchas aplicaciones eso es imposible. A veces, las reflexiones en el suelo son convenientes, para obtener el ángulo de elevación deseado para tener la respuesta máxima del lóbulo mayor.

La altura de una antena no aterrizada, sobre la superficie del suelo, también influye sobre la resistencia de radiación. Esto se debe a que las ondas reflejadas atraviesan o interceptan la antena, y alteran su corriente. Dependiendo de la fase de la onda reflejada en el suelo, la corriente de antena puede aumentar o disminuir, causando un aumento o disminución respectivos en la impedancia de entrada.

La fig. 10-14 muestra cómo varía la resistencia a la radiación de los dipolos de media onda, verticales y horizontales, en función de la distancia sobre la superficie del suelo. Se ve que una vez que la distancia es mayor que aproximadamente media longitud de onda sobre el piso, el efecto de las reflexiones se reduce mucho, y que la resistencia de radiación permanece relativamente constante.

Antena conectada a tierra

Una antena *monopolo* (un solo polo) de un cuarto de longitud de onda de largo, montada en dirección vertical con el extremo inferior conectado en forma directa al suelo, o aterrizada a través de la red de acoplamiento de la antena, se llama *antena de Marconi*. Las características de esta antena se parecen a las de la antena de Hertz, a causa de las ondas reflejadas en el suelo. La fig. 10-15a muestra las ondas estacionarias de corriente y voltaje para una antena de cuarto de onda aterrizada. Se ve que si se monta la antena de Marconi en forma directa sobre la superficie terrestre, se combinan la antena real y su *imagen* y producen exactamente las mismas distribuciones de ondas estacionarias que las de una antena de media onda no aterrizada (de Hertz). Los máximos de corriente se presentan en los extremos aterrizados, y eso hace que pase una gran corriente por el suelo. Para reducir las pérdidas, el terreno debe ser buen conductor, por ejemplo, como un suelo rico en arcilla. Si es mal conductor, como por ejemplo si es arenoso o rocoso, se

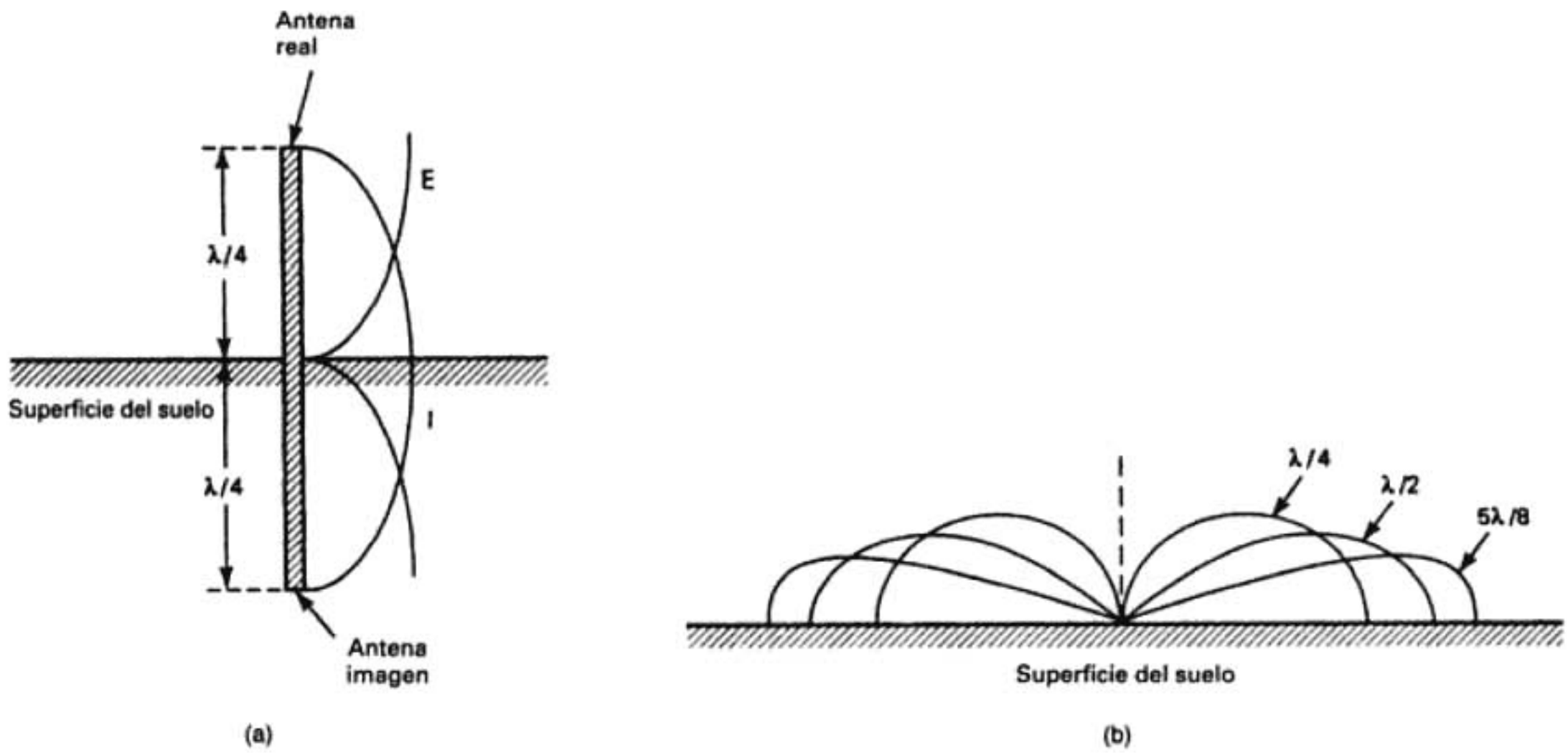


FIGURA 10-15 Antena aterrizada de cuarto de onda: (a) ondas estacionarias de voltaje y de corriente; (b) distribución de la radiación

podrá necesitar un sistema de *plano de tierra*, artificial, hecho con alambres gruesos de cobre repartidos radialmente abajo de la antena. Otra forma de mejorar la conductividad, en forma artificial, de la zona del suelo abajo de la antena, es con un *contrapeso*, *contraantena* o *antena de contrapeso*. Un contrapeso es una estructura de alambres, instalada debajo de la antena y sobre el piso. El contrapeso debe aislarse de la tierra física. Es una forma de sistema capacitivo de tierra: se forma capacitancia entre el contrapeso y la superficie terrestre.

La fig. 10-15b muestra la distribución de radiación para una antena de cuarto de onda aterrizada (de Marconi). La mitad inferior de cada lóbulo se anula por las ondas reflejadas en el suelo. Esto en general no tiene importancia, porque aumenta la radiación en dirección horizontal y se incrementa así la radiación a lo largo de la superficie terrestre (ondas terrestres) y se mejora el área abarcada. También se puede ver que al aumentar la longitud de la antena se mejora la radiación horizontal, a expensas de la propagación de ondas celestes. Esto también se ve en la fig. 10-15b. La radiación horizontal óptima se obtiene en una antena de más o menos cinco octavos de onda de longitud. Para una de un cuarto de onda, no hay propagación por ondas terrestres.

Una antena de Marconi tiene la ventaja obvia, sobre una de Hertz, de sólo tener la mitad de la longitud. La desventaja de una antena de Marconi es que debe estar cerca del suelo.

CARGA DE LA ANTENA

Hasta ahora se ha considerado la longitud de una antena en función de longitudes de onda, más que de dimensiones físicas. Por cierto, ¿qué longitud tiene una antena de un cuarto de onda, o de un cuarto de longitud de onda? Para una frecuencia de transmisión de 1 GHz, un cuarto de longitud de onda es 7.5 cm (2.95 pulg.). Sin embargo, cuando la frecuencia de transmisión es 1 MHz, un cuarto de onda es 75 m y a 100 kHz es 750 m. Es obvio que las dimensiones físicas para las antenas de baja frecuencia no son prácticas, en especial para aplicaciones de radio móvil. Sin embargo, es posible aumentar la longitud eléctrica de una antena mediante una técnica llamada *carga*. Cuando se carga una antena, su longitud física permanece inalterada, aunque su longitud eléctrica efectiva aumenta. Para cargar las antenas se usan varios métodos.

Bobinas de carga

La fig. 10-16a muestra cómo una bobina (inductor) conectada en serie con una antena de dipolo aumenta efectivamente la longitud eléctrica de una antena. Esa bobina tiene el adecuado nombre

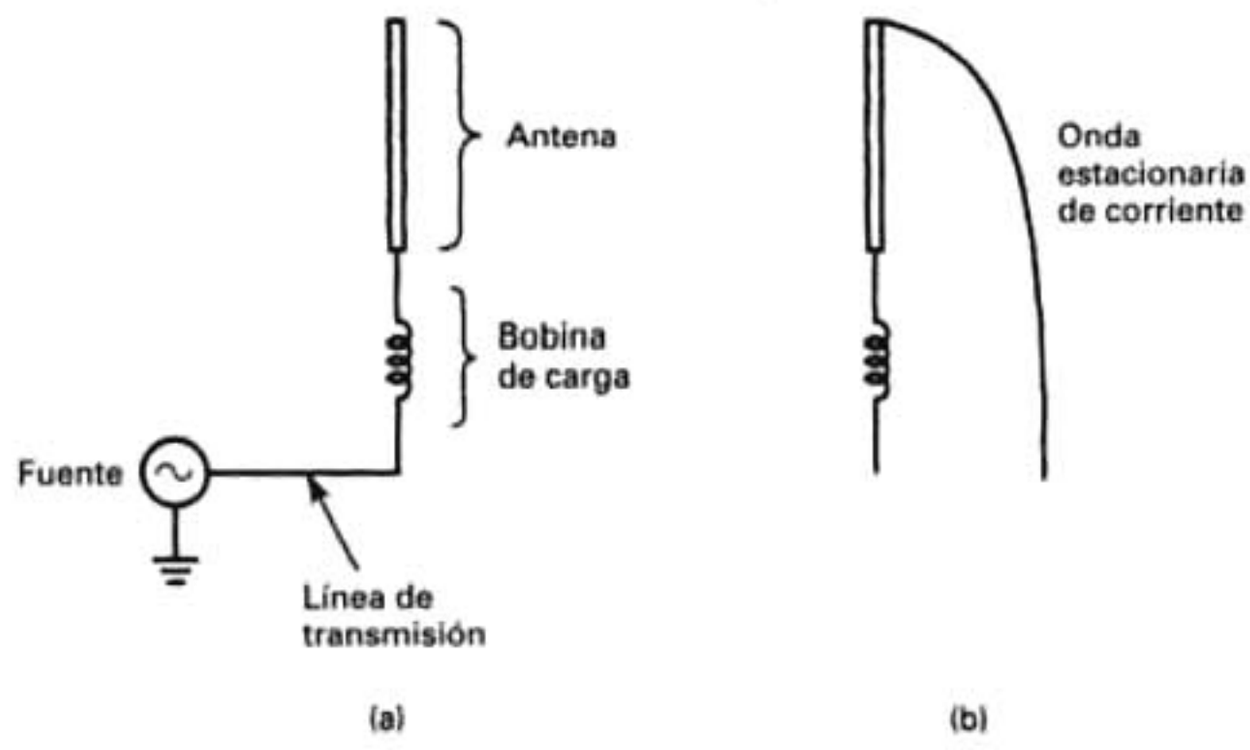


FIGURA 10-16 Bobina de carga: (a) antena con bobina de carga; (b) onda estacionaria de corriente con bobina de carga

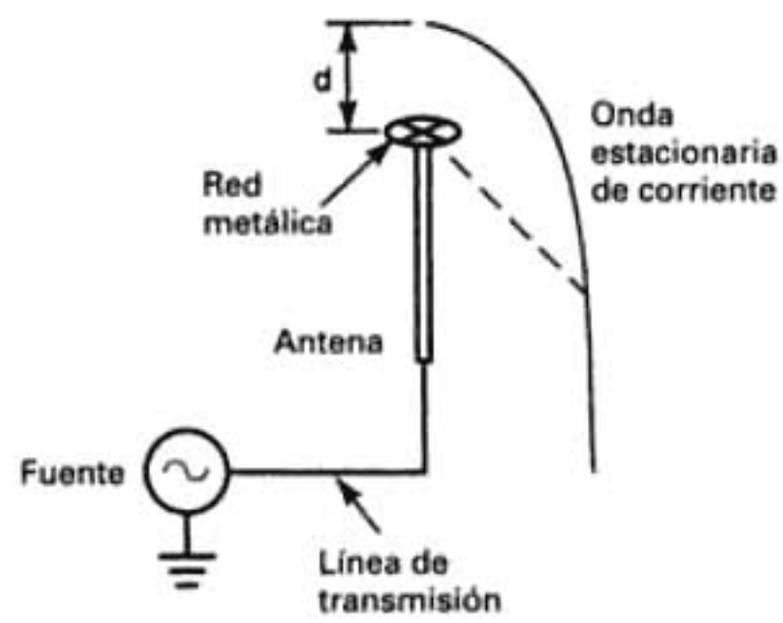


FIGURA 10-17 Antena cargada por la terminal

de *bobina de carga*. La bobina de carga anula y elimina bien el componente capacitivo de la impedancia de entrada de la antena. Así, la antena se ve como si fuera un circuito resonante, es resistiva y ya puede absorber el 100% de la potencia incidente. La fig. 10-16b muestra las pautas de onda estacionarias de corriente en una antena con una bobina de carga. La bobina de carga se suele poner en la parte inferior de la antena, para permitir que ésta se sintonice con facilidad con la resonancia. Una bobina de carga aumenta eficazmente la resistencia de radiación de la antena en unos 5 Ω. También, obsérvese que la onda estacionaria de corriente tiene valor máximo en la bobina, aumentando las pérdidas de potencia, creando condiciones de posible corona y reduciendo en forma efectiva la eficiencia de radiación de la antena.

Carga por el extremo superior

Las bobinas de carga tienen algunos inconvenientes que se pueden evitar con un método llamado *carga por el extremo superior* (o *por la terminal*) de la antena. En este caso se coloca un conjunto metálico que parece una rueda con rayos, en el extremo superior de la antena. La rueda aumenta la capacitancia en paralelo con el suelo y reduce la capacitancia general de la antena. En la fig. 10-17 se muestra la carga superior de una antena. Obsérvese que la onda estacionaria de corriente sube a lo largo de la antena, como si su longitud hubiera aumentado la distancia *d*, poniendo el máximo de corriente en la base. La carga por el extremo superior da como resultado un aumento considerable en la resistencia de radiación y la eficiencia de irradiación. También reduce el voltaje de la onda estacionaria en la base de la antena. Sin embargo, este método es muy impráctico para aplicaciones móviles.

La onda estacionaria de corriente se puede subir todavía más, aumentando todavía más la eficiencia de irradiación, si a la antena se le agrega un *techo plano*, *hoja* o *hilos horizontales*. Si

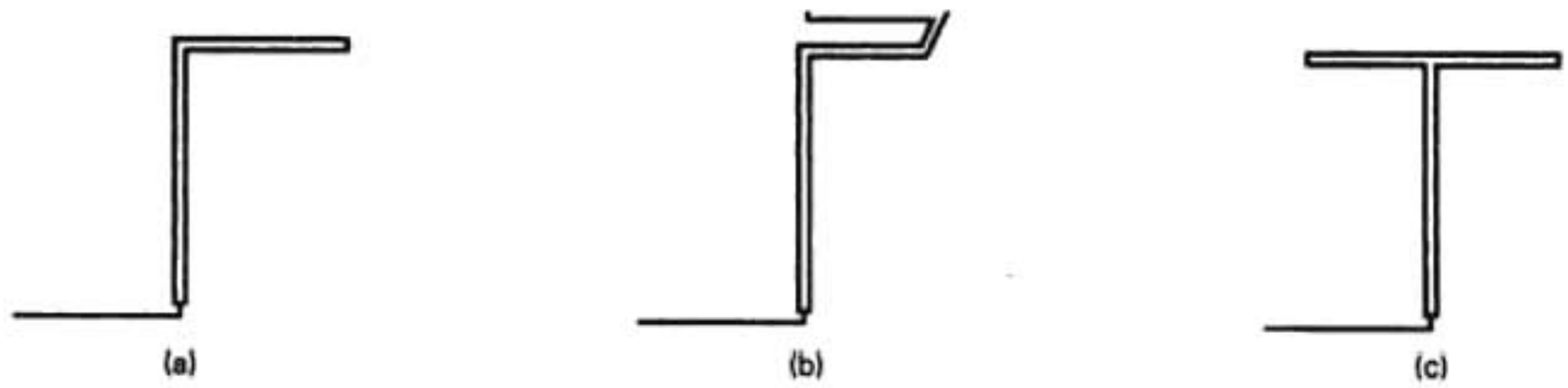


FIGURA 10-18 Carga de antena de hilos horizontales

se dobla una antena vertical en su parte superior para formar una L o una T, como se ve en la fig. 10-18, la onda de corriente estará más cerca del extremo superior del radiador. Si el techo plano y las partes verticales son de un cuarto de longitud de onda de largo cada uno, el máximo de la corriente estará en la parte superior del radiador vertical.

CONJUNTOS DE ANTENAS

Un *conjunto de antenas*, o *red de antenas*, se forma cuando se combinan dos o más elementos de antena para formar una sola antena. Un elemento de antena es un radiador individual, como por ejemplo, un dipolo de media o de un cuarto de onda. Los elementos se colocan físicamente en forma tal que sus campos de radiación interactúan y producen una distribución total de radiación que es la suma vectorial de los campos individuales. El objetivo de una red de antenas es aumentar la directividad de un sistema de antena y concentrar la potencia irradiada dentro de un área geográfica menor.

En esencia hay dos tipos de elementos de antena: los *excitados* y los *parásitos* (no excitados). Los elementos excitados se conectan en forma directa con la línea de transmisión y reciben la potencia de, o están excitados por, la fuente. Los elementos parásitos no se conectan a la línea de transmisión; sólo reciben energía mediante inducción mutua con un elemento excitado u otro elemento parásito. Un elemento parásito más largo que el elemento excitado del que recibe la energía se llama *reflector*. Un reflector reduce la intensidad de la señal en su dirección, y la aumenta en dirección contraria. En consecuencia, funciona como si fuera un espejo cóncavo. Esto se debe a que la onda que pasa por el elemento parásito induce un voltaje invertido 180° respecto a la onda que lo indujo. El voltaje inducido produce una corriente en fase, y el elemento la irradia (en realidad, reirradia la energía que acaba de recibir). La energía reirradiada establece un campo que se anula en una dirección y se refuerza en la otra. Un elemento parásito que es más corto que su elemento excitado asociado se llama *director*. Un director aumenta la intensidad de campo en su dirección, y la reduce en la dirección opuesta. En consecuencia, funciona como si fuera una lente convergente. Esto se ve en la fig. 10-19.

Red de radiación lateral

Una *red de radiación lateral* o *transversal* es uno de los tipos más sencillos de redes de antenas. Se forma tan sólo con poner varios dipolos resonantes de igual tamaño, tanto en longitud como en diámetro, en paralelo entre sí, y en línea recta (colineales). Todos los elementos se alimentan en fase desde la misma fuente. Como indica el nombre, una red de radiación lateral irradia en ángulo recto respecto al plano del conjunto, e irradia muy poco en dirección del plano. La fig. 10-20a muestra una red de radiación lateral formada por cuatro elementos excitados de media onda, separados por media onda. En consecuencia, la señal que irradia el elemento 2 ha recorrido media longitud de onda más lejos que la señal irradiada del elemento 1, es decir, las señales se irradian desfasadas 180°. Si se pone la línea de transmisión en zigzag de produce un corrimiento adicional de fase de 180°. Por lo anterior, las corrientes en todos los elementos están enfasadas y las señales irradiadas están enfasadas en forma aditiva en un plano que forma

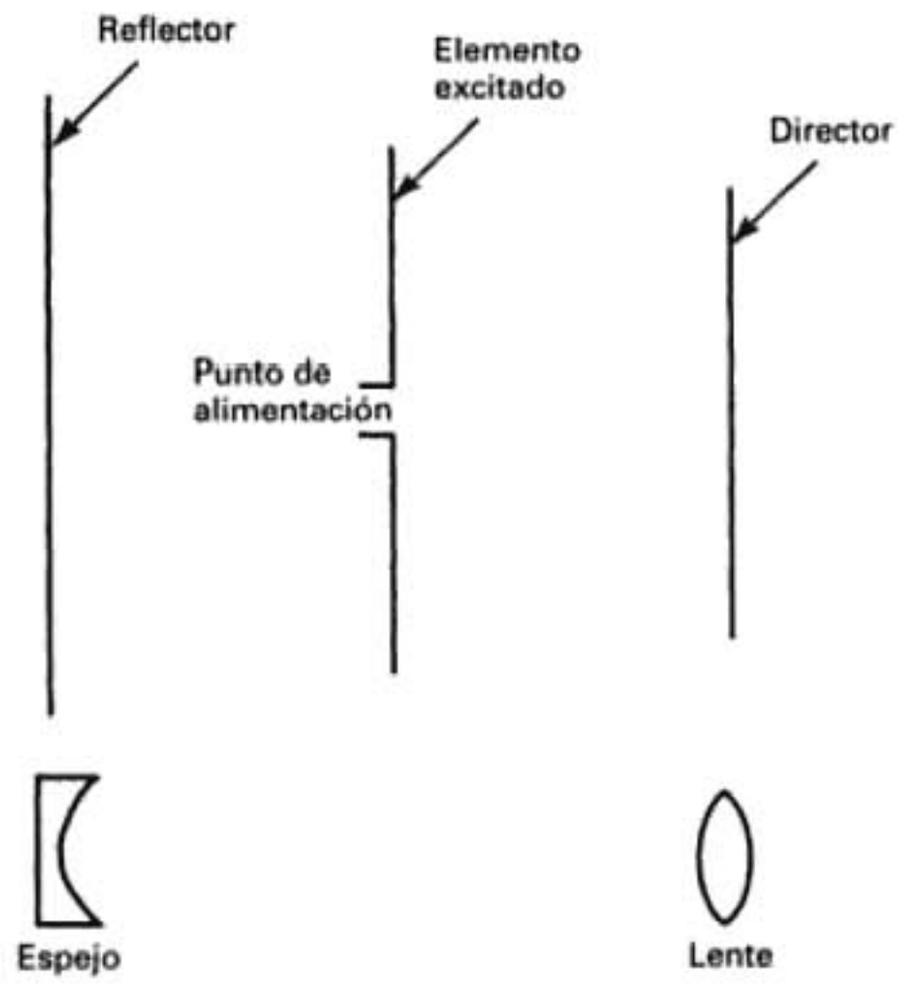


FIGURA 10-19 Red de antenas

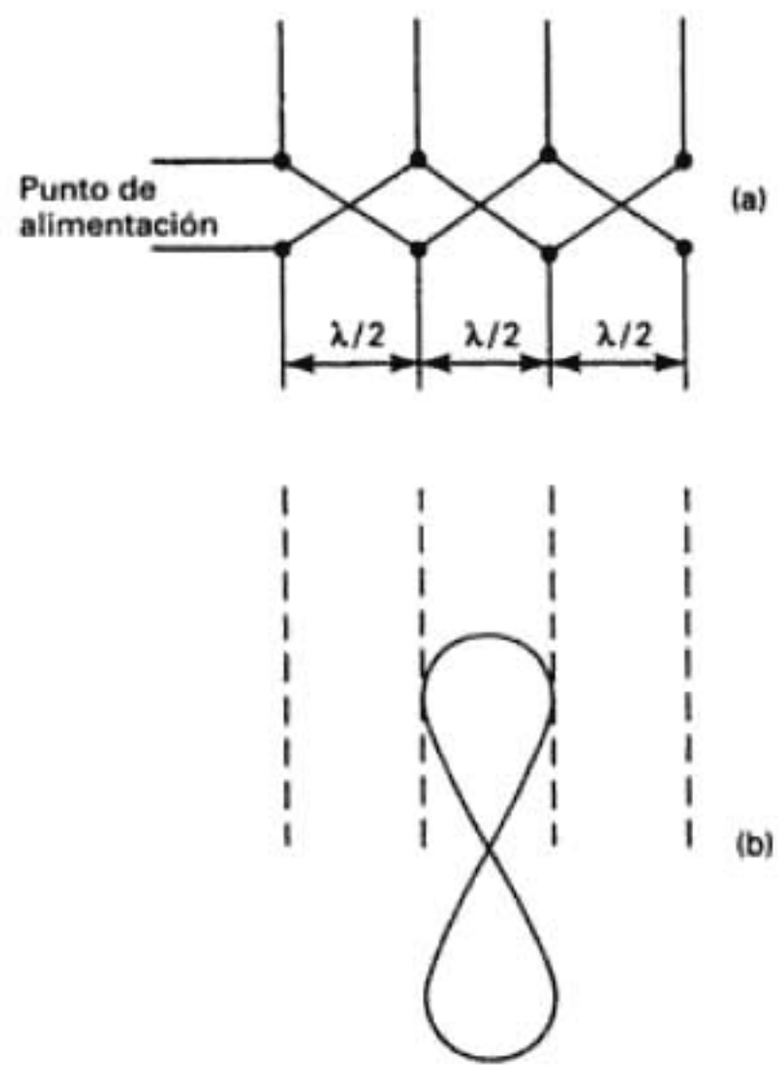


FIGURA 10-20 Antena de radiación lateral: (a) Red de radiación lateral (b) Gráfica de radiación

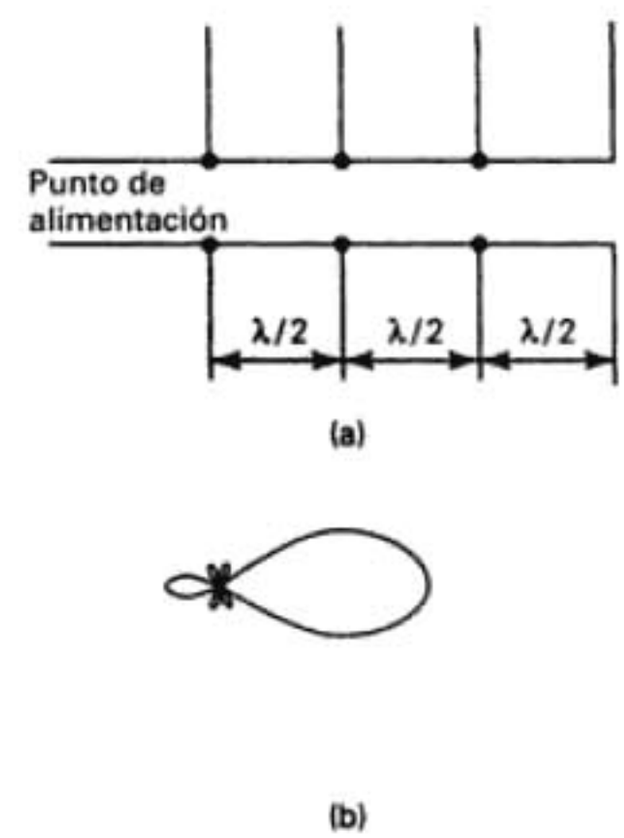


FIGURA 10-21 Antena de radiación longitudinal: (a) red de radiación longitudinal; (b) gráfica de radiación

ángulo recto con el plano de la red. Aunque de por sí la distribución horizontal de la radiación para cada elemento es omnidireccional, al combinarse, los campos producen una distribución bidireccional muy dirigida (10-20b). Se puede aumentar todavía más la directividad si se aumenta la longitud del conjunto, agregando más elementos.

Red de radiación longitudinal

Una *red de radiación longitudinal* tiene en esencia la misma configuración de elementos que la red de radiación lateral, pero la línea de transmisión no se conecta en zigzag entre los elementos. En consecuencia, los campos son aditivos y alineados con el plano del conjunto. La fig. 10-21 muestra una red de radiación longitudinal, y su gráfica resultante de radiación.

TABLA 10-1 Dipolo con directores y reflectores

Tipo	Ganancia de potencia (dB)	Amplitud de haz (°)	Relación frente a posterior (dB)
Radiador isotrópico	0	360	0
Media onda	2.16	80	0
1 elemento (director o reflector)	7.16	52	15 o 20
2 elementos (1 reflector + 1 director o 2 directores)	10.16	40	23
3 elementos (1 reflector + 2 directores)	12.06	36	24
4 elementos (1 reflector + 3 directores)	13.26	34	25
6 elementos (1 reflector + 5 directores)	14.56	30	27
10 elementos (1 reflector + 9 directores)	15.66	24	29.5

0.1 λ = espaciamiento de director
0.15 λ = espaciamiento de reflector
0.05 λ = cambio de longitud por elemento

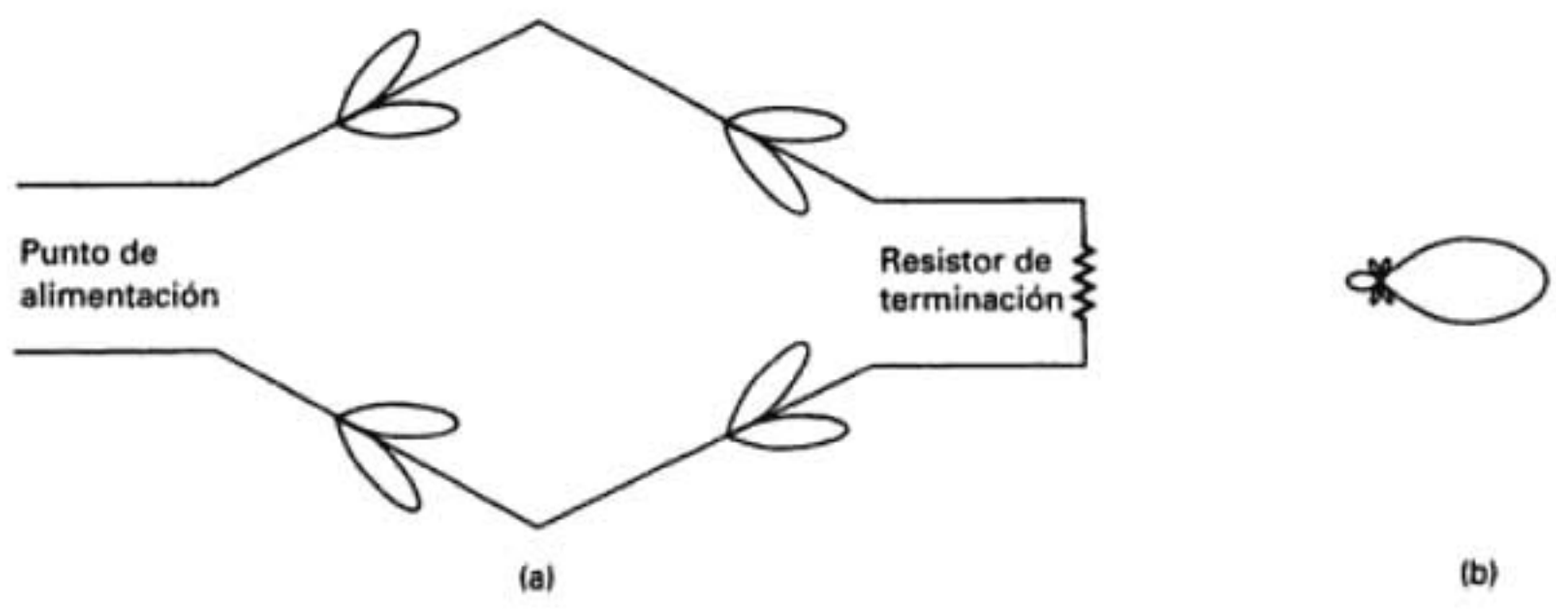


FIGURA 10-22 Antena de rombo: (a) red rómbica; (b) gráfica de radiación

La tabla 10-1 presenta los efectos de agregar directores y reflectores sobre la ganancia, ancho de haz y relación frente a posterior para un dipolo de media onda.

Conjunto no resonante: la antena rómbica

La *antena rómbica* es una antena no resonante capaz de funcionar en forma satisfactoria con un ancho de banda relativamente grande, por lo que es ideal para transmisión de HF (de 3 a 30 MHz). Está formada por cuatro elementos no resonantes, cada uno de varias longitudes de onda de largo. Todo el conjunto se termina en un resistor si se desea tener operación unidireccional. El arreglo que más se usa para la antena rómbica se asemeja a una línea de transmisión que se ha estrangulado en su parte media; se ve en la fig. 10-22. La antena se monta horizontalmente y se coloca a media longitud de onda, o más, sobre el piso. La altura exacta depende de la gráfica precisa de radiación que se desee. Cada conjunto de elementos funciona como línea de transmisión terminada en su impedancia característica; así, las ondas sólo se irradian hacia adelante. El resistor terminador absorbe más o menos la tercera parte de la potencia total de entrada a la antena. Entonces, una antena rómbica tiene aproximadamente 67% de eficiencia máxima. Se han alcanzado ganancias mayores que 40 (16 dB) con las antenas rómbicas.

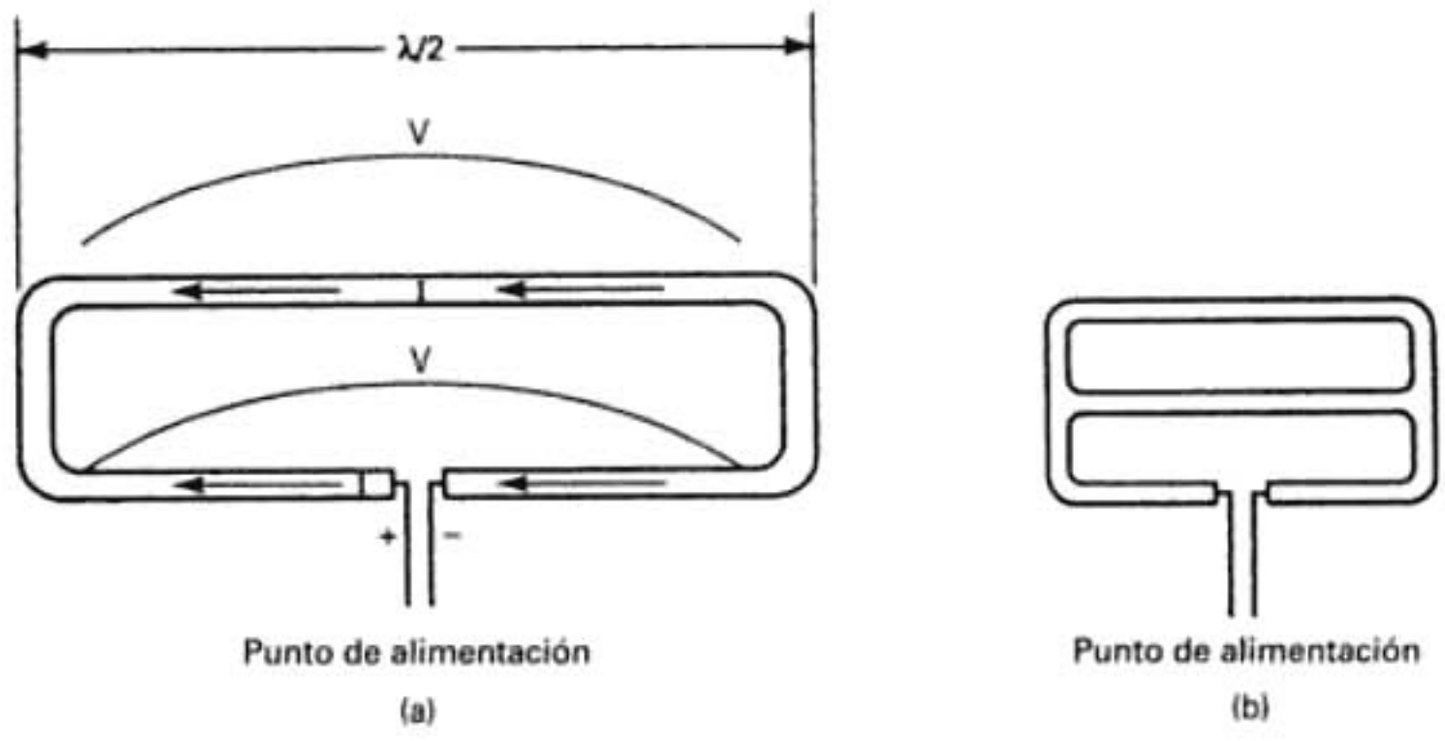


FIGURA 10-23 (a) Dipolo doblado; (b) dipolo doblado de tres elementos

ANTENAS DE USO ESPECIAL

Dipolo doblado

Un *dipolo doblado* de dos alambres, con su onda estacionaria de voltaje asociada, se muestra en la fig. 10-23a. El dipolo doblado es, en esencia, una antena única formada por dos elementos. Un elemento se alimenta en forma directa, mientras que el otro tiene acoplamiento inductivo en los extremos. Cada elemento tiene media longitud de onda de largo. Sin embargo, como puede pasar corriente por las esquinas, hay una longitud de onda completa de corriente en la antena. Así, para la misma potencia de entrada, la corriente de entrada es la mitad de la del dipolo básico de media onda, y la impedancia de entrada es cuatro veces mayor ($4 \times 72 = 288$). La impedancia de entrada de un dipolo doblado es igual a la impedancia de media onda (72 Ω) multiplicada por la cantidad de alambres doblados, elevada al cuadrado. Por ejemplo, si hay tres dipolos como se ve en la fig. 10-23b, la impedancia de entrada es $3^2 \times 72 = 648\ \Omega$. Otra ventaja del dipolo doblado sobre un dipolo básico de media onda es el mayor ancho de banda. Se puede aumentar todavía más el ancho de banda haciendo de mayor diámetro los elementos del dipolo (esa antena se llama *dipolo grueso* o *dipolo gordo*). Sin embargo, los dipolos gruesos tienen distribuciones de corriente y características de impedancia de entrada ligeramente distintas que los delgados.

Antena de Yagi-Uda. Una antena muy usada que usa con frecuencia un dipolo doblado como elemento excitado es la *antena de Yagi-Uda*, en honor de los dos científicos japoneses que la inventaron y describieron su funcionamiento. Esta antena se suele llamar *de Yagi*, simplemente. Es un conjunto lineal formado por un dipolo y dos o más elementos parásitos: un reflector y uno o más directores. En la fig. 10-24a se ve una antena Yagi sencilla de tres elementos. El elemento excitado es un dipolo doblado de media onda. A este elemento se le llama elemento excitado, porque está conectado con la línea de transmisión; sin embargo, se suele usar sólo para recibir. El reflector es una barra recta de aluminio, 5% más larga que el dipolo, y el director se corta un 5% más corto que el elemento excitado. La distancia entre los elementos suele ser de 0.1 a 0.2 longitudes de onda. La fig. 10-24b muestra la gráfica de radiación de una antena de Yagi. Su directividad típica es de 7 a 9 dB. Se puede aumentar el ancho de banda de una antena Yagi si se usa más de un dipolo doblado, cada uno cortado con longitud ligeramente distinta. En consecuencia, la antena de Yagi se suele usar para recibir TV VHF, por su gran ancho de banda: la banda de VHF TV está entre los 54 y los 216 MHz. La tabla 10-2 muestra las distancias entre elementos para conjuntos de Yagi con dos a ocho elementos.

Antena de torniquete

Una antena de torniquete se forma colocando dos dipolos en ángulo recto entre sí, desfasados 90°, como se ve en la fig. 10-25a. La pauta de radiación que se ve en la fig. 10-25b es la suma

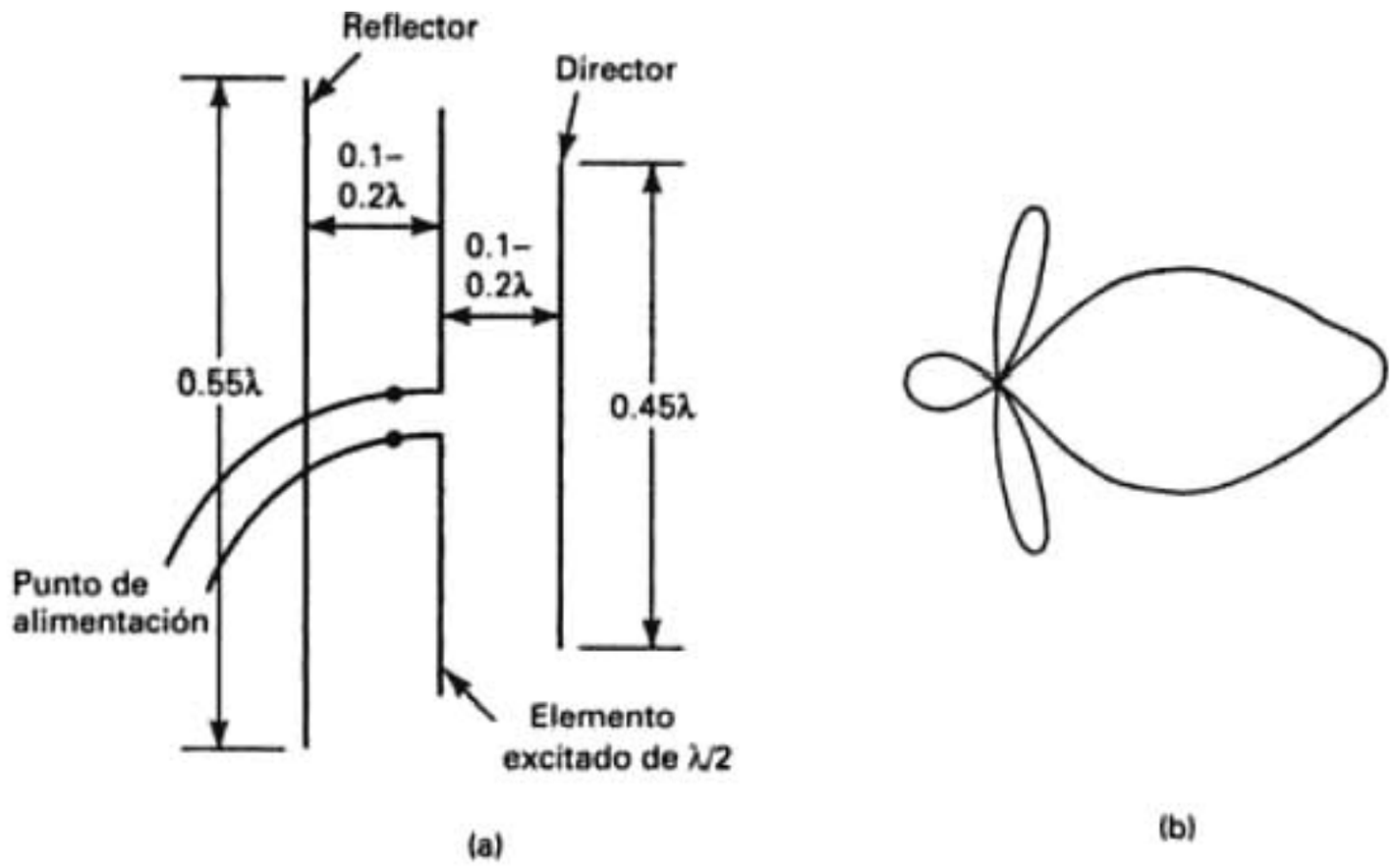


FIGURA 10-24 Antena de Yagi-Uda: (a) Yagi de tres elementos; (b) gráfica de radiación

TABLA 10-2 Distancia entre elementos para redes de Yagi (todas las unidades en longitud de onda, λ)

Espacio entre elementos	Cantidad de elementos 2	3	4	5	6	7	8
Reflector a elemento excitado	0.19	0.19	0.19	0.18	0.18	0.18	0.18
Director 1 a elemento excitado	—	0.17	0.16	0.16	0.16	0.16	0.15
Director 2 a director 1	—	—	0.16	0.18	0.20	0.21	0.22
Director 3 a director 2	—	—	—	0.20	0.25	0.30	0.30
Director 4 a director 3	—	—	—	—	0.28	0.28	0.29
Director 5 a director 4	—	—	—	—	—	0.30	0.30
Director 6 a director 5	—	—	—	—	—	—	0.35
Director 7 a director 6	—	—	—	—	—	—	—

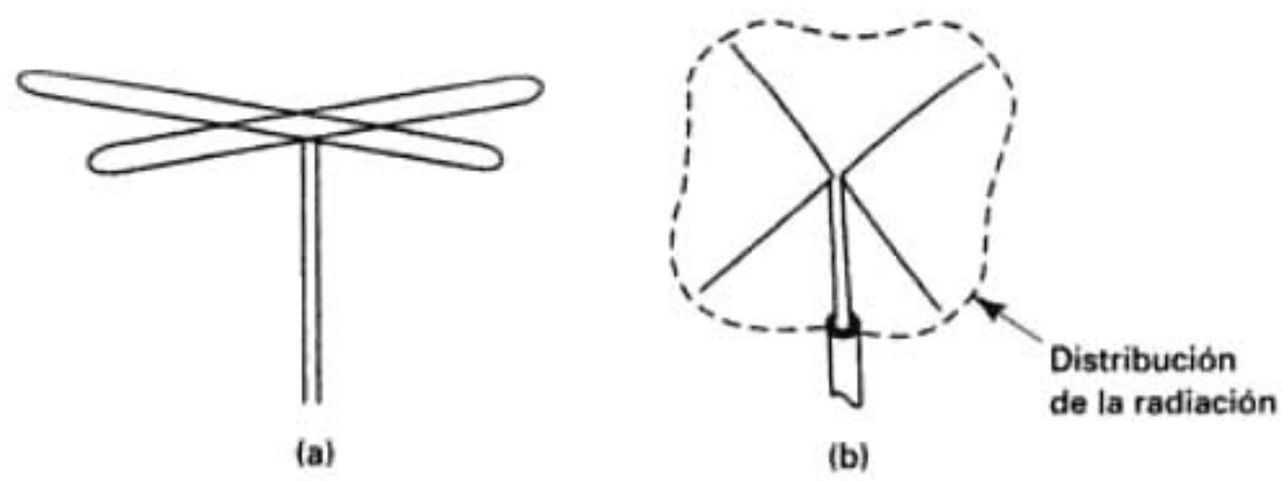


FIGURA 10-25 (a) Antena de torniquete; (b) distribución de la radiación

de las radiaciones de los dos dipolos, con lo cual se obtiene una distribución casi omnidireccional. Son frecuentes las ganancias de las antenas de torniquete de 10 o más dB.

Antena log-periódica

Una clase de antenas independientes de la frecuencia, llamadas *log-periódicas*, surgió del trabajo inicial de V. H. Rumsey, J. D. Dyson, R. H. DuHamel y D. E. Isbell en la Universidad de Illinois, en 1957. La ventaja principal de estas antenas es su independencia de resistencia de radiación y de la distribución de la radiación, respecto a la frecuencia. Las antenas log-periódicas tienen relaciones de ancho de banda de 10:1 o mayores. La relación de ancho de banda es el cociente de la frecuencia máxima entre la mínima de operación satisfactoria de una antena. Se usa la relación de ancho de banda con más frecuencia que la mención del porcentaje del ancho de banda entre la frecuencia central. Las antenas log periódicas no sólo son un tipo, sino más bien una clase de antenas porque tienen muchos tipos distintos, algunos de los cuales son bastante extraños. Las antenas log-periódicas pueden ser unidireccionales o bidireccionales, y tener una ganancia directiva de baja a moderada. También se pueden alcanzar altas ganancias usándolas como elementos de una red más complicada.

La estructura física de una antena log-periódica es repetitiva, y eso causa un comportamiento repetitivo de sus características eléctricas. En otras palabras, el diseño de una antena log-periódica consiste en una figura geométrica básica que se repite, pero con distintos tamaños. Un conjunto básico log-periódico de dipolos es quizá lo que más se acerca un periodo logarítmico a una antena convencional; se ve en la fig. 10-26. Consiste en varios dipolos de distinta longitud y distancia, que se alimentan de una sola fuente en el extremo pequeño. La línea de transmisión está en zigzag entre los puntos de alimentación de pares adyacentes de dipolos. La distribución de la radiación para una antena log-periódica básica tiene la radiación máxima alejándose del extremo pequeño. Las longitudes de los dipolos y su distancia se relacionan en tal forma que los elementos adyacentes tienen una relación constante entre sí. Las longitudes y las distancias entre los dipolos se relacionan con la fórmula

$$\frac{R_2}{R_1} = \frac{R_3}{R_2} = \frac{R_4}{R_3} = \frac{1}{\tau} = \frac{L_2}{L_1} = \frac{L_3}{L_2} = \frac{L_4}{L_3} \tag{10-18}$$

o sea

$$\frac{1}{\tau} = \frac{R_n}{R_{n-1}} = \frac{L_n}{L_{n-1}}$$

en la que R = distancia entre dipolos (pulgadas)
L = longitud del dipolo (pulgadas)
τ = relación de diseño (número menor que 1)

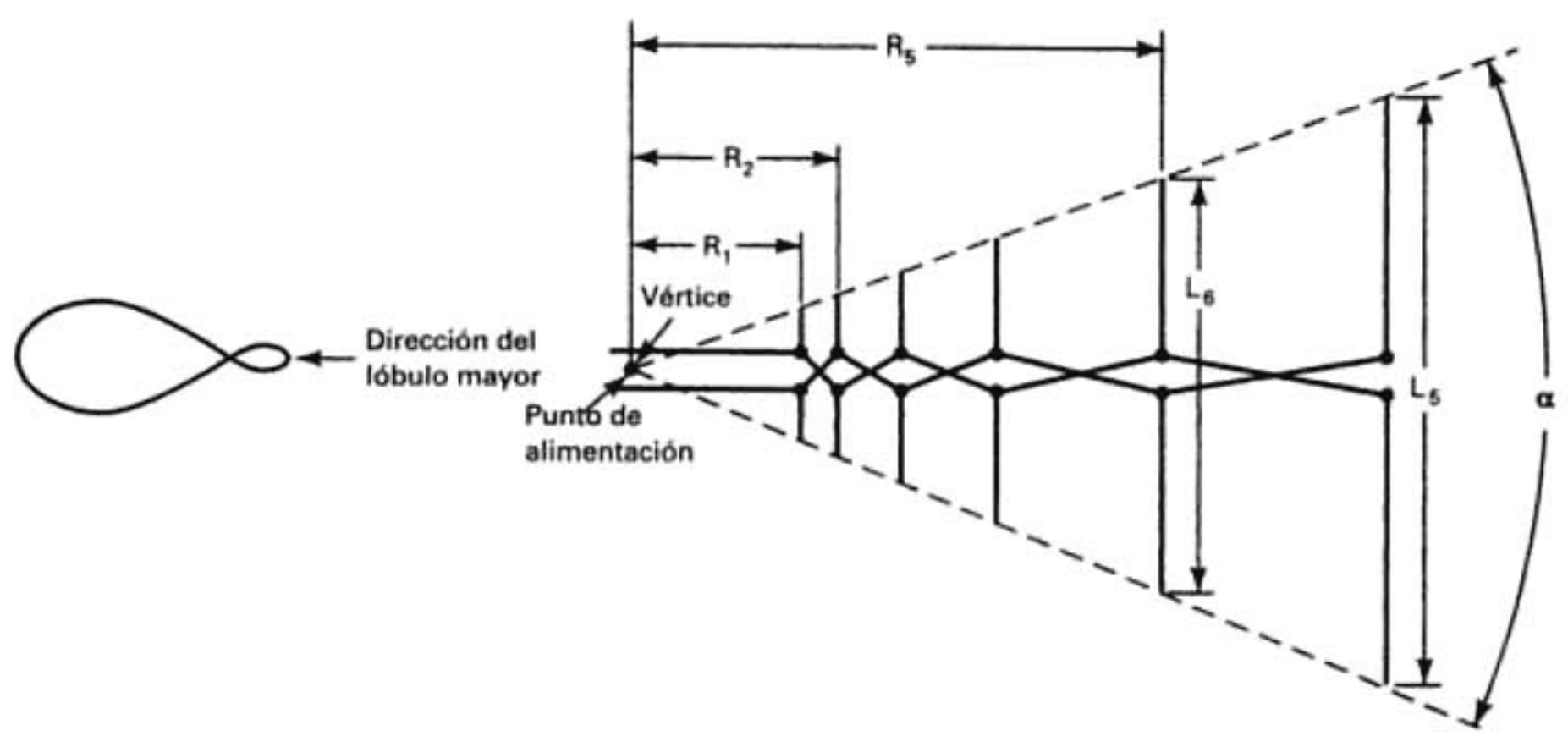


FIGURA 10-26 Antena log-periódica

Los extremos de los dipolos están en una línea recta, y el ángulo que forman se denomina α. En un diseño típico, $\tau = 0.7$ y $\alpha = 30°$. Con las anteriores estipulaciones estructurales, la impedancia de entrada de la antena varía en forma repetitiva cuando se grafica en función de la frecuencia, y cuando se grafica en función del logaritmo de la frecuencia, varía en forma periódica; de aquí el nombre "log-periódica". En la fig. 10-27 se ve una gráfica característica de la impedancia de entrada. Aunque esa impedancia varía en forma periódica, no necesariamente son senoidales las variaciones. También la distribución, directividad, ganancia de potencia y ancho de banda sufren una variación similar con la frecuencia.

La magnitud de un periodo logarítmico de frecuencia depende de la relación de diseño y si hay dos máximos sucesivos en las frecuencias f_1 y f_2, se relacionan mediante la fórmula

$$\log f_2 - \log f_1 = \log \frac{f_2}{f_1} = \log \frac{1}{\tau} \tag{10-19}$$

Por consiguiente, las propiedades medidas de una antena log-periódica a la frecuencia f tendrán propiedades idénticas a la frecuencia τf, $\tau^2 f$, $\tau^3 f$, etcétera. Las antenas log-periódicas, como las de rombo, se usan principalmente en comunicaciones HF y VHF. Sin embargo, las antenas log-periódicas no tienen un resistor de terminación y en consecuencia son más eficientes. Con mucha frecuencia, las antenas de TV anunciadas como de "alta ganancia" o de "alto rendimiento" son log-periódicas.

Antena de cuadro

La *antena de cuadro* más fundamental no es más que una bobina de una vuelta de alambre, bastante más corto que una longitud de onda, y conduce corriente de RF. Esa espira se ve en la fig. 10-28. Si el radio r es pequeño en comparación con una longitud de onda, la corriente está esencialmente enfasada por la espira. Se puede imaginar que una espira está formada por muchos dipolos elementales conectados entre sí. Los dipolos son rectos y, por consiguiente, la espira es en realidad un polígono, más que un círculo. Sin embargo, se puede tener un círculo aproximado si se supone que los dipolos son lo bastante cortos. La espira está rodeada por un campo magnético en ángulo recto con el alambre, y la pauta direccional es independiente de la forma exacta.

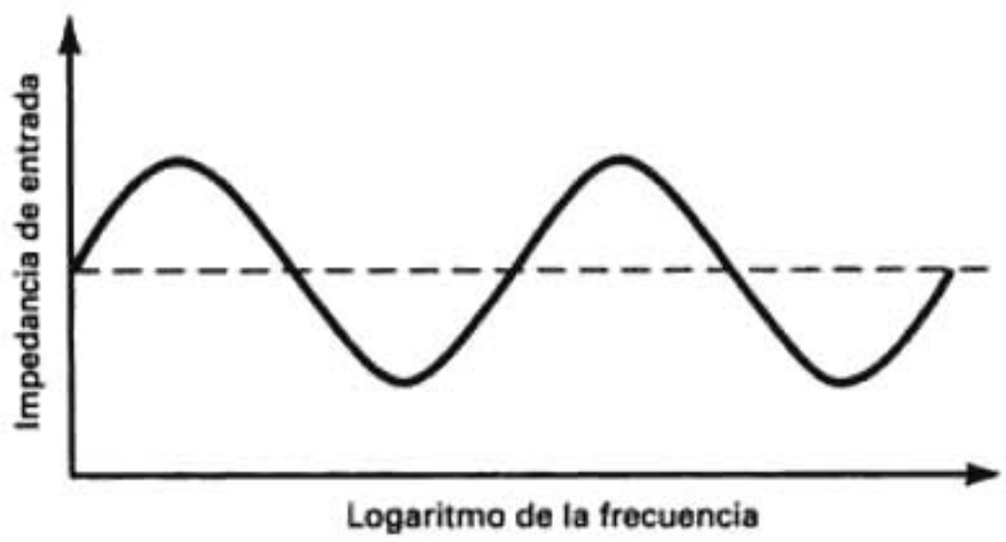


FIGURA 10-27 Impedancia de entrada log-periódica en función de la frecuencia

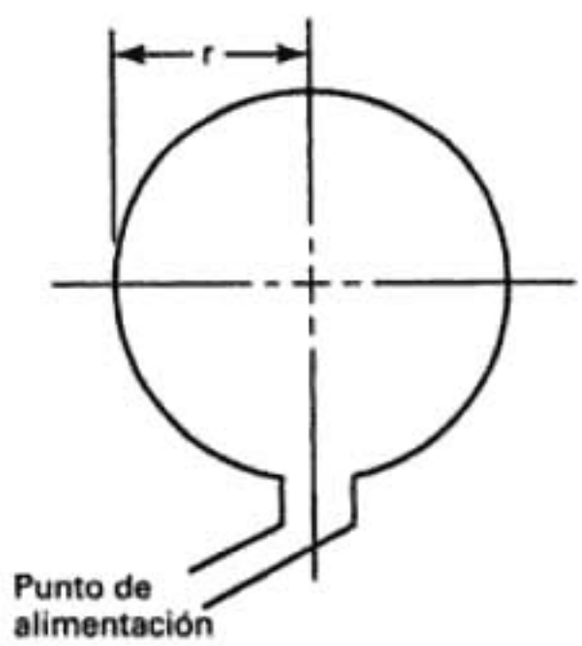


FIGURA 10-28 Antena de cuadro

En general, las espiras son circulares; sin embargo, sirve cualquier forma. La distribución de la radiación para una antena de cuadro es igual, en esencia, que la de un dipolo horizontal corto.

La resistencia de radiación para una espira pequeña es

$$R_r = \frac{31{,}200A^2}{\lambda^4} \tag{10-20}$$

en la que A es el área de la espira. Para aplicaciones en muy baja frecuencia, las espiras se hacen a menudo con más de una vuelta de alambre. La resistencia de radiación de un cuadro de varias vueltas no es más que la de una sola vuelta multiplicada por la cantidad de vueltas elevada al cuadrado. La polarización de una antena de cuadro, como la de un dipolo elemental, es lineal. Sin embargo, una espira vertical está polarizada verticalmente y una horizontal, polarizada horizontalmente.

Los cuadros pequeños polarizados verticalmente se usan mucho como antenas goniométricas. La dirección de la señal recibida se puede determinar orientando la espira hasta encontrar un valor nulo o cero. Es la dirección de la señal recibida. Los cuadros tienen una ventaja sobre casi todos los demás tipos de antenas en goniometría, porque en general son mucho menores y en consecuencia se adaptan con más facilidad a aplicaciones de comunicaciones móviles.

Antenas de conjunto enfasado

Una *antena de conjunto enfasado* es un arreglo de antenas, o un arreglo de redes de antenas que, cuando se conectan entre sí, funcionan como una sola antena cuyo ancho de banda y dirección (es decir, su gráfica de radiación) se puede cambiar en forma electrónica, sin tener que mover físicamente alguna de las antenas individuales o de los elementos de antena dentro del conjunto. La ventaja principal de las antenas de conjunto enfasado es que eliminan la necesidad de elementos de antena giratorios. En esencia, un conjunto enfasado es una antena cuya pauta de radiación se puede ajustar o cambiar en forma electrónica. La aplicación principal de los conjuntos enfasados es para radar, cuando se debe poder cambiar con rapidez las pautas de radiación para seguir un objeto en movimiento. Sin embargo, las agencias gubernamentales que transmiten señales de potencia extremadamente grande, para lugares remotos seleccionados en todo el mundo, como la Voz de América, también usan antenas de conjunto enfasado ajustable para dirigir sus transmisiones.

El principio básico de los conjuntos enfasados se basa en la interferencia entre las ondas electromagnéticas en el espacio libre. Cuando las energías electromagnéticas de diversas fuentes ocupan el mismo espacio al mismo tiempo, se combinan, a veces en forma constructiva (se suman entre sí) y a veces en forma destructiva (se oponen entre sí).

Hay dos tipos básicos de antenas de conjunto enfasado. En el primer tipo, un solo dispositivo de salida, de potencia relativamente alta, suministra la potencia de transmisión a una gran cantidad de antenas a través de un conjunto de divisores de potencia y desfasadores. Una intrincada combinación de atenuadores y demoras de tiempo ajustables determina cuánta de la potencia total de transmisión va a cada antena, así como la fase de la señal. La cantidad de pérdida en los atenuadores y el corrimiento de fase introducido en las demoras de tiempo se controlan con una computadora. Las demoras de tiempo pasan la señal de RF sin distorsionarla más que para proporcionarle una cantidad específica de demora (corrimiento de fase). El segundo tipo de antenas de conjunto enfasado usa más o menos la misma cantidad de dispositivos de salida variables, de baja potencia, que la cantidad de elementos radiadores, y la relación entre las fases de las señales de salida se controla con desfasadores. En ambos tipos de conjuntos enfasados, la distribución de la radiación se selecciona cambiando la demora de fase que introduce cada desfasador. La fig. 10-29 muestra una antena de conjunto enfasado, que emplea varios elementos idénticos de antena, cada uno con su demora ajustable de fase.

Antena helicoidal

Una *antena helicoidal* es una de VHF o de UHF, ideal para aplicaciones donde se requiere irradiar ondas electromagnéticas de polarización circular, más que de polarización horizontal o vertical. Una antena helicoidal se puede usar como antena de un solo elemento, o se puede apilar en dirección horizontal o vertical en un conjunto, para modificar su pauta de radiación, aumentando la ganancia y disminuyendo el ancho de banda del lóbulo primario.

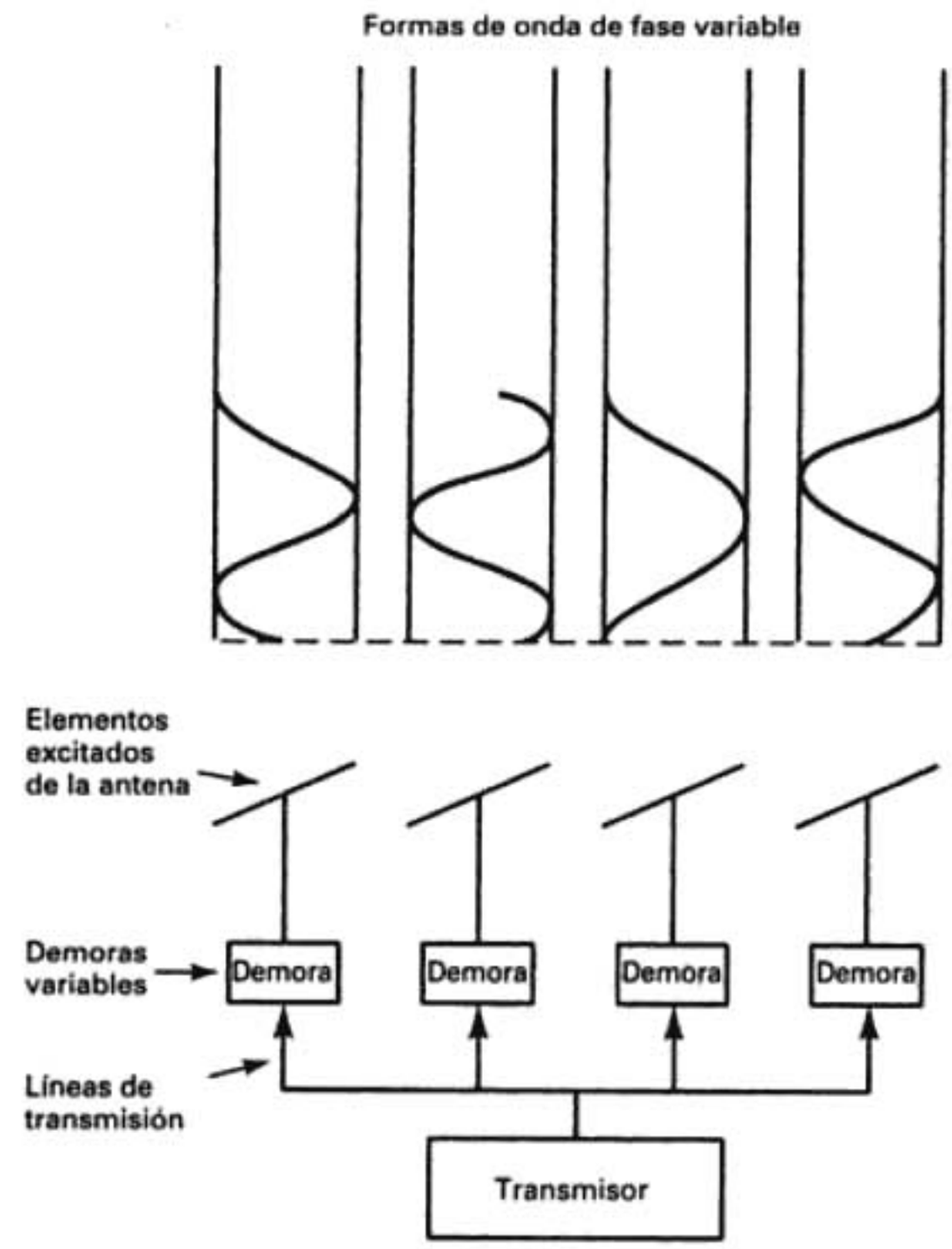


FIGURA 10-29 Antena de elementos enfasados

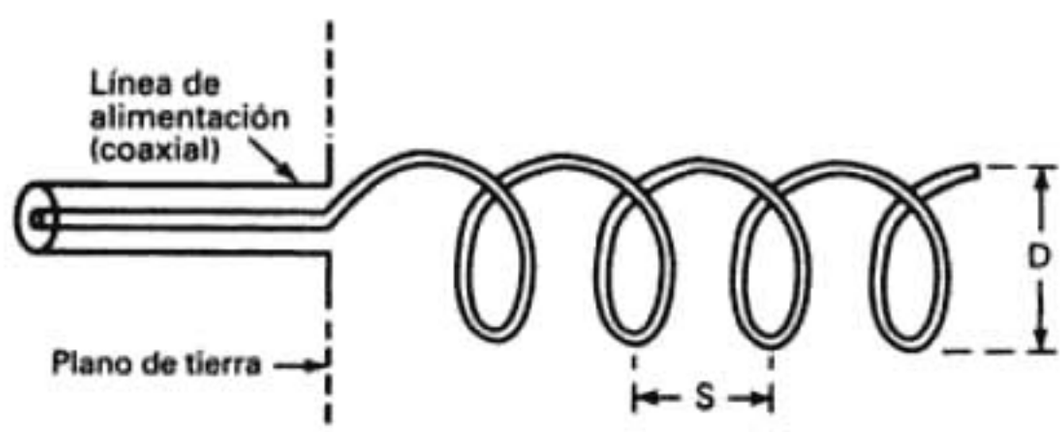


FIGURA 10-30 Antena helicoidal de radiación longitudinal

En la fig. 10-30 se ve una antena helicoidal básica de radiación longitudinal. El elemento excitado de la misma es una hélice rígida, con espiras espaciadas y con la longitud aproximada del eje igual al producto de la cantidad de vueltas por la distancia (paso) entre las vueltas. Una antena helicoidal se monta en un plano de tierra formado por metal macizo, o por una malla metálica como de tela de gallinero. Con una antena helicoidal hay dos modos de propagación: *normal* y *axial*. En el modo normal, la radiación electromagnética es en ángulo recto con el eje de la hélice. En el modo axial, la radiación va en dirección axial y produce una distribución de banda ancha y relativamente direccional. Si la circunferencia aproximada de la hélice es un cuarto de onda, las ondas viajeras se propagan por las espiras de la hélice e irradian una onda de polarización circular. Con las dimensiones de la fig. 10-30, las frecuencias dentro de ±20% de la frecuencia central producen una directividad casi de 25, y una abertura de haz de 90° entre ceros.

La ganancia de una antena helicoidal depende de varios factores, que incluyen el diámetro de la hélice, la cantidad de vuelas en la hélice, el paso o distancia entre vueltas, y la frecuencia de operación. La ecuación que define la ganancia de una antena helicoidal es

$$A_{p(\mathrm{dB})} = 10 \log\left[15\left(\frac{\pi D}{\lambda}\right)^2 \frac{(NS)}{\lambda}\right] \tag{10-21}$$

en donde $A_{p(\mathrm{dB})}$ = ganancia de potencia de la antena (dB)
D = diámetro de la hélice (metros)
N = cantidad de vueltas (entero positivo)
S = paso (metros)
λ = longitud de onda (metros por ciclo)

Una antena helicoidal típica tiene desde un mínimo de 3 o 4 hasta un máximo de 20 vueltas, y ganancias de potencia de 15 a 20 dB. Se puede determinar la abertura de haz de 3 dB para una antena helicoidal con la siguiente ecuación

$$\theta = \frac{52}{(\pi D/\lambda)(\sqrt{NS/\lambda})} \tag{10-22}$$

en la que θ = abertura del haz (grados)
D = diámetro de la hélice (metros)
N = cantidad de vueltas (entero positivo)
S = paso (metros)
λ = longitud de onda (metros por ciclo)

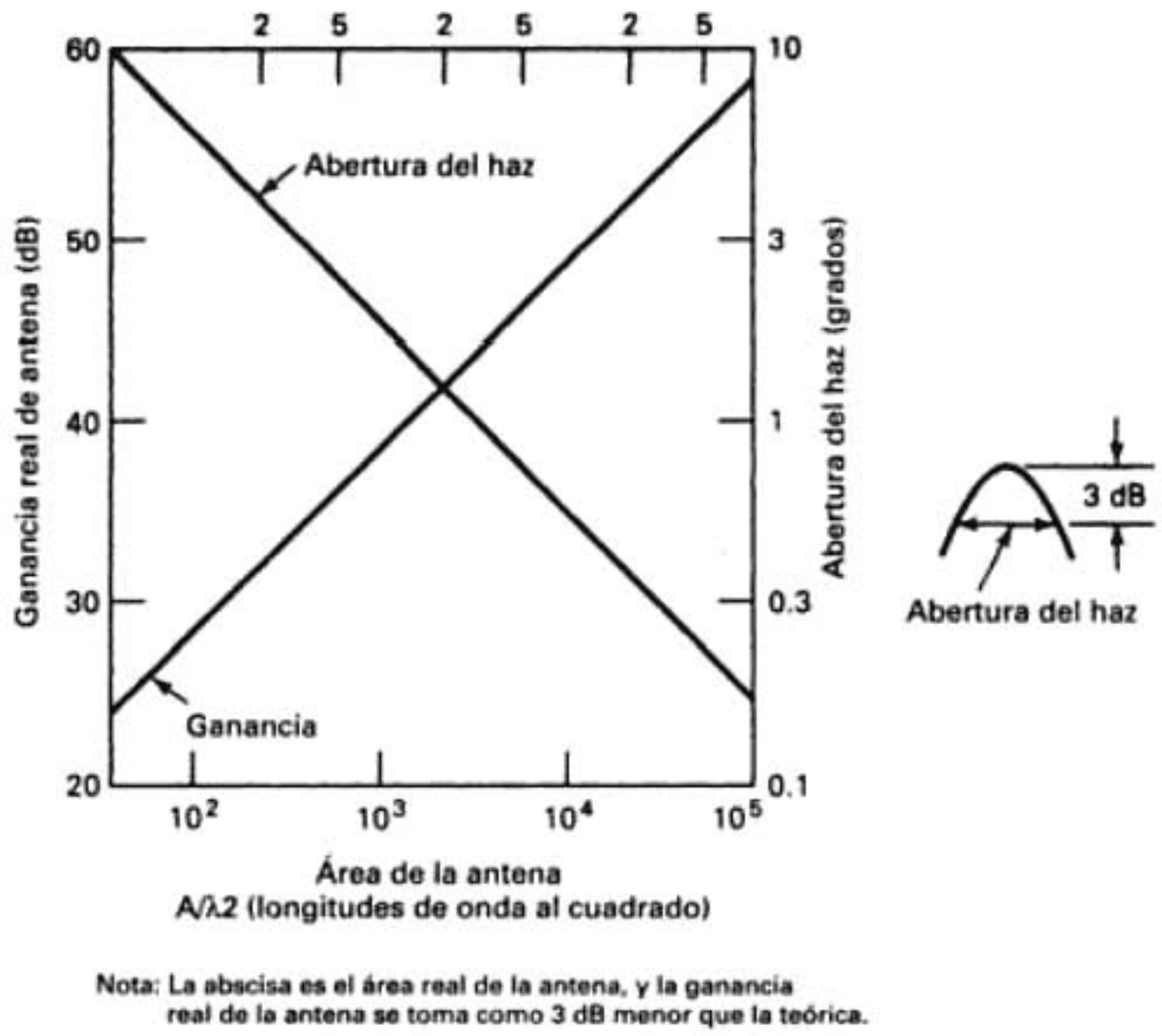
Se ve, en las ecuaciones 10-21 y 10-22, que para determinado diámetro y paso de una hélice, la ganancia de potencia aumenta en forma proporcional a la cantidad de vueltas, y que disminuye la abertura del haz. Las antenas helicoidales producen anchos de banda desde ±20% de la frecuencia central, hasta un intervalo de 2:1 entre las frecuencias máxima y mínima de operación.

ANTENAS DE UHF Y DE MICROONDAS

Las antenas para UHF (0.3 a 3 GHz) y para microondas (1 a 100 GHz) deben ser muy direccionales. Una antena tiene ganancia aparente porque concentra la potencia irradiada en un haz delgado, más que mandarlo por igual en todas direcciones; además, la abertura del haz disminuye al aumentar la ganancia de la antena. La relación entre el área, la ganancia y el ancho de banda de la antena se ve en la fig. 10-31. Las antenas de microondas suelen tener aberturas de haz a mitad de potencia del orden de 1° o menos. Un haz angosto minimiza los efectos de la interferencia debida a fuentes externas y a antenas adyacentes. Sin embargo, para transmisión por la visual, como la que se usa en la radio de microondas, un haz angosto impone varias limitaciones, como son estabilidad mecánica y desvanecimiento, que pueden causar problemas en el alineamiento de la antena.

Toda la energía electromagnética emitida por una antena de microondas no se irradia en la dirección del *lóbulo* (haz) *principal*; algo de ella se concentra en los *lóbulos menores*, llamados *lóbulos laterales*, que pueden ser fuentes de interferencia hacia o desde otras trayectorias de señal de microondas. La fig. 10-32 muestra la relación entre el haz principal y los lóbulos laterales, para una antena normal de microondas, como puede ser un reflector parabólico.

Las antenas de microondas tienen tres características importantes, que son la eficiencia direccional, acoplamiento lado a lado y acoplamiento espalda con espalda. La *eficiencia direccional*, o relación de frente a espalda de una antena, se define como la relación de su ganancia máxima en dirección delantera entre su ganancia máxima en dirección trasera. La eficiencia direccional de una antena en una instalación real puede ser menor 20 dB o más, que su valor aislado o de espacio libre, a causa de las reflexiones en el suelo del frente, de objetos en o cerca del lóbulo principal de transmisión. La eficiencia direccional de una antena de microondas es crítica en el diseño de un sistema de radio, porque las antenas de transmisión y recepción, en las estaciones repetidoras, se ubican con frecuencia opuestas entre sí, en la misma estructura (los sistemas de radio y repetidoras de microondas se describen con más detalle en el capítulo 17). Los *acoplamientos lado a lado* y *espalda con espalda* expresan, en decibelios, la pérdida de



Nota: La abscisa es el área real de la antena, y la ganancia real de la antena se toma como 3 dB menor que la teórica.

FIGURA 10-31 Relación entre la ganancia de potencia y la abertura del haz de una antena

acoplamiento entre antenas que conducen señales de salida de transmisión, y antenas cercanas que llevan señales de entrada de receptor. En forma típica, las potencias de salida de transmisor tienen una intensidad de 60 dB o más que las potencias de recepción; en consecuencia, las pérdidas de acoplamiento deben ser altas, para evitar que una señal de transmisión de una antena interfiera con una señal de recepción de otra antena.

Las antenas muy direccionales (de alta ganancia) se usan con los sistemas de microondas *de punto a punto*, o *entre puntos fijos*. Al enfocar la energía radioeléctrica en un haz angosto que se pueda dirigir hacia la antena receptora, la antena transmisora puede aumentar varios órdenes de magnitud la potencia efectiva irradiada, respecto a una antena no direccional. La antena receptora, en forma parecida a un telescopio, también puede aumentar la potencia efectiva recibida en una cantidad parecida. El tipo más común de antena de transmisión y recepción de microondas es el reflector parabólico.

Antena de reflector parabólico

Las *antenas de reflector parabólico* proporcionan ganancias y directividades extremadamente altas, y son muy usadas en los enlaces de comunicaciones por radio y satélite. Una antena parabólica consiste en dos partes principales: un *reflector parabólico* y el elemento activo, llamado *mecanismo de alimentación*. En esencia, el mecanismo de alimentación encierra la antena primaria, que normalmente es un dipolo o una red de dipolos; la antena irradia ondas electromagnéticas hacia el reflector. El reflector es un dispositivo pasivo, que tan sólo refleja la energía que le llega del mecanismo de alimentación. La reflexión produce una emisión muy concentrada y muy direccional, en la que todas las ondas individuales están enfasadas entre sí y, por consiguiente, un frente de onda enfasado.

Reflectores parabólicos. Es probable que el reflector parabólico sea el componente más fundamental de una antena parabólica. Los reflectores parabólicos se asemejan en forma a un plato o a una fuente, y en consecuencia se les llama a veces antenas de *plato parabólico* o simplemente antenas de *plato*. Para comprender cómo funciona un reflector parabólico es necesario primero comprender la geometría de una *parábola*. Una parábola es una curva en el plano que se describe matemáticamente como $y = ax^2$, y se define como el lugar geométrico de un punto que se mueve en forma tal que su distancia a otro punto (llamado *foco*), sumada a su

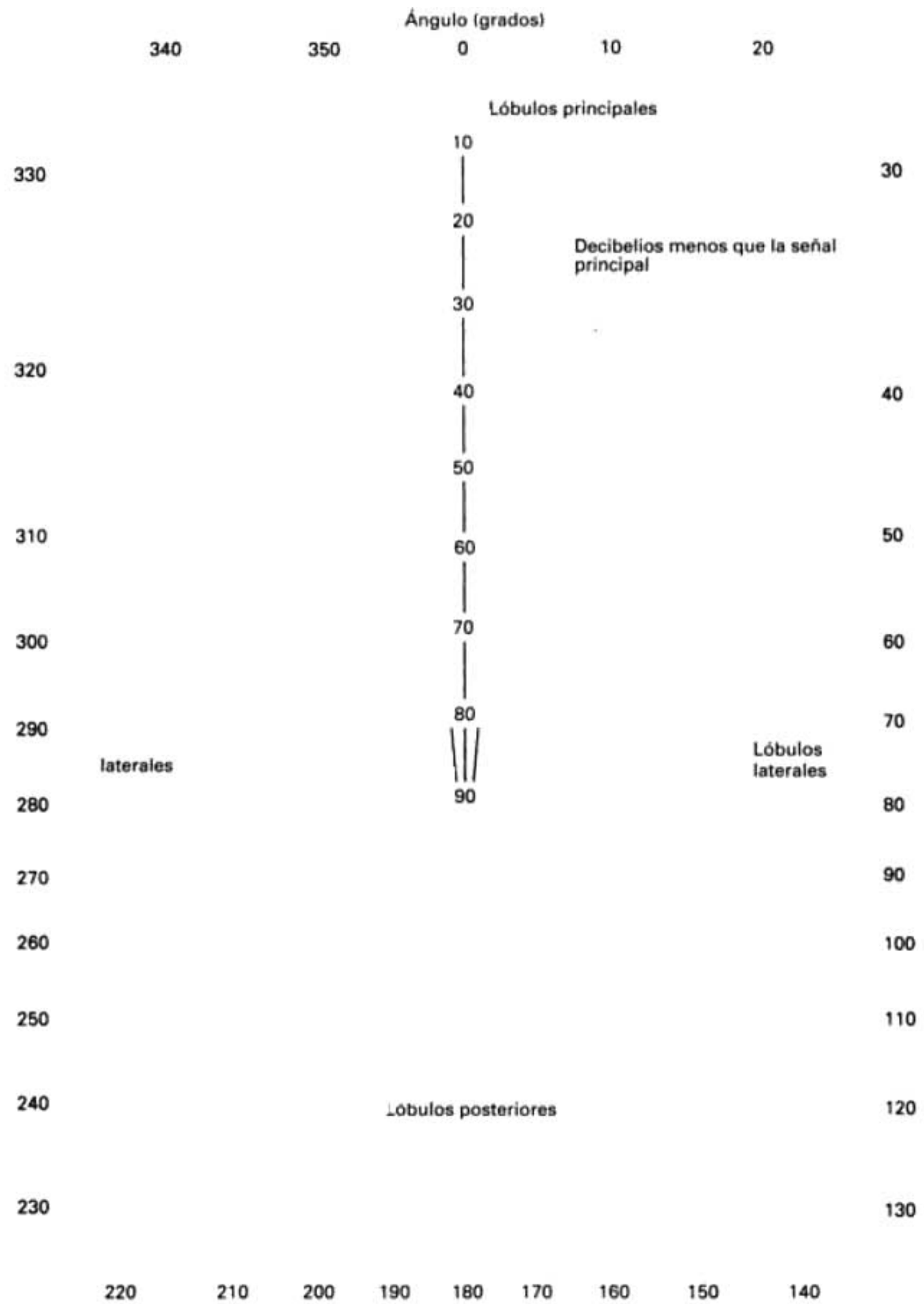


FIGURA 10-32 Haz principal y lóbulos laterales para una antena parabólica típica

distancia a una recta (llamada *directriz*) es una longitud constante. La fig. 10-33 muestra la geometría de una parábola cuyo foco está en el punto F y cuyo eje es la recta XY.

Para la parábola que se ve en la fig. 10-33 existen las siguientes relaciones

$$FA + AA' = FB + BB' = FC + CC' = k \quad \text{(una longitud constante)}$$

y además FX = distancia focal de la parábola (metros)
k = una constante para una parábola dada (metros)
WZ = longitud de la directriz (metros)

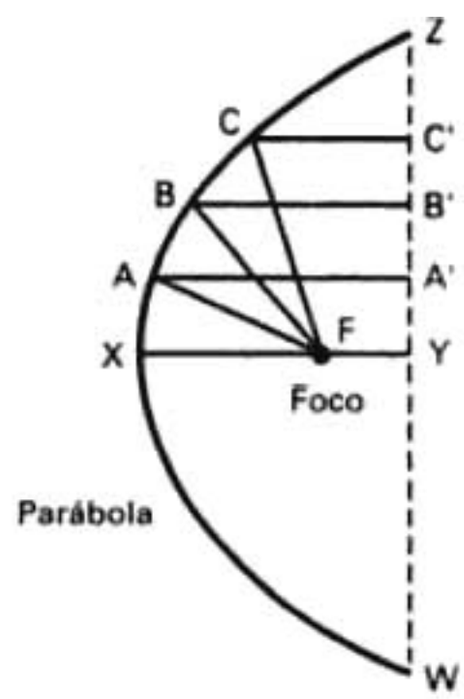


FIGURA 10-33 Geometría de una parábola

La relación de la distancia focal al diámetro de la boca de la parábola (*FX/WZ*) se llama *relación de abertura* o simplemente *abertura* de la parábola. Es el mismo término con que se describen los lentes de las cámaras. Un reflector parabólico se obtiene al girar la parábola en torno al eje *XY*. El plato de superficie curva que resulta se llama paraboloide. El reflector detrás del bulbo de una lámpara sorda o del faro de un coche tiene forma paraboloide para concentrar la luz en determinada dirección.

Una antena parabólica consiste en un reflector paraboloide iluminado por energía de microondas irradiada por un sistema alimentador ubicado en el foco. Si la energía electromagnética se irradia del foco hacia el reflector parabólico, todas las ondas irradiadas recorrerán la misma distancia cuando lleguen a la directriz (la recta *WZ*). En consecuencia, la radiación se concentra a lo largo del eje *XY* y hay anulación en otras direcciones. Un reflector parabólico que se use para recibir la energía electromagnética se comporta exactamente igual. Así, una antena parabólica posee el *principio de reciprocidad*, y funciona por igual como antena receptora para ondas que lleguen en la dirección *XY* (normales a la directriz). Los rayos recibidos de todas las demás direcciones se anulan en ese punto.

No es necesario que el plato tenga superficie metálica maciza para reflejar o recibir las señales con eficiencia. La superficie puede ser una malla, que refleja casi tanta energía como una superficie maciza, siempre que el ancho de las aberturas sea menor que 0.1 por la longitud de onda. Si se usa una malla en lugar de un conductor macizo se reduce en forma considerable el peso del reflector. También, los reflectores de malla son más fáciles de ajustar, se afectan menos por el viento y en general se obtiene con ellos una estructura mucho más estable.

Abertura del haz de una antena parabólica. La radiación tridimensional procedente de un reflector parabólico tiene un lóbulo principal que se asemeja a la forma de un puro grueso en la dirección *XY*. La abertura aproximada de haz de −3 dB para una antena parabólica, en grados, es

$$\theta = \frac{70\lambda}{D} \tag{10-23a}$$

o también

$$\theta = \frac{70c}{fD} \tag{10-23b}$$

en donde θ = abertura del haz entre puntos de mitad de potencia (grados)
λ = longitud de onda (metros)
$c = 3 \times 10^8$ metros por segundo
D = diámetro de la boca de la antena (metros)
f = frecuencia (hertz)

y $$\phi_0 = 2\theta \tag{10-24}$$

siendo ϕ_0 la abertura del haz (grados) entre ceros de la gráfica de radiación. Las ecuaciones 10-23a y b y 10-24 son exactas cuando se usan con antenas de grandes aberturas (es decir, aberturas de haz angostas).

Eficiencia de la antena parabólica (η). En un reflector parabólico, la reflectancia de la superficie del plato no es perfecta. En consecuencia, hay una pequeña parte de la señal irradiada del mecanismo de alimentación que se absorbe en la superficie del plato. Además, la energía cercana a las orillas del plato no se refleja, sino más bien se refracta en torno a la orilla. A esto se le llama *fuga* o *derrame*. Por las imperfecciones dimensionales sólo se refleja en el paraboloide del 50 al 75% de la energía que emite el mecanismo de alimentación. También, en una antena real, el mecanismo de alimentación no es una fuente puntual; ocupa un área finita frente al reflector, y obstruye un área pequeña en el centro del plato, formando una zona de sombra frente a la antena, que no es capaz ni de reunir ni de enfocar la energía. Estas imperfecciones contribuyen a que la eficiencia normal de una antena parabólica sólo sea de un 55% ($\eta = 0.55$). Esto es, sólo el 55% de la energía irradiada por el mecanismo de alimentación se propaga en realidad en forma de un haz concentrado.

Ganancia de potencia de una antena parabólica. Para una antena parabólica de transmisión, la ganancia de potencia se aproxima como sigue

$$A_p = \eta\left(\frac{\pi D}{\lambda}\right)^2 \tag{10-25a}$$

en la que A_p = ganancia de potencia con respecto a una antena isotrópica (adimensional)
D = diámetro de la boca de un reflector parabólico (metros)
η = eficiencia de la antena: potencia irradiada por la antena en relación con la potencia irradiada por el mecanismo de alimentación (adimensional)
λ = longitud de onda (metros por ciclo)

Para una eficiencia normal de antena de 0.55% ($\eta = 0.55$), la ecuación 10-25a se reduce a

$$A_p = \frac{5.4D^2f^2}{c^2} \tag{10-25b}$$

en la que c es la velocidad de propagación, 3×10^8 m/s. En forma de decibelios,

$$A_{p(\mathrm{dB})} = 20 \log f_{(\mathrm{MHz})} + 20 \log D_{(\mathrm{m})} - 42.2 \tag{10-25c}$$

en donde A_p = ganancia de potencia con respecto a una antena isotrópica (decibelios)
D = diámetro de la boca de un reflector parabólico (metros)
f = frecuencia (megahertz)
42.2 = constante (decibelios)

Para una eficiencia de antena de 100%, se agregan 2.66 dB al valor calculado con la ecuación 10-25c.

De acuerdo con las ecuaciones 10-25a, b y c, la ganancia de potencia de una antena parabólica es inversamente proporcional a la longitud de onda elevada al cuadrado. En consecuencia, el área (el tamaño) del plato es un factor importante en el diseño de antenas parabólicas. Con mucha frecuencia, el área del reflector mismo se cita en longitudes de onda al cuadrado, en lo que a veces se llama el área *eléctrica* o *efectiva* del reflector. Mientras mayor es el área, la relación de área a longitud de onda es mayor, y mayor es la ganancia de potencia.

Para una antena parabólica receptora, la superficie del reflector, de nuevo, no está totalmente iluminada y se reduce en realidad el área de la antena. En una antena receptora parabólica se llama *área de captura* y siempre es menor que el área real de la boca. El área de captura se

puede calcular comparando la potencia recibida con la densidad de potencia de la señal que se recibe. El área de captura se expresa en la forma siguiente

$$A_c = kA \tag{10-26}$$

en donde A_c = área de captura (metros cuadrados)
A = área real (metros cuadrados)
k = eficiencia de abertura, una constante que depende del tipo de antena y de su configuración (aproximadamente, 0.55 para un paraboloide alimentado con un dipolo de media onda)

Así, la ganancia de potencia para una antena parabólica receptora es

$$A_p = \frac{4\pi A_c}{\lambda^2} = \frac{4\pi kA}{\lambda^2} \tag{10-27a}$$

Sustituyendo en la ecuación 10-27a el área de la boca de un paraboloide, la ganancia de una antena parabólica de recepción, con eficiencia $\eta = 0.55$, se puede calcular con mucha aproximación como

$$A_p = 5.4\left(\frac{D}{\lambda}\right)^2 \tag{10-27b}$$

en donde D = diámetro del plato (metros)
λ = longitud de onda (metros por ciclo)

En decibelios,

$$A_{p(\mathrm{dB})} = 10 \log\left[5.4\left(\frac{D}{\lambda}\right)^2\right] \tag{10-27c}$$

El término k de la ecuación 10-26 se llama *eficiencia de abertura* (o a veces, *eficiencia de iluminación*). En la eficiencia de abertura se tienen en cuenta la distribución de la radiación del radiador primario, y también el efecto que introduce la relación de la distancia focal de la antena entre el diámetro del reflector, *f/D*. A esta relación se le llama *número de abertura*. Ese número determina la abertura angular del reflector, que en forma indirecta determina a su vez cuánta de la radiación primaria se refleja en el plato parabólico. La fig. 10-34 muestra direcciones de radiación para reflectores parabólicos (a) cuando el foco está fuera del reflector, y (b) cuando el foco está dentro del reflector.

La ganancia de potencia de transmisión, calculada con la ecuación 10-25c, y la ganancia de potencia de la antena de recepción, calculada con la ecuación 10-27c, producen los mismos resultados aproximados para una antena dada, con lo que se demuestra la reciprocidad de las antenas parabólicas.

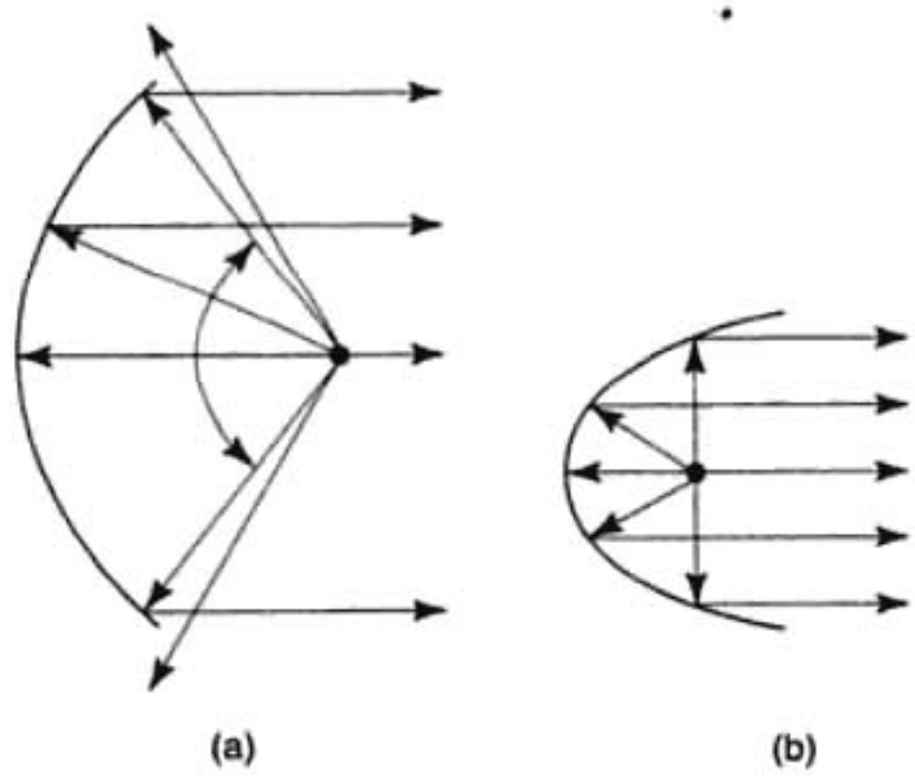


FIGURA 10-34 Direcciones de radiación en reflectores parabólicos: (a) foco fuera del reflector; (b) foco dentro del reflector

La gráfica de radiación que muestra la fig. 10-32 es característica de antenas tanto de transmisión como de recepción. La ganancia de potencia dentro del lóbulo principal es, en forma aproximada, unos 75 dB mayor que en la dirección hacia atrás, y casi 65 dB más que la ganancia máxima del lóbulo lateral.

Ejemplo 10-5

Calcular lo siguiente, para un reflector parabólico de 2 m de diámetro, con 10 W de potencia irradiada por el mecanismo de alimentación, que funciona a 6 GHz con una eficiencia de transmisión de la antena de 55%, y eficiencia de abertura de 55%:

(a) Abertura del haz.
(b) Ganancia de potencia de transmisión.
(c) Ganancia de potencia de recepción.
(d) Potencia isotrópica efectiva irradiada (EIRP, de *effective isotropic radiated power*).

Solución (a) La abertura del haz se calcula sustituyendo valores en la ecuación 10-23b

$$\theta = \frac{70(3 \times 10^8)}{(6 \times 10^9)(2)} = 1.75°$$

(b) La ganancia de potencia de transmisión se calcula con la ecuación 10-25c

$$A_{p(\mathrm{dB})} = 20 \log 6000 + 20 \log 2 - 42.2 = 39.4 \text{ dB}$$

(c) La ganancia de potencia de recepción se calcula con la ecuación 10-27c

$$\lambda = \frac{c(\mathrm{m/s})}{\text{frecuencia (Hz)}} = \frac{3 \times 10^8}{6 \times 10^9} = 0.05 \text{ m/ciclo}$$

$$A_{p(\mathrm{dB})} = 10 \log\left[5.4\left(\frac{2}{0.05}\right)^2\right] = 39.4 \text{ dB}$$

(d) La EIRP es el producto de la potencia irradiada por la ganancia de la antena de transmisión, en decibelios

$$\begin{aligned} \text{EIRP} &= A_{p(\mathrm{dB})} + P_{\text{irradiada(dBm)}} \\ &= 39.4 + 10 \log \frac{10}{0.001} \\ &= 39.4 \text{ dB} + 40 \text{ dBm} = 79.4 \text{ dBm} \end{aligned}$$

Mecanismos alimentadores. El mecanismo de alimentación en una antena parabólica es el que irradia, en realidad, la energía electromagnética, y por eso se le llama a menudo *antena primaria*. Este mecanismo de alimentación es de importancia básica, porque su función es irradiar la energía hacia el reflector. Un mecanismo ideal de alimentación debería dirigir toda su energía hacia el reflector parabólico, sin tener efecto de sombra. En la práctica es imposible lograr lo anterior, aunque si se tiene cuidado al diseñarlo, la mayor parte de la energía se puede irradiar con la dirección adecuada, y se puede reducir la sombra al mínimo. Hay tres tipos principales de antenas parabólicas: alimentación central, alimentación cónica y alimentación de Cassegrain.

Alimentación central. La fig. 10-35 muestra un diagrama de un reflector paraboloide alimentado en el centro, con un *reflector esférico* adicional. La antena primaria se coloca en el foco. La energía irradiada hacia el reflector se refleja hacia afuera, en forma de un haz concentrado. Sin embargo, la energía que no refleja el paraboloide se reparte en todas direcciones, y tiene la tendencia a perturbar la distribución general de la radiación. El reflector esférico redirige esas emisiones, hacia atrás, hacia el reflector parabólico, donde se vuelven a reflejar en la dirección correcta. Aunque el reflector esférico adicional contribuye a concentrar más energía en la dirección deseada, también tiene la tendencia de bloquear algunas de las reflexiones iniciales. En consecuencia la ventaja que logra se anula algo debido a su propio efecto de sombra, y su eficiencia general sólo es un poco mejor que cuando no se agrega el reflector esférico.

Alimentación por bocina. La fig. 10-36a muestra el diagrama de un reflector parabólico que usa alimentación por bocina, o *por cuerno*. Con un mecanismo de alimentación por bocina, la antena primaria es una antena pequeña cónica, o de embudo, y no un dipolo simple o una red de dipolos. La bocina no es más que un material abocardado de guía de ondas, que se coloca en el foco e irradia una distribución algo direccional, hacia el reflector parabólico.

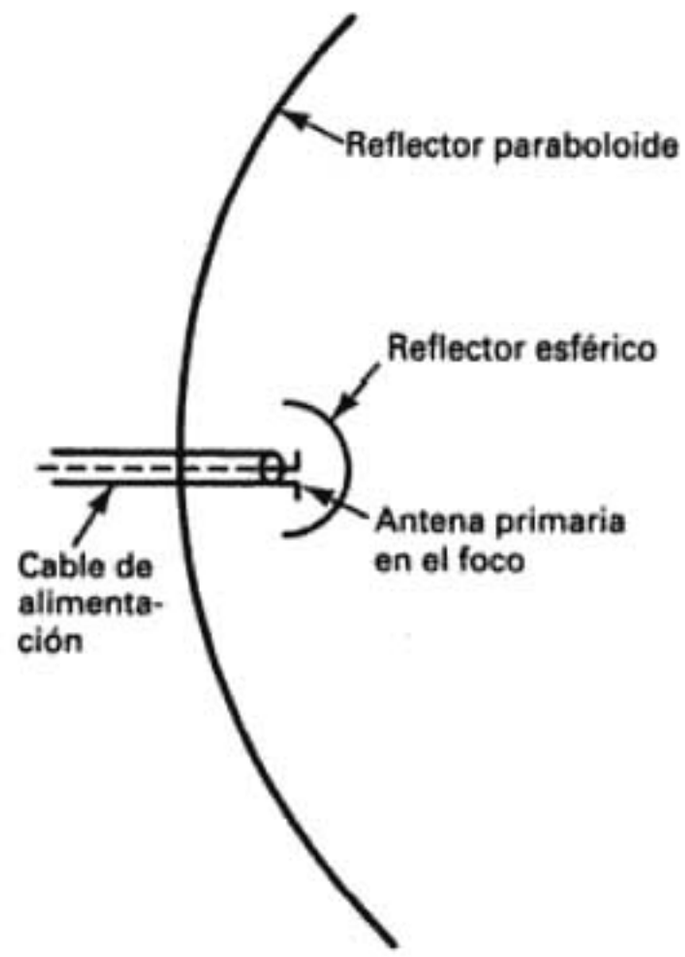


FIGURA 10-35 Antena parabólica con alimentación central

Cuando un campo electromagnético que se propaga llega a la boca del cuerno, continúa propagándose en la misma dirección general, pero de acuerdo con el principio de Huygens, se reparte en dirección lateral y al final, el frente de onda se hace esférico. La estructura cónica puede tener varias formas diferentes, como se ve en la fig. 10-36b: sectorial (abocinado sólo en una dirección), piramidal o cónica. Como en la alimentación central, una alimentación cónica representa cierta obstrucción a las ondas que se reflejan del plato parabólico.

La abertura del haz de una bocina en un plano que contiene el eje de guía es inversamente proporcional a la abertura de la boca de la bocina en ese plano. Unas fórmulas aproximadas para calcular las aberturas de haz de media potencia, de cuernos con abocinado óptimo y en los planos E y H, son las siguientes

$$\theta_E = \frac{56\lambda}{d_E} \tag{10-28a}$$

$$\theta_H = \frac{56\lambda}{d_H} \tag{10-28b}$$

en donde θ_E = abertura del haz de media potencia en el plano E (grados)
θ_H = abertura del haz de media potencia en el plano H (grados)
λ = longitud de onda (metros)
d_E = dimensión de la boca en el plano E (metros)
d_H = dimensión de la boca en el plano H (metros)

Alimentación de Cassegrain. Esta alimentación recibe el apellido de un astrónomo del siglo XVIII, y evolucionó en forma directa de los telescopios astronómicos ópticos. La fig. 10-37 muestra la geometría básica de un mecanismo de alimentación de Cassegrain. La fuente de radiación primaria está en o justo atrás de una pequeña abertura en el vértice del paraboloide, y no en el foco. La antena primaria se apunta hacia un pequeño reflector secundario (el *subreflector Cassegrain*) que está entre el vértice y el foco.

Los rayos emitidos por la antena primaria se reflejan en el subreflector de Cassegrain y a continuación iluminan el reflector parabólico principal, como si se hubieran originado en el foco. Los rayos son colimados por el reflector parabólico de la misma forma que en los mecanismos de alimentación central y por bocina. El subreflector debe tener curvatura hiperbólica para reflejar los rayos de la antena primaria en tal forma que funcione como una *fuente virtual* en el foco del paraboloide. La alimentación de Cassegrain se suele usar para recibir señales extremadamente débiles, o cuando se requieren líneas de transmisión o tramos de guías de onda extremadamente largos, y es necesario colocar preamplificadores de bajo ruido tan cerca de la antena como sea posible. Con la alimentación de Cassegrain, los preamplificadores se pueden instalar justo antes del mecanismo de alimentación, sin obstruir las ondas reflejadas.

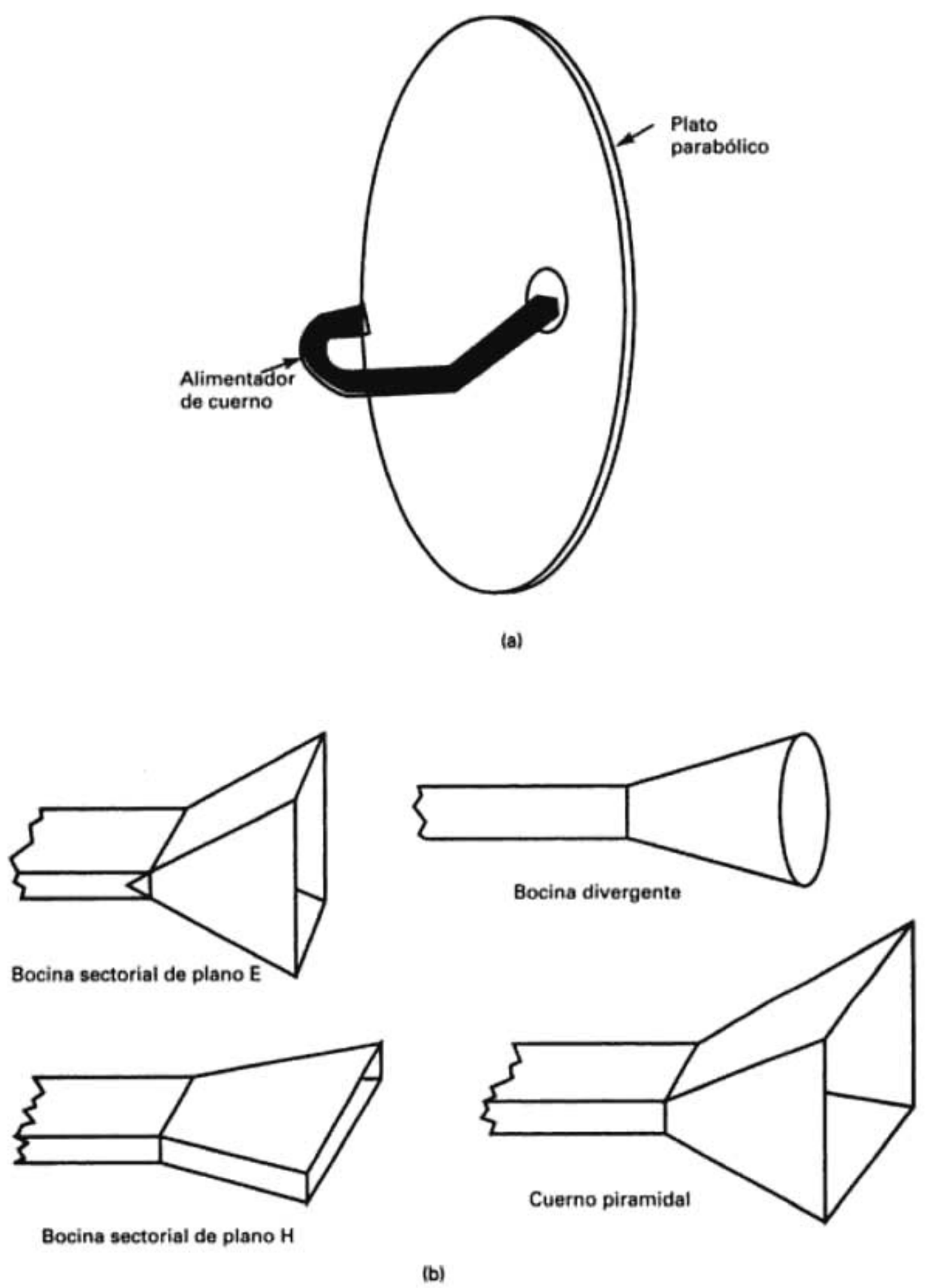


FIGURA 10-36 Antena parabólica con alimentación por cuerno: (a) alimentador de cuerno; (b) tipos de guías de onda cónicos

Antena de cuerno cónico

Una antena *de cuerno cónico* consiste en un cono que se trunca en un tramo circular de guía de ondas, como se ve en la fig. 10-38. A su vez, la guía de ondas conecta a la antena con el transmisor o con el receptor. Si el cuerno mismo se usa como antena, el *ángulo del cono* θ que a veces se llama *ángulo de conicidad*, se hace de unos 50°. En este caso, la longitud del cono truncado determina la ganancia de la antena. Cuando se usa un cuerno cónico como mecanismo de alimentación de un plato parabólico, el ángulo de conicidad y la longitud se ajustan para obtener la iluminación óptima del reflector. El mecanismo de alimentación más sencillo se obtiene cuando el cuerno cónico está en el foco del reflector.

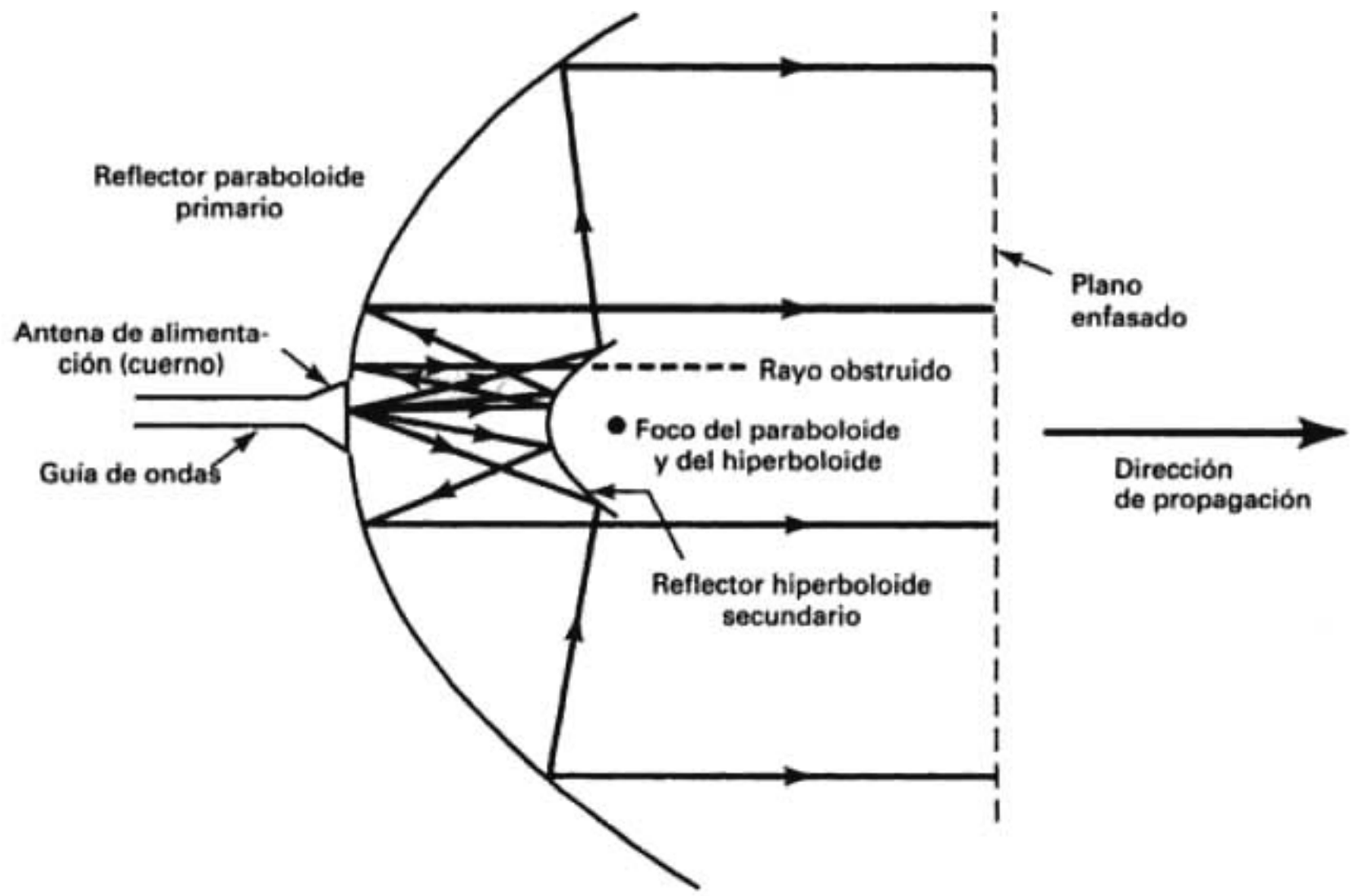


FIGURA 10-37 Antena parabólica con alimentación de Cassegrain

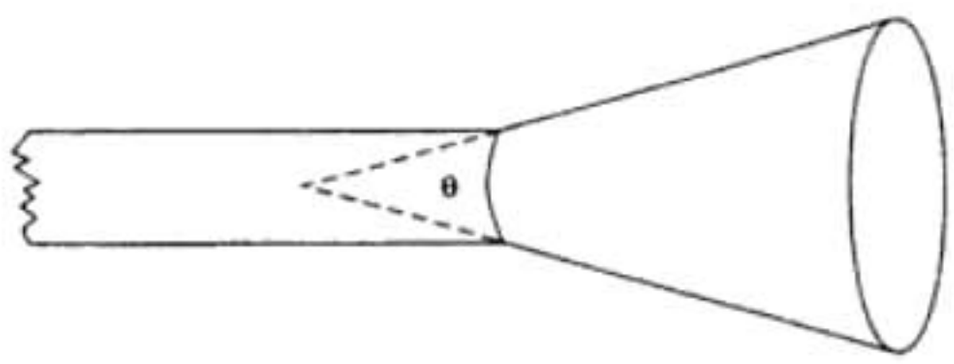


FIGURA 10-38 Antena cónica

GUÍAS DE ONDA

Las líneas de transmisión de hilos paralelos, incluyendo los cables coaxiales, no pueden propagar con eficiencia la energía electromagnética arriba de unos 20 GHz, por la atenuación causada por el efecto de superficie y las pérdidas por radiación. Además, las líneas de transmisión de hilos paralelos no se pueden usar para propagar señales de alta potencia, por los altos voltajes asociados con ellas, que causarían la disrupción del dieléctrico que separa los dos conductores. En consecuencia, las líneas de transmisión de hilos paralelos son imprácticas en muchas aplicaciones de UHF y de microondas. Hay varias alternativas, que incluyen cables de fibra óptica y guías de onda. Las fibras ópticas se describirán con detalle en el capítulo 20.

En su forma más simple, un (o *una*) *guía de ondas*, o *guíaondas*, es un tubo conductor hueco, por lo general de corte transversal rectangular, pero a veces circular o elíptico. Las dimensiones de la sección transversal se establecen de tal manera que se puedan propagar las ondas electromagnéticas por el interior de la guía (de aquí el nombre de guía de ondas). Una guía de ondas no conduce la corriente en el sentido estricto, sino más bien sirve como una frontera para confinar la energía electromagnética. Las paredes de la guía de ondas son conductores y, en consecuencia, reflejan la energía electromagnética en su superficie. Si la pared de la guía de ondas es un buen conductor y es muy delgada, pasa poca corriente a través de las paredes interiores y, por consiguiente, se disipa muy poca potencia. En una guía de ondas, la conducción

de la energía no se hace en las paredes, sino más bien por el dieléctrico del interior, que por lo general es aire deshidratado o un gas inerte. En esencia, una guía de ondas se parece a un conductor de alambre metálico, con su interior eliminado. La energía electromagnética se propaga por una guía de ondas reflejándose y rebotando, y su trayectoria es en zigzag.

Al describir el comportamiento de los guías de onda es necesario hablar citando conceptos de campo electromagnético es decir, campos eléctricos y magnéticos, y no corrientes y voltajes, como en las líneas de transmisión. El área transversal de una guía de ondas debe ser del mismo orden que la longitud de onda de la señal que propaga. En consecuencia, los guías de onda se restringen a frecuencias mayores que 1 GHz.

Guía de onda rectangular

Los guías de onda rectangulares son los que más se usan. Para comprender cómo funcionan es necesario comprender el comportamiento básico de las ondas que se reflejan en una superficie conductora.

La energía electromagnética se propaga por el espacio libre en forma de ondas electromagnéticas transversales (TEM), y su campo magnético, campo eléctrico y dirección de propagación son perpendiculares entre sí. Para que exista una onda electromagnética en una guía de ondas debe satisfacer las ecuaciones de Maxwell en su interior. Las ecuaciones de Maxwell son complicadas por necesidad, y salen del alcance de este libro. Sin embargo, un factor limitante de las ecuaciones de Maxwell es que una TEM no puede tener componente tangencial del campo eléctrico en las paredes de la guía de ondas. La onda no puede viajar por una guía de ondas en forma directa sin reflejarse en sus paredes, porque necesitaría existir el campo eléctrico junto a una pared conductora. Si eso sucediera, el campo eléctrico se pondría en corto por las mismas paredes. Para propagar bien una TEM a través de una guía de ondas, la onda debe propagarse por ella en zigzag, con el máximo del campo eléctrico en el centro de la guía y cero en la superficie de las paredes.

En las líneas de transmisión, la velocidad de la onda no depende de su frecuencia, y cuando los dieléctricos son aire o vacío, esa velocidad es igual a la velocidad en el espacio libre. Sin embargo, en los guías de onda la velocidad varía en función de la frecuencia. Además, es necesario diferenciar entre dos clases de velocidad: la *velocidad de fase* y la *velocidad de grupo*. Esta última es la velocidad a la que se propaga una onda, y la velocidad de fase es la velocidad con la cual la onda cambia de fase.

Velocidad de fase y velocidad de grupo. La velocidad de fase es la velocidad aparente de una fase determinada de la onda, por ejemplo, su cresta, o punto de máxima intensidad de campo eléctrico. La velocidad de fase es aquella con la que cambia de fase una onda, en dirección paralela a una superficie conductora que pueden ser las paredes de una guía de ondas. Se determina midiendo la longitud de una onda de determinada frecuencia y a continuación sustituyéndola en la siguiente fórmula

$$v_{ph} = f\lambda \tag{10-29}$$

en la que v_{ph} = velocidad de fase (metros por segundo)
f = frecuencia (hertz)
λ = longitud de onda (metros por ciclo)

La velocidad de grupo es la velocidad de un grupo de ondas, es decir, de un pulso. La velocidad de grupo es aquella con la que se propagan las señales de información de cualquier tipo. También, es la velocidad con la que se propaga la energía. Se puede medir determinando el tiempo necesario para que un pulso se propague por determinada longitud de la guía de ondas. Las velocidades de grupo y de fase tienen el mismo valor en el espacio libre y en las líneas de transmisión de hilos paralelos. Sin embargo, si se miden esas dos velocidades con la misma frecuencia en una guía de ondas, se encontrará que, en general, las dos velocidades no son las mismas. En ciertas frecuencias serán casi iguales, y en otras pueden ser muy distintas.

La velocidad de fase siempre es igual o mayor que la velocidad de grupo, y su producto es igual al cuadrado de la velocidad de propagación en el espacio libre. Así,

$$v_g v_{ph} = c^2 \tag{10-30}$$

donde v_{ph} = velocidad de fase (metros por segundo)
v_g = velocidad de grupo (metros por segundo)
c = 3×10^8 metros por segundo

La velocidad de fase puede ser mayor que la velocidad de la luz. Un principio básico de la física establece que ninguna forma de energía puede viajar a mayor velocidad que la de la luz (ondas electromagnéticas) en el espacio libre. Este principio no se viola, porque es la velocidad de grupo y no la de fase la que representa la velocidad de propagación de la energía.

Como la velocidad de fase en una guía de ondas es mayor que su velocidad en el espacio libre, la longitud de onda para determinada frecuencia será mayor en la guía que en el espacio libre. La relación entre la longitud de onda en el espacio libre, en la guía y la velocidad de las ondas electromagnéticas en el espacio libre es la siguiente

$$\lambda_g = \lambda_o \frac{v_{ph}}{c} \tag{10-31}$$

en donde λ_g = longitud de onda en la guía (metros por ciclo)
λ_o = longitud de onda en el espacio libre (metros por ciclo)
v_{ph} = velocidad de fase (metros por segundo)
c = velocidad de la luz en el espacio libre (3×10^8 metros por segundo)

Frecuencia de corte y longitud de onda de corte. A diferencia de las líneas de transmisión, que tienen una frecuencia máxima de operación, las guías de onda tienen una frecuencia mínima de operación que se llama *frecuencia de corte*. Es una frecuencia limitadora única; las frecuencias inferiores a la de corte no se propagarán por la guía de ondas. Al revés, las guías de onda tienen una longitud de onda mínima que se pueden propagar, que se llama *longitud de onda de corte*. Esta última se define como la mínima en el espacio libre que ya no puede propagarse en la guía de ondas. En otras palabras, sólo las frecuencias cuyas longitudes de onda sean menores que la de corte se pueden propagar por la guía de ondas. La longitud y la frecuencia de corte se determinan por las dimensiones transversales de la guía de ondas.

La relación matemática entre la longitud de onda de la guía en determinada frecuencia, y la frecuencia de corte es

$$\lambda_g = \frac{c}{\sqrt{f^2 - f_c^2}} \tag{10-32}$$

en la que λ_g = longitud de onda en la guía (metros por ciclo)
f = frecuencia de operación (hertz)
f_c = frecuencia de corte (hertz)
c = velocidad de propagación en el espacio libre (3×10^8 metros por segundo)

La ecuación 10-32 se puede ordenar en función de la longitud de onda en el espacio libre, como sigue

$$\lambda_g = \frac{\lambda_o}{\sqrt{1 - (f_c/f)^2}} \tag{10-33}$$

en donde λ_g = longitud de onda en la guía (metros por ciclo)
λ_o = longitud de onda en el espacio libre (metros por ciclo)
f = frecuencia de operación (hertz)
f_c = frecuencia de corte (hertz)

FIGURA 10-39 Perspectiva del corte de un guía de ondas rectangular

Al combinar las ecuaciones 10-31 y 10-32 y reordenar se obtiene

$$v_{ph} = \frac{c(\lambda_g)}{\lambda_o} = \frac{c}{\sqrt{1 - (f_c/f)^2}} \tag{10-34}$$

En esta ecuación es evidente que si f se hace menor que f_c, la velocidad de fase se vuelve imaginaria, lo que significa que la onda no se propaga. También, se puede ver que al tender la frecuencia de operación a la frecuencia de corte, la velocidad de fase y la velocidad en la guía se vuelven infinitas, y que la velocidad de grupo tiende a cero.

La fig. 10-39 muestra el corte transversal de una parte de una guía de ondas rectangular, con dimensiones a y b (se acostumbra representar con a a la mayor de las dos dimensiones). La dimensión a determina la frecuencia de corte de la guía de ondas, de acuerdo con la siguiente relación

$$f_c = \frac{c}{2a} \tag{10-35}$$

en la que f_c = frecuencia de corte (hertz)
a = longitud transversal (metros)

o bien, en función de la longitud de onda

$$\lambda_c = 2a \tag{10-36}$$

donde λ_c = longitud de onda de corte (metros por ciclo)
a = longitud de la sección transversal (metros)

Las ecuaciones 10-35 y 10-36 indican que el corte se presenta a la frecuencia para la cual la dimensión transversal máxima de la guía es exactamente la mitad de la longitud de onda en el espacio libre.

La fig. 10-40 muestra la vista superior de un tramo de guía de ondas rectangular, e ilustra cómo se propagan las ondas electromagnéticas por ella. Para frecuencias mayores que la de corte (figs. 10-40a, b y c), las ondas se propagan por la guía reflejándose de uno a otro lado por las paredes a diversos ángulos. La fig. 10-40d muestra lo que pasa con la onda electromagnética a la frecuencia de corte.

Ejemplo 10-6

Para una guía de ondas rectangular con una separación entre paredes de 3 cm, y una frecuencia deseada de operación de 6 GHz, determinar:

(a) La frecuencia de corte.
(b) La longitud de onda de corte.
(c) La velocidad de grupo.
(d) La velocidad de fase.

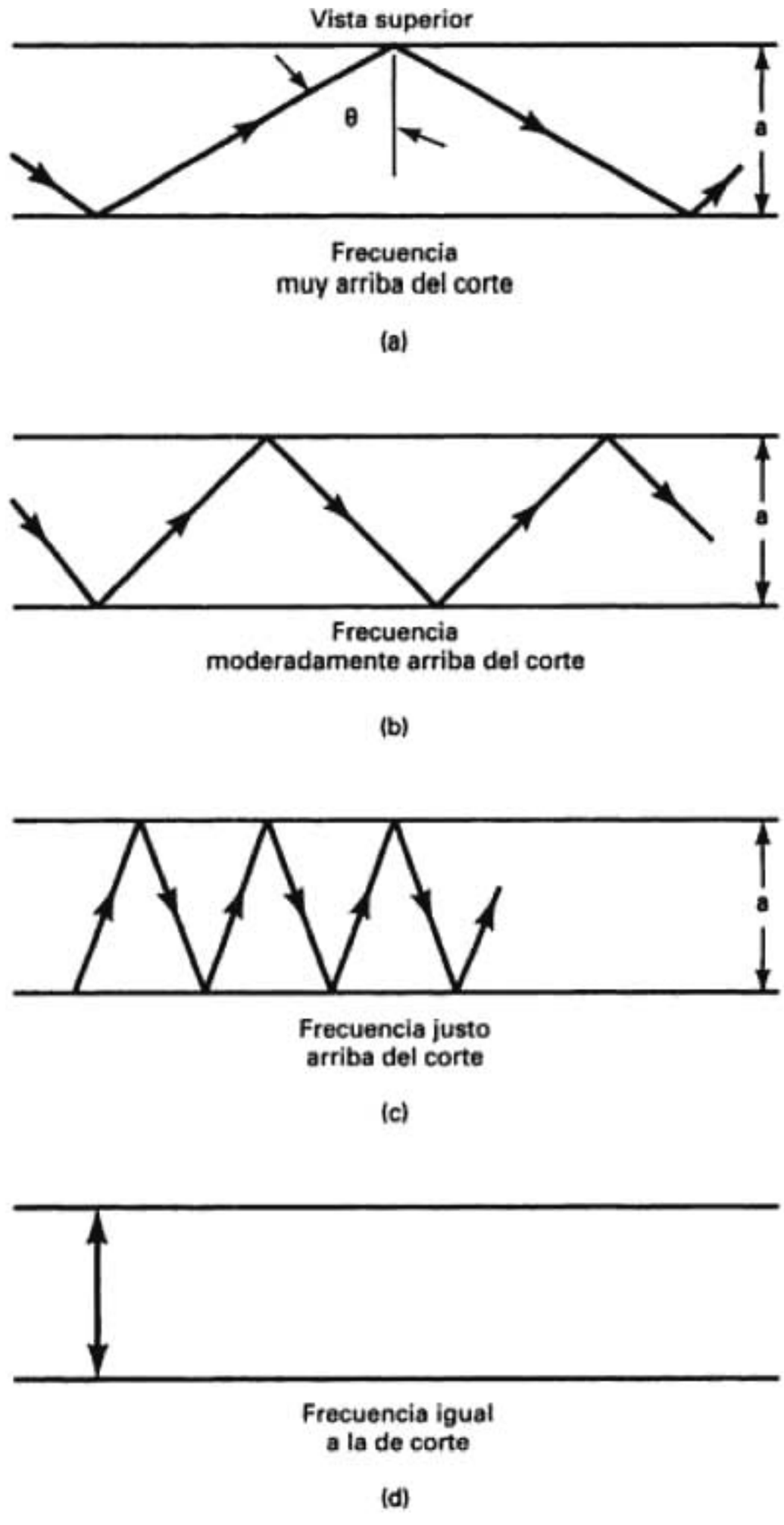


FIGURA 10-40 Propagación de ondas electromagnéticas en una guía de ondas rectangular

Solución (a) La frecuencia de corte se calcula sustituyendo los datos en la ecuación 10-35.

$$f_c = \frac{3 \times 10^8 \text{ m/s}}{2(0.03 \text{ m})} = 5 \text{ GHz}$$

(b) La frecuencia de corte se determina con la ecuación 10-36

$$\lambda_c = 2(3 \text{ cm}) = 6 \text{ cm}$$

(c) La velocidad de fase se calcula con la ecuación 10-34

$$v_{ph} = \frac{3 \times 10^8}{\sqrt{1 - (5 \text{ GHz}/6 \text{ GHz})^2}} = 5.43 \times 10^8 \text{ m/s}$$

(d) La velocidad de grupo se calcula reordenando la ecuación 10-30

$$v_g = \frac{c^2}{v_{ph}} = \frac{(3 \times 10^8)^2}{5.43 \times 10^8} = 1.66 \times 10^8 \text{ m/s}$$

Modos de propagación. Las ondas electromagnéticas viajan por una guía de ondas en configuraciones distintas, llamadas *modos* de propagación. En 1955, el Instituto de Ingenieros de Radio (IRE, *Institute of Radio Engineers*) publicó un grupo de normas. Esas normas nombraban los modos para guías de onda rectangulares como $TE_{m,n}$ para *ondas eléctricas transversales* y $TM_{m,n}$ para *ondas magnéticas transversales*. TE indica que las líneas de campo eléctrico son transversales en cualquier lugar, es decir, que son perpendiculares a las paredes de la guía, y TM indica que las líneas de campo magnético son transversales en cualquier lugar. En ambos casos, m y n son enteros que indican la cantidad de medias longitudes de onda de intensidad (eléctrica o magnética) que existen entre cada par de paredes. La cantidad m se mide a lo largo del eje x de la guía de ondas, que es el mismo eje en que se mide la dimensión a; n se mide a lo largo del eje y, igual que la dimensión b.

La fig. 10-41 muestra la distribución de campo electromagnético para una onda de modo $TE_{1,0}$. Este modo $TE_{1,0}$ se llama a veces el *modo dominante*, por que es el modo más "natural". Una guía de ondas funciona como un filtro de paso alto porque sólo pasa las frecuencias mayores que la frecuencia mínima o de corte. A frecuencias mayores que la de corte, son posibles los modos TE de propagación de orden superior, con configuraciones de campo más complicadas. Sin embargo, se prefiere no trabajar una guía de ondas a una frecuencia a la que se puedan propagar esos modos superiores. El siguiente modo superior se presenta cuando la longitud de onda en espacio libre es igual a la longitud a, es decir, al doble de la frecuencia de corte. En consecuencia, una guía de ondas rectangular se suele trabajar dentro de los márgenes de frecuencia de f_c a $2f_c$. No es deseable permitir modos superiores de propagación porque no se acoplan bien con la carga y, en consecuencia, provocan reflexiones, y creación de ondas estacionarias. También se prefiere el modo $TE_{1,0}$ porque permite tener la guía de ondas de tamaño mínimo, para determinada frecuencia de operación.

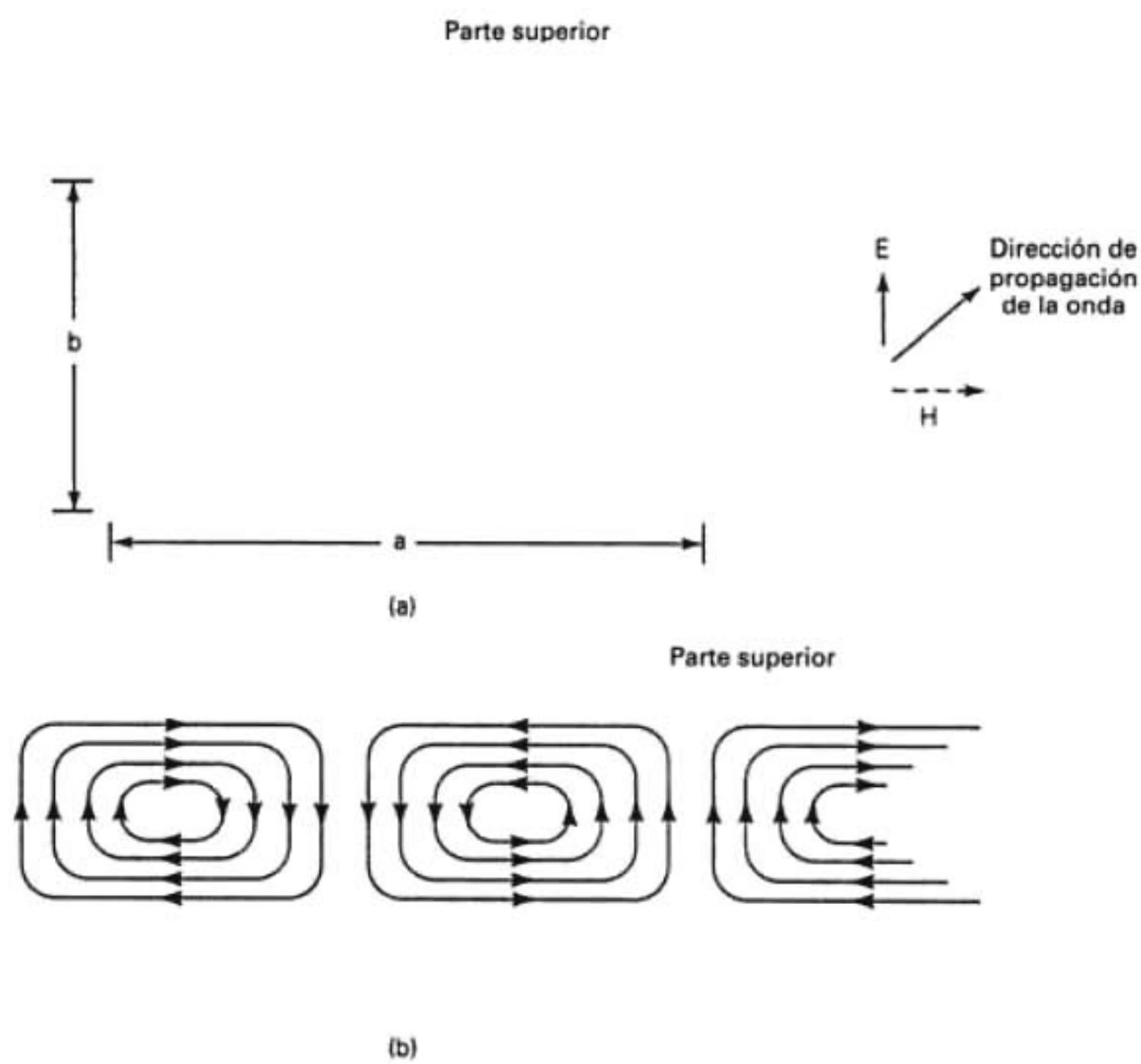


FIGURA 10-41 Vectores de campo eléctrico y magnético en una guía de ondas rectangular: (a) vista desde el extremo; (b) configuración del campo magnético en un tramo longitudinal

En la fig. 10-41a, los vectores de campo eléctrico E son paralelos entre sí, y perpendiculares a la cara larga de la guía. Su amplitud es máxima a la mitad entre las paredes cortas, y disminuye hasta cero en las paredes, en forma senoidal. Los vectores de campo magnético H, representados con líneas interrumpidas, también son paralelos entre sí, y perpendiculares a los vectores eléctricos. La intensidad magnética es constante en dirección vertical a través de la sección de la guía. La onda se propaga en la dirección longitudinal de la guía, perpendicularmente a los vectores E y H. La fig. 10-41b muestra la configuración del campo magnético en un tramo longitudinal de guía de onda, para el modo de propagación $TE_{1,0}$.

Impedancia característica. Las guías de onda tienen una impedancia característica que es análoga a la de las líneas de transmisión de hilos paralelos, y se relaciona en forma estrecha con la impedancia característica del espacio libre. La impedancia característica de una guía de ondas tiene el mismo significado que la de una línea de transmisión, con respecto a balanceo de carga, reflexiones de señal y ondas estacionarias. La impedancia característica de una guía de ondas se determina con la siguiente ecuación

$$Z_o = \frac{377}{\sqrt{1 - (f_c/f)^2}} = 377\frac{\lambda_g}{\lambda_o} \tag{10-37}$$

siendo Z_o = impedancia característica (ohms)
f_c = frecuencia de corte (hertz)
f = frecuencia de operación (hertz)

En general, Z_o es mayor que 377 Ω. De hecho, a la frecuencia de corte, Z_o se vuelve infinita, y a una frecuencia igual a $2f_c$, el doble de la frecuencia de corte, Z_o = 435 Ω. Dos guías de onda con la misma dimensión *a* pero distintas dimensiones *b* tendrán el mismo valor de frecuencia de corte y el mismo valor de impedancia característica. Sin embargo, si se interconectan estas dos guías de onda, extremo con extremo, y se propaga por ellas una onda electromagnética, se produce una discontinuidad en el empalme, y se presentarán reflexiones, aun cuando sus impedancias se balanceen.

Balanceo de impedancia. En las guías de onda se usan líneas de acoplamiento reactivas para transformación y balanceo (o compensación, o igualación) de impedancia, de la misma manera que en las líneas de transmisión de hilos paralelos. Se usan líneas de acoplamiento en cortocircuito de guías de onda en la misma forma que se usan en las líneas de transmisión.

La fig. 10-42 muestra cómo se instalan iris (o diafragmas) inductivos y capacitivos en una guía de onda rectangular, para que se comporten como si fueran susceptancias en *shunt*, o en paralelo. Los iris consisten en láminas metálicas delgadas instaladas perpendicularmente a las

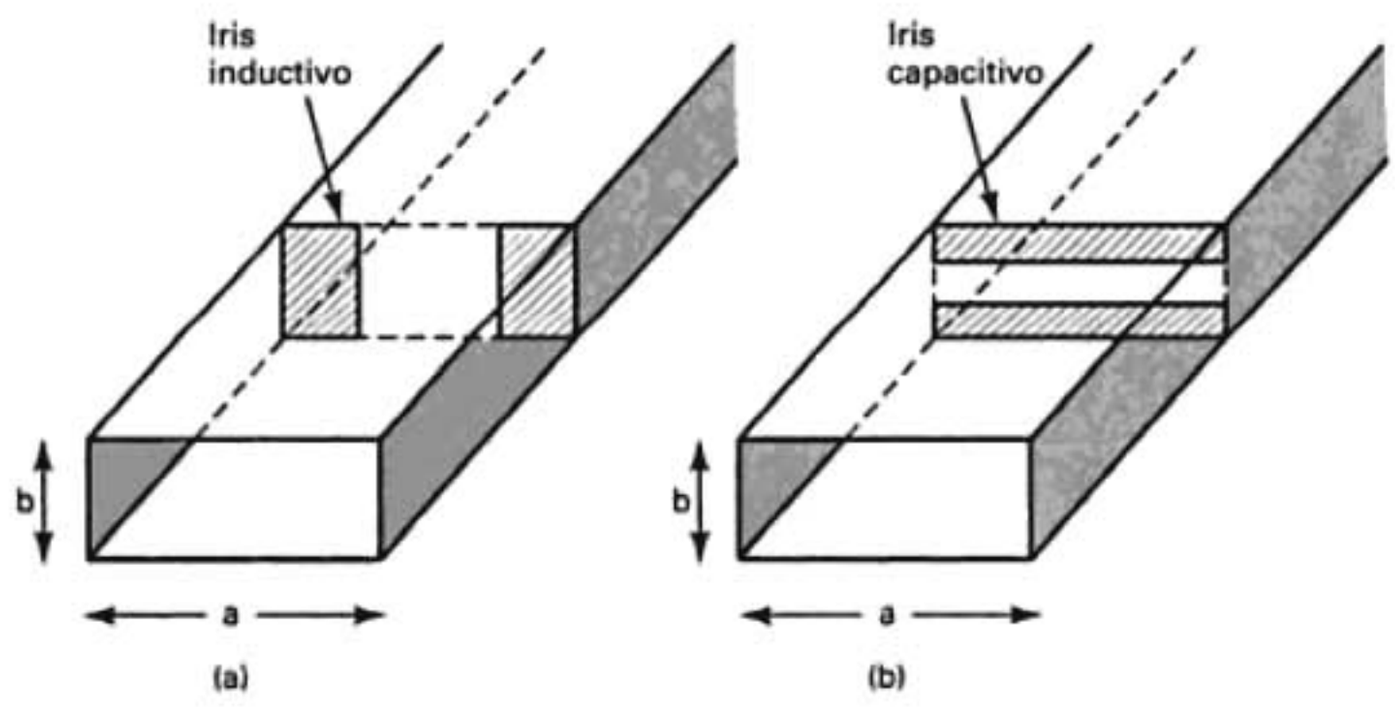


FIGURA 10-42 Balanceo de impedancia de una guía de onda: (a) iris inductivo; (b) iris capacitivo

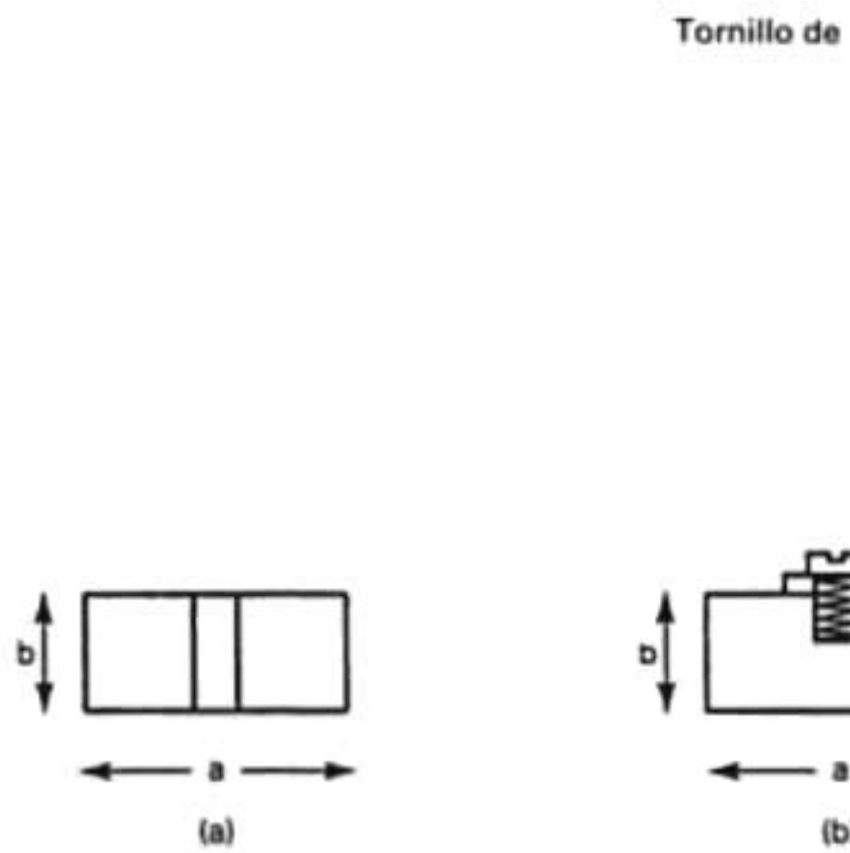


FIGURA 10-43 Balanceo de impedancia de una guía de onda: (a) poste; (b) tornillo de sintonía

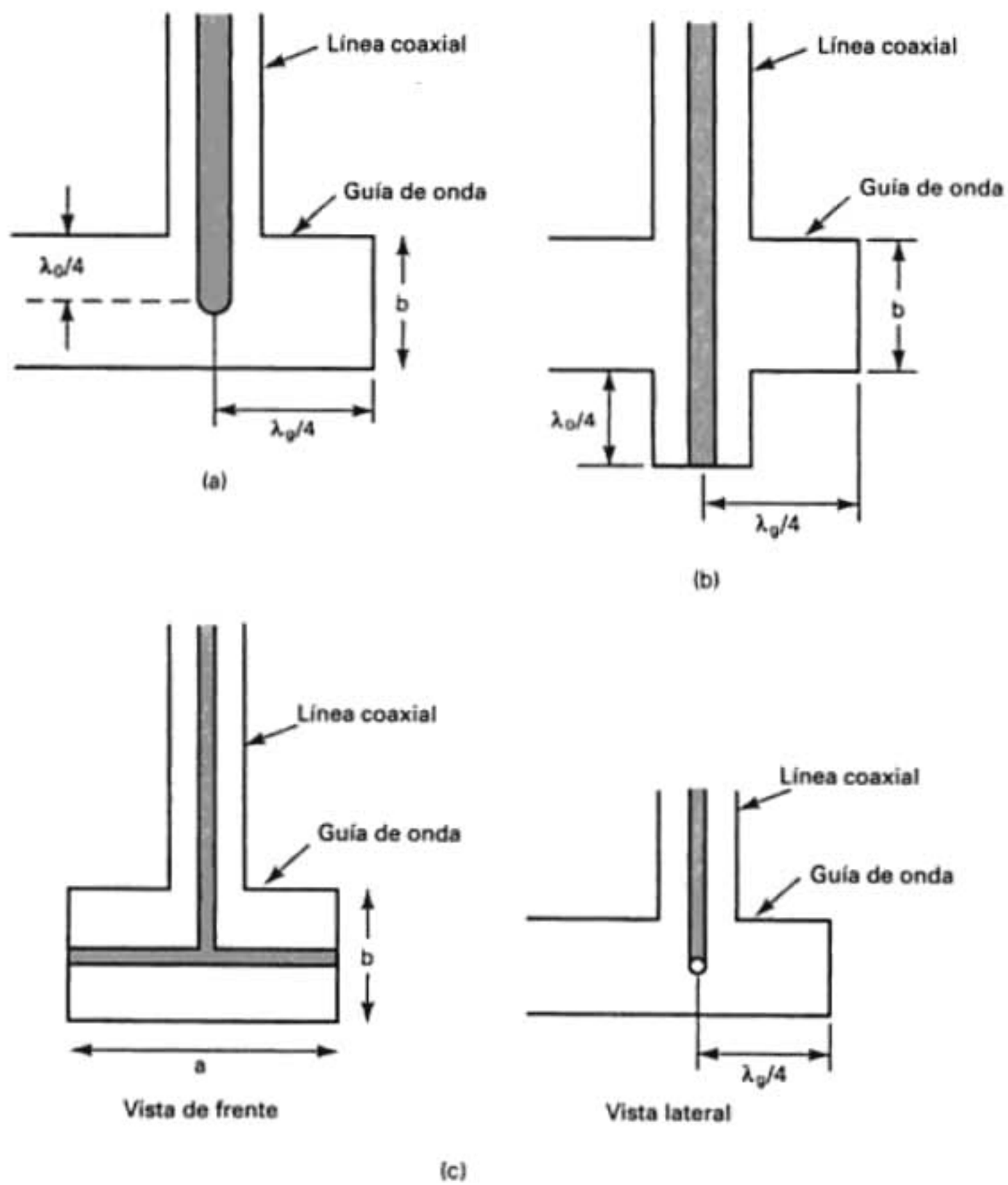


FIGURA 10-44 Acoplamiento de línea de transmisión con guía de onda: (a) acoplador de sensor de cuarto de onda; (b) acoplador recto pasante; (c) acoplador de barra transversal

TABLA 10-3 Dimensiones y características eléctricas de guías de onda rectangulares

Intervalo de frecuencia útil (GHz)	Dimensiones exteriores (mm)	Atenuación promedio teórica (dB/m)	Capacidad teórica de potencia (CW) (kW)
1.12–1.70	169 × 86.6	0.0052	14,600
1.70–2.60	113 × 58.7	0.0097	6,400
2.60–3.95	76.2 × 38.1	0.019	2,700
3.95–5.85	50.8 × 25.4	0.036	1,700
5.85–8.20	38.1 × 19.1	0.058	635
8.20–12.40	25.4 × 12.7	0.110	245
12.40–18.00	17.8 × 9.9	0.176	140
18.0–26.5	12.7 × 6.4	0.37	51
26.5–40.0	9.1 × 5.6	0.58	27
40.0–60.0	6.8 × 4.4	0.95	13
60.0–90.0	5.1 × 3.6	1.50	5.1
90.0–140	4.0 (diám.)	2.60	2.2
140–220	4.0 (diám.)	5.20	0.9
220–325	4.0 (diám.)	8.80	0.4

paredes de la guía de ondas, unidas a ella en sus orillas, y que forman una abertura entre ellas. Cuando la abertura es paralela a las paredes cortas, la susceptancia es inductiva; cuando es paralela a las paredes largas, es capacitiva. La magnitud de la susceptancia es proporcional al tamaño de la abertura.

Un poste colocado transversal, en la dirección de la dimensión corta de la guía de ondas, como muestra la fig. 10-43a, funciona como susceptancia *shunt* inductiva, cuyo valor depende de su distancia y de su posición en el plano transversal. Los tornillos de sintonía, que se ven en la fig. 10-43b, sobresalen en parte en la dirección corta de la guía, funcionan como capacitancia y se pueden ajustar.

Acoplamiento de la línea de transmisión a la guía de ondas. La fig. 10-44 muestra varias formas en que se pueden unir una guía de ondas y una línea de transmisión. Los acopladores que se ven se pueden usar como lanzadores de onda en el extremo de entrada de una guía de ondas, o como receptores de onda en el extremo de carga de la guía. Las dimensiones representadas por $\lambda_o/4$ y $\lambda_g/4$ son aproximadas. En la práctica se ajustan en forma experimental para tener los mejores resultados.

La tabla 10-3 es una lista del intervalo de frecuencias, las dimensiones y las características eléctricas de varios tipos comunes de guías de onda rectangulares.

OTROS TIPOS DE GUÍAS DE ONDA

Guía de onda circular

Las guías de onda rectangulares son, con mucho, las más comunes; sin embargo, en radar y en aplicaciones de microondas se usan guías de onda circulares, cuando es necesario o hay ventajas al propagar ondas polarizadas vertical y horizontalmente por la misma guía. La fig. 10-45 muestra dos tramos de guía de onda circular, unidas por una junta rotativa.

El comportamiento de las ondas electromagnéticas en las guías de onda circulares es igual que en las rectangulares. Sin embargo, debido a la distinta geometría, algunos de los cálculos se hacen en una forma un poco distinta.

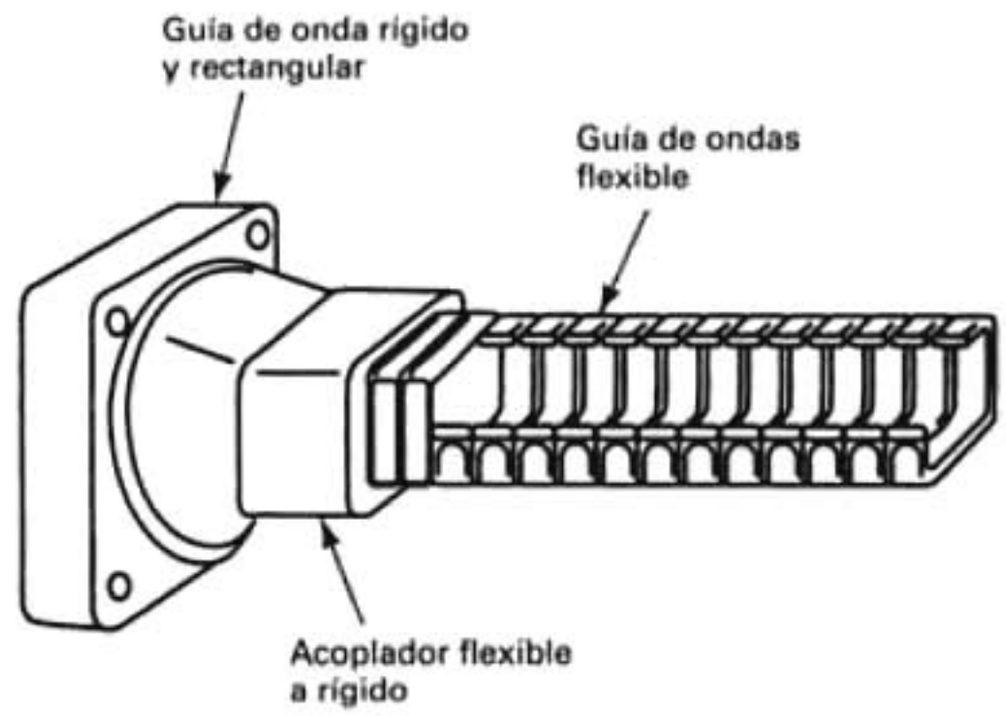
La longitud de onda de corte, para guías de onda circulares, es

$$\lambda_o = \frac{2\pi r}{kr} \tag{10-38}$$

en donde λ_o = longitud de onda de corte (metros por ciclo)
r = radio interno de la guía de onda (metros)
kr = solución de una ecuación con funciones de Bessel

Junta rotativa

FIGURA 10-45 Guía de onda circular con junta rotativa



(a) (b)

FIGURA 10-46 Guía de onda rígida: (a) un entrante; (b) doble entrante

FIGURA 10-47 Guía flexible de ondas

Como el modo de propagación con la máxima longitud de onda de corte es aquél con el valor mínimo de kr (1.84), el modo $TE_{1,1}$ es el dominante en las guías de onda circulares. Para ese modo, la longitud de onda de corte se reduce a

$$\lambda_o = 1.7d \tag{10-39}$$

siendo d el diámetro de la guía de ondas (metros).

Las guías de onda circulares son más fáciles de fabricar que las rectangulares, y más fáciles de empalmar. Sin embargo, las guías circulares tienen un área mucho mayor que una rectangular correspondiente, para llevar la misma señal. Otra desventaja de las guías de onda circulares es que el plano de polarización puede girar mientras la onda se propaga; es decir, una onda polarizada horizontalmente se puede transformar en una polarizada verticalmente, y viceversa.

Guía de onda rígida

La fig. 10-46 muestra dos clases de guía de onda rígida. Esta clase de guías es más costosa en su fabricación que las rectangulares normales; sin embargo, permite también el funcionamiento a menores frecuencias, para determinado tamaño. En consecuencia, es posible tener menores dimensiones generales de guía de ondas cuando son con entrantes. Esta característica, combinada con su mayor costo, limita su utilidad a aplicaciones especializadas.

Guía de onda flexible

La fig. 10-47 muestra un tramo de una guía de ondas rectangular flexible. Consiste de bandas en espiral, de latón o de cobre. El exterior se cubre con un recubrimiento dieléctrico suave, con frecuencia de hule, para mantener hermética la guía de onda al aire y al agua. En los sistemas de microondas se usan tramos cortos de guía de onda flexible, cuando se interconectan varios transmisores y receptores con una unidad compleja de combinación o de separación. También, se usan mucho las guías de onda flexibles en los equipos de prueba de microondas.

PREGUNTAS

10-1. Defina *antena*.

10-2. Describa el funcionamiento básico de la antena, usando ondas estacionarias.

10-3. Describa una gráfica de radiación relativa y una gráfica de radiación absoluta.

10-4. Defina la *relación frente a posterior.*

10-5. Describa una antena omnidireccional.

10-6. Defina el *campo cercano* y el *campo lejano*.

10-7. Defina la *resistencia de radiación* y la *eficiencia de antena*.

10-8. Defina y describa las diferencias entre *ganancia directiva* y *ganancia de potencia.*

10-9. ¿Qué es la directividad para una antena isotrópica?

10-10. Defina la *potencia irradiada isotrópica efectiva*.

10-11. Defina qué es *polarización de antena.*

10-12. Defina la *abertura de haz de la antena*.

10-13. Defina el *ancho de banda de la antena*.

10-14. Defina la *impedancia de entrada a la antena*. ¿Qué factores contribuyen a esa impedancia?

10-15. Describa el funcionamiento de un doblete elemental.

10-16. Describa el funcionamiento de un dipolo de media onda.

10-17. Describa los efectos del suelo sobre un dipolo de media onda.

10-18. Describa el funcionamiento de una antena aterrizada.

10-19. ¿Qué quiere decir *carga de antena*?

10-20. Describa una bobina de carga de antena.

10-21. Describa la carga por el punto superior de una antena.

10-22. Describa qué es una red de antenas.

10-23. ¿Qué quiere decir *elemento excitado* y *elemento parásito*?

10-24. Describa la distribución de la radiación para una red de radiación lateral, y para una red de radiación longitudinal.

10-25. Defina qué es una *antena no resonante*.

10-26. Describa el funcionamiento de la antena de rombo.

10-27. Describa una antena de dipolo plegado.

10-28. Describa una antena de Yagi-Uda.

10-29. Describa una antena log-periódica.

10-30. Describa el funcionamiento de una antena de cuadro.

10-31. Describa, en forma breve, cómo funciona una *antena de red enfasada*, y para qué se usa principalmente.

10-32. Describa, en forma breve, cómo funciona una antena *helicoidal*.

10-33. Defina los siguientes términos: *lóbulo principal, lóbulos laterales, acoplamiento lado a lado* y *acoplamiento espalda con espalda*.

10-34. ¿Cuáles son las dos partes principales de una *antena parabólica*?

10-35. Describa, en forma breve, cómo funciona un *reflector parabólico*.

10-36. ¿Cuál es el objetivo del *mecanismo de alimentación* en una antena de reflector parabólico?

10-37. ¿Qué quiere decir *área de captura* de una antena parabólica?

10-38. Describa cómo funciona un mecanismo de *alimentación central* con un reflector parabólico.

10-39. Describa cómo funciona un mecanismo de *alimentación cónica* con un reflector parabólico.

10-40. Describa cómo funciona una *alimentación de Cassegrain* con un reflector parabólico.

10-41. En su forma más simple ¿qué es una *guía de ondas*?

10-42. Describa la *velocidad de fase* y la *velocidad de grupo*.

10-43. Describa qué es la *frecuencia de corte* para una guía de ondas; ¿qué es la *longitud de onda de corte*?

10-44. ¿Qué quiere decir modo de propagación TE? ¿Y modo de propagación TM?

10-45. ¿Cuándo es mejor usar una guía de ondas circular?

PROBLEMAS

10-1. Para una antena con potencia de entrada $P_{ent} = 100$ W rms, corriente $I = 2$ A y resistencia efectiva $R_e = 2\ \Omega$, calcule:
- **(a)** La resistencia de radiación de la antena.
- **(b)** La eficiencia de la antena.
- **(c)** La potencia irradiada de la antena, P_{rad}.

10-2. Calcule la directividad, en decibelios, de una antena que produce una densidad de potencia $\mathcal{P} = 2\ \mu W/m^2$ en un punto en el que una antena de referencia produce $0.5\ \mu W/m^2$.

10-3. Calcule la ganancia de potencia, en decibelios, para una antena con ganancia directiva $\mathcal{D} = 40$ dB y eficiencia $\eta = 65\%$.

10-4. Calcule la potencia isotrópica efectiva irradiada para una antena con ganancia de potencia $A_p = 43$ dB y potencia irradiada $P_{rad} = 200$ W.

10-5. Calcule la potencia isotrópica efectiva irradiada para una antena con directividad $\mathcal{D} = 33$ dB, eficiencia $\eta = 82\%$ y potencia de entrada $P_{ent} = 100$ W.

10-6. Determine la densidad de potencia en un puno a 20 km de una antena que está irradiando 1000 W y tiene ganancia de potencia $A_p = 23$ dB.

10-7. Calcule la densidad de potencia en un punto a 30 km de una antena cuya potencia de entrada es $P_{ent} = 40$ W, eficiencia $\eta = 75\%$, y directividad $\mathcal{D} = 16$ dB.

10-8. Determine la densidad de potencia capturada por una antena receptora con los siguientes parámetros: entrada de la antena de transmisión $P_{ent} = 50$ W, ganancia de la antena de transmisión, $A_p = 30$ dB, distancia entre las antenas transmisora y receptora $d = 20$ km, ganancia directiva de la antena receptora $A_p = 26$ dB.

10-9. Calcule la directividad, en decibelios, de una antena que produce una densidad de potencia en un punto, que es 40 veces mayor que la densidad en el mismo punto cuando se usa la antena de referencia.

10-10. Calcule la potencia efectiva irradiada de una antena con directividad $\mathcal{D} = 400$, eficiencia $\eta = 0.60$ y potencia de entrada $P_{ent} = 50$ W.

10-11. Calcule la eficiencia de una antena con resistencia de radiación $R_r = 18.8\ \Omega$, resistencia efectiva $R_e = 0.4\ \Omega$ y ganancia directiva $\mathcal{D} = 200$.

10-12. Calcule la ganancia de potencia A_p para el problema 10-11.

10-13. Determine la eficiencia de una antena con potencia irradiada $P_{rad} = 44$ W, potencia disipada $P_d = 0.8$ W y ganancia directiva $\mathcal{D} = 400$.

10-14. Calcule la ganancia de potencia A_p para el problema 10-13.

10-15. Calcule la ganancia de potencia y la abertura de haz para una antena helicoidal de radiación longitudinal, con los siguientes parámetros: diámetro de hélice = 0.1 m, cantidad de vueltas = 10, paso = 0.05 m y frecuencia de operación 500 MHz.

10-16. Calcule la abertura del haz y las potencias de transmisión y recepción de una antena parabólica con los siguientes parámetros: diámetro de plato 2.5 m, frecuencia de operación 4 GHz y eficiencia 55%.

10-17. Para una guía de ondas rectangular con separación entre paredes de 2.5 cm y una frecuencia deseada de operación de 7 GHz, determine:
- **(a)** La frecuencia de corte.
- **(b)** La longitud de onda de corte.
- **(c)** La velocidad de grupo.
- **(d)** La velocidad de fase.

10-18. Para una antena con potencia de entrada $P_{ent} = 400$ W, corriente rms $i = 4$A y resistencia $R_{cd} = 4\ \Omega$, determine:
- **(a)** La resistencia de radiación de la antena.
- **(b)** La eficiencia de la antena.
- **(c)** La potencia irradiada de la antena, P_{rad}.

10-19. Calcule la directividad, en decibelios, para una antena que produce una densidad de potencia $\mathcal{P} = 4\ \mu W/m^2$ en un punto en el espacio, cuando una antena de referencia produce $0.4\ \mu W/^2$ en el mismo punto.

10-20. Calcule la ganancia de potencia, en decibelios, para una antena con ganancia directiva $\mathcal{D} = 50$ dB, y una eficiencia de 75%.

10-21. Calcule la potencia isotrópica efectiva irradiada, para una antena con ganancia de potencia $A_p = 26$ dB y una potencia irradiada $P_{rad} = 400$ W.

10-22. Calcule la potencia isotrópica efectiva irradiada, para una antena con directividad $\mathcal{D} = 43$ dB, eficiencia de 75% y potencia de entrada $P_{ent} = 50$ W.

10-23. Calcule la densidad de potencia en un punto a 20 km de una antena que está irradiando 1200 W y tiene ganancia de potencia $A_p = 46$ dB.

10-24. Calcule la densidad de potencia en un punto a 50 km de una antena cuya potencia de entrada $P_{ent} = 100$ W, su eficiencia es 55% y su directividad es $\mathcal{D} = 23$ dB.

10-25. Calcule la potencia capturada por una antena receptora con los siguientes parámetros:

Potencia irradiada $P_{rad} = 100$ W

Ganancia directiva de la antena transmisora $A_t = 40$ dB

Distancia entre la antena transmisora y la receptora $d = 40$ km

Ganancia directiva de la antena receptora $A_r = 23$ dB.

10-26. Calcule la directividad, en dB, de una antena que produce una densidad de potencia en un punto que es 100 veces mayor que la densidad de potencia en el mismo punto cuando se usa una antena de referencia.

10-27. Calcule la potencia efectiva irradiada para una antena con directividad $\mathcal{D} = 300$, eficiencia $= 80\%$ y potencia de entrada $P_{ent} = 2500$ W.

10-28. Calcule la eficiencia de una antena con resistencia de radiación $R_r = 22.2\ \Omega$, resistencia en cd $R_{cd} = 2.8\ \Omega$ y ganancia directiva $\mathcal{D} = 40$ dB.

10-29. Calcule A_p, la ganancia de potencia, para el problema 10-28.

10-30. Calcule la eficiencia de una antena cuya potencia irradiada $P_{rad} = 65$ W, potencia disipada $P_d = 5$ W y ganancia directiva $\mathcal{D} = 200$.

10-31. Calcule la ganancia de potencia para el problema 10-30.

CAPÍT

Imagen protegida por derechos de autor

Comunicaciones con fibra óptica

INTRODUCCIÓN

En esencia, un *sistema óptico de comunicaciones* es un sistema electrónico de comunicaciones que usa la luz como portador de información. Sin embargo, es difícil e impráctico propagar ondas luminosas por la atmósfera terrestre. En consecuencia, los *sistemas de comunicaciones con fibra óptica* usan fibras de vidrio o de plástico para "contener" las ondas luminosas y guiarlas en una forma similar a como las ondas electromagnéticas son guiadas en una guía de ondas. La optoelectrónica es la rama de la electrónica que estudia la transmisión de la luz a través de fibras ultrapuras, que se suelen fabricar con vidrio o con plástico.

La *capacidad de conducción de información* de un sistema electrónico de comunicaciones es directamente proporcional a su amplitud de banda. Para fines de comparación, se acostumbra expresar el ancho de banda de un sistema analógico de comunicaciones como un porcentaje de la frecuencia de su portadora. A esto se le llama a veces *relación de utilización del ancho de banda*. Por ejemplo, un sistema de radiocomunicaciones de VHF trabajando a una frecuencia de portadora de 100 MHz con ancho de banda de 10 MHz tiene una relación de utilización de ancho de banda de 10%. Un sistema de radio de microondas que funciona con una frecuencia de portadora, con una relación de utilización de ancho de banda igual a 10% tendría disponible 1 GHz de ancho de banda. Es obvio que mientras mayor es la frecuencia de portadora, el ancho de banda es mayor y la capacidad de conducción de información es mayor. Las frecuencias luminosas que se usan en los sistemas de comunicaciones con fibra óptica están entre 1×10^{14} y 4×10^{14} Hz (100,000 a 400,000 GHz). Una relación de utilización de ancho de banda de 10% significaría una banda de entre 10,000 y 40,000 GHz de ancho.

HISTORIA DE LAS FIBRAS ÓPTICAS

Alexander Graham Bell, en 1880, experimentó con un aparato al que llamó *fotófono*. El fotófono era un dispositivo formado con espejos y detectores de selenio, que transmitía ondas sonoras sobre un rayo de luz. Ese fotófono era muy malo, no confiable y no tenía aplicación práctica. En realidad, la luz visible era un medio principal de comunicaciones antes de que comenzaran las comunicaciones electrónicas. Durante mucho tiempo se usaron señales de humo para mandar

mensajes cortos y sencillos. Sin embargo, el concepto de Bell fue el primer intento de usar un rayo de luz para transportar información.

Es impráctica la transmisión de ondas luminosas a través de cualquier distancia útil a través de la atmósfera terrestre, porque el vapor de agua, el oxígeno y las partículas en el aire absorben y atenúan las señales en frecuencias luminosas. En consecuencia, el único tipo práctico de comunicaciones ópticas es el que usa una guía de fibra. En 1930, J. L. Baird, científico inglés, y C. W. Hansell, de Estados Unidos, obtuvieron patentes para barrer y transmitir imágenes de televisión a través de cables de fibra no recubierta. Algunos años después un científico alemán, llamado H. Lamm, transmitió bien imágenes a través de una sola fibra de vidrio. En esa época, la mayoría de las personas consideraban a la óptica de fibras más como juguete o como una gracia de laboratorio y, en consecuencia, no fue sino hasta principios de la década de 1950 que se hicieron avances sustanciales en el campo de las fibras ópticas.

En 1951, A. C. S. van Heel de Holanda, y H. H. Hopkins y N. S. Kapany de Inglaterra experimentaron con transmisión de luz a través de *haces* de fibras. Sus estudios condujeron al desarrollo del *fibroscopio flexible*, que se usa mucho en el campo de la medicina. Kapany fue quien acuño el término "fibra óptica" en 1956.

En 1958, Charles H. Townes, de Estados Unidos, y Arthur L. Schawlow, de Canadá, presentaron un trabajo donde se describía cómo era posible usar emisión estimulada para amplificar las ondas luminosas (láser) y las microondas (máser). Dos años después Theodore H. Maiman, científico de Hughes Aircraft Company, construyó el primer máser óptico.

El *láser* (de *light amplification by stimulated emission of radiation*, amplificación de luz por emisión estimulada de radiación) fue inventado en 1960. La potencia relativamente alta de salida del láser, su alta frecuencia de operación y su capacidad de portar un ancho de banda extremadamente grande, lo hacen ideal para sistemas de comunicaciones de gran capacidad. La invención del láser aceleró mucho los esfuerzos de investigación en comunicaciones con fibra óptica, aunque no fue sino hasta 1967 que K. C. Kao y G. A. Bockham, del Standard Telecommunications Laboratory de Inglaterra propusieron un medio nuevo de comunicaciones, usando cables de fibra *revestida*.

Los cables de fibra disponibles en la década de 1960 tenían *pérdidas* extremadamente grandes (más de 1000 dB/km), lo cual limitaba las transmisiones ópticas a distancias cortas. En 1970, Kapron, Keck y Maurer, de Corning Glass Works en Corning, New York, desarrollaron una fibra óptica con pérdidas menores que 2 dB/km. Fue el "gran" avance necesario para permitir los sistemas prácticos de comunicaciones con fibra óptica. A partir de 1970, la tecnología de fibras ópticas ha crecido en forma exponencial. En fecha reciente, los Laboratorios Bell transmitieron bien 1000 millones de bps por un cable de fibra de 600 mi, sin un regenerador.

A fines de la década de 1970 y principios de la década de 1980, el refinamiento de los cables ópticos, y el desarrollo de fuentes luminosas y detectores de alta calidad y económicos abrió la puerta al desarrollo de sistemas de comunicaciones de alta calidad, alta capacidad, eficientes y económicos, con fibra óptica. A fines de la década de 1980 las pérdidas en las fibras ópticas se redujeron hasta 0.16 dB/km, y en 1988, NEC Corporation estableció un récord de transmisión a gran distancia, al enviar 10 Gbits/s con 80.1 km de fibra óptica. También en 1988, el Instituto Nacional Americano de Normas (ANSI) publicó *Synchronous Optical Network* (*SONET*). A mediados de la década de 1990, las redes ópticas para voz y datos eran lugar común en Estados Unidos y en gran parte del mundo.

COMPARACIÓN DE INSTALACIONES CON FIBRAS ÓPTICAS Y CON CABLES METÁLICOS

Las comunicaciones a través de cables de fibra de vidrio o de plástico tiene varias ventajas abrumadoras sobre las comunicaciones que usan instalaciones convencionales de cable *metálico* o *coaxial*.

Ventajas de los sistemas de fibra óptica

1. Mayor capacidad de información: los sistemas de comunicaciones con fibras ópticas tienen mayor capacidad de información que los cables metálicos, debido a los anchos de banda,

inherentemente mayores con las frecuencias ópticas. Las fibras ópticas se consiguen con anchos de banda hasta de 10 GHz. Los cables metálicos tienen capacitancia entre, e inductancia a lo largo, de sus conductores, que los hacen funcionar como filtros pasabajas y eso limita sus frecuencias de transmisión, anchos de banda y capacidad de conducción de información. Los sistemas modernos de comunicaciones con fibra óptica son capaces de transmitir varios gigabits por segundo a través de cientos de millas, permitiendo combinar y propagar literalmente millones de canales individuales de datos y de voz con un solo cable de fibra óptica.

2. Inmunidad a la diafonía: los cables ópticos son inmunes a la diafonía entre cables vecinos, debida a la inducción magnética. Las fibras de vidrio o de plástico son no conductores de electricidad y, en consecuencia, no tienen campos magnéticos asociados con ellas. En los cables metálicos, la causa principal de la diafonía es la inducción magnética entre conductores ubicados físicamente cercanos entre sí.

3. Inmunidad a la interferencia por estática: los cables ópticos son inmunes al ruido de estática que causa la interferencia electromagnética (EMI) debida a rayos, motores eléctricos, luces fluorescentes y otras fuentes de ruido eléctrico. Esta inmunidad también se debe a que las fibras ópticas son no conductores de la electricidad, y a que el ruido eléctrico no afecta la energía en las frecuencias luminosas. Los cables de fibras tampoco irradian energía de RF y, en consecuencia, no pueden interferir con otros sistemas de comunicaciones. Esta característica hace que los sistemas de fibra óptica sean ideales para aplicaciones militares en las que los efectos de las armas nucleares (interferencia de pulsos electromagnéticos EMP) tienen un efecto devastador sobre los sistemas electrónicos convencionales de comunicaciones.

4. Inmunidad al ambiente: los cables ópticos son más resistentes a los extremos en el ambiente que los cables metálicos. También, los cables ópticos funcionan dentro de variaciones más amplias de temperatura y son menos afectados por los líquidos y gases corrosivos.

5. Seguridad: los cables ópticos son más seguros y fáciles de instalar y mantener que los cables metálicos. Debido a que las fibras de vidrio y de plástico son no conductoras, no se asocian con ellas corrientes ni voltajes eléctricos. Las fibras ópticas se pueden usar cerca de líquidos volátiles y de gases, sin preocuparse porque puedan causar explosiones o incendios. Las fibras ópticas son menores y mucho más ligeras que los cables metálicos. En consecuencia, es más fácil trabajar con ellas y se adaptan mucho mejor a aplicaciones aéreas. También, los cables de fibra requieren menos espacio de almacenamiento y son más fáciles de transportar.

6. Seguridad: las fibras ópticas son más seguras que los cables metálicos. Es virtualmente imposible entrar a un cable de fibra sin que sepa el usuario, y los cables de fibra óptica no se pueden detectar con buscadores de metales, a menos que tengan refuerzo de acero para tener mayor resistencia. Éstas también son las cualidades que hacen atractivas a las fibras ópticas para las aplicaciones militares.

7. Duran más: aunque todavía no se ha demostrado, se anticipa que los sistemas de fibra óptica durarán más que las instalaciones metálicas. Esta hipótesis se basa en las mayores tolerancias que tienen los cables de fibra frente a cambios en las condiciones ambientales, y en su inmunidad a las sustancias corrosivas.

8. Economía: el costo de los cables de fibra óptica es, aproximadamente, igual al de los cables metálicos. Sin embargo, los cables de fibra tienen menores pérdidas y en consecuencia requieren menos repetidoras. Esto equivale a menores costos de instalación y del sistema en general, así como a mayor confiabilidad.

Desventajas de los sistemas de fibra óptica

1. Costos de interconexión: los sistemas de fibra óptica son virtualmente inútiles por sí mismos. Para ser prácticos se deben conectar a instalaciones electrónicas normales, lo cual requiere con frecuencia interconexiones costosas.

2. Resistencia: las fibras ópticas de por sí tienen una resistencia bastante menor a la tensión que los cables coaxiales. Esto se puede mejorar recubriendo la fibra con Kevlar normal y una chaqueta protectora de PVC.

3. Potencia eléctrica remota: a veces es necesario llevar energía eléctrica a un equipo remoto de interconexión o de regeneración. Esto no se puede hacer con el cable óptico, por lo que se deben agregar más cables metálicos en el cableado.

4. No están demostrados: los sistemas de cable de fibra óptica son relativamente nuevos, y no han tenido el tiempo suficiente para demostrar su confiabilidad.

5. Herramientas, equipo y adiestramiento especializados: las fibras ópticas requieren herramientas especiales para empalmar y reparar cables, y equipos especiales de prueba para hacer medidas rutinarias. También es difícil y costoso reparar cables de fibra, y los técnicos que trabajan con cables de fibra óptica necesitan también tener destrezas y adiestramiento especiales.

EL ESPECTRO ELECTROMAGNÉTICO

El esfuerzo total de frecuencias electromagnéticas se muestra en la fig. 11-1. Se puede ver que ese espectro de frecuencias va desde las *subsónicas* (unos cuantos hertz) hasta los *rayos cósmicos* (10^{22} Hz). El espectro de frecuencias luminosas se puede dividir en tres bandas generales:

1. *Infrarrojo:* banda de longitudes de onda de luz demasiado grandes para que las vea el ojo humano.
2. *Visible:* banda de longitudes de onda de luz a las que responde el ojo humano.
3. *Ultravioleta:* banda de longitudes de onda de luz que son demasiado cortas para que las vea el ojo humano.

Cuando se manejan ondas electromagnéticas de mayor frecuencia, como las luminosas, se acostumbra usar unidades de *longitud de onda* y no de frecuencia. La longitud de onda es la distancia que ocupa en el espacio un ciclo de una onda electromagnética. La longitud de onda depende de la frecuencia de la onda y de la velocidad de la luz. La relación matemática correspondiente es

$$\lambda = \frac{c}{f} \tag{11-1}$$

en donde λ = longitud de onda (metros/ciclo)
c = velocidad de la luz (300,000,000 metros por segundo)
f = frecuencia (hertz)

Con las frecuencias luminosas, la longitud de onda se suele expresar en *micrones* o *micras* (1 micrón = 1 micrómetro) o en *nanómetros* (1 nanómetro = 10^{-9} metros, o 0.001 micrón). Sin embargo, cuando se describe el espectro visible, la unidad *angstrom* (Å) se ha usado con frecuencia para expresar longitudes de onda (1 Å = 10^{-10} metros, o 0.0001 micrón). La fig. 11-2 muestra el espectro total de longitudes de ondas electromagnéticas.

DIAGRAMA DE BLOQUES DEL SISTEMA DE COMUNICACIONES CON FIBRA ÓPTICA

La fig. 11-3 muestra un diagrama de bloques simplificado de un enlace de comunicaciones con fibra óptica. Los tres bloques principales que lo forman son el *transmisor*, el *receptor* y la *guía*

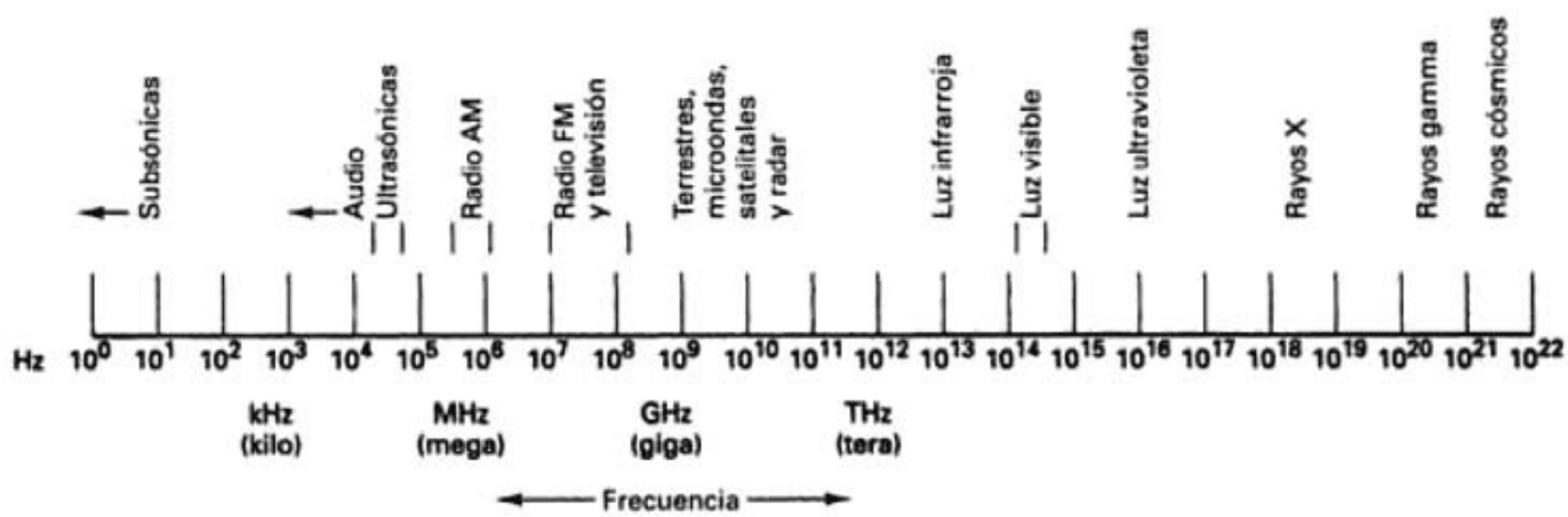


FIGURA 11-1 Espectro de frecuencias electromagnéticas

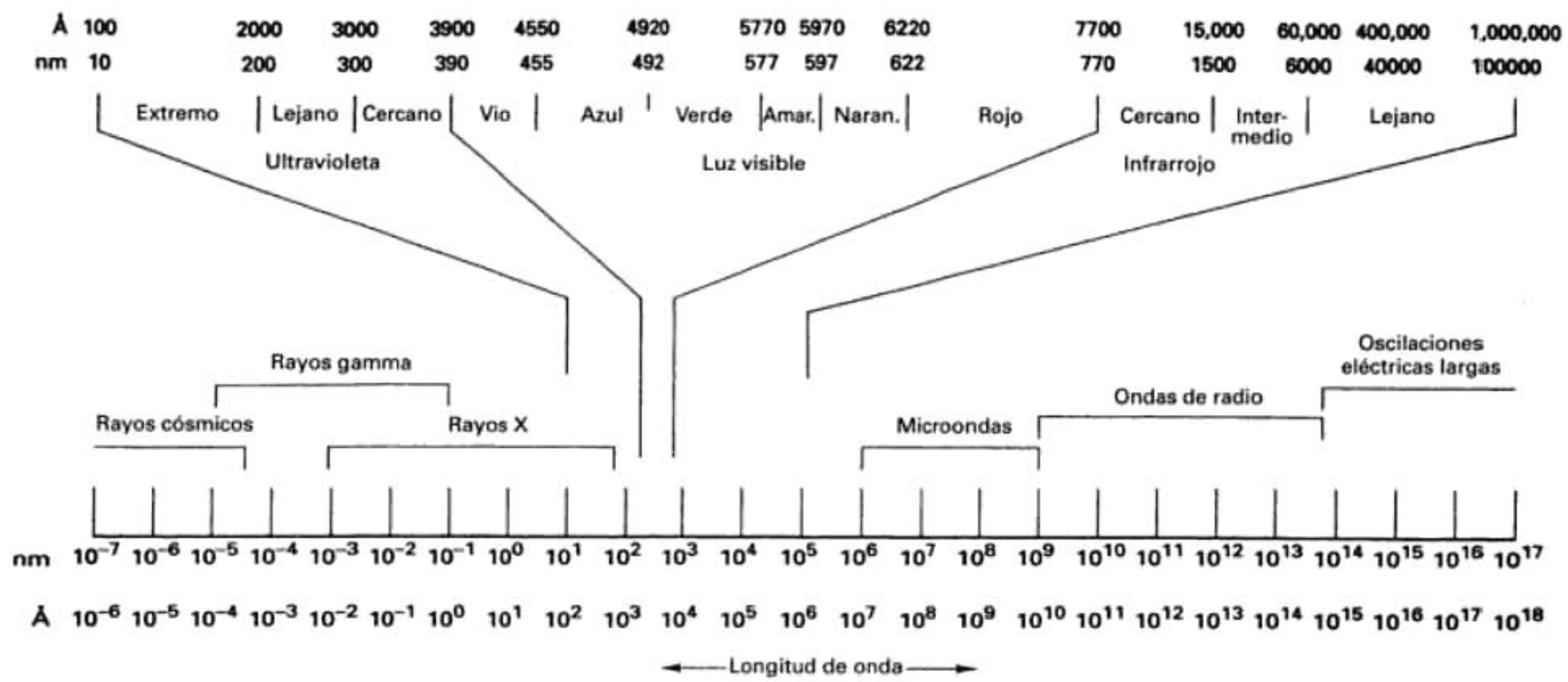


FIGURA 11-2 Espectro de longitudes de onda electromagnéticas

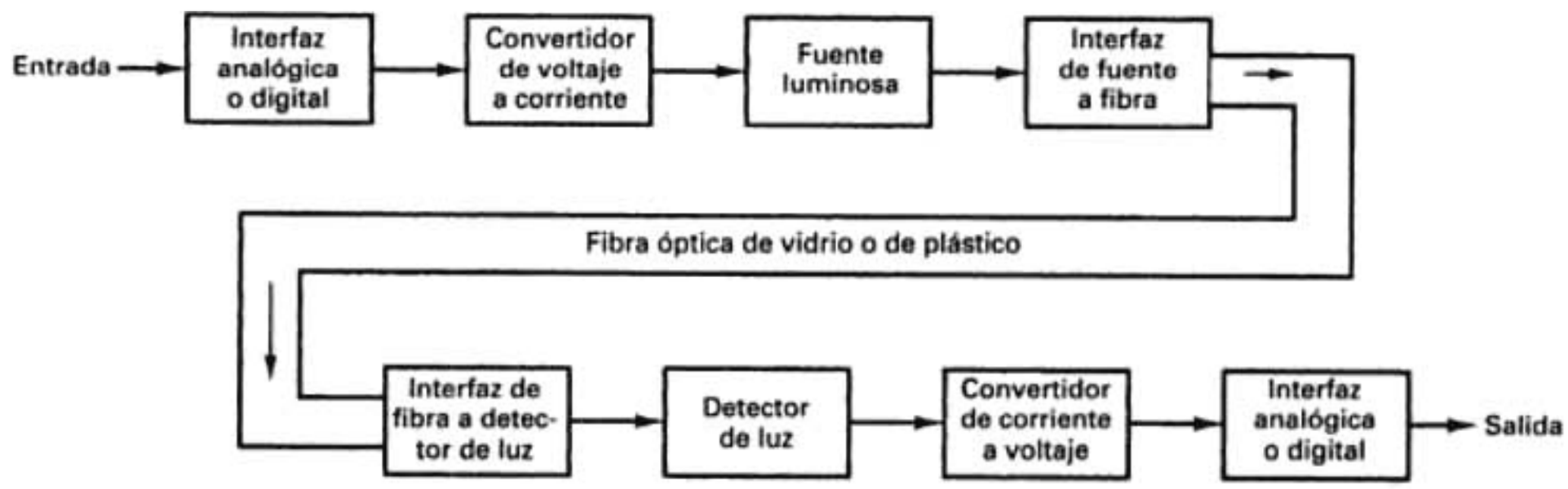


FIGURA 11-3 Enlace simplificado de comunicaciones con fibra óptica

de fibra o *fibra guía*. El transmisor consiste en una interconexión o interfaz analógica o digital, un convertidor de voltaje a corriente, una fuente luminosa y un acoplador de luz de fuente a fibra. La guía de fibra es un cable de vidrio o plástico ultrapuro. El receptor comprende un dispositivo detector acoplador de fibra a luz, un detector fotoeléctrico, un convertidor de corriente a voltaje, un amplificador y una interfaz analógica o digital.

En un transmisor de fibra óptica, la fuente luminosa se puede modular mediante una señal digital o una analógica. Para la modulación analógica, la interfaz de entrada compensa las impedancias y limita la amplitud de la señal de entrada. Para la modulación digital, la fuente original puede tener ya la forma digital o bien, si está en forma analógica, debe convertirse a una corriente de pulsos digitales. En el último caso se debe agregar un convertidor de analógico a digital en la interfaz.

El convertidor de voltaje a corriente sirve como interconexión eléctrica entre los circuitos de entrada y la fuente luminosa. Esta fuente luminosa puede ser un diodo emisor de luz (LED, de *light-emitting diode*) o un diodo de inyección láser (ILD, por *injection laser diode*). La cantidad de luz emitida por un LED o un ILD es proporcional a la cantidad de la corriente de excitación. Así, el convertidor de voltaje a corriente convierte un voltaje de señal de entrada en una corriente que se usa para activar la fuente luminosa.

El acoplador entre fuente y fibra (como por ejemplo un lente) es una interfaz mecánica. Su función es acoplar la luz que emite la fuente e introducirla al cable de fibra óptica. La fibra óptica es un núcleo de vidrio o plástico, un revestimiento y una chaqueta protectora. El dispositivo de acoplamiento detector de fibra a luz también es un acoplador mecánico. Su función es sacar tanta luz del cable de fibra como sea posible y ponerlo en el detector de luz.

El detector de luz es, con mucha frecuencia, un diodo PIN (tipo *p* tipo *n* *i*ntrínseco) o un fotodiodo de avalancha (APD, de *a*valanche *p*hoto*d*iode). Tanto el diodo APD como el PIN convierten la energía luminosa en corriente. En consecuencia, se necesita un convertidor de corriente a voltaje. El convertidor de corriente a voltaje transforma los cambios de la corriente del detector en cambios de voltaje de la señal de salida.

La interfaz analógica o digital en la salida del receptor también es una interconexión eléctrica. Si se usa modulación analógica, la interfaz compensa las impedancias y los niveles de señal con los circuitos de salida. Si se usa modulación digital, la interfaz debe incluir un convertidor de digital a analógico.

TIPOS DE FIBRA

En esencia hay tres variedades de fibra óptica que se usan en la actualidad. Las tres se fabrican con vidrio, plástico o una combinación de vidrio y plástico. Esas variedades son:

1. Núcleo y forro de plástico.
2. Núcleo de vidrio con forro de plástico (llamado con frecuencia fibra PCS, *plastic-clad silica* o sílice revestido con plástico).
3. Núcleo de vidrio y forro de vidrio (llamado con frecuencia SCS, *silica-clad silica* o sílice revestido con sílice).

En la actualidad se investiga, en Bell Laboratories, la posibilidad de usar una cuarta variedad que usa una sustancia *no silícea*, el *cloruro de zinc*. Los experimentos preliminares parecen

indicar que esta sustancia será hasta 1000 veces más eficiente que el vidrio, su contraparte a base de sílice.

Las fibras de plástico tienen varias ventajas sobre las de vidrio. La primera es que las de plástico son más flexibles y, en consecuencia, más robustas que el vidrio. Son fáciles de instalar, pueden resistir mejor los esfuerzos, son menos costosas y pesan 60% menos que el vidrio. La desventaja de las fibras de plástico es su alta atenuación característica: no propagan la luz con tanta eficiencia como el vidrio. En consecuencia, las fibras de plástico se limitan a tramos relativamente cortos, como por ejemplo dentro de un solo edificio o un complejo de edificios.

Las fibras con núcleos de vidrio tienen bajas atenuaciones características; sin embargo, las fibras PCS son un poco mejores que las SCS. Las fibras PCS también se afectan menos por la radiación y, en consecuencia, tienen mucho más atractivo en las aplicaciones militares. Las fibras SCS tienen las mejores características de propagación y son más fáciles de terminar que las PCS. Desafortunadamente, los cables SCS son los menos robustos y son más susceptibles a aumentos de atenuación cuando están expuestos a la radiación.

La selección de una fibra para determinada aplicación es función de los requisitos específicos del sistema. Siempre hay compromisos basados en la economía y la logística en una aplicación determinada.

Construcción del cable

Hay muchos diseños distintos de cable, en la actualidad. La fig. 11-4 muestra ejemplos de varias configuraciones de cables de fibra óptica. De acuerdo con la configuración, el cable puede consistir de un *núcleo*, un *revestimiento*, un *tubo protector*, *amortiguadores*, *miembros resistentes* y uno o más *forros* o *chaquetas de protección*.

En la construcción *suelta* de tubo (fig. 11-4a), cada fibra está contenida en un tubo de protección. Dentro del tubo protector hay un compuesto de poliuretano que encapsula, o sella, a la fibra, y evita la penetración del agua. Se puede presentar un fenómeno llamado *corrosión bajo tensiones*, o *fatiga estática* si la fibra de vidrio se expone durante largos periodos a humedades altas. Los cristales de dióxido de silicio interaccionan con la humedad y hacen que las pegaduras se suelten, causando fracturas espontáneas después de un periodo prolongado. Algunos de los cables de fibra tienen más de una capa protectora para asegurar que no se alteren las características de la fibra, si ésta queda expuesta a cambios extremos de temperatura. Rodeando al revestimiento de la fibra, se acostumbra tener una capa ya sea de laca, silicona o acrilato, que se aplica normalmente para sellar y preservar las características de resistencia y atenuación de la fibra.

La fig. 11-4b muestra la construcción de un cable *restringido* de fibra óptica. Hay un amortiguador primario y uno secundario, rodeando al cable de fibra. Las chaquetas amortiguadoras protegen a la fibra de influencias mecánicas externas que pudieran causar su ruptura, o demasiada atenuación óptica. El Kevlar es un material fibroso que aumenta la resistencia del cable a la tensión. De nuevo hay un tubo protector externo, lleno de poliuretano, que evita que la humedad se ponga en contacto con la fibra del núcleo.

La fig. 11-4c muestra una configuración de *hilos múltiples*. Para aumentar la resistencia a la tensión se incluyen en el paquete un miembro central de acero y una envoltura de cinta de Mylar. La fig. 11-4d muestra la configuración de *cinta*, que se ve con frecuencia en los sistemas telefónicos de fibra óptica. La fig. 11-4e representa las vistas de frente y lateral de un cable de sílice revestido con plástico.

Como se mencionó, una desventaja de los cables de fibra óptica es su falta de resistencia a la tensión, que puede ser sólo de una libra. Por esta razón, la fibra se debe reforzar con material tal que pueda resistir los esfuerzos mecánicos que probablemente se presentarán al ser jalada y maniobrada a través de los conductos subterráneos y aéreos, y al colgarse en los postes. Los materiales que se usan con frecuencia para reforzar y proteger a las fibras contra la abrasión y los esfuerzos ambientales son el acero, la fibra de vidrio, el plástico, FR-PVC (cloruro de polivinilo con retardante de llama), fibra de Kevlar y papel. El tipo de construcción de cable que se use depende de los requisitos de funcionamiento del sistema y de las restricciones económicas y ambientales.

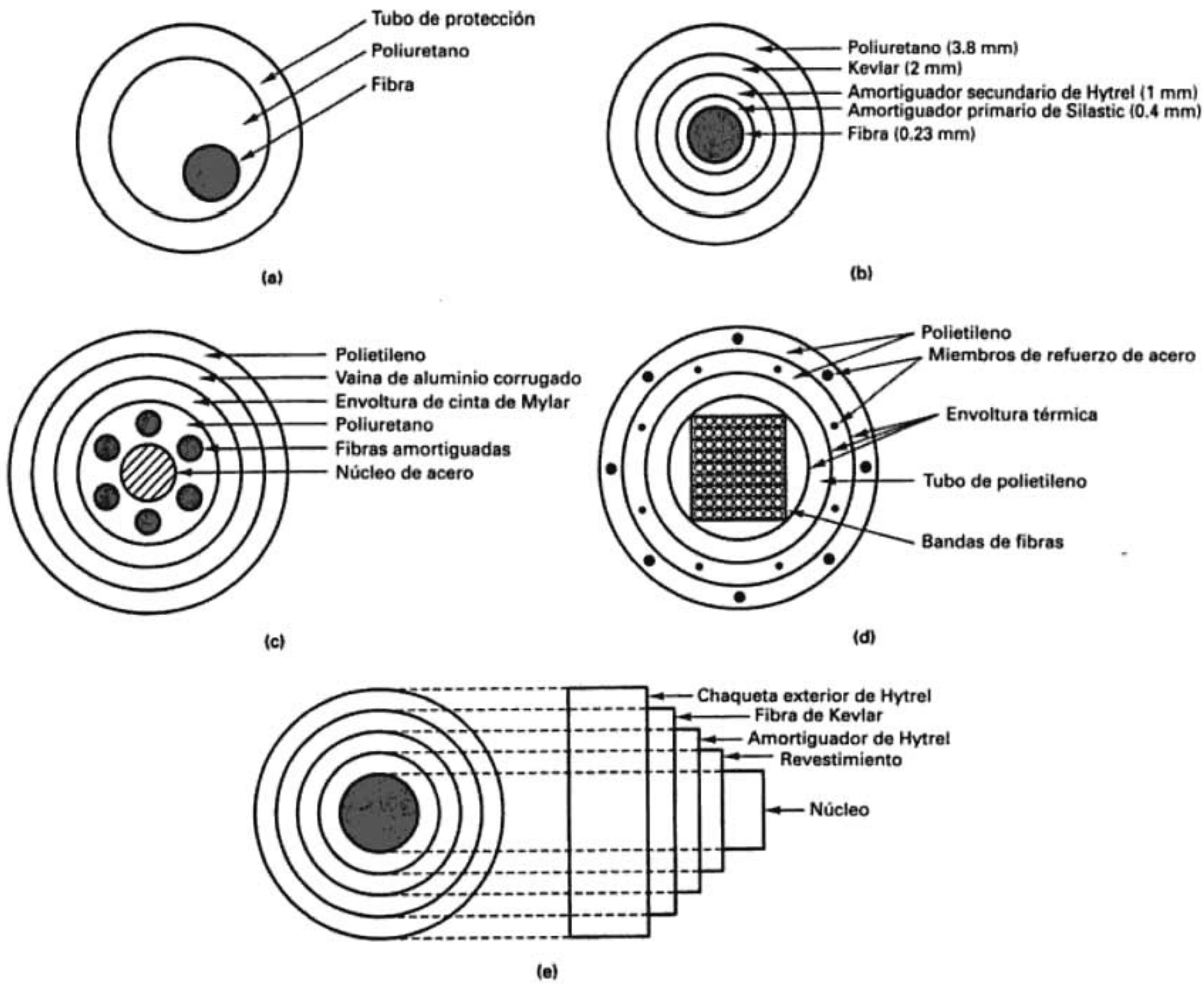


FIGURA 11-4 Configuraciones de cable de fibra óptica: (a) construcción suelta en tubo; (b) fibra restringida; (c) fibras múltiples; (d) cable telefónico; (e) cable de sílice revestida de plástico

PROPAGACIÓN DE LA LUZ

La física de la luz

Aunque se puede analizar por completo el funcionamiento de las fibras ópticas aplicando las ecuaciones de Maxwell, es una forma complicada por necesidad. Para la mayoría de las aplicaciones prácticas, se puede usar el *trazado geométrico de ondas* en lugar de las ecuaciones de Maxwell. Con el trazado de rayos se obtienen resultados con exactitud suficiente.

Un átomo tiene varios niveles de energía, o estados; el mínimo nivel es el estado fundamental. Un nivel de energía mayor que el estado fundamental se llama *estado excitado*. Si un átomo que tiene un nivel de energía decae a un nivel inferior, la pérdida de energía (en electrón volts), se emite en forma de fotón. La energía del fotón es igual a la diferencia entre las energías de los dos niveles. El proceso de decaer de un nivel de energía a otro se llama *decaimiento espontáneo* o *emisión espontánea*.

Los átomos pueden irradiarse mediante una fuente luminosa cuya energía sea igual a la diferencia entre el estado fundamental y otro nivel de energía. Esto puede hacer que un electrón

cambie de uno a otro nivel de energía, absorbiendo energía luminosa. El proceso de pasar de uno a otro nivel de energía se llama *absorción*. Cuando se hace la transición de un nivel a otro, el átomo absorbe un paquete de energía llamado *fotón*. Este proceso se parece al de la emisión.

La energía absorbida o emitida (el fotón) es igual a la diferencia entre los dos niveles de energía, esto es

$$E_2 - E_1 = E_p \tag{11-2}$$

en donde E_p es la energía del fotón. También,

$$E_p = hf \tag{11-3}$$

siendo h = constante de Planck = 6.625×10^{-34} J-s
f = frecuencia de la luz emitida (hertz)

También, la energía de los fotones se puede expresar en términos de longitud de onda. Al sustituir la ecuación 11-1 en la 11-3 se obtiene

$$E_p = hf \tag{11-4}$$

$$= \frac{hc}{\lambda}$$

Velocidad de propagación

La energía electromagnética, como la luz, recorre aproximadamente 300,000,000 m/s en el espacio libre. También, la velocidad de propagación es igual para todas las frecuencias de luz en el espacio libre. Sin embargo, se ha demostrado que en materiales más densos que el espacio libre, la velocidad se reduce. Cuando se reduce la velocidad de una onda electromagnética al pasar de uno a otro medio de un material más denso, el rayo de luz se *refracta*, o se *dobla*, hacia la normal. También, en materiales más densos que el espacio libre, todas las frecuencias de la luz no se propagan con la misma velocidad.

Refracción

La fig. 11-5a muestra cómo se refracta un rayo de luz al pasar de un material de determinada densidad a un material menos denso. En realidad, el rayo de luz no se dobla, sino que cambia de dirección en la interfaz. La fig. 11-5b muestra cómo la luz solar, que contiene todas las frecuencias de la luz, se afecta al pasar a través de un material más denso que el espacio libre. La refracción se presenta en la interfaz entre aire y vidrio. Las longitudes de onda del violeta se refractan más, y las del rojo se refractan menos. Esta separación espectral de la luz blanca se llama *refracción prismática*. Es el fenómeno que producen los arco iris: las gotitas de agua en la atmósfera funcionan como prismas pequeños que dividen la luz blanca del sol en las distintas longitudes de onda y crean un espectro visible de colores.

Índice de refracción

La cantidad de desviación o refracción que sucede en la interfaz de dos materiales de distintas densidades se puede predecir bastante bien, y depende del *índice de refracción* de los dos materiales. Este índice de refracción no es más que la relación de la velocidad de propagación de un rayo de luz en el espacio libre, entre la velocidad de propagación del rayo en determinado material. La ecuación correspondiente es

$$n = \frac{c}{v}$$

en la que c = velocidad de la luz en el espacio libre (300,000,000 metros por segundo)
v = velocidad de la luz en determinado material (metros por segundo)

Aunque el índice de refracción también es una función de la frecuencia, la variación, en la mayoría de las aplicaciones, es insignificante y en consecuencia se omite de esta descripción. Los índices de refracción de varios materiales comunes aparecen en la tabla 11-1.

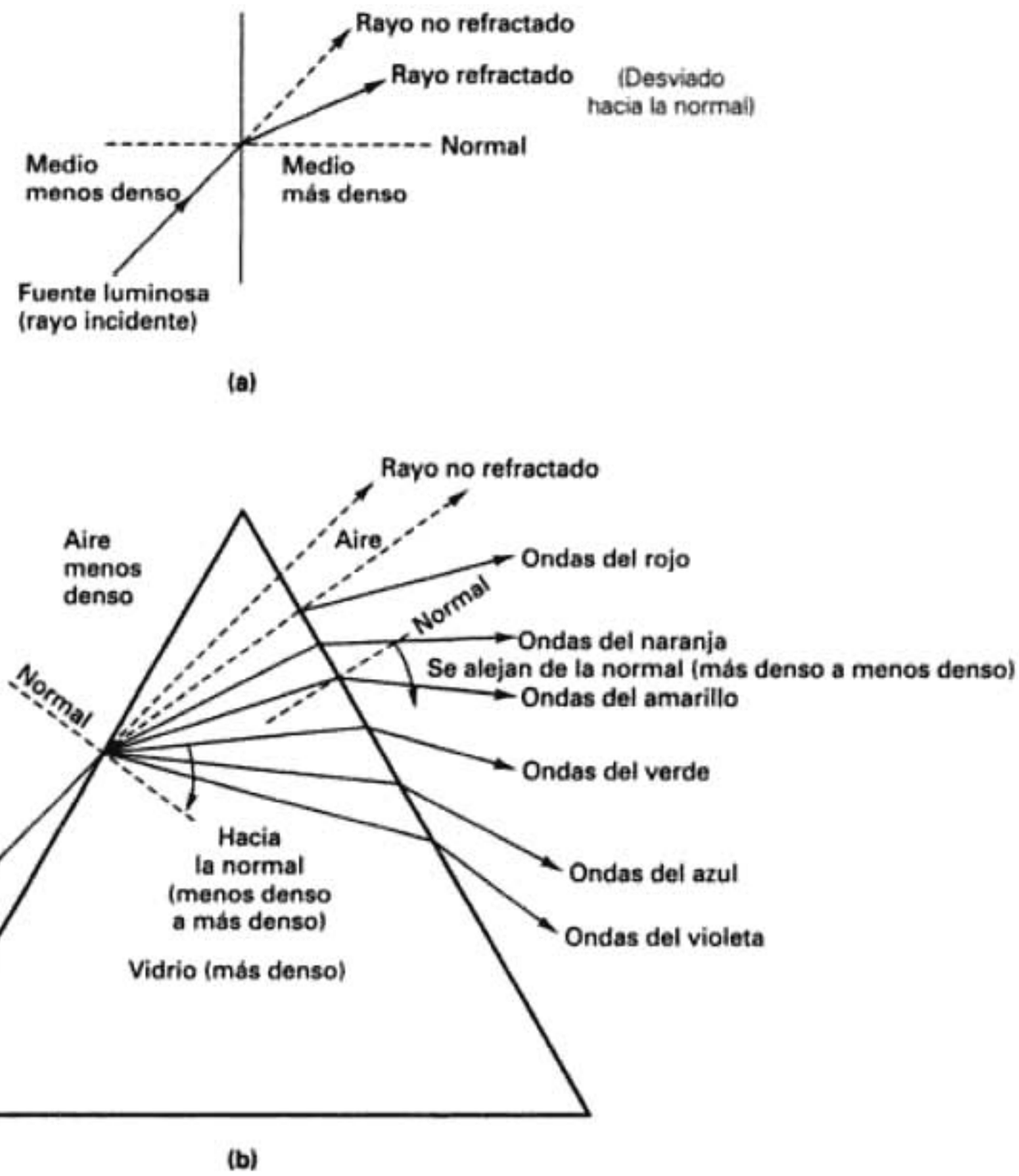


FIGURA 11-5 Refracción de la luz: (a) refracción; (b) refracción en prisma

TABLA 11-1 Algunos índices de refracción

Medio	Índice de refracción*
Vacío	1.0
Aire	1.0003 (≈1.0)
Agua	1.33
Alcohol etílico	1.36
Cuarzo fundido	1.46
Fibra de vidrio	1.5–1.9
Diamante	2.0–2.42
Silicio	3.4
Arseniuro de galio	3.6

Densidad creciente ↓

Al aumentar la densidad disminuye la velocidad de propagación y aumenta el índice de refracción (ya que el índice de refracción es inversamente proporcional a la velocidad y directamente proporcional a la densidad)

* Basado en una longitud de onda emitida por una llama de sodio (5890 Å)

Se puede explicar la forma en que un rayo de luz se refracta al encontrar la interfaz de dos materiales transmisores con índices de refracción distintos, con la ley de Snell. Esa ley establece que

$$n_1 \operatorname{sen} \theta_1 = n_2 \operatorname{sen} \theta_2 \tag{11-5}$$

en donde n_1 = índice de refracción del material 1 (adimensional)
n_2 = índice de refracción del material 2 (adimensional)
θ_1 = ángulo de incidencia (grados)
θ_2 = ángulo de refracción (grados)

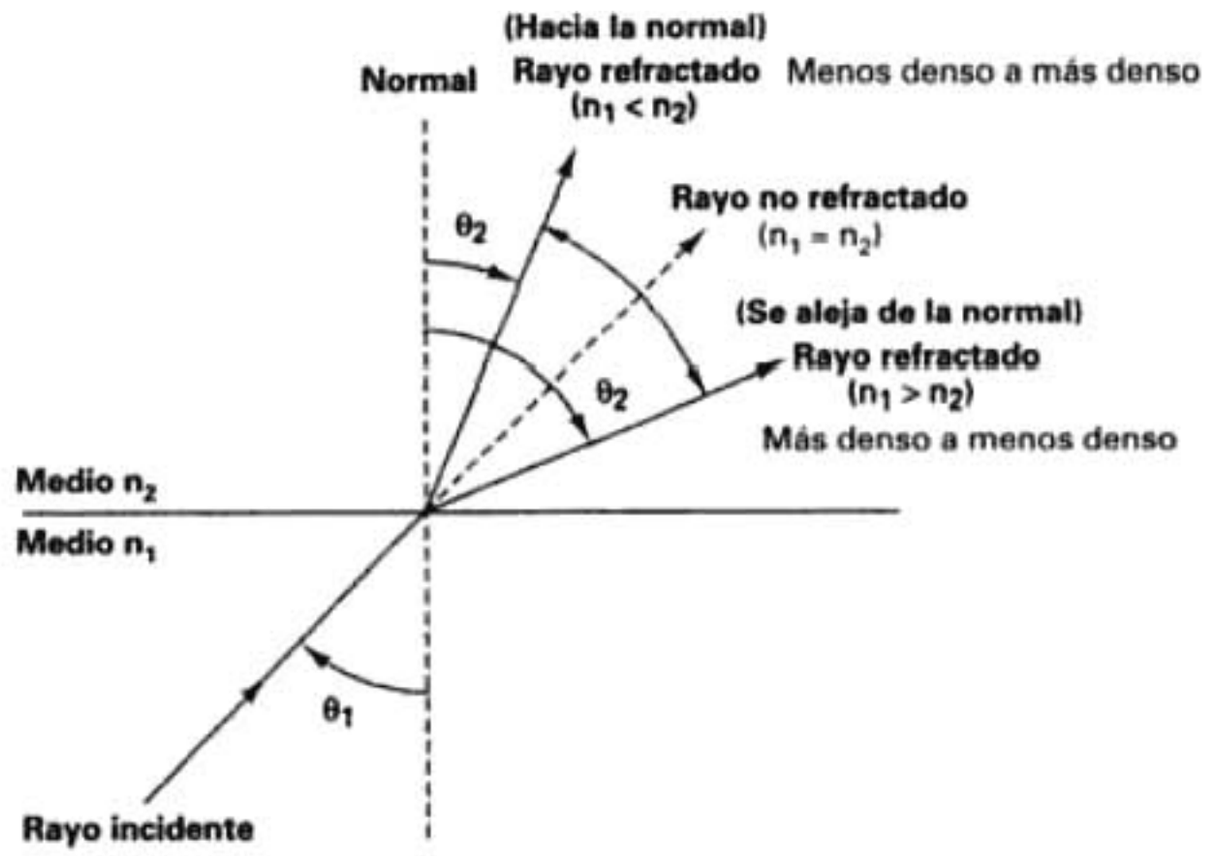


FIGURA 11-6 Modelo para la ley de Snell de refracción

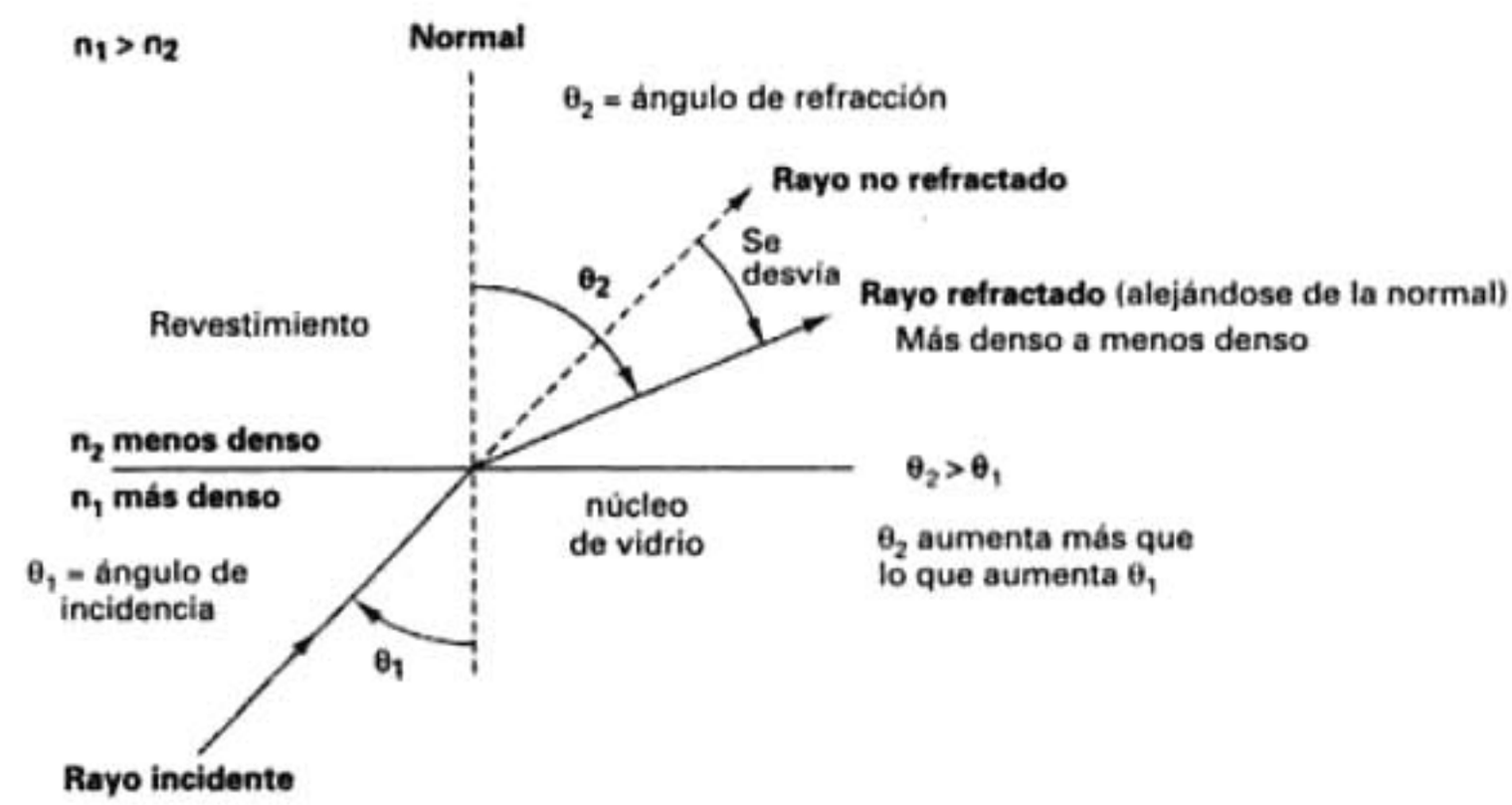


FIGURA 11-7 Rayo de luz que se refracta alejándose de la normal

En la fig. 11-6 se presenta un modelo de la ley de Snell para índices de refracción. En la interfaz, se puede refractar el rayo incidente hacia la normal, o alejándose de ella, dependiendo de si n_1 es menor o mayor que n_2.

La fig. 11-7 muestra cómo se refracta un rayo de luz al pasar de un material más denso (con mayor índice de refracción) a uno menos denso (menor índice de refracción). Se ve que el rayo de luz cambia de dirección en la interfaz, y que el ángulo de refracción es mayor que el de incidencia. En consecuencia, cuando un rayo de luz entra a un material menos denso, se dobla alejándose de la normal. La normal no es más que una línea perpendicular a la interfaz, en el punto en el que el rayo incidente llega a la interfaz. De igual manera, cuando un rayo de luz entra a un material más denso, se desvía hacia la normal.

Ejemplo 11-1

En la fig. 11-7, sea vidrio el medio 1 y alcohol etílico el medio 2. Para un ángulo de incidencia de 30°, determinar el ángulo de refracción.

Solución De acuerdo con la tabla 11-1,

$$n_1 \text{ (vidrio)} = 1.5$$
$$n_2 \text{ (alcohol etílico)} = 1.36$$

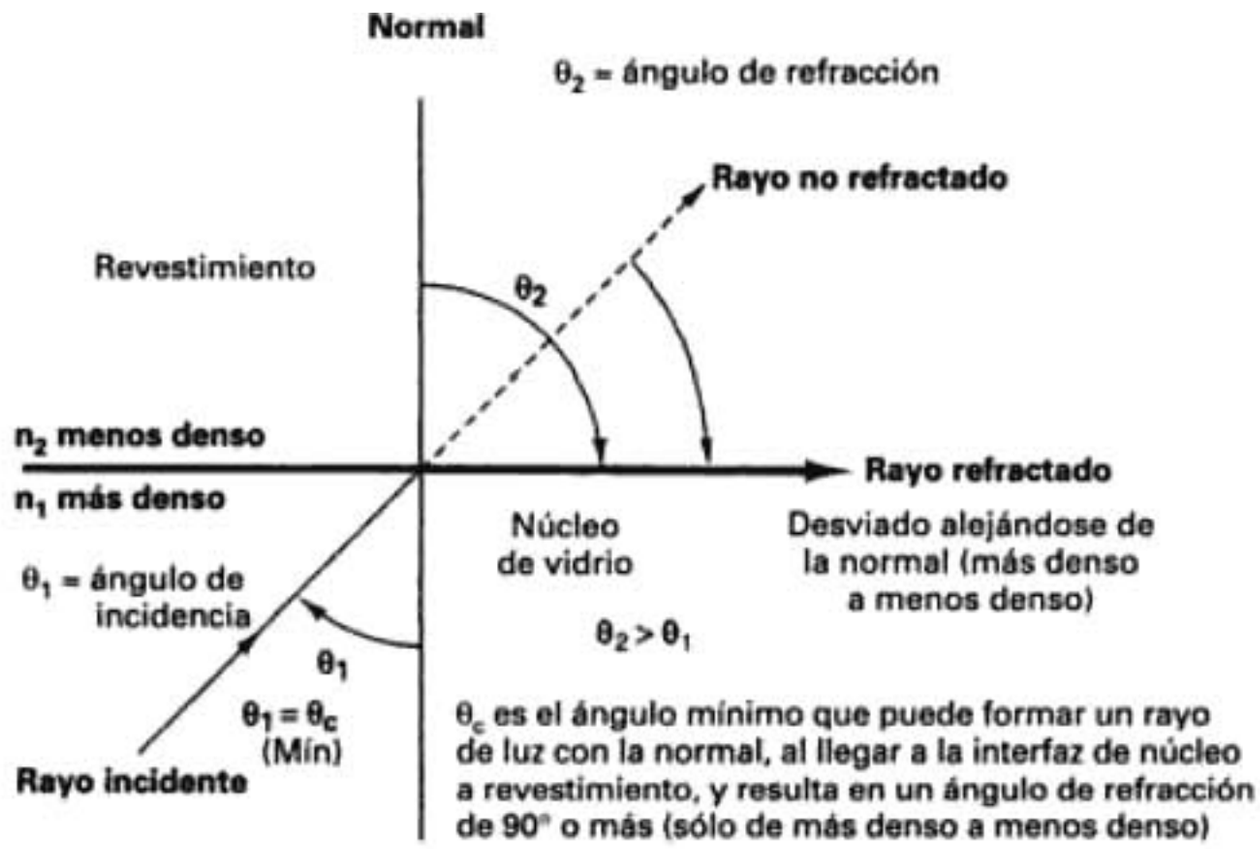


FIGURA 11-8 Refracción en ángulo crítico

Se reordena la ecuación 11-5, y al sustituir n_1, n_2 y θ

$$\frac{n_1}{n_2}\,\text{sen}\,\theta_1 = \text{sen}\,\theta_2$$

$$\frac{1.5}{1.36}\,\text{sen}\,30 = 0.5514 = \text{sen}\,\theta_2$$

$$\theta_2 = \text{sen}^{-1}\,0.5514 = 33.47°$$

Este resultado indica que el rayo de luz se refractó o cambió de dirección formando un ángulo de 33.47° con la normal a la interfaz. Como la luz pasaba de un material más denso (vidrio) a uno menos denso (alcohol), cambió su dirección alejándose de la normal.

Ángulo crítico

La fig. 11-8 muestra un caso en el que un *rayo incidente* forma un ángulo tal que el ángulo de refracción es 90°, y el rayo refractado va por la interfaz. Es importante observar que el rayo de luz va de un medio de mayor índice de refracción a un medio con menor índice. De nuevo, si se aplica la ley de Snell

$$\text{sen}\,\theta_1 = \frac{n_2}{n_1}\,\text{sen}\,\theta_2$$

Con $\theta_2 = 90°$,

$$\text{sen}\,\theta_1 = \frac{n_2}{n_1}(1) \qquad \text{o} \qquad \text{sen}\,\theta_1 = \frac{n_2}{n_1}$$

y

$$\text{sen}^{-1}\frac{n_2}{n_1} = \theta_1 = \theta_c \tag{11-6}$$

en donde θ_c es el ángulo crítico.

El *ángulo crítico* se define como el ángulo de incidencia mínimo en el cual un rayo de luz puede llegar a la interfaz entre dos medios y tener un ángulo de refracción de 90° o mayor. Esta definición sólo se aplica cuando el rayo de luz pasa de un medio más denso a uno menos denso. Si el ángulo de refracción es 90° o menor, el rayo de luz no puede penetrar en el material menos denso. En consecuencia, se produce la reflexión total en la interfaz, y el ángulo de reflexión es igual al ángulo de incidencia. La fig. 11-9 muestra una comparación entre los ángulos de refracción y de reflexión, cuando el ángulo de incidencia es menor que, o mayor que el ángulo crítico.

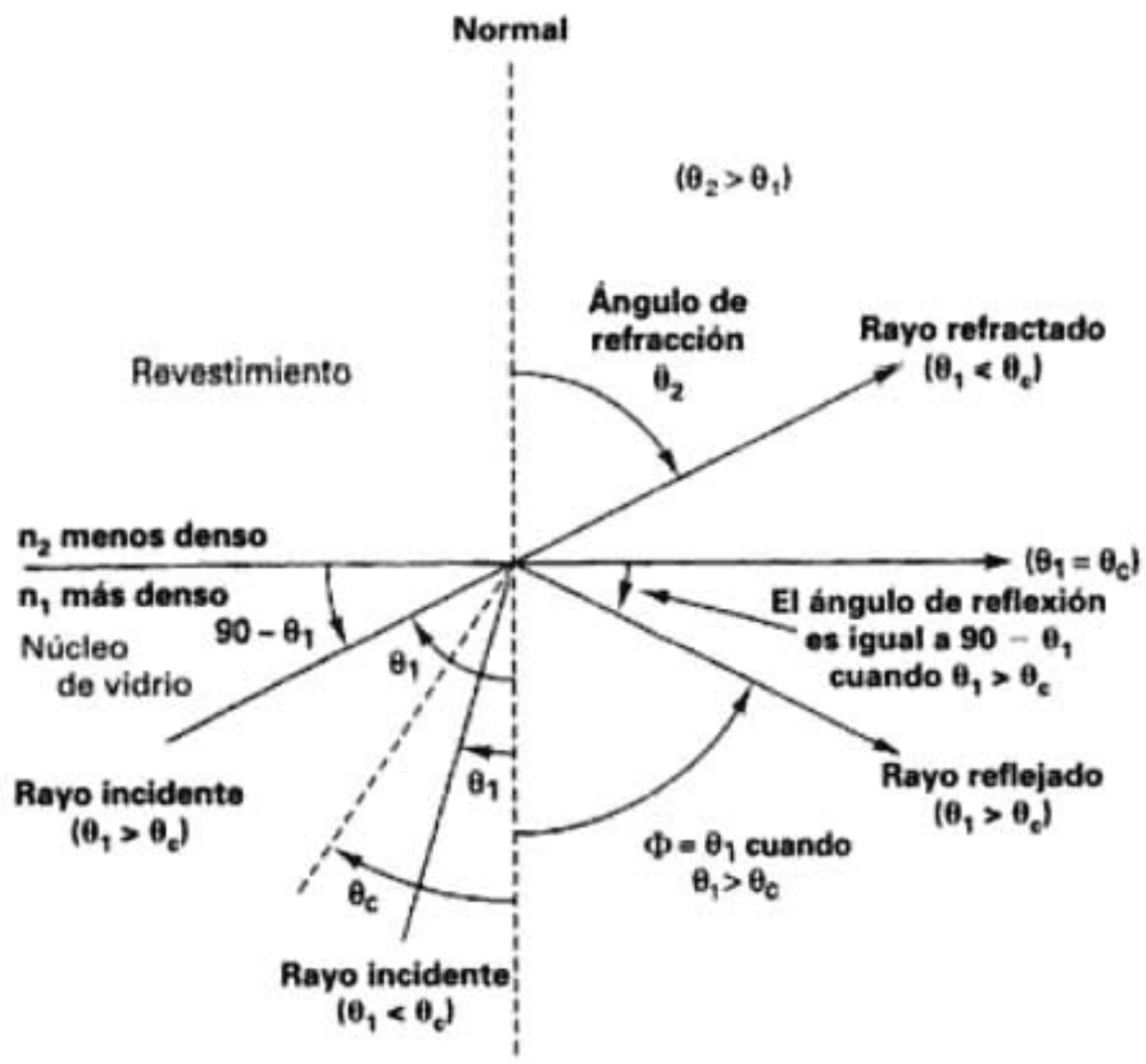


FIGURA 11-9 Ángulos de reflexión y de refracción

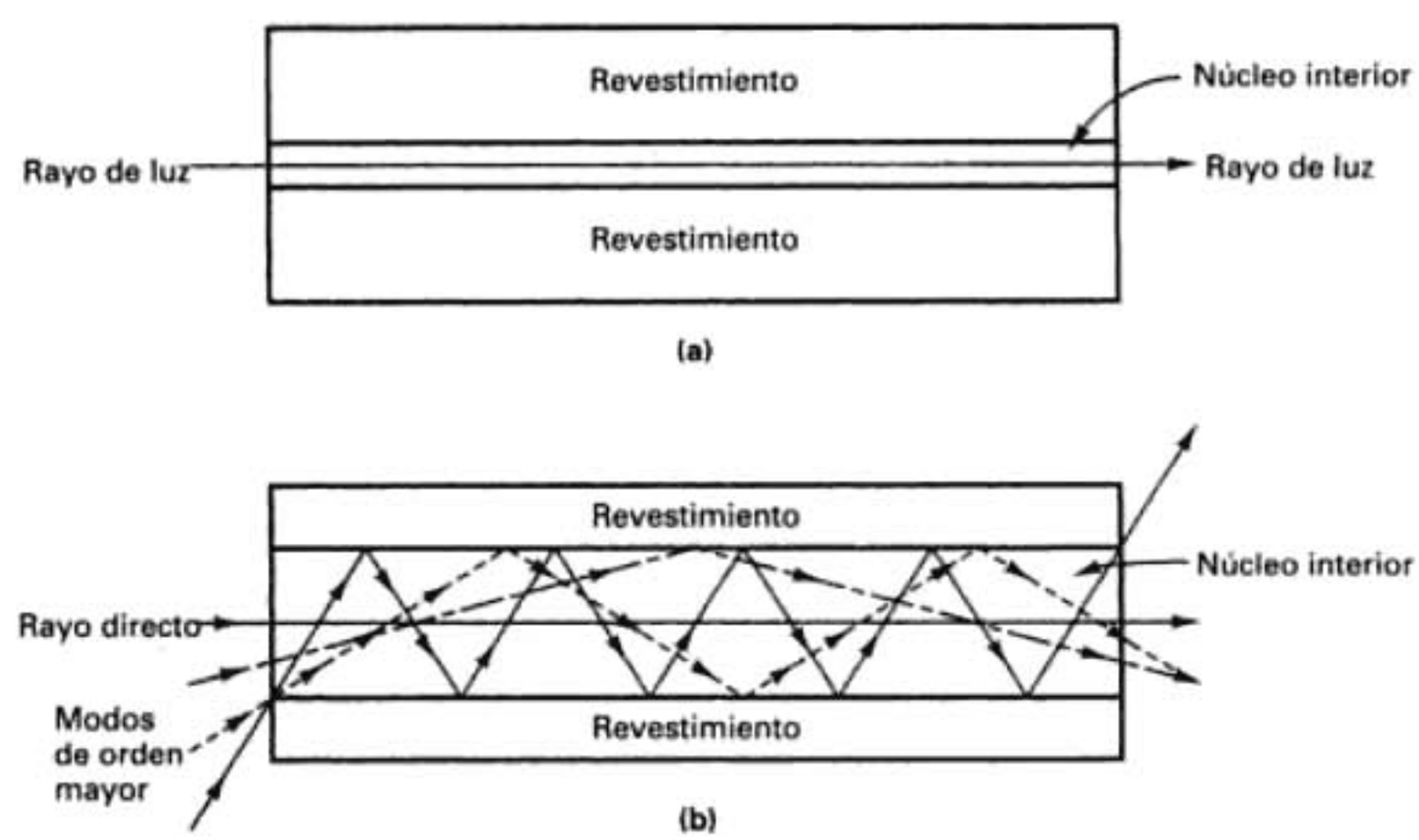


FIGURA 11-10 Modos de propagación: (a) unimodal; (b) multimodal

PROPAGACIÓN DE LA LUZ A TRAVÉS DE UNA FIBRA ÓPTICA

La luz se puede propagar por un cable de fibra óptica por reflexión o por refracción. La forma en que se propague depende del *modo de propagación* y del *perfil de índice(s)* de la fibra.

Modo de propagación

En la terminología de fibras ópticas, la palabra *modo* simplemente quiere decir camino. Si sólo hay una trayectoria que pueda tener la luz por el cable, se llama *modo único*, o *unimodal*. Si hay más de una trayectoria, se llama *modo múltiple* o *multimodal*. La fig. 11-10 muestra la propagación de luz en modo único y en multimodo, por una fibra óptica.

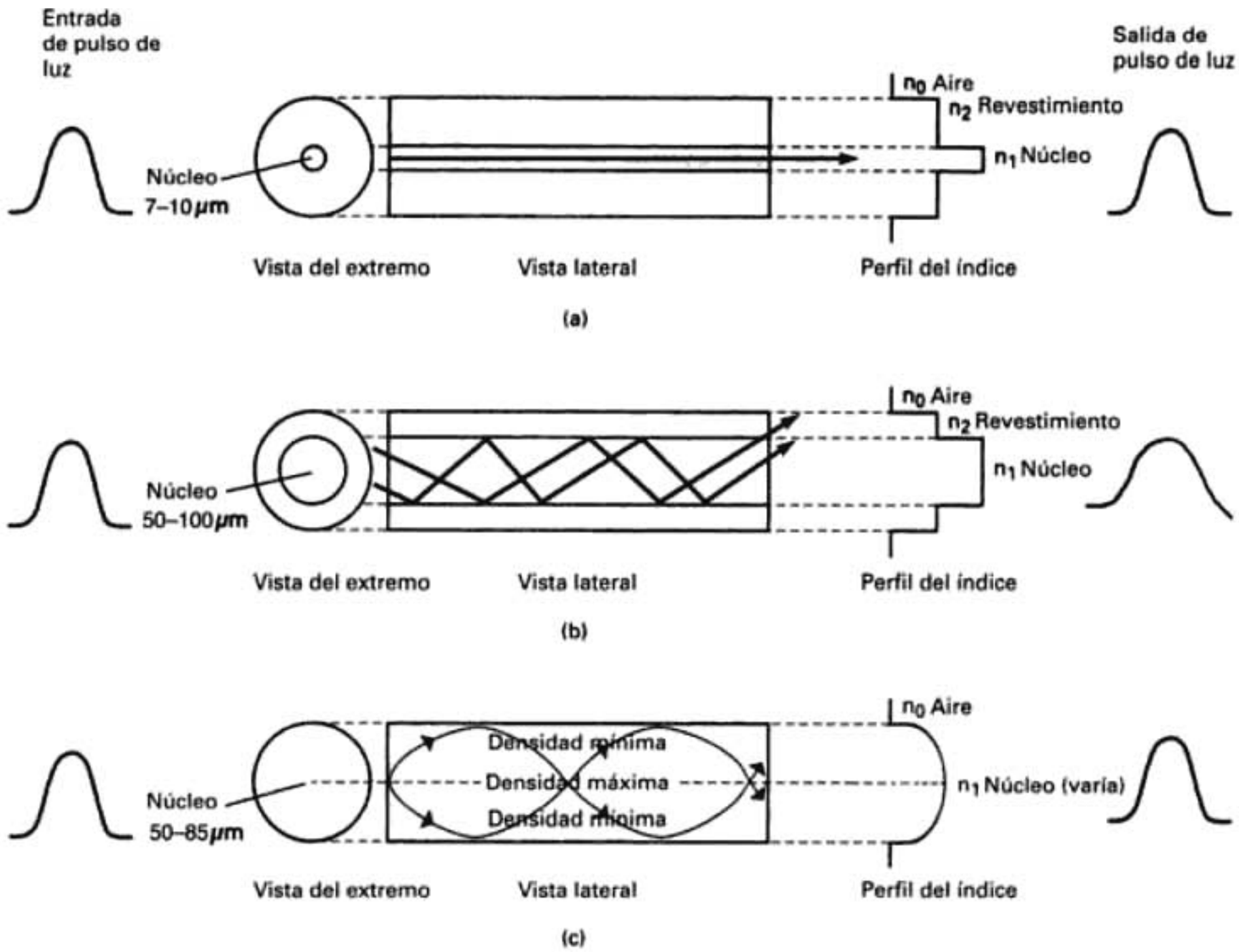


FIGURA 11-11 Perfiles de índice en el núcleo: (a) unimodal, índice escalonado; (b) multimodal, índice escalonado; (c) multimodal, índice graduado

Perfil índice

El perfil índice de una fibra óptica es una representación gráfica del índice de refracción en la sección transversal de la fibra. El índice de refracción se grafica en el eje horizontal, y el eje de la distancia radial al centro es el vertical. La fig. 11-11 muestra los perfiles índices del núcleo para tres clases de cables de fibra.

Hay dos tipos básicos de perfiles índice: escalonados y graduados. Una *fibra de índice escalonado* tiene un núcleo central con índice de refracción uniforme. Este núcleo está rodeado por un revestimiento externo con índice de refracción uniforme, pero menor que el del núcleo central. Se ve en la fig. 11-11 que en una fibra de índice escalonado hay un cambio abrupto de índice de refracción en la interfaz entre núcleo y revestimiento. En una *fibra de índice graduado* no hay revestimiento, y el índice de refracción del núcleo no es uniforme; es máximo en el centro y disminuye en forma gradual de acuerdo con la distancia hacia la orilla externa.

CONFIGURACIONES DE FIBRA ÓPTICA

En esencia hay tres clases de configuración de fibra óptica: índice escalonado unimodal, índice escalonado multimodal e índice graduado multimodal.

Fibra unimodal de índice escalonado

La *fibra unimodal de índice escalonado* tiene un núcleo central con la pequeñez suficiente como para que en esencia sólo haya una trayectoria que pueda seguir la luz para propagarse por el cable. Esta clase de fibras se ve en la fig. 11-12. En su forma más simple, el revestimiento externo sólo es de aire (vea la fig. 11-12a). El índice de refracción del núcleo de vidrio (n_1) es aproximada-

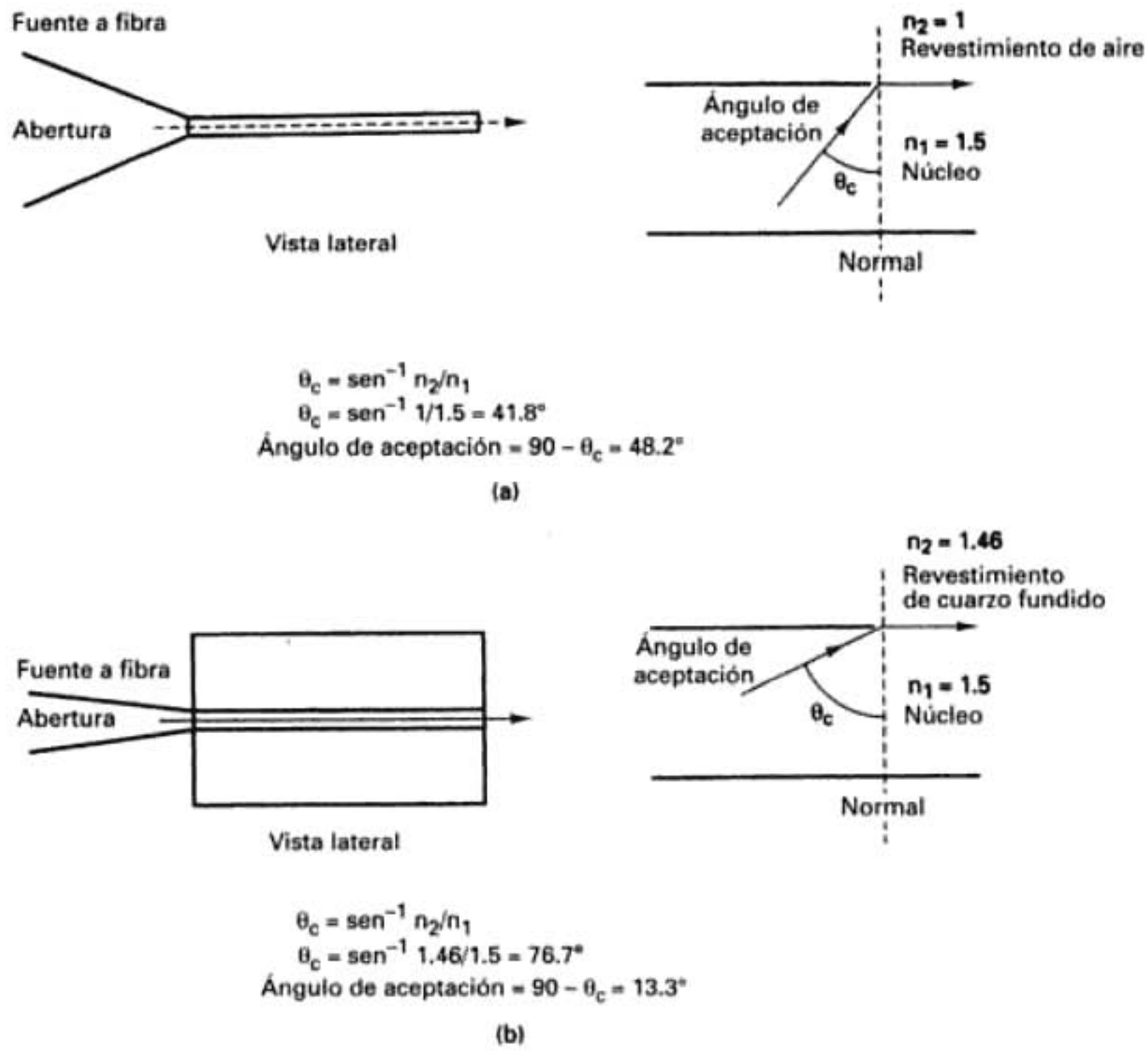


FIGURA 11-12 Fibras unimodales de índice escalonado: (a) revestimiento de aire; (b) revestimiento de vidrio

mente 1.5, y el del revestimiento de aire (n_O) es 1. La gran diferencia de índices de refracción hace que el ángulo crítico sea pequeño (unos 42°) en la interfaz vidrio-aire. En consecuencia, la fibra tomará la luz de una gran abertura. Esto simplifica algo acoplar la luz de una fuente hacia el cable. Sin embargo, este tipo de fibra suele ser muy débil y sus usos prácticos son limitados.

Una clase más práctica de fibra unimodal escalonada es la que tiene un revestimiento distinto del aire (fig. 11-12b). El índice de refracción del revestimiento (n_2) es un poco mayor que el del núcleo central (n_1) y es uniforme en todo el revestimiento. Esta clase de cables es físicamente más resistente que los de fibra con revestimiento de aire, pero el ángulo crítico también es mucho mayor, unos 77°. Lo anterior produce un ángulo de aceptación pequeño y una abertura pequeña entre la fuente y la fibra, dificultando mucho más acoplar la luz de una fuente para que pase al interior de la fibra.

Con ambos tipos de fibra unimodal de índice escalonado, la luz se propaga por la fibra por reflexión. Los rayos luminosos que entran a la fibra se propagan directamente por el núcleo o quizá se reflejen una vez. En consecuencia, todos los rayos de luz siguen más o menos la misma trayectoria por el cable, y tardan más o menos el mismo tiempo en recorrer el tramo de cable. Ésta es una ventaja abrumadora de las fibras unimodales de índice escalonado, y se explicará con más detalle más adelante en este capítulo.

Una fibra óptica unimodal transmitirá un solo modo para todas las longitudes de onda que sean mayores que su longitud de onda de corte. Esta longitud de onda de corte se calcula como sigue

$$\lambda_c = \frac{2\pi a n_1 \sqrt{2\Delta}}{2.405}\ (\mu\text{m}) \tag{11-7}$$

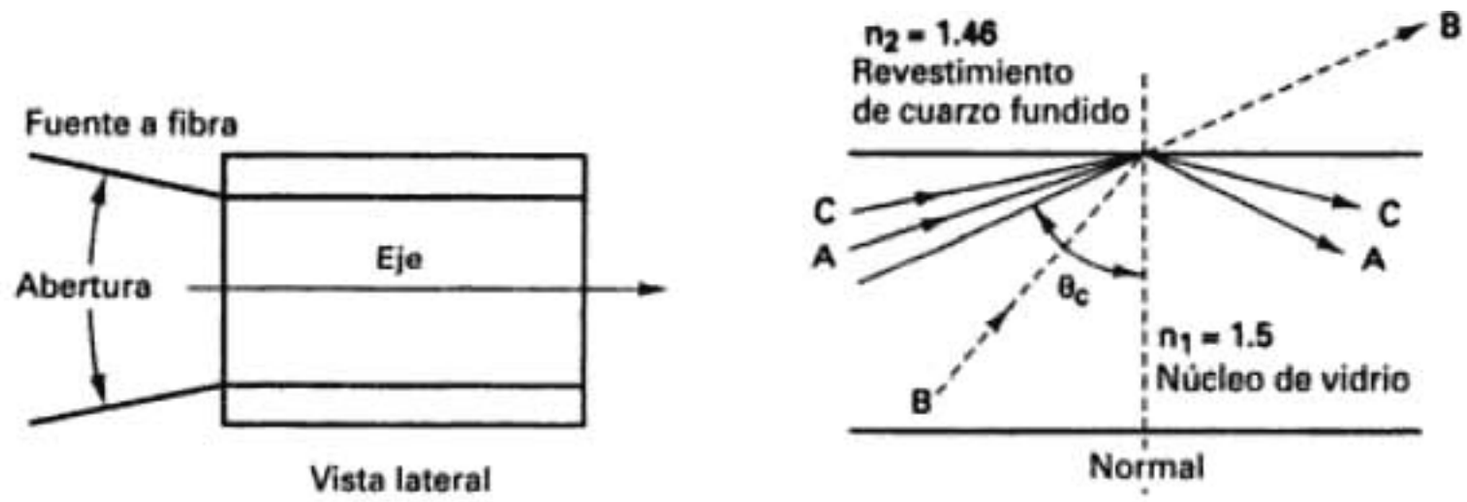


FIGURA 11-13 Fibra multimodal de índice escalonado

siendo λ_c = longitud de onda de corte (μm)
n_1 = índice de refracción del núcleo (adimensional)
n_2 = índice de refracción del revestimiento (adimensional)
a = radio del núcleo (μm)

$$\Delta = \frac{n_1 - n_2}{n_1} \text{ (adimensional)}$$

Ejemplo 11-2

Calcular la longitud de onda de corte para una fibra óptica de 3 μm de diámetro con índices de refracción de núcleo y de revestimiento de 1.545 y 1.510, respectivamente.

$$\Delta = \frac{1.545 - 1.510}{1.545} = 0.023$$

$$a = \frac{3\ \mu\text{m}}{2} = 1.5\ \mu\text{m}$$

$$\lambda_c = \frac{(2\pi)(1.5\ \mu\text{m})(1.545)\sqrt{2(0.023)}}{2.405}$$

$$= 1.29\ \mu\text{m}$$

Fibra multimodal de índice escalonado

En la fig. 11-13 se ve una *fibra multimodal de índice escalonado*. Se parece a la configuración unimodal, pero el núcleo central es mucho mayor. Esta clase de fibra tiene una abertura grande de luz a fibra y en consecuencia permite la entrada de más luz al cable. Los rayos luminosos que llegan a la interfaz entre núcleo y revestimiento formando un ángulo mayor que el crítico (rayo A) se propagan por el núcleo en zigzag, reflejándose continuamente en la interfaz. Los rayos luminosos que llegan a la interfaz entre el núcleo y el revestimiento formando un ángulo menor que el crítico (rayo B) y que entran al revestimiento se pierden. Se puede ver que hay muchas trayectorias que puede seguir un rayo de luz al propagarse por la fibra. El resultado es que no todos los rayos de luz siguen la misma trayectoria y, en consecuencia, no tardan lo mismo para recorrer la longitud de la fibra.

Fibra multimodal de índice graduado

En la fig. 11-14 se ve una *fibra multimodal de índice graduado*. Estas fibras se caracterizan por un núcleo central cuyo índice de refracción es no uniforme; es máximo en el centro y disminuye en forma gradual hacia la orilla externa. La luz se propaga por esta clase de fibras por refracción. Al propagarse un rayo de luz en dirección diagonal por el núcleo hacia el centro, pasa continuamente de una fase menos densa a una más densa. En consecuencia, los rayos luminosos se refractan en forma constante, y eso produce su deflexión continua. La luz entra a la fibra formando muchos ángulos diferentes. Al propagarse por la fibra, los rayos que viajan en la zona

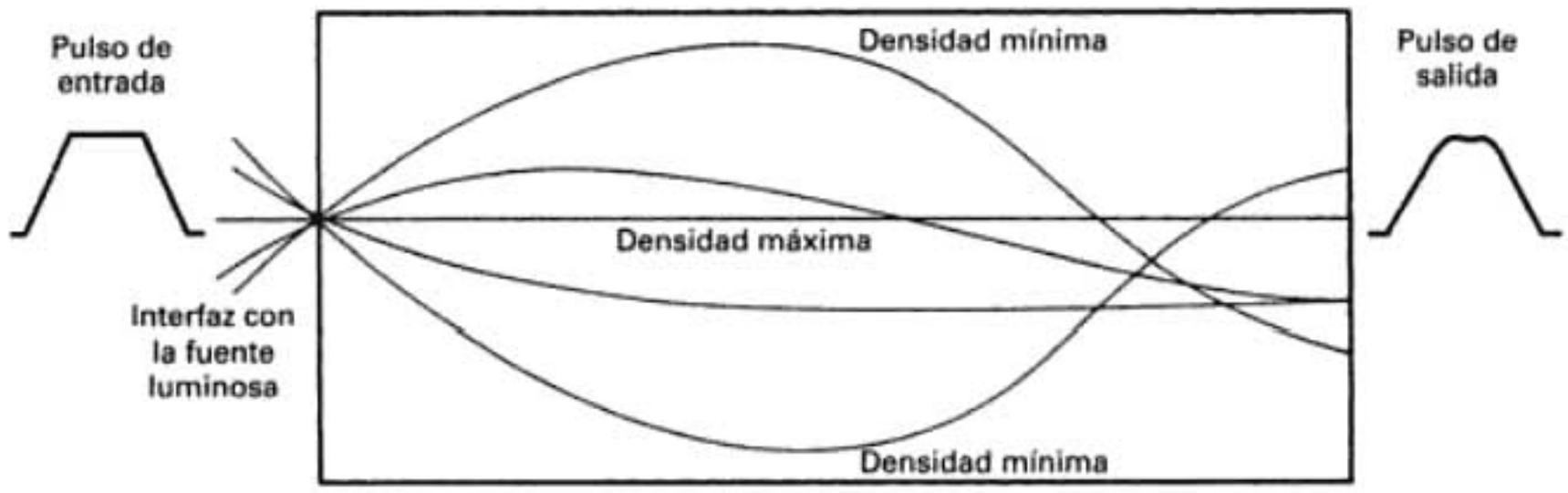


FIGURA 11-14 Fibra multimodal de índice graduado

más externa de la fibra recorren mayor distancia que los que van cerca del centro. En vista de que el índice de refracción disminuye con la distancia al centro, y que la velocidad es inversamente proporcional al índice de refracción, la luz que va más alejada del centro se propaga con mayor velocidad. En consecuencia, se tardan aproximadamente lo mismo en recorrer la longitud de la fibra.

Comparación de los tres tipos de fibra óptica

Fibra unimodal de índice escalonado

Ventajas

1. Hay dispersión mínima. Como todos los rayos que se propagan por la fibra toman aproximadamente la misma trayectoria, tardan aproximadamente el mismo tiempo para recorrer el cable. En consecuencia, un pulso de luz que entra al cable se puede reproducir con mucha exactitud en el extremo de recepción.
2. Debido a la gran exactitud de reproducción de los pulsos transmitidos en el extremo de recepción, son posibles mayores anchos de banda y mayores capacidades de transmisión de información con las fibras unimodales de índice escalonado que con los otros tipos de fibra.

Desventajas

1. Debido a que el núcleo central es muy pequeño, es difícil acoplar la luz hacia adentro y hacia afuera de esta clase de fibra. La abertura de la fuente a la fibra es la más pequeña de todos los tipos de fibra.
2. También debido al pequeño núcleo central, se requiere una fuente luminosa muy direccional, como por ejemplo un láser, para acoplar la luz en una fibra unimodal de índice escalonado.
3. Las fibras unimodales de índice escalonado son costosas y difíciles de fabricar.

Fibra multimodal de índice escalonado

Ventajas

1. Las fibras multimodales de índice escalonado son poco costosas, y su fabricación es sencilla.
2. Es fácil acoplar la luz hacia adentro y hacia afuera de las fibras multimodales de índice escalonado; tienen una abertura grande de la fuente a la fibra.

Desventajas

1. Los rayos luminosos siguen muchas trayectorias distintas por la fibra, lo que da como resultado grandes diferencias en sus tiempos de propagación. Por eso, los rayos que recorren esta clase de fibras tienden a extenderse y, en consecuencia, un pulso de luz que se propague por una fibra multimodal de índice escalonado se distorsiona más que en otros tipos de fibra.
2. El ancho de banda y la capacidad de transferencia de información posibles con este tipo de cables es menor que con los demás tipos.

Fibra multimodal de índice graduado. En esencia, este tipo de fibra no tiene ventajas ni desventajas sobresalientes. Las fibras multimodales de índice escalonado son más fáciles de

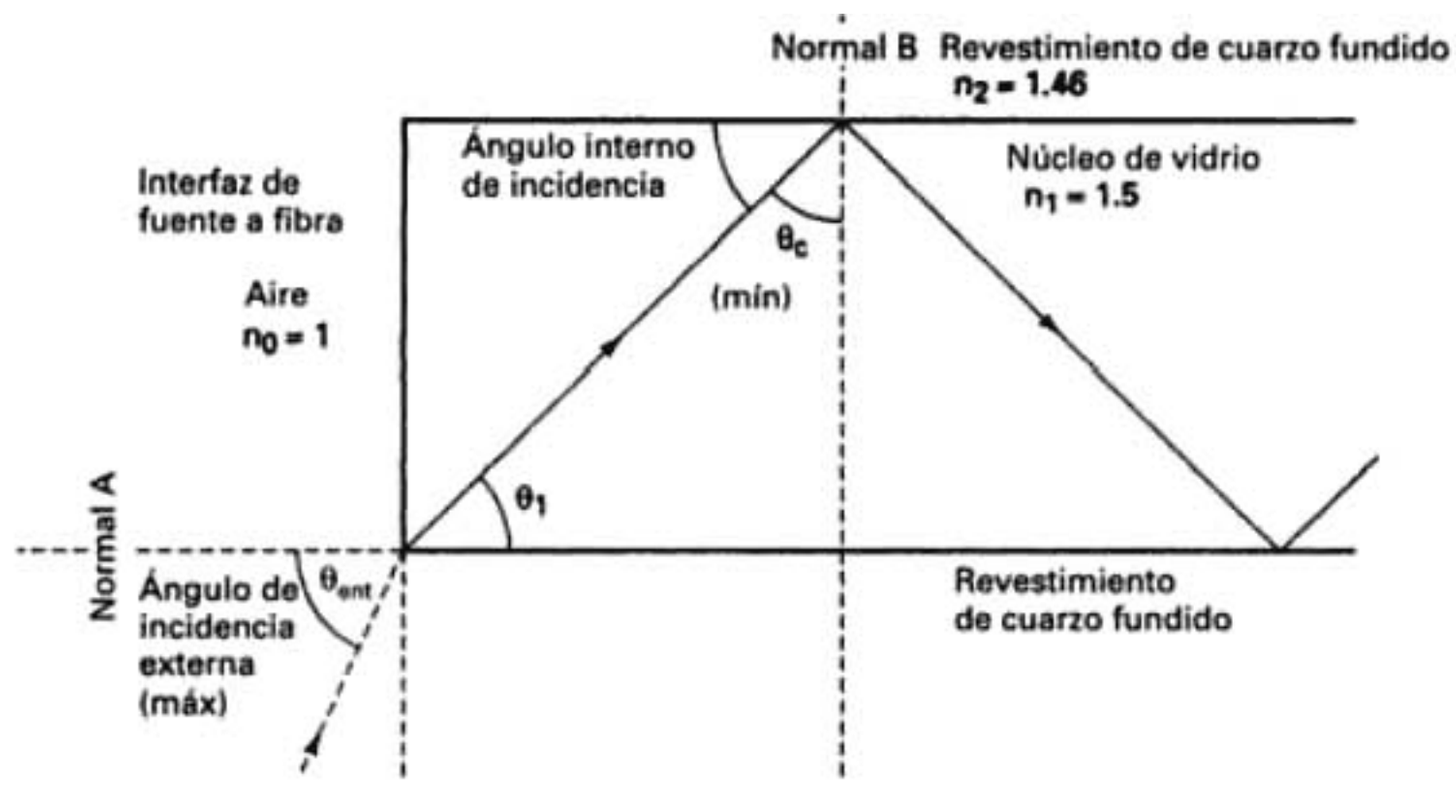


FIGURA 11-15 Propagación de un rayo al entrar y recorrer un cable de fibra óptica

acoplar la luz que les entra y que les sale, en comparación con las fibras unimodales de índice escalonado, pero más difíciles de acoplar en comparación con las multimodales de índice escalonado. La distorsión debida a trayectorias múltiples de propagación es mayor que en las fibras unimodales de índice escalonado, pero menor que en las multimodales de índice escalonado. Son más fáciles de fabricar las fibras de índice graduado que las unimodales de índice escalonado, pero más difíciles que las multimodales de índice escalonado. La fibra multimodal de índice graduado se considera intermedia en comparación con los otros tipos.

ÁNGULO DE ACEPTACIÓN Y CONO DE ACEPTACIÓN

En las descripciones anteriores se mencionó la *abertura de la fuente a la fibra* en varias ocasiones, y se explicaron los ángulos *crítico* y *de aceptación* en el punto en el que un rayo de luz llega a la interfaz de núcleo y revestimiento. La siguiente descripción se refiere a la capacidad de recogida de luz que tiene la fibra, la capacidad de acoplar la luz de la fuente hacia el interior del cable de fibra.

La fig. 11-15 representa el extremo del cable de fibra que da a la fuente. Cuando los rayos de luz entran a la fibra, llegan a la interfaz aire-vidrio en la normal A. El índice de refracción del aire es 1, y el del núcleo de vidrio es 1.5. En consecuencia, la luz que entra a la interfaz aire-vidrio se propaga desde un medio menos denso a uno más denso. Bajo estas condiciones, y de acuerdo con la ley de Snell, los rayos de luz se refractan acercándose a la normal. Eso hace que los rayos cambien de dirección y se propaguen en dirección diagonal por el núcleo, en el ángulo θ_c, distinto del ángulo externo de incidencia θ_{ent} en la interfaz aire-vidrio. Para que un rayo de luz se propague por el cable debe llegar a la interfaz entre el núcleo y revestimiento formando un ángulo que sea mayor que θ_c, el ángulo crítico.

Si se aplica la ley de Snell al ángulo externo de incidencia se obtiene la siguiente ecuación

$$n_0 \operatorname{sen} \theta_{ent} = n_1 \operatorname{sen} \theta_1 \tag{11-8}$$

y

$$\theta_1 = 90 - \theta_c$$

Por lo que

$$\operatorname{sen} \theta_1 = \operatorname{sen} (90 - \theta_c) = \cos \theta_c \tag{11-9}$$

Se sustituye la ecuación 11-9 en la 11-8 y se obtiene lo siguiente

$$n_0 \operatorname{sen} \theta_{ent} = n_1 \cos \theta_c$$

De aquí se despeja θ_{ent} y se obtiene

$$\operatorname{sen} \theta_{ent} = \frac{n_1}{n_0} \cos \theta_c \tag{11-10}$$

La fig. 11-16 muestra la relación geométrica de la ecuación 11-10.

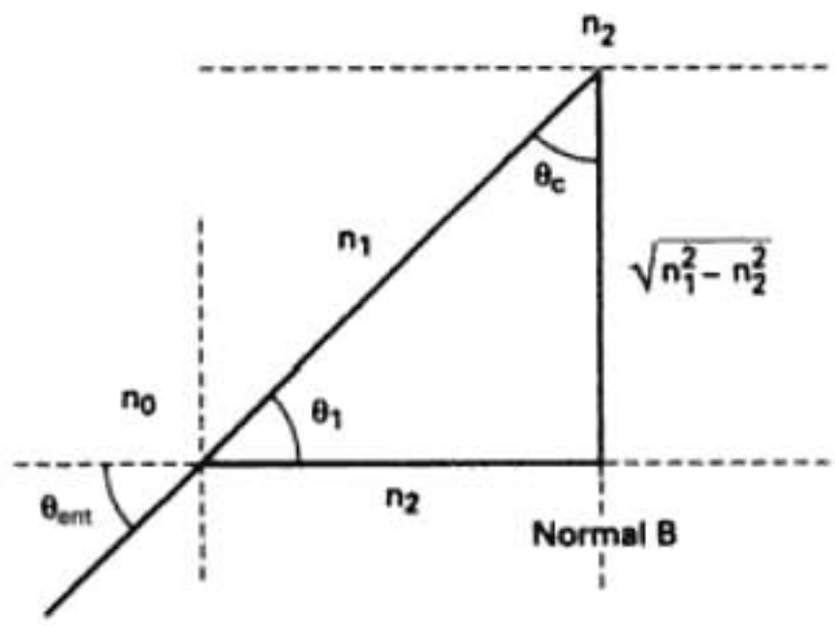


FIGURA 11-16 Relaciones geométricas en la ecuación 11-10

De acuerdo con la fig. 11-16 y con el teorema de Pitágoras

$$\cos \theta_c = \frac{\sqrt{n_1^2 - n_2^2}}{n_1} \tag{11-11}$$

Se sustituye la ecuación 11-11 en la 11-10 y se obtiene

$$\text{sen}\ \theta_{ent} = \frac{n_1}{n_0} \frac{\sqrt{n_1^2 - n_2^2}}{n_1}$$

Simplificando, resulta

$$\text{sen}\ \theta_{ent} = \frac{\sqrt{n_1^2 - n_2^2}}{n_0} \tag{11-12}$$

y

$$\theta_{ent} = \text{sen}^{-1} \frac{\sqrt{n_1^2 - n_2^2}}{n_0} \tag{11-13}$$

Como en general los rayos de luz entran a la fibra desde el aire, n_0 es igual a 1. Con eso se simplifica la ecuación 11-13 como sigue

$$\theta_{ent(máx)} = \text{sen}^{-1} \sqrt{n_1^2 - n_2^2} \tag{11-14}$$

A θ_{ent} se le llama *ángulo de aceptación,* o medio ángulo del *cono de aceptación.* Define al ángulo máximo que pueden formar los rayos luminosos externos al llegar a la interfaz aire-fibra para poder propagarse por la fibra, con una respuesta no mayor de 10 dB menos que el valor máximo. En la fig. 11-17a se ve el ángulo de aceptación. Al girar este ángulo en torno al eje de la fibra se obtiene el *cono de aceptación* de la entrada de la fibra, que se ve en la fig. 11-17b.

Abertura numérica

La *abertura numérica* (NA, de *numerical aperture*) se relaciona en forma estrecha con el ángulo de aceptación y es la figura de mérito que se suele usar para medir la magnitud del ángulo de aceptación. En esencia, se usa la abertura numérica para describir la capacidad de reunir la luz que tiene una fibra óptica. Mientras mayor sea la magnitud de la abertura numérica, la fibra acepta mayor cantidad de luz externa. Para fibras unimodales o multimodales, la abertura numérica se define como el seno del ángulo máximo que un rayo de luz que entra a la fibra puede formar con el eje de la fibra, y propagarse por el cable por reflexión interna, es decir, el seno del medio ángulo de aceptación. Los rayos de luz que entran al cable desde el exterior del cono de aceptación entrarán al revestimiento y, en consecuencia, no se propagarán por el cable. Así, para la luz que entra a la fibra desde el espacio libre, la abertura numérica es

$$\text{NA} = \text{sen}\ \theta_{ent} \tag{11-15}$$

y

$$\text{NA} = \sqrt{n_1^2 - n_2^2} \tag{11-16}$$

También,

$$\text{sen}^{-1}\ \text{NA} = \theta_{ent} \tag{11-17}$$

en donde NA = abertura numérica (adimensional)

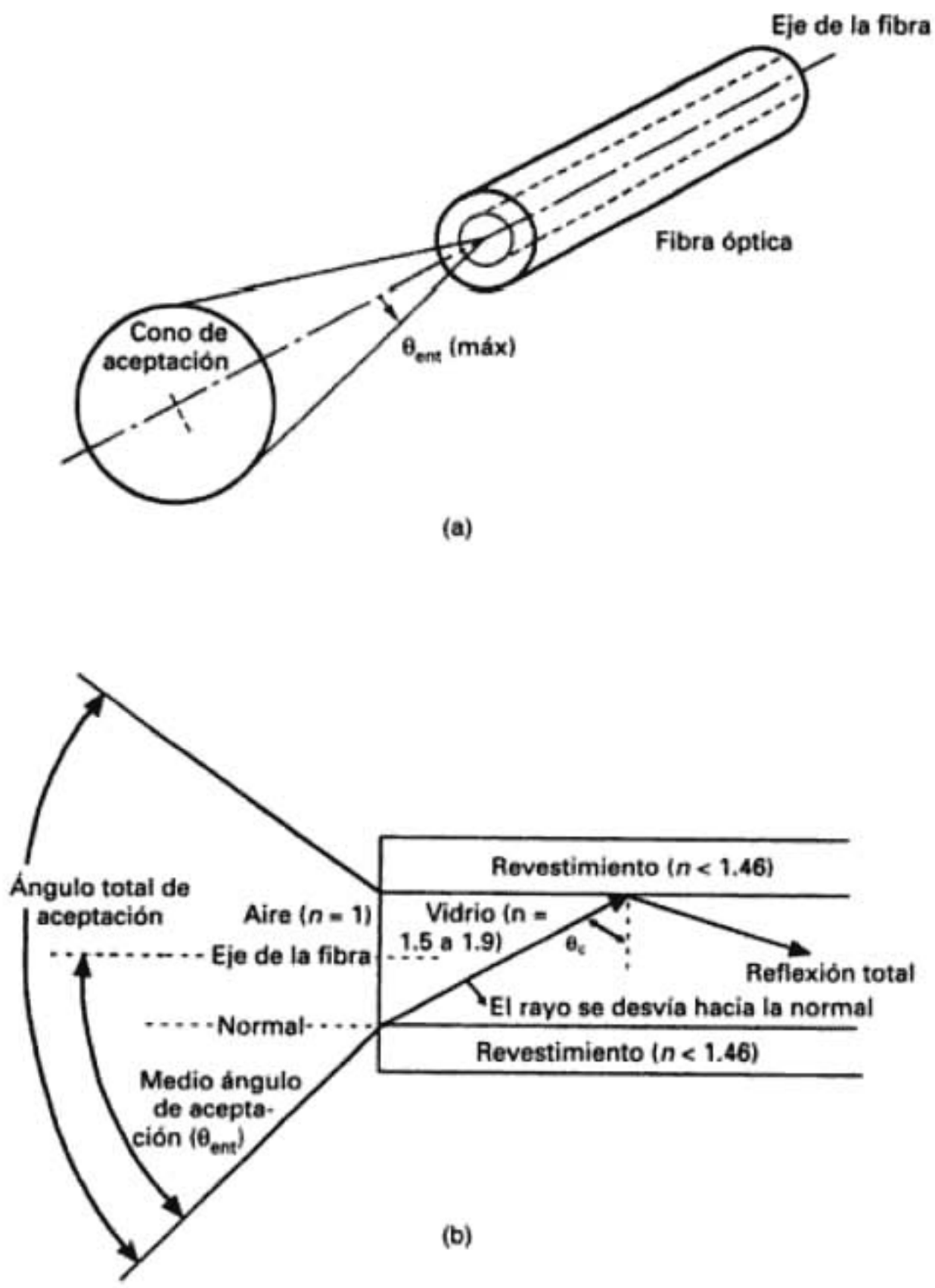


FIGURA 11-17 (a) Cono de aceptación; (b) ángulo de aceptación

n_1 = índice de refracción del núcleo de vidrio (adimensional)
n_2 = índice de refracción del revestimiento (adimensional)
θ_{ent} = ángulo de aceptación (grados)

Un diámetro mayor de núcleo no necesariamente produce una mayor abertura numérica; sin embargo, en la práctica los núcleos de fibra grande tienden a tener aberturas numéricas mayores, de 0.3 o más. En general, las fibras de índice graduado tienen aberturas numéricas de 0.2 a 0.3, y las fibras unimodales de 0.1 a 0.15. La abertura numérica de las fibras unimodales no suele ser tan importante como en las multimodales, porque la luz que sale de una fibra que funciona en modo único no se abre en la misma forma que lo hace de las fibras multimodales.

La abertura numérica se puede calcular con la ecuación 11-15, pero en la práctica se suele medir viendo la salida de una fibra. Debido a que las propiedades de guía de luz que tiene un cable de fibra son simétricas, la luz sale del cable y se reparte, pero formando un ángulo igual al ángulo de aceptación.

Ejemplo 11-3

En este ejemplo, véase la fig. 11-15. Calcular θ_c, el ángulo crítico, θ_{ent}, el ángulo de aceptación y la abertura numérica de una fibra multimodal de índice escalonado con núcleo de vidrio ($n_1 = 1.5$) y revestimiento de cuarzo fundido ($n_2 = 1.46$). El medio entre la fuente y la fibra es aire.

Solución Se sustituyen valores en la ecuación 11-6 y se obtiene

$$\theta_c = \text{sen}^{-1}\frac{n_2}{n_1} = \text{sen}^{-1}\frac{1.46}{1.5} = 76.7°$$

Al sustituir en la ecuación 11-14 resulta

$$\theta_c = \text{sen}^{-1}\sqrt{n_1^2 - n_2^2} = \text{sen}^{-1}\sqrt{1.5^2 - 1.46^2} = 20.2°$$

De acuerdo con la ecuación 11-15

$$\text{NA} = \text{sen}\ \theta_{ent} = \text{sen}\ 20.2 = 0.344$$

El diámetro máximo que puede tener una fibra óptica unimodal es proporcional a la longitud de onda del rayo de luz que entra al cable, y a la abertura numérica de la fibra. El radio máximo del núcleo de una fibra unimodal es

$$r_{máx} = \frac{0.383\ \lambda}{\text{NA}} \tag{11-18}$$

en la que $r_{máx}$ = radio máximo del núcleo (metros)
NA = abertura numérica (adimensional)
λ = longitud de onda del rayo de luz (metros)

PÉRDIDAS EN LOS CABLES DE FIBRA ÓPTICA

Es probable que la pérdida de potencia en un cable de fibra óptica sea la característica más importante del cable. Con frecuencia se llama *atenuación* a la pérdida de potencia, y produce una pérdida de potencia de la onda luminosa al atravesar el cable. La atenuación tiene varios efectos adversos sobre el funcionamiento, que incluyen la reducción del ancho de banda del sistema, la rapidez de transmisión de información, la eficiencia y la capacidad general del sistema.

La fórmula normal con la que se expresa la pérdida total de potencia en un cable de fibra es

$$A(\text{dB}) = 10 \log \frac{P_{sal}}{P_{ent}} \tag{11-19}$$

en la que A(dB) = reducción total de potencia (atenuación)
P_{sal} = potencia de salida del cable (watts)
P_{ent} = potencia de entrada al cable (watts)

En general, las fibras multimodales tienden a tener mayores pérdidas de atenuación que los cables unimodales, debido principalmente a la mayor dispersión de la onda luminosa, producida por las impurezas. La tabla 11-2 muestra la potencia de salida como porcentaje de la potencia de entrada para un cable de fibra óptica a distintos valores de pérdida en decibelios. Un cable con pérdida de 3 dB reduce la potencia de salida a 50% de la potencia de entrada.

Aunque la pérdida total de potencia es de principal importancia, la atenuación de un cable óptico se expresa, en general, en decibelios de pérdida por unidad de longitud. La atenuación se representa con un valor positivo de dB, porque por definición es una pérdida. La tabla 11-3 es una lista de atenuaciones, en dB/km, para diversos tipos de cables de fibra.

La potencia óptica, en watts, medida a determinada distancia a una fuente de potencia, se puede calcular con la siguiente ecuación

$$P = P_t \times 10^{-Al/10} \tag{11-20a}$$

TABLA 11-2 % de potencia de salida en función de la pérdida en dB

Pérdida (dB)	Potencia de salida (%)
1	79
3	50
6	25
9	12.5
10	10
13	5
20	1
30	0.1
40	0.01
50	0.001

TABLA 11-3 Atenuación en el cable de fibra óptica

Tipo de cable	Diámetro del núcleo (μm)	Diámetro del revestimiento (μm)	NA (adimensional)	Atenuación (dB/km)
Unimodal	8	125	—	0.5 @ 1300 nm
	5	125	—	0.4 @ 1300 nm
Índice graduado	50	125	0.2	4 @ 850 nm
	100	140	0.3	5 @ 850 nm
Índice escalonado	200	380	0.27	6 @ 850 nm
	300	440	0.27	6 @ 850 nm
PCS	200	350	0.3	10 @ 790 nm
	400	550	0.3	10 @ 790 nm
Plástico	—	750	0.5	400 @ 650 nm
	—	1000	0.5	400 @ 650 nm

en la que P = valor medido de la potencia (watts)
P_t = valor de la potencia de transmisión (watts)
A = pérdidas de potencia en el cable (dB/km)
l = longitud del cable (km)

De igual manera, la potencia óptica, en decibelios, es

$$P(\text{dBm}) = P_{\text{ent}}(\text{dBm}) - Al \tag{11-20b}$$

en la que P = valor medido de la potencia (dBm)
P_{ent} = potencia de transmisión (dBm)
A = pérdidas de potencia en el cable (dB/km)

Ejemplo 11-4

Para un cable óptico unimodal con 0.25 dB/km, calcular la potencia óptica a 100 km de una fuente de 0.1 mW.

Solución Al sustituir en la ecuación 11-20a se obtiene

$$\begin{aligned} P &= 0.1\ \text{mW} \times 10^{-[(0.25)(100)/10]} \\ &= 1 \times 10^{-4} \times 10^{-[(0.25)(100)/10]} \\ &= [1 \times 10^{-4}]\,[1 \times 10^{-2.5}] \\ &= 0.316\ \mu\text{W} \end{aligned}$$

y

$$\begin{aligned} P(\text{dBm}) &= 10 \log \frac{0.316\mu\text{W}}{0.001} \\ &= -35\ \text{dBm} \end{aligned}$$

o bien, sustituyendo en la ecuación 11-20b

$$P(\text{dBm}) = 10 \log \frac{0.1 \text{ mW}}{0.001} - [(100 \text{ km})(0.25 \text{ dB/km})]$$
$$= -10 \text{ dBm} - 25 \text{ dB}$$
$$= -35 \text{ dBm}$$

Las pérdidas de transmisión en los cables de fibra óptica son una de las características más importantes de la fibra. Las pérdidas en la fibra causan una reducción de la potencia luminosa y, en consecuencia, reducen el ancho de banda del sistema, la rapidez de transmisión de información, la eficiencia y la capacidad general del sistema. Las principales pérdidas en la fibra son:

1. Pérdidas por absorción.
2. Pérdidas por dispersión en material o de Rayleigh.
3. Dispersión cromática, o de longitudes de onda.
4. Pérdidas por radiación.
5. Dispersión modal.
6. Pérdidas por acoplamiento.

Pérdidas por absorción

La *pérdida por absorción* en las fibras ópticas es análoga a la disipación de potencia en los cables de cobre; las impurezas en la fibra absorben la luz y la convierten en calor. El vidrio ultrapuro que se usa para fabricar las fibras ópticas tiene, aproximadamente, 99.9999% de pureza. Sin embargo, son normales las pérdidas por absorción de 1 a 1000 dB/km. En esencia, hay tres factores que contribuyen a las pérdidas por absorción en las fibras ópticas: absorción de ultravioleta, absorción de infrarrojo y absorción de resonancia iónica.

Absorción ultravioleta. La absorción ultravioleta se produce por los electrones de valencia en el material silíceo con el que se fabrican las fibras. La luz *ioniza* los electrones de valencia y los pasa a la capa de conducción. La ionización equivale a una pérdida en el campo luminoso total y en consecuencia contribuye a las pérdidas de transmisión en la fibra.

Absorción infrarroja. Es un resultado de los *fotones* de luz que son absorbidos por los átomos de las moléculas del núcleo de vidrio. Los fotones absorbidos se convierten en vibraciones mecánicas aleatorias, características del calentamiento.

Absorción por resonancia de iones. Esta absorción se debe a los iones OH^- en el material. La fuente de los iones OH^- son moléculas de agua atrapadas en el vidrio durante el proceso de fabricación. La absorción de iones también se debe a la presencia de moléculas de hierro, cobre y cromo.

La fig. 11-18 muestra las pérdidas características en los cables de fibra óptica, debido a las absorciones ultravioleta, infrarroja y de resonancia de iones.

Pérdidas en material, o por dispersión de Rayleigh

Durante el proceso de manufactura, el hilo se trefila formando fibras largas de un diámetro muy pequeño. Durante ese proceso, el vidrio se encuentra en un estado plástico: ni sólido ni líquido. La tensión aplicada al vidrio durante ese proceso hace que en el enfriamiento se desarrollen irregularidades submicroscópicas, que se incorporan a la fibra en forma permanente. Cuando los rayos de luz que se están propagando por una fibra chocan con una de esas impurezas, se *difractan*. La difracción hace que la luz se disperse o se abra en muchas direcciones. Algo de la luz difractada continúa recorriendo la fibra, y algo escapa a través del revestimiento. Los rayos luminosos que salen representan una pérdida de potencia. A todo esto se le llama *pérdida por dispersión de Rayleigh*. La fig. 11-19 muestra, en forma gráfica, la relación entre la longitud de onda y la pérdida por dispersión de Rayleigh.

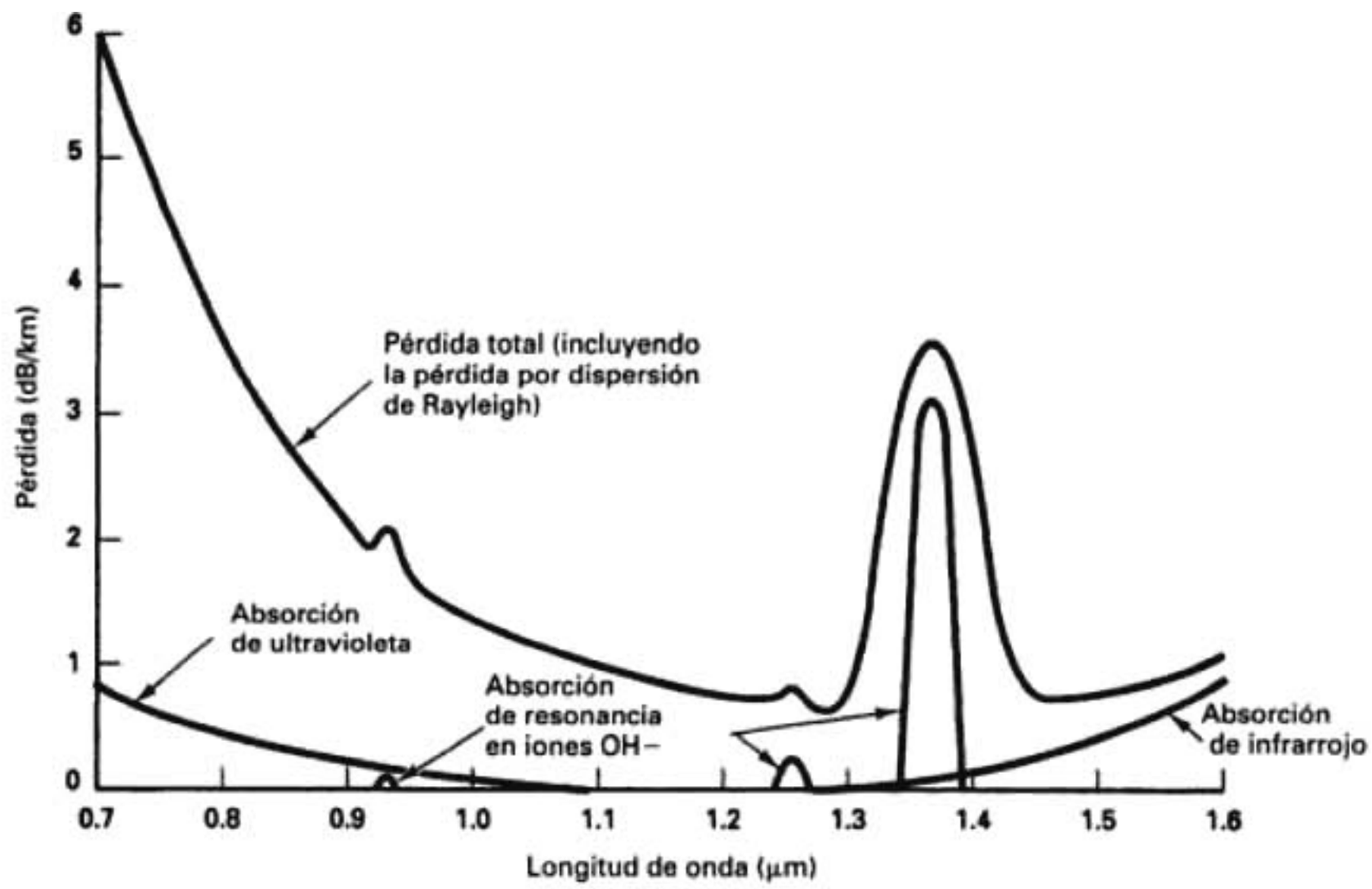


FIGURA 11-18 Pérdidas por absorción en las fibras ópticas

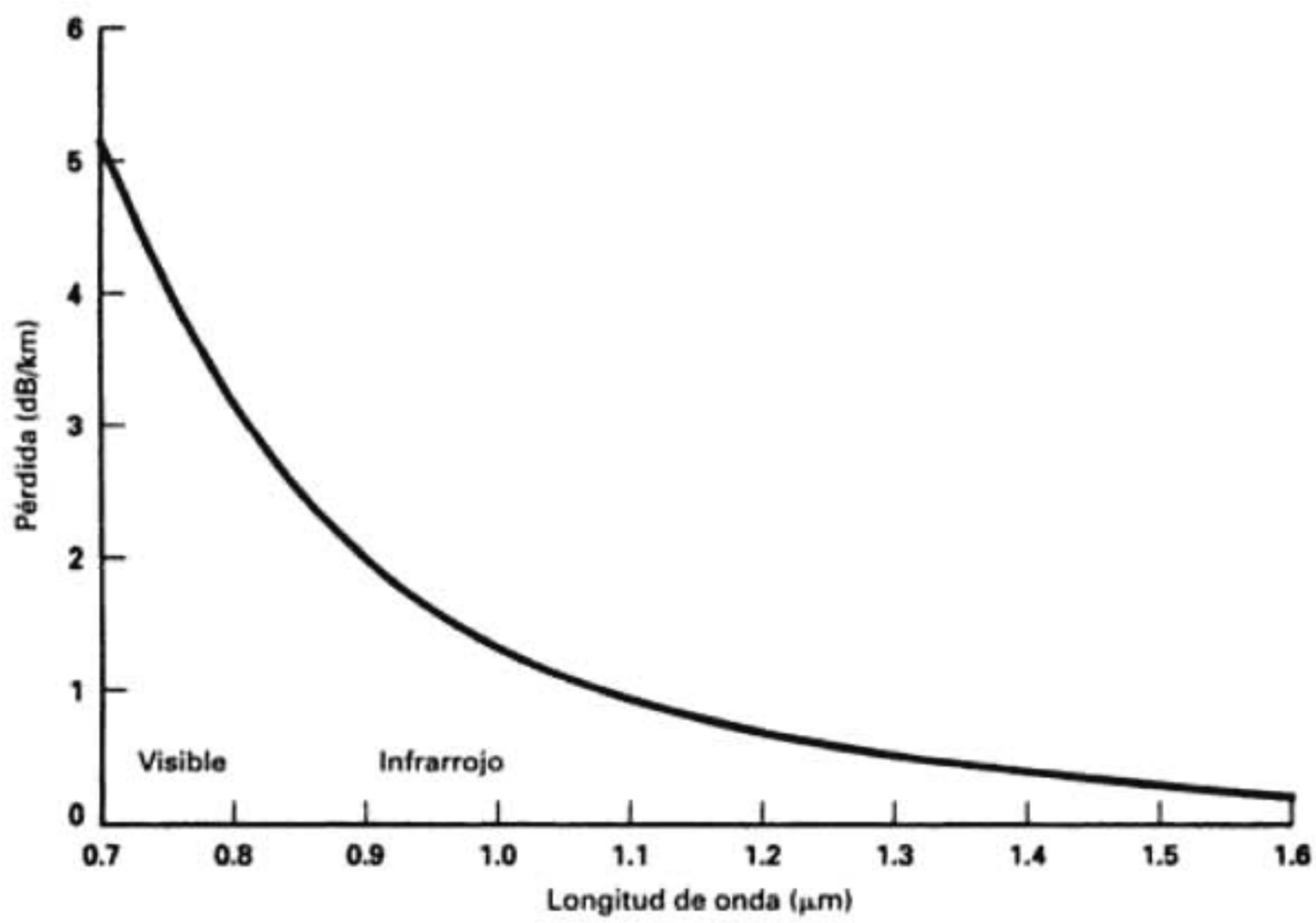


FIGURA 11-19 Pérdida por dispersión de Rayleigh en función de la longitud de onda

Dispersión cromática o de longitudes de onda

Como se dijo antes, el índice de refracción de un material depende de la longitud de onda. Los *diodos emisores de luz* (LED) emite luz que contiene una combinación de longitudes de onda. Cada una de ellas, en la señal de luz compuesta, viaja a distinta velocidad. En consecuencia, los rayos de luz que emite al mismo tiempo un LED y se propagan por una fibra óptica no llegan al extremo opuesto al mismo tiempo. Esto da como resultado una señal recibida distorsionada; la distorsión se llama *distorsión cromática*. Se puede eliminar usando una fuente monocromática, como por ejemplo un *diodo láser de inyección* (ILD). La dispersión cromática o de longitudes de onda sólo se presenta en fibras con transmisión unimodal.

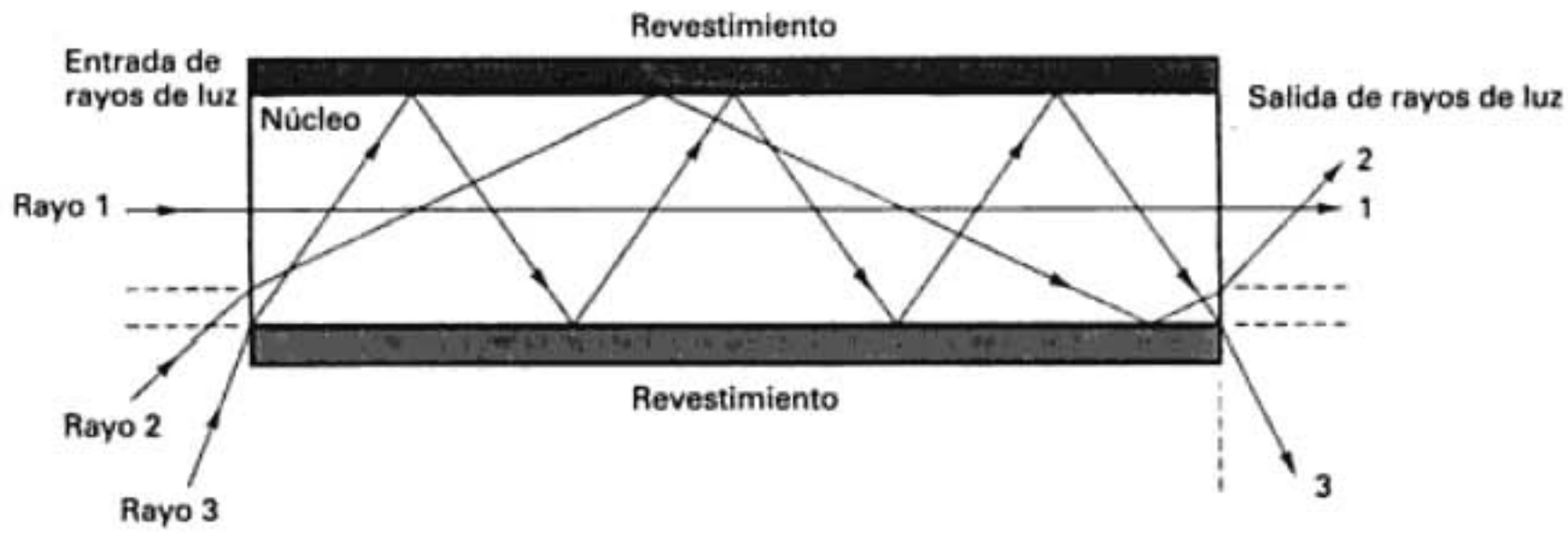


FIGURA 11-20 Propagación de la luz por una fibra multimodal de índice escalonado

Pérdidas por radiación

Las pérdidas por radiación se deben principalmente a pequeños cambios de dirección y dobleces de la fibra. En esencia hay dos tipos de dobleces: los microdobleces y los dobleces con radio constante. El *microdoblamiento* se debe a diferencias en las velocidades de contracción térmica del núcleo y del material de revestimiento. Un microdoblez es un doblez o imperfección geométrica en miniatura del eje de la fibra, que representa una discontinuidad en ella en donde puede presentarse la dispersión de Rayleigh. Las pérdidas en microdobleces en general constituyen menos de 20% de la atenuación en una fibra. Los *dobleces de radio constante* se deben a demasiada presión y tensión y, en general, se presentan cuando se doblan las fibras durante su manejo o instalación.

Dispersión modal

La *dispersión modal*, o *ensanchamiento del pulso* se debe a la diferencia en los tiempos de propagación de rayos de luz que van por diferentes trayectorias en una fibra. Es obvio que la dispersión modal sólo puede presentarse en las fibras multimodales. Se puede reducir en forma considerable usando fibras de índice graduado, y se puede eliminar casi con el uso de fibras unimodales de índice escalonado.

La dispersión modal puede causar que un pulso de energía luminosa se reparta al propagarse por una fibra. Si el ensanchamiento del pulso tiene la gravedad suficiente, el pulso puede quedar encimado al siguiente (un ejemplo de interferencia entre símbolos). En una fibra multimodal de índice escalonado, un rayo de luz que se propaga en forma directa por el eje de la fibra tarda el menor tiempo en recorrer la fibra. Un rayo de luz que llega a la interfaz núcleo-revestimiento, en el ángulo crítico, sufrirá la cantidad máxima de reflexiones internas y, por consiguiente, tardará el mayor tiempo en recorrer la fibra.

Para la propagación multimodal, la dispersión se expresa con frecuencia como un *producto de abertura de banda por longitud* (BLP, de *bandwidth length product*) o como *producto de abertura de banda por distancia* (BDP, de *bandwidth distance product*). BLP indica qué frecuencias de señal se pueden propagar a una determinada distancia de cable de fibra, y se expresa matemáticamente como el producto de la abertura de banda (que a veces, con frecuencias ópticas, se llama *ancho de raya*) por la distancia. Los productos de amplitud de banda por longitud se expresan con frecuencia en unidades de MHz-km.

La fig. 11-20 muestra tres rayos de luz que se propagan por una fibra multimodal de índice escalonado. El modo de orden mínimo (rayo 1) recorre una trayectoria paralela al eje de la fibra. El modo de orden intermedio (rayo 2) rebota varias veces en la interfaz para recorrer la longitud de la fibra. El modo de orden máximo (rayo 3) hace muchos cruces de la fibra al propagarse por toda la distancia. Se puede ver que el rayo 3 recorre una distancia bastante mayor que el rayo 1 al propagarse por la fibra. En consecuencia, si se emitieran tres rayos de luz hacia la fibra, al mismo tiempo, y representaran un pulso de energía luminosa, los tres llegarían al otro extremo de la fibra en tiempos distintos, y causarían una repartición de energía luminosa a través del tiempo. A esto se le llama dispersión modal, y causa que un pulso se estire y también se reduzca su amplitud en la salida de la fibra. Los tres rayos de luz se propagan a través del mismo material y a la misma velocidad, pero el rayo 3 debe recorrer mayor distancia y, en consecuencia, tarda más tiempo en propagarse por la fibra.

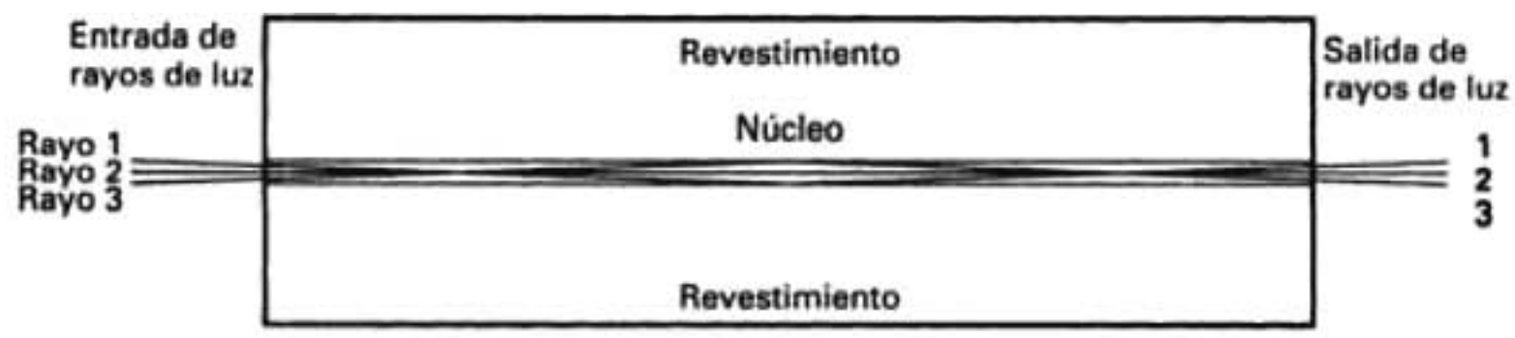


FIGURA 11-21 Propagación de la luz por una fibra unimodal de índice escalonado

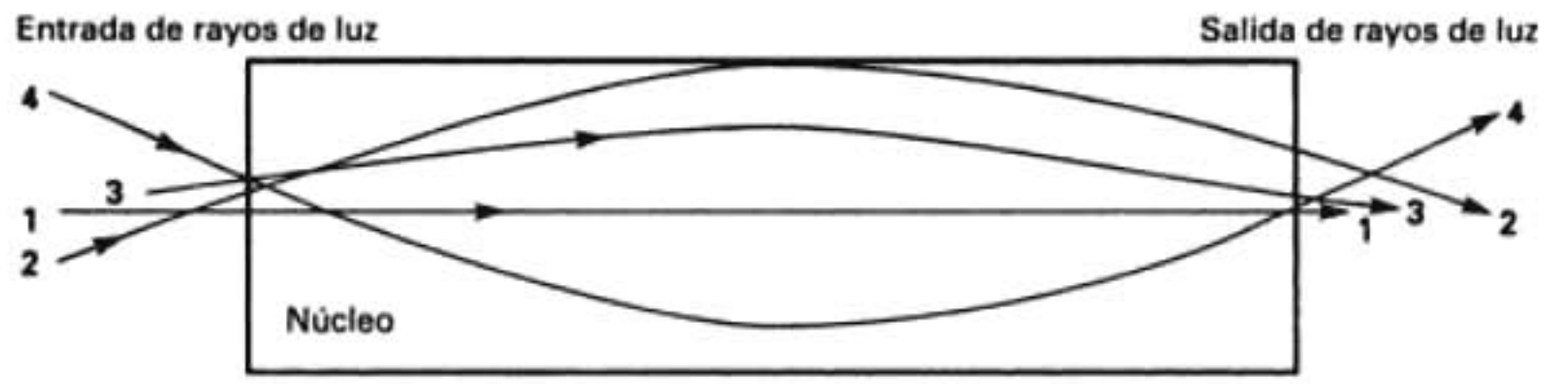


FIGURA 11-22 Dispersión del ancho del pulso en un cable de fibra óptica

La fig. 11-21 muestra los rayos de luz propagándose por una fibra unimodal de índice escalonado. Como la dimensión radial de la fibra es suficientemente pequeña, sólo hay una trayectoria que puede seguir cada uno de los rayos al recorrer la longitud de la fibra. En consecuencia, cada rayo de luz viaja la misma distancia en determinado tiempo, y los rayos tienen exactamente la misma relación en el otro extremo que la que tenían cuando entraron al cable. El resultado es que *no hay dispersión modal* o *ensanchamiento del pulso*.

La fig. 11-22 representa la luz propagándose por una fibra multimodal de índice graduado. Se muestran tres rayos que viajan en tres modos distintos. Cada uno viaja por una trayectoria diferente, pero todos tardan aproximadamente el mismo tiempo en recorrer la longitud de la fibra. Esto se debe a que el índice de refracción de la fibra disminuye en función de la distancia al centro, y la velocidad a la que viaja un rayo es inversamente proporcional al índice de refracción. Así, cuanto más se apartan los rayos 2 y 3 del centro de la fibra, se propagan con más rapidez.

La fig. 11-23 muestra la relación entre tiempo y energía de un pulso de luz al propagarse por un cable de fibra. Se aprecia que al propagarse el pulso por la fibra, los rayos de luz que forman el pulso se reparten en el tiempo, y eso causa una reducción correspondiente de la amplitud del pulso y un estiramiento de su ancho. A esto se le llama *ensanchamiento de pulso* o *dispersión de ancho de pulso*, y causa errores en la transmisión digital. También se puede ver que cuando la energía luminosa de un pulso se retrasa en el tiempo, interfiere con el siguiente pulso y causa interferencia entre símbolos ("*intersímbolos*").

La fig. 11-24a es de una transmisión digital unipolar de retorno a cero (UPRZ, de *unipolar return-to-zero*). En esta transmisión, suponiendo que el pulso es muy angosto, si la energía luminosa del pulso A se retrasara (*ensanchara*) un bit de tiempo (t_b), interferiría con el pulso B, y cambiaría lo que era un 0 lógico a un 1 lógico. La fig. 11-24b muestra una transmisión digital unipolar de no retorno a cero (UPNRZ, de *unipolar nonreturn-to-zero*), en donde cada pulso es igual al bit de tiempo. Con la transmisión UPNRZ, si la energía del pulso A se retrasara la mitad de un bit de tiempo, interferiría con el pulso B. En consecuencia, las transmisiones UPRZ pueden tolerar el doble de la demora, o ensanchamiento, que las transmisiones UPNRZ.

La diferencia entre los tiempos absolutos de demora de los rayos de luz más rápidos y más lentos que se propagan por una fibra de longitud unitaria se llama *constante de ensanchamiento del pulso* (Δt) y, en general, se expresa en nanosegundos por kilómetro (ns/km). El ensanchamiento total de pulso, ΔT, es igual, entonces, a la constante de ensanchamiento de pulso, Δt, por la longitud total de la fibra, L, esto es,

$$\Delta T_{(\text{ns})} = \Delta t_{(\text{ns/km})} \times L_{(\text{km})} \tag{11-21}$$

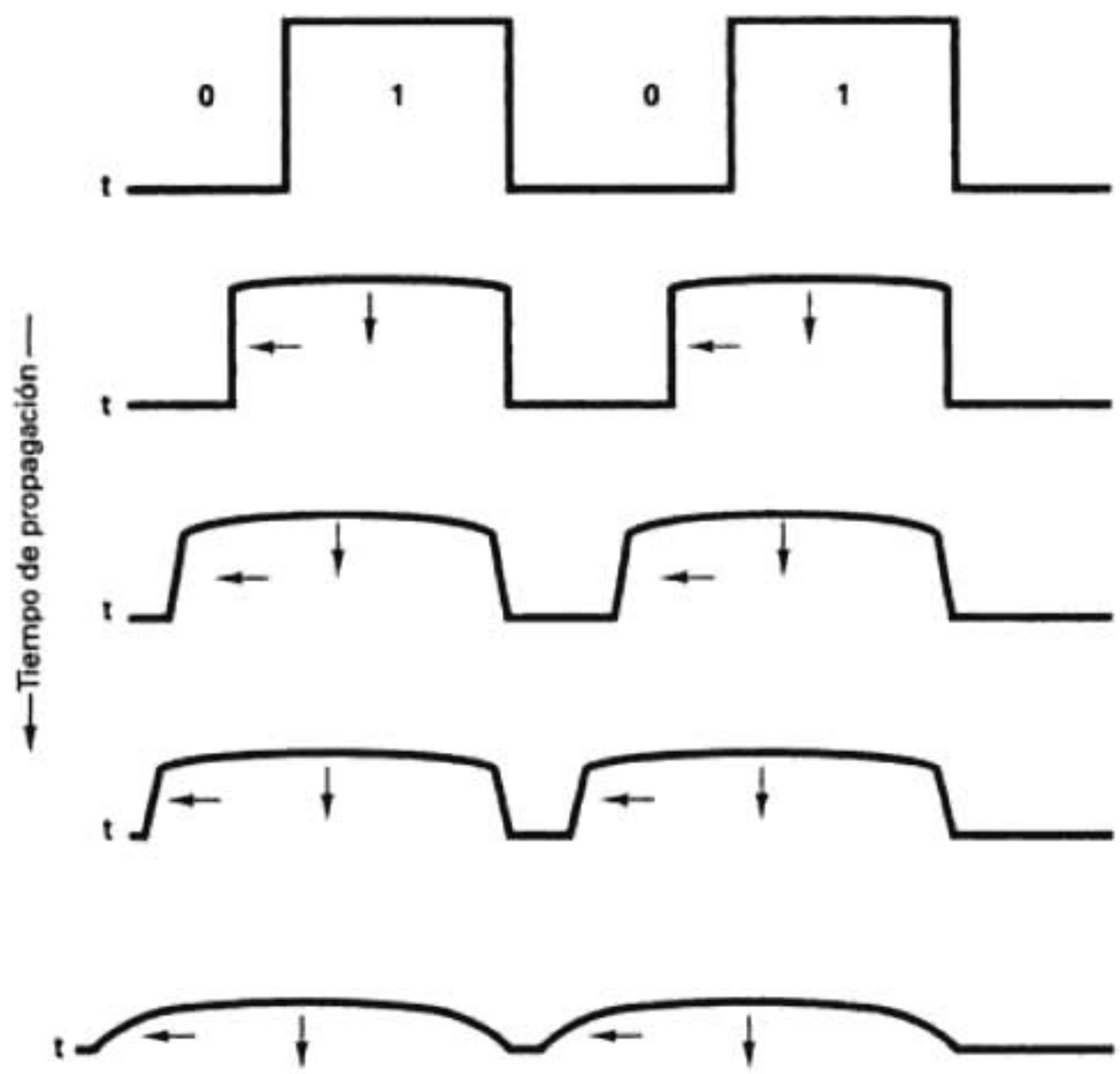


FIGURA 11-23 Dispersión del ancho del pulso en un cable de fibra óptica

Para las transmisiones UPRZ, la rapidez máxima de transmisión de datos, en bits por segundo (bps), se expresa como sigue

$$f_{b(\text{bps})} = \frac{1}{\Delta t \times L} \tag{11-22}$$

y para las transmisiones UPNRZ, la rapidez máxima de transmisión es

$$f_{b(\text{bps})} = \frac{1}{2\,\Delta t \times L} \tag{11-23}$$

Ejemplo 11-5

Determinar la rapidez máxima de transmisión digital, con una fibra óptica de 10 km de longitud y constante de ensanchamiento de pulso de 5 ns/km, para:

(a) Transmisión con regreso a cero.

(b) Transmisión sin regreso a cero.

Solución (a) Al sustituir en la ecuación 11-22 se obtiene

$$f_b = \frac{1}{5\text{ ns/km} \times 10\text{ km}} = 20\text{ Mbps}$$

(b) De acuerdo con la ecuación 11-23,

$$f_b = \frac{1}{(2 \times 5\text{ ns/km}) \times 10\text{ km}} = 10\text{ Mbps}$$

Estos resultados indican que la rapidez posible de transmisión digital para esta fibra óptica es el doble (20 Mbps contra 10 Mbps) para la transmisión UPRZ en comparación con la UPNRZ.

Pérdidas en acoplamiento

En los cables de fibra pueden presentarse pérdidas por acoplamiento en cualquiera de los tres tipos siguientes de uniones ópticas: conexiones de fuente luminosa a fibra, conexiones de fibra

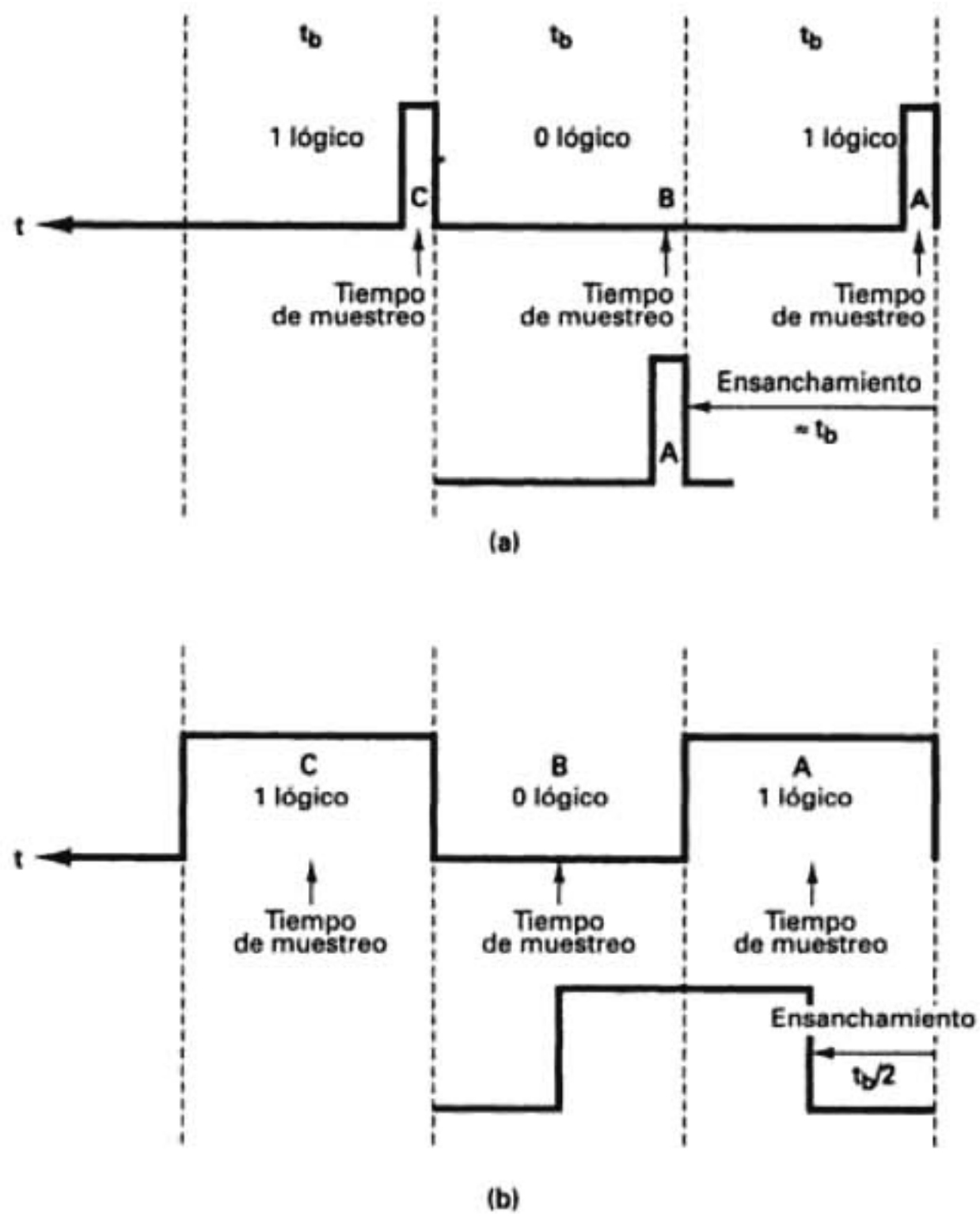


FIGURA 11-24 Ensanchamiento del pulso en transmisiones digitales: (a) UPRZ; (b) UPNRZ

a fibra y conexiones de fibra a fotodetector. Las pérdidas en las uniones se deben, con más frecuencia, a uno de los siguientes problemas de alineación: desalineamiento lateral, desalineamiento de entrehierro, desalineamiento angular y acabados superficiales imperfectos. Estas imperfecciones se ven en la fig. 11-25.

Desalineamiento lateral. Éste se ve en la fig. 11-25a, y es el desplazamiento lateral o axial entre dos tramos de cables de fibra adjuntos. La cantidad de pérdida puede ser desde un par hasta varias décimas de decibelio, o hasta varios decibelios. Esta pérdida suele ser despreciable si los ejes de las fibras se alinean a menos de 5% del diámetro de la fibra menor.

Desalineamiento de entrehierro. Se ve en la fig. 11-25b, y a veces se llama *separación entre extremos*. Cuando se hacen *empalmes* en fibras ópticas, las fibras se deben tocar realmente. Mientras más alejadas queden, la pérdida de luz será mayor. Si dos fibras se unen con un conector, los extremos no se deben tocar, porque al frotarse entre sí en el conector se podrían dañar una o las dos fibras.

Desalineamiento angular. Éste se ve en la fig. 11-25c, y a veces se llama *desplazamiento angular*. Si el desplazamiento angular es menor que 2°, la pérdida será menor que 0.5 dB.

Acabado superficial imperfecto. Se ve en la fig. 11-25d. Los extremos de las dos fibras adjuntas deben estar muy pulidos y asentar entre sí. Si los extremos de las fibras están menos de 3° fuera de la perpendicular, las pérdidas serán menores que 0.5 dB.

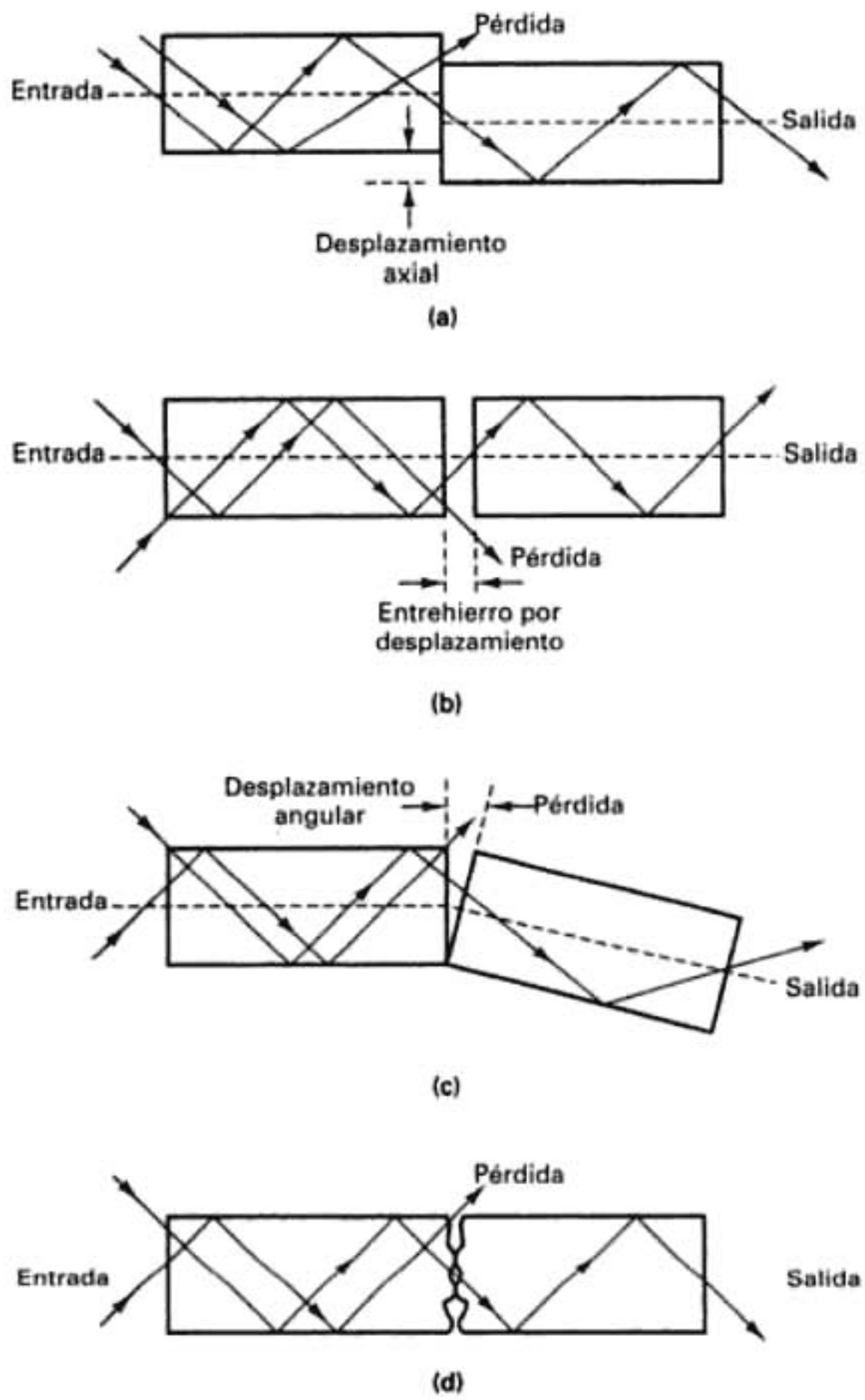


FIGURA 11-25 Defectos en el alineamiento de las fibras: (a) desalineamiento lateral; (b) desalineamiento de entrehierro; (c) desalineamiento angular; (d) acabado superficial

FUENTES LUMINOSAS

El intervalo de frecuencias luminosas que detecta el ojo humano ocupa un segmento muy angosto del espectro total de frecuencias electromagnéticas. Por ejemplo, la luz azul ocupa las longitudes de onda más cortas de la luz visible, y los tonos rojos ocupan las longitudes de onda más largas. La fig. 11-26 muestra la distribución de longitudes de onda de la luz producida por un foco de tungsteno, y el intervalo de longitudes de onda perceptibles al ojo humano. Como se ve allí, el ojo humano puede detectar sólo aquellas longitudes de onda entre 380 nm y 780 nm. Además, la luz consiste en muchos tonos de los colores, que se relacionan en forma directa con la energía calorífica que se irradia. También se ve en la figura que se produce más luz visible a medida que aumenta la temperatura del foco.

Las fuentes luminosas que se usan en los sistemas de fibra óptica deben tener longitudes de onda que se propaguen con eficiencia en la fibra óptica. Además se debe considerar el intervalo de longitudes de onda, porque mientras más amplio sea el intervalo, será más probable que se presente la dispersión cromática. También, las fuentes luminosas deben producir la potencia

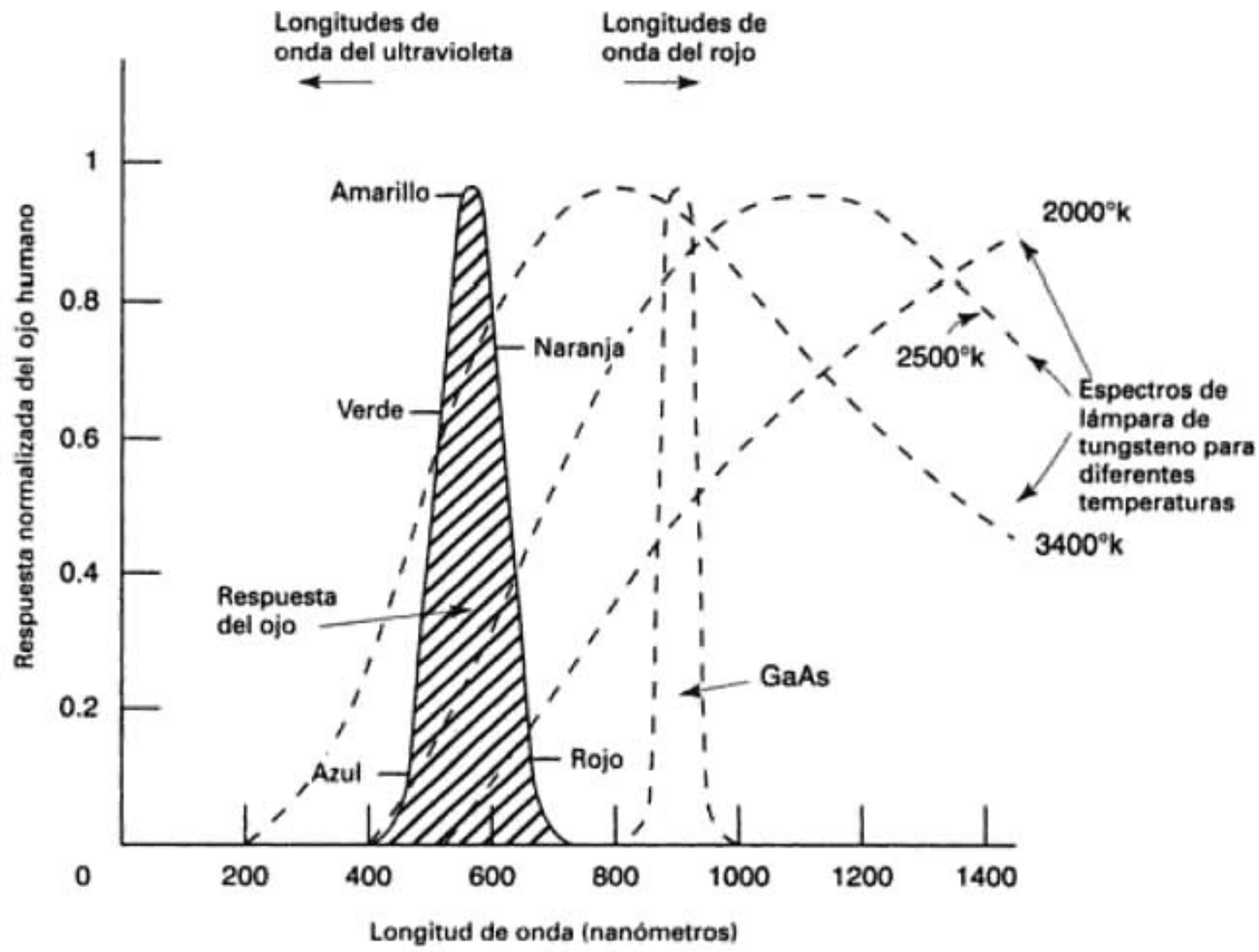


FIGURA 11-26 Radiación de una lámpara de tungsteno y respuesta del ojo humano

suficiente para permitir que la luz se propague por la fibra sin causar distorsiones no lineales en ella ni en el receptor. Por último, la fuente luminosa se debe fabricar de tal manera que su salida se pueda acoplar con la fibra en forma eficiente.

Potencia óptica

La *intensidad luminosa* es un concepto bastante complejo, que se puede expresar en términos tanto *fotométricos* como *radiométricos*. La fotometría es la ciencia de medir sólo las ondas luminosas visibles al ojo humano. Por otro lado, la radiometría mide la luz en todo el espectro electromagnético. En términos fotométricos, la intensidad luminosa se describe con más frecuencia en términos de *densidad de flujo* luminoso, y se mide en lúmenes por unidad de área. Sin embargo, con frecuencia son más útiles los términos radiométricos para el ingeniero y el técnico. En términos radiométricos, la *potencia óptica* mide la tasa con la que las ondas electromagnéticas transfieren energía luminosa. En términos sencillos, la potencia óptica se describe como el flujo de energía luminosa que atraviesa determinado punto en un tiempo especificado. La potencia óptica se define con la siguiente ecuación

$$Potencia = \frac{d(\text{energía})}{d(\text{tiempo})}$$

o sea

$$P = \frac{dQ}{dt} \text{ (watts)} \tag{11-24}$$

en donde P = potencia óptica (watts)
dQ = carga instantánea (joules)
dt = cambio instantáneo de tiempo (segundos)

A veces, a la potencia óptica se le llama *flujo radiante* (ϕ), que equivale a joules por segundo, y es la misma potencia que se mide eléctrica o térmicamente en watts. Los términos radiométri-

cos se usan en general con fuentes luminosas cuyas potencias van de decenas de microwatts hasta más de 100 miliwatts. La potencia óptica se suele expresar en decibelios relativos a un valor definido de potencia, como 1 mW (dBm) o 1 μW (dBμ). Las ecuaciones correspondientes son

$$\text{dBm} = 10 \log \frac{P(\text{watts})}{0.001 \text{ watt}} \tag{11-25}$$

y

$$\text{dB}\mu = 10 \log \frac{P(\text{watts})}{0.000001 \text{ watt}} \tag{11-26}$$

Ejemplo 11-6

Calcular la potencia óptica, en dBm y dBμ para valores de potencia de:
(a) 10 mW, **(b)** 0.1 mW y **(c)** 20 μW.

Solución (a) Se sustituye en las ecuaciones 11-25 y 11-26, para obtener lo siguiente

$$\text{dBm} = 10 \log \frac{10 \text{ mW}}{1 \text{ mW}} = 10 \text{ dBm}$$

$$\text{dB}\mu = 10 \log \frac{10 \text{ mW}}{1 \ \mu\text{W}} = 40 \text{ dB}\mu$$

(b) Al sustituir en las ecuaciones 11-25 y 11-26, se obtiene

$$\text{dBm} = 10 \log \frac{0.1 \text{ mW}}{1 \text{ mW}} = -10 \text{ dBm}$$

$$\text{dB}\mu = 10 \log \frac{0.1 \text{ mW}}{1 \ \mu\text{W}} = 20 \text{ dB}\mu$$

(c) Con las mismas ecuaciones se obtiene

$$\text{dBm} = 10 \log \frac{20 \ \mu\text{W}}{1 \text{ mW}} = -17 \text{ dBm}$$

$$\text{dB}\mu = 10 \log \frac{20 \ \mu\text{W}}{1 \ \mu\text{W}} = 13 \text{ dB}\mu$$

FUENTES ÓPTICAS

En esencia sólo hay dos dispositivos que se usan con frecuencia para generar luz en sistemas de comunicaciones con fibra óptica: diodos emisores de luz (LED) y diodos de láser de inyección (ILD, de *injection laser diode*). Ambos dispositivos se fabrican con materiales semiconductores, y tienen sus ventajas y desventajas. Los LED normales tienen anchos espectrales de 30 a 50 nm, mientras que los láseres de inyección sólo tienen anchos espectrales de 1 a 3 nm (1 nm corresponde a una frecuencia aproximada de 178 GHz). Por consiguiente, una fuente luminosa de 1320 nm con *ancho de raya* espectral de 0.0056 nm tiene una amplitud de banda de frecuencias aproximada de 1 GHz. El ancho de raya es el equivalente, en longitudes de onda, del ancho de banda.

La preferencia hacia un dispositivo emisor de luz respecto a otro se determina con los requisitos económicos y de funcionamiento del sistema. El mayor costo de los diodos de láser se compensa con una mayor eficiencia, mientras que los diodos emisores de luz, normalmente, tienen menor costo y menor eficiencia.

Diodos emisores de luz

Un *diodo emisor de luz* (LED, por *light-emitting diode*) es un *diodo de unión p-n*, fabricado casi siempre con un material semiconductor como el arseniuro de aluminio y galio (AlGaAs) o el arseniuro fosfuro de galio (GaAsP). Los LED emiten luz por emisión espontánea: la luz se emite como resultado de la recombinación de electrones con huecos. Cuando tienen polarización directa, los portadores minoritarios se inyectan a través de la unión *p-n*. Una vez atravesada la unión, esos portadores minoritarios se recombinan con portadores mayoritarios y desprenden energía en forma de luz. Este proceso es esencialmente el mismo que en un diodo semiconduc-

TABLA 11-4 Longitudes de onda producidas con materiales semiconductores

Material	Longitud de onda (nm)
AlGaInP	630–680
GaInP	670
GaAlAs	620–895
GaAs	904
InGaAs	980
InGaAsP	1100–1650
InGaAsSb	1700–4400

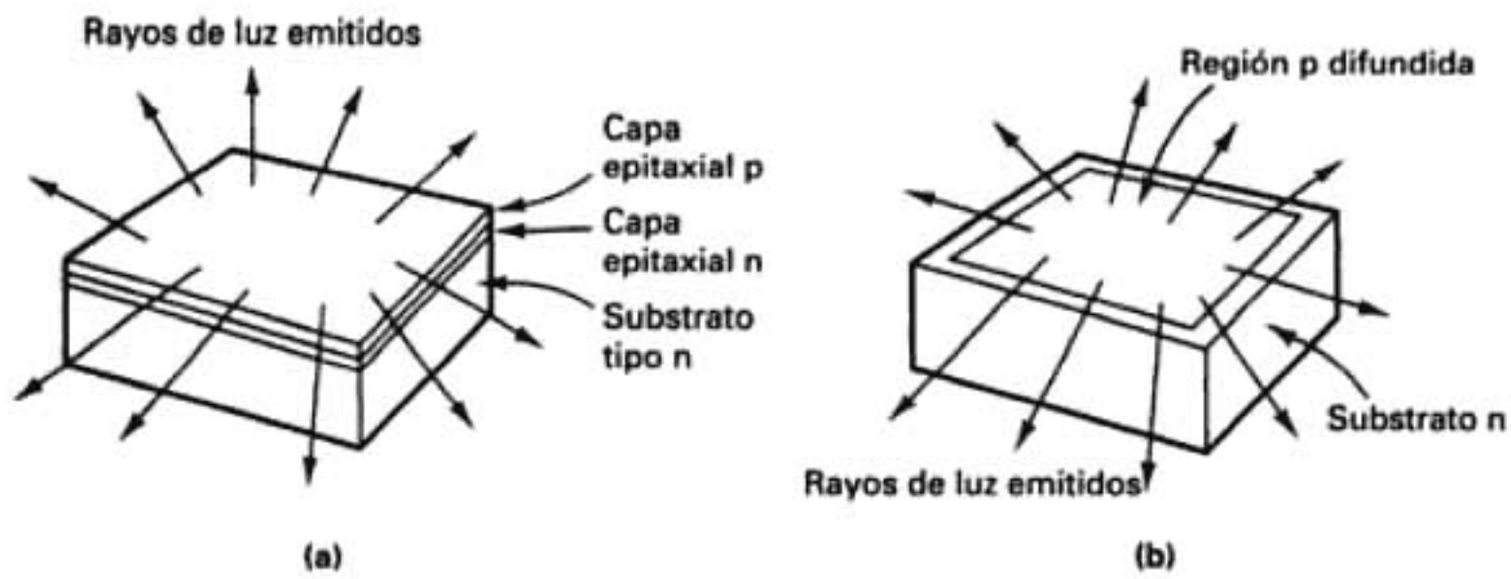


FIGURA 11-27 Estructuras de LED de homounión: (a) arseniuro de galio dopado con silicio; (b) difusión plana

tor convencional, pero en los LED se eligen ciertos materiales semiconductores y dopantes tales que el proceso es radiativo; esto es, que se produce un fotón. Éste es un cuanto de energía de onda electromagnética. Los fotones son partículas que viajan a la velocidad de la luz, pero que en reposo no tienen masa. En los diodos de semiconductores convencionales (por ejemplo, de germanio y de silicio), el proceso es principalmente no radiativo, y no se generan fotones. La banda prohibida del material que se usa para fabricar un LED determina el color de la luz que emite, y si la luz emitida es visible al ojo humano.

Para producir los LED, se forman semiconductores con materiales cuyos átomos tienen tres o cinco electrones de valencia; se conocen como átomos del grupo III o del grupo V, respectivamente, por su lugar en la tabla periódica de los elementos. Para producir longitudes de onda de luz en la región de 800 nm, los LED son de átomos del grupo III, como por ejemplo, galio (GA) y aluminio (Al), y un átomo del grupo V, como el arsénico (As). La unión que se forma se abrevia GaAlAs, que indica arseniuro de galio y aluminio. Para longitudes de onda mayores, el galio, junto con el indio (In), un átomo del grupo III, se combinan con átomos de fósforo (P) y arsénico, del grupo V, con lo que se forma un arseniuro fosfuro de galio e indio (GaInAsP). La tabla 11.4 es una lista de algunos materiales semiconductores comunes que se usan para fabricar LED, con sus longitudes de onda respectivas.

LED de homounión. Una unión *p-n* formada con dos mezclas distintas de igual tipo de átomos se llama estructura de homounión. Las estructuras más sencillas de LED son las de homounión y de crecimiento epitaxial, o dispositivos semiconductores de un solo semiconductor difundido, como los dos que se ven en la fig. 11-27. Los LED de *crecimiento epitaxial* se fabrican en general con arseniuro de galio dopado con silicio (fig. 11-27a). Una longitud de onda normal emitida con esta estructura es 940 nm, y la potencia normal de salida es aproximadamente, 2 mW (3 dBm) a 100 mA de corriente directa. Las ondas luminosas producidas en fuentes de homounión no producen una luz muy útil para una fibra óptica. La luz se emite en todas direcciones por igual y, en consecuencia, sólo una pequeña fracción del total de luz producida se

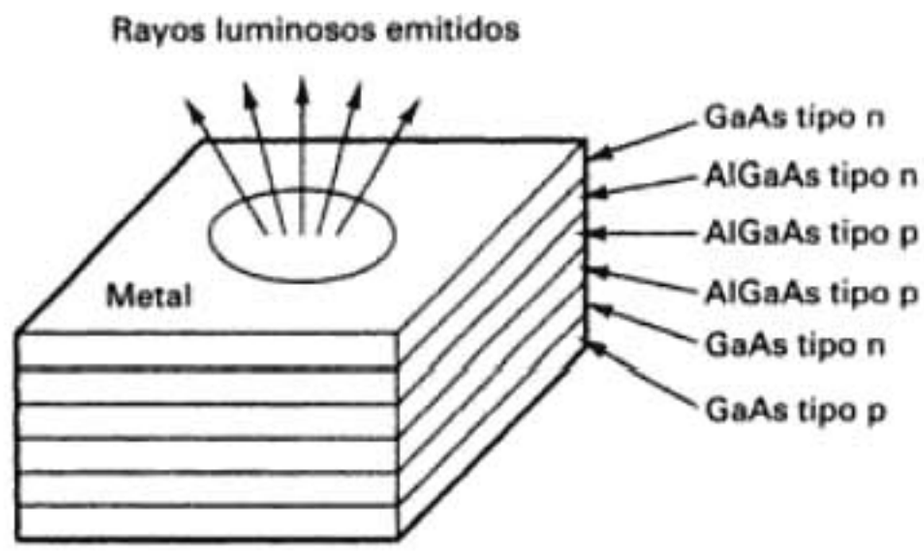


FIGURA 11-28 LED de heterounión plana

acopla en la fibra. Además, la relación de electricidad convertida en luz es muy baja. A los dispositivos de homounión se les llama con frecuencia *emisores superficiales*.

Los LED de homounión de *difusión plana* (fig. 11-27b) producen más o menos 500 μW a 900 nm de longitud de onda. La principal desventaja de los LED de homounión es la no direccionalidad de su luz emitida, lo que hace que sean malas opciones como fuente luminosa para sistemas de fibra óptica.

LED de heterounión. Los LED de heterounión se fabrican con material semiconductor tipo *p* de un conjunto de átomos, y material semiconductor tipo *n*, de otro conjunto. Los dispositivos de heterounión están estratificados (normalmente dos capas) de tal manera que se amplía el efecto de concentración. Así, se produce un dispositivo que confina a los electrones, los huecos portadores y la luz, en un área mucho menor. La unión se suele fabricar sobre un substrato de material de respaldo, para después emparedarse entre contactos metálicos con los que se conecta el dispositivo a una fuente de electricidad.

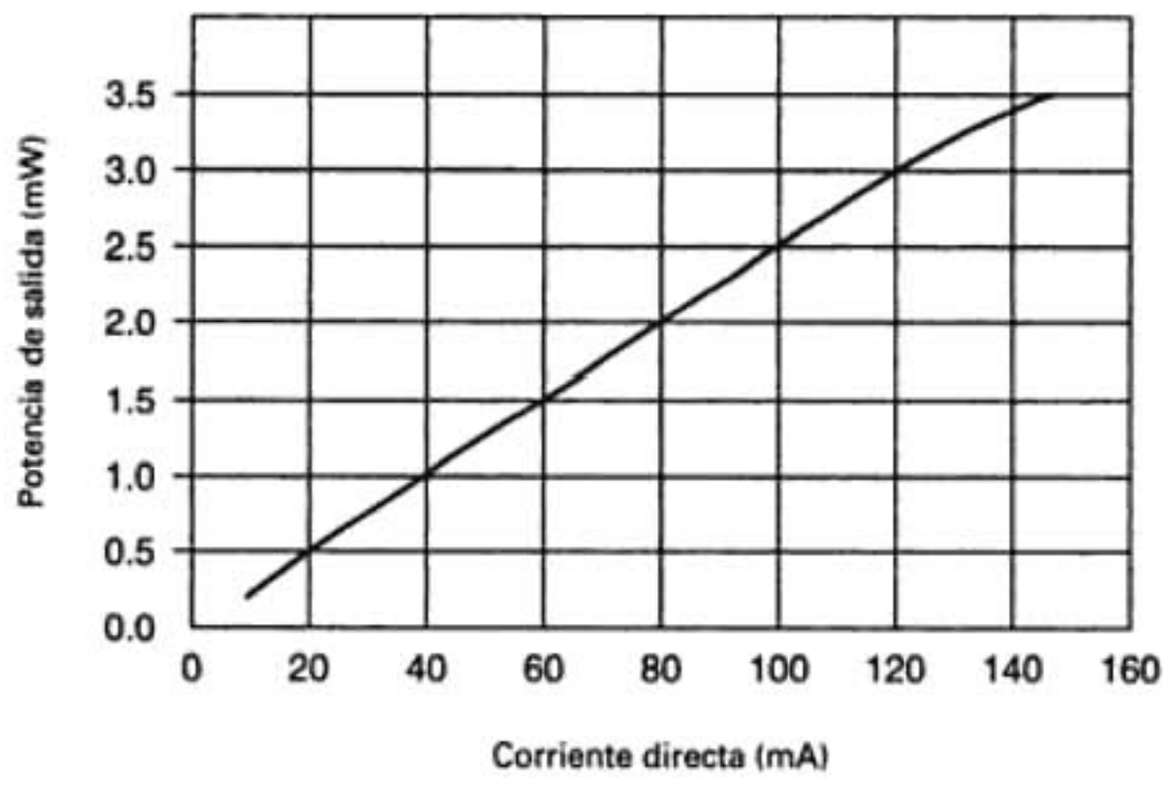
Con los dispositivos de heterounión, la luz se emite desde la orilla del material y, en consecuencia, se les llama *emisores de borde*. Un *LED de heterounión plana* (fig. 11-28) se parece mucho al LED de crecimiento epitaxial, excepto que el diseño geométrico es tal que se concentra la corriente directa en un área muy pequeña de la capa activa.

Los dispositivos de heterounión tienen las siguientes ventajas sobre los de homounión:

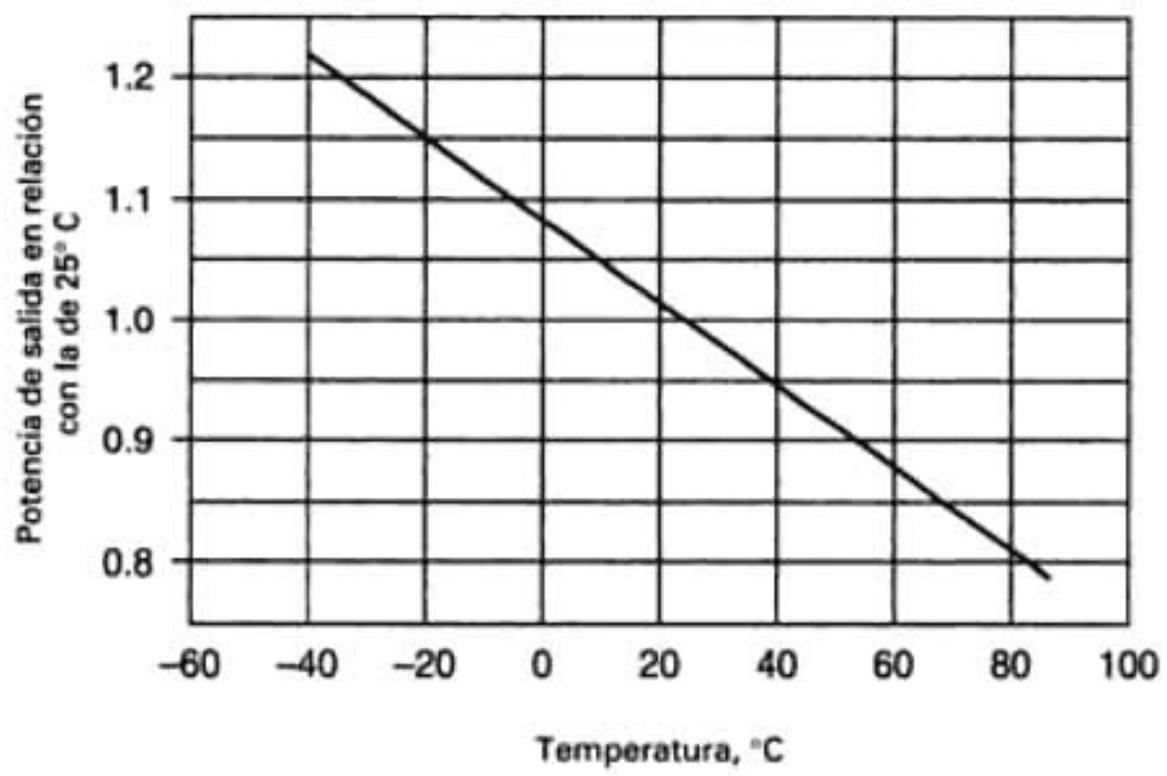
1. El aumento de densidad de corriente genera una mancha luminosa más brillante.
2. La menor área emisora facilita acoplar la luz emitida a una fibra.
3. La pequeña área efectiva tiene menor capacitancia, lo que permite usar el LED de heterounión a mayores velocidades.

La fig. 11-29 muestra las características eléctricas típicas de un diodo emisor de luz infrarroja de bajo costo. La parte (a) de esta figura representa la potencia de salida en función de la corriente directa. Se puede ver que la potencia de salida varía en forma lineal dentro de un amplio margen de corriente de entrada [0.5 mW (–3 dBm) a 20 mA hasta 3.4 mW (5.3 dBm) a 140 mA]. La fig. 11-29b muestra la potencia de salida en función de la temperatura. Se aprecia que la potencia de salida tiene una variación inversa respecto a la temperatura, dentro del intervalo de –40° C a 80° C. La fig. 11-29c muestra la potencia relativa de salida en función de la longitud de la onda de salida. Para este ejemplo particular, la potencia máxima de salida se alcanza a una longitud de 825 nm a la salida.

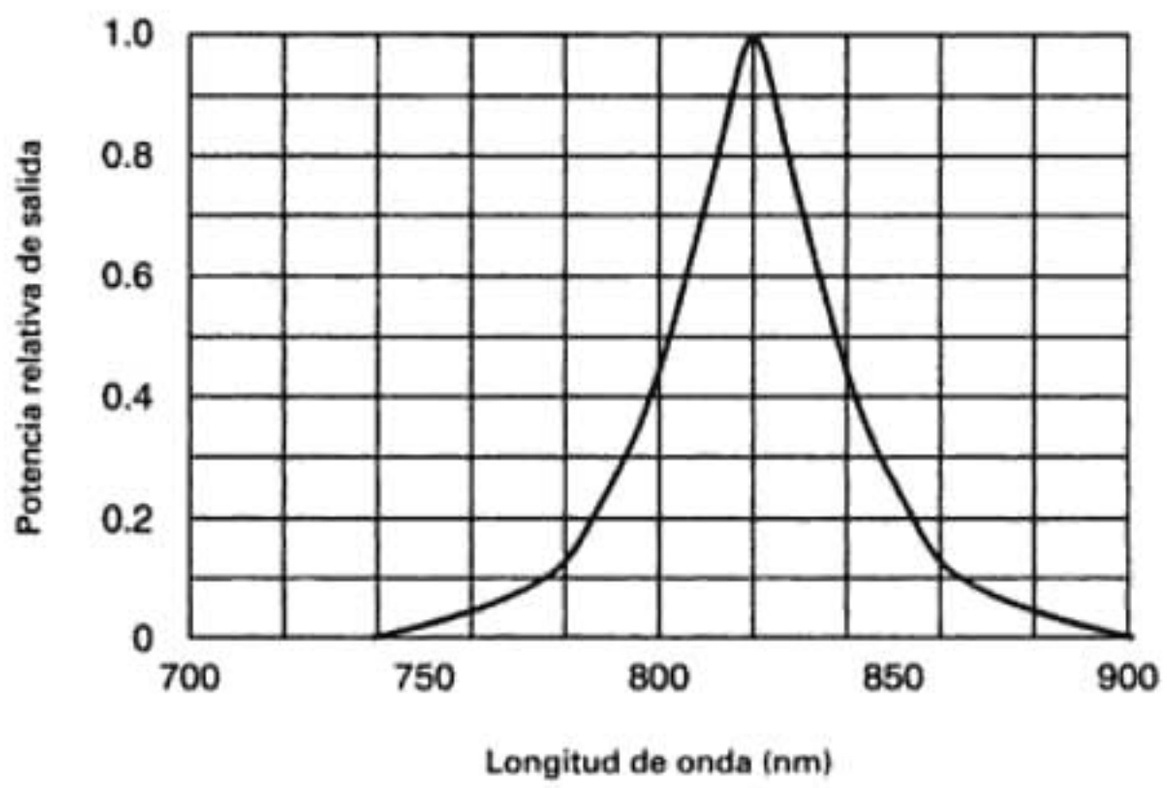
LED de superficie emisora y pozo grabado de Burrus. Para las aplicaciones más prácticas, como telecomunicaciones, se requieren velocidades de datos mayores que 100 Mbps. Para estas aplicaciones se desarrolló el LED de pozo grabado, por parte de Burrus y Dawson, de los Bell Laboratories. Es un LED de superficie emisora, y se ve en la fig. 11-30. Emite luz en muchas direcciones, y ayuda a concentrar la luz emitida en un área muy pequeña. También se puede poner lentes en domo, sobre la superficie emisora, para dirigir la luz hacia un área menor. Estos dispositivos son más eficientes que los emisores normales de superficie, y permiten acoplar más potencia a la fibra óptica, pero su fabricación es más difícil y costosa.



(a)



(b)



(c)

FIGURA 11-29 Características eléctricas típicas de los LED: (a) potencia de salida en función de la corriente directa; (b) potencia de salida en función de la temperatura; (c) potencia de salida en función de la longitud de onda

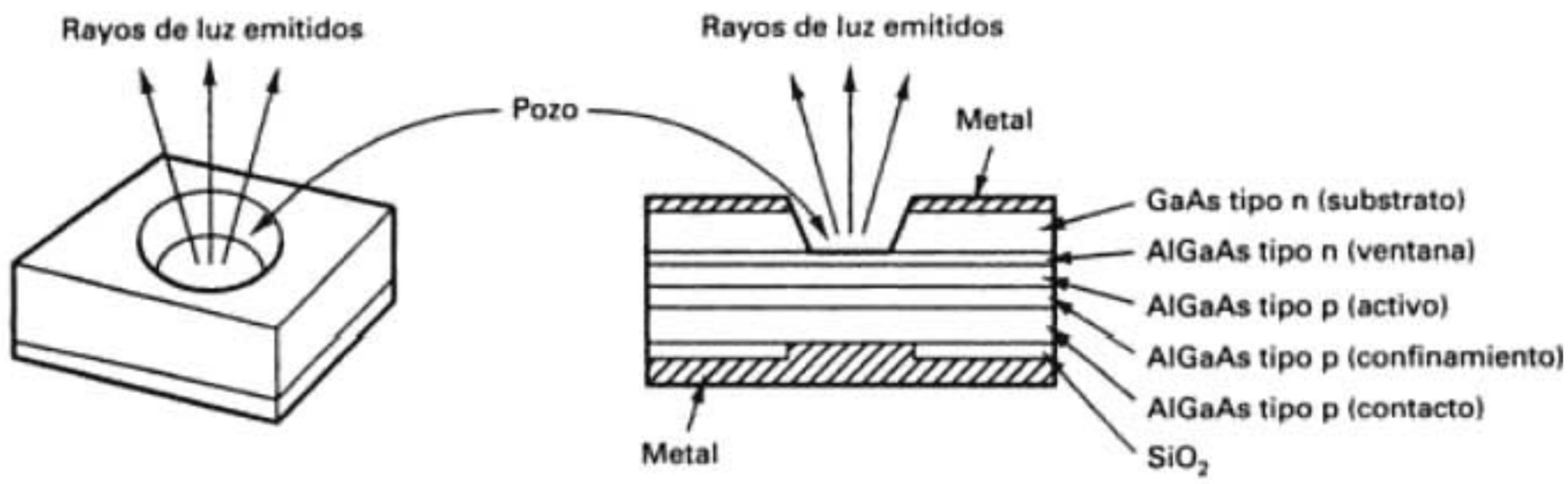


FIGURA 11-30 LED de Burrus, de superficie emisora y pozo grabado

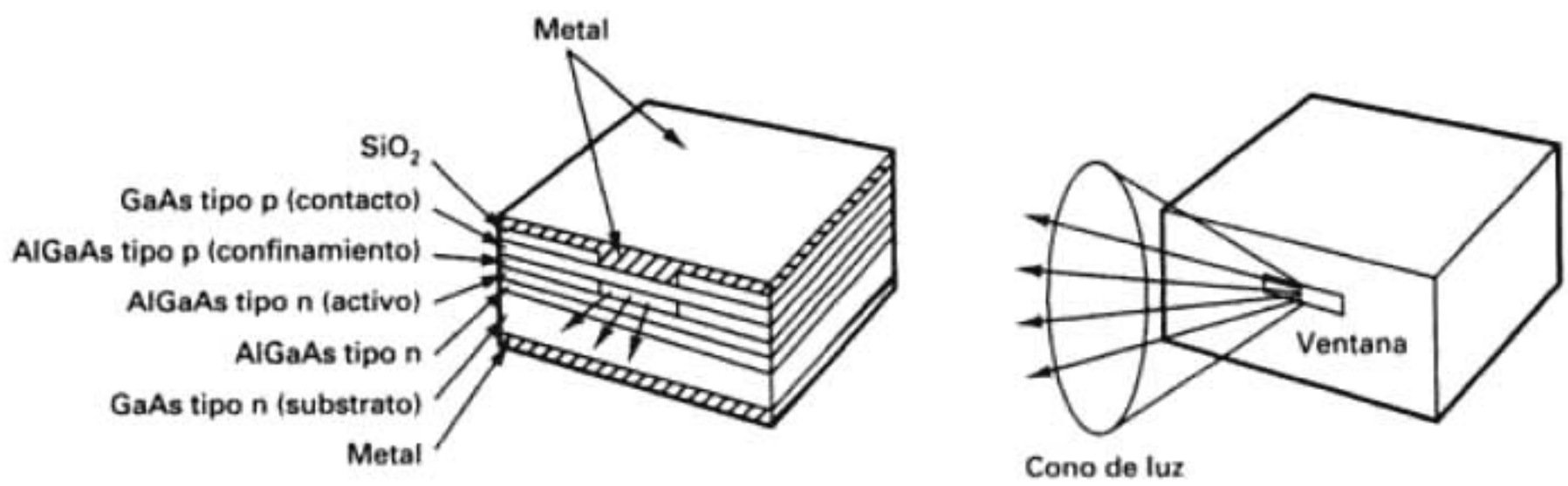


FIGURA 11-31 LED emisor de borde

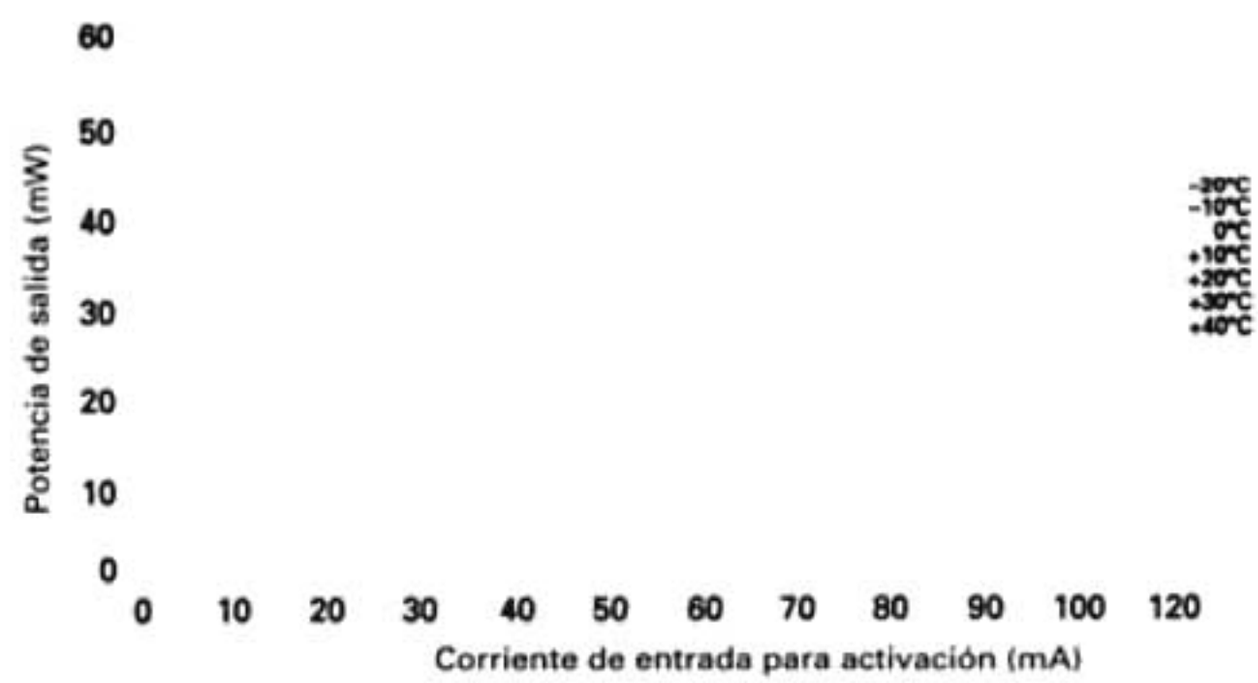


FIGURA 11-32 Potencia de entrada en función de corriente directa y de temperatura de funcionamiento de un LED

LED emisores de borde. El LED emisor de borde, desarrollado por la RCA, se ve en la fig. 11-31. Estos LED emiten una distribución más direccional de luz que los LED de superficie emisora. La fabricación se parece a los diodos planos y de Burrus, pero la superficie emisora es una banda, más que un área circular confinada. La luz se emite de una banda activa y forma un haz elíptico. Los LED de superficie emisora se usan con más frecuencia que los emisores de borde, porque emiten más luz. Sin embargo, las pérdidas por el acoplamiento de los emisores de superficie son mayores, y tienen anchos de banda más angostos.

La potencia luminosa *radiante* que emite un LED es una función lineal de la corriente directa que pasa por el dispositivo (fig. 11-32). También se ve que la potencia óptica de salida de un LED es, en parte, una función de la temperatura de funcionamiento.

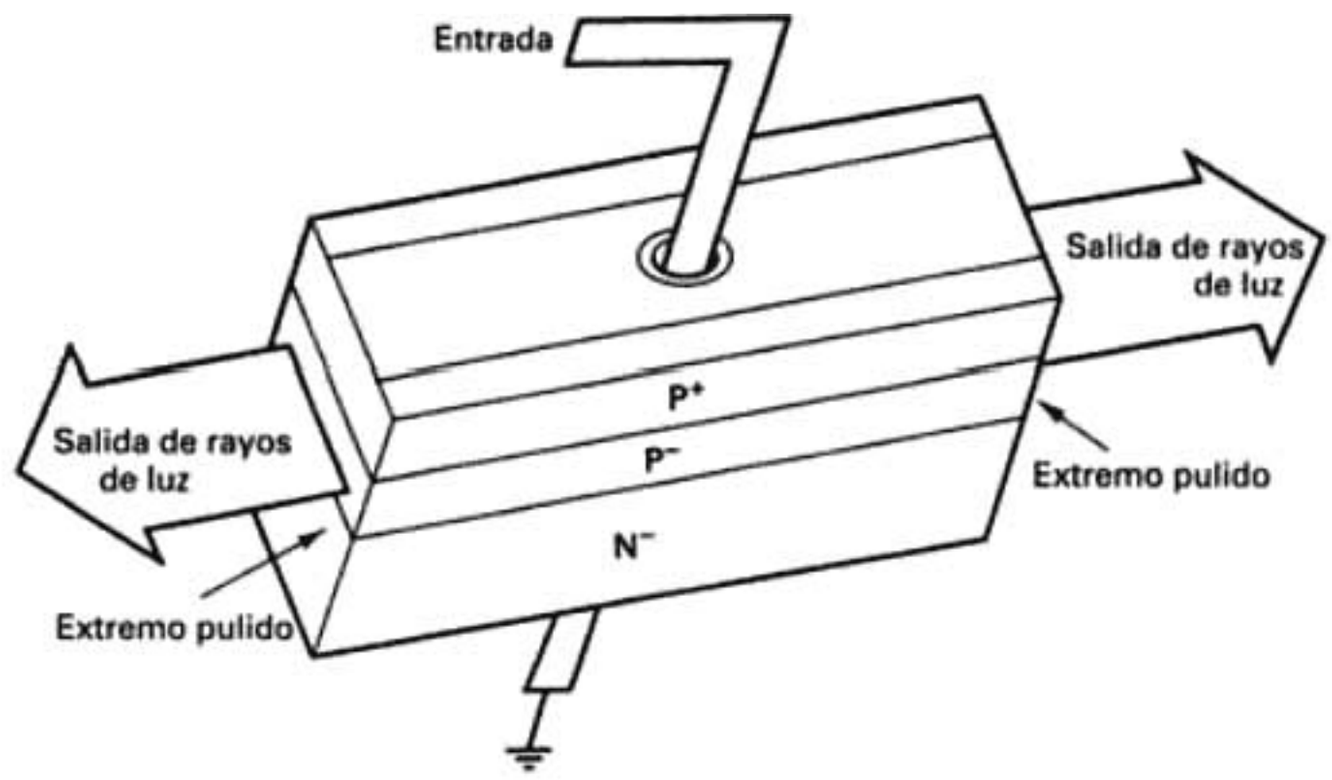


FIGURA 11-33 Construcción de un diodo de láser de inyección

Diodo de láser de inyección

Los láseres se fabrican con muchos y diversos materiales, que incluyen gases, líquidos y sólidos, aunque el tipo de láser que se usa con más frecuencia para comunicaciones con fibra óptica es el láser de semiconductor.

El *diodo de láser de inyección* (ILD, de *injection laser diode*) se parece al LED. De hecho, por debajo de cierta corriente umbral, un ILD funciona en forma parecida a un LED. Arriba de la corriente de umbral, un ILD oscila y se produce la emisión láser. Al pasar la corriente por un diodo de unión *p-n* de polarización directa, se emite luz por emisión espontánea, a una frecuencia determinada por la banda prohibida del material semiconductor. Cuando se llega a determinado valor de la corriente, la cantidad de portadores minoritarios y de fotones que se producen en ambos lados de la unión *p-n* llega a un valor en el que comienzan a chocar con portadores minoritarios ya excitados. Esto causa un aumento en el nivel de la energía de ionización y hace que los portadores sean inestable. Al suceder eso, un portador normal se recombina con un portador del tipo contrario a un valor de la energía que es mayor que su valor normal antes del choque. En el proceso se crean dos fotones; uno es estimulado por otro. En esencia, se realiza una ganancia en la cantidad de fotones. Para que eso suceda se requiere una gran corriente directa que pueda suministrar muchos portadores (huecos y electrones).

La fabricación de un ILD se parece a la de un LED (fig. 11-33), excepto que los extremos están muy pulidos. Los extremos con acabado de espejo atrapan los fotones en la región activa y, al reflejarse de un lado a otro, estimulan a electrones libres, para recombinarse con huecos a un valor de energía mayor que el normal. Éste es el proceso llamado *láser*, o estimulación de emisión.

La potencia luminosa radiante de salida de un ILD típico se ve en la fig. 11-34. Se aprecia que se obtiene muy poca potencia de salida hasta que se llega a la corriente de umbral; entonces se produce la estimulación. Después de comenzada la estimulación, la potencia óptica de salida aumenta en forma dramática, con pequeños aumentos en la corriente de activación. También se ve que la magnitud de la potencia óptica de salida del ILD depende más de la temperatura de funcionamiento que en el caso del LED.

La fig. 11-35 muestra las distribuciones de irradiación de luz normales en un LED y en un ILD. Como la luz se irradia saliendo del extremo de un ILD en forma de haz angosto y concentrado, tiene una distribución de radiación más dirigida.

Ventajas de los ILD

1. Como los ILD tienen una distribución de irradiación más dirigida, es más fácil de acoplar su luz en una fibra óptica. Esto reduce las pérdidas por acoplamiento y permite usar fibras más pequeñas.

2. La potencia radiante de salida de un ILD es mayor que la de un LED. Una potencia normal de salida de un ILD es 5 mW (7 dBm), en comparación con 0.5 mW (–3 dBm) para los LED. Eso permite que los ILD proporcionen una mayor potencia de activación, y usarlos en sistemas que funcionen a través de mayores distancias.

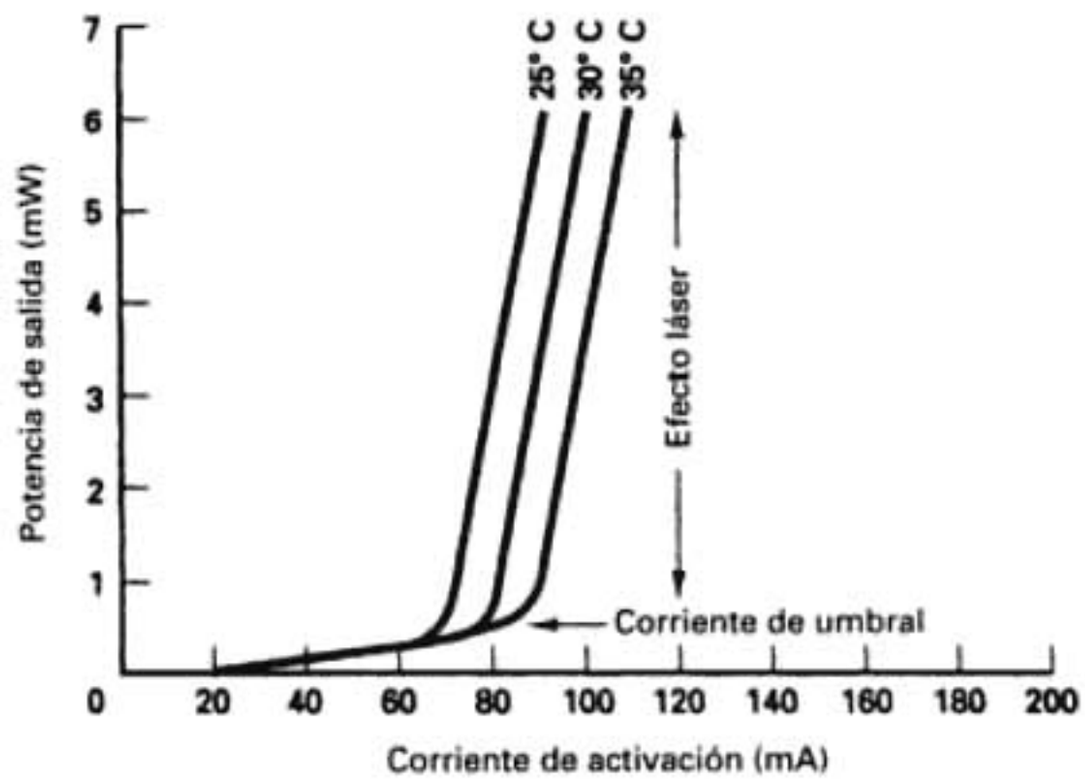


FIGURA 11-34 Potencia de salida en función de la corriente directa y la temperatura, para un ILD

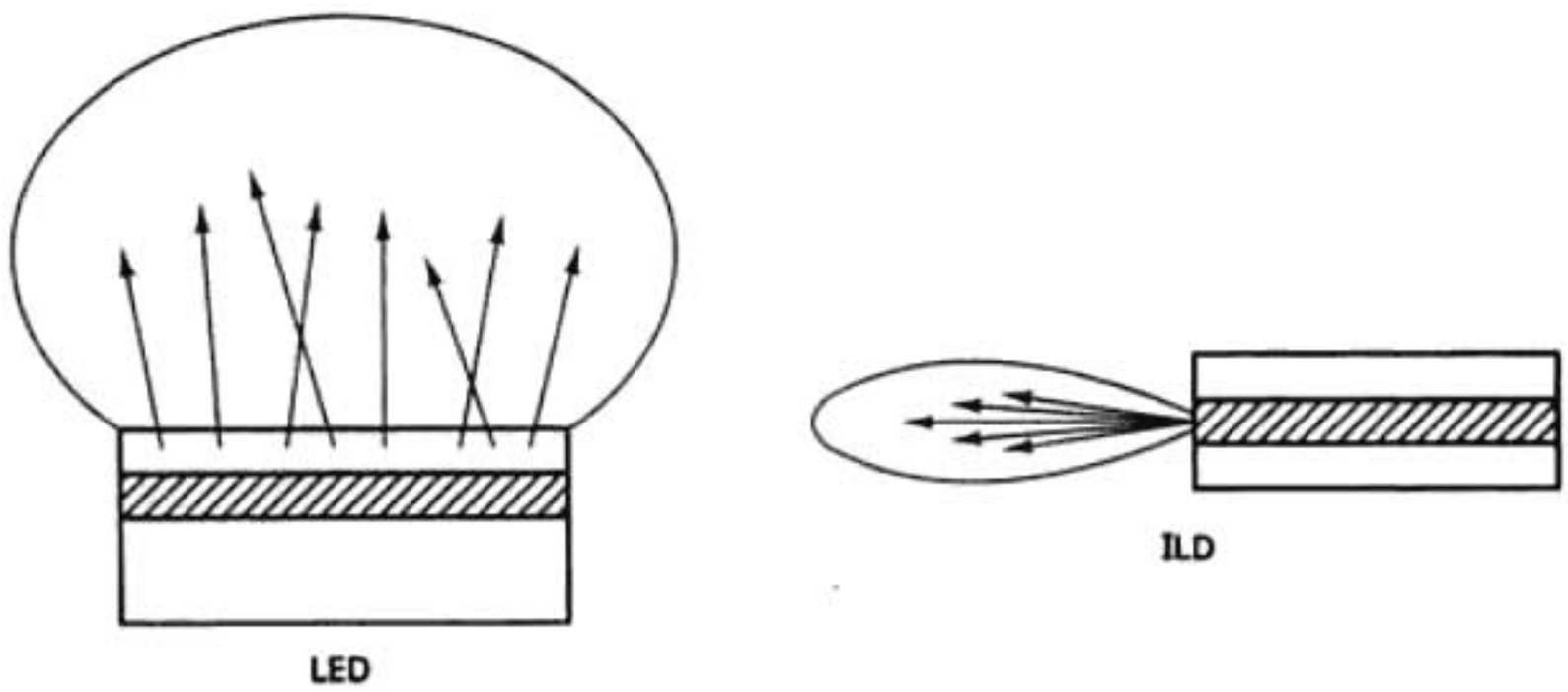


FIGURA 11-35 Distribuciones de radiación de un LED y de un ILD

3. Los ILD se pueden usar a frecuencias mayores de bits que los LED.

4. Los ILD generan luz monocromática, lo cual reduce la dispersión cromática o de longitudes de onda.

Desventajas de los ILD

1. Los ILD cuestan normalmente 10 veces más que los LED.

2. Como los ILD trabajan con mayores potencias, suelen tener duraciones mucho menores que los LED.

3. Los ILD dependen más de la temperatura que los LED.

DETECTORES DE LUZ

Hay dos dispositivos que se suelen usar para detectar la energía luminosa en los receptores de comunicación con fibra óptica: los diodos PIN (tipo *p*-tipo *n* intrínseco) y los APD (fotodiodos de avalancha).

Diodos PIN

Un *diodo PIN* es un *fotodiodo de capa de agotamiento*, y es probable que sea el dispositivo más usado como detector de luz en los sistemas de comunicaciones con fibra óptica. La fig. 11-36

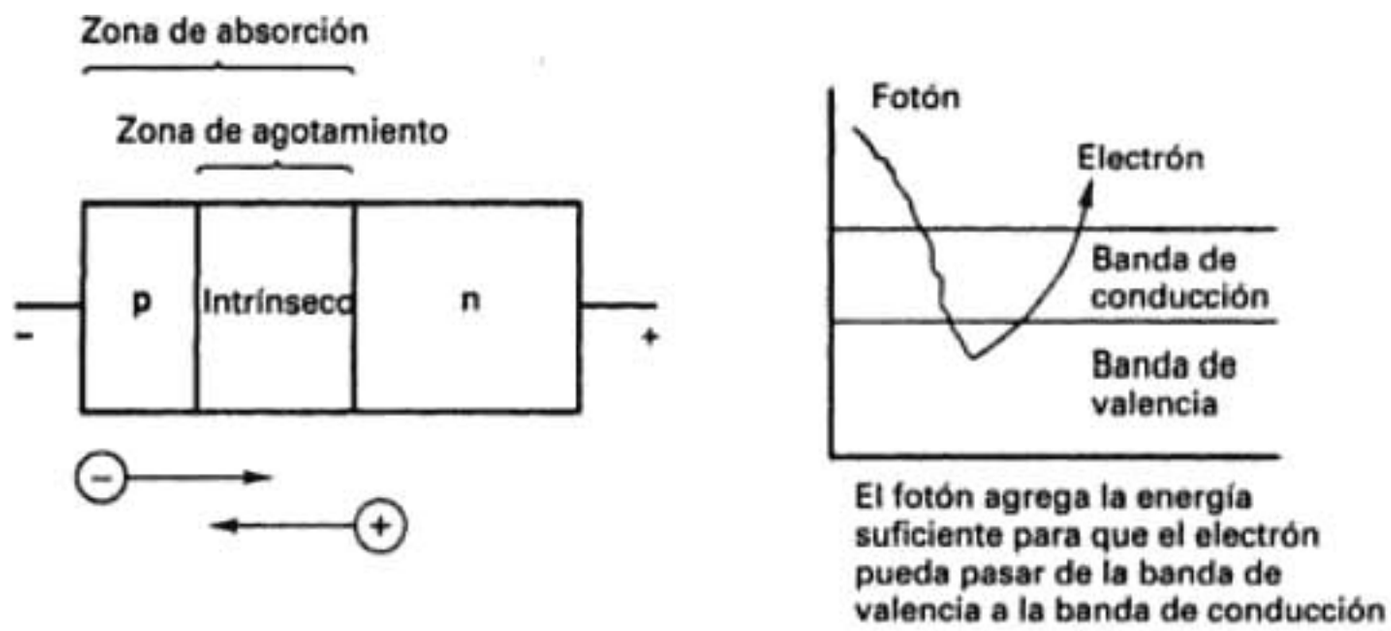


FIGURA 11-36 Construcción de un fotodiodo PIN

muestra el esquema básico de un diodo PIN. Una capa dopada muy ligeramente (casi pura o intrínseca) de material semiconductor tipo *n* se encierra entre la unión de las dos áreas de contacto tipo *p* y tipo *n*, muy dopadas. La luz entra al dispositivo a través de una ventana muy pequeña y cae sobre el material intrínseco carente de portadores. El material intrínseco se hace lo suficientemente grueso como para que la mayoría de los fotones que entren al dispositivo queden absorbidos por esta capa. En esencia, el fotodiodo PIN funciona exactamente al contrario que un LED. La mayoría de los fotones quedan absorbidos por electrones de la banda de valencia del material intrínseco. Cuando se absorben los fotones agregan la energía suficiente para generar portadores en la región de agotamiento, y permiten el paso de la corriente por el dispositivo.

Efecto fotoeléctrico. La luz que entra por la ventana de un diodo PIN se absorbe en el material intrínseco y agrega la energía suficiente para hacer que los electrones pasen de la banda de valencia a la banda de conducción. El aumento en la cantidad de electrones que pasan a la banda de conducción produce un aumento en la cantidad de huecos en la banda de valencia. Para hacer que la corriente pase por un fotodiodo, se debe absorber luz de la energía suficiente para comunicar a los electrones de valencia la energía suficiente para que salten la banda prohibida. La banda prohibida, para el silicio, es 1.12 eV (electrón volts). La descripción matemática es la siguiente.

La equivalencia de unidades de energía es

$$1\ \text{eV} = 1.6 \times 10^{-19}\ \text{J}$$

Así, la banda prohibida para el silicio, es

$$E_g = (1.12\ \text{eV})\left(1.6 \times 10^{-19}\ \frac{\text{J}}{\text{eV}}\right) = 1.792 \times 10^{-19}\ \text{J}$$

y $\quad$ energía $(E) = hf$ $\quad$ (11-27)

en donde $\quad h$ = constante de Planck = 6.6256×10^{-34} J/Hz
$\quad\quad\quad f$ = frecuencia (hertz)

De esta ecuación se despeja f

$$f = \frac{E}{h} \tag{11-28}$$

Para un fotodiodo de silicio,

$$f = \frac{1.792 \times 10^{-19}\ \text{J}}{6.6256 \times 10^{-34}\ \text{J/Hz}} = 2.705 \times 10^{14}\ \text{Hz}$$

Al pasar a longitud de onda se obtiene

$$\lambda = \frac{c}{f} = \frac{3 \times 10^{8}\ \text{m/s}}{2.705 \times 10^{14}\ \text{Hz}} = 1109\ \text{nm/ciclo}$$

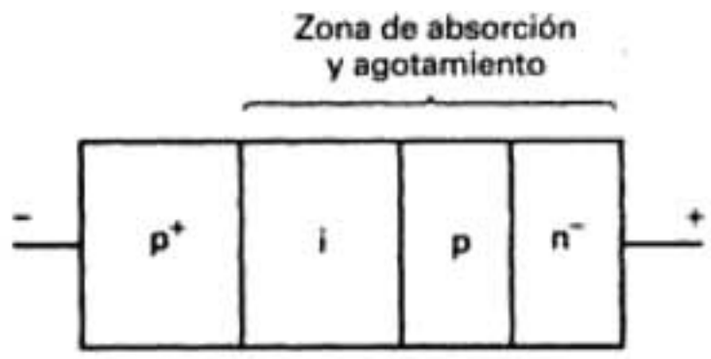


FIGURA 11-37 Construcción de un fotodiodo de avalancha

Así, la luz de longitudes de onda de 1109 nm o menores, o de frecuencias de 2.705×10^{14} Hz o mayores, es la que se requiere para hacer que haya electrones suficientes que salten la banda prohibida de un fotodiodo de silicio.

Fotodiodos de avalancha

La fig. 11-37 muestra la construcción básica de un *fotodiodo de avalancha* (APD, de *avalanche photodiode*). Un APD es una estructura *pipn*. La luz entra al diodo y se absorbe en la capa *n*, delgada y muy dopada. Entre la unión *i-p-n* se desarrolla una gran intensidad de campo eléctrico, por polarización inversa, que causa ionización por impacto. Durante la ionización por impacto, un portador puede adquirir la energía suficiente para ionizar otros electrones enlazados. Estos portadores ionizados, a su vez, provocan más ionizaciones. El proceso continúa como en una avalancha y es, de hecho, equivalente a una ganancia o multiplicación interna de portadores. En consecuencia, los APD son más sensibles que los diodos PIN y requieren menos amplificación adicional. Las desventajas de los APD son los tiempos de tránsito relativamente grandes, y ruido adicional, generado internamente, debido al factor de multiplicación por avalancha.

Características de los detectores de luz

Las características más importantes de los detectores de luz son:

1. *Responsividad.* La responsividad es una medida de la eficiencia de conversión de un fotodetector. Es la relación de la corriente de salida de un fotodiodo entre la potencia óptica que le entra, y tiene unidades de amperes/watt. En general, se menciona la responsividad a una determinada longitud de onda o frecuencia.

2. *Corriente oscura.* La corriente oscura es la corriente de reposo que pasa por un fotodiodo cuando no hay entrada luminosa. Se debe a portadores generados térmicamente en el diodo.

3. *Tiempo de tránsito.* Es el tiempo que tarda un portador inducido por la luz en cruzar la región de agotamiento. Este parámetro determina la máxima frecuencia posible de bits con determinado fotodiodo.

4. *Respuesta espectral.* Es el intervalo de longitudes de onda que se puede usar con determinado fotodiodo. En general, la respuesta espectral relativa se grafica en función de la longitud de onda o de la frecuencia. La fig. 11-38 es un ejemplo ilustrativo de una curva de respuesta espectral. Se puede ver que este fotodiodo en particular absorbe con mayor eficiencia la energía dentro del intervalo de 800 a 820 nm.

5. *Sensibilidad a la luz.* En esencia, esta sensibilidad es la potencia óptica mínima que puede recibir un detector para producir una señal eléctrica útil de salida. La sensibilidad a la luz se acostumbra mencionar para determinada longitud de onda, ya sea en dBm o en dBμ.

LÁSERES

La tecnología láser se centra en la concentración de luz en forma de haces muy pequeños y poderosos. El acrónimo se generalizó cuando la tecnología pasó de las microondas a las ondas de luz.

El primer láser fue desarrollado por Theodore H. Maiman, científico de Hughes Aircraft Company, de California. Dirigió un rayo de luz a cristales de rubí, con una lámpara de xenón, y midió la radiación emitida por el rubí. Descubrió que cuando la radiación emitida aumentaba de cierto umbral, se volvía en extremo intensa, y muy direccional. Los láseres de uranio se de-

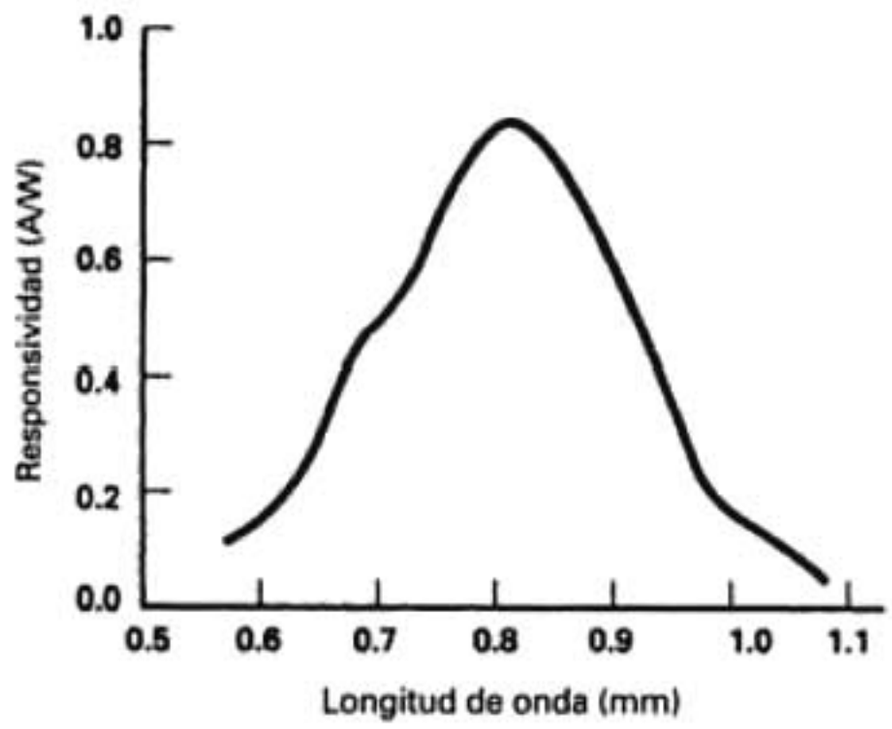


FIGURA 11-38 Curva espectral de respuesta

sarrollaron en 1960, junto con los de otros materiales de tierras raras. También en 1960, A. Javin de los Bell Laboratories desarrolló el láser de helio. Los láseres de semiconductor (diodos de inyección láser) se fabricaron en 1962, en General Electric, IBM y Lincoln Laboratories.

Tipos de láser

Básicamente hay cuatro tipos de láser: de gas, líquido, sólido y de semiconductor.

1. *Láseres de gas.* Los láseres de gas usan una mezcla de helio y neón encerrada en un tubo de vidrio. Cuando se descarga una corriente en el gas se emite un flujo de ondas luminosas coherentes (de una frecuencia) a través del acoplador de salida. La salida de onda luminosa continua es monocromática (de un color).

2. *Láseres de líquido.* En estos láseres se usan colorantes orgánicos encerrados en un tubo de vidrio, como medio activo. El colorante se hace circular con una bomba dentro del tubo. Un pulso poderoso de luz excita al colorante orgánico.

3. *Láseres de sólido.* En ellos se usa un cristal cilíndrico y macizo, como el rubí, como medio activo. Cada extremo del rubí está pulido y paralelo al otro. El rubí se excita con una lámpara de tungsteno conectada a una fuente de poder de corriente alterna. La salida del láser es una onda continua.

4. *Láseres de semiconductor.* Estos láseres se fabrican con semiconductores de unión *p-n* y se suelen llamar *diodos de láser de inyección* (ILD, de *injection laser diode*). El mecanismo de excitación es un suministro de corriente directa que controla la cantidad de corriente al medio activo. La luz de salida de un ILD se modula con facilidad, y eso lo hace muy útil en muchas aplicaciones de comunicaciones electrónicas.

Características de los láseres

Todas las clases de láser tienen varias características comunes: 1) todos usan un material activo que convierte la energía en luz láser, 2) una fuente de bombeo que proporciona la potencia o energía, 3) el sistema óptico para dirigir el haz que se va a amplificar a través del material, 4) el sistema óptico para dirigir el haz en forma de un cono de divergencia angosto y potente, 5) un mecanismo de retroalimentación para obtener funcionamiento continuo, y 6) un acoplador de salida, para transmitir la potencia que sale del láser.

La radiación de un láser es extremadamente intensa y direccionada. Cuando se enfoca en forma de un haz capilar muy fino, puede concentrar su potencia. Si se permitiera diverger al haz de luz, perdería la mayor parte de su potencia.

Construcción de los láseres

La fig. 11-39 muestra la construcción de un láser básico. Una fuente de poder se conecta a un tubo de destello, que está enrollado sobre un tubo de vidrio que contiene al medio activo. Un extremo del tubo de vidrio es una cara pulida a espejo, 100% reflejante, para tener reflexión interna. El tubo de destello se energiza con un pulso de disparo y produce un destello de luz de

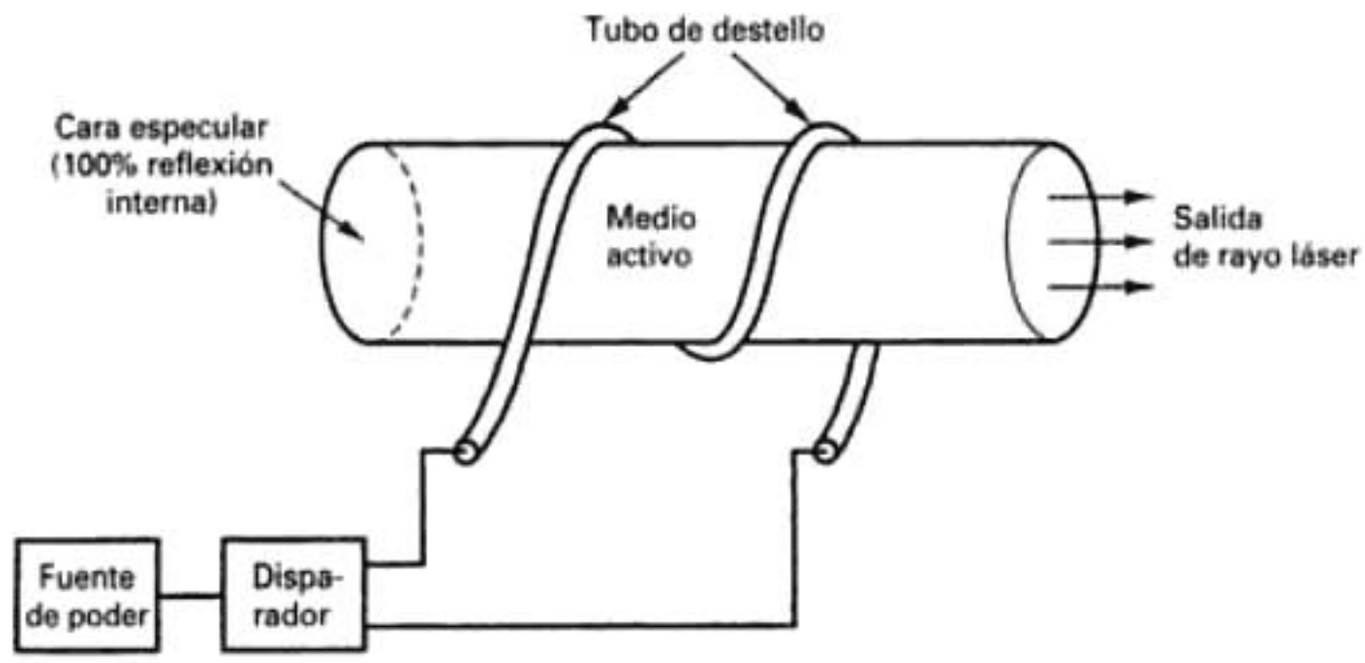


FIGURA 11-39 Construcción de un láser

gran intensidad (parecido a una lámpara de destello en fotografía). El destello excita los átomos de cromo en el interior de la estructura cristalina. El proceso de bombeo eleva la energía de los átomos de cromo desde el estado fundamental hasta un estado excitado de energía. A continuación, los iones decaen y pasan a un nivel intermedio de energía. Cuando la población de iones en el nivel intermedio es mayor que la del estado fundamental, sucede una inversión de poblaciones, que causa la acción de láser (emisión). Después de cierto tiempo, los átomos excitados de cromo caen al nivel fundamental de energía. En esta ocasión se emiten fotones. Un fotón es un paquete de energía radiante. Los fotones emitidos chocan con átomos y causan la emisión de otros dos fotones (de aquí el término "emisión estimulada"). La frecuencia de la energía determina la intensidad de los fotones: mayores frecuencias producen fotones de mayor intensidad.

PROVISIÓN PARA ENLACE EN EL SISTEMA DE FIBRA ÓPTICA

Como en cualquier sistema de comunicaciones, los sistemas de fibra óptica consisten en una fuente y un destino, que están separados por numerosos componentes y dispositivos que introducen diversas cantidades de pérdida o de ganancia a la señal, al propagarse a través del sistema. La fig. 11-40 muestra dos configuraciones normales de sistema de comunicaciones con fibra óptica. En la fig. 11-40a se ve un sistema sin repetidoras, en donde la fuente y el destino se interconectan a través de uno o más tramos de cable óptico. En un sistema sin repetidoras no hay amplificadores ni regeneradores entre la fuente y el destino.

La fig. 11-40b representa un sistema de fibra óptica que incluye una repetidora que puede amplificar o regenerar la señal. Los sistemas con repetidora se usan, claro está, cuando la fuente y el destino están separados por grandes distancias.

Las necesidades de enlace se calculan en general entre una fuente luminosa y un detector de luz; en consecuencia, para nuestro ejemplo, se buscan provisiones para un sistema sin repetidoras. Éste consiste en una fuente luminosa, como un LED o un ILD, y un detector de luz, como por ejemplo un APD, conectados con fibra óptica y conectores. En consecuencia, la provisión de enlace consiste en una fuente de potencia luminosa, un detector de luz y diversas pérdidas en el cable y conectores. Las pérdidas características de los enlaces de fibra óptica incluyen las siguientes:

1. *Pérdidas en el cable*. Las pérdidas en el cable dependen de la longitud, del material y de la pureza del material del cable. En general, se especifican en dB/km, y pueden variar desde décimas de dB hasta varios dB por km.
2. *Pérdidas en el conector.* A veces se usan conectores mecánicos para unir dos tramos de cable. Si la conexión mecánica no es perfecta, puede escapar energía luminosa y

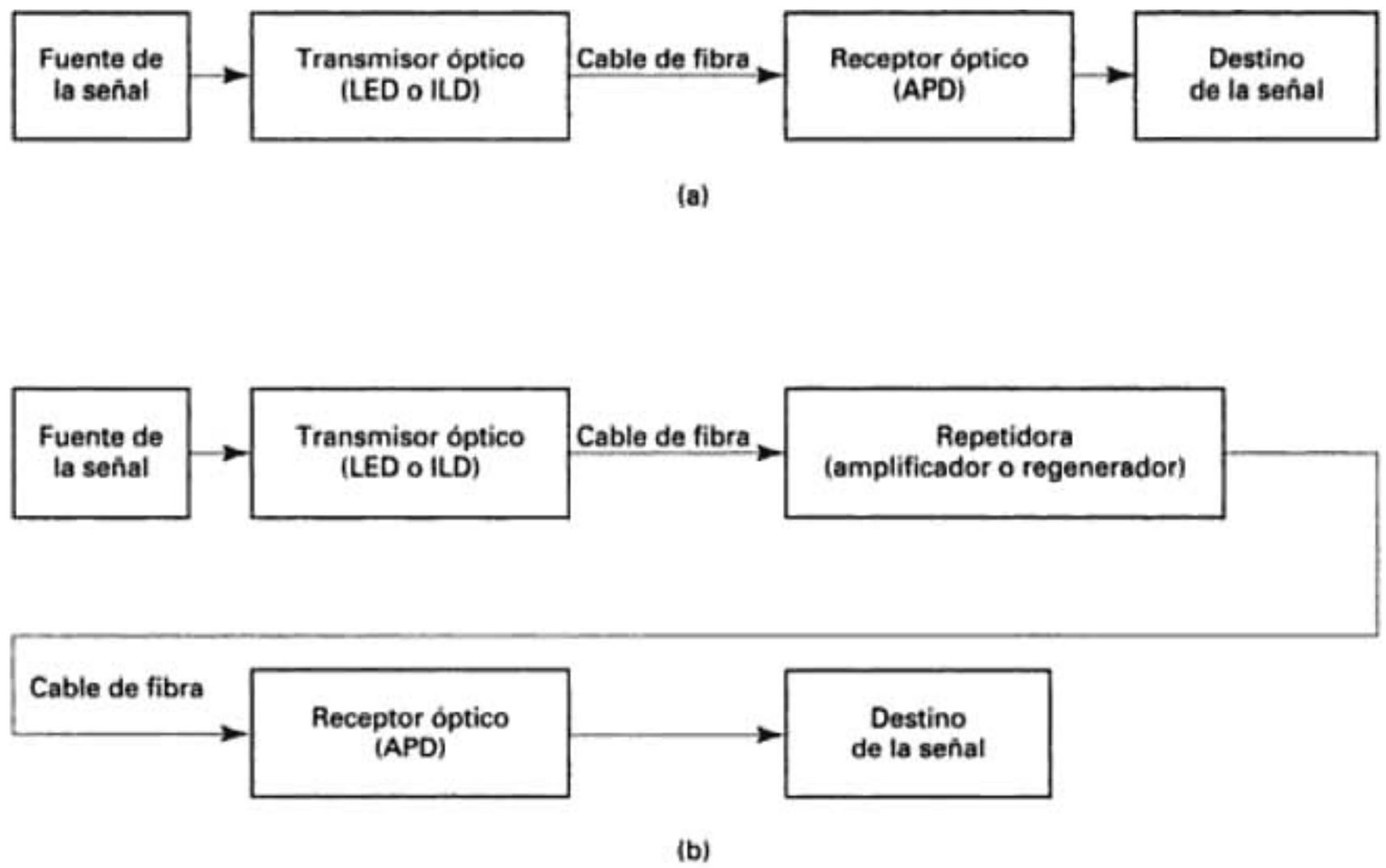


FIGURA 11-40 Sistemas de comunicaciones por fibra óptica: (a) sin repetidoras; (b) con repetidoras

causar una reducción de la potencia óptica. Las pérdidas en el conector suelen variar entre unas décimas de dB hasta 2 dB para cada conector.

3. *Pérdida en interfaz de fuente a cable.* La interfaz mecánica que se usa para contener la fuente de luz y fijarla al cable casi nunca es perfecta. En consecuencia, un pequeño porcentaje de la potencia óptica no se acopla al cable, y representa una pérdida de potencia en el sistema de algunas décimas de dB.
4. *Pérdida en interfaz de cable a detector de luz.* La interfaz mecánica para contener el detector de luz y fijarlo al cable tampoco es perfecta y, en consecuencia, evita que un pequeño porcentaje de la potencia que sale del cable entre al detector de luz. Esto, naturalmente, representa una pérdida para el sistema, que suele ser de algunas décimas de dB.
5. *Pérdida en empalmes.* Si se requiere más que un tramo continuo de cable, se pueden fundir (empalmar) los tramos. Como los empalmes no son perfectos, se pueden introducir pérdidas que van desde un par de décimas de dB hasta varios dB en la señal.
6. *Dobleces del cable.* Cuando un cable óptico se dobla formando un ángulo muy agudo, sus características internas cambian radicalmente. Si los dobleces son grandes, ya no se lograrán las reflexiones totales de algunos de los rayos de luz, y se refractarán. La luz refractada en la interfaz entre el núcleo y el revestimiento entra al revestimiento y causa una pérdida neta de la señal, de algunas décimas de dB hasta varios dB.

Como en cualquier provisión de enlace o de sistema, la potencia útil disponible en el receptor depende de la potencia de transmisión y de las pérdidas en el enlace. La ecuación correspondiente es

$$P_r = P_t - \text{pérdidas} \tag{11-29}$$

en donde P_r = potencia recibida (dBm)
P_t = potencia transmitida (dBm)
pérdidas = suma de todas las pérdidas (dB)

Ejemplo 11-7

Calcular la potencia óptica recibida, en dBm y en watts, con un enlace de 20 km de fibra óptica con los siguientes parámetros:

Potencia de salida del LED, 30 mW
Cuatro tramos de 5 km de cable óptico, cada uno con pérdida de 0.5 dB/km
Tres conectores de cable a cable con una pérdida de 2 dB en cada uno
No hay empalmes de cable
Pérdida en interfaz de fuente luminosa a fibra de 1.9 dB
Pérdida entre fibra y detector de luz de 2.1 dB
No hay pérdidas por dobleces de cable

Solución La potencia de salida del LED se convierte a dBm aplicando la ecuación 11-25

$$P_{sal} = 10 \log \frac{30 \text{ mW}}{1 \text{ mW}}$$
$$= 14.8 \text{ dBm}$$

La pérdida en el cable no es más que el producto de la longitud total del cable, en km, por la pérdida en dB/km. Cuatro tramos de 5 km de cable hacen una longitud total de 20 km; en consecuencia

$$\text{pérdida total en el cable} = 20 \text{ km} \times 0.5 \text{ dB/km}$$
$$= 10 \text{ dB}$$

Las pérdidas en los conectores de cable son tan sólo el producto de la pérdida, en dB por conector, por la cantidad de conectores. La cantidad máxima de conectores siempre es uno menos que la cantidad de tramos de cable. Entonces, se requieren los conectores para cuatro tramos de cable y en consecuencia

$$\text{pérdida total en los conectores} = 3 \text{ conectores} \times 2 \text{ dB/conector}$$
$$= 6 \text{ dB}$$

Las pérdidas de la fuente al cable y del cable al detector de luz fueron dados: 1.9 dB y 2.1 dB, respectivamente. Así

$$\begin{aligned}\text{pérdida total} &= \text{pérdida en el cable} + \text{pérdidas en los conectores} + \text{pérdida de fuente a cable} \\ &\quad + \text{pérdida de cable a detector de luz} \\ &= 10 \text{ dB} + 6 \text{ dB} + 1.9 \text{ dB} + 2.1 \text{ dB} \\ &= 20 \text{ dB}\end{aligned}$$

La potencia de recepción se determina sustituyendo valores en la ecuación 11-29

$$\begin{aligned}P_r &= 14.8 \text{ dBm} - 20 \text{ dB} \\ &= -5.2 \text{ dBm} \\ &= 0.302 \text{ mW}\end{aligned}$$

PREGUNTAS

11-1. Defina un sistema de fibra óptica.

11-2. ¿Cuál es la relación entre capacidad de información y amplitud de banda?

11-3. ¿Qué desarrollo, en 1951, fue un gran avance en el campo de la óptica de fibras? ¿Y en 1960? ¿Y en 1970?

11-4. Describa las ventajas y desventajas de los cables de fibra óptica y los cables metálicos.

11-5. Describa las unidades constructivas primarias de un sistema de fibra óptica.

11-6. Describa las diferencias entre los cables de fibra de vidrio y de fibra de plástico.

11.7. Describa, en forma breve, la construcción de un cable de fibra óptica.

11-8. Defina los siguientes términos: *velocidad de propagación*, *refracción* e *índice de refracción*.

11-9. Enuncie la ley de Snell de la refracción y describa su importancia en los cables de fibra óptica.

11-10. Defina qué es el *ángulo crítico*.

11-11. Describa qué quiere decir *modo de operación* y *perfil de índice*.

11-12. Describa un cable de fibra de índice escalonado, y uno de índice graduado.

11-13. Describa las ventajas y desventajas del índice escalonado, el índice graduado, la propagación unimodal y la propagación multimodal.

11-14. ¿Por qué es imposible la propagación unimodal en las fibras de índice graduado?

11-15. Describa qué es abertura de fuente a fibra.

11-16. ¿Qué son el *ángulo de aceptación* y el *cono de aceptación* para un cable de fibra?

11-17. Defina qué es *abertura numérica.*

11-18. Haga una lista de las pérdidas asociadas con los cables de fibra, y descríbalas en forma breve.

11-19. ¿Qué es el *ensanchamiento de pulso*?

11-20. Defina la *constante de ensanchamiento de pulso*.

11-21. Haga una lista de las diversas pérdidas por acoplamiento, y descríbalas en forma breve.

11-22. Describa en forma breve el funcionamiento de un diodo emisor de luz.

11-23. ¿Cuáles son los dos tipos primarios de LED?

11-24. Describa en forma breve el funcionamiento de un diodo de láser de inyección.

11-25. ¿Qué es el efecto láser?

11-26. Describa las ventajas y desventajas de los ILD y los LED.

11-27. Describa en forma breve la función de un fotodiodo.

11-28. Describa el efecto fotoeléctrico.

11-29. Explique la diferencia entre un diodo PIN y un APD.

11-30. Haga una lista de las características primarias de los detectores de luz y descríbalas.

PROBLEMAS

11-1. Calcule las longitudes de onda, en nanómetros y angstroms, de las siguientes frecuencias:
(a) 3.45×10^{14} Hz.
(b) 3.62×10^{14} Hz.
(c) 3.21×10^{14} Hz.

11-2. Calcule la frecuencia de las siguientes longitudes de onda luminosa:
(a) 670 nm.
(b) 7800 Å.
(c) 710 nm.

11-3. Para una interfaz vidrio (n = 1.5)/cuarzo (n = 1.38) y un ángulo de incidencia de 35°, calcule el ángulo de refracción.

11-4. Calcule el ángulo crítico para la fibra que se menciona en el problema 11-3.

11-5. Calcule el ángulo de aceptación para el cable que se menciona en el problema 11-3.

11-6. Calcule la abertura numérica del cable que se describe en el problema 11-3.

11-7. Calcule la frecuencia máxima de bits con codificación RZ y NRZ, para las siguientes constantes de ensanchamiento de pulso y longitudes de cable:
(a) Δt = 10 ns/m, L = 100 m.
(b) Δt = 20 ns/m, L = 1000 m.
(c) Δt = 2000 ns/km, L = 2 km.

11-8. Calcule la mínima frecuencia luminosa que puede detectar un fotodiodo con banda prohibida de 1.2 eV.

11-9. Calcule las longitudes de onda, en nanómetros y angstroms, de las siguientes frecuencias luminosas:
(a) 3.8×10^{14} Hz.
(b) 3.2×10^{14} Hz.
(c) 3.5×10^{14} Hz.

11-10. Calcule las frecuencias de la luz de las siguientes longitudes de onda:
(a) 650 nm.
(b) 7200 Å.
(c) 690 nm.

CAPÍ

Imagen protegida por derechos de autor

Comunicaciones digitales

INTRODUCCIÓN

En esencia, las comunicaciones electrónicas son la transmisión, recepción y procesamiento de *información* usando circuitos electrónicos. Se define a la información como el conocimiento o las señales inteligentes comunicados o recibidos. La fig. 12-1 muestra un diagrama simplificado de bloques de un sistema de comunicaciones electrónicos, que comprende tres secciones principales: una *fuente*, un *destino* y un *medio de transmisión*. La información se propaga a través de un sistema de comunicaciones en forma de símbolos que pueden ser *analógicos* (proporcionales), como la voz humana, la información de las imágenes de video, o la música; o bien pueden ser *digitales* (discretos), como los números codificados en sistema binario, los códigos alfanuméricos, los símbolos gráficos, los códigos de operación de microprocesadores o la información de bases de datos. Sin embargo, y con mucha frecuencia, la información de la fuente no es adecuada para transmitirse en su forma original y se debe convertir a una forma más adecuada antes de transmitirla. Por ejemplo, con los sistemas de comunicaciones digitales, la información analógica se convierte a forma digital antes de transmitirla, y con los sistemas de comunicaciones analógicos, los datos digitales se convierten en señalas analógicas antes de transmitirlos.

Los sistemas tradicionales de comunicaciones electrónicas, que usan técnicas convencionales de modulación analógica, como los de *modulación de amplitud* (AM), *modulación de frecuencia* (FM) y *modulación de fase* (PM) se están sustituyendo rápidamente por sistemas de comunicación digital, más modernos, que tienen varias y notables ventajas sobre los sistemas analógicos tradicionales: facilidad de procesamiento, facilidad de multiplexado e inmunidad al ruido.

Sin embargo, *comunicaciones digitales* es un término bastante ambiguo que podría significar cosas totalmente distintas a personas distintas. En el contexto de este libro, los sistemas digitales de comunicaciones (o sistemas de comunicaciones digitales) incluyen a aquellos en los que hay portadoras analógicas de frecuencia relativamente alta, que se modulan mediante señales de información digital de relativamente baja frecuencia, y a los sistemas que manejan la transmisión de pulsos digitales.

FIGURA 12-1 Diagrama de bloques simplificado de un sistema de comunicaciones electrónicas

COMUNICACIONES DIGITALES

El término *comunicaciones digitales* abarca una gran área de técnicas de comunicaciones, que incluyen la *transmisión digital* y el (o la) radio digital. Se aplica a la transmisión de pulsos digitales entre dos o más puntos en un sistema de comunicaciones. La radio digital es la transmisión de portadoras analógicas moduladas digitalmente entre dos o más puntos de un sistema de comunicaciones. Los sistemas digitales de transmisión requieren una instalación física entre el transmisor y el receptor, como un par de hilos metálicos, un cable coaxial o un cable de fibra óptica. En los sistemas digitales de radio, el medio de transmisión podría ser el espacio libre, la atmósfera terrestre o una instalación física, como un cable metálico o de fibra óptica.

La fig. 12-2 muestra diagramas simplificados de bloques de un sistema de transmisiones digitales y uno de radio digital. En el primero, la fuente original de información puede estar en forma digital o analógica. Si está en forma analógica debe convertirse en pulsos digitales antes de la transmisión, y reconvertirse a la forma analógica en el extremo de recepción. En un sistema digital de radio, la señal moduladora de entrada y la señal desmodulada de salida son pulsos digitales. Éstos se podrían originar en un sistema digital de transmisión, o en la fuente original digital, como puede ser una computadora central, o bien estar en la codificación binaria de una señal analógica.

LÍMITE DE SHANNON DE CAPACIDAD DE INFORMACIÓN

La *capacidad de información* de un sistema de comunicaciones representa la cantidad de símbolos independientes que pueden transportarse por el sistema en determinada unidad de tiempo. El símbolo binario más básico es el *dígito binario* o bit. En consecuencia, conviene con frecuencia expresar la capacidad de información de un sistema en *bits por segundo*, o bps. En 1928, R. Hartley, de Bell Telephone Laboratories, desarrolló una relación útil entre la amplitud de banda, el tiempo de transmisión y la capacidad de información. El planteamiento sencillo de la *ley de Hartley* es

$$I \propto B \times T \tag{12-1a}$$

en donde I = capacidad de información (bits por segundo)
B = amplitud de banda (hertz)
T = tiempo de transmisión (segundos)

Se ve, en la ecuación 12-1a, que la capacidad de información es una función lineal de la amplitud de banda y del tiempo de transmisión, y es directamente proporcional a ambos. Si cambia el ancho de información o el tiempo de transmisión, se tendrá un cambio directamente proporcional de capacidad de información.

C. E. Shannon (también de Bell Telephone Laboratories) publicó en 1948 un trabajo en el *Bell System Technical Journal* donde se relacionaba la capacidad de información de un canal de comunicaciones con el ancho de banda y la relación de señal a ruido. El enunciado matemático del *límite de Shannon, de capacidad de información*, es

$$I = B \log_2\left(1 + \frac{S}{N}\right) \tag{12-1b}$$

o sea

$$I = 3.32\, B \log_{10}\left(1 + \frac{S}{N}\right) \tag{12-1c}$$

en donde I = capacidad de información (bits por segundo)
B = amplitud de banda (hertz)
$\frac{S}{N}$ = relación de potencias de señal a ruido (adimensional)

Para un canal normal de comunicaciones de voz, con relación de potencias de señal a ruido de 1000 (30 dB) y amplitud de banda de 2.7 kHz, el límite de Shannon de capacidad de información es

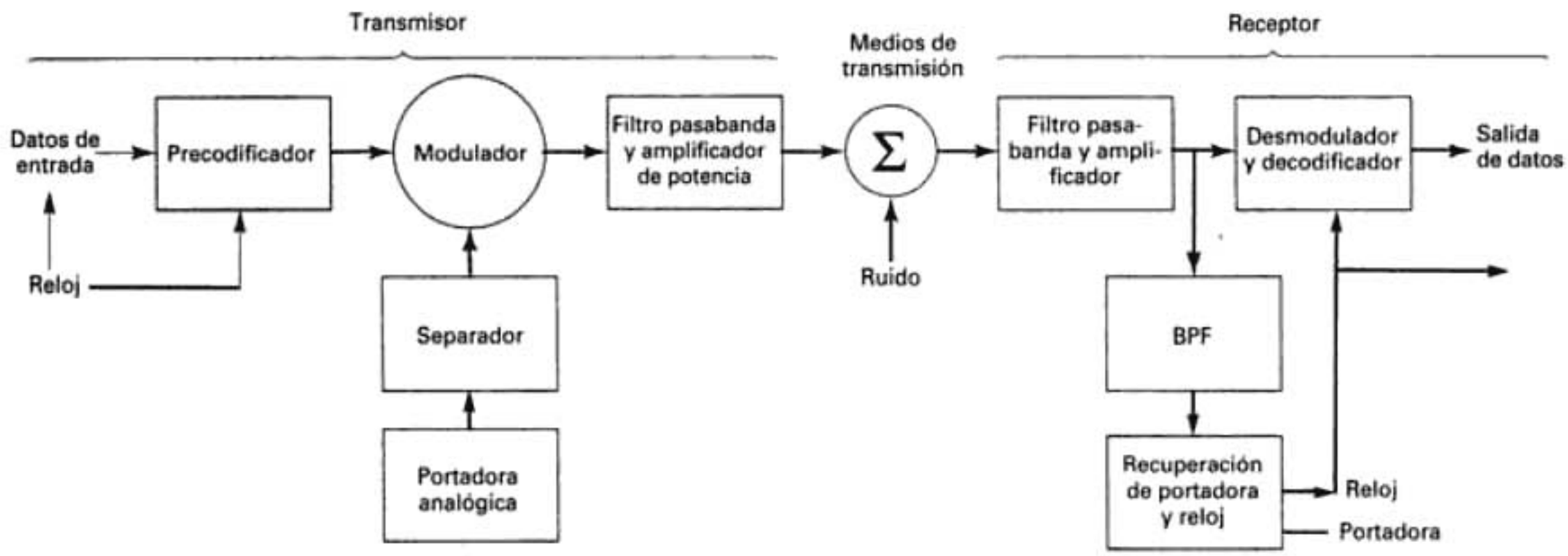


FIGURA 12-3 Diagrama de bloques simplificado de un sistema de radio digital

de salida. En los siguientes capítulos se describen algunos de esos sistemas, que incluyen técnicas de modulación tanto analógica como digital, y la transmisión de señales tanto analógicas como digitales.

RADIO DIGITAL

La propiedad que distingue un sistema de radio digital (o sistema digital de radio) de un sistema convencional de radio AM, FM o PM es la naturaleza de la señal moduladora. Los sistemas digitales y analógicos de radio usan, ambos, portadoras; sin embargo, en la modulación analógica la señal moduladora es analógica, y en la modulación digital la señal moduladora es digital. Sin embargo, téngase en cuenta que tanto en la modulación analógica como en la digital, la fuente de información original podría haber sido analógica o digital.

La fig. 12-3 muestra un diagrama de bloques simplificado de un sistema digital de radio. En el transmisor, el precodificador hace la conversión de nivel, y a continuación codifica, o agrupa, los datos que llegan en una palabra de control que modula a la portadora analógica. La portadora modulada se conforma (se filtra), se amplifica y a continuación se transmite por el medio de transmisión hasta el receptor. En el receptor, la señal de entrada se filtra, amplifica y se aplica a continuación al circuito desmodulador, que reproduce la información de la fuente original. Los circuitos de reloj y de recuperación de portadora eliminan la información de la portadora y del reloj de sincronización de la señal modulada que entra.

MODULACIÓN DIGITAL DE AMPLITUD

La técnica de modulación digital más sencilla es la *modulación digital de amplitud*, que no es más que modulación de amplitud con portadora completa y doble banda lateral. La ecuación que describe la modulación digital de amplitud mediante una señal binaria es

$$v_{am}(t) = [1 + v_m(t)]\left[\frac{A}{2}\cos(\omega_c t)\right] \tag{12-2}$$

en la que $v_{am}(t)$ = voltaje de la onda de amplitud modulada
$A/2$ = amplitud de la portadora no modulada (volts)
$v_m(t)$ = señal binaria moduladora (volts)
ω_c = frecuencia de la portadora en radianes (radianes por segundo)

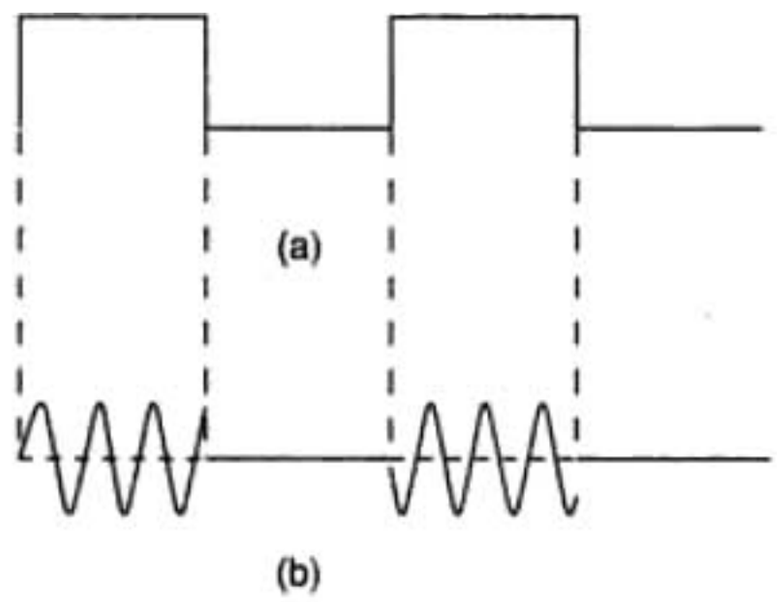


FIGURA 12-4 Modulación digital de amplitud: (a) entrada binaria; (b) forma de onda de salida OOK

En la ecuación 12-2, la señal moduladora [$v_m(t)$] es una forma de onda binaria normalizada, en la que +1 V = 1 lógico, y −1 V = 0 lógico. Por consiguiente, para una entrada de 1 lógico, $v_m(t) = +1$, y la ecuación 12-2 se reduce a

$$v_{am}(t) = [1 + 1]\left[\frac{A}{2}\cos(\omega_c t)\right]$$
$$= A\cos(\omega_c t)$$

y para una entrada de 0 lógico, $v_m(t) = -1$, y la ecuación 12-2 se reduce a

$$v_{am}(t) = [1 - 1]\left[\frac{A}{2}\cos(\omega_c t)\right]$$
$$= 0$$

Así, para 100% de modulación, $v_{am}(t)$ es $A\cos(w_c t)$ o 0. Por consiguiente, la portadora está "encendida" o "apagada", y es la causa de que a la modulación digital de amplitud se le suela llamar modulación por manipulación *encedido-apagado,* o *todo o nada* (OOK, de *on-off keying*). A veces, a la modulación por modulación de amplitud se le llama *de onda continua* (CW, de *continuous wave*) porque cuando se transmite la portadora (es decir, está encendida), tiene amplitud constante, frecuencia constante y fase constante.

La fig. 12-4 muestra las formas de onda de entrada y salida para un transmisor digital con modulación de amplitud. Una forma de onda OOK se puede desmodular en forma coherente o incoherente con poca diferencia en funcionamiento. El uso de portadoras analógicas de amplitud modulada para transportar información digital es un tipo de radio digital de relativamente baja calidad y bajo costo y, en consecuencia, rara vez se usa en sistemas de comunicaciones de gran capacidad y alta eficiencia.

MANIPULACIÓN POR DESPLAZAMIENTO DE FRECUENCIA

La *manipulación por desplazamiento de frecuencia* (FSK, de *frequency-shift keying*) es otro tipo relativamente sencillo y de baja eficiencia de modulación digital. La FSK binaria es una forma de modulación de ángulo, de amplitud constante, parecido a la modulación convencional de frecuencia (FM), pero la señal moduladora es una señal binaria que varía entre dos valores discretos de voltaje, y no es una forma de onda analógica que cambie continuamente. La ecuación general de la FSK binaria es

$$v_{fsk}(t) = V_c\cos\{2\pi[f_c + v_m(t)\Delta f]t\} \tag{12-3}$$

en donde $v_{fsk}(t)$ = forma de onda binaria FSK
V_c = amplitud de la portadora (volts)
f_c = frecuencia central de la portadora (hertz)
Δf = desviación máxima de frecuencia (hertz)
$v_m(t)$ = señal moduladora de entrada binaria (±1)

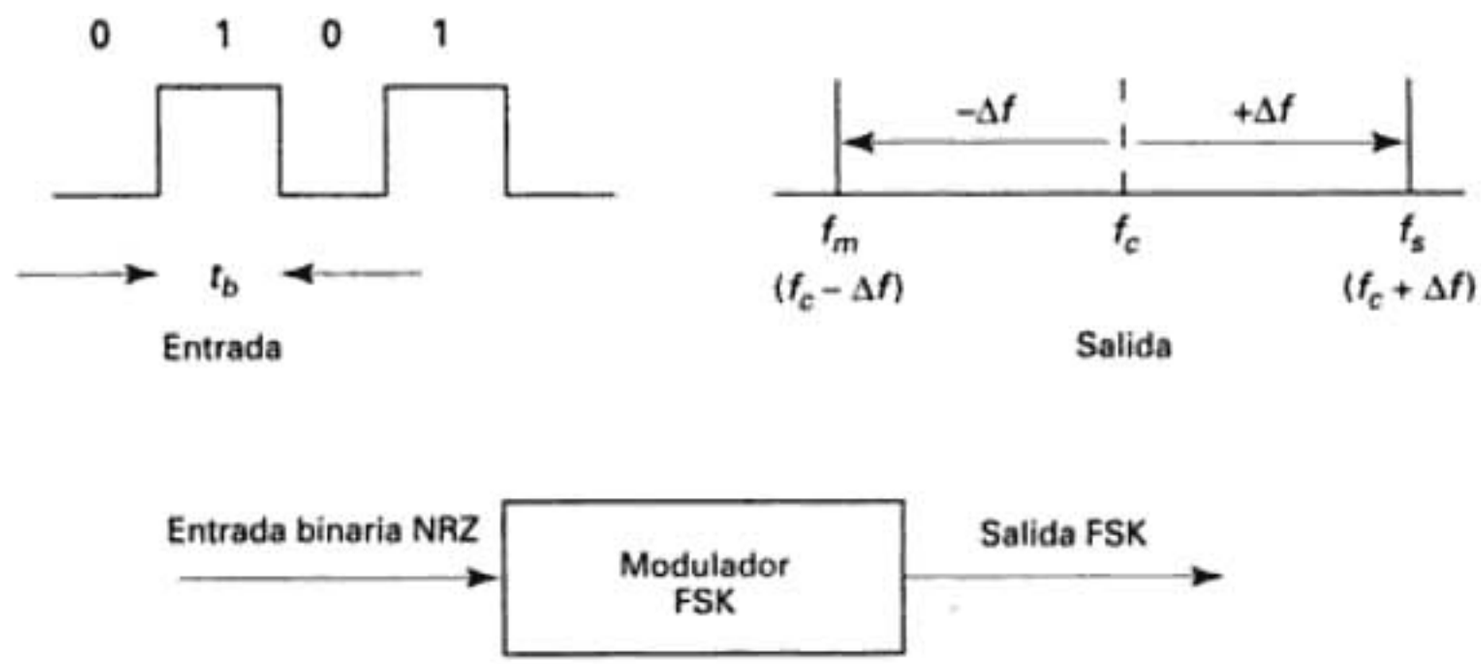


FIGURA 12-7 Desviación de frecuencia en FSK

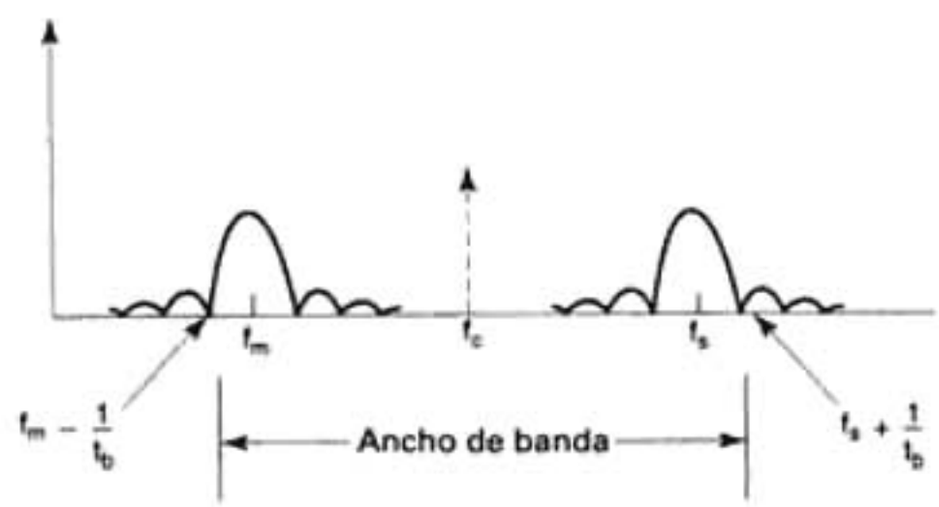


FIGURA 12-8 Espectro de frecuencias FSK

Consideraciones de FSK respecto del ancho de banda

La salida de un modulador FSK se relaciona con la entrada binaria en la forma que se ve en la fig. 12-7, donde un 0 lógico corresponde a la frecuencia de espacio f_s, un 1 lógico corresponde a la frecuencia de marca f_m, y f_c es la frecuencia de portadora. La desviación máxima de frecuencia se determina con

$$\Delta f = \frac{|f_m - f_s|}{2} \tag{12-5}$$

en la que Δf = desviación máxima de frecuencia (hertz)
f_m = frecuencia de marca (hertz)
f_s = frecuencia de espacio (hertz)

Se puede ver en la fig. 12-7 consiste en dos ondas senoidales pulsadas, de frecuencia f_m y f_s. Las ondas senoidales pulsadas tienen espectros de frecuencia que son funciones *sen x/x*. Por consiguiente, se puede representar el espectro de salida de una señal FSK como se ve en la fig. 12-8. Suponiendo que los máximos del espectro de potencia contienen la mayor parte de la energía, el ancho de banda mínimo para pasar una señal FSK se puede aproximar como sigue

$$\begin{aligned} B &= |(f_s + f_b) - (f_m - f_b)| \\ &= (|f_s - f_m|) + 2f_b \end{aligned} \tag{12-6}$$

y en vista de que $|f_s - f_m|$ es igual a $2\Delta f$, el ancho mínimo de banda aproximado es

$$\begin{aligned} B &= 2\Delta f + 2f_b \\ &= 2(\Delta f + f_b) \end{aligned} \tag{12-7}$$

en donde B = ancho mínimo de banda (hertz)
Δf = desviación mínima de frecuencias máximas (hertz)
f_m = frecuencia de marca (hertz)
f_s = frecuencia de espacio (hertz)

Ejemplo 12-2

Con una tabla de funciones de Bessel, determinar el ancho mínimo de banda para la señal FSK que se describió en el ejemplo 12-1, con una frecuencia de marca de 49 kHz, frecuencia de espacio de 51 kHz y rapidez de bits de entrada de 2 kbps.

Solución El índice de modulación se calcula sustituyendo valores en la ecuación 12-10, es decir

$$h = \frac{|49 \text{ kHz} - 51 \text{ kHz}|}{2 \text{ kbps}}$$

o

$$= \frac{2 \text{ kHz}}{2 \text{ kbps}}$$

$$= 1$$

De acuerdo con una tabla de funciones de Bessel, se producen tres conjuntos de bandas laterales significativas con un índice de modulación igual a uno. Por consiguiente, el ancho de banda se puede calcular como sigue

$$B = 2(3 \times 1000)$$

$$= 6000 \text{ Hz}$$

El ancho de banda determinado en este ejemplo, usando la tabla de funciones de Bessel, es idéntico al determinado en el ejemplo 12-1.

Receptor FSK

La desmodulación FSK es bastante sencilla, con un circuito como el de la fig. 12-9. La señal de entrada FSK se aplica en forma simultánea a las entradas de los dos filtros pasabanda (BPF) a través de un divisor de potencia. El filtro respectivo sólo pasa la frecuencia de marca o sólo la de espacio a su respectivo detector de envolvente. A su vez, los detectores de envolvente indican la potencia total en cada banda pasante, y el comparador responde a la mayor de las dos potencias. A este tipo de detección FSK se le llama detección no coherente; no hay frecuencia, implicada en el proceso de desmodulación, que esté sincronizada en fase, o en frecuencia o en ambas, con la señal FSK que entra.

La fig. 12-10 muestra el diagrama de bloques de un receptor FSK coherente. La señal FSK de entrada se multiplica por una señal de portadora recuperada que tenga exactamente la misma

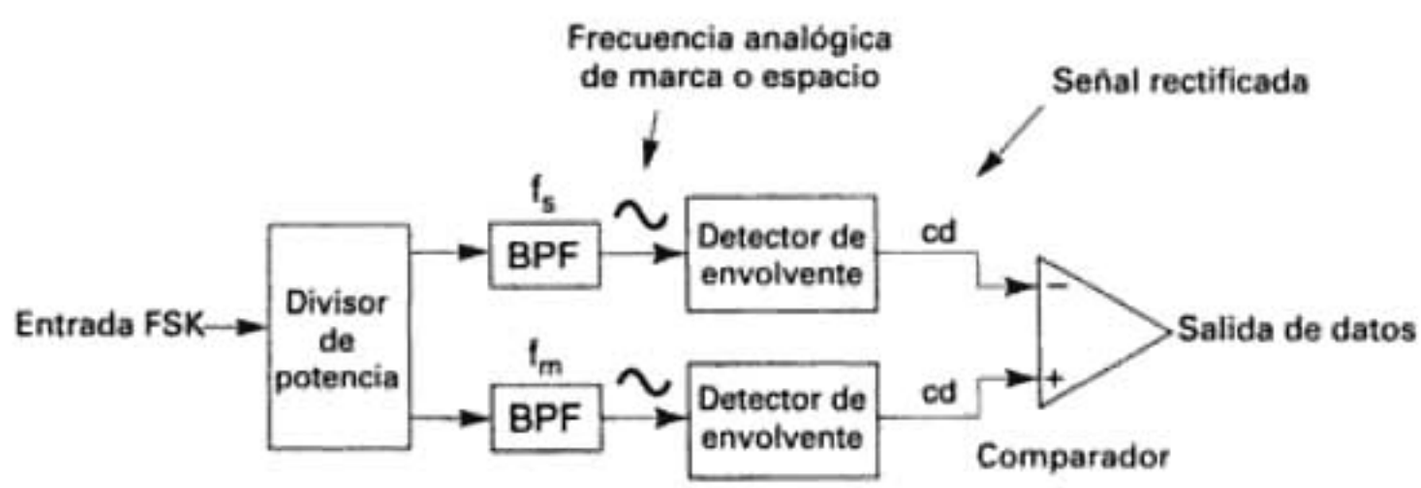


FIGURA 12-9 Desmodulador FSK no coherente

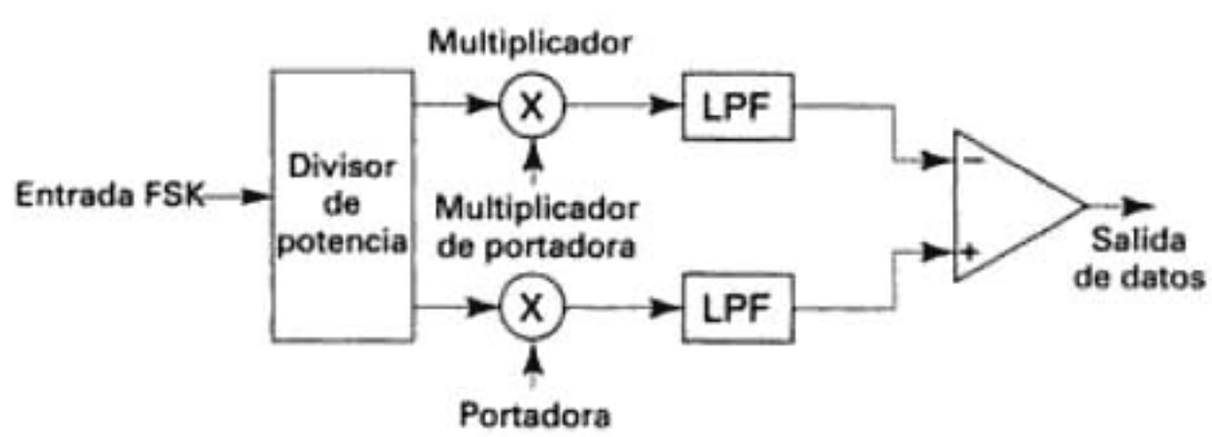


FIGURA 12-10 Desmodulador FSK coherente

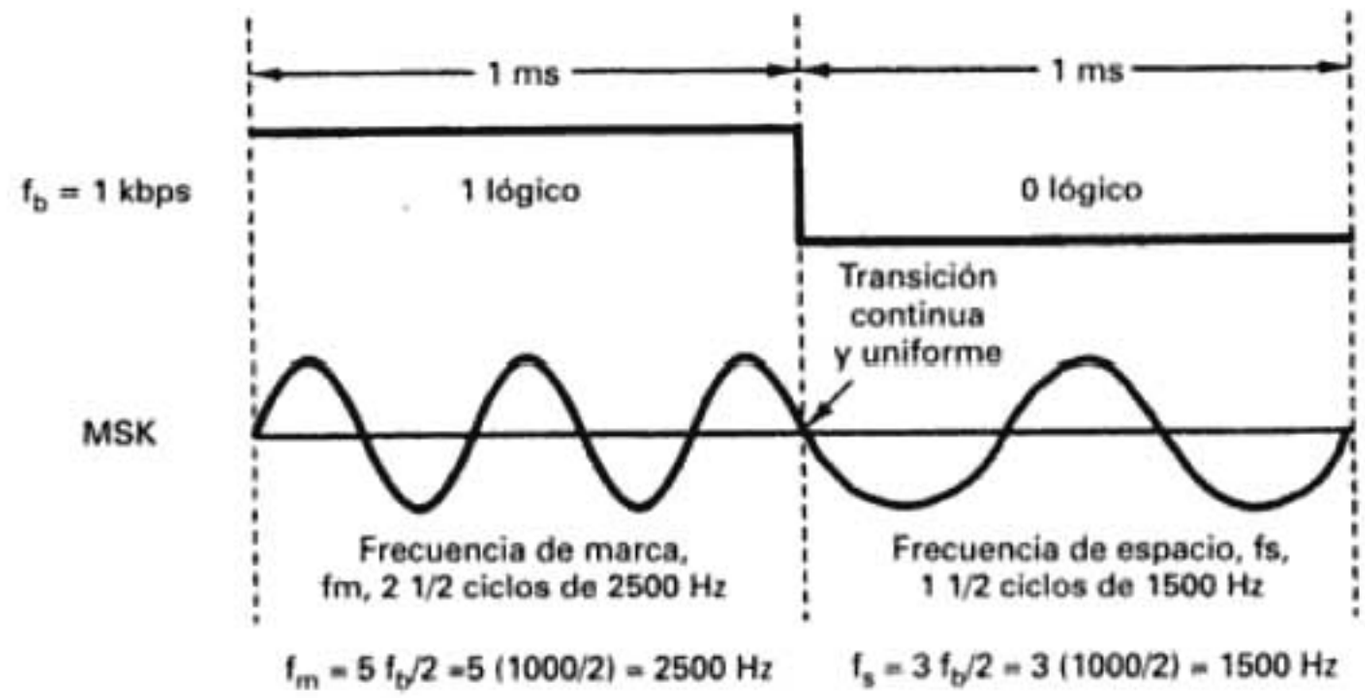


FIGURA 12-13 Forma continua de onda MSK

cambia de un 1 lógico a un 0 lógico y viceversa, hay una discontinuidad abrupta de fase en la señal analógica. Cuando eso sucede, el desmodulador tiene problemas para seguir el cambio de frecuencia y por consiguiente puede tener errores.

La fig. 12-13 muestra una forma de onda FSK de fase continua. Obsérvese que cuando cambia la frecuencia de salida, hay una transición uniforme y continua. En consecuencia, no hay discontinuidades de fase. La CP-FSK tiene mejor eficiencia de errores de bits que la FSK binaria convencional, para determinada relación de señal a ruido. La desventaja de la CP-FSK es que requiere circuitos de sincronización y, por consiguiente, su implementación es más costosa.

Si la diferencia entre las frecuencias de marca y de espacio es la mitad de la rapidez de bits (es decir, $f_s - f_m = 0.5 f_b$), el índice de modulación es $h = 0.5$. Cuando éste es el caso, hay una diferencia mínima entre las frecuencias de marca y de espacio. A esta forma particular de CP-FSK se le llama *manipulación por desplazamiento mínimo* (MSK, por *minimum-shift keying*).

MANIPULACIÓN POR DESPLAZAMIENTO DE FASE

La *manipulación por desplazamiento de fase* (PSK, por *phase-shift keying*) es otra forma de modulación digital angular de amplitud constante. Se parece a la modulación convencional de fase, excepto que en la PSK la señal de entrada es una señal digital binaria, y es posible tener una cantidad limitada de fases de salida.

Manipulación por desplazamiento binario de fase

En la *manipulación por desplazamiento binario de fase* (BPSK, de *binary phase shift keying*), son posibles dos fases de salida para una sola frecuencia portadora ("binario" significa "2"). Una fase de salida representa un 1 lógico, y la otra un 0 lógico. Cuando la señal de entrada digital cambia de estado, la fase de la portadora de salida varía entre dos ángulos que están desfasados 180°. Otros nombres del BPSK son *manipulación por inversión de fase* (PRK, de *phase reversal keying*) y *modulación en bifase*. La BPSK es una forma de modulación de onda cuadrada con portadora suprimida de una señal de onda continua (CW, de *continuous wave*).

Transmisor BPSK. La fig. 12-14 muestra un diagrama simplificado de bloques de un transmisor BPSK. El modulador balanceado funciona como conmutador de inversión de fase. Dependiendo de la condición lógica de la entrada digital, la portadora se transfiere a la salida, sea enfasada o bien desfasada 180° con referencia al oscilador de la portadora.

La fig. 12-15a muestra el esquema de un modulador de anillo balanceado. Este modulador balanceado tiene dos entradas: una portadora que está enfasada con el oscilador de referencia, y los datos digitales binarios. Para que funcione bien el modulador balanceado, el voltaje de entrada digital debe ser mucho mayor que el máximo de la portadora. Así se asegura que la entrada digital controle el estado encendido/apagado de los diodos D1 a D4. Si la entrada binaria

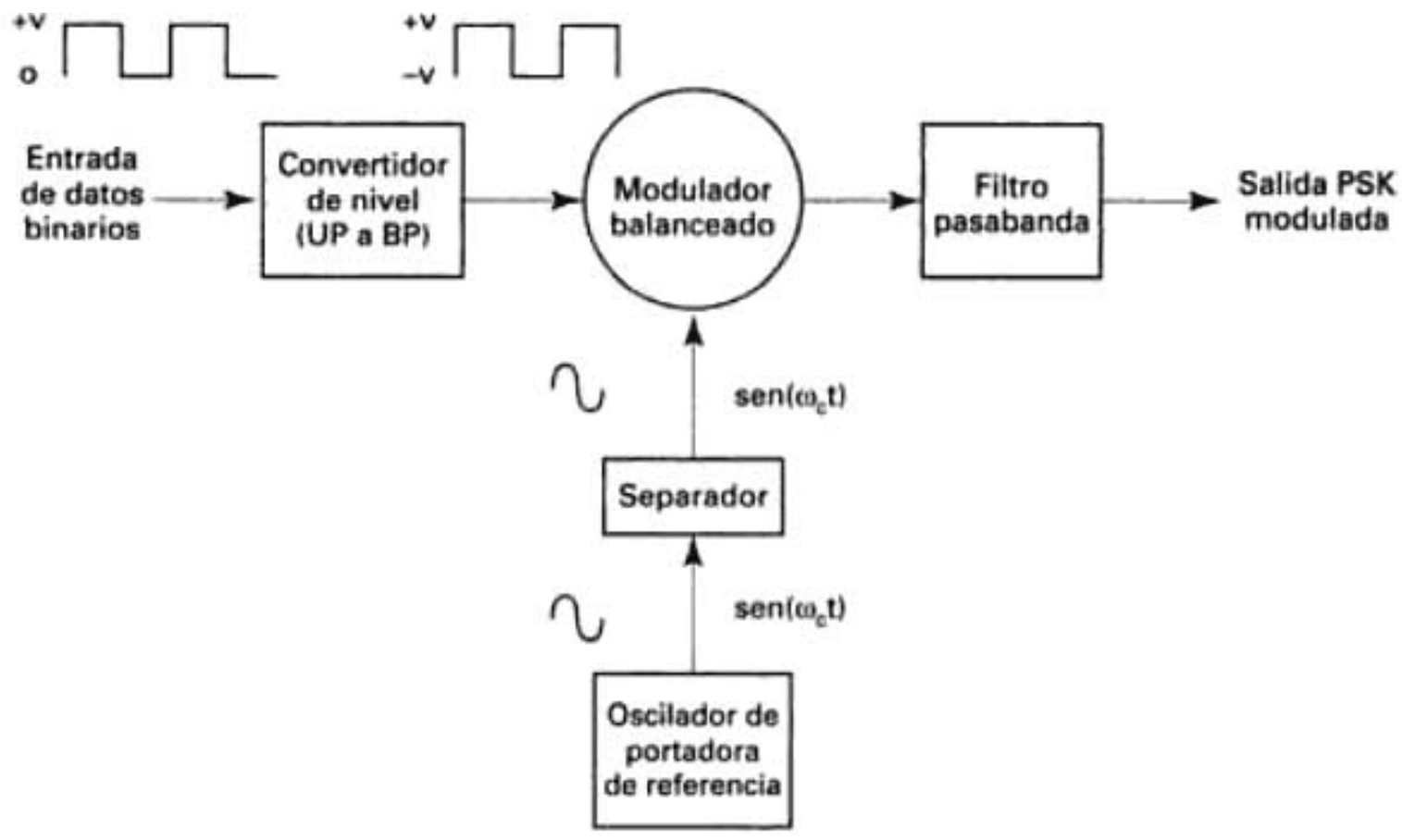


FIGURA 12-14 Transmisor BPSK

es un 1 lógico (voltaje positivo), los diodos D1 y D2 tienen polarización directa y están encendidos, mientras que los diodos D3 y D4 tienen polarización inversa y están apagados (fig. 12-15b). Con las polaridades indicadas se desarrolla el voltaje de la portadora a través del transformador T2 en fase con el voltaje de la portadora a través de T1. En consecuencia, la señal de salida está en fase con el oscilador de referencia.

Si la entrada binaria es un 0 lógico (voltaje negativo), los diodos D1 y D2 tienen polarización inversa y están apagados, mientras que los diodos D3 y D4 tienen polarización directa y están encendidos (fig. 12-15c). En consecuencia, el voltaje de la portadora se desarrolla a través del transformador T2, desfasado 180° con respecto al voltaje de portadora a través de T1. En consecuencia, la señal de salida está desfasada con respecto al oscilador de referencia. La fig. 12-16 muestra la tabla de verdad, el diagrama fasorial y el diagrama de constelación para un modulador BPSK. Un *diagrama de constelación*, que a veces se le llama *diagrama de estado-espacio de señal*, se parece a un diagrama fasorial, pero no se dibuja todo el fasor. En un diagrama de constelación sólo se muestran las posiciones relativas de los máximos de los fasores.

Consideraciones de ancho de banda en BPSK. Un modulador balanceado es un *modulador de producto*; la señal de salida es el producto de las dos señales de entrada. En un modulador BPSK, la señal portadora de entrada se multiplica por los datos binarios. Si se asigna +1 V al 1 lógico y −1 V al 0 lógico, la portadora de entrada, sen $\omega_c t$ se multiplica por +1 o por −1. En consecuencia, la señal de salida puede ser +1 sen $\omega_c t$ o −1 sen $\omega_c t$; el primer producto representa una señal que está *enfasada* con el oscilador de referencia, y el último producto, una señal que está desfasada 180° respecto al oscilador de referencia. Cada vez que cambia la condición lógica de la entrada, cambia la fase de salida. En consecuencia, para la BPSK, la tasa de cambio de la salida (baudios) es igual a la tasa de cambio de la entrada (bps), y el ancho de banda máximo de salida se presenta cuando los datos binarios en la entrada son una secuencia de 1 y 0 alternados. La frecuencia fundamental f_a de una secuencia alternativa de bits 1 y 0 es igual a la mitad de la frecuencia de bits ($f_b/2$). La ecuación de la salida de un modulador BPSK es proporcional a

$$\text{Salida BPSK} = [\text{sen}(2\pi f_a t)] \times [\text{sen}(2\pi f_c t)] \tag{12-11}$$

en donde f_a = frecuencia fundamental máxima de la entrada binaria (hertz)
f_c = frecuencia de portadora de referencia (hertz)

Al despejar el producto de dos funciones de seno de la identidad trigonométrica

$$\frac{1}{2}\cos[2\pi(f_c - f_a)t] - \frac{1}{2}\cos[2\pi(f_c + f_a)t]$$

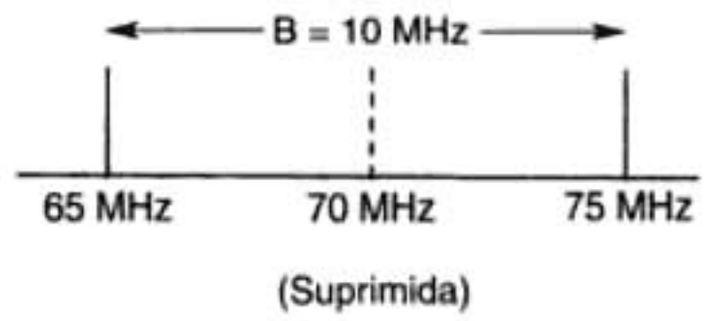
Frecuencia mínima de lado inferior (LSF)

$$LSF = 70\ MHz - 5\ MHz = 65\ MHz$$

Frecuencia máxima de lado superior (USF):

$$USF = 70\ MHz + 5\ MHz = 75\ MHz$$

Por consiguiente, el espectro de salida para las condiciones de entrada binaria en el peor de los casos es el siguiente



El ancho mínimo de banda de Nyquist, f_N, es

$$f_N = 75\ MHz - 65\ MHz = 10\ MHz$$

y los baudios $= f_b$, es decir, 10 megabaudios.

Codificación *M*-aria

M-*ario* (eme ario) es un término derivado de la palabra *binario*. *M* sólo es un dígito que representa la cantidad de condiciones o combinaciones posibles para determinada cantidad de variables binarias. Las dos técnicas de modulación digital que se han descrito hasta ahora (FSK binaria y BPSK) son sistemas binarios; codifican bits individuales y sólo hay dos condiciones posibles de salida. La FSK produce 1 lógico o frecuencia de marca, o un 0 lógico o frecuencia de espacio, y la BPSK produce una fase de 1 lógico o una fase de 0 lógico. Los sistemas FSK y BPSK son *M*-arios en los que $M = 2$.

Muchas veces conviene, en la modulación digital, codificar a un nivel mayor que el binario (que a veces se dice *más allá del binario*, o *más alto que el binario*). Por ejemplo, un sistema PSK (PSK = manipulación por desplazamiento de fase) con cuatro fases de salida posibles es un sistema *M*-ario en el que $M = 4$. Si hay ocho fases posibles de salida, $M = 8$, etcétera. La cantidad de condiciones de salida se calcula con la ecuación

$$N = \log_2 M \tag{12-12}$$

en donde N = cantidad de bits codificados
M = cantidad de condiciones posibles de salida con N bits

Por ejemplo, con FSK binaria, cada bit de entrada actúa en forma independiente sobre la portadora y produce una de las dos frecuencias posibles de salida. Así,

$$N = \log_2 2$$
$$2^N = 2$$

esto se pasa a logaritmos y se despeja N

$$\log 2^N = \log 2$$
$$N \log 2 = \log 2$$
$$N = \frac{\log 2}{\log 2}$$
$$N = 1$$

Con la BPSK, cada bit de entrada también actúa en forma independiente sobre la portadora y, en consecuencia, $N = 1$.

Si entran dos bits codificados juntos, y a continuación se dejan modificar la portadora en forma simultánea, la cantidad de condiciones de salida es

$$M = 2^2$$
$$= 4$$

o sea $$\text{sen}^2\,\omega_c t = \frac{1}{2}(1 - \cos 2\omega_c t) = \frac{1}{2} - \frac{1}{2}\underbrace{\cos 2\,\omega_c t}_{\text{(se filtra y se elimina)}}$$

y solo queda $$\text{salida} = +\frac{1}{2}\text{V} = 1 \text{ lógico}$$

Se puede ver que la salida del modulador balanceado contiene un voltaje positivo, [[+(1/2)V] y una onda cosenoide del doble de la frecuencia de la portadora, $2\omega_c$. El filtro de pasabajas tiene una frecuencia de corte mucho menor que $2\omega_c$, y por ello bloquea la segunda armónica de la portadora y sólo pasa el componente constante positivo. Un voltaje positivo representa un 1 lógico desmodulado.

Para una señal de entrada BPSK de $-\text{sen}\,\omega_c t$ (un 0 lógico), la salida del modulador balanceado es

$$\text{salida} = (-\text{sen}\,\omega_c t)(\text{sen}\,\omega_c t) = -\text{sen}^2\,\omega_c t$$

o sea $$-\text{sen}^2\,\omega_c t = -\frac{1}{2}(1 - \cos 2\omega_c t) = -\frac{1}{2} + \frac{1}{2}\underbrace{\cos 2\,\omega_c t}_{\text{(se filtra y se elimina)}}$$

y queda $$\text{salida} = -\frac{1}{2}\text{V} = 0 \text{ lógico}$$

La salida del modulador balanceado contiene un voltaje negativo, $-(1/2)$V, y una onda cosenoide del doble de la frecuencia de la portadora, es decir, $2\,\omega_c$. De nuevo, el filtro pasabajas bloquea la segunda armónica de la portadora y sólo pasa el componente constante negativo. Un voltaje negativo representa un 0 lógico desmodulado.

Manipulación por desplazamiento cuaternario de fase

La *manipulación por desplazamiento cuaternario de fase* (QPSK, por *quaternary phase shift keying*), o *PSK de cuadratura* como a veces se le llama, es otra forma de modulación digital angular y de amplitud constante. La QPSK es una técnica M-aria de codificación en la que $M = 4$ (de ahí el nombre "cuaternario", que significa "4"). Con esta codificación, son posibles cuatro fases de salida para una sola frecuencia de portadora. Como hay cuatro fases distintas de salida, debe haber cuatro condiciones distintas de entrada. Ya que la entrada digital a un modulador PQSK es una señal binaria (base 2), para producir cuatro condiciones distintas de entrada se necesita más de un solo bit de entrada. Con dos bits hay cuatro condiciones posibles: 00, 01, 10 y 11. En consecuencia, en la QPSK, los datos binarios de entrada se combinan en grupos de dos bits, llamados *dibits*. Cada dibit de código genera una de las cuatro fases posibles de salida. Así, para cada dibit de dos bits sincronizado en el modulador, se obtiene un solo cambio en la salida. Entonces, la rapidez de cambio en la salida (baudios) es la mitad de la rapidez de entrada de bits.

Transmisor QPSK. En la fig. 12-19 se muestra un diagrama de bloques de un modulador QPSK. Dos bits (un dibit) se sincroniza en el divisor de bits. Después de que han entrado ambos bits en serie, salen en forma simultánea y en paralelo. Un bit se dirige al canal I y el otro al canal Q. El bit I modula una portadora enfasada con el oscilador de referencia ("I" proviene de canal "*in phase*", enfasado), y el bit Q modula una portadora que está desfasada 90°, o en cuadratura ("Q" de "*quadrature*") respecto a la portadora de referencia.

Se puede ver que una vez que un dibit se ha dividido en los canales I y Q, la operación es igual que en un modulador BPSK. En esencia, un modulador QPSK es dos moduladores BPSK en paralelo. Otra vez, para un 1 lógico = +1 V y un 0 lógico = −1 V, son posibles dos fases en la salida del modulador balanceado I: $+\text{sen}\,\omega_c t$ y $-\text{sen}\,\omega_c t$, y son posibles dos fases en la salida del modulador balanceado Q: $+\cos\omega_c t$ y $-\cos\omega_c t$. Cuando el sumador lineal combina las dos señales en cuadratura, es decir, desfasadas 90°, hay cuatro fasores resultantes posibles, definidos por las siguientes expresiones: 1) $+\text{sen}\,\omega_c t + \cos\omega_c t$, 2) $+\text{sen}\,\omega_c t - \cos\omega_c t$, 3) $-\text{sen}\,\omega_c t + \cos\omega_c t$, y 4) $-\text{sen}\,\omega_c t - \cos\omega_c t$.

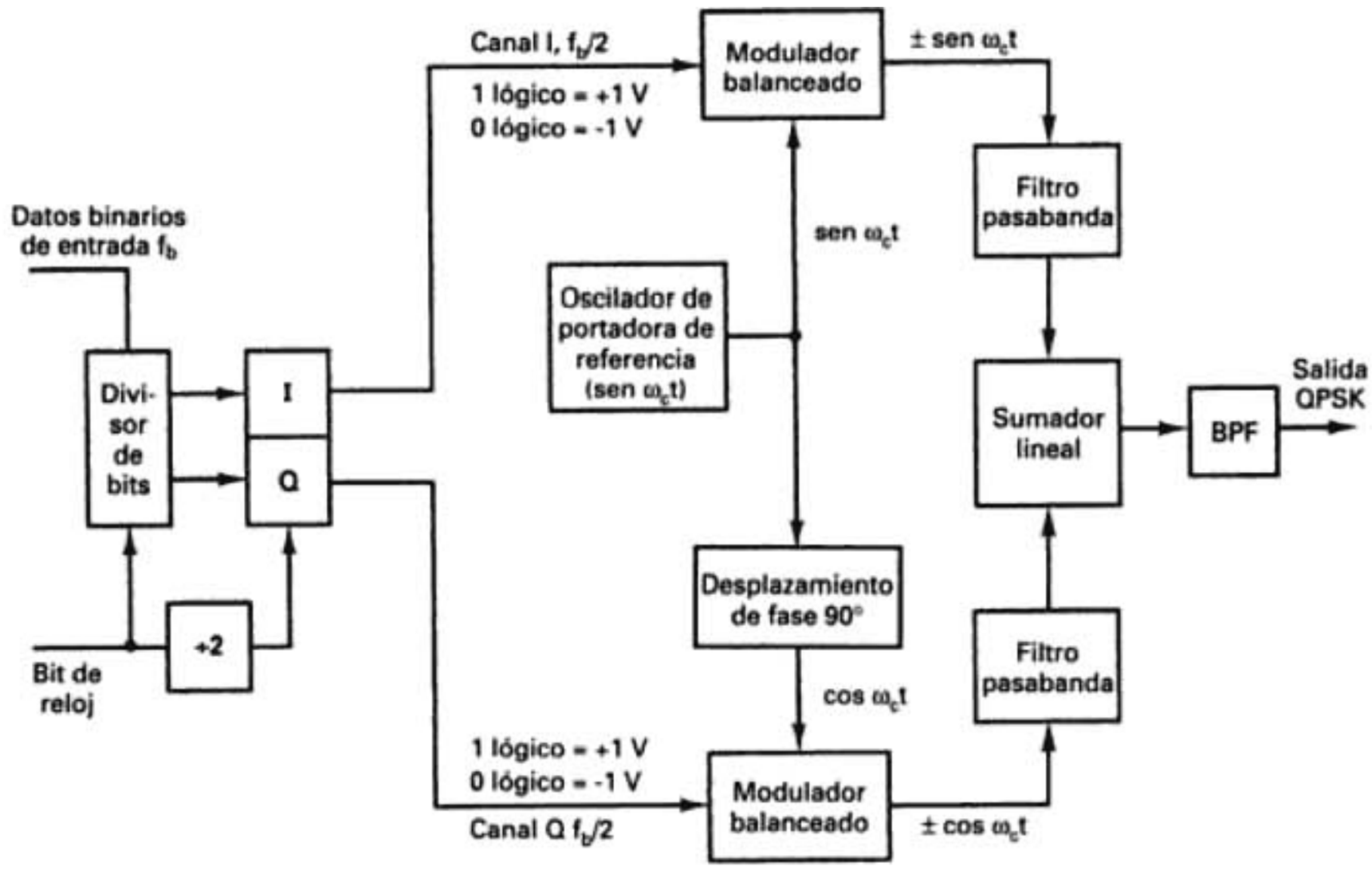


FIGURA 12-19 Modulador QPSK

Ejemplo 12-4

Determinar la tabla de verdad, el diagrama fasorial y el diagrama de constelación, para el modulador QPSK de la fig. 12-19.

Solución Para una entrada binaria de datos de Q = 0 e I = 0, las dos entradas al modulador balanceado I son −1 y sen $\omega_c t$, y las dos entradas al modulador balanceado Q son −1 y cos $\omega_c t$. En consecuencia, las salidas son

$$\text{Modulador balanceado I} = (-1)(\text{sen } \omega_c t) = -1 \text{ sen } \omega_c t$$

$$\text{Modulador balanceado Q} = (-1)(\cos \omega_c t) = -1 \cos \omega_c t$$

y la salida del sumador lineal es

$$-1 \cos \omega_c t - 1 \text{ sen } \omega_c t = 1.414 \text{ sen}(\omega_c t - 135°)$$

Para los códigos restantes de dibit (01, 10 y 11), el procedimiento es igual. Los resultados se ven en la fig.12-20a.

En las figuras 12-20b y c, se ve que con la QPSK cada uno de los cuatro fasores posibles de salida tiene exactamente la misma amplitud. Así, la información binaria se debe codificar por completo en la fase de la señal de salida. Esta característica de amplitud constante es la más importante de la PSK, que la distingue de la QAM que se explicará más adelante en este capítulo. También, se ve en la fig. 12-20b que la separación angular entre dos fasores adyacentes cualesquiera en la QPSK es 90°. En consecuencia, una señal QPSK puede sufrir cuando mucho un desplazamiento de fase de +45° o −45° durante la transmisión y retener en forma correcta la información codificada, cuando se desmodule en el receptor. La fig. 12-21 muestra la relación de fase de salida en función del tiempo para un modulador QPSK.

Consideraciones de ancho de banda con QPSK. En la QPSK, como los datos de entrada se dividen en dos canales, la rapidez de bits en cualquiera de los canales I o Q es igual a la mitad de la rapidez de entrada de bits, $f_b/2$. En esencia, el divisor de bits estira los bits I y Q al doble de su longitud en la entrada. En consecuencia, la mayor frecuencia fundamental presente en la entrada de datos al modulador balanceado I o Q es igual a la cuarta parte de la rapidez de entrada de datos (la mitad de $f_b/2$ es $f_b/4$). En consecuencia, la salida de los moduladores balanceados I y Q requiere un ancho de banda bilateral de Nyquist mínimo igual a la mitad de la

Entrada binaria		Fase de salida QPSK
Q	I	
0	0	−135°
0	1	−45°
1	0	+135°
1	1	+45°

(a)

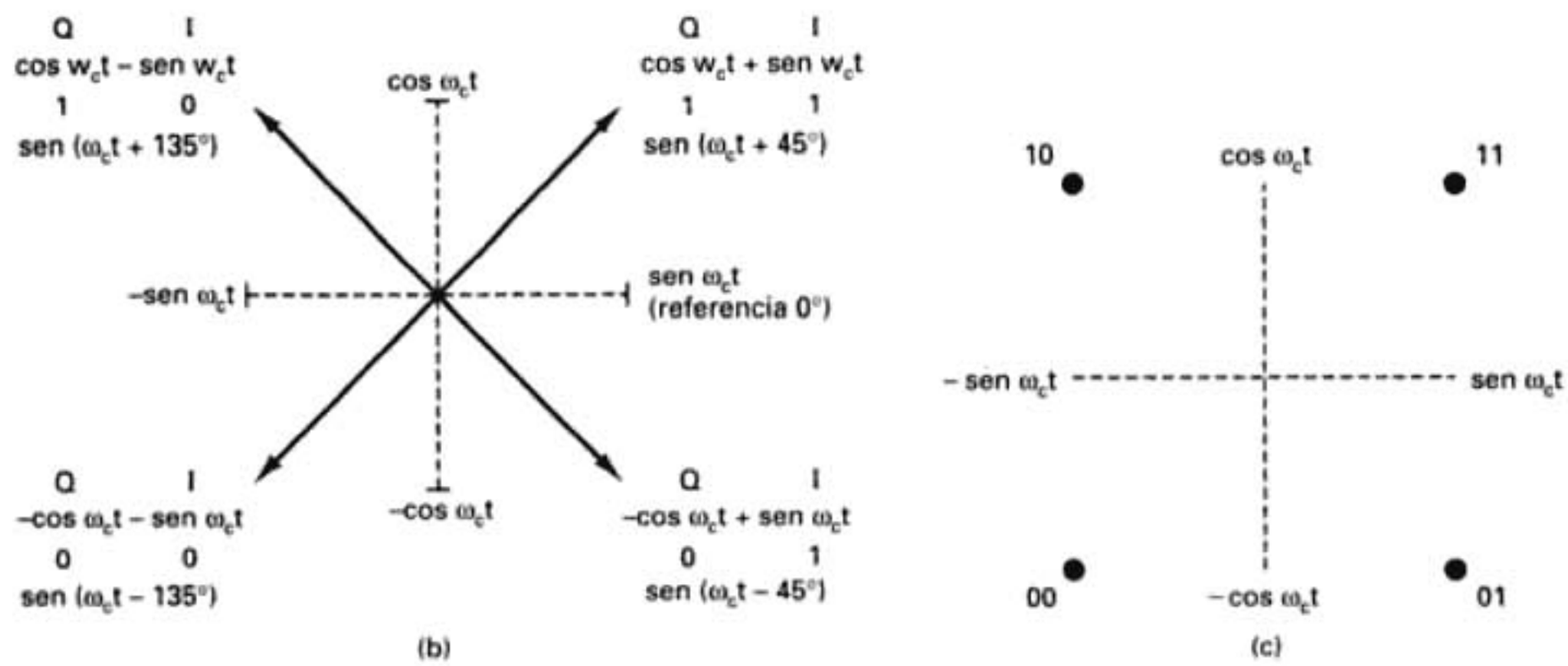


FIGURA 12-20 Modulador QPSK: (a) tabla de verdad; (b) diagrama fasorial; (c) diagrama de constelación

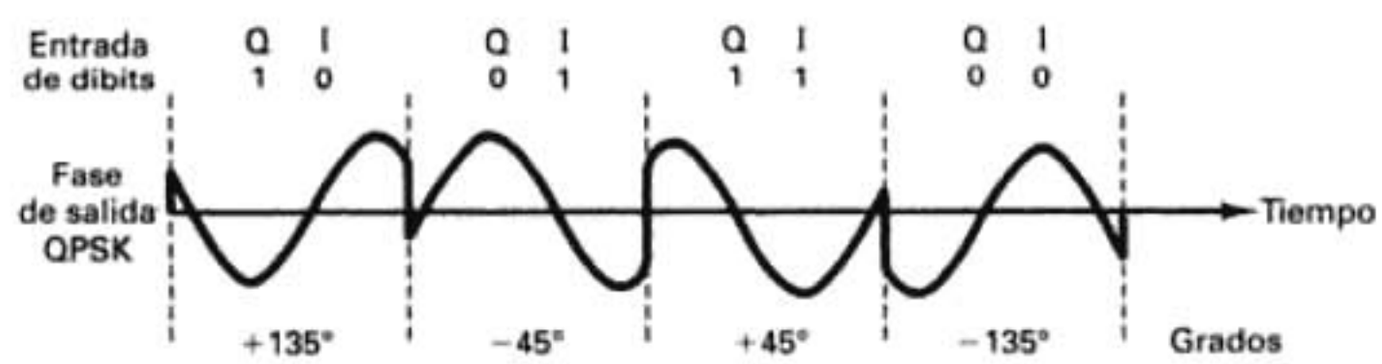


FIGURA 12-21 Fase de salida en función de tiempo para un modulador QPSK

frecuencia de bits que entran: f_N = el doble de $f_b/4 = f_b/2$. Así, con la QPSK se obtiene una compresión de ancho de banda: el ancho mínimo de banda es menor que la rapidez de bits que llegan. También, como la señal de salida QPSK no cambia de fase sino hasta que han entrado dos bits (un dibit) al divisor de bits, la rapidez máxima de cambio (los baudios) en la salida también es igual a la mitad de la rapidez de bits en la entrada. Como en el caso de la BPSK, el ancho mínimo de banda y los baudios son iguales. Esta relación se ve en la fig. 12-22.

La fig. 12-22 muestra que la condición del peor de los casos en la entrada al modulador balanceado I o Q es una secuencia alternativa de 1 y 0, que se presenta cuando los datos binarios de entrada tienen una pauta repetitiva 1100. Un ciclo de la transición binaria más rápida (una secuencia de 1/0) en el canal I o Q tarda igual que cuatro bits de datos de entrada. Entonces, la frecuencia fundamental máxima en la entrada, y la rapidez máxima de cambio en la salida de los moduladores balanceados es igual a la cuarta parte de la rapidez de entrada binaria.

La salida de los moduladores balanceados se puede representar con la siguiente ecuación

$$\text{salida} = (\text{sen } \omega_a t)(\text{sen } \omega_c t)$$

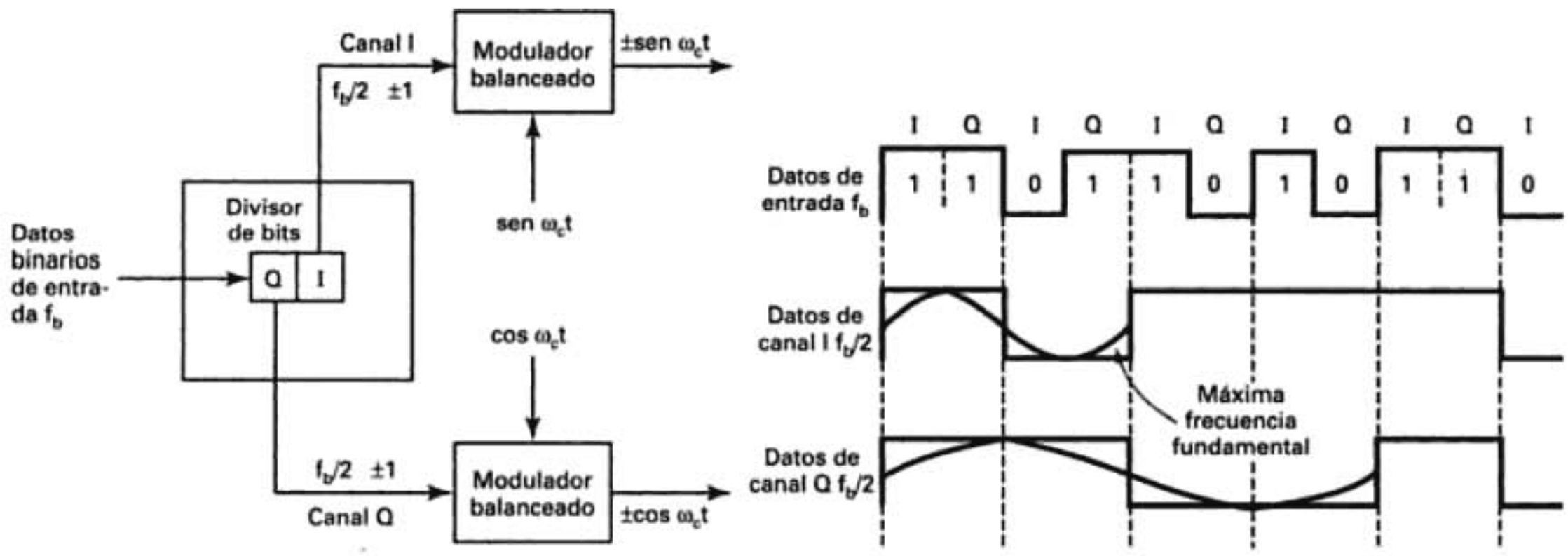


FIGURA 12-22 Consideraciones de ancho de banda en un modulador QPSK

donde
$$\underbrace{\omega_a t = 2\pi \frac{f_b}{4} t}_{\text{fase moduladora}} \quad \text{y} \quad \underbrace{\omega_c t = 2\pi f_c t}_{\text{fase de portadora no modulada}}$$

Así,
$$\text{salida} = \left(\text{sen } 2\pi \frac{f_b}{4} t\right)(\text{sen } 2\pi f_c t)$$

$$\frac{1}{2}\cos 2\pi\left(f_c - \frac{f_b}{4}\right)t - \frac{1}{2}\cos 2\pi\left(f_c + \frac{f_b}{4}\right)t$$

El espectro de frecuencias de salida va desde $f_c + f_b/4$ hasta $f_c - f_b/4$, y el ancho mínimo de banda, f_N, es

$$\left(f_c + \frac{f_b}{4}\right) - \left(f_c - \frac{f_b}{4}\right) = \frac{2f_b}{4} = \frac{f_b}{2}$$

Ejemplo 12-5

Calcular el ancho mínimo bilateral de Nyquist, f_N y los baudios, para un modulador QPSK con rapidez de entrada de datos f_b = 10 Mbps y una frecuencia de portadora de 70 MHz. También comparar los resultados con los obtenidos con el modulador BPSK del ejemplo 12-3. Usar el diagrama de bloques de QPSK de la fig. 12-19 como modelo de modulador.

Solución La rapidez de bits en los canales I y Q es igual a la mitad de la rapidez de transmisión de bits, es decir

$$f_{bQ} = f_{bI} = \frac{f_b}{2} = \frac{10 \text{ Mbps}}{2} = 5 \text{ Mbps}$$

La frecuencia fundamental máxima que se presenta a cualquiera de los moduladores balanceados es

$$f_a = \frac{f_{bQ}}{2} \text{ o } \frac{f_{bI}}{2} = \frac{5 \text{ Mbps}}{2} = 2.5 \text{ MHz}$$

La onda de salida de cada modulador balanceado es

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

$$\frac{1}{2}\cos 2\pi(f_c - f_a)t - \frac{1}{2}\cos 2\pi(f_c + f_a)t$$

$$\frac{1}{2}\cos 2\pi[(70 - 2.5) \text{ MHz}]t - \frac{1}{2}\cos 2\pi[(70 + 2.5) \text{ MHz}]t$$

$$\frac{1}{2}\cos 2\pi(67.5 \text{ MHz})t - \frac{1}{2}\cos 2\pi(72.5 \text{ MHz})t$$

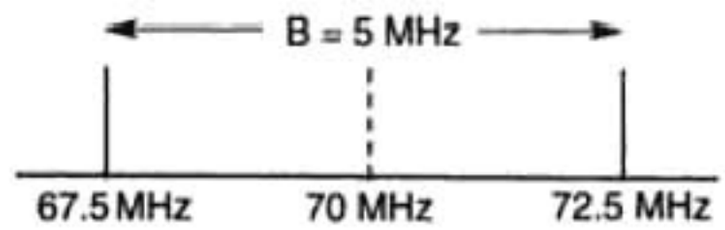
El ancho mínimo de banda de Nyquist es

$$f_N = (72.5 - 67.5)\ \text{MHz} = 5\ \text{MHz}$$

La rapidez de símbolos es igual al ancho de banda, y entonces

$$\text{rapidez de símbolos} = 5\ \text{megabaudios}$$

El espectro de salida es el siguiente



$$f_N = 5\ \text{MHz}$$

Se puede ver que para la misma rapidez de entrada de bits, el ancho mínimo de banda necesario para pasar la salida del modulador QPSK es igual a la mitad del que requiere el modulador BPSK del ejemplo 12-3. También, la rapidez de baudios para el modulador QPSK es la mitad que para el modulador BPSK.

El ancho mínimo de banda para el sistema QPSK del ejemplo 12-5 también se puede determinar sólo con sustituir en la ecuación 12-13b

$$B = \frac{10\ \text{Mbps}}{2}$$
$$= 5\ \text{MHz}$$

Receptor QPSK. En la fig. 12-23 aparece el diagrama de bloques de un receptor QPSK. El divisor de potencia dirige la señal de entrada QPSK a los detectores de producto I y Q, y al circuito de recuperación de portadora. Este último circuito reproduce la señal original del oscilador de portadora. La portadora recuperada debe ser coherente, en frecuencia y en fase, con la portadora de referencia de transmisión. La señal QPSK se desmodula en los detectores de producto I y Q, que generan los bits originales de datos I y Q. Las salidas de los detectores de productos se alimentan al circuito combinador de bits, donde se convierten de canales I y Q paralelos a una sola corriente de salida de datos binarios.

La señal QPSK que entra puede ser cualquiera de las cuatro fases posibles de salida que se muestran en la fig. 12-20. Para ilustrar el proceso de desmodulación, sea la señal QPSK de entrada $-\text{sen}\ \omega_c t + \cos\ \omega_c t$. La descripción matemática del modelo de desmodulación es la siguiente:

La señal QPSK de recepción, $-\text{sen}\ \omega_c t + \cos\ \omega_c t$, es una de las entradas al detector de producto I. La otra entrada es la portadora recuperada, $\text{sen}\ \omega_c t$. La salida del detector de producto I es

$$\text{I} = \underbrace{(-\text{sen}\ \omega_c t + \cos\ \omega_c t)}_{\text{señal de entrada QPSK}}\ \underbrace{(\text{sen}\ \omega_c t)}_{\text{portadora}}$$

$$= (-\text{sen}\ \omega_c t)(\text{sen}\ \omega_c t) + (\cos\ \omega_c t)(\text{sen}\ \omega_c t)$$

$$= -\text{sen}^2\ \omega_c t + (\cos\ \omega_c t)(\text{sen}\ \omega_c t)$$

$$= -\frac{1}{2}(1 - \cos 2\omega_c t) + \frac{1}{2}\text{sen}\ (\omega_c + \omega_c)t + \frac{1}{2}\text{sen}(\omega_c - \omega_c)t$$

(se filtra y se elimina) (igual a 0)

$$\text{I} = -\frac{1}{2} + \frac{1}{2}\cos 2\omega_c t + \frac{1}{2}\text{sen}\ 2\omega_c t + \frac{1}{2}\text{sen}\ 0$$

$$= -\frac{1}{2}\ \text{V (lógico 0)}$$

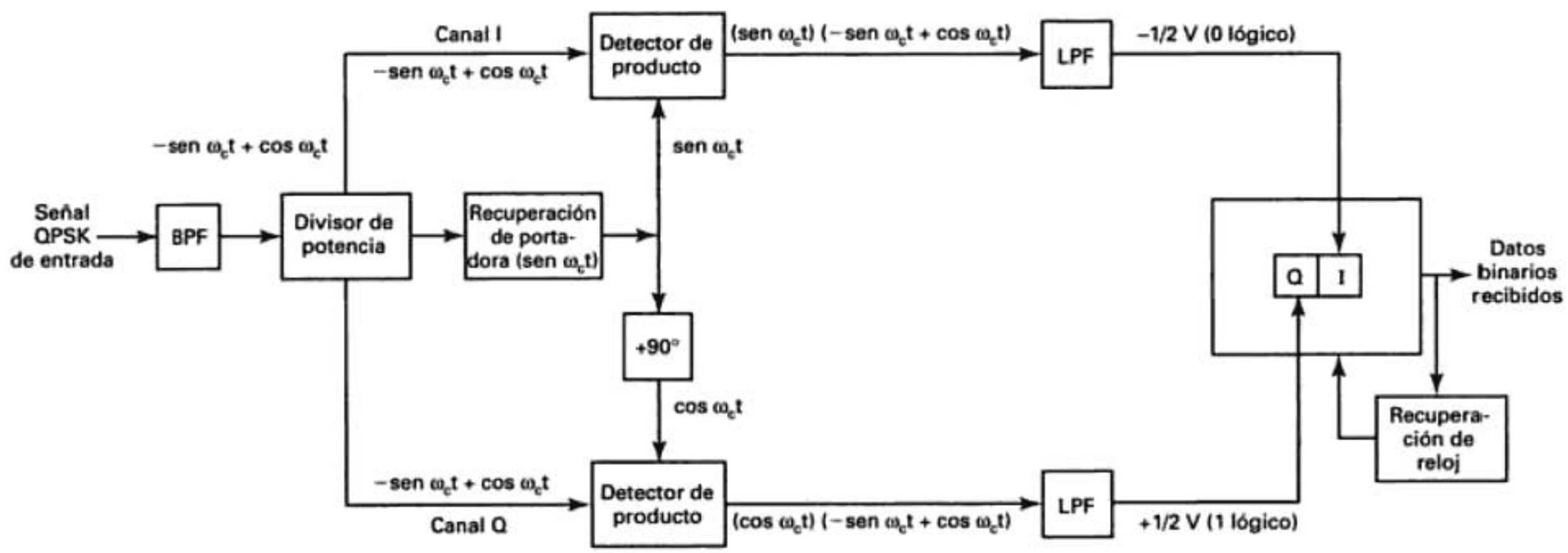


FIGURA 12-23 Receptor QPSK

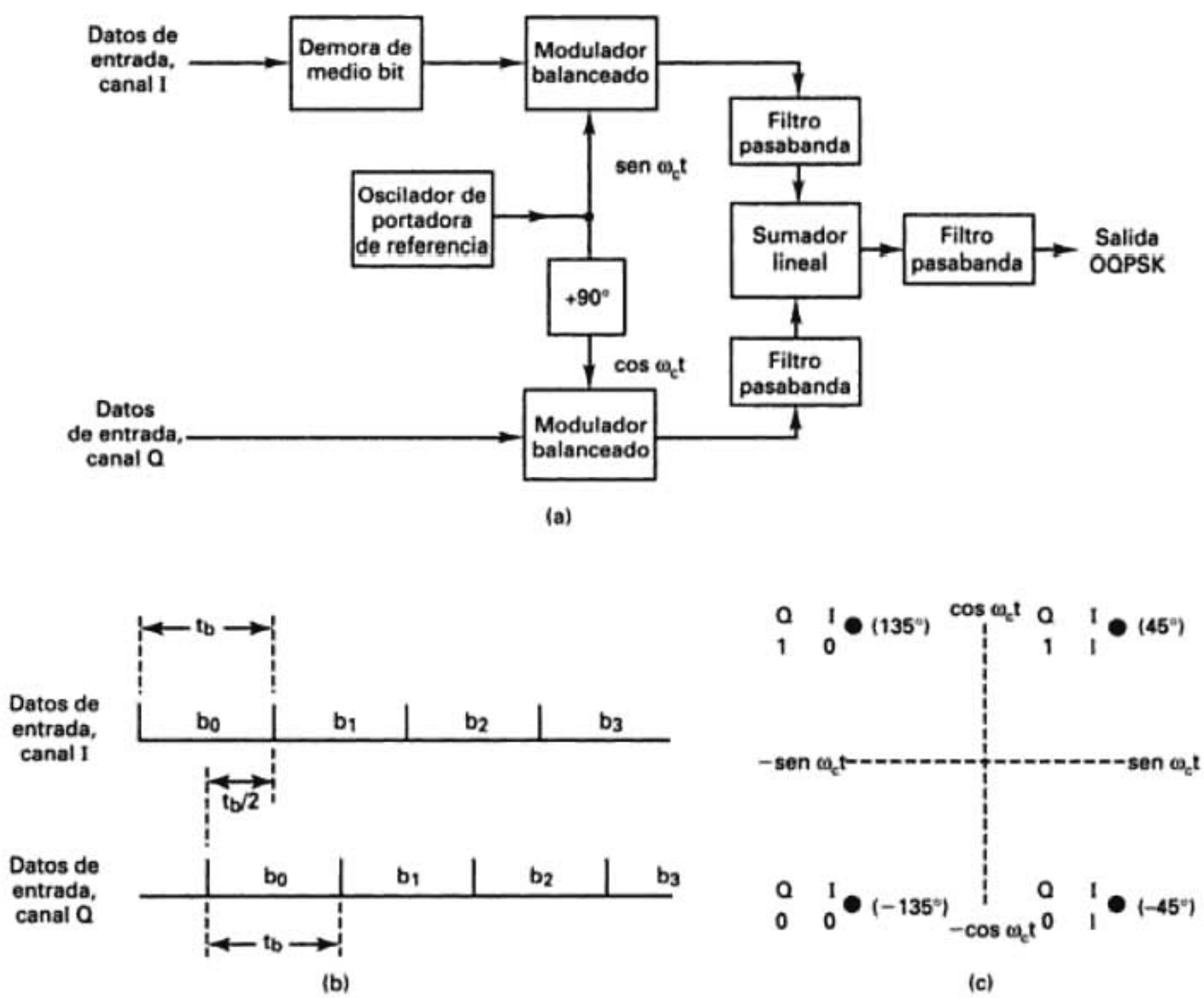


FIGURA 12-24 Manipulación QPSK compensada (OQPSK): (a) diagrama de bloques; (b) alineación de bits; (c) diagrama de constelación

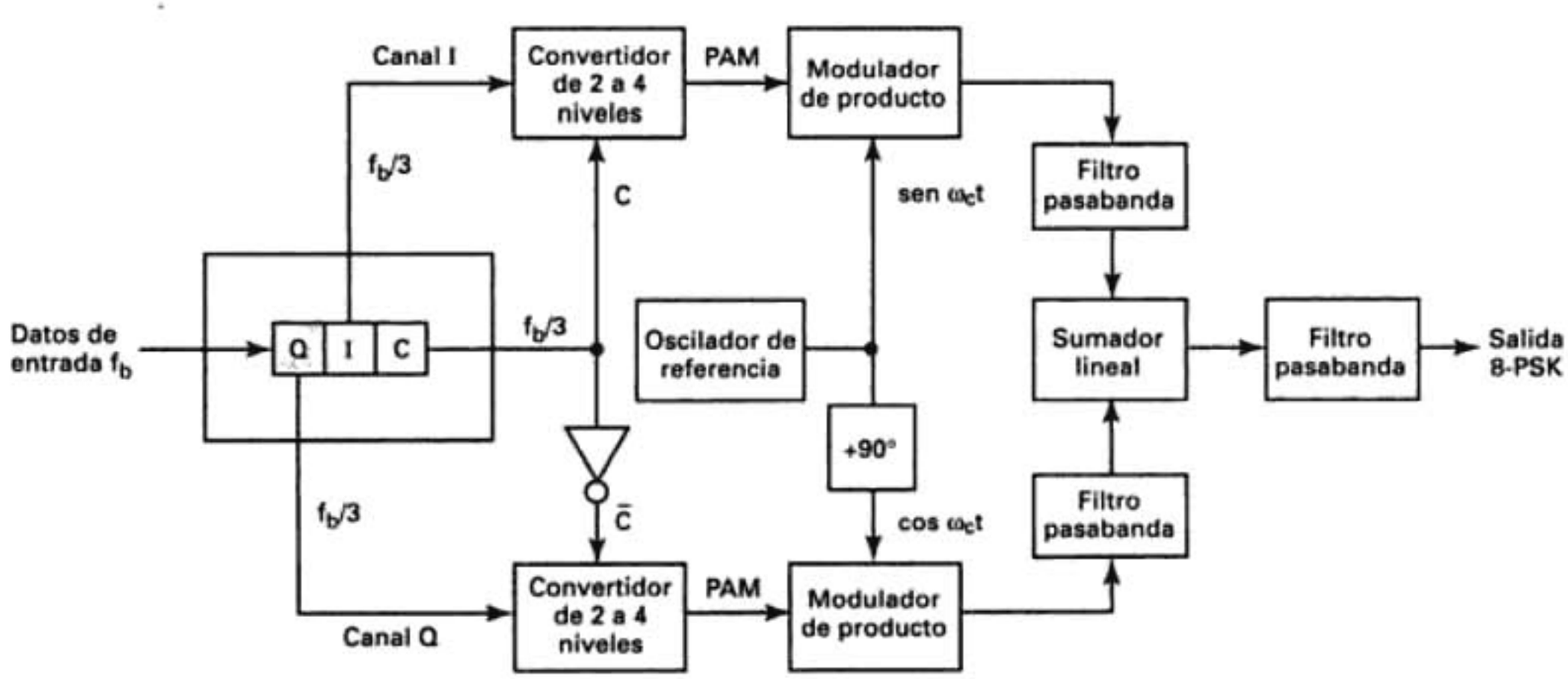
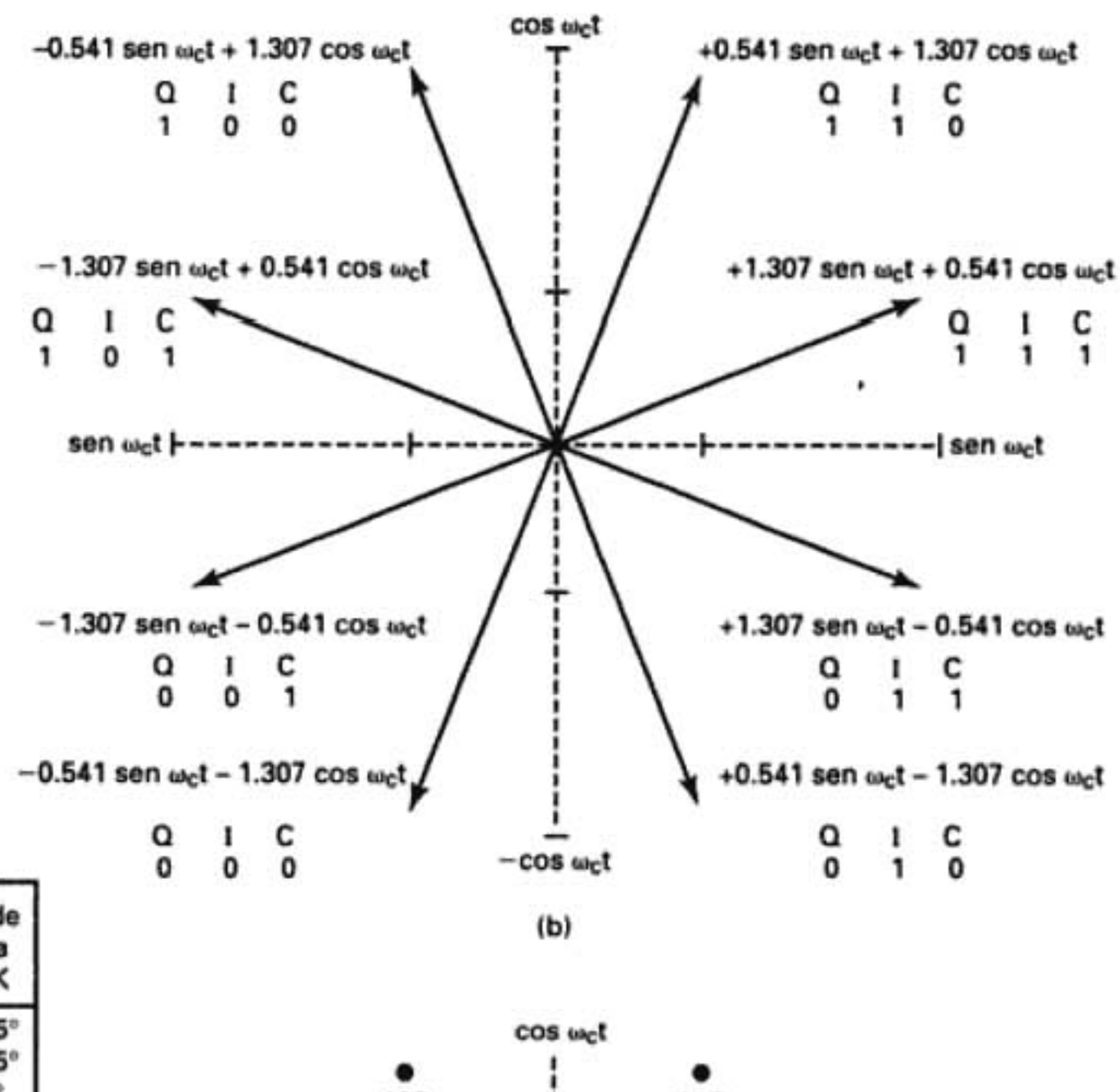


FIGURA 12-25 Modulador 8-PSK



Entrada binaria			Fase de salida 8-PSK
Q	I	C	
0	0	0	−112.5°
0	0	1	−157.5°
0	1	0	−67.5°
0	1	1	−22.5°
1	0	0	+112.5°
1	0	1	+157.5°
1	1	0	+67.5°
1	1	1	+22.5°

(a)

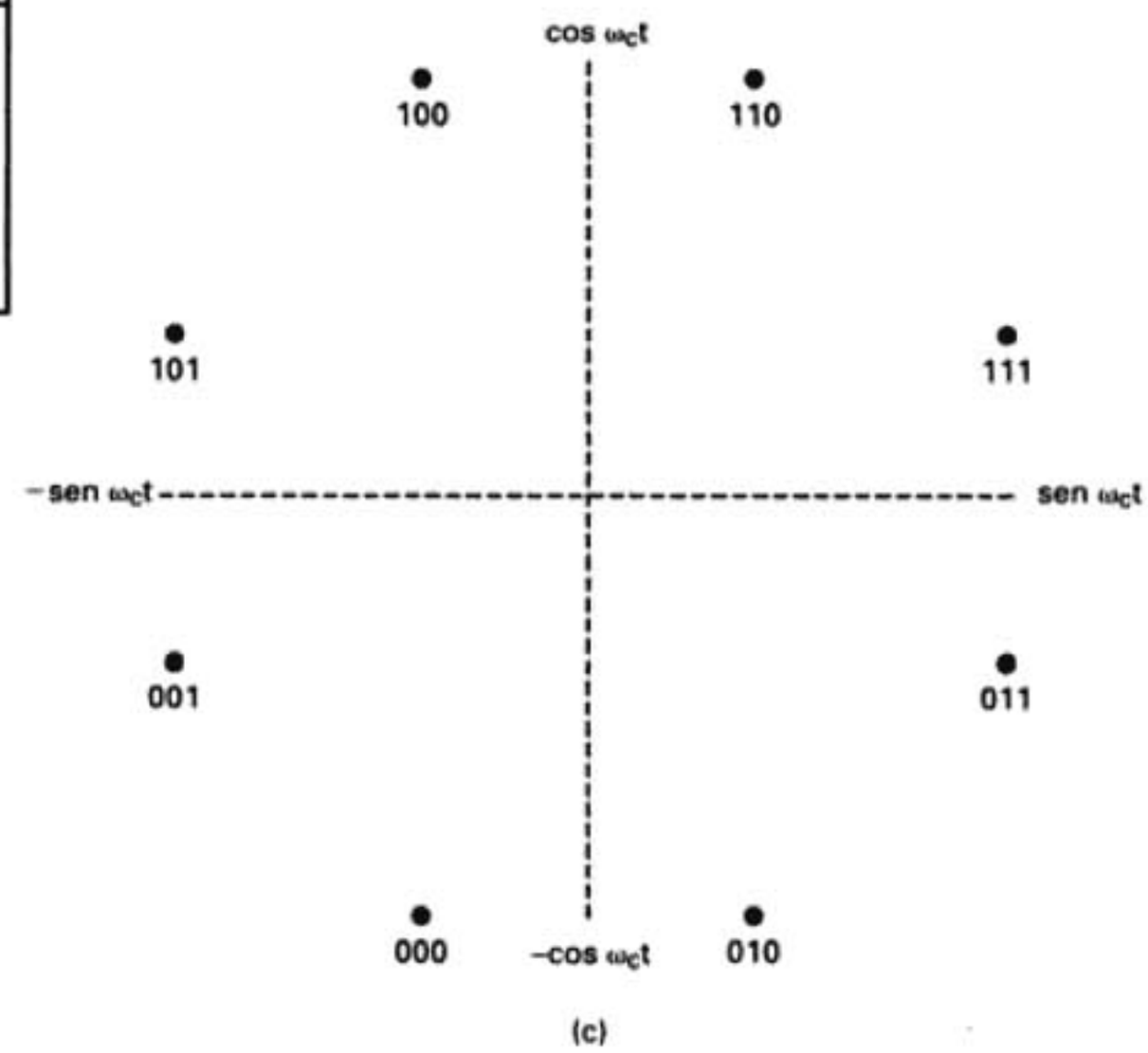


FIGURA 12-27 Modulador 8-PSK: (a) tabla de verdad; (b) diagrama fasorial; (c) diagrama de constelación

Consideraciones de ancho de banda con 8-PSK. En la 8-PSK, como los datos están divididos en tres canales, la rapidez de bits en el canal I, Q o C es igual a un tercio de la rapidez de entrada de datos binarios, $f_b/3$. El divisor de bits estira los bits I, Q y C hasta el triple de su longitud en la entrada. Como los bits I, Q y C salen en forma simultánea y en paralelo, los convertidores de 2 a 4 niveles también ven un cambio en sus entradas y, en consecuencia, en sus salidas, con una rapidez igual a $f_b/3$.

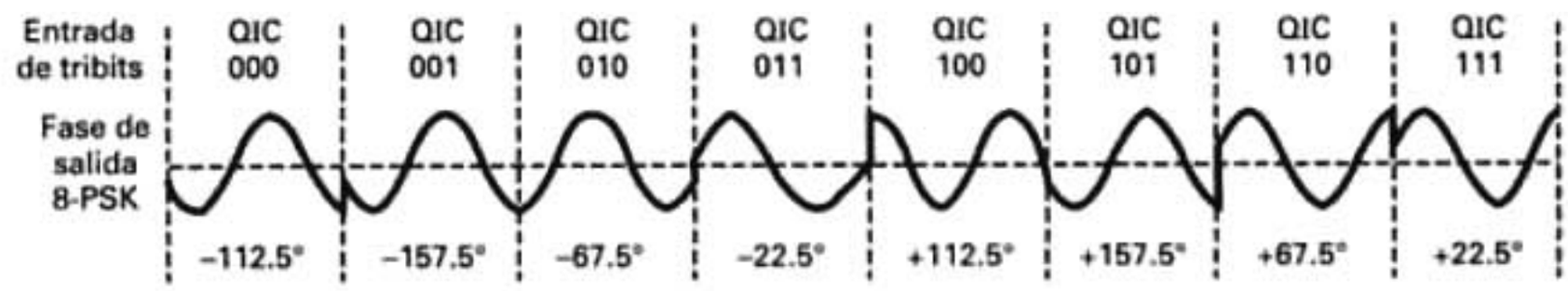


FIGURA 12-28 Fase de salida en función del tiempo, para un modulador 8-PSK

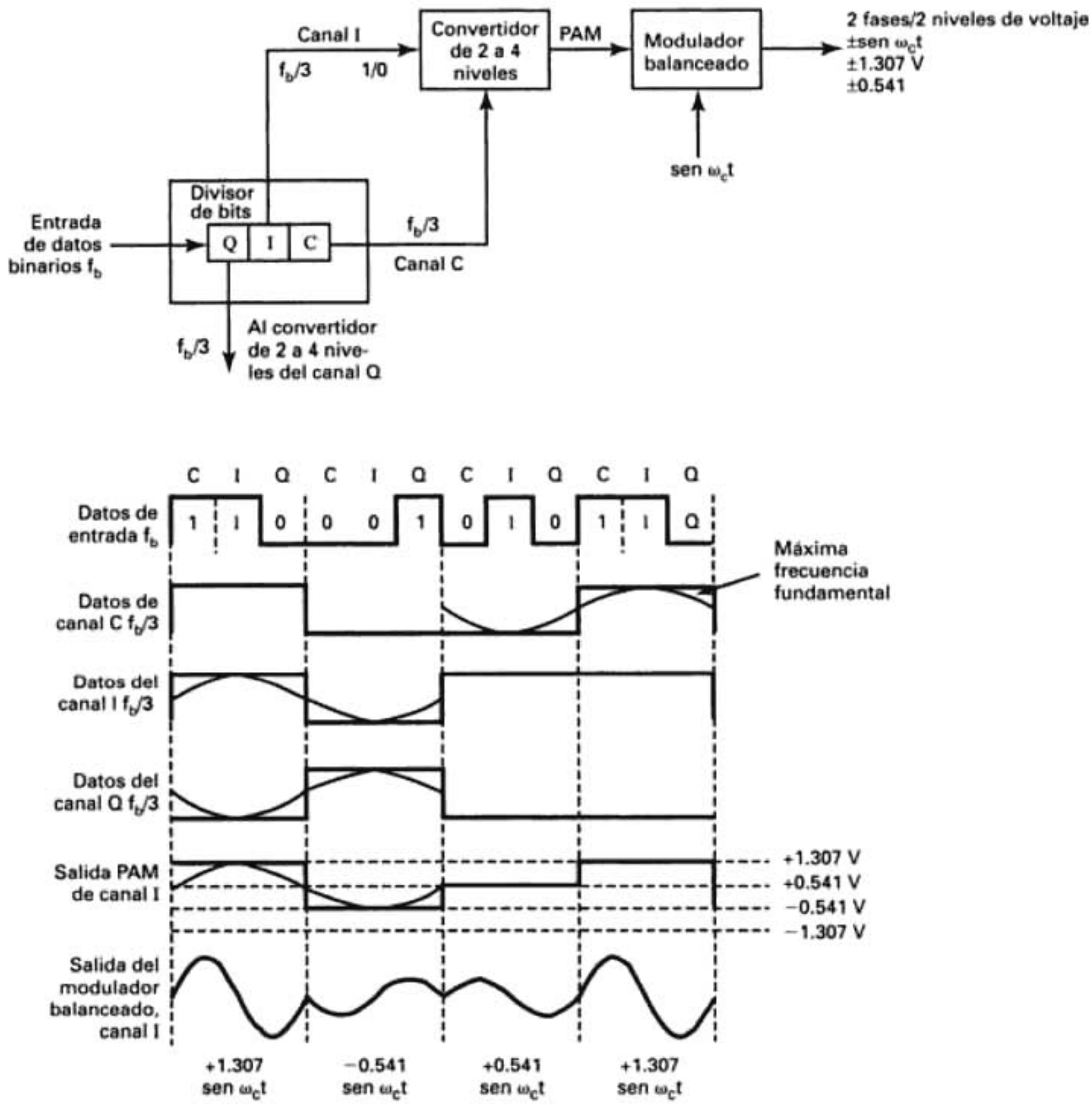


FIGURA 12-29 Consideraciones de ancho de banda de un modulador 8-PSK

La fig. 12-29 muestra la relación de sincronización de bits entre los datos de entrada binaria, los datos del canal I, Q y C y las señales de modulación de amplitud de pulso (PAM) I y Q. Se ve que la frecuencia fundamental máxima en el canal I, Q o C, es igual a la sexta parte de la rapidez de bits en la entrada binaria (un ciclo en el canal I, Q o C, tarda lo mismo que seis bits de entrada). También, la máxima frecuencia fundamental en cualquier señal PAM es igual a un sexto de la rapidez de entrada de bits binarios.

Con un modulador 8-PSK hay un cambio de fase en la salida, por cada tres bits de datos de entrada. En consecuencia, los baudios en el 8-PSK son $f_b/3$, igual que el ancho mínimo de banda. También los moduladores balanceados son moduladores de producto; sus salidas son igual al producto de la portadora por la señal PAM, es decir

$$\theta = (X \text{ sen } \omega_a t)(\text{sen } \omega_c t)$$

en donde

$$\underbrace{\omega_a t = 2\pi \frac{f_b}{6} t}_{\text{señal moduladora}} \quad \text{y} \quad \underbrace{\omega_c t = 2\pi f_c t}_{\text{portadora}}$$

y

$$X = \pm 1.307 \text{ o } \pm 0.541$$

por lo anterior,

$$\theta = \left(X \text{ sen } 2\pi \frac{f_b}{6} t\right)(\text{sen } 2\pi f_c t)$$

$$= \frac{X}{2} \cos 2\pi\left(f_c - \frac{f_b}{6}\right)t - \frac{X}{2} \cos 2\pi\left(f_c + \frac{f_b}{6}\right)t$$

El espectro de frecuencias de salida va desde $f_c + f_b/6$ hasta $f_c - f_b/6$, y el ancho mínimo de banda, f_N, es

$$\left(f_c + \frac{f_b}{6}\right) - \left(f_c - \frac{f_b}{6}\right) = \frac{2f_b}{6} = \frac{f_b}{3}$$

Ejemplo 12-7

Calcular el ancho mínimo de banda bilateral de Nyquist, f_N, y los baudios, para un modulador 8-PSK con una rapidez de entrada de bits f_b = 10 Mbps y una frecuencia de portadora de 70 MHz. También, comparar los resultados con los encontrados para los moduladores BPSK y QPSK de los ejemplos 12-3 y 12-5. Usar el diagrama de bloques de 8-PSK de la fig. 12-25 como modelo de modulador.

Solución La rapidez de bits en los canales I, Q y C es igual a la tercera parte de la rapidez de entrada de bits, es decir

$$f_{bC} = f_{bQ} = f_{bI} = \frac{10 \text{ Mbps}}{3} = 3.33 \text{ Mbps}$$

Por consiguiente, la máxima rapidez de cambio y la máxima frecuencia fundamental que se presentan a cualquiera de los moduladores balanceados es

$$f_a = \frac{f_{bC}}{2} \text{ o } \frac{f_{bQ}}{2} \text{ o } \frac{f_{bI}}{2} = \frac{3.33 \text{ Mbps}}{2} = 1.667 \text{ Mbps}$$

La onda de salida de los moduladores balanceados es

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

$$\frac{1}{2} \cos 2\pi(f_c - f_a)t - \frac{1}{2} \cos 2\pi(f_c + f_a)t$$

$$\frac{1}{2} \cos 2\pi[(70 - 1.667) \text{ MHz}]t - \frac{1}{2} \cos 2\pi[(70 + 1.667) \text{ MHz}]t$$

$$\frac{1}{2} \cos 2\pi(68.333 \text{ MHz})t - \frac{1}{2} \cos 2\pi(71.667 \text{ MHz})t$$

El ancho mínimo de banda de Nyquist es

$$f_N = (71.667 - 68.333) \text{ MHz} = 3.333 \text{ MHz}$$

El ancho mínimo de banda para el 8-PSK también se puede calcular tan sólo con sustituir en la ecuación 12-13b

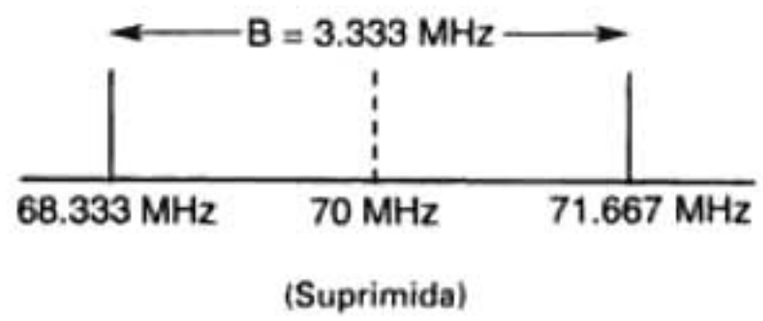
$$B = \frac{10 \text{ Mbps}}{3}$$

$$= 3.33 \text{ MHz}$$

Otra vez, los baudios son iguales al ancho de banda, y entonces

baudios = 3.333 megabaudios.

El espectro de salida es el siguiente



$$f_N = 3.333 \text{ MHz}$$

Se aprecia que para la misma rapidez de entrada de bits, el ancho mínimo de banda requerido para pasar la salida de un modulador 8-PSK es la tercera parte del necesario para un modulador BPSK del ejemplo 12-3, y 50% menor que el necesario para el modulador QPSK en el ejemplo 12-5. También, en cada caso los baudios se redujeron en las mismas proporciones.

Receptor 8-PSK. La fig. 12-30 contiene un diagrama de bloques de un receptor 8-PSK. El divisor de potencia dirige la señal 8-PSK de entrada a los detectores de producto I y Q, y al circuito de recuperación de portadora. Este circuito reproduce la señal original del oscilador de referencia. La señal 8-PSK de entrada se mezcla con la portadora recuperada en el detector I de producto, y con la portadora en cuadratura en el detector de producto Q. Las salidas de los detectores de producto son señales PAM de 4 niveles que se alimentan a los *convertidores analógico a digital* (ADC, de *analog-to-digital converter*) de 4 a 2 niveles. Las salidas del convertidor de 4 a 2 niveles del canal I son los bits I y C, mientras que las del convertidor de 4 a 2 niveles del canal Q son los bits Q y $\overline{C}$. El circuito lógico de paralelo a serie convierten los pares I/C y Q/$\overline{C}$ en corrientes de salida de datos I, Q y C en serie.

PSK de dieciséis fases

La *manipulación por desplazamiento de fase de dieciséis fases* (16-PSK, de *sixteen-phase phase shift keying*) es una técnica de codificación M-aria, en la que $M = 16$; hay 16 fases posibles de salida. Un modulador 16-PSK actúa sobre los datos de entrada, en grupos de cuatro bits ($2^4 = 16$) llamados *cuadbits*. La fase de salida no cambia sino hasta que hayan entrado cuatro bits al modulador. En consecuencia, la rapidez de cambio de salida (los baudios) y el ancho mínimo de banda son iguales a la cuarta parte de la rapidez de bits que llegan, $f_b/4$. La tabla de verdad y el diagrama de constelación para un transmisor 16-PSK se ven en la fig. 12-31.

En la 16-PSK, la separación angular entre las fases adyacentes sólo es 22.5° y, en consecuencia, una señal 16-PSK puede sufrir un desplazamiento de fase hasta de ±11.25° durante su transmisión y seguir reteniendo su integridad. Por esta causa, la 16-PSK es muy susceptible a irregularidades introducidas en el medio de transmisión, así es que, casi no se usa.

MODULACIÓN DE AMPLITUD EN CUADRATURA

La *modulación de amplitud en cuadratura* (QAM, por *quadrature amplitude modulation*) es una forma de modulación digital, donde la información digital está contenida tanto en la amplitud como en la fase de la portadora transmitida.

Ocho QAM

La ocho QAM (8-QAM) es una técnica de codificación M-aria en la que $M = 8$. A diferencia de la 8-PSK, la señal de salida de un modulador 8-QAM no es una señal de amplitud constante.

Transmisor 8-QAM. La fig. 12-32a muestra el diagrama de bloques de un transmisor 8-QAM. Allí se ve que la única diferencia entre los transmisores 8-QAM y el 8-PSK de la fig. 12-25 es la omisión del inversor entre el canal C y el modulador de producto Q. Como en la 8-PSK, los datos que llegan se dividen en grupos de tres bits (tribits): las corrientes I, Q y C de bits, cada una con rapidez de bits igual a la tercera parte de la rapidez de entrada de datos.

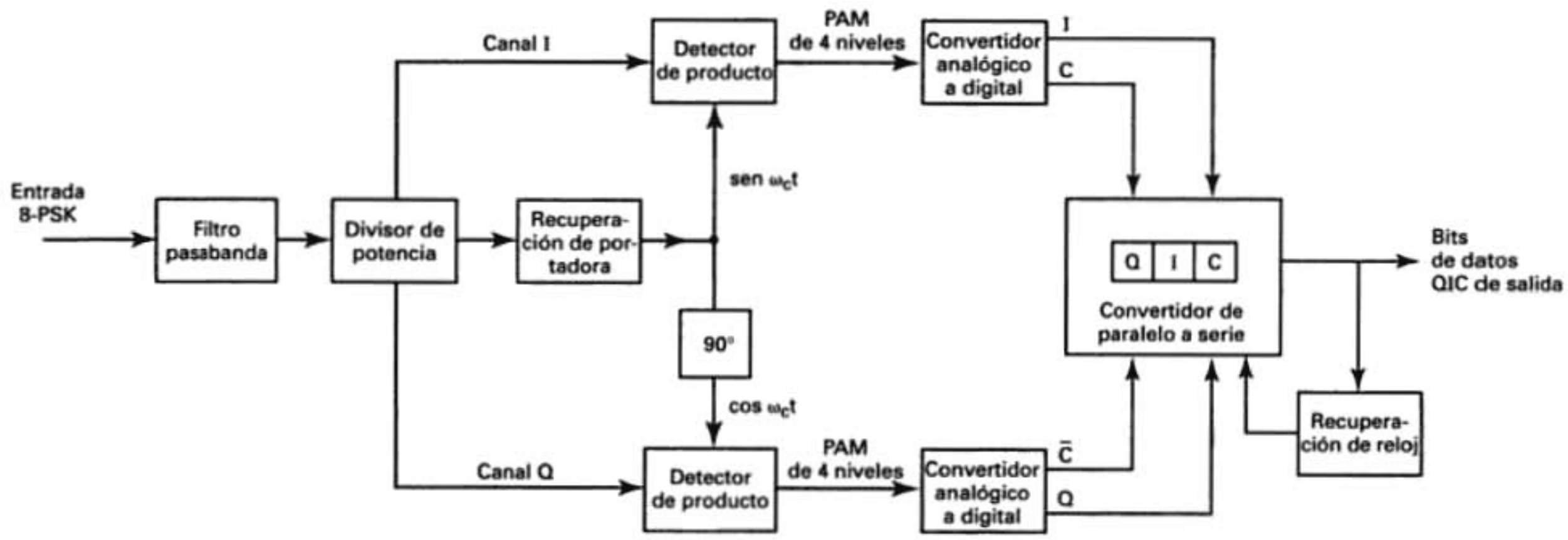
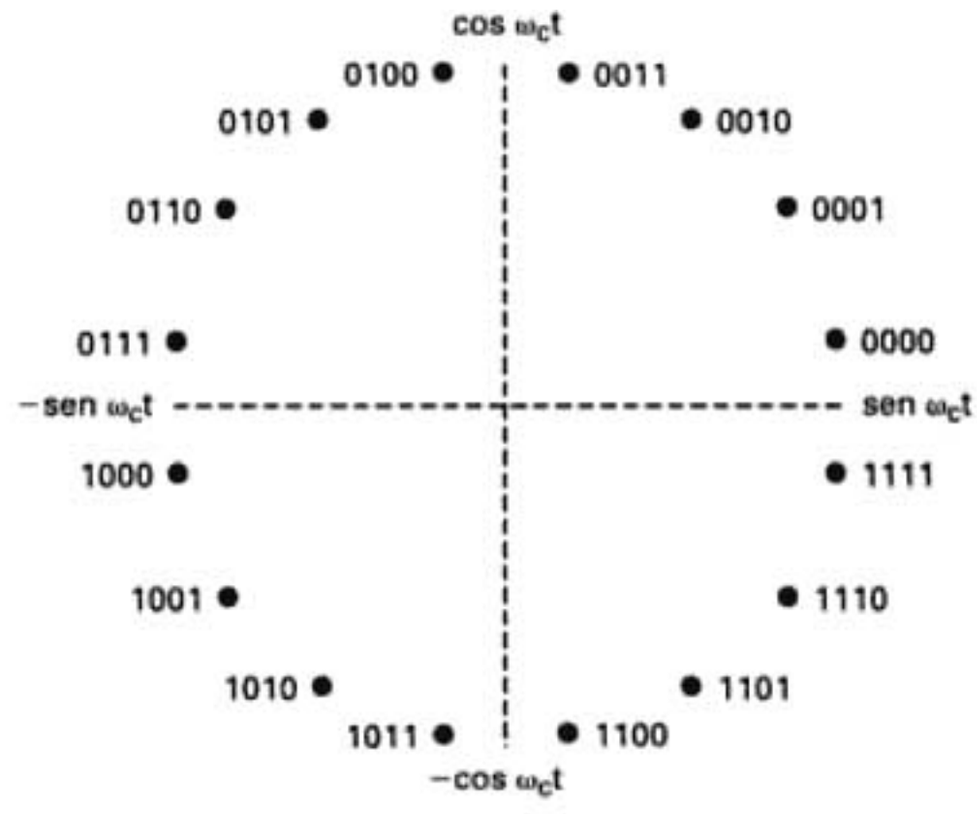


FIGURA 12-30 Receptor 8-PSK

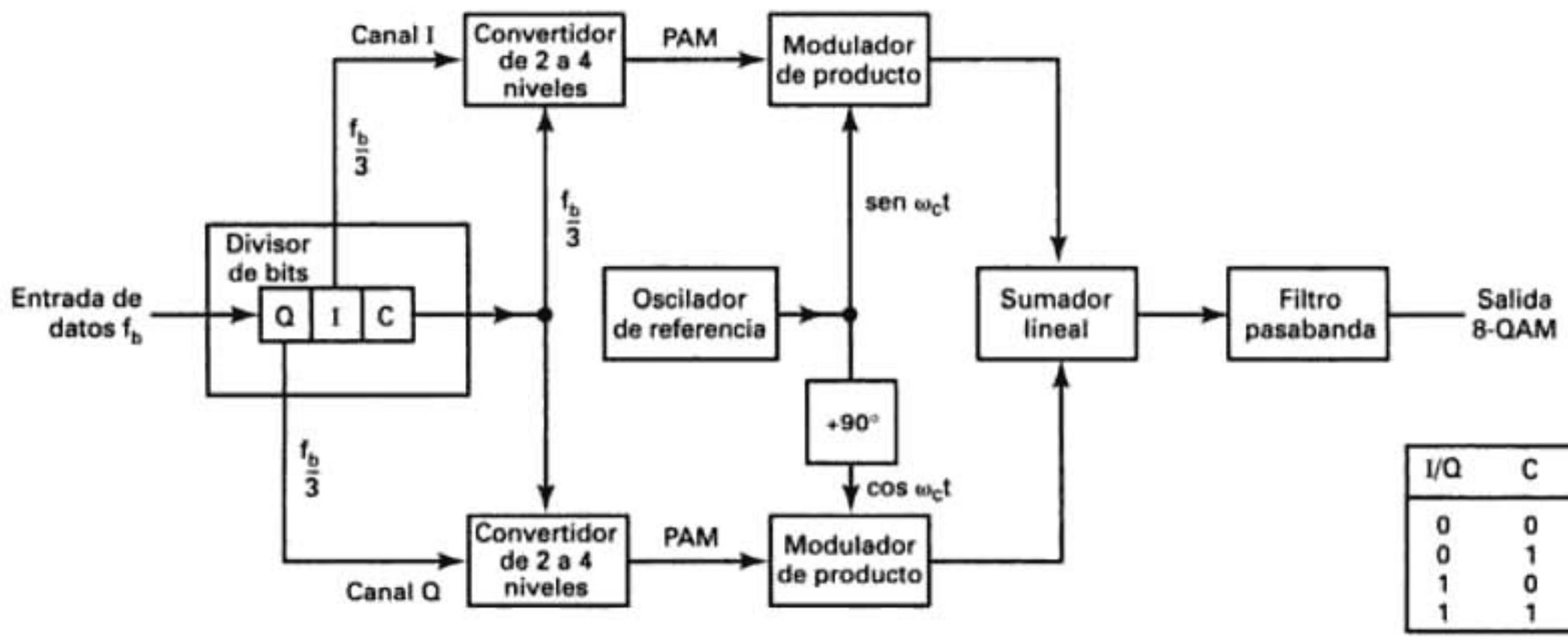
Bits de código	Fase	Bits de código	Fase
0000	11.25°	1000	191.25°
0001	33.75°	1001	213.75°
0010	56.25°	1010	236.25°
0011	78.75°	1011	258.75°
0100	101.25°	1100	281.25°
0101	123.75°	1101	303.75°
0110	146.25°	1110	326.25°
0111	168.75°	1111	348.75°

(a)



(b)

FIGURA 12-31 16-PSK: (a) tabla de verdad; (b) diagrama de constelación



(a)

I/Q	C	Salida
0	0	−0.541 V
0	1	−1.307 V
1	0	+0.541 V
1	1	+1.307 V

(b)

FIGURA 12-32 Transmisor 8-QAM: (a) diagrama de bloques; (b) tabla de verdad, convertidores de 2 a 4 niveles

También, aquí los bits I y Q determinan la polaridad de la señal PAM en la salida de los convertidores de 2 a 4 niveles, y el canal C determina la magnitud. Como el bit C se alimenta sin inversión a ambos convertidores de 2 a 4 niveles de los canales I y Q, las magnitudes de las señales PAM, I y Q siempre son iguales. Sus polaridades dependen del estado lógico de los bits I y Q y, en consecuencia, pueden ser distintas. La fig. 12-32b muestra la tabla de verdad para los convertidores de 2 a 4 niveles, del canal I y del canal Q; son idénticos.

Ejemplo 12-8

Calcular la amplitud y la fase de salida del transmisor 8-QAM de la fig. 12-32a, para una entrada de tribits Q = 0, I = 0 y C = 0 (000).

Solución Las entradas al convertidor de 2 a 4 niveles del canal I son I = 0 y C = 0. De acuerdo con la fig. 12-32b, la salida es −0.541 V. Las entradas al convertidor de 2 a 4 niveles del canal Q son Q = 0 y C = 0. También, según la fig. 12-32b, la salida es −0.541 V.

Así, las dos entradas al modulador de producto del canal I son -0.541 y sen $\omega_c t$. La salida es

$$I = (-0.541)(\text{sen } \omega_c t) = -0.541 \text{ sen } \omega_c t$$

Las dos entradas al modulador de producto del canal Q son -0.541 y cos $\omega_c t$. La salida es

$$Q = (-0.541)(\cos \omega_c t) = -0.541 \cos \omega_c t$$

Las salidas de los moduladores de producto de los canales I y Q se combinan en el sumador lineal y producen la siguiente salida

$$\begin{aligned} \text{salida del sumador} &= -0.541 \text{ sen } \omega_c t - 0.541 \cos \omega_c t \\ &= 0.765 \text{ sen}(\omega_c t - 135°) \end{aligned}$$

Para los códigos restantes de tribits (001, 010, 011, 100, 101, 110 y 111), el procedimiento es el mismo. Los resultados se ven en la fig. 12-33.

La fig. 12-34 muestra la relación de fase de salida en función del tiempo para un modulador 8-QAM. Nótese que hay dos amplitudes de salida, y sólo son posibles cuatro fases.

Consideraciones de ancho de banda con 8-QAM. En el 8-QAM, la rapidez de bits en los canales I y Q es la tercera parte de la rapidez de entrada binaria, igual que en 8-PSK. En consecuencia, la máxima frecuencia moduladora fundamental, y la máxima rapidez de cambio de salida necesaria para 8-QAM es $f_b/3$, igual que para 8-PSK.

Entrada binaria			Salida 8-QAM	
Q	I	C	Amplitud	Fase
0	0	0	0.765 V	−135°
0	0	1	1.848 V	−135°
0	1	0	0.765 V	−45°
0	1	1	1.848 V	−45°
1	0	0	0.765 V	+135°
1	0	1	1.848 V	+135°
1	1	0	0.765 V	+45°
1	1	1	1.848 V	+45°

(a)

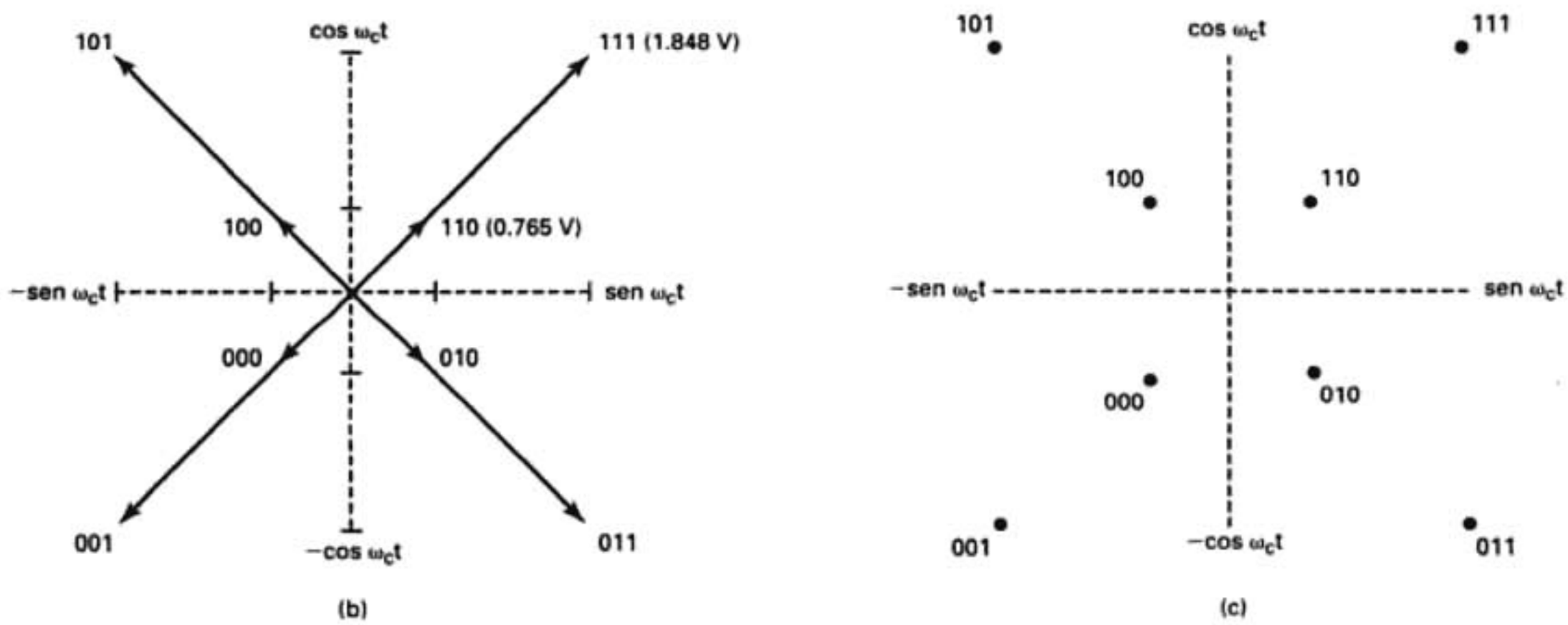


FIGURA 12-33 Modulador 8-QAM: (a) tabla de verdad; (b) diagrama fasorial; (c) diagrama de constelación

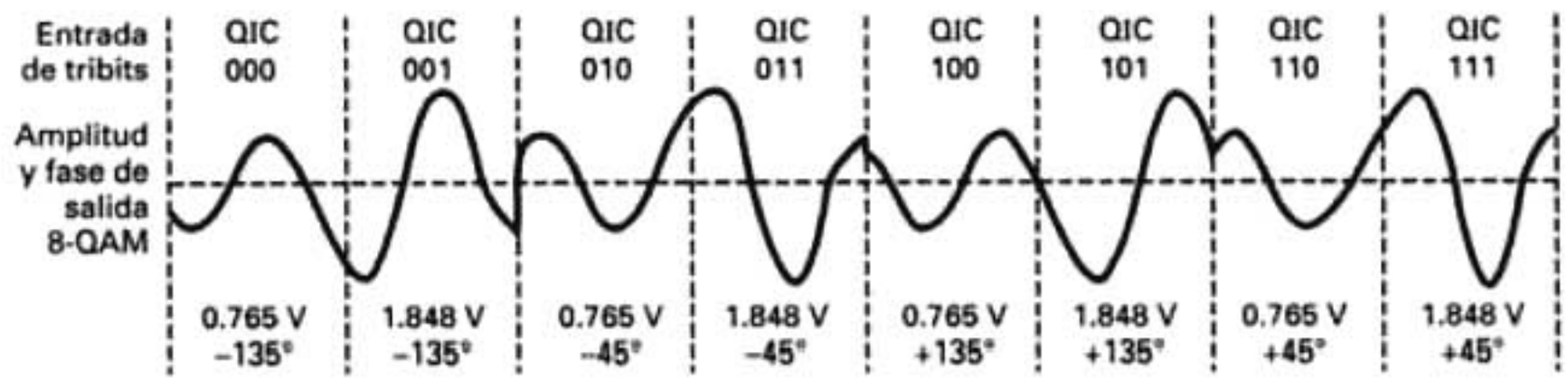


FIGURA 12-34 Fase y amplitud de salida en función del tiempo, para 8-QAM

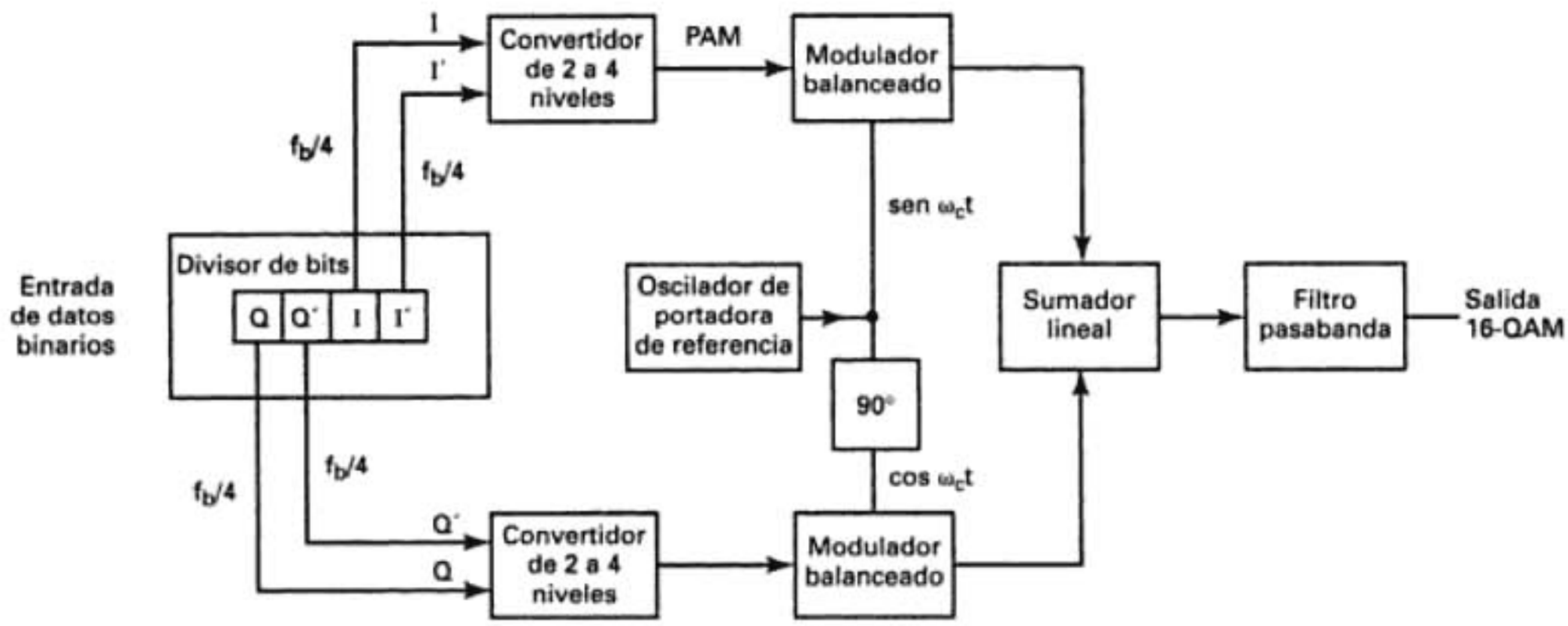


FIGURA 12-35 Diagrama de bloques de un transmisor 16-QAM

Receptor 8-QAM. Un receptor 8-QAM es casi idéntico al receptor 8-PSK de la fig. 12-30. Las diferencias son los niveles de modulación de amplitud de pulso (PAM) en la salida de los detectores de producto y las señales binarias en la salida de los convertidores analógico a digital. Como hay dos amplitudes de transmisión posibles con 8-QAM, que son distintas que las que se obtienen con 8-PSK, los cuatro niveles PAM desmodulados en el 8-QAM son distintos de los de 8-PSK. En consecuencia, el factor de conversión para los convertidores analógico a digital también debe ser distinto. También, en 8-QAM, las señales binarias de salida del convertidor analógico a digital del canal I son los bits I y C, y las señales de salida binaria del convertidor analógico a digital del canal Q son los bits Q y C.

Dieciséis QAM

Como en la manipulación 16-PSK, la 16-QAM es un sistema M-ario en el que $M = 16$. Los datos de entrada se manejan en grupos de cuatro ($2^4 = 16$). Como en 8-QAM, se varían tanto la fase como la amplitud de la portadora de transmisión.

Transmisor 16-QAM. En la fig. 12-35 se muestra un diagrama de bloques de un transmisor 16-QAM. Los datos binarios de entrada se dividen en cuatro canales: I, I′, Q y Q′. La rapidez de bits en cada canal es igual a la cuarta parte de la rapidez de entrada de bits ($f_b/4$). Se sincronizan cuatro bits en serie en el divisor de bits; a continuación, salen en forma simultánea y en paralelo con los canales I, I′, Q y Q′. Los bits I y Q determinan la polaridad en la salida de los convertidores de 2 a 4 niveles (1 lógico = positivo y 0 lógico = negativo). Los bits I′ y Q′ determinan la magnitud (1 lógico = 0.821 V y 0 lógico = 0.22 V). En consecuencia, los convertidores de 2 a 4 niveles generan una señal PAM de 4 niveles. Son posibles dos polaridades y dos magnitudes en la salida de cada convertidor de 2 a 4 niveles, que son ±0.22 V y ±0.821 V. Las señales PAM modulan las portadoras en fase y en cuadratura en los moduladores de producto.

I	I'	Salida
0	0	−0.22 V
0	1	−0.821 V
1	0	+0.22 V
1	1	+0.821 V

(a)

Q	Q'	Salida
0	0	−0.22 V
0	1	−0.821 V
1	0	+0.22 V
1	1	+0.821 V

(b)

FIGURA 12-36 Tablas de verdad para los convertidores de 2 a 4 niveles, canales I y Q: (a) canal I; (b) canal Q

Para cada modulador de producto son posibles cuatro salidas. Para el modulador de producto I son $+0.821$ sen $\omega_c t$, -8.821 sen $\omega_c t$, $+0.22$ sen $\omega_c t$ y -0.22 sen $\omega_c t$. Para el modulador de producto Q son $+0.821$ cos $\omega_c t$, $+0.22$ cos $\omega_c t$, -0.821 cos $\omega_c t$ y -0.22 cos $\omega_c t$. El sumador lineal combina las salidas de los moduladores de producto de canal I y Q, y produce las 16 condiciones de salida necesarias para 16-QAM. La fig. 12-36 muestra la tabla de verdad para los convertidores de 2 a 4 niveles de canal I y Q.

Ejemplo 12-9

Para una entrada de cuadribits de I = 0, I′ = 0, Q = 0 y Q′ = 0 (0000), Determinar la amplitud y la fase de salida en el modulador 16-QAM de la fig. 12-35.

Solución Las entradas al convertidor de 2 a 4 niveles del canal I son I = 0 e I′ = 0. De acuerdo con la fig. 12-36, la salida es −0.22V. Las entradas al convertidor de 2 a 4 niveles del canal Q son Q = 0 y Q′ = 0. También, según la fig. 12-36, la salida es −0.22 V.

Así, las dos entradas al modulador de producto del canal I son −0.22 V y sen $\omega_c t$. La salida es

$$I = (-0.22)(\sin \omega_c t) = -0.22 \sin \omega_c t$$

Las dos entradas al modulador de producto del canal Q son −0.22 V y sen $\omega_c t$. La salida es

$$Q = (-0.22)(\cos \omega_c t) = -0.22 \cos \omega_c t$$

Las salidas de los moduladores de producto de los canales I y Q se combinan en el sumador lineal y producen una salida modulada de

$$\begin{aligned}\text{salida del sumador} &= -0.22 \text{ sen } \omega_c t - 0.22 \cos \omega_c t \\ &= 0.311 \text{ sen}(\omega_c t - 135°)\end{aligned}$$

Para las siguientes claves de cuadribits, el procedimiento es igual. Los resultados se ven en la fig. 12-37.

Consideraciones de ancho de banda con 16-QAM. En 16-QAM, como los datos de entrada se dividen en cuatro canales, la rapidez de bits en el canal I, I′, Q o Q′ es igual a la cuarta parte de la rapidez de entrada de datos binarios, $f_b/4$, porque el divisor de bits estira los bits I, I′, Q y Q′ hasta cuatro veces su longitud de entrada. También, como esos bits salen en forma simultánea y en paralelo, los convertidores de 2 a 4 niveles ven un cambio en sus entradas y salidas con una rapidez igual a la cuarta parte de la rapidez de entrada de datos.

La fig. 12-38 muestra la relación de sincronización de bits entre los datos de entrada binaria, los datos del canal I, I′ Q y Q′ y la señal PAM I. Se ve que la frecuencia fundamental máxima en los canales I, I′, Q o Q′ es igual a la octava parte de la rapidez de bits en la entrada de datos binarios (un ciclo el el canal I, I′, Q o Q′ tarda igual que ocho bits de entrada). También, la máxima frecuencia fundamental de cualquier señal PAM es igual a la octava parte de la rapidez de entrada de bits.

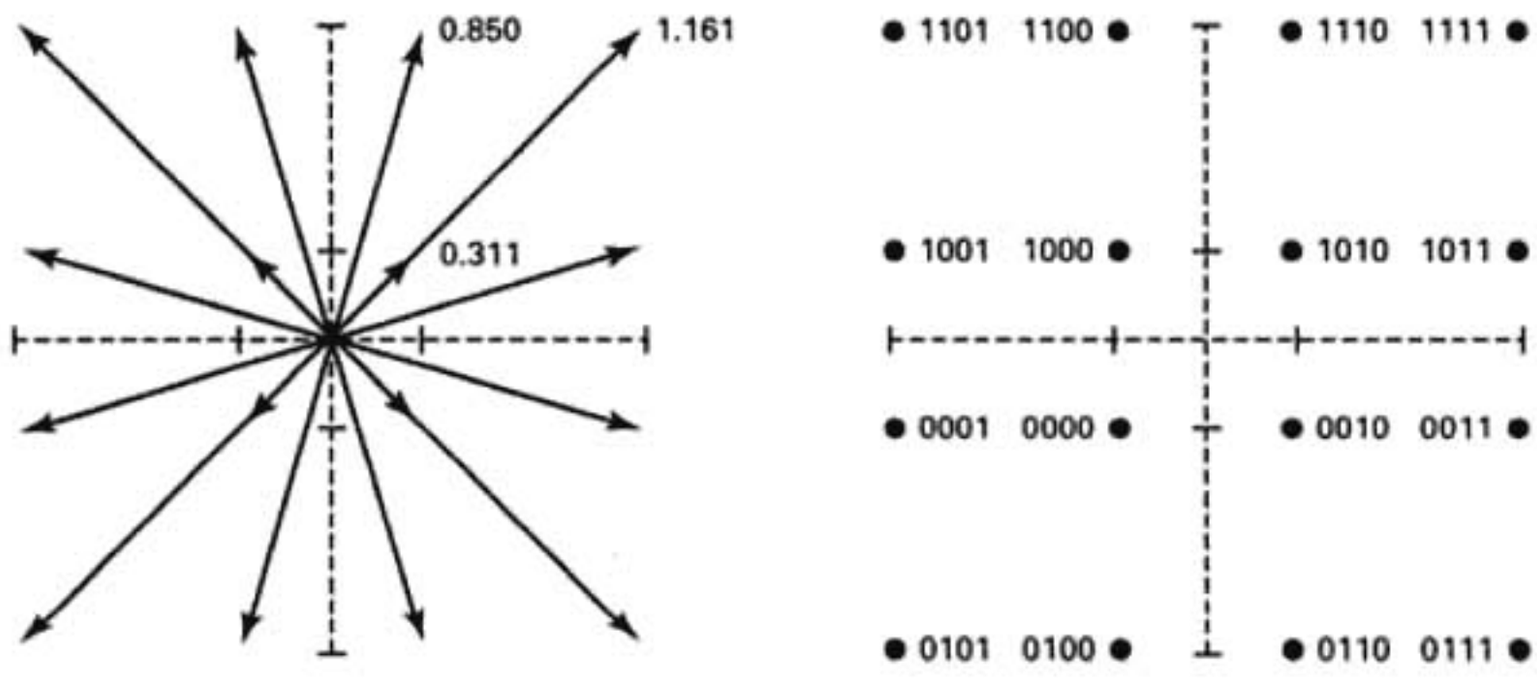
Con un modulador 16-QAM, hay un cambio en la señal de salida, sea en su fase, su amplitud o en ambas cosas, para cada cuatro bits de datos de entrada. Por lo anterior, los baudios son iguales a $f_b/4$, igual que el ancho mínimo de banda.

También en este caso, los moduladores balanceados son moduladores de producto, y se pueden representar sus salidas con la siguiente ecuación

$$\text{salida} = (X \text{ sen } \omega_a t)(\text{sen } \omega_c t) \tag{12-15}$$

Entrada binaria				Salida 16-QAM	
Q	Q'	I	I'		
0	0	0	0	0.311 V	−135°
0	0	0	1	0.850 V	−165°
0	0	1	0	0.311 V	−45°
0	0	1	1	0.850 V	−15°
0	1	0	0	0.850 V	−105°
0	1	0	1	1.161 V	−135°
0	1	1	0	0.850 V	−75°
0	1	1	1	1.161 V	−45°
1	0	0	0	0.311 V	135°
1	0	0	1	0.850 V	165°
1	0	1	0	0.311 V	45°
1	0	1	1	0.850 V	15°
1	1	0	0	0.850 V	105°
1	1	0	1	1.161 V	135°
1	1	1	0	0.850 V	75°
1	1	1	1	1.161 V	45°

(a)



(b) (c)

FIGURA 12-37 Modulador 16-QAM: (a) tabla de verdad; (b) diagrama fasorial; (c) diagrama de constelación

en donde $$\underbrace{\omega_a t = 2\pi \frac{f_b}{8} t}_{\text{fase de la señal moduladora}} \quad \text{y} \quad \underbrace{\omega_c t = 2\pi f_c t}_{\text{fase de la portadora}}$$

y $$X = \pm 0.22 \quad \text{o} \quad \pm 0.821$$

Así, $$\text{salida} = \left(X \operatorname{sen} 2\pi \frac{f_b}{8} t\right)(\operatorname{sen} 2\pi f_c t)$$

$$= \frac{X}{2} \cos 2\pi\left(f_c - \frac{f_b}{8}\right)t = \frac{X}{2} \cos 2\pi\left(f_c + \frac{f_b}{8}\right)t$$

El espectro de frecuencias de salida se extiende desde $f_c + f_b/8$ hasta $f_c - f_b/8$, y el ancho mínimo de banda, f_N, es

$$\left(f_c + \frac{f_b}{8}\right) - \left(f_c - \frac{f_b}{8}\right) = \frac{2f_b}{8} = \frac{f_b}{4}$$

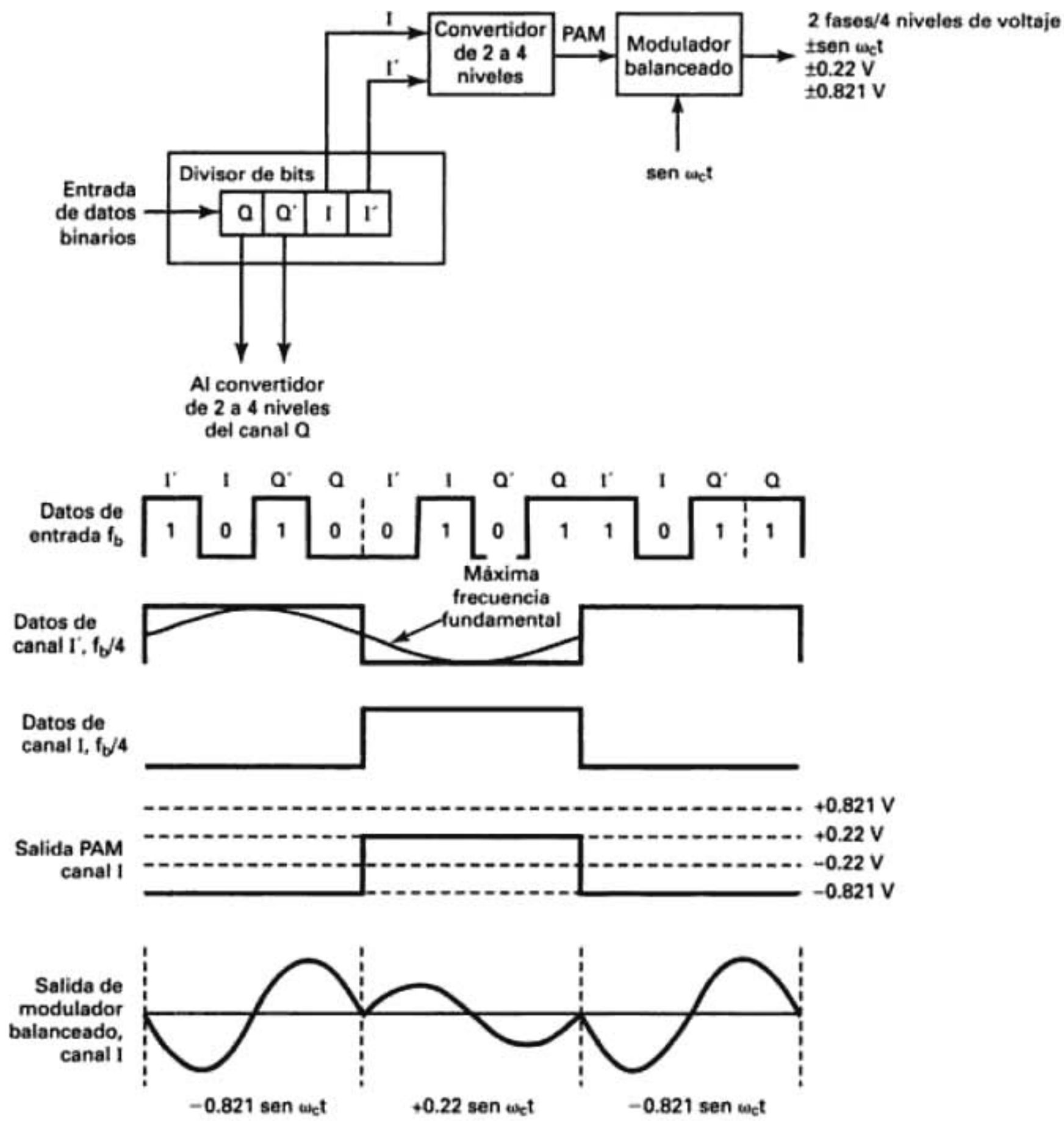


FIGURA 12-38 Consideraciones de ancho de banda de un modulador 16-QAM

Ejemplo 12-10

Calcular la frecuencia bilateral mínima de Nyquist f_N, y los baudios para un modulador 16-QAM con rapidez de entrada de datos f_b = 10 Mbps y frecuencia de portadora de 70 MHz. También, comparar los resultados con los obtenidos con los moduladores BPSK, QPSK y 8-PSK en los ejemplos 12-3, 12-5 y 12-7. Usar el diagrama de bloques del 16-QAM de la fig. 12-35, como modelo de modulador.

Solución La rapidez de bits en los canales I, I', Q y Q' es igual a la cuarta parte de la rapidez de entrada de bits, es decir,

$$f_{bI} = f_{bI'} = f_{bQ} = f_{bQ'} = \frac{f_b}{4} = \frac{10 \text{ Mbps}}{4} = 2.5 \text{ Mbps}$$

Entonces, la máxima rapidez de cambio y la máxima frecuencia fundamental que se presentan a cualquiera de los moduladores balanceados es

$$f_a = \frac{f_{bI}}{2} \text{ o } \frac{f_{bI'}}{2} \text{ o } \frac{f_{bQ}}{2} \text{ o } \frac{f_{bQ'}}{2} = \frac{2.5 \text{ Mbps}}{2} = 1.25 \text{ MHz}$$

La onda de salida del modulador balanceado es

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

$$\frac{1}{2}\cos 2\pi(f_c - f_a)t - \frac{1}{2}\cos 2\pi(f_c + f_a)t$$

$$\frac{1}{2}\cos 2\pi[(70 - 1.25)\text{ MHz}]t - \frac{1}{2}\cos 2\pi[(70 + 1.25)\text{ MHz}]t$$

$$\frac{1}{2}\cos 2\pi(68.75\text{ MHz})t - \frac{1}{2}\cos 2\pi(71.25\text{ MHz})t$$

El ancho mínimo de banda de Nyquist es

$$f_N = (71.25 - 68.75)\text{ MHz} = 2.5\text{ MHz}$$

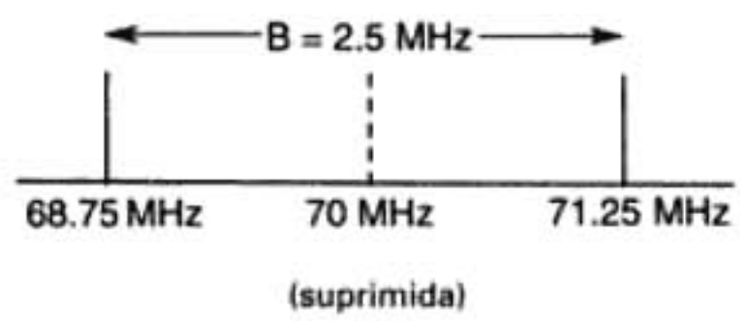
El ancho mínimo de banda para la modulación 16-QAM también se puede calcular en forma sencilla sustituyendo valores en la ecuación 12-13b:

$$B = \frac{10\text{ Mbps}}{4}$$
$$= 2.5\text{ MHz}$$

La rapidez de símbolos es igual al ancho de banda, y entonces

rapidez de símbolos = 2.5 megabaudios

El espectro de salida es el siguiente:



$$f_N = 2.5\text{ MHz}$$

Para la misma rapidez de entrada de bits, el ancho mínimo de banda necesario para pasar la salida de un modulador 16-QAM es igual a la cuarta parte de la necesaria para el modulador BPSK, la mitad de la de QPSK y 25% menos que con 8-PSK. Para cada técnica de modulación los baudios también se reducen en la misma proporción.

EFICIENCIA DE ANCHO DE BANDA

La *eficiencia de ancho de banda*, o *densidad de información*, como a veces se le llama, se usa con frecuencia para comparar el funcionamiento de dos técnicas de modulación digital. En esencia, es la relación de rapidez de transmisión de bits entre el ancho mínimo de banda necesario para determinado esquema de modulación. En general, la eficiencia de ancho de banda se normaliza a un ancho de banda de 1 Hz y en consecuencia indica la cantidad de bits que se pueden propagar a través de un medio, por cada hertz de ancho de banda. La definición matemática del ancho de banda es

$$\text{Eficiencia de AB} = \frac{\text{rapidez de transmisión (bps)}}{\text{ancho mínimo de banda (Hz)}} \tag{12-16}$$

$$= \frac{\text{bits/segundo}}{\text{hertz}} = \frac{\text{bits/segundo}}{\text{ciclos/segundo}} = \frac{\text{bits}}{\text{ciclo}}$$

Ejemplo 12-11

Determinar las eficiencias de ancho de banda para los siguientes esquemas de modulación: BPSK, QPSK, 0-PSK y 16-QAM.

Solución Recuérdese que en los ejemplos 12-3, 12-5, 12-7 y 12-10, los anchos mínimos de banda necesarios para propagar una rapidez de transmisión de 10 Mbps son, para los siguientes esquemas de modulación:

Esquema de modulación	Ancho de banda mínimo (MHz)
BPSK	10
QPSK	5
8-PSK	3.33
16-QAM	2.5

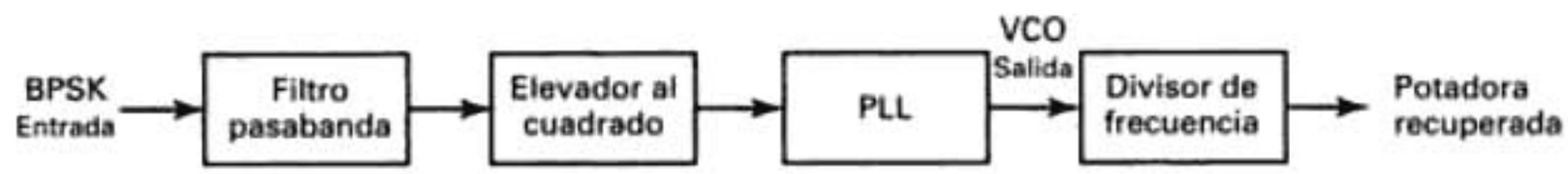


FIGURA 12-39 Circuito de recuperación de portadora con lazo cuadrático

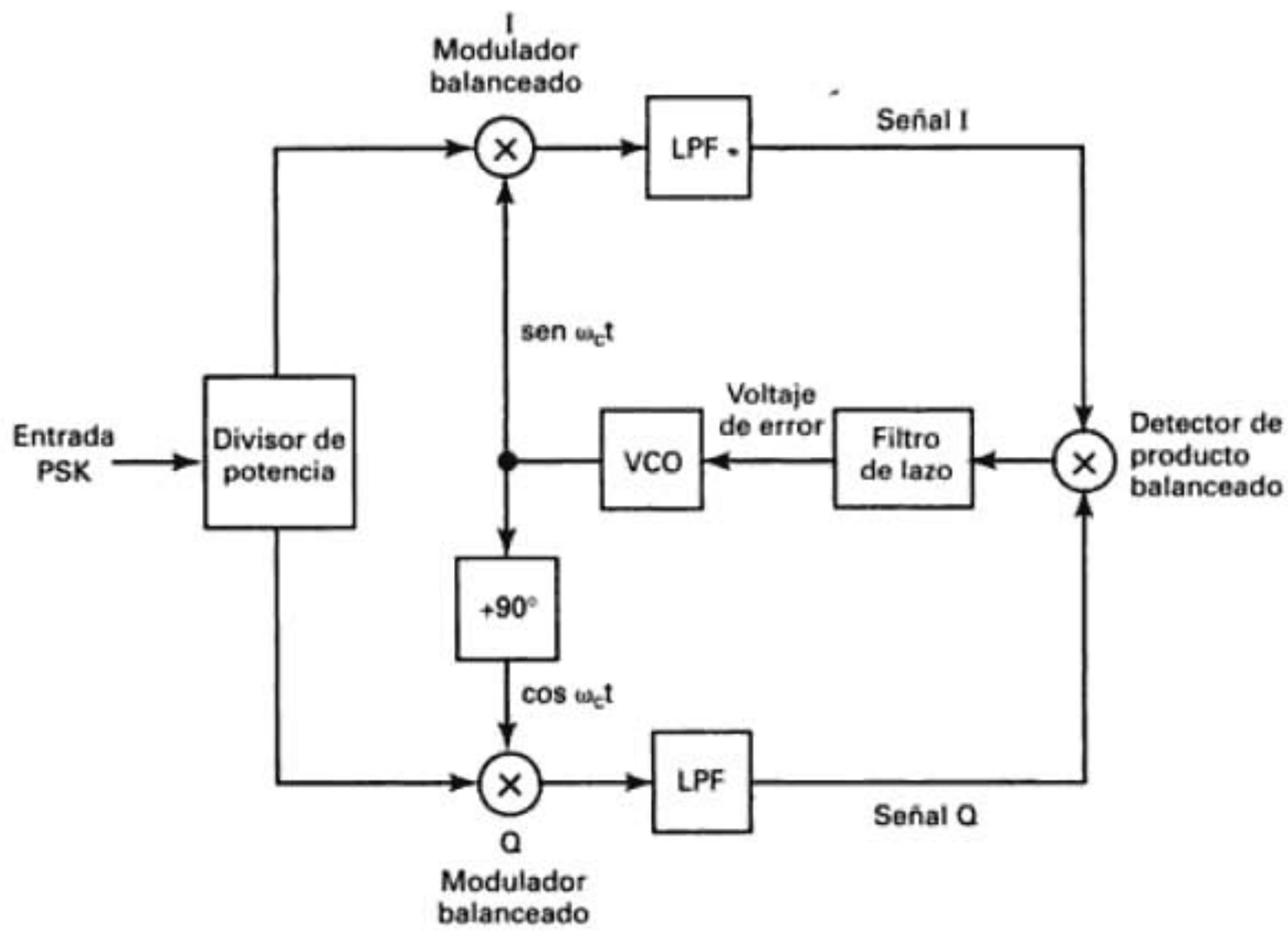


FIGURA 12-40 Circuito de recuperación de portadora con lazo de costas

Con el sistema BPSK sólo puede haber dos fases de salida: $+\text{sen}\ \omega_c t$ y $-\text{sen}\ \omega_c t$. El funcionamiento del circuito cuadrático se puede describir con ecuaciones matemáticas. Para una señal de recepción de $+\text{sen}\ \omega_c t$, la salida del circuito cuadrático es

$$\text{salida} = (+\text{sen}\ \omega_c t)(+\text{sen}\ \omega_c t) = +\text{sen}^2\ \omega_c t$$

$$= \frac{1}{2}(1 - \cos 2\omega_c t) = \underbrace{\frac{1}{2}}_{\text{(se filtra y se elimina)}} - \frac{1}{2}\cos 2\omega_c t$$

Para una señal recibida de $-\text{sen}\ \omega_c t$, la salida del circuito cuadrático es

$$\text{salida} = (-\text{sen}\ \omega_c t)(-\text{sen}\ \omega_c t) = +\text{sen}^2\ \omega_c t$$

$$= \frac{1}{2}(1 - \cos 2\omega_c t) = \underbrace{\frac{1}{2}}_{\text{(se filtra y se elimina)}} - \frac{1}{2}\cos 2\omega_c t$$

Se ve que en ambos casos, la salida del lazo de respuesta cuadrada contiene un voltaje constante (+1/2 V) y una señal con el doble de frecuencia que la portadora, $\cos 2\omega_c t$. El voltaje constante se elimina por filtración y sólo queda $\cos 2\ \omega_c t$.

Lazo de costas

Un segundo método de recuperación de portadora es el lazo de costas, o de cuadratura, que se ve en la fig. 12-40. El lazo de costas produce los mismos resultados que un circuito de respuesta cuadrada seguido de un PLL ordinario en lugar del filtro BPF. Este esquema de recuperación

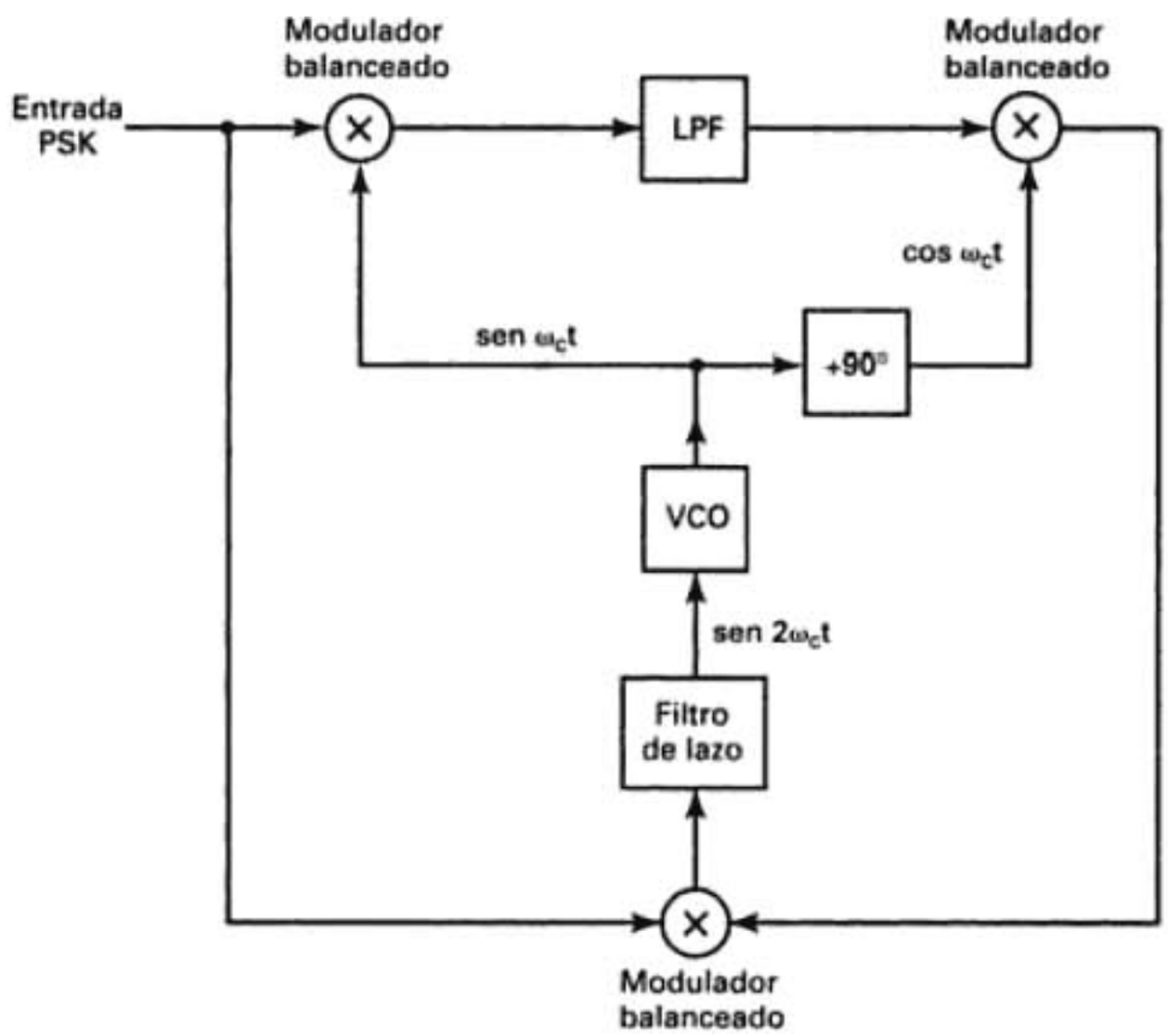


FIGURA 12-41 Circuito de recuperación de portadora con lazo remodulador

usa dos lazos paralelos de rastreo (I y Q) al mismo tiempo para obtener el producto de los componentes I y Q de la señal que activa al VCO. El lazo I enfasado usa al oscilador controlada por voltaje (VCO) como un PLL, y el lazo Q de cuadratura usa una señal de VCO desplazada 90°. Una vez que la frecuencia del VCO es igual a la de la portadora suprimida, el producto de las señales I y Q producirá un voltaje de error proporcional a cualquier error de fase en el VCO. El voltaje de error controla la fase y, en consecuencia, la frecuencia del VCO.

Remodulador

Un tercer método de recuperar una portadora de fase y frecuencia coherentes es el del remodulador, que se ve en la fig. 12-41. Este remodulador produce un voltaje de error de lazo que es proporcional al doble del error de fases entre la señal de entrada y la señal del VCO. El remodulador tiene un tiempo más corto de adquisición que los lazos cuadrático o el de costas.

Los circuitos de recuperación de portadora, para técnicas de codificación más grandes que binaria, se parecen a los de BPSK, excepto que se usan circuitos que elevan la señal de referencia a la cuarta, octava y otras potencias mayores.

MANIPULACIÓN POR DESPLAZAMIENTO DIFERENCIAL DE FASE

La *manipulación por desplazamiento diferencial de fase* (DPSK, por *differential phase shift keying*) es una forma alterna de modulación digital, en donde la información binaria de la entrada está contenida en la diferencia entre las fases de dos elementos sucesivos de señalización, y no en la fase absoluta. Con la DPSK no es necesario recuperar una portadora de fase coherente. En lugar de ella, un elemento de señalización recibido se demora por un intervalo de frecuencia del elemento de señalización, y a continuación se compara con el siguiente elemento de señalización recibido. La diferencia en la fase de los dos elementos de señalización determina la condición lógica de los datos.

BPSK diferencial

Transmisor DBPSK. La fig. 12-42a muestra un diagrama simplificado de bloques de un transmisor de *manipulación por desplazamiento diferencial de fase binaria* (DBPSK, de *differen-*

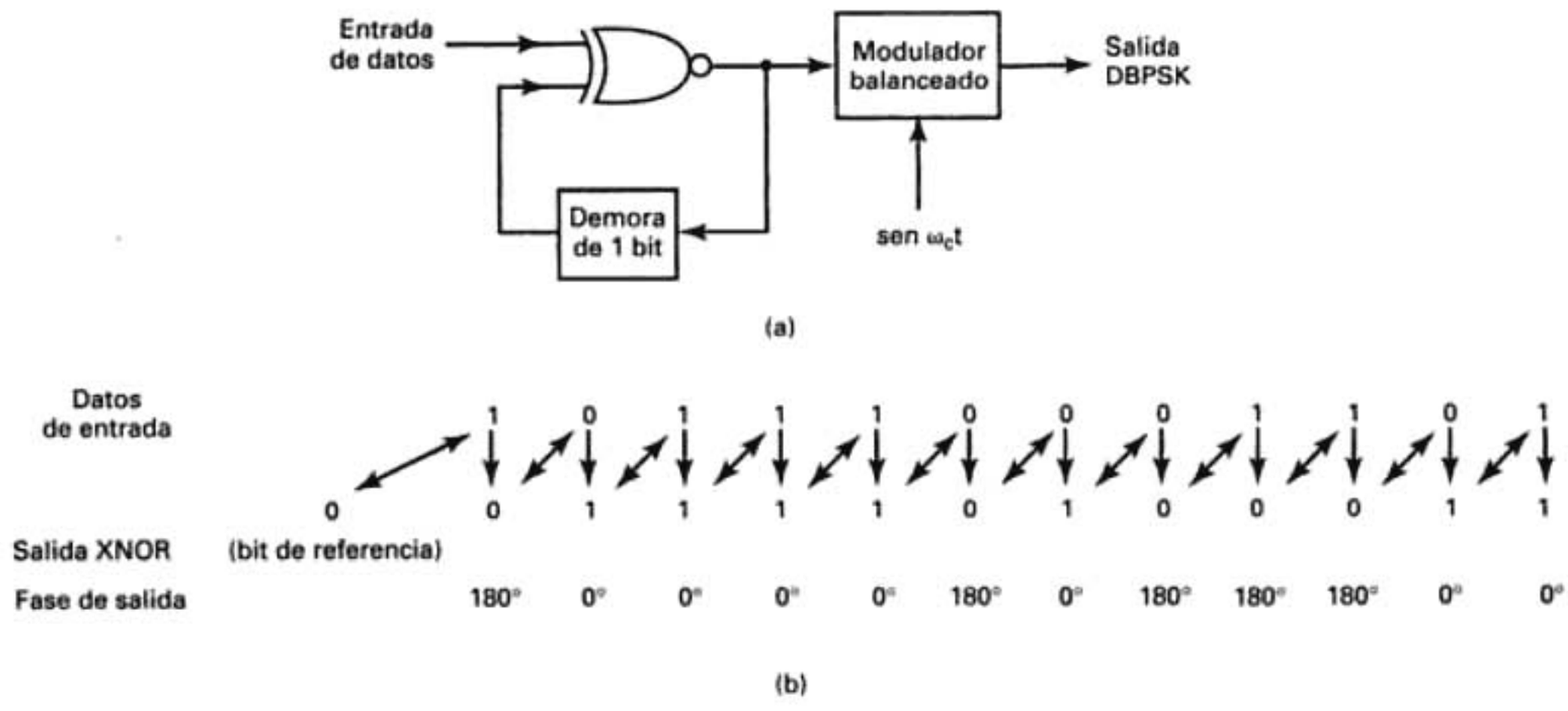


FIGURA 12-42 Modulador DBPSK: (a) diagrama de bloques; (b) diagrama de sincronización

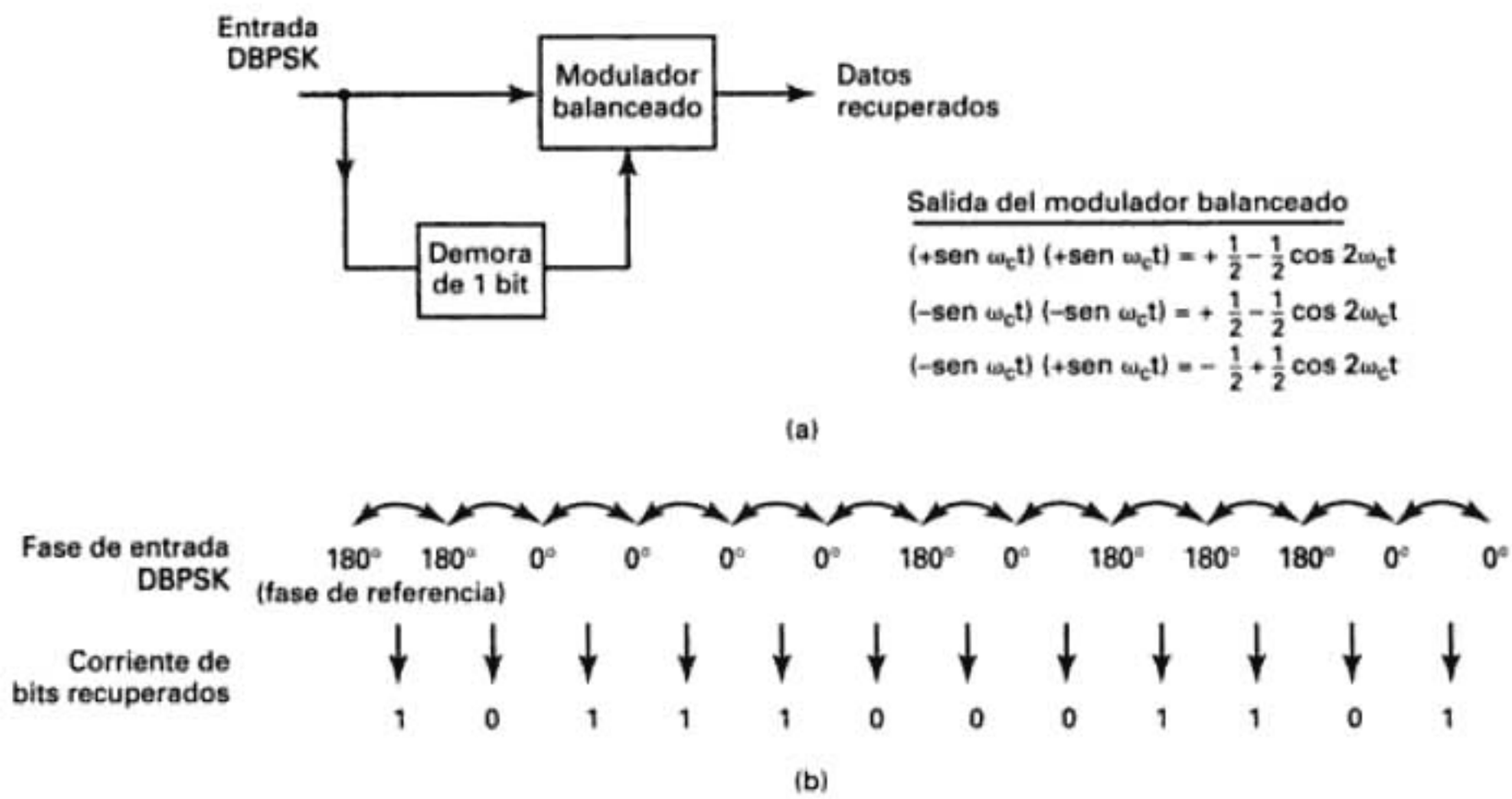


FIGURA 12-43 Desmodulador DBPSK: (a) diagrama de bloques; (b) secuencia de sincronización

tial binary phase shift keying). Un bit de información que llega se reconoce XNOR con el bit anterior, antes de entrar al modulador BPSK (modulación balanceado). Para el primer bit de datos no hay bit anterior con qué compararlo y, en consecuencia, se supone un bit inicial de referencia. La fig. 12-42b muestra la relación entre los datos de entrada, los datos de salida XNOR y la fase en la salida del modulador balanceado. Si se supone que el bit inicial de referencia es un 1 lógico, la salida del circuito XNOR no es más que el complemento del indicado.

En la fig. 12-42b el primer bit de datos se compara XNOR con el de referencia. Si son iguales, la salida XNOR es un 1 lógico; si son distintos, la salida XNOR es un 0 lógico. El modulador balanceado funciona igual que un modulador BPSK convencional; un 1 lógico produce $+\text{sen}\ \omega_c t$ en la salida, y un 0 lógico produce $-\text{sen}\ \omega_c t$.

Receptor DBPSK. La fig. 12-43 muestra el diagrama de bloques y la secuencia de sincronización de un receptor DBPSK. La señal recibida se demora un bit de tiempo y a continuación se compara con el siguiente elemento de señalización en el modulador balanceado. Si son iguales se genera un 1 lógico (voltaje +). Si son distintas, se genera un 0 lógico (voltaje −). Si

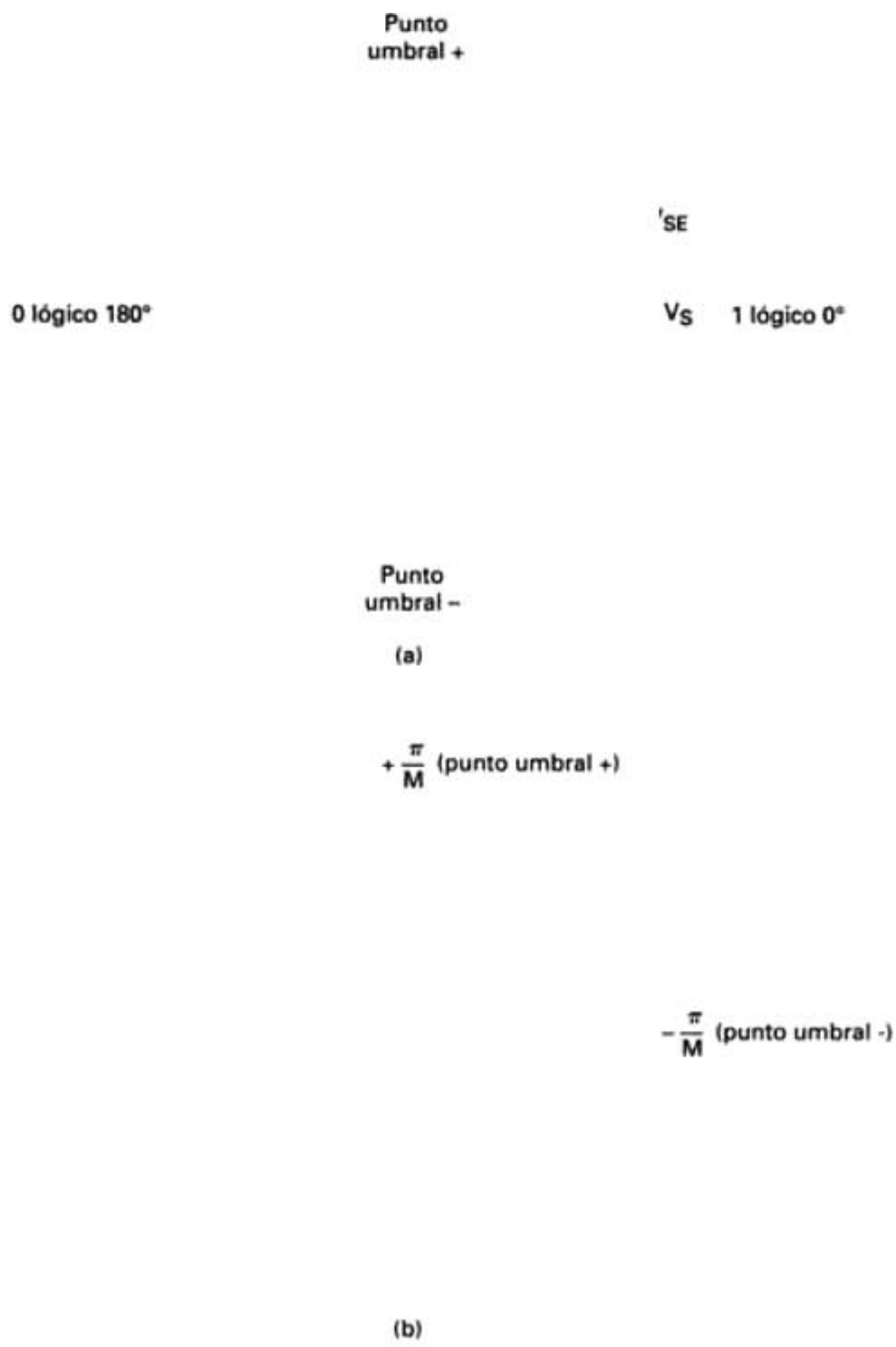


FIGURA 12-45 Región de error en PSK: (a) BPSK; (b) QPSK

en donde d = distancia entre errores
M = cantidad de fases
D = máxima amplitud de la señal

La ecuación 12-24 se reordena para despejar d, y se obtiene

$$d = \left(2 \text{ sen } \frac{180°}{M}\right) \times D \tag{12-25}$$

La fig. 12-45b muestra el diagrama estado-espacio de señal para el sistema QPSK. Se puede ver en esta figura, y en la ecuación 12-24, que el QPSK sólo puede tolerar un desplazamiento de fase de ±45°. De acuerdo con la ecuación 12-23, el desplazamiento máximo de fase en 8-PSK y 16-PSK es ±22.5° y ±11.5°, respectivamente. En consecuencia, los mayores niveles de modulación, es decir, los valores mayores de M, requieren mayor relación de energía por bit a densidad de potencia de ruido, para reducir el efecto de la interferencia de ruido. Por consiguiente, mientras mayor sea el nivel de modulación, la separación angular entre los puntos de señal será menor y la distancia entre errores será menor.

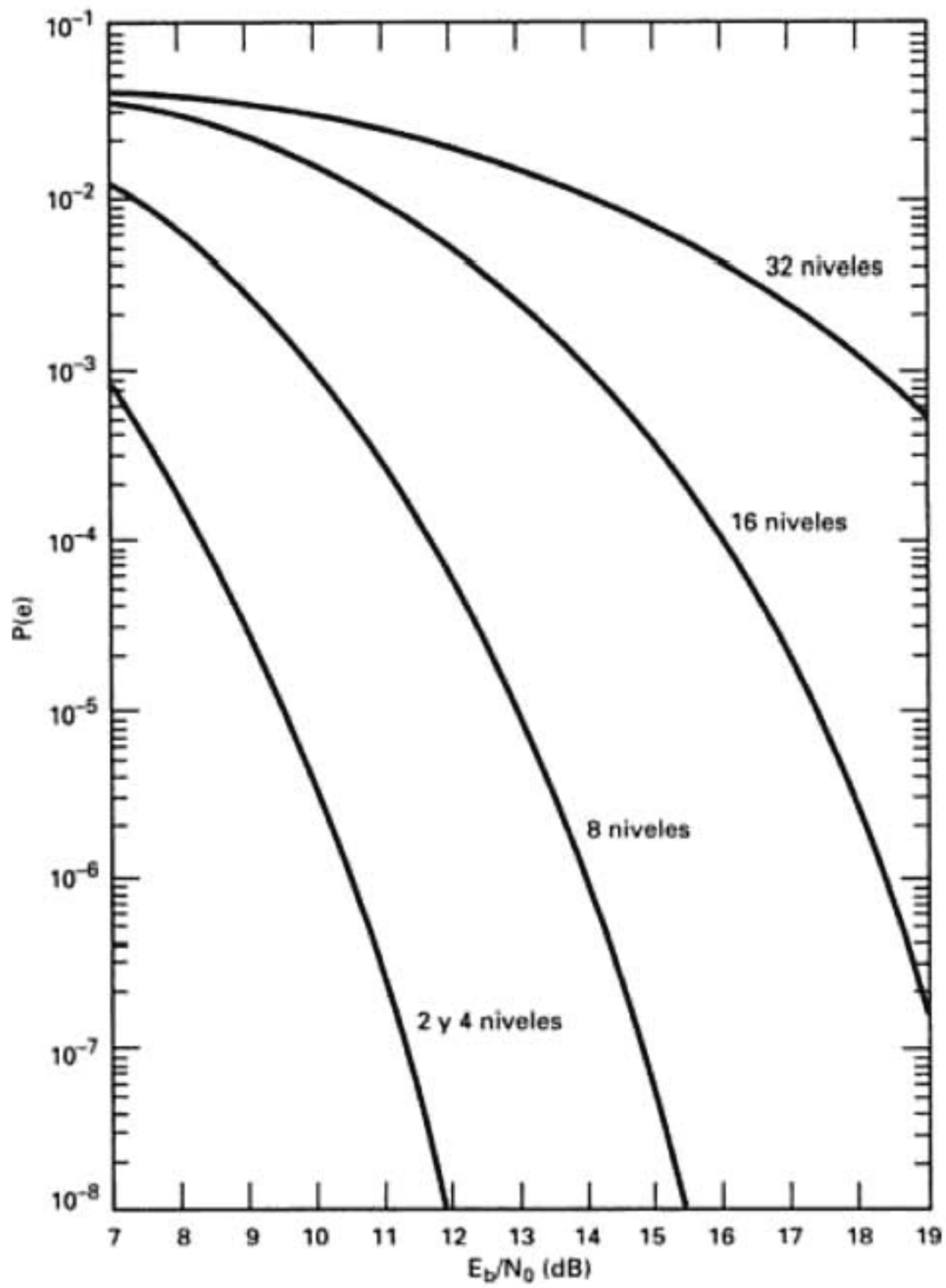


FIGURA 12-46 Tasas de error para sistemas de modulación PSK

La ecuación general de la probabilidad de error de bit, para un sistema PSK M-fásico es

$$P(e) = \frac{1}{\log_2 M} \operatorname{erf}(z) \tag{12-26}$$

en donde erf = función de error

$$z = \operatorname{sen}(\pi/M)(\sqrt{\log_2 M})(\sqrt{E_b/N_0})$$

Al sustituir en la ecuación 12-26 se puede demostrar que el sistema QPSK proporciona igual presencia de errores que BPSK. Esto se debe a que la reducción de 3 dB en la distancia de error para el QPSK se compensa con la disminución de 3 dB en su ancho de banda (además de la distancia entre errores, se deben tener en cuenta también las amplitudes relativas de las bandas de ruido). Así, ambos sistemas tienen la misma eficiencia en cuanto a errores. La fig. 12-46 muestra la presencia de errores para los sistemas de 2, 4, 8, 16 y 32-PSK en función de E_b/N_0.

Ejemplo 12-13

Calcular el ancho mínimo de banda necesario para alcanzar una $P(e) = 10^{-7}$ para un sistema 8-PSK trabajando a 10 Mbps con una relación de potencia de portadora a ruido de 11.7 dB.

Solución De acuerdo con la fig. 12-46, la relación mínima E_b/N_0 para lograr una $P(e)$ de 10^{-7}, para un sistema 8-PSK, es 14.7 dB. El ancho mínimo de banda se calcula reordenando la ecuación 12-22b.

$$\frac{B}{f_b} = \frac{E_b}{N_0} - \frac{C}{N}$$
$$= 14.7 \text{ dB} - 11.7 \text{ dB} = 3 \text{ dB}$$
$$\frac{B}{f_b} = \text{antilog } 3 = 2$$
$$B = 2 \times 10 \text{ Mbps} = 20 \text{ MHz}$$

Errores en QAM

Para una gran cantidad de puntos de señal, es decir, con sistemas M-arios mayores de 4, la QAM es mejor que la PSK. Esto se debe a que la distancia entre los puntos de señalización en un sistema PSK es menor que la distancia entre puntos en un sistema QAM comparable. La ecuación general de la distancia entre los puntos adyacentes de señalización para un sistema QAM con L niveles en cada eje es

$$d = \frac{\sqrt{2}}{L-1} \times D \tag{12-27}$$

en donde d = distancia entre errores
L = cantidad de niveles en cada eje
D = amplitud máxima de la señal

Al comparar las ecuaciones 12-25 y 12-27 se ve que los sistemas QAM aventajan a los sistemas PSK con el mismo nivel máximo de potencia de la señal.

La ecuación general de la probabilidad de error de bit, en un sistema QAM de L niveles es

$$P(e) = \frac{1}{\log_2 L}\left(\frac{L-1}{L}\right)\text{erfc}(z) \tag{12-28}$$

en donde erfc(z) es la función error complementario, y

$$z = \frac{\sqrt{\log_2 L}}{L-1}\sqrt{\frac{E_b}{N_0}}$$

La fig. 12-47 muestra la probabilidad de error en sistemas 4, 16, 32 y 64-QAM, en función de E_b/N_0.

En la tabla 12-2 se ve una lista de relaciones mínimas de potencia de portadora a ruido, y de energía por bit entre densidad de potencia de ruido que se requieren para tener una probabilidad de error de 10^{-6}, para varios esquemas de modulación.

Ejemplo 12-14

¿Cuál sistema requiere la máxima relación E_b/N_0 para obtener una probabilidad de error de 10^{-6}, un sistema QAM de 4 niveles o uno 8-PSK?

TABLA 12-2 Comparación de presencia de errores en diversos esquemas de modulación digital (BER = 10^{-6} de probabilidad de error)

Técnica de modulación	Relación C/N (dB)	Relación E_b/N_0 (dB)
BPSK	10.6	10.6
QPSK	13.6	10.6
4-QAM	13.6	10.6
8-QAM	17.6	10.6
8-PSK	18.5	14
16-PSK	24.3	18.3
16-QAM	20.5	14.5
32-QAM	24.4	17.4
64-QAM	26.6	18.8

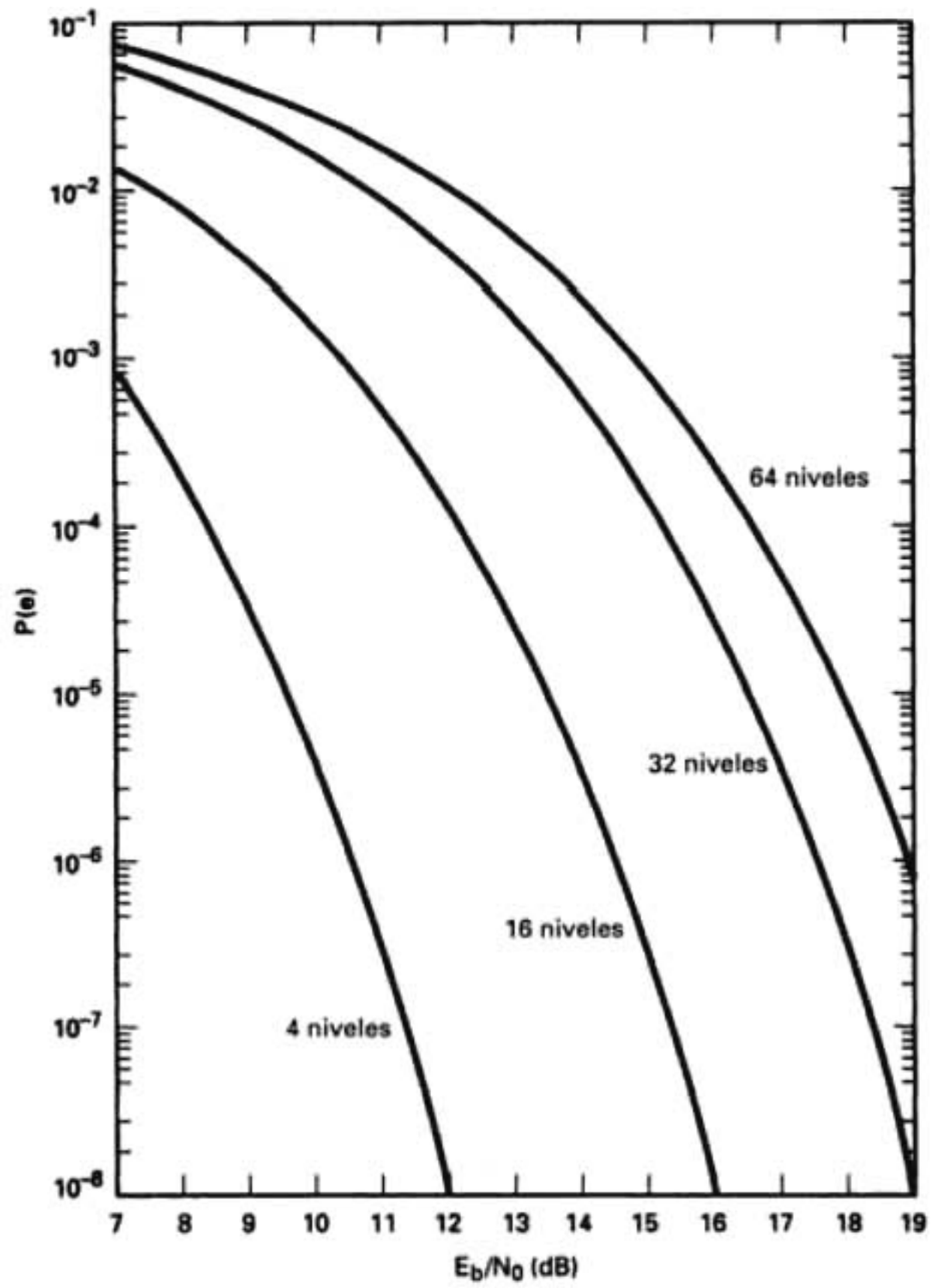


FIGURA 12-47 Tasas de error para sistemas de modulación QAM

Solución De acuerdo con la fig. 12-47, la relación E_b/N_0 mínima necesaria para un sistema QAM de cuatro niveles es 10.6 dB. Según la fig. 12-46, la relación E_b/N_0 mínima necesaria para un sistema 8-PSK es 14 dB. Por consiguiente, para lograr una $P(e)$ de 10^{-6}, un sistema QAM necesitaría la relación E_b/N_0 3.4 dB menor.

Errores en FSK

La probabilidad de error en los sistemas FSK se evalúa en forma algo distinta que en los PSK o QAM. En esencia, sólo hay dos tipos de sistema FSK: no coherente (asíncrono) y coherente (síncrono). En el FSK no coherente, el transmisor y el receptor no tienen sincronización de fase ni de frecuencia. En FSK coherente, las señales de referencia del receptor local están amarradas en fase y frecuencia con las señales transmitidas. La probabilidad de error para FSK no coherente es

$$P(e) = \frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right) \tag{12-29}$$

La probabilidad de error en FSK coherente es

$$P(e) = \operatorname{erfc}\sqrt{\frac{E_b}{N_0}} \tag{12-30}$$

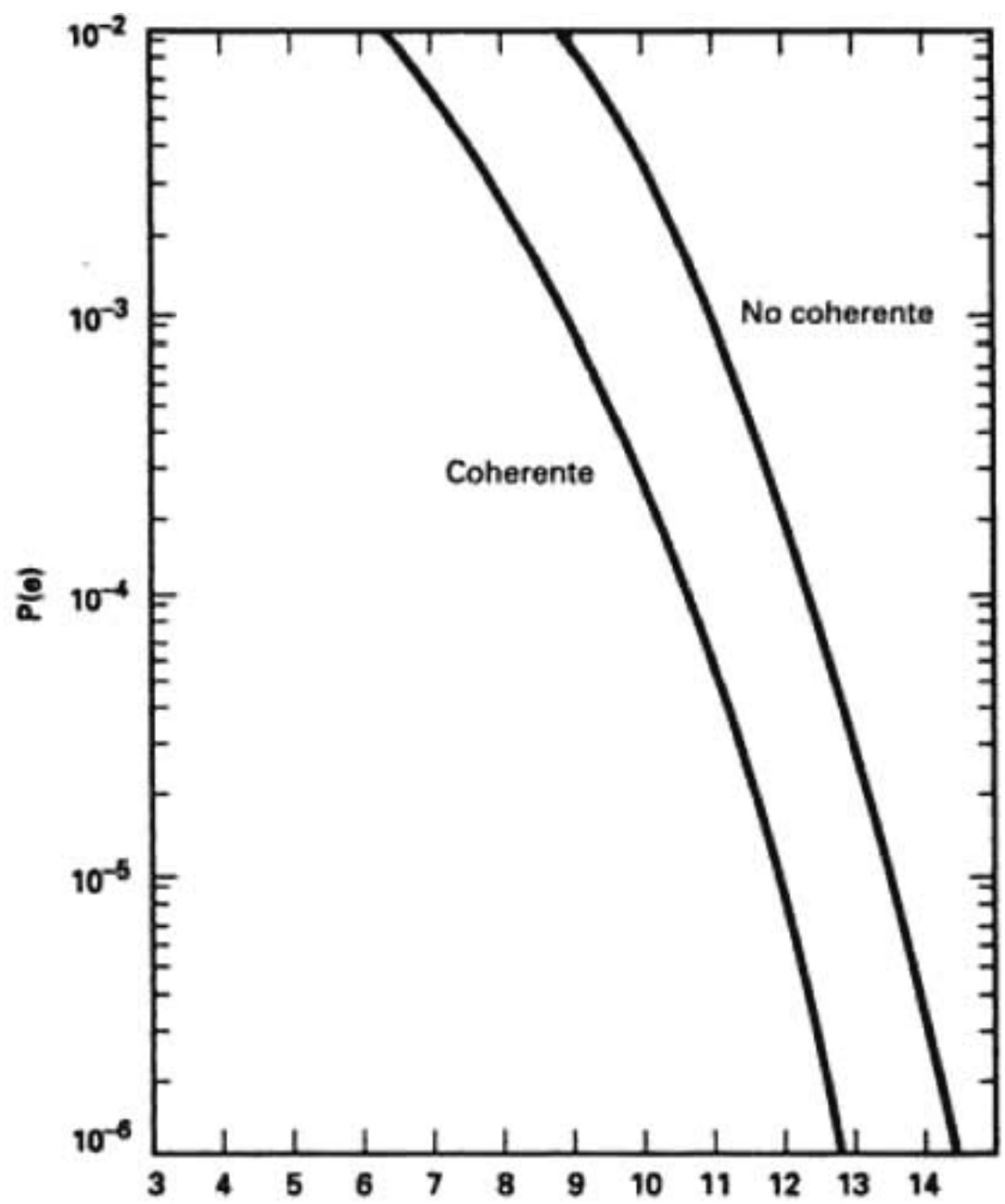


FIGURA 12-48 Tasas de error para sistemas de modulación FSK

La fig. 12-48 muestra las curvas de probabilidad de error para FSK coherente y no coherente, en función de valores de E_b/N_0. Se puede calcular, con las ecuaciones 12-29 y 12-30, que la probabilidad de error para FSK no coherente es mayor que para FSK coherente de igual relación de energía por bit a potencia de ruido.

CODIFICACIÓN TRELLIS

Para lograr rapidez de transmisión mayor que 9600 bps en los canales telefónicos normales de banda de voz se requiere, claro está, un esquema de codificación mucho más allá de los cuadribits que se usan en el 16-PSK o el 16-QAM, es decir, M debe ser bastante mayor que 16. Como cabría esperar, los mayores esquemas de codificación requieren mayores relaciones de señal a ruido, S/N. Si se aplica el límite de Shannon de capacidad de información (ecuación 12-1c), una rapidez de transmisión de datos de 28.8 kbps que use un canal de comunicaciones con 3200 Hz de ancho de banda requiere una relación de señal a ruido de

$$I(\text{bps}) = (3.32 \times B)\log(1 + S/N)$$

por consiguiente,

$$28.8 \text{ kbps} = (3.32)(3200)\log(1 + S/N)$$

$$28{,}800 = 10{,}624\log(1 + S/N)$$

$$\frac{28{,}800}{10{,}624} = \log(1 + S/N)$$

$$2.71 = \log(1 + S/N)$$

así que

$$10^{2.71} = 1 + S/N$$

$$513 = 1 + S/N$$

$$512 = S/N$$

como errores de transmisión. Así, la TCM puede mejorar el funcionamiento en lo que a error se refiere restringiendo la manera en la que se permite transición de señales. Para valores de N mayores que 2, el QAM es el esquema de modulación que se escoge, frente al TCM; sin embargo, para simplificar, se usará la manipulación PSK para explicar, porque es más fácil de ilustrar.

La fig. 12-50 muestra un esquema TCM que usa 8-PSK de dos estados, que en esencia son dos constelaciones QPSK compensadas 45°. Una constelación de cuatro estados se identifica con 0-4-2-6, y la otra con 1-5-3-7. Para esta explicación, las identificaciones 0 a 7 de puntos de señal no son para representar las condiciones reales de datos, sino más bien para indicar un método cómodo de identificar los diversos puntos de señal. Cada dígito representa uno de cuatro puntos de señal permitidos dentro de cada una de las dos constelaciones QPSK. Cuando se está en la constelación 0-4-2-6 y se transmite un 0 o 4, el sistema permanece en la misma constelación. Sin embargo, cuando se transmite un 2 o un 6, el sistema cambia a la constelación 1-5-3-7. Una vez en esa constelación y cuando se transmite un 3 o un 7, el sistema permanece en ella, y si se transmite un 1 o un 5, el sistema pasa a la constelación 0-4-2-6. Recuérdese que cada símbolo representa a dos bits, por lo que el sistema sufre un desplazamiento de fase de 45° siempre que cambia de una a otra constelación. Un análisis completo de errores de QPSK normal en comparación con QPSK con TCM indicaría una ganancia de codificación, para TCM de 2 a 1, 1 a 3 dB. La tabla 12-3 muestra la lista de las ganancias de codificación logradas con esquemas de codificación TCM con varios estados distintos de Trellis.

La rapidez máxima de datos alcanzable con determinado ancho de banda se puede calcular rearreglando la ecuación 12-13b.

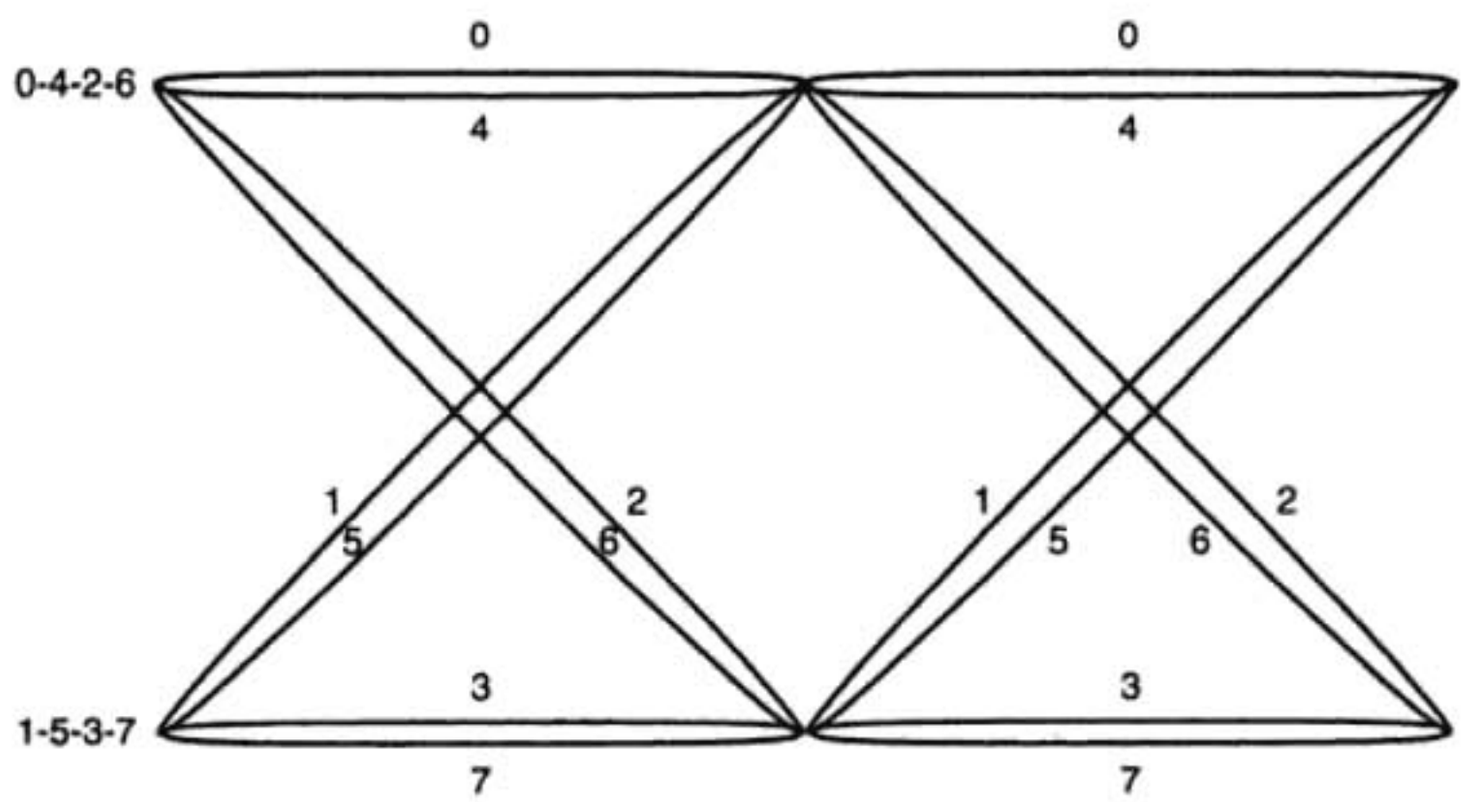


FIGURA 12-50 Constelaciones TCM de 8-PSK

TABLA 12-3 Ganancia en codificación de Trellis

Cantidad de estados de Trellis	Ganancia de codificación (dB)
2	3.0
4	5.5
8	6.0
16	6.5
32	7.1
64	7.3
128	7.3
256	7.4

FIGURA 12-51 Constelación TCM de QAM, de 32 puntos

$$N \times B = f_b$$

en donde N = cantidad de bits codificados (bits)
B = ancho de banda (hertz)
f_b = rapidez de transmisión de bits (bits por segundo)

Recuérdese que con sistemas QAM o PSK M-arios, los baudios son igual al ancho de banda mínimo requerido. Por consiguiente, un ancho de banda de 3200 Hz con código Trellis de nueve bits produce una señal de 3200 baudios y cada baudio lleva nueve bits. En consecuencia, la rapidez de transmisión es $f_b = 9 \times 3200 = 28.8$ kbps.

Se tiene la idea de que la modulación TCM es un esquema de codificación que mejora la QAM normal. El primer sistema TCM usaba un código de cinco bits, que incluía cuatro bits QAM (un cuadribit) y un quinto bit que se usaba para ayudar a decodificar el cuadribit. Para transmitir cinco bits dentro de un solo elemento de señalización se requiere producir 32 señales discernibles. La fig. 12-51 muestra una constelación QAM de 32 puntos, que usa cuatro amplitudes y ocho fases para producir los 32 estados discretos de señal.

Una señal de 3200 baudios que use una codificación TCM de nueve bits produce 512 códigos distintos. Los nueve bits de datos más un bit redundante para codificación TCM requieren una constelación de 960 puntos. La fig. 12-52 ilustra la cuarta parte de la superconstelación de 960 puntos, que muestra 240 puntos de señal. Toda la superconstelación se puede obtener girando 0°, 90°, 180° y 270° los 240 puntos que se ven.

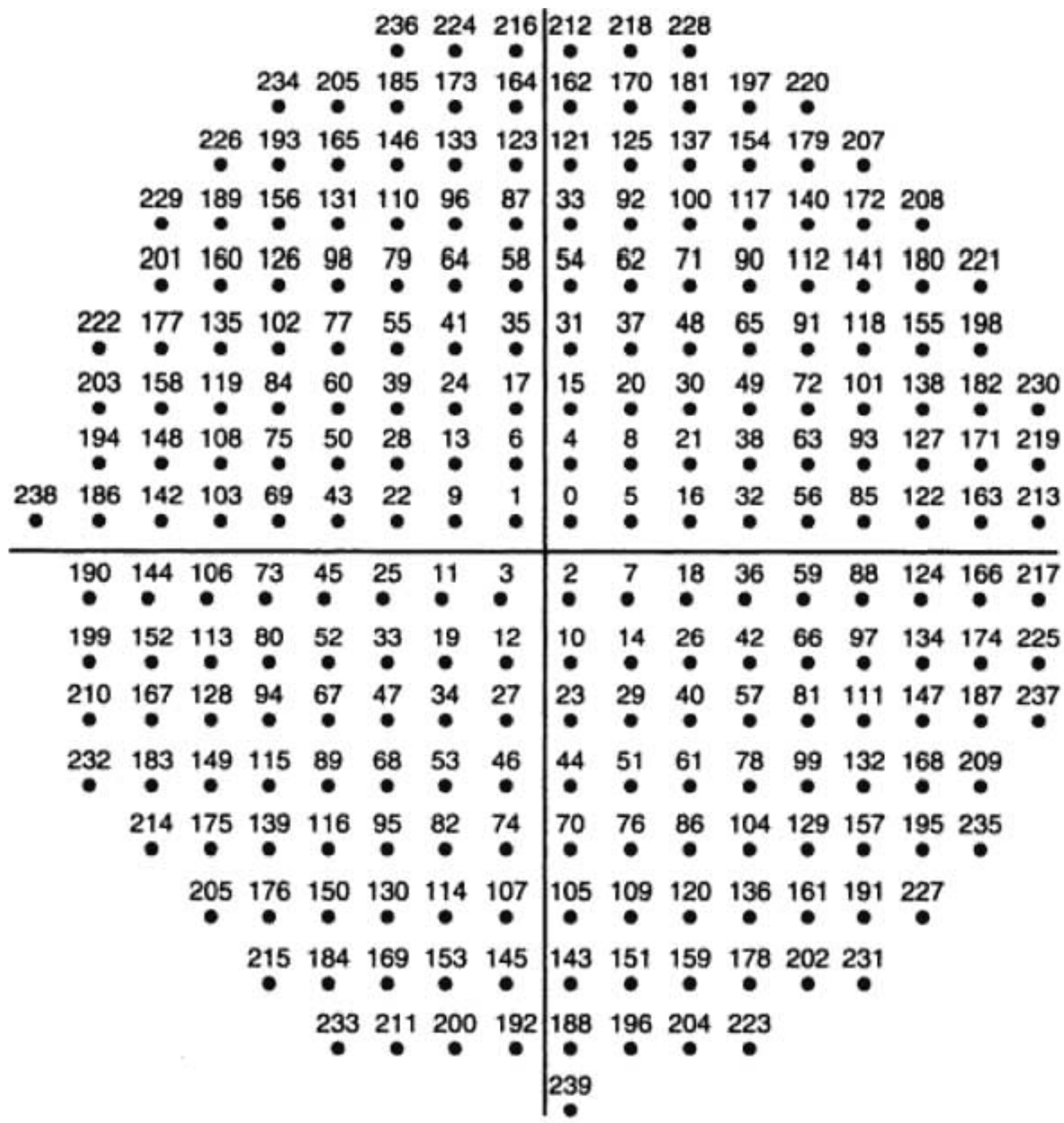


FIGURA 12-52 La cuarta parte de una constelación TCM de QAM, con 960 puntos

PREGUNTAS

12-1. Explique la *transmisión digital* y la *radio digital*.

12-2. Defina la *capacidad de información*.

12-3. ¿Cuáles son los tres esquemas de modulación predominantes que se usan en los sistemas de radio digital?

12-4. Explique la relación entre bits por segundo y baudios para un sistema FSK.

12-5. Defina los siguientes términos de modulación FSK: *desviación de frecuencia, índice de modulación* y *relación de desviación*.

12-6. Explique la relación entre a) el ancho mínimo de banda necesario para un sistema FSK y la rapidez de bits, y b) las frecuencias de marca y espacio.

12-7. ¿Cuál es la diferencia entre FSK normal y MSK? ¿Cuál es la ventaja de MSK?

12-8. Defina *PSK*.

12-9. Explique la relación entre bits por segundo y baudios para un sistema BPSK.

12-10. ¿Qué es un diagrama de constelación, y cómo se usa con PSK?

12-11. Explique la relación entre el ancho mínimo de banda necesario para un sistema BPSK y la rapidez de bits.

12-12. Explique qué es *M*-ario.

12-13. Explique la relación entre bits por segundo y baudios para un sistema QPSK.

12-14. Explique la importancia de los canales I y Q en un modulador QPSK.

12-15. Defina qué es *dibit*.

12-16. Explique la relación entre el ancho mínimo de banda requerido para un sistema QPSK y la rapidez de bits.

12-17. ¿Qué es un desmodulador coherente?

12-18. ¿Qué ventaja tiene OQPSK sobre la QPSK convencional? ¿Cuál es la desventaja de la manipulación OQPSK?

12-19. Explique la relación entre bits por segundo y baudios, para un sistema 8-PSK.

12-20. Defina qué es un *tribit*.

12-21. Explique la relación entre el ancho mínimo de banda necesario para un sistema 8-PSK y la rapidez de bits.

12-22. Explique la relación entre bits por segundo y baudios para un sistema 16-PSK.

12-23. Defina qué es un *cuadribit*.

12-24. Defina la *QAM*.

12-25. Explique la relación entre el ancho mínimo de banda necesario para un sistema 16-QAM y la rapidez de bits.

12-26. ¿Cuál es la diferencia entre PSK y QAM?

12-27. Defina la *eficiencia de ancho de banda.*

12-28. Defina qué es *recuperación de portadora.*

12-29. Explique las diferencias entre PSK absoluto y PSK diferencial.

12-30. ¿Cuál es el objetivo de un circuito recuperador de reloj? ¿Cuándo se usa?

12-31. ¿Cuál es la diferencia entre probabilidad de error y la frecuencia de errores de bits?

PROBLEMAS

12-1. Calcule el ancho de banda y los baudios para una señal FSK con frecuencia de marca 32 kHz, frecuencia de espacio 24 kHz y rapidez de bits de 4 kbps.

12-2. Determine la máxima rapidez de bits para una señal FSK con 48 kHz de frecuencia de marca, 52 kHz de frecuencia de espacio y 10 kHz de ancho de banda disponible.

12-3. Calcule el ancho de banda y los baudios para una señal FSK con 99 kHz de frecuencia de marca, 101 kHz de frecuencia de espacio y rapidez de bits de 10 kbps.

12-4. Calcule la máxima rapidez de bits para una señal FSK con 102 kHz de frecuencia de marca, 104 kHz de frecuencia de espacio y 8 kHz de ancho de banda disponible.

12-5. Determine el ancho de banda mínimo y los baudios para un modulador BPSK con 40 MHz de frecuencia de portadora y 500 kbps de rapidez de entrada de bits. Haga un esquema del espectro de salida.

12-6. Para el modulador QPSK de la fig. 12-19, cambie la red de deplazamiento de fase de $+90°$ a una de $-90°$ y trace el nuevo diagrama de constelación.

12-7. Para el desmodulador QPSK de la fig. 12-23, determine los bits I y Q para una señal de entrada sen $\omega_c t - \cos \omega_c t$.

12-8. Calcule el ancho mínimo de banda bilateral de Nyquist, F_N, y los baudios, para un modulador 8-PSK con 20 Mbps de rapidez de entrada de bits, f_b, y 100 MHz de frecuencia de portadora. Trace un esquema del espectro de salida.

12-9. En el modulador 8-PSK de la fig. 12-25, cambie el oscilador de referencia a $\cos \omega_c t$ y trace el nuevo diagrama de constelación.

12-10. Para un modulador 16-QAM con 20 Mbps de rapidez de entrada de bits (f_b) y 100 MHz de frecuencia de portadora, determine el ancho mínimo de banda bilateral de Nyquist, f_N, y los baudios. Trace un esquema del espectro de salida.

12-11. En el modulador 16-QAM de la fig. 12-35, cambie el oscilador de referencia a $\cos \omega_c t$ y determine las ecuaciones de salida con lo siguiente: condiciones de entrada de I, I', Q y Q': 0000, 1111, 1010 y 0101.

12-12. Determine la eficiencia de ancho de banda en los siguientes moduladores:

(a) QPSK, $f_b = 10$ Mbps.
(b) 8-PSK, $f_b = 21$ Mbps.
(c) 16-QAM, $f_b = 20$ Mbps.

12-13. Determine la secuencia de fase de salida para la siguiente secuencia de bits de entrada: 00110011010101, para el modulador DBPSK de la fig. 12-42a. Suponga que el bit de referencia es 1.

12-14. Para un sistema QPSK con los parámetros mencionados abajo, determine:

(a) Potencia de portadora, en dBm.
(b) Potencia de ruido, en dBm.
(c) Densidad de potencia de ruido, en dBm.
(d) Energía por bit, en dBJ.
(e) Relación de potencia de portadora a ruido.
(f) Relación E_b/N_0.

$C = 10^{-13}$W $\quad f_b = 30$ kbps
$N = 0.06 \times 10^{-15}$W $\quad B = 60$ kHz

12-15. Calcule el ancho mínimo de banda requerido para lograr una $P(e)$ de 10^{-6} para un sistema 8-PSK que funciona a 20 Mbps, con una relación de potencia de portadora a ruido de 11 dB.

12-16. Calcule el ancho mínimo de banda y los baudios para un modulador BPSK con frecuencia de portadora de 80 MHz y rapidez de entrada de bits $f_b = 1$ Mbps. Trace el espectro de salida.

12-17. En el modulador QPSK de la fig. 12-19, cambie el oscilador de referencia a cos $\omega_c t$ y trace el nuevo diagrama de constelación.

12-18. Para el desmodulador QPSK de la fig. 12-23, determine los bits I y Q para una señal de entrada $-\text{sen}\ \omega_c t + \cos \omega_c t$.

12-19. Para un modulador 8-PSK con rapidez de entrada de bits $f_b = 10$ Mbps y frecuencia de portadora $f_c = 80$ MHz, calcule el ancho mínimo de banda de Nyquist y los baudios. Trace un esquema del espectro de salida.

12-20. Para el modulador 8-PSK de la fig. 12-25, cambie el circuito de desplazamiento de fase de $+90°$ a un cambiador de fase de $-90°$, y trace el nuevo diagrama de constelación.

12-21. Para un modulador 16-QAM con rapidez de entrada de bits $f_b = 10$ Mbps y frecuencia de portadora $f_c = 60$ MHz, calcule la frecuencia bilateral mínima de Nyquist y los baudios. Trace un esquema del espectro de salida.

12-22. En el modulador 16-QAM de la fig. 12-35, cambie la red de desplazamiento de fase de 90° por un desplazador de fase de $-90°$ y determine las ecuaciones de salida de las siguientes condiciones de entrada I, I', Q y Q': 0000, 1111, 1010 y 0101.

12-23. Calcule la eficiencia de ancho de banda para los siguientes moduladores:

(a) QPSK, $f_b = 20$ Mbps.
(b) 8-PSK, $f_b = 28$ Mbps.
(c) 16-PSK, $f_b = 40$ Mbps.

12-24. Para el modulador DBPSK de la fig. 12-42, determine la secuencia de fase de salida para la siguiente secuencia de bits de entrada: 11001100101010. Suponga que el bit de referencia es un 1 lógico.

C A P Í

Imagen protegida por derechos de autor

Comunicación de datos

INTRODUCCIÓN

En el mundo de las *comunicaciones de datos*, se define *datos* en general como información que se almacena en forma digital. La palabra *datos* es plural; una sola unidad de datos se llama *dato*. La *comunicación de datos* es el proceso de transferir *información* digital, por lo general, en forma binaria, entre dos o más puntos. Se define la información como conocimiento, noticia o información secreta. La información que se ha procesado, organizado y guardado se llama datos. Los datos pueden ser de naturaleza alfabética, numérica o simbólica, y están formados por cualquiera de los siguientes símbolos, o una combinación de ellos: alfanuméricos codificados en binario, programas de microprocesador, códigos de control, direcciones de usuarios, datos de programa o información de base de datos. Tanto en la fuente como en el destino, los datos están en forma digital; sin embargo, durante la transmisión, pueden estar en forma digital o analógica.

Una red de comunicación de datos puede ser tan simple como dos computadoras personales conectadas a través de una red pública de telecomunicaciones, o puede abarcar una red compleja de una o más computadoras centrales y cientos, o hasta miles, de terminales remotas, computadoras personales y estaciones de trabajo. Hoy, las redes de comunicación de datos se usan para interconectar casi todas las clases de equipos de cómputo digital, como por ejemplo los cajeros automáticos con las computadoras de los bancos, las computadoras personales con las carreteras de información como Internet, y estaciones de trabajo con computadoras centrales. Las redes de comunicaciones de datos también se usan en sistemas de reservaciones de aerolíneas y de hoteles, y para medios masivos y redes noticiosas, como la Associated Press (AP) o la United Press International (UPI). La lista de aplicaciones para las redes de comunicación de datos aumenta casi en forma infinita.

HISTORIA DE LA COMUNICACIÓN DE DATOS

Es muy probable que las comunicaciones de datos comenzaron mucho antes del registro histórico, en forma de señales de humo o de batir del tambor, aunque no es probable que estuvieran en código binario. Si se limita el propósito de la comunicación de datos a métodos que usan señales eléctricas para transmitir información codificada en binario, la comunicación de datos

comenzó en 1837, con el invento del *telégrafo* y el desarrollo de la *clave de Morse*, por Samuel F. B. Morse. Con el telégrafo se transmiten puntos y rayas (equivalentes a los 1 y 0 binarios) por un hilo, usando la inducción electromecánica. Se usaron varias combinaciones de esos puntos y rayas para representar los códigos binarios de letras, números y signos de puntuación. En realidad, el primer telégrafo fue inventado en Inglaterra, por Sir Charles Wheatstone y Sir William Cooke, pero su sistema necesitaba seis hilos para una sola línea telegráfica. En 1840 Morse obtuvo una patente en Estados Unidos para su telégrafo, y en 1844 se estableció la primera línea telegráfica entre Baltimore y Washington, D. C. En 1849 se inventó la primera impresora telegráfica de baja velocidad, pero no fue sino hasta 1860 que se pudo disponer de impresoras de alta velocidad (15 bps). En 1850 se fundó la Western Union Telegraph Company en Rochester, New York, para enviar mensajes codificados entre dos personas.

En 1874 Emile Baudot inventó un *multiplexor* telegráfico, que permitió transmitir al mismo tiempo señales hasta de seis máquinas telegráficas por un solo hilo. El teléfono fue inventado en 1876 por Alexander Graham Bell, y después se hicieron muy pocos avances hasta 1899, cuando Marconi pudo mandar mensajes radiotelegráficos. El telégrafo fue el único medio de mandar información a través de grandes extensiones de agua, hasta 1920, cuando se instalaron las primeras estaciones comerciales de radio.

Los Laboratorios Bell desarrollaron la primera computadora especial en 1940, con relevadores electromecánicos. La primera computadora de propósito general fue una calculadora automática controlada por secuencia, desarrollada en conjunto por la Universidad de Harvard y la International Business Machines (IBM) Corporation. La computadora UNIVAC, construida en 1951 por Remington Rand Corporation (ahora Sperry Rand) fue la primera computadora electrónica producida en gran escala. Desde 1951 la cantidad de computadoras centrales, computadoras comerciales pequeñas, computadoras personales y terminales de computadora ha aumentado en forma exponencial, dando lugar a una situación en la que cada vez más personas tienen la necesidad de intercambiar información digital entre sí. En consecuencia, también ha aumentado en forma exponencial la necesidad de comunicar datos.

Hasta 1968 la tarifa de operación de AT&T sólo permitía conectarse con sus líneas con equipo suministrado por AT&T. En 1968 se emitió una histórica decisión en la Suprema Corte de Estados Unidos, la decisión Carterfone, que permitió a empresas distintas a Bell interconectarse con la vasta red de comunicaciones de AT&T. Esta decisión dio comienzo a la *industria de la interconexión*, que ha permitido tener ofertas competitivas para comunicaciones, por parte de una gran cantidad de empresas independientes.

ORGANIZACIONES NORMATIVAS PARA COMUNICACIÓN DE DATOS

Durante la última década la industria de comunicación de datos creció a una tasa astronómica. En consecuencia, también aumentó la necesidad de proporcionar comunicaciones entre sistemas distintos de cómputo. Así, para asegurar una transferencia ordenada de información entre dos o más sistemas de comunicación de datos que usen distintos equipos para distintas necesidades, se reúne un consorcio de organizaciones, fabricantes y usuarios, con regularidad, para establecer lineamientos y normas. Se trata de que todos los usuarios de comunicación de datos se apeguen a esas normas. A continuación se describen algunas de esas organizaciones.

La *Organización Internacional de Normas* (ISO). La ISO es la organización internacional de normalización. Establece los conjuntos de reglas y normas para tecnologías gráficas, de intercambio de documentos y otras afines. Es responsable de patrocinar y coordinar el trabajo de las demás organizaciones de normas.

Comité Consultivo Internacional Telefónico y Telegráfico (CCITT, por *Consultative Committe for International Telephony and Telegraphy*). La membresía del CCITT es de autoridades oficiales y representantes de muchos países. Hoy, la CCITT es la organización de normas en las Naciones Unidas, y desarrolla los conjuntos recomendados de reglas y normas para comunicaciones telefónicas y telegráficas. La CCITT ha desarrollado tres conjuntos de especificaciones: la serie V para interconectar módems, la serie X para

comunicaciones de datos y las series I y Q, para los Servicios Integrados de Red Digital (ISDN, de *Integrated Services Digital Network*).

American National Standards Institute (ANSI). El ANSI es la agencia oficial de normas en Estados Unidos, y es el representante de ese país ante la ISO.

Institute of Electrical and Electronics Engineers (IEEE). Es una organización de profesionistas, en Estados Unidos, de ingenieros en electrónica, informática y comunicaciones.

Asociación de Industrias Electrónicas (EIA, de *Electronics Industries Association*). Es una organización estadounidense que establece y recomienda normas industriales. La EIA es responsable de desarrollar la serie RS (*recommended standard*, norma recomendada) de normas para comunicaciones de datos y telecomunicaciones.

Standards Council of Canada (SCC). Es la agencia oficial de normas en Canadá, con responsabilidades parecidas a las de ANSI.

CIRCUITOS DE COMUNICACIÓN DE DATOS

La fig. 13-1 muestra un diagrama simplificado de bloques de una red de comunicaciones de datos. Como allí se ve, hay una fuente de información digital (estación *primaria*), un medio de transmisión (instalación) y un destino (estación *secundaria*). La estación *primaria* (o anfitriona) es, con mucha frecuencia, una computadora central con su propio juego de terminales locales y equipo periférico. Para simplificar, en la figura sólo hay una estación secundaria (o remota). Las estaciones secundarias son los usuarios de la red. La cantidad de estaciones secundarias que haya, y la forma en que se interconectan con ellos y con la estación primaria varía mucho, dependiendo del sistema y sus aplicaciones. Hay muchas clases distintas de medios de transmisión, que incluyen la radiotransmisión por el espacio libre (terrestre y satelital por microondas), instalaciones de cables metálicos (en sistemas tanto analógicos como digitales) y cables de fibra óptica (propagación de ondas luminosas).

El *equipo de terminal de datos* (DTE, de *data terminal equipment*) es un término general que describe al equipo de interconexión que se usa en las estaciones para adaptar las señales digitales de las computadoras y las terminales, a una forma más adecuada para su transmisión. En esencia, cualquier equipo entre la computadora central y el módem o el equipo de la estación y su módem, se clasifica como equipo de terminal de datos. El *equipo de comunicación de datos* (DCE, por *data communications equipment*) es un término general que describe al equipo que convierte las señales digitales en analógicas, e interconecta al equipo de terminal de datos con el medio de transmisión analógica. En esencia, un equipo de comunicación de datos es un *módem* (*mod*ulador/*dem*odulador). Un módem convierte señales digitales binarias en señales analógicas, como FSK, PSK y QAM, y viceversa.

Transmisión de datos en serie y en paralelo

La información binaria se puede transmitir en forma paralela o en serie ("*serial*"). La fig. 13-2a muestra cómo se transmite el código binario 0110 del lugar A al lugar B, en paralelo. Como allí se ve, cada posición de bit (A_0 a A_3) tiene su propia línea de transmisión. En consecuencia, los cuatro bits se pueden transmitir en forma simultánea durante el tiempo de un solo pulso del reloj (T). A esta clase de transmisión se le llama *paralela a nivel de bit* o *en serie a nivel de carácter*.

La fig. 13-2b muestra cómo se transmite el mismo código binario en serie. Como se ve, hay una sola línea de transmisión y, en consecuencia, sólo se puede transmitir un bit cada vez. Por lo anterior, se requieren cuatro pulsos de reloj ($4T$) para transmitir toda la palabra. A esta clase de transmisión se le llama con frecuencia *en serie a nivel de bit*.

Es obvio que la principal opción de alternativas entre la transmisión en serie y en paralelo es entre velocidad y sencillez. La transmisión de datos se puede hacer con mucha mayor rapidez usando el sistema en paralelo; sin embargo, en él se requieren más líneas entre la fuente y el destino. Por regla general, se usa transmisión en paralelo para comunicaciones en distancias cortas, y dentro de una computadora. La transmisión en serie se usa para comunicaciones a gran distancia.

Los *cuatro hilos* implican un medio de transmisión que usa cuatro conductores (dos para señales que se propagan en direcciones opuestas, y dos son conductores de referencia), o una configuración que equivale a tener cuatro conductores. En la operación con cuatro hilos las señales que se propagan en direcciones opuestas están separadas físicamente y, por consiguiente, pueden ocupar los mismos anchos de banda sin interferirse. La operación con cuatro hilos permite más aislamiento, y se prefiere a la de dos hilos, aunque requiere el doble de conductores y, en consecuencia, el doble de costo en conductores.

Un transmisor y su receptor asociado equivalen a un circuito de dos hilos. Un transmisor y un receptor con ambas direcciones de propagación equivalen a un circuito de cuatro hilos. Con la transmisión dúplex por una línea de dos hilos, se debe dividir a la mitad el ancho de banda disponible, con lo cual se reduce la capacidad de información en ambas direcciones, a la mitad del valor del semidúplex. En consecuencia, la operación dúplex con líneas de dos hilos requiere el doble de tiempo para transferir la misma cantidad de información.

CÓDIGOS DE COMUNICACIÓN DE DATOS

Los *códigos de comunicación de datos* son secuencias predeterminadas de bits, para codificar caracteres y símbolos. En consecuencia, a las claves de comunicación de datos se les llama con frecuencia *conjunto de caracteres, códigos de caracteres, claves de caracteres, códigos de símbolo* o *lenguajes de caracteres*. En esencia, sólo hay tres clases de caracteres que se usan en las claves de comunicación de datos: *caracteres de control de eslabón de datos*, que se usan para facilitar el flujo ordenado de datos desde una fuente hasta un destino; *caracteres de control gráfico*, que implican la sintaxis o presentación de datos en la terminal de recepción, y los caracteres *alfa/numéricos*, para representar los diversos símbolos empleados en las letras, números y signos de puntuación.

El primer código de comunicación de datos que se difundió mucho fue la clave Morse. Esta clave usaba tres símbolos de longitud diferente (punto, raya y espacio) para representar los caracteres alfanuméricos, signos de puntuación y una palabra de interrogación.

La clave de Morse es inadecuada para el moderno equipo de computadoras digitales, porque todos los caracteres no tienen la misma cantidad de símbolos ni ocupan el mismo tiempo de transmisión, y cada operador de clave Morse transmite a una velocidad distinta. También, la clave Morse tiene un surtido insuficiente de caracteres de control gráfico y de enlace, que faciliten la transmisión y la presentación normal de los datos que se usan en las aplicaciones modernas de cómputo.

Los tres conjuntos más comunes de caracteres que se usan en la actualidad para codificar caracteres son el código de Baudot, el ASCII (de *American Standard Code for Information Interchange*) y el EBCDIC (de *Extended-Binary-Coded Decimal Interchange Code*).

Código de Baudot

El código de Baudot, llamado a veces *código Télex*, fue el primer código de caracteres de longitud fija. Fue desarrollado en 1875 por Thomas Murray, ingeniero postal francés, y recibe el nombre de Emile Baudot, un precursor de la impresión telegráfica. Es un código de caracteres de cinco bits, que se usa principalmente en equipos de teletipo de baja velocidad, como por ejemplo el sistema TWX/Télex. Con un código de cinco bits sólo hay 2^5, o 32, combinaciones posibles, insuficientes para representar las 26 letras del alfabeto, los 10 dígitos y los diversos signos de puntuación y caracteres de control. En consecuencia, en el código Baudot se usan caracteres de paso a *figuras* y de paso a *letras*, para aumentar su capacidad a 58 caracteres. La CCITT recomienda a la última versión del código de Baudot como alfabeto internacional No. 2. La Western Union Company sigue usando el código de Baudot para los sistemas TWX y Telex, de teletipo. Los servicios noticiosos de la AP y la UPI usaron durante muchos años el código Baudot para mandar noticias por todo el mundo. La versión más reciente de este código se ve en la tabla 13-1.

TABLA 13-1 Código de Baudot

Mayúsculas		Código binario				
Letra	Figura	Bit: 4	3	2	1	0
A	—	1	1	0	0	0
B	?	1	0	0	1	1
C	:	0	1	1	1	0
D	$	1	0	0	1	0
E	3	1	0	0	0	0
F	!	1	0	1	1	0
G	&	0	1	0	1	1
H	#	0	0	1	0	1
I	8	0	1	1	0	0
J	'	1	1	0	1	0
K	(	1	1	1	1	0
L	)	0	1	0	0	1
M	.	0	0	1	1	1
N	,	0	0	1	1	0
O	9	0	0	0	1	1
P	0	0	1	1	0	1
Q	1	1	1	1	0	1
R	4	0	1	0	1	0
S	bel	1	0	1	0	0
T	5	0	0	0	0	1
U	7	1	1	1	0	0
V	;	0	1	1	1	1
W	2	1	1	0	0	1
X	/	1	0	1	1	1
Y	6	1	0	1	0	1
Z	"	1	0	0	0	1
Mayúscula de signo		1	1	1	1	1
Mayúscula de letra		1	1	0	1	1
Espacio		0	0	1	0	0
Avance de renglón (LF)		0	1	0	0	0
Blanco (nulo)		0	0	0	0	0

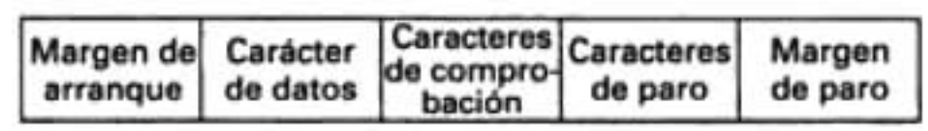
Código ASCII

En 1963, tratando de normalizar los códigos de comunicación de datos, en Estados Unidos se adoptó el código de teletipo modelo 33, de Bell System, como el código patrón en ese país, para intercambio de información (USASCII, *United States of America Standard Code for Information Interchange*), que se conoce más como ASCII-63. Desde su adopción, el ASCII ha evolucionado en sus versiones de 1965, 1967 y 1977; la CCITT recomienda la versión de 1977 como alfabeto internacional No. 5. El ASCII es un conjunto de caracteres de siete bits, y tiene 2^7, o 128 combinaciones. Con el ASCII, el bit menos significativo se llama b_0, y el más significativo se llama b_6. El b_7 no es parte del código ASCII, sino se reserva, en general, como bit de paridad, que se explicará más adelante en este capítulo. En realidad, con cualquier conjunto de caracteres, todos los bits son igualmente significativos, porque la clave no representa a un número binario ponderado. Se acostumbra, en los códigos de caracteres, citar los bits por su orden; b_0 es el bit de orden cero, b_1 es el bit de primer orden, b_7 es el bit de séptimo orden, etcétera. En la transmisión en serie, al bit que se transmite primero se le llama menos significativo (LSB, de *least-significant bit*). En el ASCII, el bit de bajo orden, b_0, es el LSB y se transmite primero. Es probable que el ASCII sea el código que más se usa hoy. En la tabla 13-2 se presenta el código ASCII, en su versión de 1977.

TABLA 13-2 (*Continuación*)

	Código binario										Código binario								
Bit:	7	6	5	4	3	2	1	0	Hex	Bit:	7	6	5	4	3	2	1	0	Hex
9	1	0	1	1	1	0	0	1	39	y	0	1	1	1	1	0	0	1	79
:	1	0	1	1	1	0	1	0	3A	z	0	1	1	1	1	0	1	0	7A
;	0	0	1	1	1	0	1	1	3B	{	1	1	1	1	1	0	1	1	7B
<	1	0	1	1	1	1	0	0	3C	|	0	1	1	1	1	1	0	0	7C
=	0	0	1	1	1	1	0	1	3D	}	1	1	1	1	1	1	0	1	7D
>	0	0	1	1	1	1	1	0	3E	~	1	1	1	1	1	1	1	0	7E
?	1	0	1	1	1	1	1	1	3F	DEL	0	1	1	1	1	1	1	1	7F

NUL = nulo
SOH = inicio de encabezado
STX = inicio de texto
ETX = fin de texto
EOT = fin de transmisión
ENQ = pregunta
ACK = reconocer
BEL = timbre
BS = retroceder
HT = tabulador horizontal
NL = nuevo renglón
VT = tabulador vertical
FF = avance de forma
CR = retorno de carro
SO = quitar mayúsculas
SI = poner mayúsculas
DLE = escape de enlace de datos
DC1 = control de dispositivo 1
DC2 = control de dispositivo 2
DC3 = control de dispositivo 3
DC4 = control de dispositivo 4
NAK = reconocimiento negativo
SYN = síncrono
ETB = fin del bloque de transmisión
CAN = cancelar
SUB = sustituir
ESC = escapar
FS = separador de campos
GS = separador de grupos
RS = separador de registros
US = separador de unidades
SP = espacio
DEL = eliminar



(a) (b)

FIGURA 13-4 a) Código de barras; (b) estructura del código de barras

Código EBCDIC

El código EBCDIC es uno de ocho caracteres desarrollado por IBM. Con ocho bits son posibles 2^8, o 256 combinaciones, y esto hace que sea el código más poderoso de caracteres. Nótese que en el EBCDIC el bit menos significativo es el b_7, y el más significativo es el b_0. Por consiguiente, en el EBCDIC se transmite primero el bit de mayor orden, b_7, y al último se transmite el bit de menor orden, b_0. Este código no facilita el uso de un bit de paridad.

Códigos de barras

Los *códigos de barras* son las omnipresentes etiquetas con bandas negras y blancas que aparecen en casi todos los artículos al consumidor, en casi todas las tiendas de Estados Unidos y en la mayor parte del resto del mundo moderno. Es una serie de barras negras, separadas por espacios blancos. Los anchos de las bandas, junto con sus propiedades reflectoras representan unos y ceros binarios que identifican el costo del artículo. Además, pueden contener información acerca de la administración y control del inventario, acceso de seguridad, salidas y entradas, conteo de producción, procesamiento de documentos y pedidos, facturación automática y muchas otras aplicaciones. En la fig. 13-4a se ve un código normal de barras.

La fig. 13-4b muestra la distribución de los campos en un código característico de barras. El campo de inicio consiste en una secuencia exclusiva de barras y espacios, para identificar el inicio del campo de datos. Los caracteres de datos corresponden a la simbología o el formato del código de barras empleado. Los datos en serie, codificados en el campo de caracteres de datos se extraen de la tarjeta con un *escáner* o lector óptico. Este lector reproduce condiciones

TABLA 13-3 Conjunto de caracteres del código 39

Carácter	Palabra binaria	Barras	Espacios	Valor del carácter de verificación
0	000110100	00110	0100	0
1	100100001	10001	0100	1
2	001100001	01001	0100	2
3	101100000	11000	0100	3
4	000110001	00101	0100	4
5	100110000	10100	0100	5
6	001110000	01100	0100	6
7	000100101	00011	0100	7
8	100100100	10010	0100	8
9	001100100	01010	0100	9
A	100001001	10001	0010	10
B	001001001	01001	0010	11
C	101001000	11000	0010	12
D	000011001	00101	0010	13
E	100011000	10100	0010	14
F	001011000	01100	0010	15
G	000001101	00011	0010	16
H	100001100	10010	0010	17
I	001001100	01010	0010	18
J	000011100	00110	0010	19
K	100000011	10001	0001	20
L	001000011	01001	0001	21
M	101000010	11000	0001	22
N	000010011	00101	0001	23
O	100010010	10100	0001	24
P	001010010	01100	0001	25
Q	000000111	00011	0001	26
R	100000110	10010	0001	27
S	001000110	01010	0001	28
T	000010110	00110	0001	29
U	110000001	10001	1000	30
V	011000001	01001	1000	31
W	111000000	11000	1000	32
X	010010001	00101	1000	33
Y	110010000	10100	1000	34
Z	011010000	01100	1000	35
–	010000101	00011	1000	36
.	110000100	10010	1000	37
(espacio)	011000100	01010	1000	38
*	010010100	00110	1000	—
$	010101000	00000	1110	39
/	010100010	00000	1101	40
+	010001010	00000	1011	41
%	000101010	00000	0111	42

lógicas que corresponden a la diferencia de reflectividad de las barras impresas y los espacios en blanco. Para leer la información tan sólo se recorre la barra impresa con un movimiento uniforme. Un fotodetector en el lector siente la luz reflejada y la convierte en señales eléctricas, unos y ceros, para su decodificación.

Hay muchos formatos de código de barras, que varían desde los símbolos numéricos hasta todo el código ASCII. Uno de los códigos más comunes fue desarrollado en 1974, y se llama simplemente *código 39*. Este código, que también se llama *código 3 de 9*, usa una clave alfanumérica parecida al ASCII, y se muestra en la tabla 13-3. Está formado por 36 claves exclusivas que representan los 10 dígitos y 24 letras mayúsculas. Hay otros siete códigos para caracteres especiales, y un carácter exclusivo de arranque y paro, en forma de un asterisco (*).

Cada carácter contiene nueve elementos (barras y espacios). La condición lógica (1 o 0) de cada elemento se codifica variando el ancho de la barra o el espacio. Un elemento ancho, sea barra o espacio, representa un 1 lógico, y uno angosto, un 0 lógico. Para diferenciar los unos de los ceros se usa una relación de ancho de 3 a 1, es decir, una barra o espacio que representa un 1 lógico tiene tres veces el ancho que los que representan un 0 lógico.

Tres de los nueve elementos del código 39 deben ser unos lógicos, y los seis restantes deben ser ceros lógicos. Además, de los tres unos lógicos, dos deben ser barras y uno espacio. Cada carácter comienza y termina con una barra negra, y hay barras blancas que alternan entre las negras. Todos los caracteres están separados por un espacio. Los códigos que usan espacios entre caracteres se clasifican como aquellos cuyos caracteres están separados por huecos. Los huecos, en general, tienen un elemento de ancho.

El formato del código 39 comprende dos caracteres opcionales de comprobación, para detección de errores. Cada carácter tiene un valor de comprobación asignado. El valor de los caracteres dentro de un mensaje se suman y se dividen entre una constante. El residuo se redondea a uno o dos lugares decimales y a continuación se convierte a un carácter de código 39 y se agrega al final del campo de caracteres. Por ejemplo, si un campo de datos contuviera el mensaje "CODE 39", los caracteres de comprobación se determinarían como sigue

Mensaje	C	O	D	E	espacio	3	9		
Valor de comprobación			12	24	13	14	38	3	9

Suma de valores de comprobación = 113

Al dividir entre la constante 43 se obtiene 113/43 = 2 y restan 27

Mensaje final: CODE 39R

donde R es igual al valor del carácter de comprobación para 27.

CONTROL DE ERRORES

Un circuito de comunicación de datos puede ser tan corto como de algunos metros, o tan largo como varios miles de millas, y el medio de transmisión puede ser tan sencillo como un trozo de alambre, o tan complicado como una microonda, un satélite o un sistema de fibra óptica. En consecuencia, por las características no ideales de transmisión que hay en cualquier sistema de comunicaciones, es inevitable que haya errores, y que sea necesario desarrollar e implementar procedimientos de control de errores. El control de errores se puede dividir en dos categorías generales: detección de errores y corrección de errores.

Detección de errores

La *detección de errores* no es más que el proceso de vigilar los datos recibidos y determinar cuándo ha habido un error de transmisión. Las técnicas de detección de errores no identifican cuál o cuáles bits están equivocados; sólo indican que hubo un error. El objetivo de la detección de errores no es evitar que ocurran errores, sino evitar que haya errores sin detectar. La forma en que un sistema reacciona a los errores de transmisión depende del sistema y varía mucho. Las técnicas más comunes para los circuitos de comunicación de datos son redundancia, ecoplex, codificación de cuenta exacta, paridad, suma de comprobación, comprobación de redundancia vertical y horizontal y comprobación de redundancia cíclica.

Redundancia. La *redundancia* implica la transmisión de un carácter dos veces. Si no se recibe el mismo carácter dos veces seguidas, se ha presentado un error de transmisión. Se puede usar el mismo concepto para los mensajes. Si no se recibe la misma sucesión de caracteres dos veces seguidas, exactamente en el mismo orden, ha sucedido un error de transmisión.

Ecoplex. Es un esquema relativamente sencillo de detección de errores, que se usa en forma casi exclusiva en sistemas de comunicación de datos en los que operadores humanos capturan datos en forma manual con un teclado. Para el ecoplex se requiere operación dúplex,

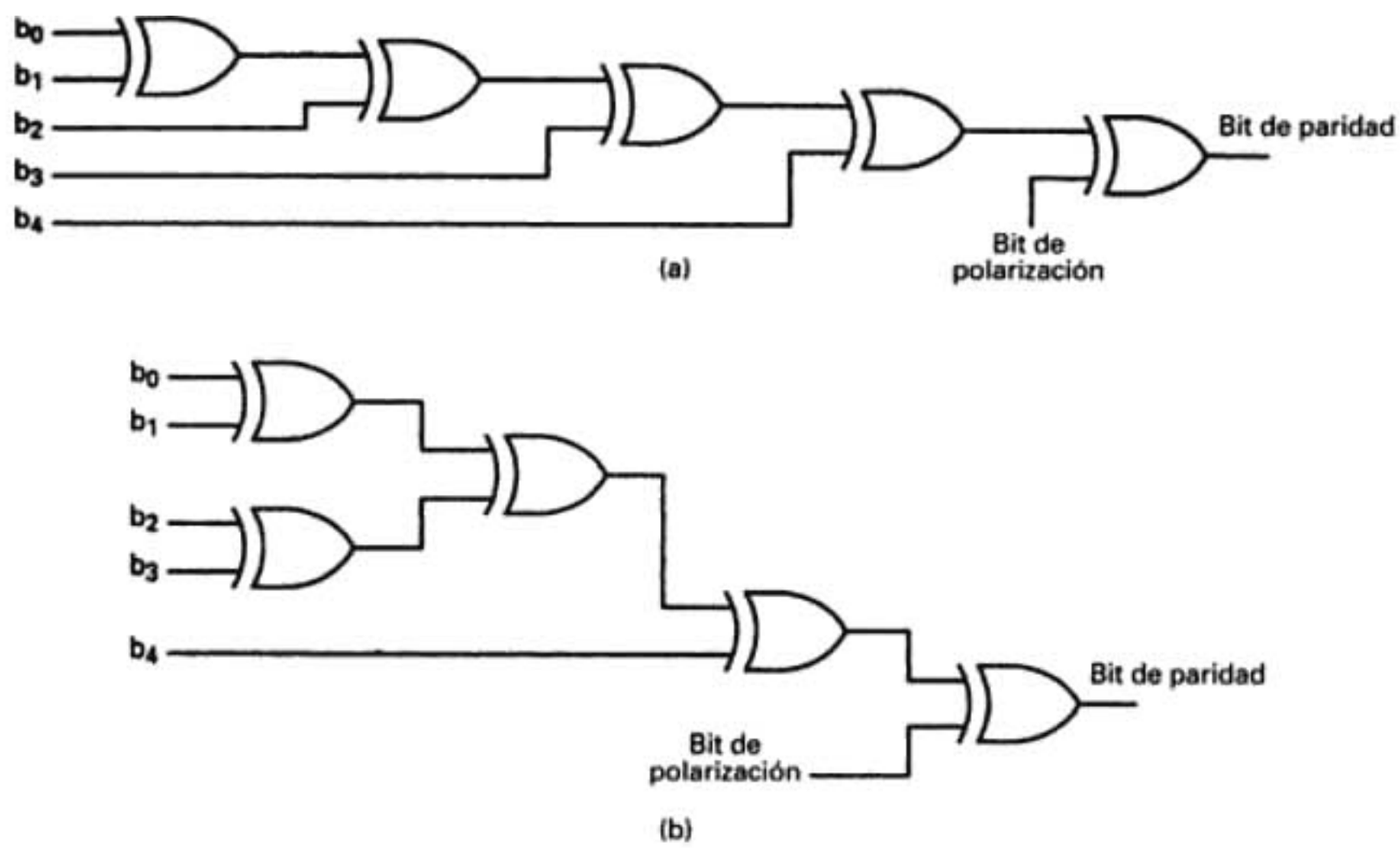


FIGURA 13-5 Generadores de paridad: (a) serie; (b) paralelo. 1 = paridad impar; 2 = paridad par

consecuencia, si sólo se cuenta la cantidad de unos recibida en cada carácter se puede determinar si ha ocurrido un error.

Paridad. Es probable que la *paridad* sea el esquema más sencillo de detección de errores que se usa en sistemas de comunicación de datos, y se usa junto con comprobación de redundancia, tanto vertical como horizontal. En la paridad se añade un solo bit (llamado *bit de paridad*) a cada carácter, para obligar a que la cantidad total de unos en el carácter, incluyendo el bit de paridad, sea un número impar (paridad impar) o bien un número par (paridad par). Por ejemplo, el código ASCII de la letra *C* es hexadecimal 43, o P1000011 binario, y la P representa al bit de paridad. Si se usa paridad impar, el bit P se iguala a 0, para mantener la cantidad total de unos en 3, que es un número impar. Si se usa paridad par, el bit P se iguala a 1, y la cantidad total de unos es 4, un número par.

Si se examina con más detalle la paridad, se verá que el bit de paridad es independiente de la cantidad de ceros en el código, y no lo afectan los pares de unos. Para la letra *C*, si se quitaran todos los bits 0, el código es P1———11. Para la paridad impar, el bit P sigue siendo un 0, y para la paridad par sigue siendo 1. Si se excluyen también pares de unos, el código puede ser P1——, P———1, o P———1—. De nuevo, para la paridad impar el bit P es un 0 y para la paridad impar el bit P es un 1.

La definición de paridad es *equivalencia* o *igualdad*. Una compuerta lógica que determine cuándo todas sus entradas son iguales es la XOR (operación lógica "o incluyente, equis o"). Con una compuerta XOR, si todas las entradas son iguales (todas 0 o todas 1), la salida es 0. Si no son iguales todas las entradas, la salida es 1. La fig. 13-5 muestra dos circuitos que se usan mucho para generar un bit de paridad. En esencia, ambos circuitos pasan por un proceso de comparación que elimina los ceros y pares de unos. El circuito de la fig. 13-5a usa comparación *secuencial* (*serial o en serie*), mientras que el de la fig. 13-5b usa comparación *compuesta* (*paralela*). En la paridad secuencial, el generador b_0 se compara XOR con b_1, el resultado se compara XOR con b_2, y así sucesivamente. El resultado de la última operación XOR se compara con un bit de *polarización*. Si se desea paridad par, el bit de polarización se iguala con 0 lógico. Si se desea paridad impar, el bit de polarización se hace 1 lógico. La salida del circuito es el bit de paridad, que se agrega a los caracteres codificados. En el generador de bit de paridad, las comparaciones se hacen en capas o niveles. Los pares de bits (b_0 y b_1, b_2 y b_3, etc., se comparan

XOR. Los resultados de las compuertas XOR de primer nivel se comparan XOR entre sí. El proceso continúa hasta que queda sólo un bit, que se compara XOR con el bit de polarización. Nuevamente, si se desea paridad par, el bit de polarización se hace 0 lógico, y si se desea paridad impar, el bit de polarización se hace igual a 1 lógico.

Los circuitos de la fig. 13-5 también se pueden usar en el comprobador de paridad del receptor. Un comprobador de paridad usa el mismo procedimiento que un generador de paridad, pero la condición lógica de la comparación final es la que se usa si ha sucedido una violación de la paridad; por ejemplo, en la paridad impar un 1 indica error y un 0 indica sin error; en la paridad par un 1 indica error y un 0 indica sin error.

La ventaja principal de la paridad es su sencillez. La desventaja es que cuando se reciben varios bits equivocados, podría no detectarlos el comprobador de paridad; es decir, si cambian las condiciones lógicas de dos bits, la paridad queda igual. En consecuencia, la paridad, dentro de un tiempo prolongado, sólo detectará el 50% de los errores de transmisión; para esto se supone que hay probabilidades iguales de que haya cantidades pares o impares de bits equivocados.

Suma de comprobación. La *suma de comprobación* es una forma muy sencilla de detección de error. Una suma de comprobación no es más que el byte menos significativo de la suma aritmética de los datos binarios que se transmiten. Mientras se transmiten los datos, cada carácter se suma con la suma acumulada de los que se transmitieron antes. Cuando se llega al final del mensaje, el sumador ha acumulado la suma de todos los caracteres que hay en el mensaje que se acaba de mandar. El byte menos significativo de esta suma se agrega al final del mensaje, y se transmite. La terminal receptora hace de nuevo la operación de suma y determina su propia suma y su propio carácter de suma de comprobación. El byte menos significativo de la suma del receptor se compara con la suma de comprobación del final del mensaje. Si son iguales, es muy probable que no haya habido error de transmisión. Si son distintas es seguro que se presentó un error. Cuando se detecta un error se pide una retransmisión de todo el mensaje.

Comprobación de redundancia vertical y horizontal. La *comprobación de redundancia vertical* (VRC, de *vertical redundancy checking*) es un esquema de detección de errores que usa la paridad para determinar si ha sucedido un error de transmisión, dentro de un carácter. En consecuencia, la VRC a veces se llama *paridad de caracteres*. En la VRC cada carácter tiene agregado un bit de paridad, antes de la transmisión. Se pueden usar paridad par o impar. El ejemplo que se ve en el tema "paridad" para el carácter "C" en ASCII es un ejemplo de cómo se usa la comprobación de redundancia vertical.

La *comprobación de redundancia horizontal* o *longitudinal* (HRC, de *horizontal redundancy checking*) o (LRC, de *longitudinal redundancy checking*) es un esquema de detección de errores que usa la paridad para determinar si se ha presentado un error de transmisión en un mensaje y, en consecuencia, a veces se le llama *paridad de mensaje*. Con la LRC, cada posición de bit tiene un bit de paridad. En otras palabras, se compara XOR el b_0 de cada carácter del mensaje con b_0 de los demás caracteres del mensaje. En forma parecida, se comparan XOR b_1, b_2, etc. con sus bits respectivos de los demás caracteres en el mensaje. En esencia, la LRC es el resultado de comparaciones XOR de los "caracteres" que forman un mensaje, mientras que la VRC es la comparación XOR de los bits dentro de un solo carácter. Con la LRC sólo se usa paridad par.

La secuencia de bits LRC se calcula en el transmisor antes de mandar los datos; a continuación se transmite como si fuera el último carácter del mensaje. En el receptor, se vuelve a calcular la LRC a partir de los datos, y el valor recalculado se compara con la LRC transmitida con el mensaje. Si son iguales, se supone que no se han presentado errores de transmisión. Si son diferentes, debe haberse presentado un error de transmisión.

En el ejemplo 13-1 se indica cómo se determinan la VRC y LRC.

Ejemplo 13-1

Determinar la VRC y LRC para el siguiente mensaje codificado en ASCII: THE CAT. Usar paridad impar para VRC y paridad par para LRC.

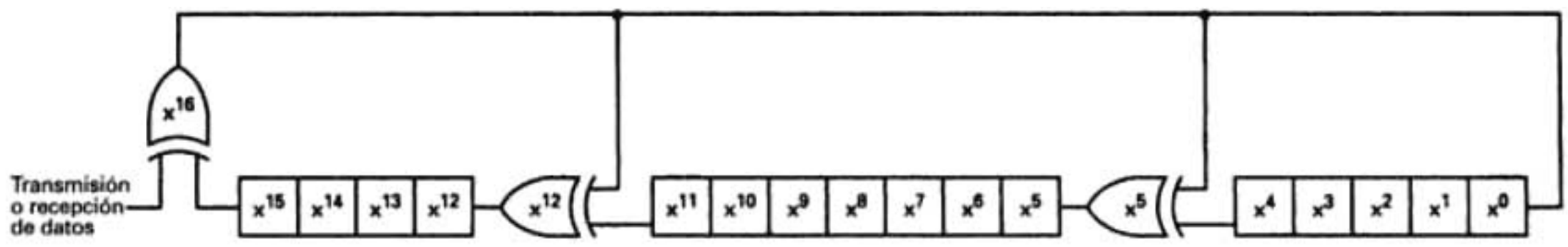


FIGURA 13-6 Circuito generador CRC-16 (CCITT V.41)

receptor, la corriente de datos recibidos, incluyendo la BCS, se dividen entre la misma función generadora $P(x)$. Si no se han presentado errores de transmisión, el residuo será cero.

El polinomio generador para la CRC-16 es

$$P(x) = x^{16} + x^{12} + x^5 + x^0$$

siendo $x^0 = 1$.

La cantidad de bits en el código CRC es igual al máximo exponente del polinomio generador. Los exponentes identifican las posiciones de bit que contienen un 1. En consecuencia, b_{16}, b_{12}, b_5 y b_0 son unos, y las demás posiciones de bit son cero.

La fig. 13-6 muestra el diagrama de bloques de un circuito para generar una BCS de CRC-16, para la norma CCITT V.41. Nótese que para cada posición de bit del polinomio generador, donde hay un 1, se pone una compuerta XOR, excepto para x^0.

Ejemplo 13-2

Determinar la BCS para los siguientes datos y polinomios generadores CRC

$$\text{datos } G(x) = x^7 + x^5 + x^4 + x^2 + x^1 + x^0 \quad \text{o} \quad 10110111$$
$$\text{CRC } P(x) = x^5 + x^4 + x^1 + x^0 \text{ or } 110011$$

Solución Primero se multiplica $G(x)$ por la cantidad de bits en el código CRC, que es 5.

$$x^5(x^7 + x^5 + x^4 + x^2 + x^1 + x^0) = x^{12} + x^{10} + x^9 + x^7 + x^6 + x^5$$
$$= 1011011100000$$

A continuación el resultado se divide entre $P(x)$.

```
                 11010111
110011 ) 1011011100000
         110011
          111101
          110011
            111010
            110011
              100100
              110011
               101110
               110011
                111010
                110011
                 01001  = CRC
```

La CRC se agrega a los datos para producir la siguiente corriente de datos transmitidos

$G(x)$	CRC
10110111	01001

En el receptor, los datos transmitidos se vuelven a dividir entre $P(x)$.

```
                    11010111
110011 ) 1011011101001
         110011
          111101
          110011
            111010
            110011
              100110
              110011
               101010
               110011
                110011
                110011
                000000      Residuo = 0
                            No hubo error
```

Corrección de errores

En esencia, hay tres métodos para corregir errores: sustitución de símbolo, retransmisión y corrección de error en avance.

Sustitución de símbolo. La *sustitución de símbolo* se diseñó para usarse en ambientes humanos, cuando hay un ser humano en la terminal de recepción, que analice los datos recibidos y tome decisiones sobre su integridad. En la sustitución de símbolo, si se recibe un carácter equivocado, más que pasarlo a un nivel superior de corrección de errores, o mostrar el carácter incorrecto, un carácter indefinido en el código de caracteres, como por ejemplo un signo de interrogación invertido (؟), sustituye al carácter con error. Si ese carácter con error no puede ser interpretado por el operador, se pide su retransmisión, es decir, la sustitución de símbolo es una forma de retransmisión selectiva. Por ejemplo, si el mensaje "nombre" tuviera un error en el primer carácter, se mostraría como " ؟ombre". Un operador puede interpretar el mensaje correcto por inspección, y es innecesaria la retransmisión. Sin embargo, si se recibiera el mensaje "$ ؟,000.00" el operador no puede determinar cuál es el carácter equivocado y se pide la retransmisión.

Retransmisión. Como su nombre lo dice, la *retransmisión* es la vuelta a mandar un mensaje cuando se recibe con errores, y la terminal de recepción pide en forma automática la retransmisión de todo el mensaje. La retransmisión se llama ARQ, con frecuencia, que es un término antiguo de radiocomunicaciones que quiere decir *automatic request for retransmission*, petición automática de retransmisión. La ARQ es, probablemente, el método más seguro de corrección de errores, aunque no siempre sea el más eficiente. Las irregularidades en los medios de transmisión se presentan como explosiones. Si se usan mensajes cortos, será pequeña la probabilidad de que haya una irregularidad durante una transmisión. Sin embargo, los mensajes cortos requieren más reconocimientos y cambios de dirección de línea que los mensajes largos. Los reconocimientos y cambios de dirección de línea para controlar errores son formas de *indirectos*; son caracteres que no son de datos y que se deben transmitir. Con los mensajes grandes se necesita menos tiempos de cambio de línea, aunque aumenta la probabilidad de que haya un error de transmisión, respecto a los mensajes cortos. Se puede demostrar, con métodos estadísticos, que los bloques de mensajes entre 256 y 512 caracteres tienen el tamaño óptimo cuando se usa ARQ para corrección de errores.

Corrección de error en sentido directo. Se le conoce como FEC (de *forward error correction*) y es el único esquema de corrección de errores que realmente detecta y corrige los errores de transmisión en la recepción, sin pedir la retransmisión.

En el sistema FEC se agregan bits al mensaje antes de transmitirlo. Un código muy difundido para corregir errores es el *código Hamming*, desarrollado por R. W. Hamming en los Laboratorios Bell. La cantidad de bits en el código Hamming depende de la cantidad de bits en el carácter de datos. La cantidad de bits de Hamming que se deben agregar a un carácter se calcula con la siguiente ecuación

$$2^n \geq m + n + 1 \tag{13-1}$$

en donde n = cantidad de bits de Hamming
m = cantidad de bits en el carácter de datos

Ejemplo 13-3

Para la cadena de 12 bits de datos 101100010010, calcular la cantidad de bits de Hamming necesarios, incrustar en forma arbitraria esos bits en la cadena de datos, determinar la condición de cada bit de Hamming, suponer un error arbitrario de transmisión de un solo bit, y demostrar que el código Hamming detecta el error.

Solución De acuerdo con la ecuación 13-1, la cantidad de bits de Hamming es

$$2^n \geq m + n + 1$$

Para $n = 4$, $2^4 = 16 \geq m + n + 1 = 12 + 4 + 1 = 17$

Como 16 < 17, cuatro bits de Hamming son insuficientes.

Para $n = 5$, $2^5 = 32 \geq m + n + 1 = 12 + 5 + 1 = 18$

32 > 18, y, en consecuencia, cinco bits de Hamming son suficientes para satisfacer el criterio de la ecuación 13-1. Así, la corriente de datos estará formada por un total de 12 + 5 = 17 bits.

Se colocan en forma arbitraria cinco bits de Hamming en la corriente de datos

17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
H	1	0	1	H	1	0	0	H	H	0	1	0	H	0	1	0

Para determinar la condición lógica de los bits de Hamming, se expresan todas las posiciones de bit que contienen un uno como número binario de cinco bits, y se combinan entre sí con XOR.

Posición de bit	Número binario
2	00010
6	00110
XOR	00100
12	01100
XOR	01000
14	01110
XOR	00110
16	10000
XOR	10110 = código de Hamming

$$b_{17} = 1, b_{13} = 0, b_9 = 1, b_8 = 1, b_4 = 0$$

La corriente de datos codificada en 17 bits se transforma en

H			H					H	H				H			
1	1	0	1	0	1	0	0	1	1	0	1	0	0	0	1	0

Se supone que durante la transmisión se presenta un error en el bit de la posición 14. La corriente de datos recibidos es

1	1	0	<u>0</u>	0	1	0	0	1	1	0	1	0	0	0	1	0

En el receptor, para determinar la posición del bit erróneo, se extraen los bits de Hamming y se comparan XOR con el código binario para cada posición de bit que contiene un 1.

Posición de bit	Número binario
Código de Hamming	10110
2	00010
XOR	10100
6	00110
XOR	10010
12	01100
XOR	11110
16	10000
XOR	01110 = 14 binario

Se recibió erróneo el bit 14. Para corregirlo, sólo se complementa ese bit.

El código Hamming aquí descrito sólo detectará errores de un solo bit. No se puede usar para identificar errores de varios bits, ni errores en los mismos bits de Hamming. El código Hamming, así como todos los códigos FEC (corrección directa de errores), requiere que se agreguen bits a los datos y, en consecuencia, se alargue el mensaje transmitido. El objetivo de los códigos FEC es reducir o eliminar el tiempo desperdiciado en retransmisiones. Sin embargo, la adición de los bits FEC a cada mensaje desperdicia el mismo tiempo de transmisión. Es obvio que se debe hacer un balance entre los códigos ARQ y FEC, y las necesidades del sistema, para determinar cuál método se adapta mejor a determinado sistema. El FEC se usa con frecuencia para transmisiones símplex a muchos receptores, cuando son imprácticos los reconocimientos.

SINCRONIZACIÓN

Sincronizar quiere decir coincidir o concordar en el tiempo. En comunicaciones de datos hay cuatro clases de sincronización que deben lograrse: de bits o reloj, de módem o portadora, de caracteres y de mensaje. Los circuitos de recuperación de reloj y de portadora, que se describieron en el capítulo 12, logran la sincronización de bits y de portadora, y la sincronización de mensaje se describirá en el capítulo 14.

Sincronización de caracteres

La sincronización de reloj asegura que el transmisor y el receptor coincidan en un intervalo preciso (muesca) de tiempo para que se presente un bit. Cuando se recibe una cadena continua de datos, es necesario identificar cuáles bits pertenecen a cuáles caracteres, y cuál bit es el menos significativo de los datos, el bit de paridad y el bit de paro. En esencia, eso es la sincronización de caracteres: identificar el comienzo y el fin de un código de caracteres. En los circuitos de comunicación de datos hay dos formatos que se usan para lograr la sincronización de caracteres: asíncronos y síncronos.

Formato de datos asíncronos. Con *datos asíncronos*, cada carácter se enmarca entre un bit *de inicio* o *arranque*, y uno *de paro*. La fig. 13-7 muestra el formato para encuadrar un carácter en una transmisión de datos asíncronos. El primer bit que se transmite es de arranque, que siempre es un 0 lógico. Los bits de código de caracteres se transmiten después, comenzando con el menos significativo y avanzando hasta el más significativo. El bit de paridad, si se usa, se transmite en forma directa después del bit más significativo del carácter. El último bit que se transmite es el de paro, y siempre es un 1 lógico. Pueden haber 1, 1.5 o 2 bits de paro.

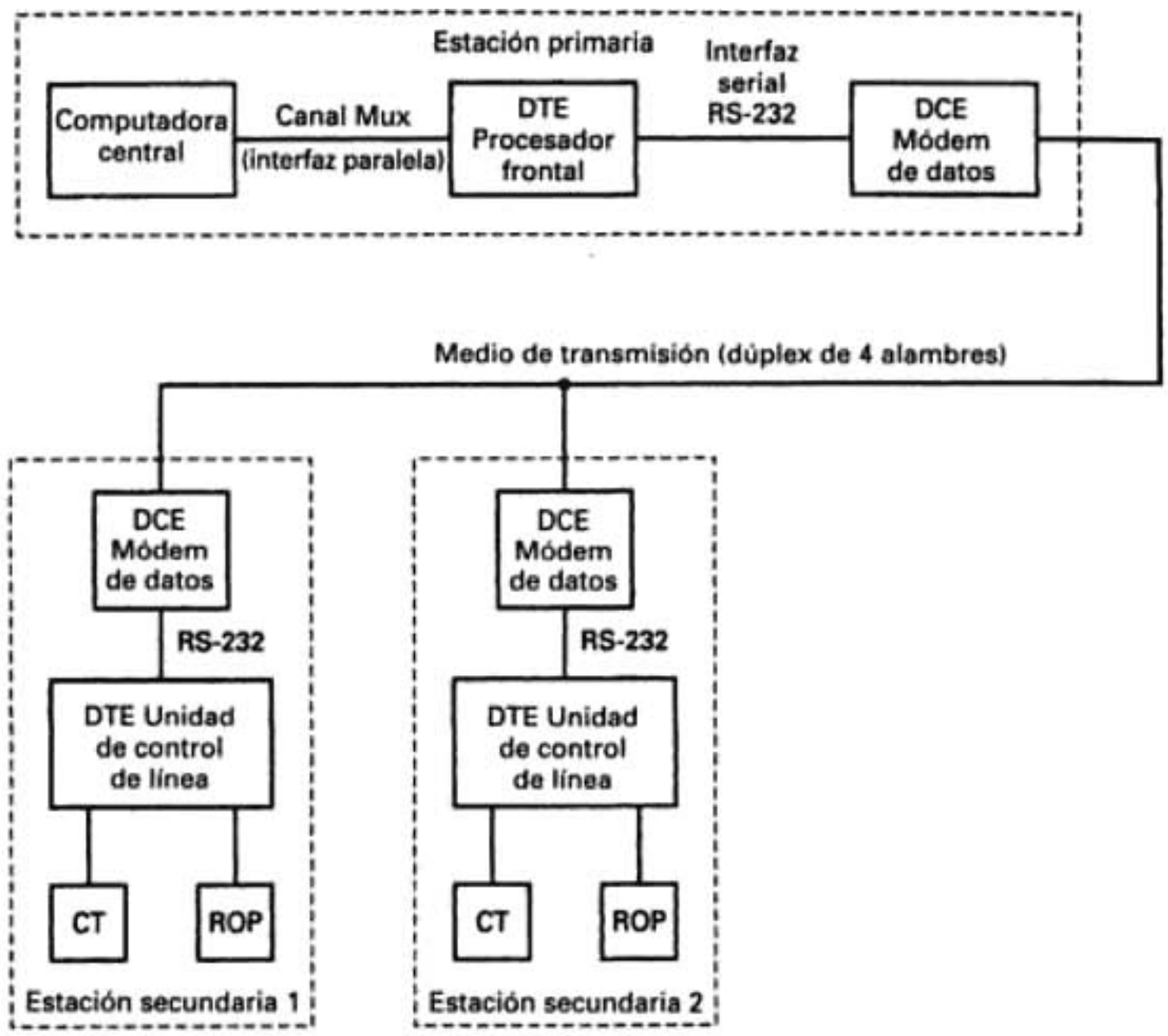


FIGURA 13-8 Diagrama de bloques de un circuito de comunicaciones de datos multipunto

la transmisión, es decir, la relación de bits de información a bits totales transmitidos. Los datos síncronos tienen dos caracteres SYN, que equivalen a 16 bits de indirectos, agregados a cada mensaje. Por consiguiente, los datos asíncronos son más eficientes para mensajes cortos y los datos síncronos son más eficientes para mensajes largos.

COMPONENTES EN LA COMUNICACIÓN DE DATOS

La fig. 13-8 muestra el diagrama de bloques de un circuito multipunto de comunicación de datos que usa topología de bus. Este arreglo es una de las configuraciones más comunes que se usan en circuitos de comunicación de datos. En una estación hay una computadora central, y en cada una de las otras dos estaciones hay un *grupo* de terminales de cómputo. Los componentes y los circuitos asociados que conectan la computadora anfitrión con las terminales remotas se llama *enlace de comunicación de datos*. La estación de la computadora central se llama *anfitrión* o *primaria*, y las demás estaciones se llaman *secundarias* o simplemente *remotas*. Un arreglo como este se llama *red centralizada*; hay una estación ubicada en el centro (el anfitrión) cuya responsabilidad es asegurar un flujo ordenado de datos entre ella y las estaciones remotas. El flujo de datos se controla con un programa de aplicaciones, guardado en la estación primaria.

En la estación primaria hay una computadora central, una *unidad de control de línea* (LCU, de *line control unit*) y un *módem de datos*, o simplemente *módem*. En cada estación secundaria hay un módem, una unidad de control de línea y equipo de terminal, como por ejemplo terminales de computadora e impresoras. La central es el anfitrión de la red, y es donde se guarda el programa de aplicación para cada circuito al que da servicio. Para simplificar, la fig. 13-8 sólo muestra un circuito servido por la primaria, aunque pueden haber muchos circuitos distintos servidos por una computadora central. La estación primaria tiene la capacidad de guardar, procesar o retransmitir los datos que recibe de las estaciones secundarias. También guarda los programas para la administración de base de datos.

La unidad de control de línea (LCU) de la estación primaria es más complicada que las de las estaciones secundarias. En la estación primaria, la LCU dirige el tráfico de datos hacia y

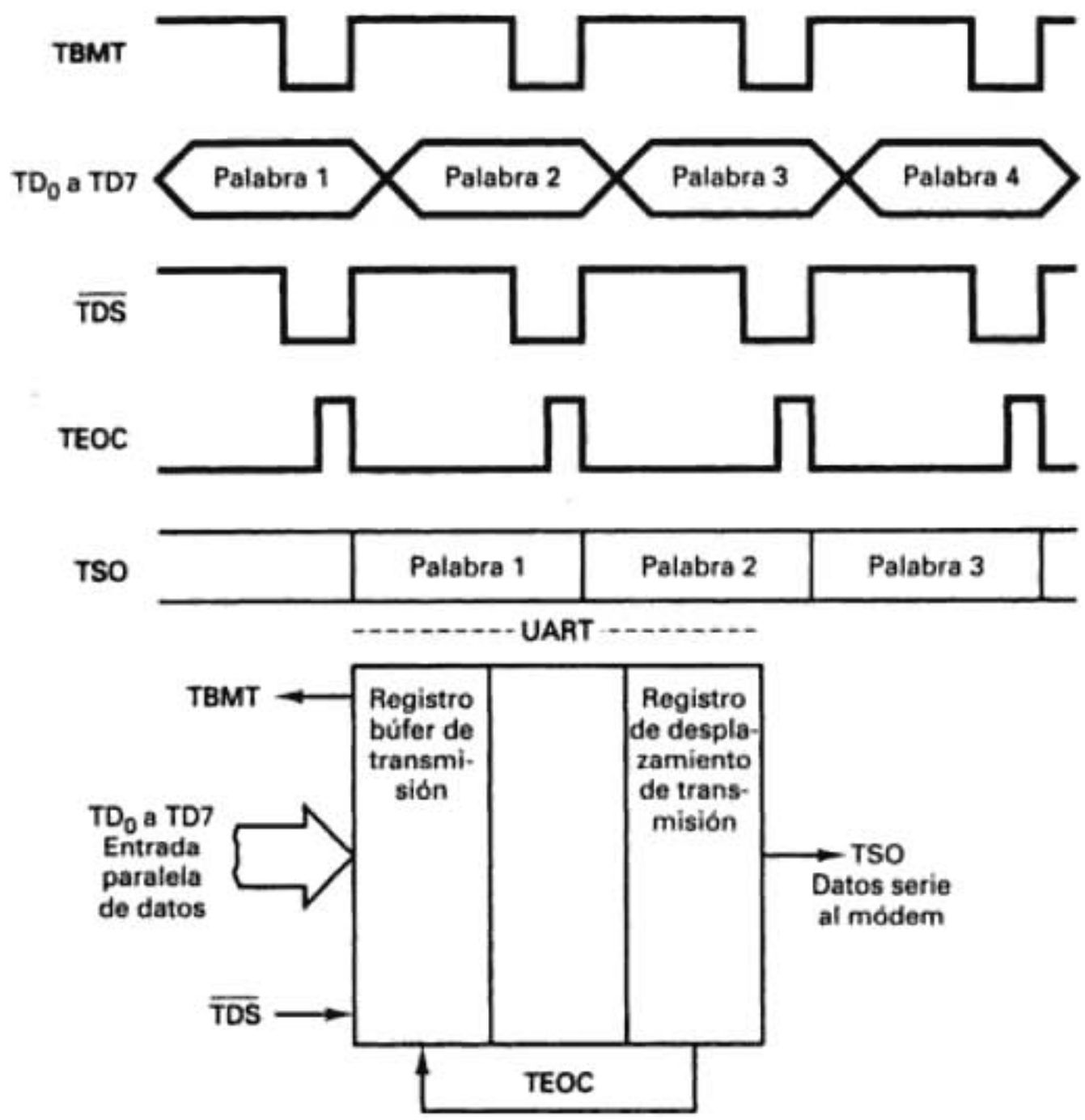


FIGURA 13-10 Diagrama de sincronización: transmisor UART

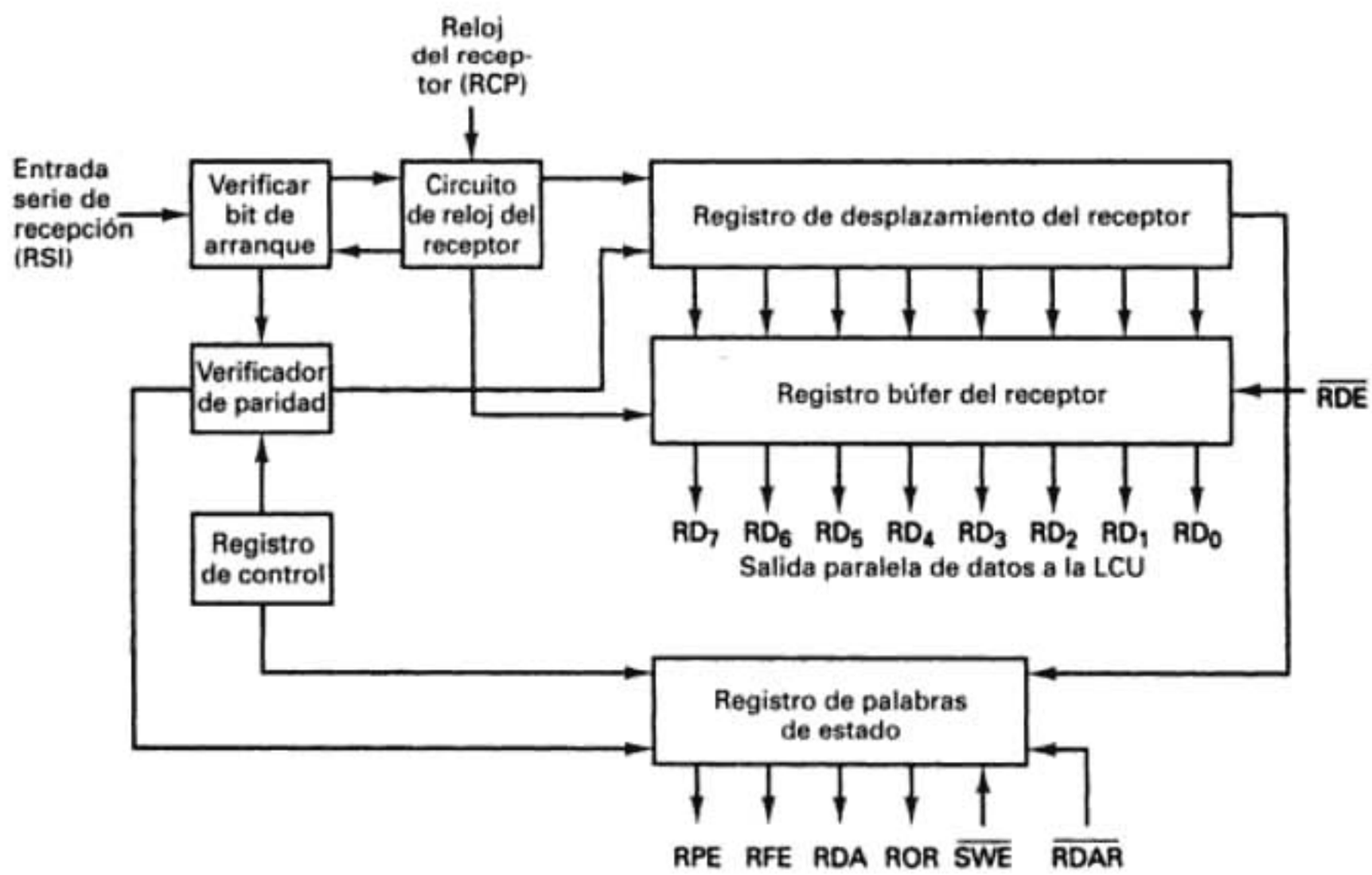


FIGURA 13-11 Diagrama simplificado de bloques de un receptor UART

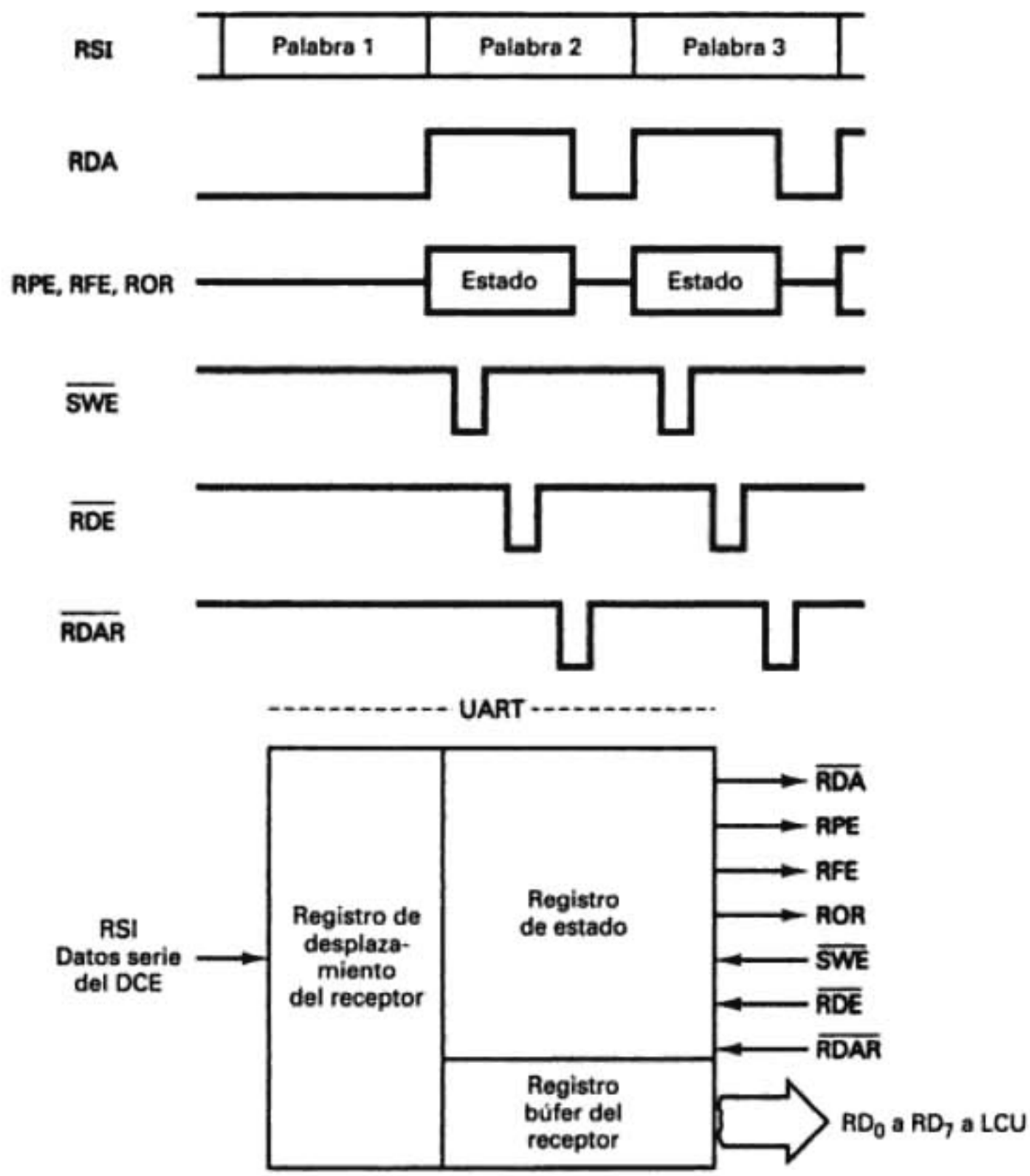


FIGURA 13-12 Diagrama de sincronización: receptor UART

El receptor del UART ignora los unos de línea inactiva. Cuando el circuito de verificación de bit de arranque detecta un bit válido de arranque, el carácter de datos se sincroniza en serie en el registro de desplazamiento del receptor. Si se usa paridad, el circuito de comprobación de paridad la verifica. Después de haber cargado un carácter completo de datos en el registro de desplazamiento, el carácter se transfiere en paralelo al registro búfer y se pone la bandera de recepción de datos disponible (RDA, de *receive data available*), en el registro de palabra de estado. Para leer el registro de estado, el DTE vigila el activamiento de la palabra de estado ($\overline{\text{SWE}}$, de *status word enable*) y, si está activa, lee el carácter del registro búfer colocando una condición activa en la terminal de activar recepción de datos (RDE, de *receive data enable*). Después de leer los datos, el DTE pone una señal activa en la terminal de restablecer disponibilidad de recepción de datos ($\overline{\text{RDAR}}$, de *receive data available reset*), con lo cual se restablece la terminal RDA. Mientras tanto se recibe y se sincroniza el siguiente carácter, en el registro de desplazamiento del receptor, y se repite el proceso hasta que se hayan recibido todos los datos. Esta secuencia se ve en la fig. 13-12.

También, el registro de palabra de estado se usa para información de diagnóstico. Se pone la bandera de error de paridad en recepción (RPE, de *receive parity error*) cuando un carácter recibido tiene un error de paridad. Se pone la bandera de error de trama de bit (RFE, de *receive framing error*) cuando se recibe un carácter sin bits de paro, o una cantidad incorrecta de ellos. Se pone la bandera de desbocamiento de recepción (ROR, de *receive overrun*) cuando un bit del registro búfer se encima con otro carácter; es decir, cuando el DTE no emitió una condición activa de RDA antes de que el siguiente carácter se recibiera en el registro de desplazamiento.

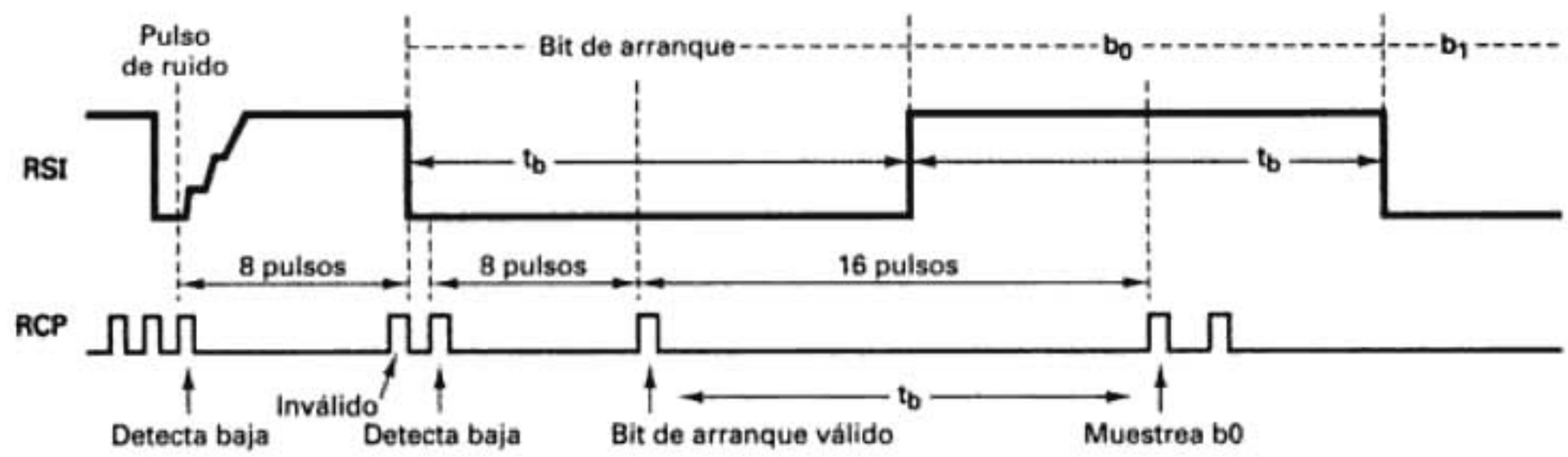


FIGURA 13-13 Verificación del bit de arranque

El reloj del receptor del UART (RCP) tiene una frecuencia 16 veces mayor que la de recepción de datos. Esto permite que el circuito de verificación del bit de arranque determine si una transición de alto a bajo en los datos recibidos es, en realidad, un bit válido de arranque, y no tan sólo un pico negativo de ruido. La fig. 13-13 muestra como se logra esto. Los unos de línea inactiva que entran (condición alta continua) se muestrean con una frecuencia 16 veces mayor que la frecuencia real de bits. Así se asegura que se va a detectar una transición de alto a bajo en 1/16 del tiempo después de ocurrir. Una vez detectado un nivel bajo, el circuito de verificación cuenta siete pulsos de reloj y a continuación vuelve a muestrear los datos. Si todavía están bajos, se supone que se ha detectado un bit válido de arranque. Si ha cambiado a la condición alta, se supone que la transición de alto a bajo sólo fue un pulso de ruido y, en consecuencia, se ignora. Una vez detectado y verificado un bit de arranque válido, el circuito de verificación muestrea los datos que entran una vez cada 16 ciclos del reloj, con frecuencia igual a la de los datos. También, el muestreo a 16 veces la frecuencia de bits, establece el tiempo de muestreo a menos de 1/16 del tiempo de un bit, a partir del centro de un bit.

Receptor/transmisor síncrono universal (USRT). El USRT (de *universal synchronous receiver/transmitter*) es para transmitir datos síncronos del DTE al DCE. Transmisión síncrona quiere decir que hay información de sincronización transferida del USRT al módem, y que cada transmisión comienza con un carácter SYN único. Las funciones primarias del USRT son:

1. Convertir datos, de serie a paralelo y de paralelo a serie.
2. Detectar errores insertando y comprobando bits de paridad.
3. Insertar y detectar los caracteres SYN.

En la fig. 13-14a se muestra el diagrama de bloques del USRT; funciona en forma muy parecida al UART y, en consecuencia, sólo se explicarán las diferencias. En el USRT no se permiten los bits de arranque y paro. En lugar de ellos, se cargan caracteres SYN únicos en los registros SYN de transmisión y recepción, antes de transferir datos. La información de programación para la palabra de control se ve en la fig. 13-14b.

Transmisor USRT. La señal del reloj del transmisor (TCP, de *transmit clock pulse*) se ajusta a la frecuencia deseada de bits y el carácter SYN deseado se carga desde las terminales de entrada paralela (DB_0 a DB_7) al registro SYN de transmisión, pulsando el estrobo (sincronizador) SYN de transmitir (TSS, de *transmit SYN strobe*). Los datos se cargan en el registro de datos de transmisión desde DB_0 a DB_7 pulsando el estrobo de datos de transmisión (TDS, de *transmit data strobe*). El siguiente carácter para transmitir se extrae del registro de datos de transmisión, siempre que el pulso TDS se presente durante el carácter que se esté transmitiendo. Si no se pulsa el TDS, el siguiente carácter transmitido se extrae del registro SYN de transmisión, y se pone el carácter SYN transmitido (SCT, de *SYN character transmitted*). La señal de búfer vacío de transmisión (TBMT) se usa para pedir el siguiente carácter al DTE. Los datos de salida serie aparecen en la terminal de salida serie de transmisión (TSO, de *transmit serial output*).

Receptor USRT. La señal del reloj del receptor (RCP, de *receive clock pulse*) se ajusta a la frecuencia deseada de bits y se carga el carácter SYN deseado al registro SYN de recepción, desde DB_0 a DB_7, pulsando el estrobo SYN de recepción (RSS, de *receive SYN strobe*). En una

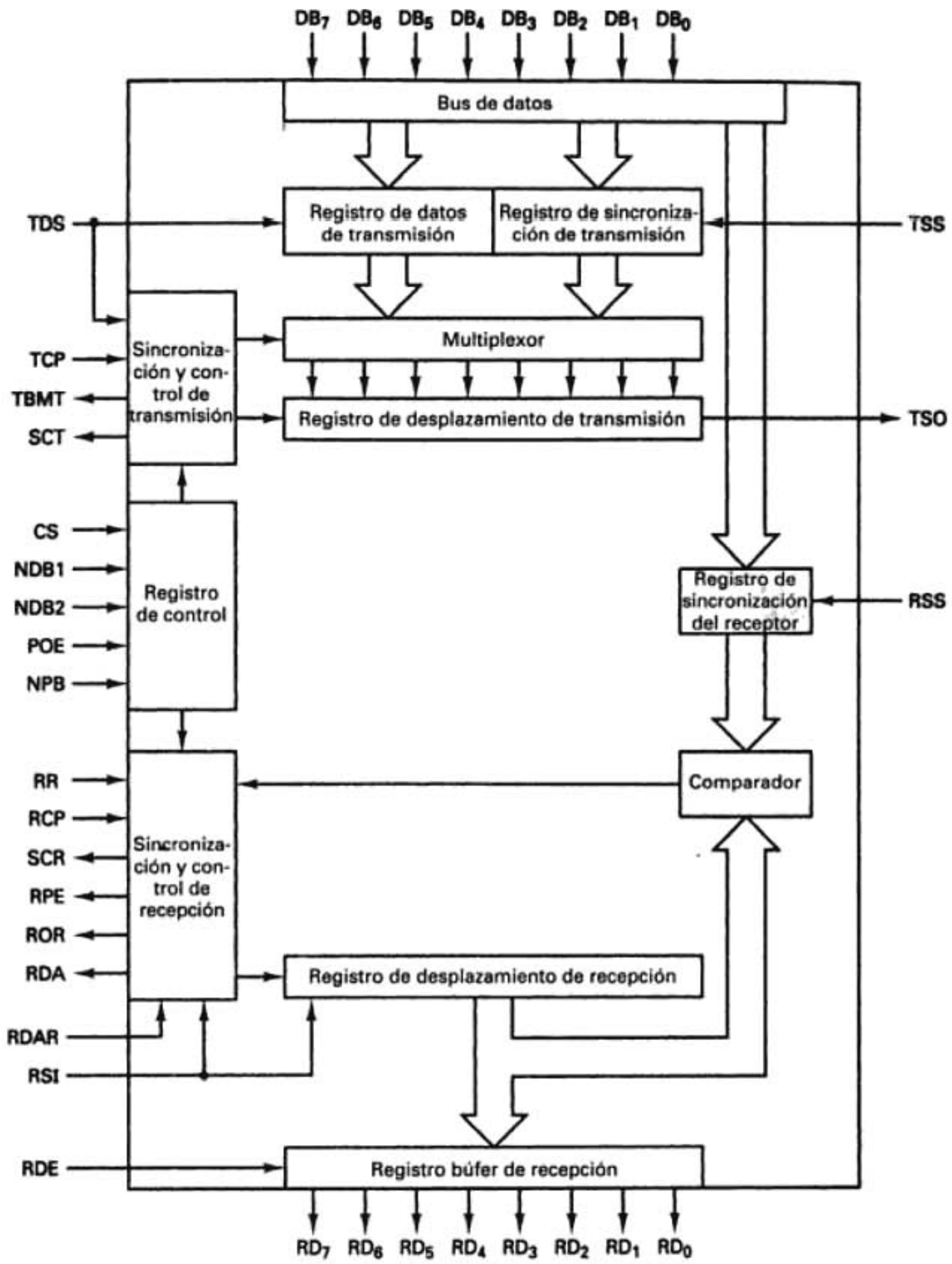


FIGURA 13-14 Transceptor USRT: (a) diagrama de bloques; (b) palabra de control

transición de alto a bajo de la entrada de reposo del receptor (RR, de *receiver rest*), el receptor se pone en el modo de búsqueda (fase de bit). En el modo de búsqueda, los datos recibidos en serie se examinan uno por uno hasta que se encuentra un carácter SYN. Después de sincronizar cada bit en el registro de desplazamiento de recepción, se compara el contenido del registro con el del registro SYN de recepción. Si son idénticos, se ha encontrado un carácter SYN y se pone la salida del carácter SYN de recepción (SCR, de *SYN character receive*). Este carácter se transfiere al registro búfer de recepción y el de recepción se pone en modo de carácter. En el modo de carácter, los datos de recepción se examinan uno por uno y se ponen banderas, en el receptor, de datos de recepción disponibles (RDA, de *receive data available*), desbocamiento del receptor (ROR, de *receiver overrun*), error de paridad del receptor (RPE, de *receive parity error*), y SCR en el registro de estado de palabra. Los datos de recepción paralela se descargan en el DTE por RD_0 a RD_7.

INTERFACES EN MODO SERIE

Para asegurar un flujo ordenado de datos entre la unidad de control de línea y el módem, entre ellos se pone una *interfaz en modo serie*. Esta interfaz coordina el flujo de datos, las señales de control y la información de sincronización entre el DTE y el DCE.

Antes de haberse normalizado las interfaces en modo serie, cada empresa fabricante de equipos de comunicación de datos usaba una configuración distinta de interfaz. En forma más específica, el cableado entre el DTE y el DCE, el tipo y el tamaño de los conectores usados y los valores de voltaje variaban mucho de uno a otro proveedor. Para interconectar equipos fabricados por empresas distintas había que fabricar convertidores de nivel, cables y conectores especiales.

En 1962, la EIA, Asociación de Industriales Electrónicos, estableció un conjunto de normas, llamada especificaciones RS-232, para tratar de normalizar el equipo de interfaces entre equipos de terminales de datos (DTE) y equipos de comunicaciones de datos (DCE). En 1969 se publicó la tercera modificación, RS-232C, y quedó como norma industrial hasta 1987, cuando se introdujo la RS-232D. Esta última versión de la interfaz es compatible con la versión C. La diferencia principal entre las dos versiones es la adición de tres circuitos de prueba en la versión D.

Las especificaciones RS-232 identifican la descripción mecánica, eléctrica, funcional y de procedimientos para la interfaz entre el DTE y el DCE. La interfaz RS-232 se parece a las normas combinadas V.28 (especificaciones eléctricas) y V.24 (descripción funcional) del CCITT, y es para transmitir datos en modo serie hasta a 20 kbps, a una distancia aproximada de 15 m (50 pies).

INTERFAZ RS-232

La interfaz RS-232 especifica un cable de 25 conductores con un conector compatible con DB25P/DB25S. La fig. 13-15 muestra las características eléctricas de esta interfaz. La capacitancia de carga de terminal en este cable se especifica como de 2500 pF, lo cual incluye la capacitancia del cable. La impedancia en el extremo de terminación debe ser de 3000 a 7000 Ω, y la impedancia de salida se especifica como mayor que 300 Ω. Con estas especificaciones eléctricas, y para una frecuencia máxima de bits de 20,000 bps, la longitud nominal máxima de la interfaz RS-232 es de unos 15 m (50 pies).

Aunque la interfaz RS-232 no es más que un cable con dos conectores, la norma también especifica limitaciones para los valores de voltaje que pueden mandar o recibir del cable el DTE y el DCE. En ambos equipos hay circuitos que convierten sus valores lógicos internos a valores RS-232. Por ejemplo, si un DTE usa lógica TTL (transistor-transistor) y se interconecta con un DCE que usa lógica ECL (transistor acoplado por el emisor), no son compatibles. Los circuitos de nivelación de voltaje convierten los valores internos de voltaje del DTE y del DCE a valores RS-232. Si tanto el DCE como el DTE producen y toman valores RS-232, son compatibles eléctricamente, independientemente de qué familia lógica usen internamente. Un nivelador se llama *excitador* si produce un voltaje de señal al cable, y un *terminador* si acepta un voltaje de señal del cable. La tabla 13-5 muestra los límites de voltaje para excitadores y terminadores.

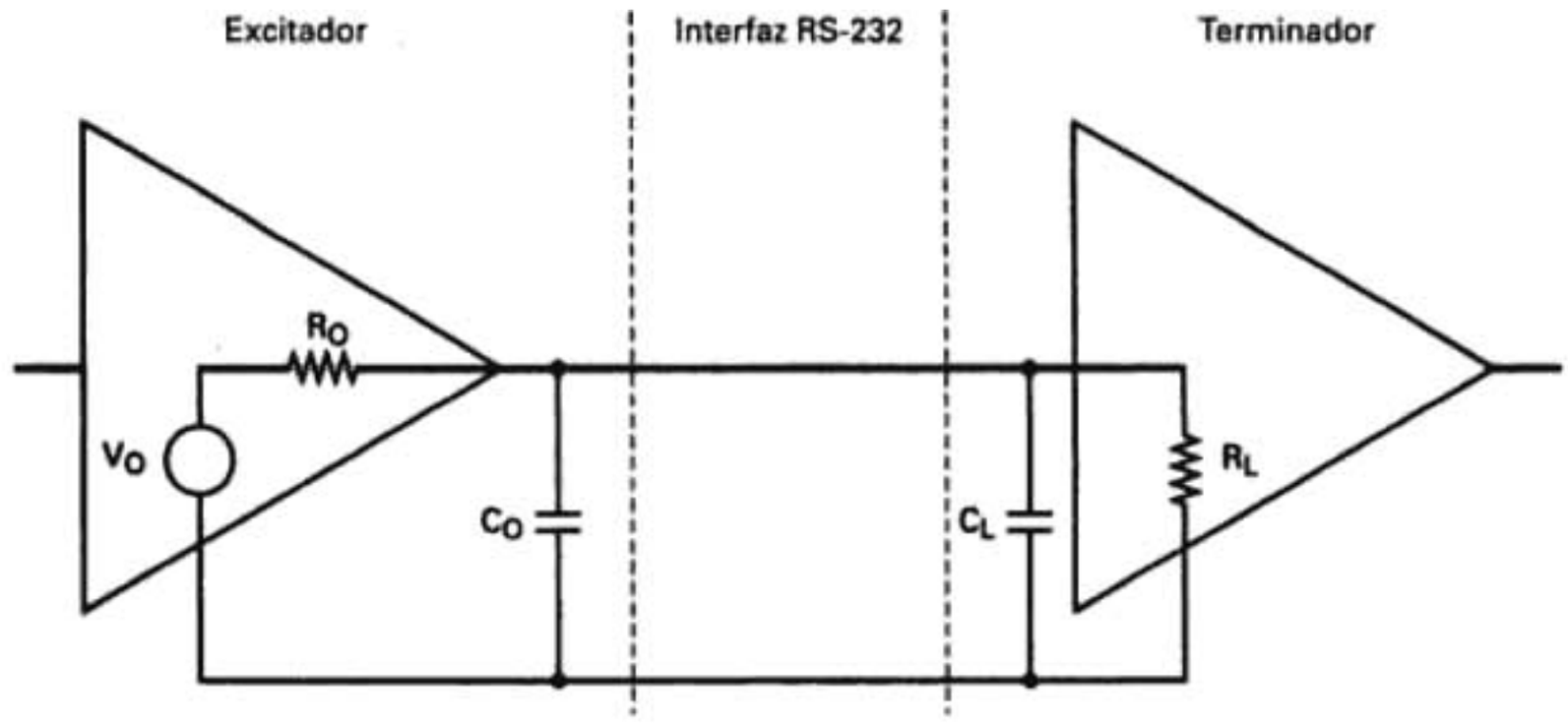


FIGURA 13-15 Especificaciones eléctricas RS-232

TABLA 13-5 Especificaciones de voltaje (V cd) en RS-232

	Terminales de datos		Terminales de control	
	1 lógico	0 lógico	Activar "Enc"	Desactivar "Apag"
Excitador	−5 a −15	+5 a +15	+5 a +15	−5 a −15
Terminador	−3 a −25	+3 a +25	+3 a +25	−3 a −25

Nótese que las líneas de datos usan lógica negativa, y que las líneas de control usan lógica positiva.

Se puede ver en la tabla 13-5 que los límites para un excitador son más incluyentes que los de un terminador. El excitador puede producir cualquier voltaje entre +5 V y +15 V, o −5 V y −15 V cd, y un terminador acepta cualquier voltaje entre +3 V y +25 V, y −3 V y −25 V cd. La diferencia entre los valores de voltaje entre excitador y terminador se llama *margen de ruido*. Este margen reduce la susceptibilidad de la interfaz a los transitorios de ruido en el cable. Los voltajes normales que se usan en las señales de datos y de control son ±7 V cd y ±10 V cd.

Las terminales del cable de interfaz RS-232 se clasifican funcionalmente como tierra, datos, control (acuerdo) o de sincronización. Todas las terminales son unidireccionales: las señales sólo se propagan del DTE al DCE o viceversa. La tabla 13-6 es una lista de las 25 terminales de la interfaz RS-232, sus nombres y la dirección de propagación de la señal, es decir, hacia el DTE o hacia el DCE. Las especificaciones RS-232 designan a las terminales de tierra, datos, control y sincronización con A, B, C y D, respectivamente. Son descripciones "en clave." Es más práctico y útil usar acrónimos para designarlas, que reflejen sus funciones. La tabla 13-6 es una lista de las designaciones CCITT y EIA, y la nomenclatura que se usa con más frecuencia en los Estados Unidos.

Funciones EIA de las terminales RS-232. Veinte de las 25 terminales de la interfaz RS-232 se designan con fines o funciones específicas. Las terminales 9, 10, 11 y 18 no están asignadas; la 1 y 7 son tierras; la 2, 3, 14 y 16 son de datos; la 15, 17 y 24 son de sincronización, y todas las demás terminales asignadas se reservan para señales de control o de acuerdo. Hay dos canales de datos dúplex disponibles en la interfaz RS-232; uno es para datos primarios (información real) y el segundo es para datos secundarios (información de diagnóstico y señales de acuerdo). Las funciones de las 25 terminales se resumen a continuación:

Terminal 1: Tierra de protección. Esta terminal se aterriza en el chasís y se usa para protección contra choques eléctricos. Se debe conectar al tercer conductor de tierra en el sistema eléctrico de ca, en un extremo del cable (sea en el DTE o en el DCE, pero no en ambos extremos).

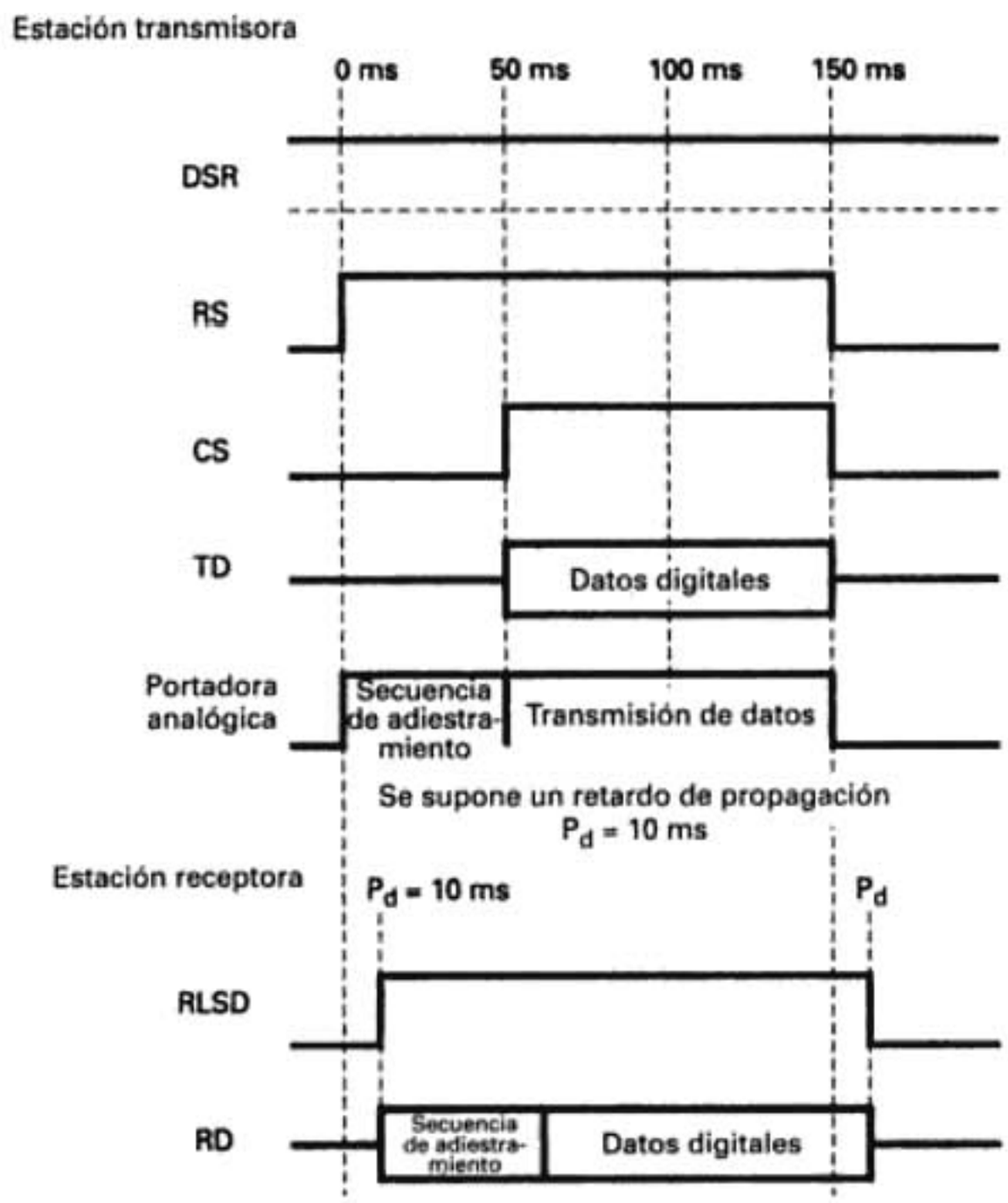


FIGURA 13-16 Diagrama de sincronización: funcionamiento básico de la interfaz RS-232

Retroalimentación remota (RL, de *remote loopback).* Una señal de control del DTE al DCE que pone al DCE (módem) en la condición de anillo remoto. Una retroalimentación remota conecta la terminal de datos de recepción del módem, en su terminal de transmisión de datos, permitiendo que la terminal del extremo remoto haga pruebas de localización de fallas en el módem.

Terminal 22: Indicador de timbrado (RI, de *ring indicator).* Esta terminal se usa con líneas de discado para que el DCE indique al DTE que está entrando una llamada.

Terminal 23: Selector de rapidez de señal de datos (DSRS, de *data signal rate selector).* El DTE usa esta terminal para seleccionar la rapidez de bits (frecuencia de reloj) en la transmisión del DCE.

Terminal 24: Sincronización de elemento de señal de transmisión (SCTE, de *transmit signal element timing).* Las señales del reloj de transmisión se mandan del DTE al DCE por esta terminal, cuando el reloj oscilador maestro está en el DTE.

Terminal 25: Modo de prueba (TM, de *test mode).* Señal de control del DCE al DTE que se activa siempre que se esté haciendo una prueba de anillo local o remoto.

Las terminales 1 a 8 se usan con los módems asíncronos y síncronos. Las terminales 15, 17 y 24 se usan sólo con módems síncronos. Las terminales 12, 13, 14, 16 y 19 sólo se usan cuando el equipo DCE tiene un canal secundario. Las terminales 19 y 22 sólo se usan para conexiones telefónicas de discar.

El funcionamiento básico de la interfaz RS-232 se ve en la fig. 13-16, y su descripción es la siguiente: cuando el DTE tiene datos primarios por mandar, activa la petición de transmitir (t = 0 ms). Después de un retardo predeterminado (50 ms), se activa la terminal CS. Durante el retardo de RS a CS, el módem produce una portadora analógica que se modula por una secuencia única de bits, llamada *secuencia de adiestramiento.* Esta secuencia es para inicializar la línea de comunicaciones y sincronizar los circuitos de portadora y de recuperación de reloj en el

TABLA 13-7 Designaciones de terminales de la interfaz RS-530

Nombre de la señal	Número(s) de terminal
Blindaje	1
Datos transmitidos[a]	2, 14
Datos recibidos[a]	3, 16
Petición de transmitir[a]	4, 19
Borrar para transmitir[a]	5, 13
Listo DCE[a]	6, 22
Listo DTE[a]	20, 23
Tierra de señal	7
Detección de señal recibida en línea[a]	8, 10
Sincronización de elemento de señal de transmisión (fuente DCE)[a]	15, 12
Sincronización de elemento de señal en receptor (fuente DCE)[a]	17, 9
Retroalimentación local[b]	18
Retroalimentación remota[b]	21
Sincronización de elemento de señal de transmisión (fuente DTE)[a]	24, 11
Modo de prueba[b]	25

[a]Circuitos de categoría I (RS-422A)
[b]Circuitos de categoría II (RS-423A)

módulo de recepción. Después del retardo RS/CS, se activa TD y el DTE comienza a transmitir datos. Después de que el DTE de recepción detecta una portadora analógica, se activa RD. Cuando la transmisión se completa (t = 150 ms), RS pasa a bajo, apagando la portadora analógica y apagando CS.

Normas de las interfaces RS-449 y RS-530

Las velocidades actuales de datos han rebasado la capacidad de la interfaz RS-232. Por consiguiente, en 1977 introdujo la EIA la especificación RS-449, con la intención de reemplazar a la RS-232. La interfaz RS-449 usa un conector de 37 terminales que permite tener más funciones, mayores velocidades de transmisión de datos y funcionamiento a mayor distancia. Sin embargo, en la industria nunca se adoptó la norma RS-449, que llegó y se fue casi sin que la notara la mayor parte de la industria de comunicaciones de datos. En consecuencia, en 1987, la EIA introdujo la norma RS-530, que pretendía trabajar a velocidades de datos de 20 kbps a 2 Mbps, con el mismo conector DB-25 de 25 terminales que usaba la interfaz RS-232. La tabla 13.7 describe las 25 terminales de la interfaz RS-530, y sus designaciones.

Sin embargo, la norma RS-530 no incluyó especificaciones eléctricas. En lugar de ello, las que se usan en RS-530 están especificadas por una de las normas RS-422A o RS-423A. La norma RS-422A especifica un cable de interfaz balanceado que funciona a frecuencias de bits hasta de 10 Mbps, y con longitudes hasta de 1200 m. Sin embargo, eso no quiere decir que se puedan transmitir 10 Mbps a 1200 m. A 10 Mbps, la distancia máxima aproximada son unos 15 m, y 1200 m es la máxima distancia a la que se puede transmitir con 90 kbps. La norma RS-423A especifica un cable de interfaz no balanceado que funciona a una rapidez máxima de línea de 100 kbps, y a 90 m de distancia máxima.

La fig. 13-17 muestra el circuito de interfaz digital *balanceada* para la RS-422A, y en la fig. 13-18 se ve el circuito de interfaz digital *desbalanceada* para la RS-423A.

Una interfaz balanceada, como la RS-422A, transfiere información a una *línea de transmisión balanceada*. Con una de éstas, ambos conductores conducen corriente, pero esa corriente pasa en direcciones contrarias en los dos hilos. Con una línea bidireccional *desbalanceada*, un conductor está al potencial de tierra y las corrientes en los dos conductores pueden ser distintas. Las corrientes que van en direcciones contrarias en un par balanceado de conductores se llaman corrientes *de circuito metálico*. Las corrientes que van en la misma dirección se llaman corrientes *longitudinales*. Un par balanceado tiene la ventaja de que la mayor parte de la interferencia de ruido se induce por igual en ambos hilos y se producen corrientes longitudinales que se

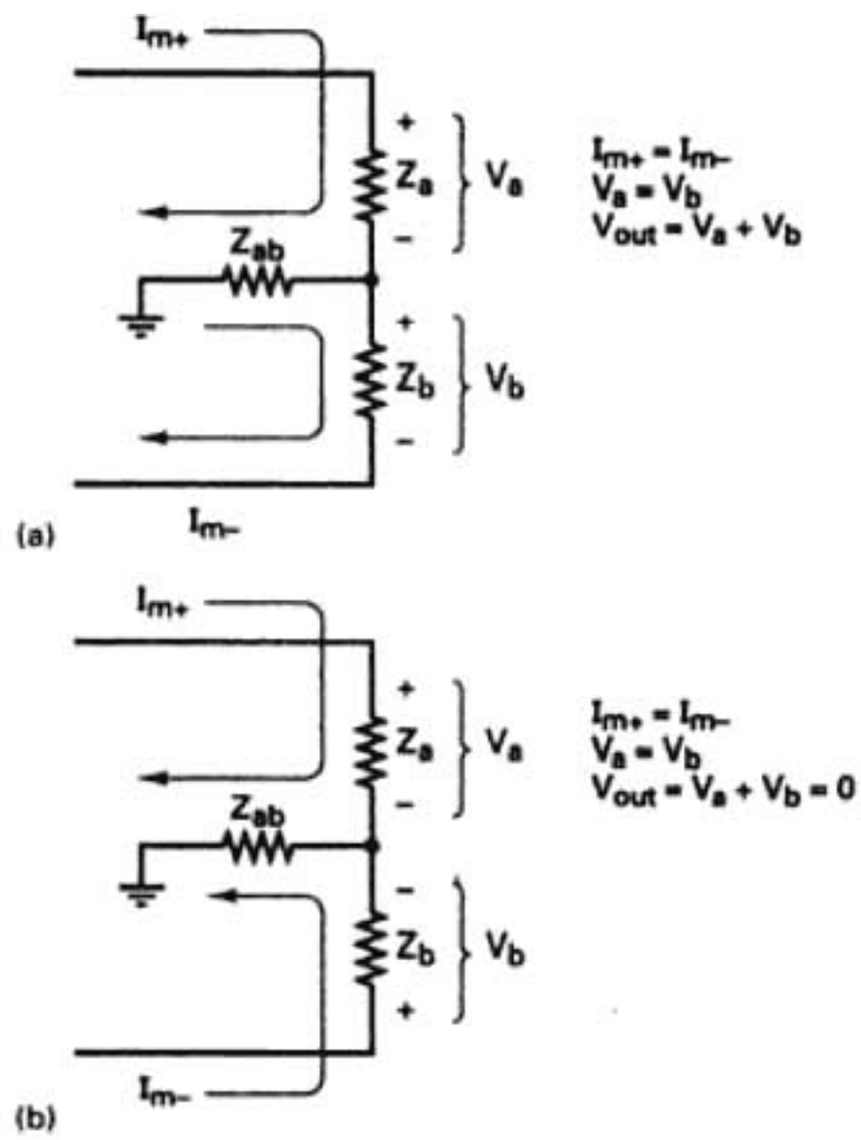


FIGURA 13-19 Resultados de las corrientes metálicas y longitudinales en una línea de transmisión balanceada: (a) corrientes metálicas debidas a voltajes de señal; (b) corrientes longitudinales debidas a voltajes de ruido

TABLA 13-8 Designaciones de terminales de la interfaz CCITT X.21

Circuito de intercambio	Nombre	Dirección
G	Tierra de señal	*
GA	Retorno común de DTE	DTE a DCE
T	Transmitir	DTE a DCE
R	Recibir	DCE a DTE
C	Control	DTE a DCE
I	Indicación	DCE a DTE
S	Sincronización de elemento de señal	DCE a DTE
B	Sincronización de byte	DCE a DTE

*Véanse las recomendaciones X.24

La X.21 sólo especifica seis señales, que se ven en la tabla 13-8. Los datos se transmiten hacia el módem, por la línea de transmisión, y el módem regresa datos por la línea de recepción. Las líneas de control y de indicación son canales de control para las dos direcciones de transmisión. La línea del elemento de sincronización conduce la señal de sincronización de bits (el reloj) y la línea de sincronización de bytes conduce la información de sincronización de caracteres. Hay especificaciones eléctricas para la X.21 en la recomendación X.26 (balanceada) o en la X.27 (desbalanceada).

La ventaja principal de la norma X.21 sobre la RS-232 y RS-530 es que las señales X.21 se codifican en forma digital en serie, lo que prepara la escena para proporcionar nuevos servicios especiales de comunicaciones entre computadoras.

INTERFACES PARALELAS

Las *interfaces paralelas* transfieren datos entre dos dispositivos, con ocho o más bits al mismo tiempo. Esto es, toda una palabra de datos se transmite y se recibe al mismo tiempo, en contraste con bit por bit en las interfaces en modo serie. A veces, la transmisión en paralelo se llama *en serie a nivel de palabra*. Una ventaja obvia de la transmisión en paralelo es que los datos se transmiten con mucha mayor rapidez que con la transmisión serie. Esto, naturalmente, es porque hay una ruta de transmisión para cada bit de la palabra. Por ejemplo, un sistema que use

TABLA 13-9 Líneas de interfaz de impresora Centronics

Terminal número	Terminal de retorno	Señal	Abreviatura	Activo	Tipo	Dirección
1	19	Estrobo	$\overline{STB}$	Bajo	Control	A la impresora
2	20	Bit 0 de datos	d_0		Datos	A la impresora
3	21	Bit 1 de datos	d_1		Datos	A la impresora
4	22	Bit 2 de datos	d_2		Datos	A la impresora
5	23	Bit 3 de datos	d_3		Datos	A la impresora
6	24	Bit 4 de datos	d_4		Datos	A la impresora
7	25	Bit 5 de datos	d_5		Datos	A la impresora
8	26	Bit 6 de datos	d_6		Datos	A la impresora
9	27	Bit 7 de datos	d_7		Datos	A la impresora
10	28	Reconocimiento	$\overline{ACK}$	Bajo	Estado	A la computadora
11	29	Ocupado	BUSY	Alto	Estado	A la computadora
12		Sin papel	PO	Alto	Estado	A la computadora
13		Seleccionar	SLCT	Alto	Estado	A la computadora
14		Alimentación automática	$\overline{AF}$	Bajo	Control	A la impresora
15		No se usa	—	—	—	—
16		Tierra de señal	SG	—	—	—
17		Tierra de chasís	FG	—	—	—
18		+5 V	—	—	—	—
31	30	Inicializar	$\overline{PRIME}$	Bajo	Control	A la impresora
32		Error	$\overline{ERROR}$	Bajo	Estado	A la computadora
33		Tierra de señal	SG	—	—	—
34		No se usa	—	—	—	—
35		No se usa	—	—	—	—
36		Seleccionar	$\overline{SLCTIN}$	Bajo	Control	A la impresora

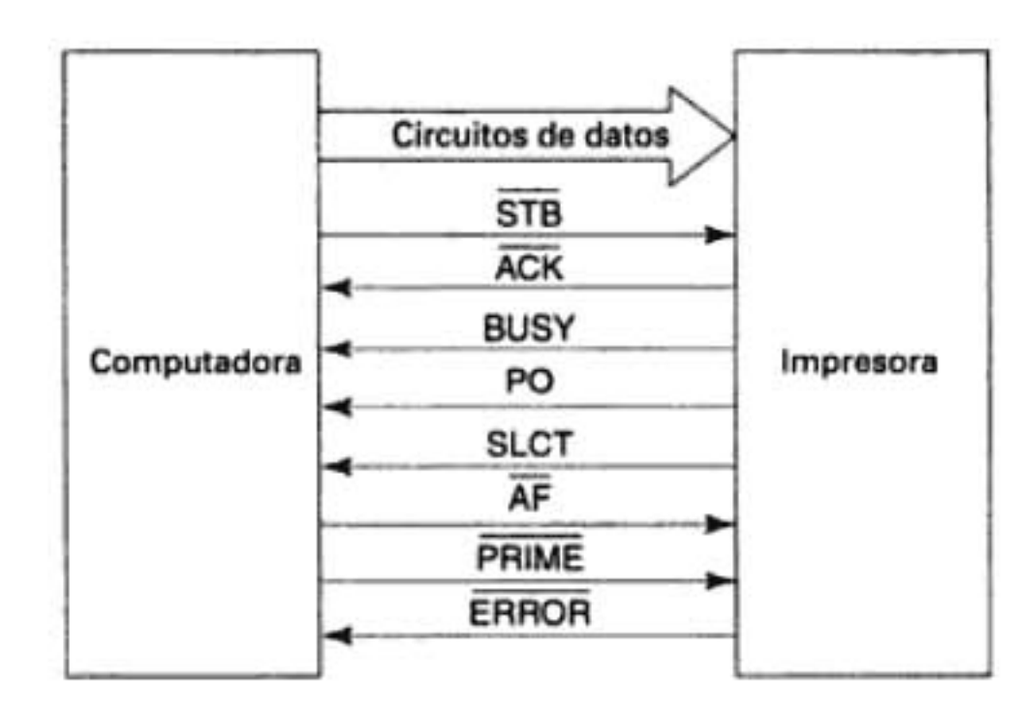


FIGURA 13-21 Líneas de control, de datos y de estado para la interfaz paralela de Centronics

La línea de control *alimentación automática* ($\overline{AF}$) es para una señal activa en bajo, que indica si la impresora efectúa automáticamente una función de alimentación de línea después de recibir un carácter de retorno de carro de la computadora. En el caso normal, el carácter de retorno de carro sólo hace que la cabeza de impresión regrese al lado izquierdo del papel, y un carácter separado de alimentación de línea hace que el papel avance al siguiente renglón. La línea de alimentación automática permite que una señal efectúe ambas funciones. Si la computadora mantiene baja $\overline{AF}$, la impresora responde al carácter de regreso de carro efectuando un regreso de carro y una alimentación de línea.

La línea de control *prime* ($\overline{PRIME}$) (que a veces se llama *inicializar*) es una señal activa en bajo que produce la computadora para borrar la memoria de la impresora, incluyendo su programación y la memoria (*búfer*) de impresión. Al detectar una condición activa en la línea de inicializar, la impresora regresa al estado en que estaba cuando se le conectó la corriente. La línea de control de inicializar se puede usar para abortar un trabajo de impresión mientras está haciéndose.

La cuarta línea de control se llama *seleccionar* ($\overline{\text{SLCTIN}}$). Esta línea casi no se usa. Al hacerlo, la impresora debe ver una señal baja en ella para aceptar datos de la computadora. Muchas impresoras tienen un interruptor desactivador interno, que siempre conecta a tierra la línea $\overline{\text{SLCTIN}}$.

Líneas de estado. Todas las líneas de estado son unidireccionales, y llevan información de la impresora a la computadora. Las usa la impresora para decir a la computadora qué está haciendo. Por ejemplo, la línea de estado *reconocer* ($\overline{\text{ACK}}$, de *acknowledge*) es una de bajo activo para la respuesta de la impresora al recibir una señal $\overline{\text{STB}}$ activa de la computadora. La impresora usa $\overline{\text{ACK}}$ para indicar a la computadora que ha procesado el último byte de datos que recibió por las líneas de datos de la interfaz, y que está lista para otro carácter.

La línea de estado *ocupado* (BUSY) es una de señal de control con activo alto, que se vuelve alta cuando la impresora está ocupada y no puede aceptar datos de la computadora. Hay cuatro condiciones que pueden hacer que la impresora esté ocupada:

1. La impresora está ingresando datos de las líneas de datos, o la memoria de la impresora está llena y no puede aceptar más datos.
2. La impresora está imprimiendo o procesando datos en alguna forma.
3. La impresora está apagada o *fuera de línea* en alguna forma. Muchas impresoras tienen un interruptor fuera de línea que evita que reciban datos de impresión.
4. La línea $\overline{\text{ERROR}}$ de la impresora está en bajo.

La línea de control *falta de papel* (PO, de *paper out*) se explica por sí misma; pasa a activo siempre que se acaba el papel en la impresora. Cuando se activa PO también se activa la línea $\overline{\text{ERROR}}$.

La línea de control *seleccionar* (SLCT) es una línea de estado con alto activo, que sólo indica si la impresora está seleccionada. La impresora mantiene alta la señal SLCT cuando está encendida y baja cuando está fuera de línea. La línea de estado de $\overline{\text{ERROR}}$ es de señal con activo bajo, para indicar que hay un problema con la impresora. La impresora mantiene alta $\overline{\text{ERROR}}$ durante su funcionamiento normal. Las siguientes condiciones pueden causar que se active $\overline{\text{ERROR}}$:

1. La impresora está fuera de línea.
2. La impresora no tiene papel.
3. Hay algún otro problema no definido en la impresora, que le evita funcionar en forma correcta.

Líneas diversas. Las terminales 20 a 27 de la interfaz paralela Centronics son de retorno de señal, para las terminales de datos de d_0 a d_7, respectivamente. Las terminales 19, 28, 29 y 30 son las trayectorias de retorno de las líneas $\overline{\text{STB}}$, $\overline{\text{ACK}}$, BUSY y $\overline{\text{PRIME}}$. Las terminales 16 y 33 son líneas de tierra de señal, que sirven como trayectoria común de retorno para las líneas que no tienen sus retornos dedicados. La terminal 17 es la tierra del chasís (FG, de *frame ground*), que conecta eléctricamente el chasís de la computadora con el de la impresora, para reducir la posibilidad de choques eléctricos. La terminal 18 es de suministro de corriente de +5 V cd, de la computadora a la impresora. Esta corriente de +5 V cd se usa para energizar los circuitos de estado de la impresora, cuando está apagada, pero no es para suministrar toda la energía necesaria para el funcionamiento general de la impresora. En general, la terminal 18 no se usa.

Interconexión. La interfaz Centronics se diseñó originalmente para usar un conector de 36 terminales. Sin embargo, las primeras computadoras personales usaron conectores DB de 25 terminales en su puerto de impresora. En consecuencia, hubo que omitir 11 terminales de la interfaz. Hoy, los cables de impresora se venden con un conector Amphenol de 36 terminales en el cable que da a la impresora, y el conector DB-25 en el lado de la computadora.

La tabla 13-10 muestra las asignaciones de terminales para los conectores Amphenol de 36 terminales y el DB-25 adaptado. Allí se ve que las terminales Centronics que no se usan se han omitido, y que sólo se usan cuatro de las trayectorias individuales de retorno (tierras). Tampoco

TABLA 13-10 Conexión de un cable Centronics a un conector DB-25

Conector Amphenol	Señal	Conector DB-25
1	Estrobo	1
2	Bit 0 de datos	2
3	Bit 1 de datos	3
4	Bit 2 de datos	4
5	Bit 3 de datos	5
6	Bit 4 de datos	6
7	Bit 5 de datos	7
8	Bit 6 de datos	8
9	Bit 7 de datos	9
10	Reconocimiento	10
11	Ocupado	11
12	Falta papel	12
13	Seleccionar	13
14	Avance automático	14
15	No se usa	No conectado
16	Tierra de señal	No conectado
17	Tierra de chasís	No conectado
18	+5 V	No conectado
19	Regreso de estrobo	19
20	Regreso de bit 0 de datos	No conectado
21	Regreso de bit 1 de datos	20
22	Regreso de bit 2 de datos	No conectado
23	Regreso de bit 3 de datos	21
24	Regreso de bit 4 de datos	No conectado
25	Regreso de bit 5 de datos	22
26	Regreso de bit 6 de datos	No conectado
27	Regreso de bit 7 de datos	23
28	Regreso de reconocimiento	No conectado
29	Regreso de ocupado	24
30	Regreso de inicializar	25
31	Inicializar	16
32	Error	15
33	Tierra de señal	18
34	No se usa	No conectado
35	No se usa	No conectado
36	Seleccionar	17

están conectados la tierra de señal y de chasís, la trayectoria de retorno de señal de reconocimiento, y el suministro de +5 V cd. Como las impresoras son dispositivos relativamente lentos, la versión Centronics para computadora personal es más que adecuada para mandar datos a la impresora.

Bus IEEE 488

El bus IEEE 488 usa ocho líneas bidireccionales de datos conectadas en paralelo, para interconectar hasta a 15 dispositivos remotos (por lo general, equipos de prueba controlados por computadora). Como los diversos equipos interconectados con el IEEE 488 están cercanos entre sí, la interfaz fue diseñada para una distancia máxima de 2.10 m entre dispositivos adyacentes, y la longitud máxima de todo el bus es de unos 20 metros (65 pies). Los dispositivos se conectan al bus con un conector de cable de cinta con 24 terminales. La tabla 13.11 resume las características del bus IEEE 488.

En este bus, el *controlador* determina qué dispositivo puede transmitir datos, y cuáles son los dispositivos receptores. El controlador selecciona a un solo dispositivo para mandar datos por el bus. El dispositivo seleccionado se llama *usuario* y sólo puede haber un usuario en cualquier momento. Se puede seleccionar cualquier cantidad de receptores (llamados *escuchas*) en cualquier momento. Los dispositivos que no están programados como usuarios, escuchas o

TABLA 13-11 Características del bus IEEE 488

Niveles lógicos	TTL
Líneas de datos	8, bidireccionales
Cantidad máxima de dispositivos	15
Distancia entre dispositivos adyacentes	2 m máximo
Longitud total máxima	20 m
Conector	Cinta, 24 terminales
Velocidad máxima con excitadores de colector abierto	24 kbps
Velocidad máxima con excitadores triestado	1 MBPs

TABLA 13-12 Dispositivos IEEE 488

Dispositivo	Función
Controlador	A cargo del bus. Especifica cuáles dispositivos se programan como usuarios, escuchas o están en el modo de reserva. Sólo puede haber un controlador al mismo tiempo.
Usuario	Manda datos por el bus a los escuchas. Sólo puede haber un usuario a la vez.
Escuchas	Reciben datos del usuario. Pueden haber muchos escuchas.
Reserva	Todos los dispositivos que no estén programados como controlador, usuario o escucha, están en modo de reserva.

controladores, se ponen en el *modo de reserva*. Las características de las cuatro clases de dispositivos se resumen en la tabla 13.12. El controlador puede cambiar la función de los dispositivos que supervisa. Por ejemplo, puede reprogramar al usuario como escucha y nombrar nuevo usuario a un dispositivo distinto. El controlador hasta puede rehusar el control del bus y designar a otro dispositivo para que sea el controlador. En este caso, el anterior controlador se puede transformar en un usuario, un escucha o se puede poner en modo de reserva.

Modos de operación. El bus IEEE sólo tiene dos modos de operación: *orden* y *datos*. El controlador conmuta la interfaz entre los dos modos, usando la línea de control de administración de bus llamada *atención* ($\overline{\text{ATN}}$). Cuando el controlador pone baja $\overline{\text{ATN}}$, todos los dispositivos conectados a la interfaz pasan al modo de orden (o *comando*). Cuando el controlador pone alta $\overline{\text{ATN}}$, todos los dispositivos pasan al modo de datos.

Cuando está en modo de órdenes, el controlador puede programar a otros dispositivos conectados con la interfaz. Por ejemplo, podría usar el modo de orden para programar la frecuencia de un generador de funciones, o el intervalo de voltajes de un voltímetro digital. El controlador también podría programar que un dispositivo funcionara como usuario o como escucha. Cada dispositivo conectado con la interfaz tiene una dirección exclusiva. En el modo de orden, el controlador dirige las órdenes a un dispositivo específico, mandando la dirección del mismo por el bus de datos. Cuando está en el modo de orden, toda la información que pasa por el bus de datos es una orden o una dirección. Cuando está en modo de datos, la única información que se permite pasar por el bus de datos es de datos, que van del usuario al o a los escuchas.

Líneas de la interfaz IEEE 488. La tabla 13-13 es una lista de las asignaciones de terminal para la interfaz IEEE 488. En esencia, hay tres grupos de circuitos (líneas): de datos, de acuerdo (que a veces se llaman *líneas de transferencia de bytes*) y de circuitos de administración de interfaz.

Circuitos de datos. Hay ocho líneas bidireccionales de datos, numeradas de DI01 a DI08, siendo DI08 la más significativa. La tarea de las líneas de datos es conducir datos de ocho bits en paralelo, del usuario al o a los escuchas, o llevar direcciones y órdenes del controlador a los demás dispositivos conectados con la interfaz.

Circuitos de acuerdo. En el bus IEEE 448 hay tres circuitos de acuerdo: no listo para datos ($\overline{\text{NRFD}}$, de *not ready for data*), datos no aceptados ($\overline{\text{NDAC}}$), y datos válidos ($\overline{\text{DAV}}$). La función de los circuitos de acuerdo es coordinar la transferencia de datos del usuario al o a los escuchas. La señal de control $\overline{\text{NRFD}}$ es de activo bajo, que produce el escucha cuando no está listo para aceptar datos de los circuitos de datos. Las terminales de salida $\overline{\text{NRFD}}$ en los

TABLA 13-13 Asignación de terminales de la interfaz IEEE 488

Terminal	Nombre de la línea	Abreviatura	Tipo
1	Bit 1 de datos (Menos significativo)	DI01	Datos
2	Bit 2 de datos	DI02	Datos
3	Bit 3 de datos	DI03	Datos
4	Bit 4 de datos	DI04	Datos
5	Fin de identificar	$\overline{\text{EOI}}$	Admón. de interfaz
6	Datos disponibles	$\overline{\text{DAV}}$	Acuerdo
7	No listo para datos	$\overline{\text{NRFD}}$	Acuerdo
8	Sin datos aceptados	$\overline{\text{NDAC}}$	Acuerdo
9	Borrar interfaz	$\overline{\text{IFC}}$	Admón. de interfaz
10	Petición de servicio	$\overline{\text{SRQ}}$	Admón. de interfaz
11	Atención	$\overline{\text{ATN}}$	Admón. de interfaz
12	Blindaje (tierra física)		
13	Bit 5 de datos	DI05	Datos
14	Bit 6 de datos	DI06	Datos
15	Bit 7 de datos	DI07	Datos
16	Bit 8 de datos (más significativo)	DI08	Datos
17	Activar remoto	$\overline{\text{REN}}$	Admón. de interfaz
18	Tierra de retorno para DAV		
19	Tierra de retorno para NRFD		
20	Tierra de retorno para NDAC		
21	Tierra de retorno para IFC		
22	Tierra de retorno para SRQ		
23	Tierra de retorno para ATN		
24	Tierra de señal		

dispositivos se combinan con OR en sus conductores, de modo que cualquier escucha conectado al bus pueda cambiar la línea $\overline{\text{NRFD}}$ a bajo. Si hay más de un escucha, la configuración OR en conductor asegura que la línea $\overline{\text{NRFD}}$ permanezca baja mientras el escucha más lento no esté listo para recibir datos.

El circuito de datos no aceptados ($\overline{\text{NDAC}}$, de *not data accepted*) es de señal activa baja, controlado por el o los escuchas. Las señales $\overline{\text{NDAC}}$ indican al usuario que todavía los escuchas no han aceptado los datos, y que debe continuar manteniéndolos en el bus. También, $\overline{\text{NDAC}}$ se combina con OR en los conductores, para todos los dispositivos de la interfaz, para que cualquier escucha pueda mantener baja la línea hasta que el más lento haya aceptado los datos del bus de interfaz.

La línea datos disponibles ($\overline{\text{DAV}}$, de *data available*) es una línea de control con estado activo bajo, que usa el usuario en el modo de datos, o el controlador en el modo de orden, para indicar a los demás dispositivos en el bus que hay por leer un byte de datos, una dirección o una orden en el bus de datos.

Circuitos de administración de interfaz. El bus IEEE 488 tiene cinco circuitos de administración de interfaz, que controlan ciertas operaciones de los dispositivos conectados, y administran el flujo de información entre ellos.

1. *Atención* ($\overline{ATN}$). El controlador usa $\overline{\text{ATN}}$ para intercambiar los modos de datos y de órdenes en el bus. Cuando $\overline{\text{ATN}}$ está alto, la interfaz está en el modo datos, y el usuario puede mandar datos por el bus a los escuchas. Cuando $\overline{\text{ATN}}$ está bajo, la interfaz IEEE está en modo de órdenes, y en este estado el controlador sólo puede mandar direcciones y órdenes por el bus de datos.
2. *Borrar interfaz* ($\overline{IFC}$, de *interface clear*). Cuando el controlador activa $\overline{\text{IFC}}$, todos los dispositivos conectados a la interfaz suspenden sus operaciones, y el bus pasa al estado inactivo, en cuyo caso no pueden hacerse comunicaciones. $\overline{\text{IFC}}$ funciona en forma parecida a una línea de restablecer.
3. *Activar remoto* ($\overline{REN}$, de *remote enable*). Es una señal con activo bajo para que el controlador cambie a los dispositivos conectados con la interfaz, entre operación

local y remota. Cuando se activa $\overline{REN}$, todos los dispositivos se ponen en operación remota, y funcionan bajo las órdenes del controlador. Cuando $\overline{REN}$ está alta (inactiva), todos los dispositivos (por lo general, equipos de prueba) están bajo control local, y sólo responden a las perillas de control de sus propios tableros, permitiendo así a los operadores controlarlos en forma directa, sin tener que programarlos a través de una computadora.

4. *Petición de servicio* ($\overline{SRQ}$, de *service request*). Todo dispositivo conectado con la interfaz puede activar $\overline{SRQ}$ para interrumpir el funcionamiento normal de ésta, e indicar al controlador que requiere su atención. Cuando el controlador siente una señal de $\overline{SRQ}$ activa, determina cuál dispositivo pidió el servicio, y por qué, poniendo la interfaz en modo de órdenes y encuestando a cada dispositivo, en forma individual. Cuando el controlador encuentra el dispositivo que activo la línea $\overline{SRQ}$, ese dispositivo responde mandando claves por el bus de datos, que indican la acción que debe tomar el controlador.
5. *Fin de identificar* ($\overline{EOI}$, de *end of identify*). El circuito $\overline{EOI}$ tiene dos funciones. Cuando está en modo de datos, el usuario tiene el control de $\overline{EOI}$ y lo activa cuando manda su último byte de datos en determinado mensaje. Cuando está en el modo de órdenes, el controlador puede activar $\overline{EOI}$ para efectuar una encuesta de grupo, para obtener información de varios dispositivos al mismo tiempo.

LA RED TELEFÓNICA

En su forma más simple, la comunicación de datos es la transmisión de información digital entre dos DTE (DTE = equipo transmisor de datos). Los DTE pueden estar separados por algunos metros o por varios miles de millas. En la actualidad hay una cantidad insuficiente de medios de transmisión que lleven información digital desde la fuente hasta el destino, en forma digital. Por consiguiente, la alternativa más cómoda es usar la red telefónica pública actual (PTN, de *public telephone network*) como medio de transmisión para circuitos de comunicaciones de datos. Desafortunadamente, la PTN fue diseñada (y en su mayor parte construida) mucho antes del advenimiento de las comunicaciones en gran escala. Era para transferir señales telefónicas de voz, y no datos digitales. Por consiguiente, para usar la PTN para comunicar datos, éstos se deben convertir a una forma más adecuada para transmitirse en sistemas de portadora analógica.

La siguiente explicación se limita a las comunicaciones usando la PTN. Se definirá a *Telco* como todas las empresas telefónicas que forman la PTN. Telco ofrece dos categorías generales de servicio: *marcado directo a distancia* (DDD) y *servicio de línea privada.* Originalmente el DDD incluía sólo las centrales e instalaciones necesarias para hacer una llamada telefónica de larga distancia sin la ayuda de un operador Telco. Ahora, el DDD incluye a toda la red pública conmutada, esto es, cualquier servicio asociado a un número telefónico. Los servicios de línea privada están dedicados a un solo usuario.

Red DDD

La red DDD, llamada con frecuencia *red conmutada* o *red de marcar* impone fuertes limitaciones que se deben superar para poderla usar en comunicación de datos. Es conveniente tener una comprensión básica del funcionamiento eléctrico de la red telefónica. La red DDD se puede dividir en cuatro secciones principales: instrumentos, matrices de conmutación, lazos locales y circuitos troncales. Un *instrumento* es el dispositivo que se usa para originar y recibir señales, como puede ser un teléfono. El instrumento se llama con frecuencia *equipo de estación*, y el lugar del instrumento se llama *estación*. Una *matriz de conmutación* es una matriz programada que proporciona una trayectoria temporal de señal. Un *lazo local* es la trayectoria de transmisión dedicada entre un instrumento y la matriz de conmutación más cercana. Un *circuito troncal* es una trayectoria de transmisión entre dos matrices de conmutación. Las matrices de conmutación están en las oficinas centrales de Telco y se clasifican como locales, de dos secciones o suburbanas. Una matriz de conmutación *local* da servicio a un área limitada. El tamaño del área queda determinado por la cantidad de teléfonos que se necesiten o se deseen en determinada zona geográfica. Telco llama central de área a esas áreas. Una central de área es una matriz de

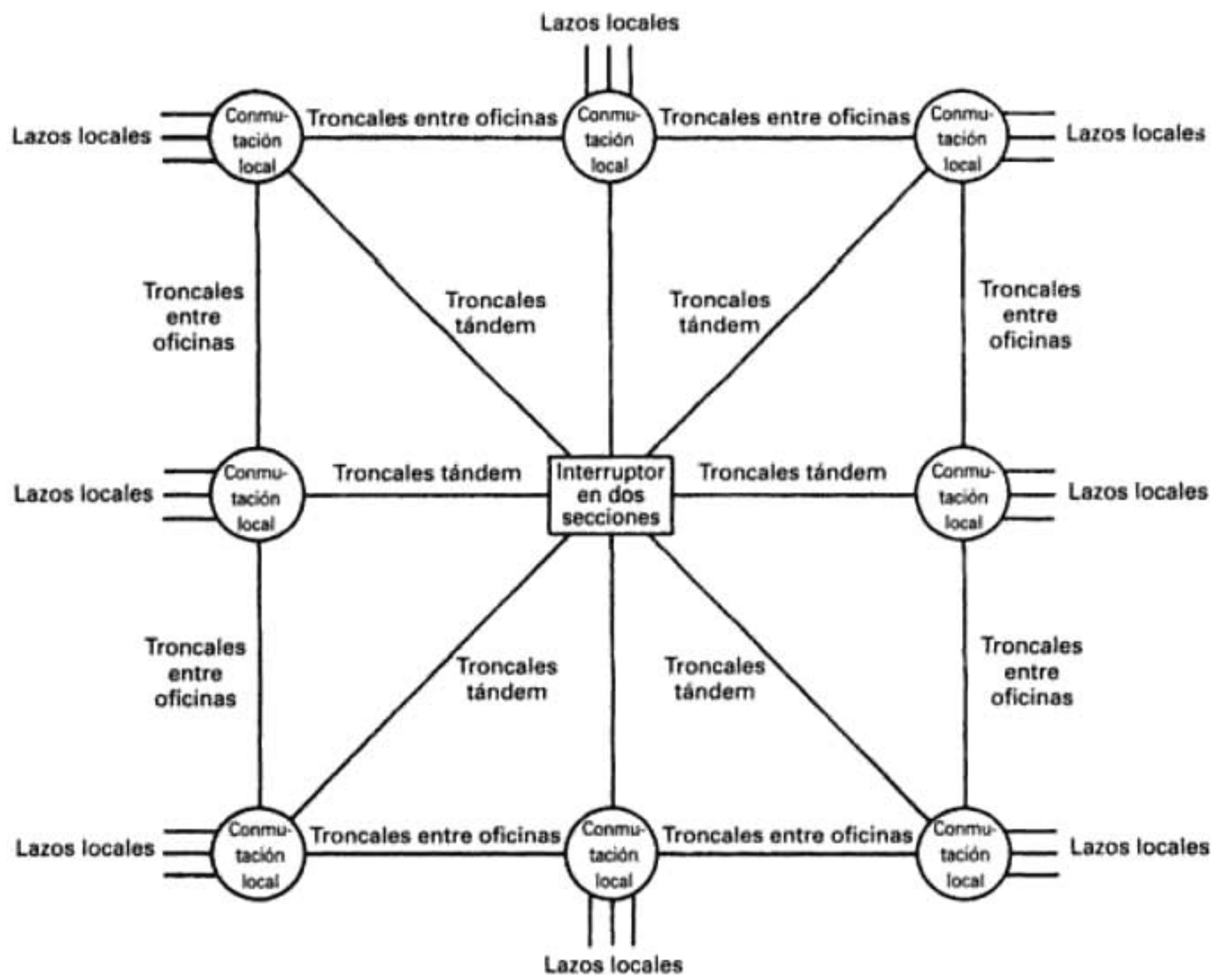


FIGURA 13-23 Sistema telefónico público mostrando el uso de un interruptor en dos secciones para facilitar las llamadas interzonales

la señal por la matriz de conmutación se mantiene mientras ambos lazos estén cerrados. Cuando alguno de los instrumentos pasa a colgado, se interrumpe la trayectoria de la señal.

¿Y si los números llamador y llamado no reciben servicio de la misma matriz de conmutación? En general, una comunidad sólo tiene servicio de una sola compañía telefónica local. La comunidad se divide en zonas; cada zona recibe servicio de una matriz de conmutación distinta. La cantidad de zonas establecidas en determinada comunidad queda determinada por la cantidad de estaciones a las que se da servicio y de su densidad. Si un suscriptor de una zona desea llamar a una estación en otra zona, se requiere un mínimo de dos matrices de conmutación. La estación llamadora recibe supervisión de descolgado, y los pulsos de marcado se producen como se describió arriba. La matriz de conmutación en la zona llamadora reconoce que el prefijo del número telefónico está servido por otra matriz de conmutación. Hay dos formas con las que puede completar la llamada la matriz servidora de conmutación. Puede localizar un circuito troncal directo (entre oficinas) hasta la matriz de conmutación de la zona de destino, o puede encauzar la llamada por un interruptor de dos secciones. Un *interruptor de dos secciones* es el interruptor del interruptor. Es una matriz de conmutación para interconectar matrices de conmutación. Los circuitos troncales que terminan en interruptores en dos secciones se usan para interconectar las matrices de conmutación. Los circuitos troncales que terminan en los interruptores de dos secciones se llaman *troncales en dos secciones*, o *troncales en tándem*. En el caso normal, sólo hay circuitos troncales directos entre zonas adyacentes. Si una llamada debe pasar a través de más de una zona, se debe usar un interruptor de dos secciones (vea la fig. 13-23). Si no hay troncales directas entre las matrices de conmutación de origen y de término, y si no hay un interruptor común de dos secciones, la llamada se clasifica como *llamada suburbana* y no se puede completar tal como se marcó. Las llamadas suburbanas implican cuotas adicionales, y el número marcado debe estar precedido de un "1".

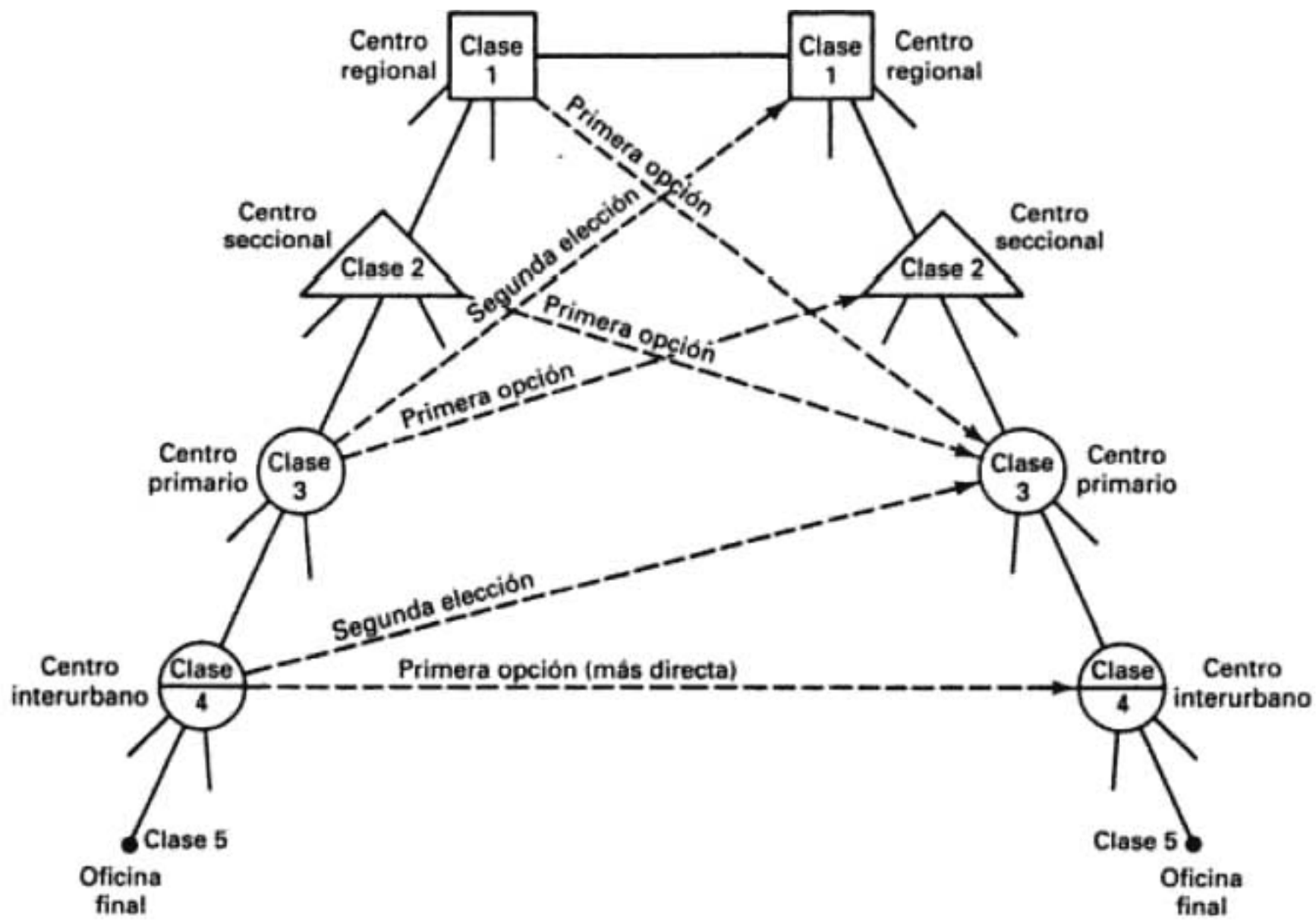


FIGURA 13-24 Jerarquía de conmutación de telefonía pública, mostrando algunas opciones posibles de trayectoria para completar una llamada entre dos oficinas finales

El número telefónico prefijo identifica qué matriz de conmutación específica da servicio a una estación. Se pueden generar 1000 prefijos con tres dígitos, del 000 al 999, para las matrices de conmutación. Una sola central metropolitana puede dar servicio de 20 a 30 prefijos. En Estados Unidos hay más de 20,000 matrices locales de conmutación. Es obvio que se requiere más codificación para diferenciar entre el mismo prefijo y la extensión, en dos partes distintas del país. En Estados Unidos se asigna un código de área adicional, de tres dígitos, a cada número telefónico. El código de área antecede al prefijo, pero sólo se debe incluir cuando las llamadas se dirigen fuera del área de la estación de origen. Cuando un número telefónico marcado está precedido de un "1", se encauza de la matriz de conmutación local a una interurbana, mediante una *troncal de conexión interurbana* o *suburbana*. El plan actual de Telco incluye cinco rangos o clases de centros de conmutación. Desde la clasificación más alta a la más baja, son el centro regional, el centro seccional, el centro primario, el centro interurbano y la oficina. Las matrices locales de conmutación se llaman oficinas extremas (centrales). Todas las matrices de conmutación interurbanas son capaces de funcionar como interruptores de dos secciones hacia otras matrices interurbanas.

El plan de conmutación de Telco incluye una jerarquía que permite cierto grado de selección de ruta cuando se establece una llamada de larga distancia (véase la fig. 13-24). La opción no se ofrece al suscriptor, sino son las matrices interurbanas, con programas de traslación, las que seleccionan la ruta óptima disponible en el momento de hacer la llamada. La ruta óptima no necesariamente es la más corta. Es la que requiere la cantidad mínima de matrices de conmutación. Si no se puede completar una llamada por no estar disponibles los circuitos troncales necesarios, la matriz de conmutación local manda una señal de equipo ocupado a la estación que llama. Una señal de equipo ocupado se parece a la de estación ocupada, pero se repite con una frecuencia de 120 ppm. La condición peor que se puede encontrar al completar una llamada de larga distancia es cuando se requieren siete troncales interurbanas de dos secciones. De acuerdo con las estadísticas de Telco, la probabilidad de que eso suceda es 1 en 100,000. Como las traslaciones programadas en las máquinas de conmutación automática permiten usar rutas alternas, y como cada ruta incluye muchos circuitos troncales distintos, la probabilidad de usar las mismas instalaciones en llamadas idénticas es escasa. Esta es una desventaja obvia cuando se usa la red telefónica PTN para transmitir datos, porque se introducen inconsistencias en los parámetros

de una llamada a otra. Telco garantiza que los parámetros de transmisión de cada lazo local y circuito troncal son mejores que los requisitos mínimos de un canal de comunicaciones de grado de voz (VG, de *voice-grade*). Sin embargo, se pueden necesitar de dos a nueve instalaciones separadas en tándem (en dos secciones), para completar una llamada telefónica (véase la fig. 13-25). Como los deterioros de transmisión son aditivos, es muy posible que los parámetros generales de transmisión de una conexión telefónica establecida a través de la red pública conmutada, puedan ser menores que la norma. Como las rutas de transmisión varían de una llamada a otra es difícil compensar los deterioros en la línea. Los suscriptores a la red DDD (DDD = marcado directo a distancia) pueden alquilar un lazo dedicado, de su estación hasta la matriz de conmutación más cercana de Telco. Todas las instalaciones adicionales que se requieren para que los suscriptores completen una llamada son de ellos, en forma temporal. El suscriptor usa esas instalaciones sólo durante la llamada, y otros usuarios de la red pueden tenerlas a su disposición después. A esas instalaciones temporales se les llama *troncales de uso común*, porque se comparten entre todos los suscriptores de la red.

Los estados transitorios de conmutación asociados con las matrices de conmutación son otra de las desventajas de usar la red pública conmutada para transmisión de datos. Las matrices de conmutación más antiguas eran electromecánicas. Se usaban contactos de relevadores mecánicos para establecer una trayectoria de señal. Los cierres de contacto en las máquinas de conmutación inducían interferencia estática que se pasaba a las trayectorias de señal adyacentes. La electricidad estática causaba ruido de pulsos que producía errores de transmisión en las señales de datos. Telco está cambiando con rapidez a los *sistemas electrónicos de conmutación*, o *sistemas de conmutación electrónica* (ESS, de *electronic switching systems*). De ningún modo las máquinas ESS son silenciosas, pero son una tremenda mejora respecto a las máquinas electromecánicas anteriores.

Para que se complete una llamada interurbana, el número telefónico marcado se debe transferir de una a otra matriz de conmutación. Al último, la matriz de conmutación en la oficina de destino final pide el prefijo y los números de extensión, para establecer la conexión final. Las rutas de transmisión entre máquinas de conmutación son, con mucha frecuencia, sistemas de portadora: enlaces de microondas, cables coaxiales o portadoras T digitales. Los enlaces de microondas y los cables coaxiales usan portadoras analógicas, y se acoplan a la ca. En consecuencia, no se pueden

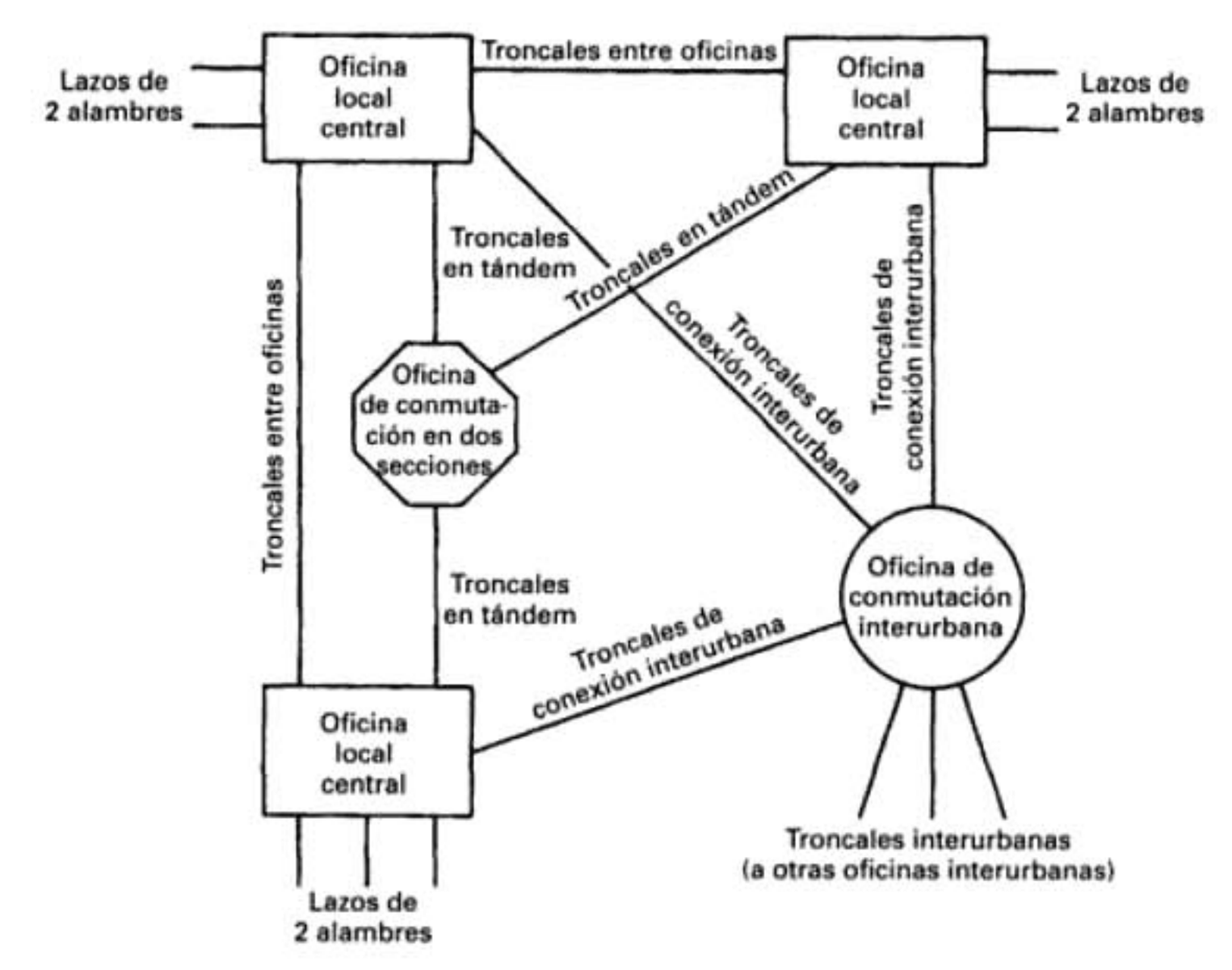


FIGURA 13-25 Esquema normal de conmutación mostrando las relaciones entre las conmutaciones locales, en dos secciones e interurbanas, en la red telefónica pública

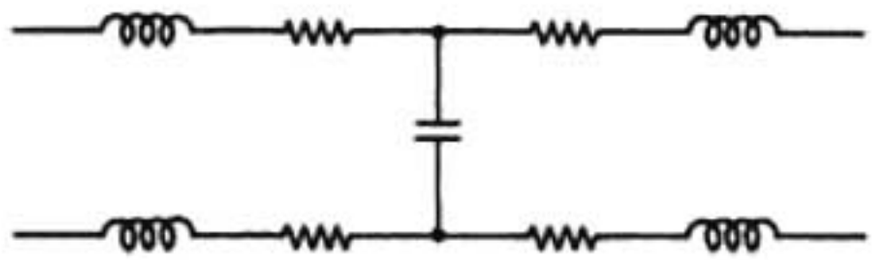


FIGURA 13-27 Características eléctricas de una línea de transmisión

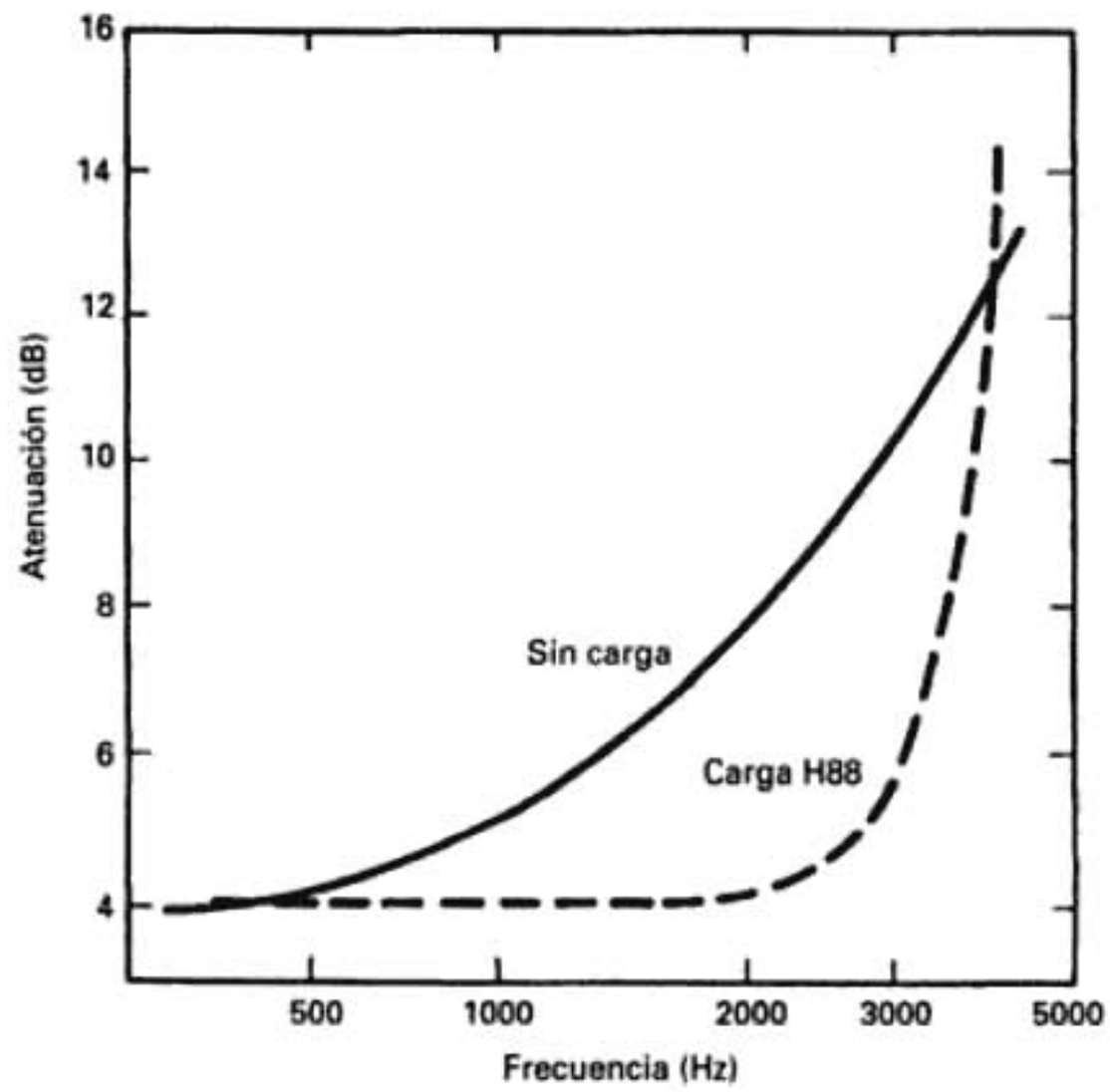


FIGURA 13-28 Frecuencia en función de la atenuación, para 12000 pies de alambre calibre 26

una señal de 3000 Hz sufre una atenuación 6 dB mayor que una de 500 Hz en la misma línea. El cable actúa como filtro de paso bajo para la señal. Con estudios extensos de la atenuación en pares de cables se ha demostrado que se puede obtener una reducción de la atenuación si se agregan inductores en serie con el alambre. A esto se le llama *carga*. Un cable cargado se identifica agregando las letras H, D o B, y un valor de inductancia al calibre del alambre. H, D y B indican que las bobinas de carga están a cada 6000 pies, 4500 pies y 3000 pies, respectivamente. En general, la cantidad de inductancia agregada en serie es 44 mH, 88 mH o 135 mH. Un par de cable con la designación 26H88 es de alambre calibre 26, con 88 mH de inductancia en serie a cada 6000 pies. Si se usa un cable cargado, una señal de 3000 Hz sólo sufrirá una atenuación 1.5 dB mayor que una de 500 Hz. Nótese que las características de pérdidas en función de frecuencia, para un cable cargado, son relativamente planas hasta unos 2000 Hz.

Las características de filtro pasabajas de un cable afectan también las características de distorsión de fase en función de la frecuencia de una señal. La cantidad de *distorsión de fase* es proporcional a la *longitud* y al *calibre* del alambre. También, al cargar un cable se afectan las características de la fase de una línea. Con frecuencia, la empresa telefónica debe agregar igualadores de ganancia y de retardo a un circuito, para satisfacer los requisitos mínimos. Los igualadores introducen discontinuidades u ondulaciones pequeñas en las características de paso de banda de un circuito. Los igualadores automáticos en los módems son sensibles a esta condición y con mucha frecuencia, un circuito sobreigualado causa tantos problemas a una señal de datos como uno subigualado.

TABLA 13-14 Límites de parámetros con ancho de banda tipo C

	Distorsión por atenuación (respuesta de frecuencia) en relación con 1004 Hz		Distorsión por retardo de envolvente	
Acondicionamiento de canal	Intervalo de frecuencias (Hz)	Variación (dB)	Intervalo de frecuencias (Hz)	Variación (μs)
Básico	500–2500	+2 a −8	800–2600	1750
	300–3000	+3 a −12		
C1	1000–2400	+1 a −3	1000–2400	1000
	300–2700	+2 a −6	800–2600	1750
	300–3000	+3 a −12		
C2	500–2800	+1 a −3	1000–2600	500
	300–3000	+2 a −6	600–2600	1500
			500–2800	3000
C3 (línea de acceso)	500–2800	+.5 a −1.5	1000–2600	110
	300–3000	+.8 a −3	600–2600	300
			500–2800	650
C3 (troncal)	500–2800	+.5 a −1	1000–2600	80
	300–3000	−.8 a −2	600–2600	260
			500–2800	500
C4	500–3000	+2 a −3	1000–2600	300
	300–3200	+2 a −6	800–2800	500
			600–3000	1500
			500–3000	3000
C5	500–2800	+.5 a −1.5	1000–2600	100
	300–3000	+1 a −3	600–2600	300
			500–2800	600

Hay cinco clasificaciones en el acondicionamiento tipo C:

1. C1 y C2, acerca de circuitos de dos puntos y multipuntos.
2. C4, acerca de circuitos de dos puntos y arreglos de multipuntos con un máximo de cuatro lugares de estación.
3. Las especificaciones C5 sólo son para circuitos de dos puntos.
4. El acondicionamiento C3 es para líneas de acceso y circuitos troncales asociados con redes conmutadas privadas.

Las redes conmutadas privadas son sistemas telefónicos dedicados a un solo cliente, por lo general, con una gran cantidad de estaciones. Un ejemplo sería el de una gran empresa con oficinas y complejos en distintos lugares geográficos, con instalación local de centrales de abonado (PBX) en cada lugar. Una PBX es una matriz de conmutación de baja capacidad, donde los suscriptores se limitan, en general, a estaciones dentro del mismo complejo de construcciones. Para interconectar las PBX se requieren líneas de acceso de uso común y circuitos troncales. Sólo son comunes para los suscriptores de la red privada, y no para toda la red telefónica pública.

La tabla 13-14 muestra los diversos límites establecidos por las distintas clases de acondicionamiento tipo C, para la distorsión por atenuación y por retardo de envolvente. Las figs. 13-29 a 13-33 tienen una representación gráfica de algunos de los límites del parámetro ancho de banda.

Distorsión por atenuación y acondicionamiento tipo C. Los límites de distorsión por atenuación para un canal básico 3002 requieren que la ganancia del circuito a cualquier frecuencia entre 500 y 2500 Hz no sea mayor que 2 dB mayor que la ganancia del circuito a 1004 Hz, y no menos de 3 dB de la ganancia del circuito a 1004 Hz (véase la fig. 13-29). Para la distorsión por atenuación, siempre se usa la ganancia del circuito a 1004 Hz como referencia. También, dentro de las bandas de frecuencia de 300 a 499 Hz, y de 2501 a 3000 Hz, la ganancia del circuito no debe ser más de 3 dB mayor ni menos de 12 dB menor de la ganancia a 1004 Hz.

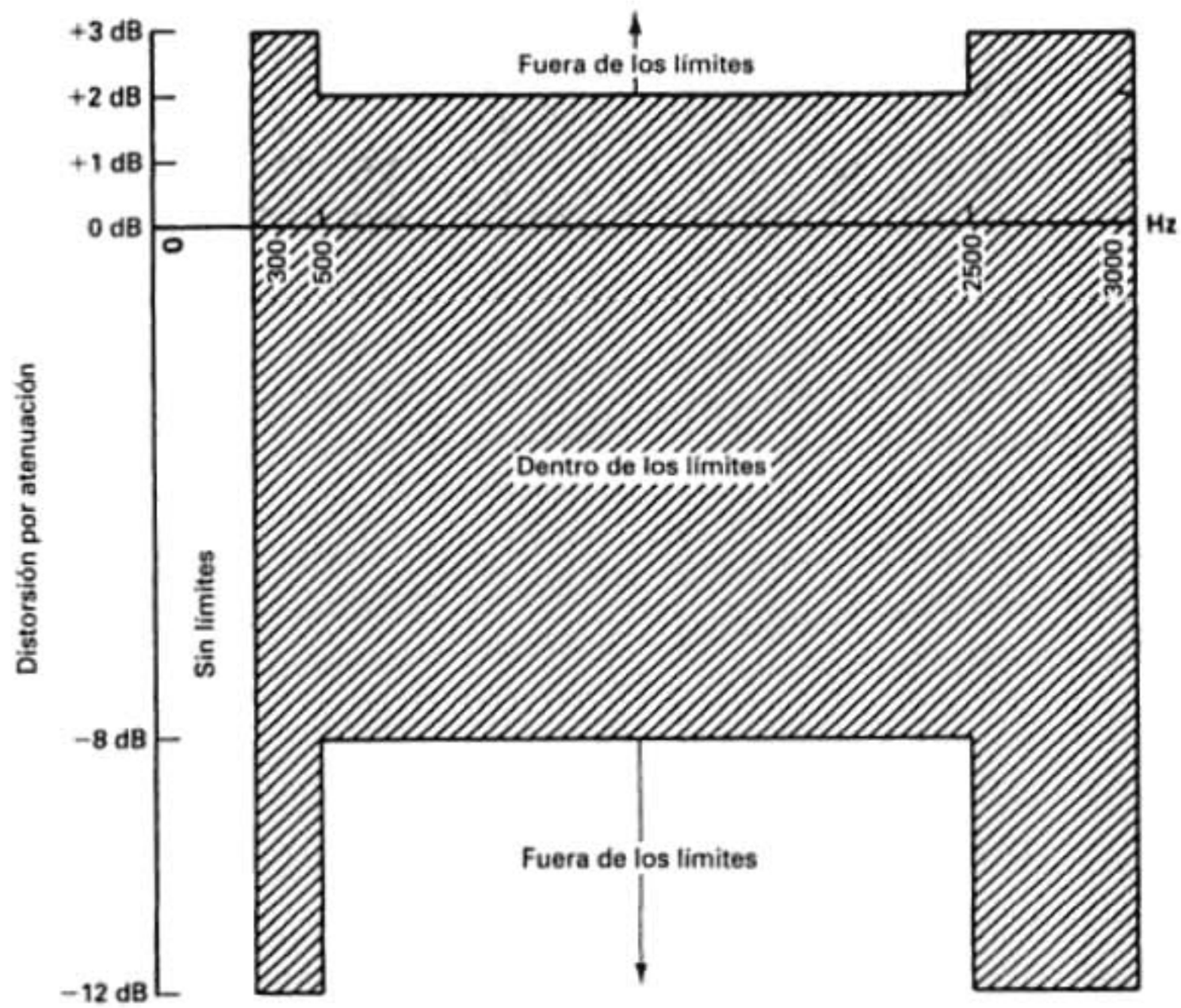


FIGURA 13-29 Presentación gráfica de los límites de distorsión por atenuación en un canal básico 3002

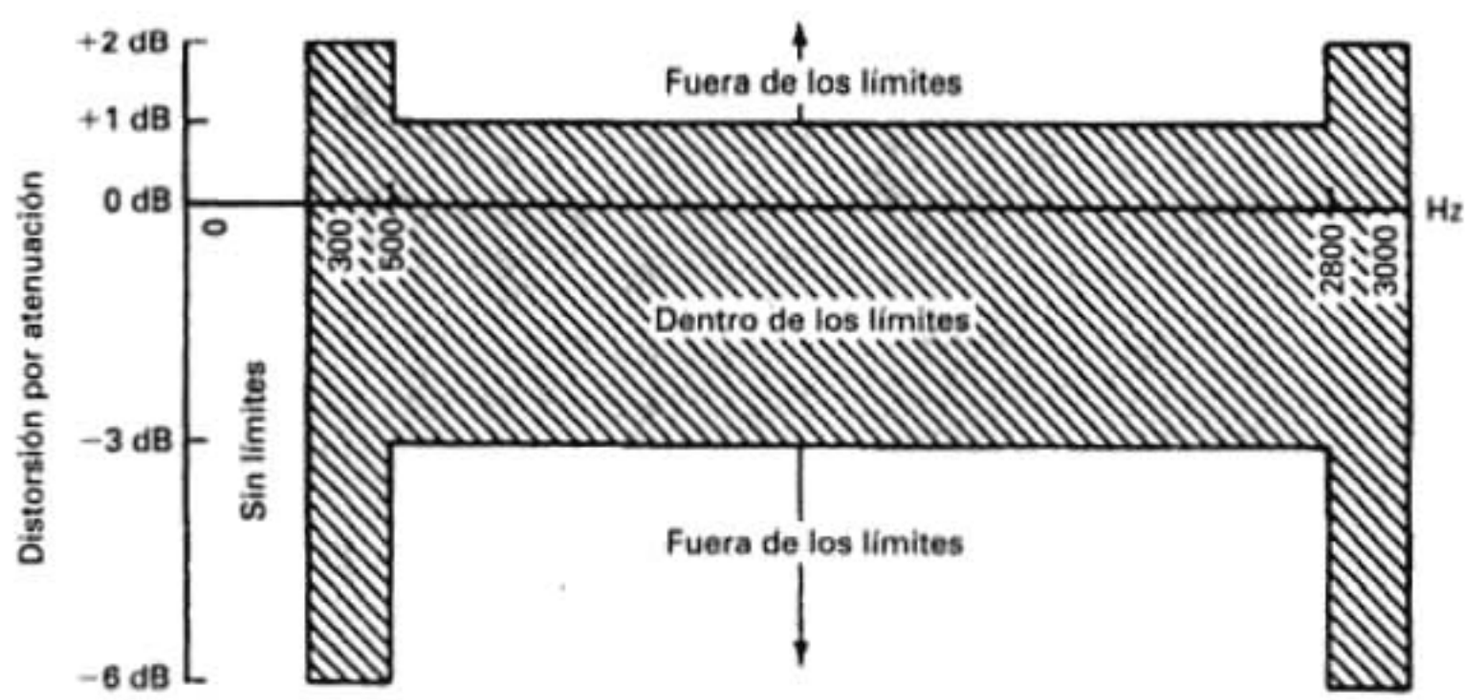


FIGURA 13-30 Representación gráfica de los límites de distorsión por atenuación en un canal con acondicionamiento C2

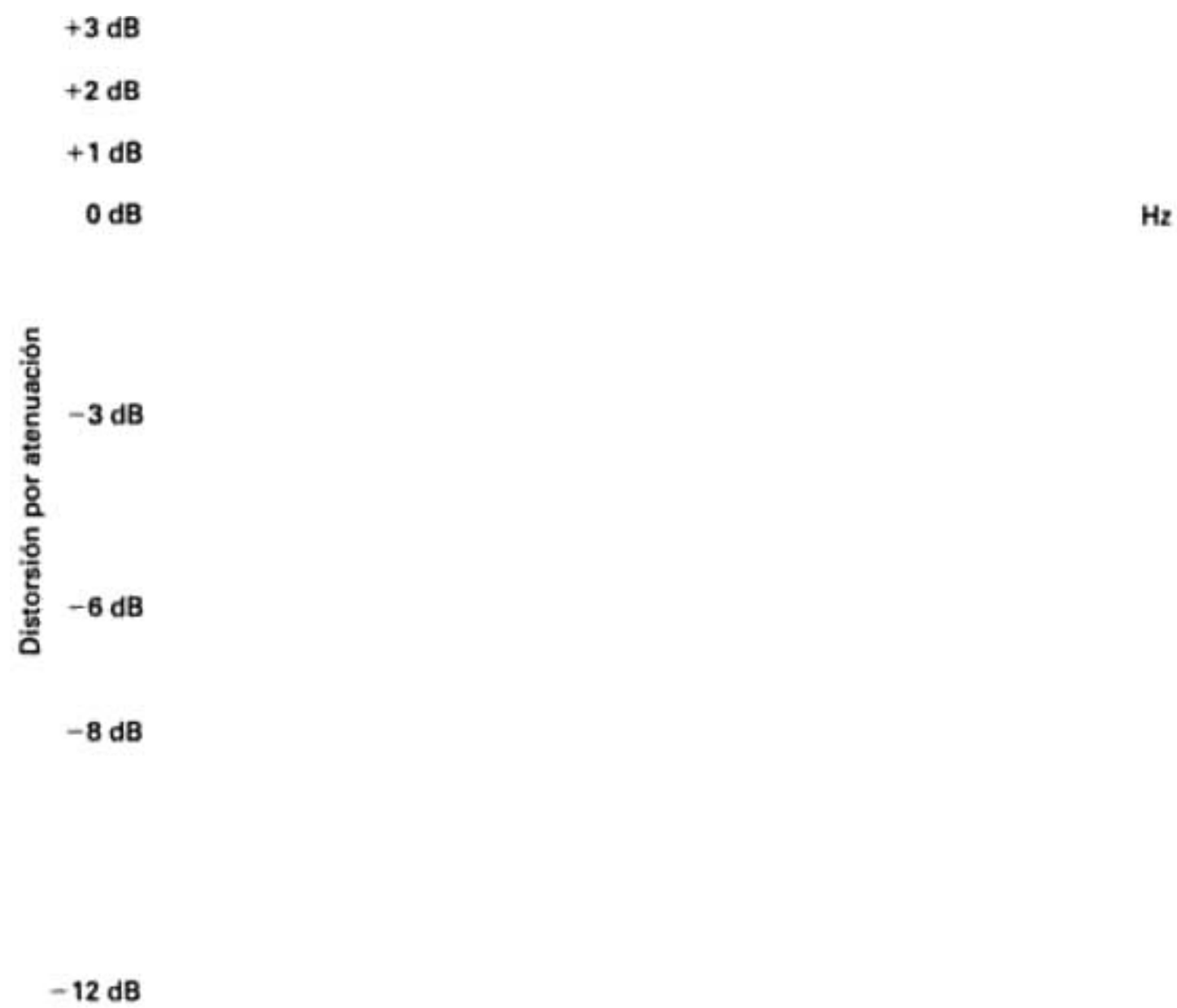


FIGURA 13-31 Superposición de las figuras 13-29 y 13-30 para demostrar los requisitos más estrictos impuestos por el acondicionamiento C2

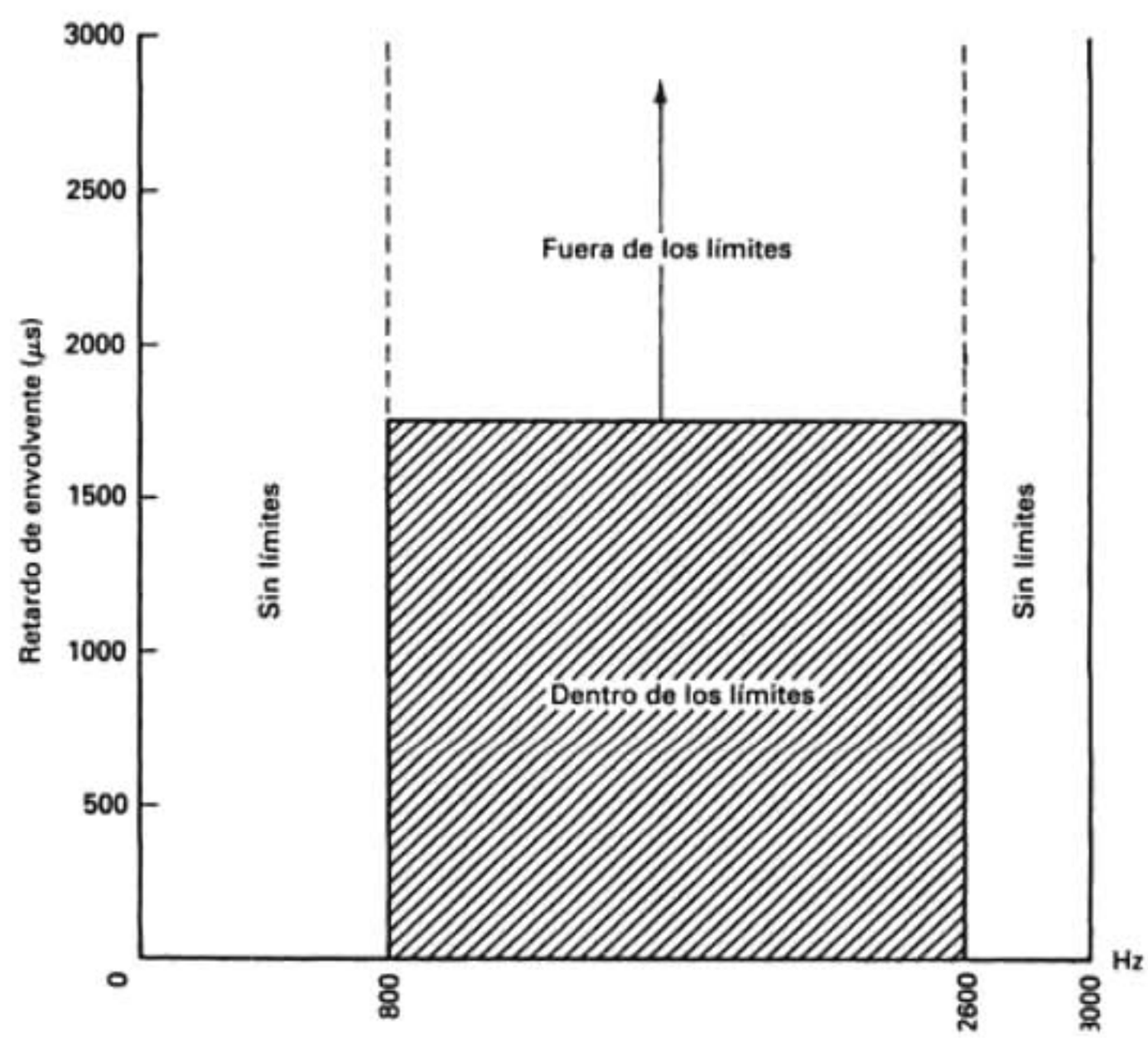


FIGURA 13-32 Presentación gráfica de los límites para el retardo de envolvente en un canal básico 3002

Por definición, el retardo de envolvente es la primera derivada de la fase con respecto a la frecuencia:

$$\text{retardo de envolvente} = \frac{d\phi(\omega)}{d\omega}$$

En realidad, el retardo de envolvente sólo se aproxima, eso sí bastante, a *df/dw*. Las mediciones de retardo de envolvente no evalúan las características reales de fase en función de frecuencia, sino más bien la fase de una onda, que es la resultante de una banda angosta de frecuencias. Es un error común confundir la distorsión real de fase (llamada también distorsión por retardo) con la distorsión por retardo de envolvente (EDD, por *envelope delay distortion*). El *retardo de envolvente* es el tiempo necesario para propagar un cambio en una envolvente de AM a través de un medio de transmisión. Para medir el retardo de envolvente se transmite una portadora de amplitud modulada de banda angosta, cuya frecuencia se varía dentro de la banda útil de voz. (La tasa de modulación AM normalmente es de 25 a 100 Hz.) En el receptor se miden las variaciones de fase de las envolventes de baja frecuencia. La diferencia de fases a las diversas frecuencias de portadora es la *distorsión por retardo de envolvente*. La frecuencia de portadora que produce el retardo mínimo de envolvente se establece como referencia y se normaliza a cero. Por consiguiente, las mediciones de EDD sólo dan valores positivos, e indican los retardos relativos de envolvente de las diversas frecuencias de portadora con respecto a la frecuencia de referencia. La frecuencia de referencia de un circuito característico de banda de voz es 1800 Hz, aproximadamente.

Las mediciones EDD no obtienen los retardos reales de fase, ni determinan las relaciones entre retardos reales de fase. Las mediciones EDD se usan para determinar una aproximación estrecha a las características relativas de retardo de fase de un circuito.

El límite EDD de un canal básico 3002 es de 1750 ms entre 800 y 2600 Hz, como se ve en la tabla 13-14. Esto quiere decir que la diferencia máxima de retardo de envolvente entre dos frecuencias cualquiera de portadora, dentro de esos márgenes, no puede ser mayor que 1750 μs.

Ejemplo 13-7

Una prueba EDD de un canal básico 3002 indicó que una portadora de 1800 Hz sufre el retardo mínimo absoluto de 400 μs. Por consiguiente, es la referencia. El retardo absoluto de envolvente sufrido por cualquier frecuencia dentro de la banda de 800 a 2600 Hz no puede ser mayor de 2150 μs (= 400 μs + 1750 μs).

El retardo absoluto de una señal entre dos puntos cualesquiera de Estados Unidos en el continente nunca debe ser mayor que 100 ms, lo cual no es suficiente para causar problemas. En consecuencia se miden valores relativos, y no absolutos, del retardo de la envolvente. Para el ejemplo anterior, siempre que las pruebas EDD den como resultado valores relativos menores que +1750 μs, el circuito está dentro de los límites aceptados.

Parámetros de interfaz. Las dos consideraciones principales de los parámetros de interfaz son:

1. Protección eléctrica de la red telefónica y de su personal
2. Normalización de los arreglos de diseño.

A continuación se resumirán esas consideraciones. Las impedancias del equipo de estación deben ser 600 Ω resistivos sobre la banda útil de voz, y el equipo de estación debe estar aislado de tierra por un mínimo de 20 MΩ en cd y 50 kΩ en ca. El circuito telefónico básico para grado de voz es un canal 3002; tiene un ancho de banda útil de 300 Hz a 3000 Hz. La ganancia de circuito a 3000 Hz es 3 dB menor que la potencia de señal en banda especificada. La ganancia a 4 kHz debe ser cuando menos 15 dB menor que la ganancia a 3 kHz. La potencia máxima de señal transmitida para un circuito de línea privada es 0 dBm. La potencia de señal transmitida para circuitos de marcar se establece para cada lazo, para que la señal se reciba en la oficina central de Telco a −12 dBm.

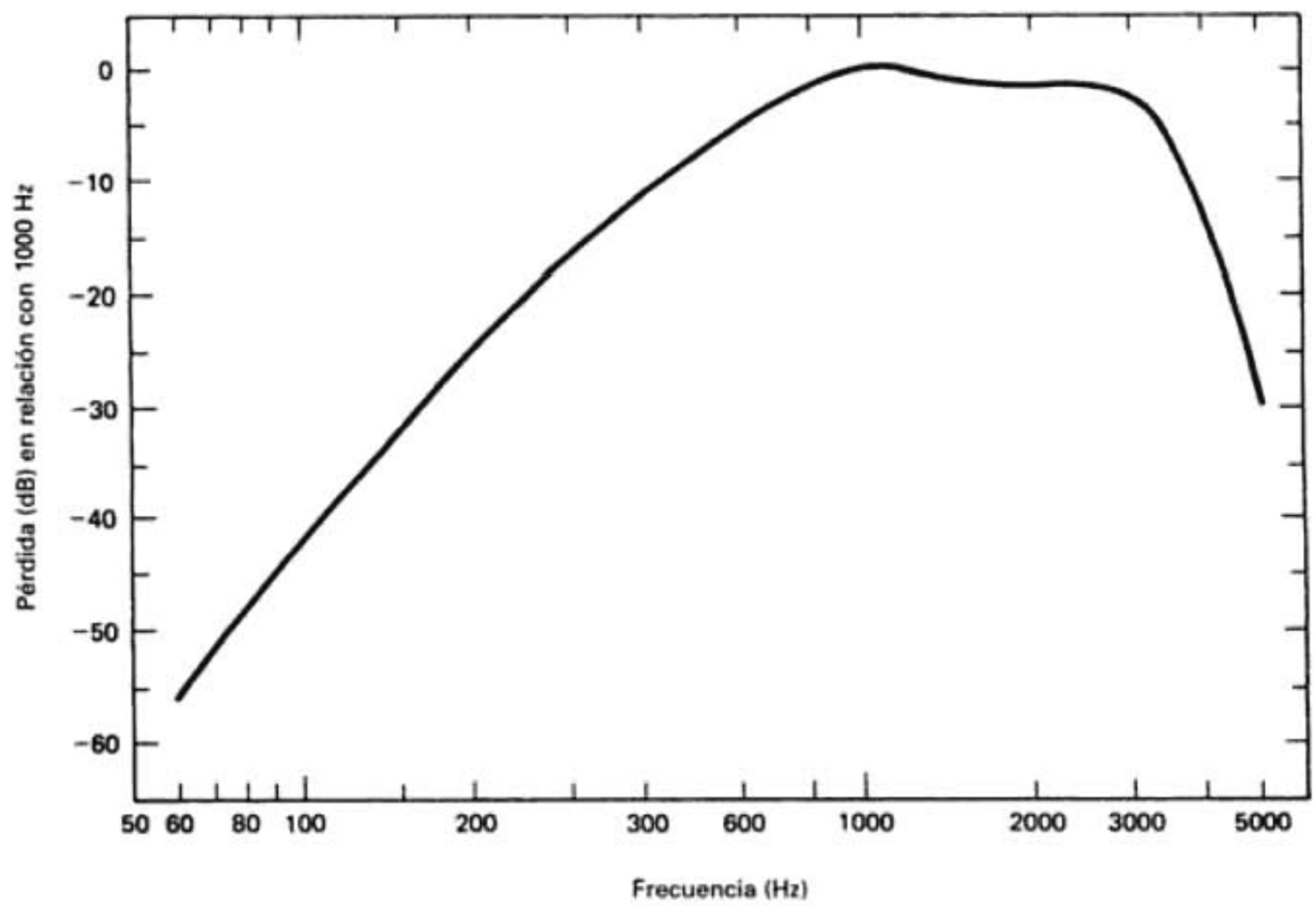


FIGURA 13-34 Curva de ponderación de mensaje-C

Parámetros de la instalación. Los parámetros de la instalación representan deterioros potenciales de la señal de datos. Estos deterioros los causa el equipo de Telco, y los límites especificados pertenecen a todos los circuitos de banda de voz, independientemente de su acondicionamiento.

Variación a 1004 Hz. Telco ha establecido que 1004 Hz es la frecuencia patrón para tono de prueba. Se seleccionaron 1004 Hz por su ubicación relativa en la pasabanda de un circuito normal de banda de voz. El objeto de este tono de prueba es simular la potencia de la señal combinada de una transmisión normal de banda de voz. La pérdida en canal de 1004 Hz, para un circuito de línea privada, es 16 dB. Un tono de prueba de 1004 Hz, aplicado al lazo local de transmitir y a 0 dBm, debe aparecer en la salida del lazo del destino a −16 dBm. Las variaciones a largo plazo de la ganancia de las instalaciones de transmisión no deben salirse de ±4 dB; la potencia de la señal recibida debe estar entre los límites de −12 a 200 dBm.

Ruido. El *ruido* se puede definir, en general, como toda energía indeseable presente en la pasabanda útil de un canal de comunicaciones. El ruido puede ser correlacionado o no correlacionado. *Correlación* es una relación entre la señal y el ruido. El ruido no correlacionado es la energía que hay, como por ejemplo el ruido térmico, en ausencia de una señal. El ruido correlacionado es energía no deseada que se presenta como resultado directo de la señal, como por ejemplo, la distorsión no lineal.

Ponderado del ruido. La interferencia de una señal por el ruido se clasifica en términos de molestia y de inteligibilidad. El ruido puede ser molesto al escucha, pero no al grado de no poder comprender la conversación. En Western Electric Company se hicieron experimentos en los que se pidió a varios grupos de escuchas calificar la molestia causada por 14 distintas frecuencias audibles, entre 180 y 3500 Hz. Estas frecuencias fueron presentadas a los escuchas a través de un teléfono normal, tipo 500 (los antiguos de discado, de los abuelos). Los escuchas primero compararon la molestia en cada frecuencia con la molestia de una frecuencia de referencia de 1000 Hz, sin potencia de voz. A continuación, se repitieron los mismos experimentos con conversación simultánea. Se promediaron y alisaron los resultados de las dos pruebas, y se obtuvo la *curva de ponderación de mensaje-C*. Esta curva se ve en la fig. 13-34.

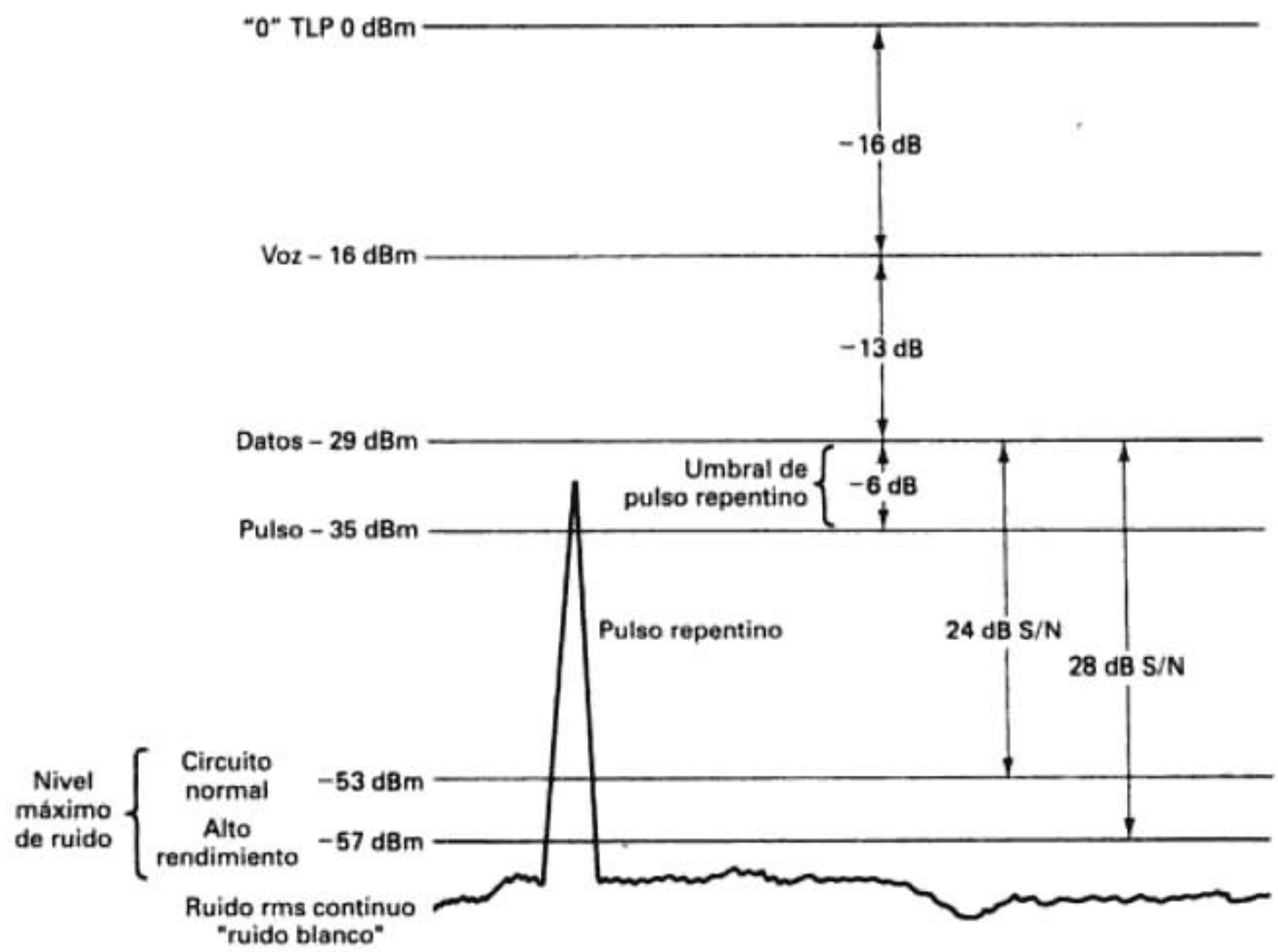


FIGURA 13-37 Ruido angosto C y de pulso

del ruido en esta forma. Las características generales del circuito, en ausencia de una señal, no son necesariamente las mismas que cuando está presente una señal. La diferencia se debe al uso de compresores-expansores y de dispositivos automáticos de ganancia. Los compresores-expansores son dispositivos que tienen características de ganancia que dependen de la frecuencia (se explicarán con detalle en el capítulo 15). Por esta razón se desarrollaron las mediciones de ruido angosto C.

Las mediciones de ruido angosto C difieren de las normales para mensaje-C sólo en que se aplica un tono continuo (por lo general de 1004 o de 2804 Hz) al extremo transmisor del circuito, durante la prueba. El tono continuo asegura que el funcionamiento del circuito simule una transmisión con carga de voz o de datos. En comunicaciones, "cargado" es un término que indica la presencia de una potencia de señal comparable a la de una transmisión real de mensaje. El tono continuo se filtra (se saca con muesca) en el medidor de ruido, antes del filtro de mensaje-C. Las características de pasabanda del filtro de banda muy angosta son tales que sólo se elimina el tono continuo. La potencia de ruido en la pasabanda útil se mide con un equipo normal de medición de ruido de mensaje-C. Esta prueba (véase la fig. 13-36) asegura que las indicaciones de ruido que se obtuvieron reflejen en realidad las características de circuito con carga en una transmisión normal en banda de voz.

La construcción física de un circuito de línea privada puede requerir el uso de varios circuitos troncales en tándem. Cada troncal individual puede ser una instalación analógica, digital, con o sin compresión-expansión. Se pueden usar diversas combinaciones de estas instalaciones para configurar un circuito. Telco ha establecido requisitos de ruido angosto C para cada clase de instalación y para diversas longitudes de troncal. Estos requisitos ayudan a que Telco evalúe el funcionamiento de una instalación, y facilitan los procedimientos de aislamiento de falla. Un suscriptor de la red telefónica sólo se debe ocupar del requisito general (de extremo a extremo) del ruido angosto C. Los circuitos normales de datos en línea privada, que funcionan a menos de 9600 bps, requieren una relación de señal a ruido angosto C de 24 dB. Los circuitos

transitorios debidos a conmutación de instalaciones de radio, durante el curso normal de un día. Los desvanecimientos atmosféricos causan la necesidad de conmutar las instalaciones de radio. Una *caída de ganancia* es una disminución de la ganancia de un circuito en más de 12 dB, que dure más de 4 ms. Las caídas repentinas de ganancia son características de una condición temporal de circuito abierto, causada por grandes desvanecimientos de radio, o de actividades de mantenimiento en Telco. Las caídas de ganancia se presentan con una frecuencia aproximada de una por hora. Las ganancias repentinas y las caídas de ganancia (véase la fig. 13-38) se detectan vigilando el nivel de recepción de un tono de prueba de 1004 Hz.

Perturbaciones de fase. Las *perturbaciones de fase* (véase la fig. 13-39) son cambios repentinos y aleatorios en la fase de una señal transmitida. Pueden ser variaciones temporales en la fase de una señal, que duren más de 4 ms. Por lo general, no se registran las perturbaciones de fase, a menos que rebasen un máximo de $\pm 20°$. Las perturbaciones de fase, al igual que las ganancias repentinas, se deben a estados transitorios que se producen cuando se conmutan las instalaciones de radio.

Inestabilidad de fase. La *inestabilidad de fase* (véase la fig. 13-39) es una forma de modulación incidental de fase: una variación continua y no controlada en los cruces de una señal con cero. En general, se presenta inestabilidad de fase a frecuencias menores que 300 Hz y su causa principal es una ondulación residual de baja frecuencia, de ca, en las plantas de suministro de corriente de Telco. La cantidad de fuentes de corriente que requiere un circuito es directamente proporcional a la cantidad de circuitos troncales y de oficinas de Telco que forman un canal de mensaje. Cada instalación troncal tiene una especificación separada de inestabilidad de fase. La inestabilidad máxima de fase, de extremo a extremo, que se permite, es de 10° de máximo a máximo, independientemente de la cantidad de enlaces de radio, instalaciones de cable o tramos de portadora digital que se usen en el circuito. La limitación de la cantidad de circuitos troncales es una consideración primaria del diseño de un circuito de datos. La inestabilidad de fase se mide observando los cruces con cero de un tono de prueba de 1004 Hz.

Interferencia por una sola frecuencia. Esta interferencia es la presencia de uno o más tonos continuos, no deseados, en un canal de mensaje. Los tonos se llaman *tonos espurios*, y con frecuencia se deben a diafonía o modulación cruzada entre canales distintos de un sistema de portadora. Los tonos espurios se miden terminando el lado de transmisión de un circuito y a continuación buscando en el espectro del canal con un voltímetro de frecuencia selectiva, u observando la pasabanda del canal con un analizador de espectro. Los tonos espurios pueden producir el mismo comportamiento perjudicial en el circuito que el ruido blanco.

Desplazamientos de frecuencia. Los sistemas analógicos de portadora, que usan las empresas telefónicas, funcionan con banda lateral única y portadora suprimida (SSBSC) y, en consecuencia, requieren demodulación coherente. En la *demodulación coherente*, se debe recuperar la frecuencia de la portadora suprimida, y se debe reproducir con exactitud en el receptor. Si no se hace lo anterior, la señal demodulada se desplazará en frecuencia, en una cantidad igual a la diferencia entre las frecuencias de transmisión y la portadora de recepción. El desplazamiento de frecuencia se mide transmitiendo un tono de prueba de 1004 Hz, para a continuación medir la frecuencia del tono en el receptor.

Distorsión de intercepción de fase. Esta distorsión se presenta en sistemas SSBSC coherentes, cuando la portadora recibida no se reinserta con la relación exacta de pulsos a la señal recibida, que la que tenía la portadora de transmisión. Este degradamiento causa un desplazamiento constante de fase hacia todas las frecuencias. Tiene poca importancia con los módems de datos que usan manipulación por desplazamiento de frecuencia, por desplazamiento diferencial de fase, o modulación de amplitud por cuadratura. Como son los métodos más comunes de modulación, la empresa telefónica no ha establecido límites para la distorsión por intercepción de fase.

Distorsión no lineal. La *distorsión no lineal* es un ejemplo de ruido correlacionado. El ruido debido a la distorsión no lineal tiene la forma de presencia de tonos adicionales, causados por la amplificación no lineal de una señal; ¡si no hay señal, no hay ruido! La distorsión no lineal produce ondas senoidales distorsionadas. Hay dos clases de distorsión no lineal:

1. *Distorsión armónica*–múltiplos indeseados de las frecuencias transmitidas.
2. *Distorsión por intermodulación*–productos cruzados (sumas y diferencias) de las frecuencias de entrada.

TABLA 13-16 Resumen de los límites de parámetros de instalación

Parámetro	Límite	
1. Variación de pérdida en 1004 Hz	No más de ± 4 dB a largo plazo	
2. Ruido en mensaje-C	rms máximo ruido en el módem del receptor (nominal: punto de −16 dBm)	
Millas de instalación	*dBm*	*dBrncO*
0–50	−61	32
51–100	−59	34
101–400	−58	35
401–1000	−55	38
1001–1500	−54	39
1501–2500	−52	41
2501–4000	−50	43
4001–8000	−47	46
8001–16,000	−44	49
3. Ruido angosto-C	(valores mínimos)	
(a) Canal normal de banda de voz	24 dB de señal a ruido angosto-C	
(b) Línea de alto rendimiento	28 dB de señal a ruido angosto-C	
4. Interferencia en una frecuencia	Al menos 3 dB menos que los límites de ruido en mensaje-C	
5. Ruido de pulso		
Umbral con respecto al tono continuo de 1004 Hz	*Cuentas máximas arriba del umbral, permitidas en 15 minutos*	
0 dB	15	
+4 dB	9	
+8 dB	5	
6. Desplazamiento de frecuencia	±5 Hz de extremo a extremo	
7. Distorsión por intercepción de fase	Sin límites	
8. Inestabilidad de fase	No más de 10° de máximo a máximo (requisito de extremo a extremo)	
9. Distorsión no lineal (sólo circuitos acondicionados D)		
Señal a segundo orden	Cuando menos 35 dB	
Señal a tercer orden	Cuando menos 40 dB	
10. Relación de máximo a promedio	Indicación mínima de 50 de extremo a extremo con medidor PAR estándar	
11. Perturbaciones de fase	8 o menos en cualquier periodo de 15 min., mayores que un pico de ±20°	
12. Ganancias repentinas	8 o menos en cualquier periodo de 15 min., mayores que ±3 dB	
13. Caídas de ganancia	2 o menos en cualquier periodo de 15 minutos, mayores que 12 dB	

3. *TLP (transmission level point*, o *punto de nivel de transmisión*). Casi no tiene utilidad conocer la potencia de la señal en algún punto. Una señal en determinado punto puede tener 10 dBm. ¿Es buena o mala? Se puede contestar a eso si se conoce cuál debería ser la intensidad de la señal en ese punto. Es lo que hace el TLP. La referencia del TLP es 0 dBm. Un TLP de −15 dBm indica que, en este punto especificado, la señal debería ser −15 dBm. Un TLP de 0 es aquel en el que la potencia de la señal debería ser 0 dBm. El TLP no dice nada acerca de la señal misma.
4. *dBmO*. Es una medida de potencia, ajustada a 0 dBm, que indica cuál sería la potencia si se midiera a TLP 0. Este valor compara la señal real en un punto con lo que debería ser esa señal en ese punto (TLP). Una señal de +4 dBm medida a un TLP de −16 dBm tiene +20 dBmO. Es 20 dB más fuerte que la de referencia. Una señal de −2 dBm medida a un TLP de −10 dBm es igual a +8 dBmO.
5. *rn* (*reference noise*, o *ruido de referencia*). Este valor es el de los dB que se usan como referencia de ruido. *Siempre* es −90 dBm, o 1 pW. Este valor se toma como referencia, porque el valor de la señal en cualquier punto nunca debe ser menor que él.
6. *dBrnc*. Es el valor de dB del ruido con respecto al ruido de referencia con ponderación de mensaje-C. Mientras mayor sea este valor, la condición empeora porque hay mayor cantidad de ruido.

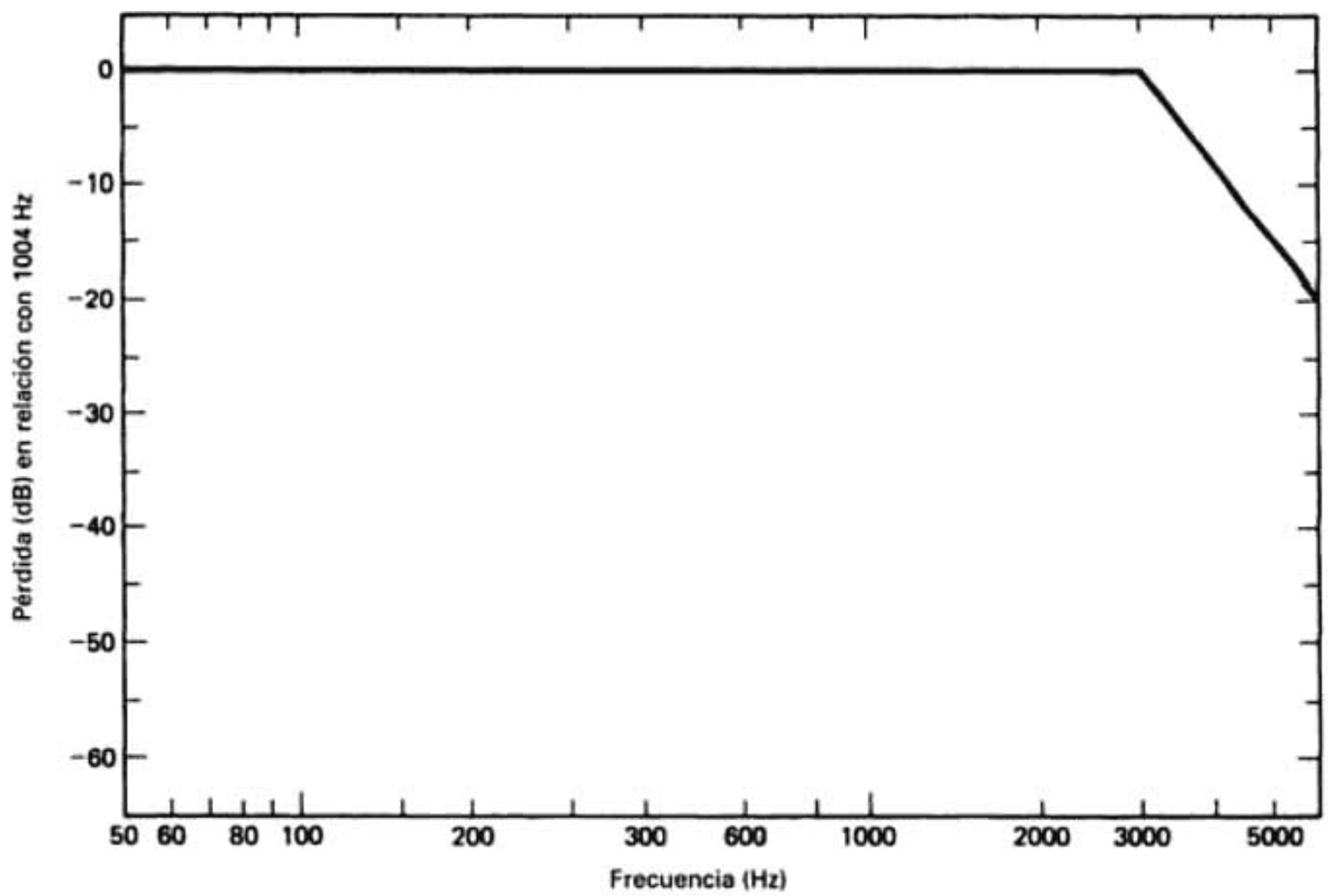


FIGURA 13-42 Filtro plano de 3 kHz

En esencia, un módem no es más que una repetidora que convierte señales eléctricas de forma digital a forma analógica, y viceversa. Físicamente se localiza entre una parte de equipo de cómputo y un canal de comunicaciones analógicas. En el transmisor del sistema de comunicaciones de datos un módem acepta pulsos digitales discretos (normalmente en forma binaria) de una interfaz digital serie, como la RS-232, y los convierte a señales analógicas de variación continua. A continuación, las señales analógicas salen a un canal de comunicaciones analógicas, donde se transfieren por el sistema hasta un destino lejano. En el destino, o extremo de recepción del sistema de comunicación de datos, un módem acepta las señales analógicas, del canal de comunicaciones, y las convierte en pulsos digitales. Después, los pulsos digitales salen hacia una interfaz digital serie.

En general, los módems se clasifican como asíncronos o síncronos, y usan una de las siguientes técnicas de modulación: manipulación por desplazamiento de amplitud (ASK), manipulación por desplazamiento de frecuencia (FSK), manipulación por desplazamiento de fase (PSK) o modulación de amplitud en cuadratura (QAM). En los módems síncronos se incorpora la información del reloj en el receptor, pero en los asíncronos no. Los módems asíncronos usan, en general, ASK o FSK, y se restringen a aplicaciones de velocidad relativamente baja (casi siempre menos de 2.4 kbps. Los módems síncronos usan PSK y QAM, y son para aplicaciones de velocidad intermedia y alta (hasta 57.6 kbps).

Módems asíncronos

Durante algún tiempo Bell Systems dominó el mercado de los módems. Debido a ello, los parámetros de funcionamiento de los diversos módems Bell son los patrones desde donde evolucionaron las normas internacionales modernas de la CCITT. Hoy, la CCITT se llama International Standards Union-Telecommunication Standardization Sector (ITU-TS).

Los módems asíncronos se usan principalmente con dircuitos de datos de baja velocidad. Hay varios diseños normales modernos, que se usan para transmitir datos asíncronos a baja velocidad. Para la operación semidúplex con la red telefónica pública de dos hilos, o la operación dúplex con cuatro hilos y circuitos dedicados de línea privada, la norma más difundida es el

TABLA 13-17 Resumen del módem Bell System

Designación Bell System	Instalación de línea	Modo de operación	Hilos	Sincronización	Modulación	Frecuencia de bits (bps)
103	Marcar	FDM/FDX	2	Asíncrono	FSK	300
113A/B	Marcar	FDM/FDX	2	Asíncrono	FSK	300
201B	Marcar	HDX	2	Síncrono	QPSK	2400
201C	Rentada	FDX	4	Síncrono	QPSK	2400
202S	Marcar	HDX	2	Asíncrono	FSK	1200
202T	Rentada	FDX	4	Asíncrono	FSK	1800
208A	Rentada	FDX	4	Síncrono	8-PSK	4800
208B	Marcar	HDX	2	Síncrono	8-PSK	4800
209A	Rentada	FDX	4	Síncrono	16-QAM	9600
209B	Marcar	HDX	2	Síncrono	16-QAM	9600
212A	Marcar	HDX	2	Asíncrono	FSK	600
212B	Rentada	FDX	4	Síncrono	QPSK	1200

Marcar—red telefónica pública
Rentado—circuito de línea privada
FDM—multiplexado por división de frecuencia
FDX—dúplex
HDX—semidúplex
FSK—manipulación por desplazamiento de frecuencia
QPSK—manipulación por desplazamiento de fase en cuadratura
8-PSK—manipulación por desplazamiento de 8 fases
16-QAM—modulación por amplitud en cuadratura 16

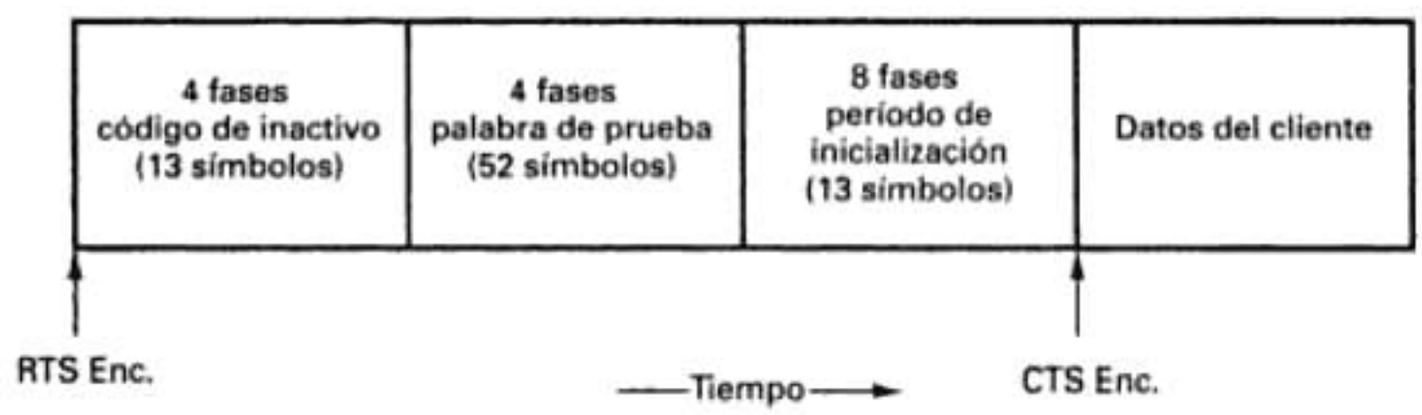


FIGURA 13-44 Secuencia de adiestramiento para un módem 208

ganancia de los amplificadores de CAG y desactivar todas las supresiones de eco en aplicaciones de marcado.

Módems de velocidad media y alta

Los módems de velocidad media y alta se usan cuando se requieren velocidades de transmisión de 2400 bps o mayores. Para transmitir con estas velocidades se usa modulación PSK o QAM, que requiere que los osciladores de portadora del receptor al menos tengan frecuencia coherente (y posiblemente, fase coherente). Como esos módems son síncronos, se debe efectuar la recuperación por el módem receptor. Estos módems contienen *circuitos desorganizadores y organizadores*, e *igualadores adaptativos* o *automáticos*.

Adiestramiento. El tipo de modulación y la técnica de codificación que se usen determina la cantidad necesaria de bits, y por consiguiente, la duración de la secuencia de adiestramiento. El módem 208 es síncrono, de 4800 bps que usa 8-DPSK. Su secuencia de adiestramiento se ve en la fig. 13-44. Cada símbolo representa tres bits (un tribit), y dura 0.625 ms. El código de inactivo, de cuatro fases, pasa por cuatro de los ocho desplazamientos posibles de fase. Eso permite que el receptor recupere con rapidez la información de la portadora y de la sincronización del reloj. La palabra de prueba de cuatro fases permite llegar a su ajuste final al igualador

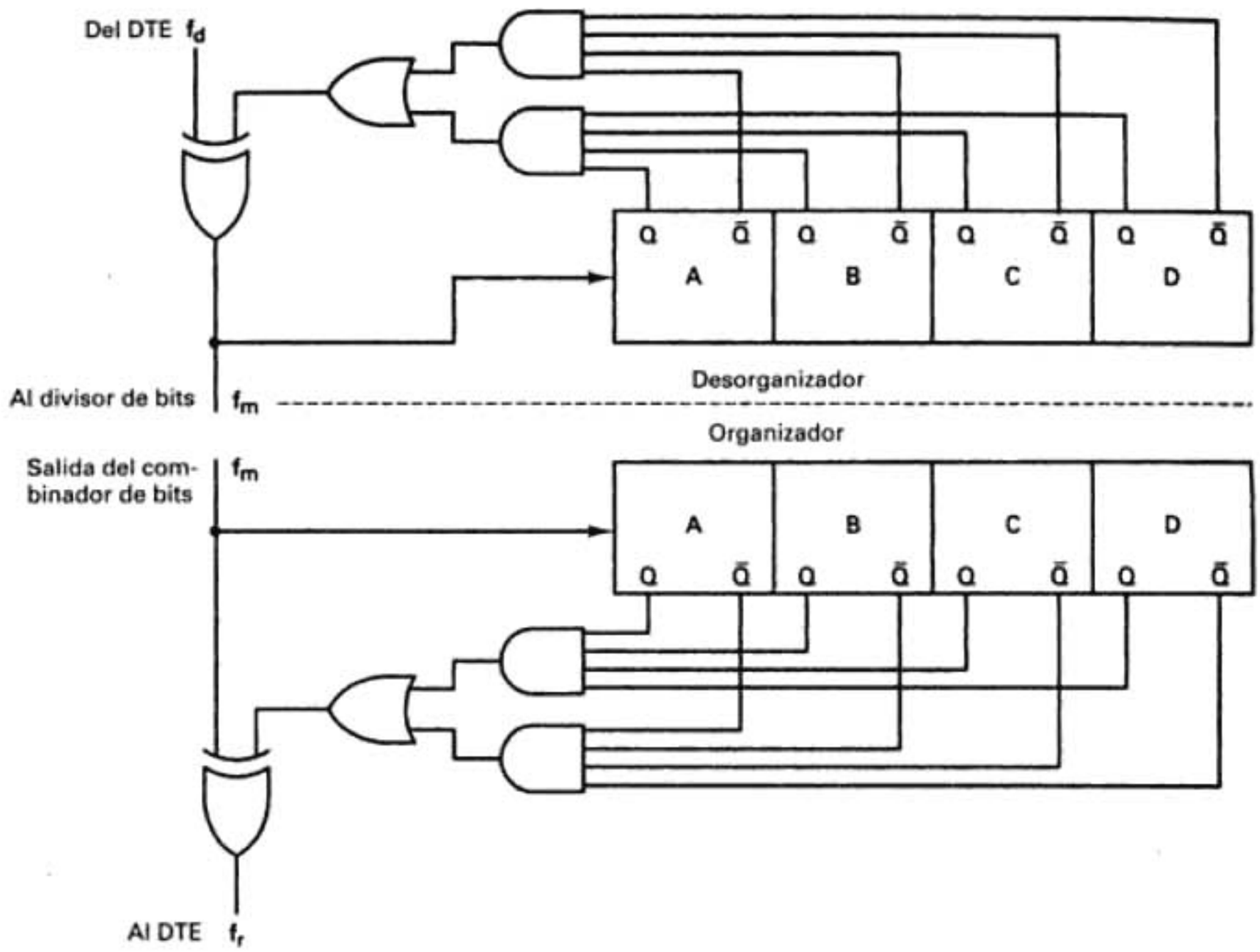


FIGURA 13-47 Circuito desorganizador y circuito organizador

de mandar los datos al DTE. El objetivo de un desorganizador no sólo es aleatorizar la secuencia transmitida de bits, sino detectar la ocurrencia de una secuencia no deseada de bits y convertirla en una más aceptable.

En la fig. 13-47 se muestra un diagrama de bloques de un circuito desorganizador y uno organizador. Estos circuitos están incompletos, porque se necesitaría una compuerta más para detectar una secuencia variable que diera origen a una de todos unos o todos ceros, en un canal modulador, después de dividir los bits.

El circuito desorganizador se inserta antes del divisor de bits en el modulador QPSK, y el organizador se inserta después del combinador de bits en el demodulador QPSK. En general, la salida de la compuerta OR del desorganizador o del organizador es A B C D + A′ B′ C′ D′.

$$f_m = f_d \oplus (A\,B\,C\,D + A'\,B'\,C'\,D') \qquad \text{compuerta XOR superior}$$
$$f_r = f_m \oplus (A\,B\,C\,D + A'\,B'\,C'\,D') \qquad \text{compuerta XOR inferior}$$

Al sustituir f_m en la segunda ecuación resulta

$$f_r = f_d \oplus (A\,B\,C\,D + A'\,B'\,C'\,D') \oplus (A\,B\,C\,D + A'\,B'\,C'\,D')$$

Como una identidad combinada XOR consigo misma es cero,

$$f_r = f_d \oplus 0$$
$$f_r = f_d$$

Esto sólo muestra que los datos originales transmitidos (f_d) se recuperan por completo en el receptor.

La salida de cada una de las compuertas OR será un 1 si el registro de cuatro bits contiene sólo unos o sólo ceros. Ninguna de las dos es una secuencia deseable. Si la salida de la compuerta OR es un 1, f_m será el complemento (el opuesto) de f_d, o f_r será el complemento de f_m. Se pretende crear transiciones en una corriente prolongada de datos que sea todos unos o todos

TABLA 13-19 (*Continuación*)

Designación CCITT	Especificación
V.52	Aparatos para medir distorsión y tasas de error en transmisión de datos
V.53	Límites de deterioro para los circuitos de datos
V.54	Dispositivos de prueba de lazo para módems
V.55	Equipo de medición de pulsos de ruido
V.56	Prueba comparativa de módems
V.57	Conjunto detallado de pruebas para transmisión de datos de alta velocidad
V.100	Interconexión entre redes públicas de datos y redes telefónicas públicas conmutadas
V.110	Adaptación ISDN de terminal
V.120	Adaptación ISDN de terminal, con multiplexado estadístico
V.230	Interfaz general de comunicación de datos, capa 1 de ISO

de 1700 Hz. Los datos se sincronizan en el módem en grupos de cuatro bits, llamados cuadribits, y resulta una rapidez de transmisión de 2400 baudios. A veces, se usan módems compatibles con V.29 en el modo semidúplex con líneas telefónicas conmutadas de dos conductores. Se puede tener un funcionamiento pseudodúplex con líneas de dos alambres con un método llamado *ping-pong*. En este método, los datos que mandan los DTE en cualquiera de los extremos del circuito, al módem, se guardan y se intercambian en forma automática en el enlace de datos, encendiendo y apagando rápidamente a las portadoras en sucesión.

La operación pseudodúplex en una línea de dos alambres también se puede hacer mediante *duplexado estadístico*. En este método se usa un canal inverso de datos de 300 bps. El canal inverso permite que un operador de datos capture datos en el teclado y al mismo tiempo que se reciba un archivo del módem distante. La dirección de transmisión de datos se puede determinar vigilando las memorias de datos internas del módem, y se pueden invertir los canales de alta y baja velocidad.

Recomendación CCITT V.32 para módems

La especificación CCITT V.32 se aplica a la velocidad de transmisión de datos a 9600 bps con operación dúplex sobre líneas privadas alquiladas de cuatro alambres, o a líneas telefónicas conmutadas de dos alambres. También se aplica a velocidades de datos de 2400 y 4800 bps. La V.32 especifica modulación QAM con frecuencia de portadora de 1800 Hz. Esta norma se parece a la V.29, pero en ella se especifica la técnica avanzada de codificación llamada *codificación trellis* (véase el capítulo 12). La codificación trellis produce una mejor relación de señal a ruido, porque divide la corriente entrante de datos en grupos de cinco bits, llamados *quintibits* (M-arios, donde $M = 2^5 = 32$). El diagrama de constelación para la codificación trellis de 32 estados fue desarrollado por el Dr. Ungerboeck del Laboratorio de Investigación de IBM en Zürich, y en él se combina la codificación y la modulación para mejorar la eficiencia de errores de bits. El concepto básico de la codificación trellis es introducir una redundancia controlada, que reduzca las frecuencias de error en el canal, duplicando la cantidad de puntos de señal en la constelación QAM. La constelación de codificación trellis que se usa con la norma V.32 se ve en la fig. 13-48.

En la V.32 se hace la operación dúplex sobre líneas telefónicas conmutadas de dos alambres usando una técnica llamada *cancelación de eco*. Implica agregar una réplica invertida de la señal transmitida a la señal recibida. Esto permite que los datos transmitidos de cada módem usen en forma simultánea la misma frecuencia de portadora, el mismo esquema de modulación y el mismo ancho de banda.

Recomendaciones CCITT V.32bis y V.32terbo para módems

La recomendación V.32bis de CCITT fue introducida en 1991, y estableció un nuevo patrón en la industria de los módems, permitiendo frecuencias de bits de 14.4 kbps por los canales telefónicos normales de banda de voz. En la V.32bis se usa una constelación de señal de 64 puntos, donde cada condición de señalización representa seis bits de datos. En la fig. 13-49 se ve el diagrama de constelación para la V.32. La frecuencia de transmisión de bits en la V.32 es 6 bits/código $\times$ 24 códigos/segundo = 14,400 bps. Los baudios de señalización son 2400.

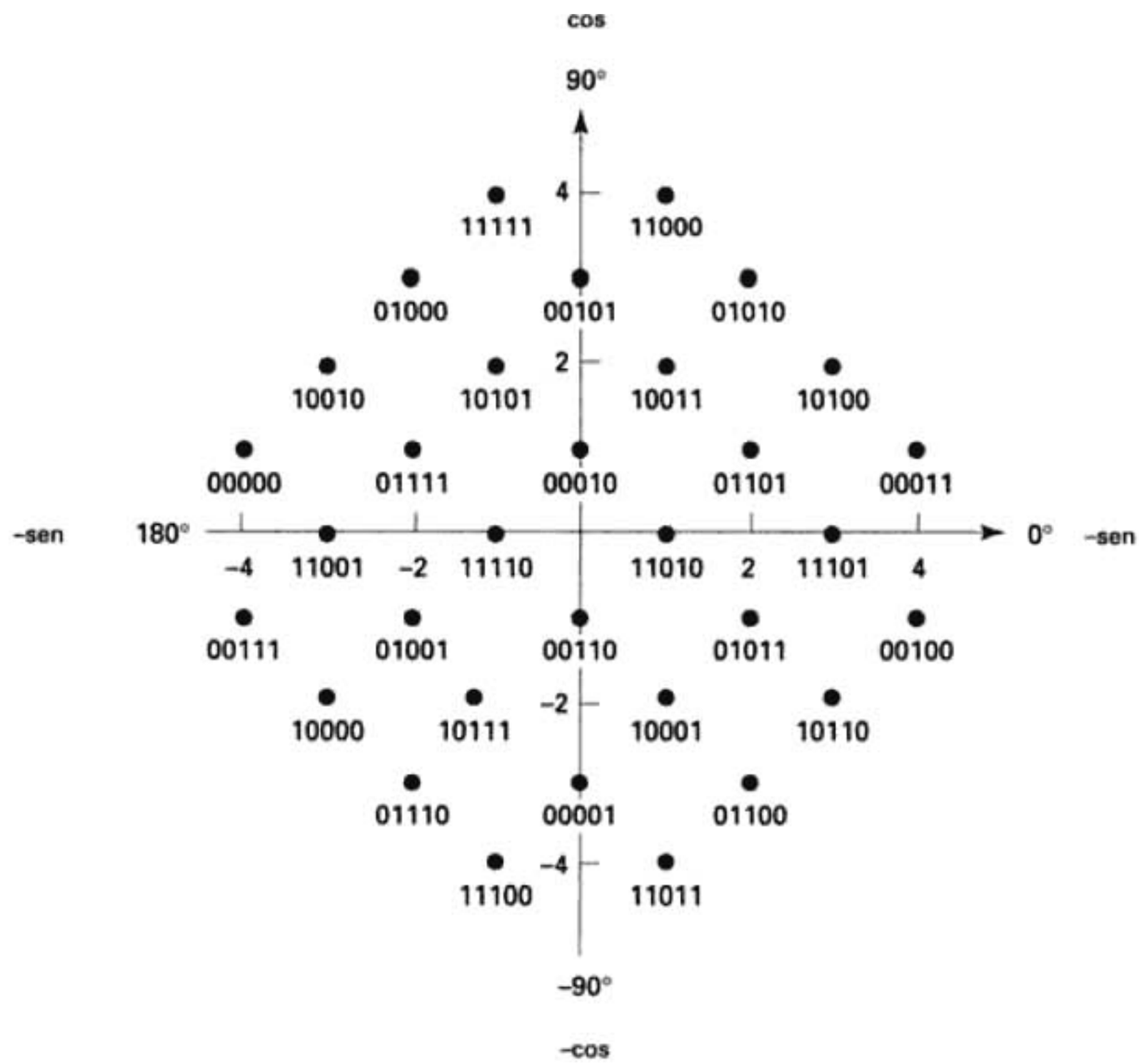


FIGURA 13-48 Diagrama de constelación V.32 con uso de codificación trellis

También, la V.32bis incluye funciones de *caída hacia adelante* y *caída hacia atrás*, que permiten cambiar al módem su transmisión, para adaptarse a cambios en la calidad de la línea de comunicaciones. La función de caída hacia atrás reduce lentamente la frecuencia de transmisión de bits a 12.2 kbps, 9.6 kbps o 4.8 kbps, si se degrada la calidad de la línea de comunicaciones. La función de caída hacia adelante da la opción, al módem, de regresar a una mayor frecuencia de transmisión cuando mejora la calidad del canal de comunicaciones. Con la V.32bis funciona el fax del grupo III, que es la norma de transmisión que describe los procedimientos de conexión entre dos máquinas o módems de fax. También, la V.32bis especifica el procedimiento de compresión de datos que se usa durante las transmisiones.

En agosto de 1993, U.S. Robotics introdujo la V.32terbo, que incluye todas las funciones de la V.32bis más una tecnología exclusiva llamada *nivelación adaptativa de velocidad*. La V.32terbo incluye dos categorías de funciones nuevas: mayores velocidades de datos y más amplia capacidad de fax. También describe la nueva rapidez de transmisión de datos, de 19.2 kbps, desarrollada por AT&T.

Recomendación CCITT V.33 para módems

La especificación V.33 de CCITT es para módems que funcionan en circuitos dedicados, de dos puntos, de línea privada y con cuatro conductores. En la V.33 se usa codificación trellis y se parece a la V.32, pero se incluye un elemento de señalización incluye seis bits de información y uno resultante, que produce una rapidez de transmisión de datos de 14.4 kbps, 2400 baudios y portadora de 1800 Hz. En la fig. 13.49 se ve el diagrama de constelación, con 128 puntos, que se usa con la V.33.

Recomendación CCITT V.42 y V.42bis para módems

En 1988, la ITU adoptó la norma V.42, de *procedimientos de corrección de errores para DCE* (para los módems). Las especificaciones V.42 se refieren a conversiones de asíncrono a síncrono en

Entre las innovaciones de la V.34 están:

1. Codificación no lineal, que compensa los efectos adversos de las no linealidades del sistema que producen ruido por distorsión armónica y por intermodulación, y ruido proporcional a la amplitud.
2. Codificación multidimensional y conformación de constelación, con mayor inmunidad de datos al ruido del canal.
3. Menor complejidad en los decodificadores de los receptores.
4. Precodificación de datos para tener mayor uso del ancho de banda disponible en el canal de comunicaciones, mejorando la transmisión de los datos en los límites externos del canal, donde la amplitud, frecuencia y distorsión de fase son las peores.
5. Investigación de línea, que es una técnica que usan los módems receptores para determinar con rapidez la mejor corrección compensadora para los deterioros de la línea de transmisión.

Recomendación CCITT V.34+ para módems

La V.34+ es una norma ampliada, adoptada por la ITU en 1996. Agrega las frecuencias de 31.2 y 33.6 kbps a la especificación V.34. Teóricamente, con la V.34+ se aumenta 17% la velocidad de transmisión; sin embargo, todavía no tiene la importancia suficiente para merecer una consideración seria, por el momento.

PREGUNTAS

13-1. Defina qué es *comunicación de datos*.

13-2. ¿Cuál es la importancia de la decisión Carterfone?

13-3. Explique la diferencia entre un circuito de dos puntos y uno multipunto.

13-4. ¿Qué es una topología de comunicaciones de datos?

13-5. Defina los cuatro modos de transmisión para circuitos de comunicación de datos.

13-6. ¿Cuál de los cuatro modos de transmisión sólo se puede usar en los circuitos multipunto?

13-7. Explique las diferencias entre los circuitos de dos y de cuatro alambres.

13-8. ¿Qué es un código de comunicación de datos? ¿Cuáles son algunos de los nombres de esos códigos?

13-9. ¿Cuáles son los tres tipos de caracteres que usan los códigos de comunicación de datos?

13-10. ¿Cuál código de comunicación de datos es el más potente? ¿Por qué?

13-11. ¿Cuáles son las dos categorías generales para controlar errores? ¿Cuál es la diferencia entre ellas?

13-12. Explique las siguientes técnicas de detección de errores: redundancia, codificación de cuenta exacta, paridad, verificación de redundancia vertical, verificación de redundancia longitudinal y verificación de redundancia cíclica.

13-13. ¿Cuál técnica de detección de errores es la más simple?

13-14. ¿Cuál técnica de detección de errores es la más confiable?

13-15. Explique las siguientes técnicas de corrección de errores: sustitución de símbolos, retransmisión y corrección de error en sentido directo.

13-16. ¿Cuál técnica de corrección de errores está diseñada para usarse en un ambiente con humanos?

13-17. ¿Cuál técnica de corrección de errores es la más confiable?

13-18. Defina lo que es *sincronización de caracteres*.

13-19. Describa el formato de datos asíncronos.

13-20. Describa el formato de datos síncronos.

13-21. ¿Cuál formato de datos se adapta mejor a los mensajes largos? ¿Por qué?

13-22. ¿Qué es una *unidad de asignación*?

13-23. Describa las funciones de una unidad de control.

13-24. ¿Cuál es el objetivo de un módem de datos?

13-25. ¿Cuáles son las funciones principales del UART?

13-26. ¿Cuál es la cantidad máxima de bits que puede formar un solo carácter con un UART?

13-27. ¿Qué indican las señales de estado RPE, RFE y ROR?

13-28. ¿Por qué el reloj de recepción de un UART funciona con una rapidez 16 veces mayor que la frecuencia de recepción de bits?

13-29. ¿Cuáles son las mayores diferencias entre un UART y un USRT?

13-30. ¿Cuál es el objeto de la interfaz serie?

13-31. ¿Cuál es la interfaz serie más notable en Estados Unidos?

13-32. ¿Por qué la EIA estableció la interfaz RS-232?

13-33. ¿Cuál es la longitud nominal máxima para la interfaz RS-232?

13-34. ¿Cuáles son las cuatro clasificaciones más generales de las terminales de la interfaz RS-232?

13-35. ¿Cuál es el voltaje positivo máximo que produce un excitador?

13-36. ¿Qué clasificación de terminales usa lógica negativa?

13-37. ¿Cuál es la diferencia principal entre la interfaz RS-449A y la RS-232?

13-38. ¿Son posibles mayores frecuencias de bits con un cable de interfaz: *balanceado*, o *desbalanceado*?

13-39. ¿Quién proporciona el medio de transmisión de uso más común en los circuitos de comunicaciones? ¿Por qué?

13-40. Explique las diferencias en los circuitos DDD y los de línea privada.

13-41. Defina los siguientes términos: *lazo local, troncal, uso común* y *matriz de conmutación*.

13-42. ¿Qué es un DCE?

13-43. ¿Cuál es la diferencia principal entre un módem síncrono y uno asíncrono?

13-44. ¿Qué es necesario para operación dúplex con un circuito de dos conductores?

13-45. ¿Qué significan *modo de origen* y *modo de respuesta*?

13-46. ¿Qué esquema de modulación se usa en aplicaciones de baja velocidad? ¿En aplicaciones de velocidad intermedia? ¿En aplicaciones de alta velocidad?

13-47. ¿Por qué se requieren módems síncronos para aplicaciones en velocidades intermedia y alta?

13-48. En general, las frecuencias mayores sufren más atenuación que las menores, cuando pasan por un cable de longitud fija. ¿Cierto o falso?

13-49. Un número telefónico de identificación de cable cargado indica tres características de ese cable. ¿Cuáles son?

13-50. Un cable metálico funciona como, ¿cuál tipo de filtro?

13-51. Un cable se carga agregándole _____ en serie con él.

13-52. Desviación a 1004 Hz es sinónimo de retardo de envolvente. ¿Cierto o falso?

13-53. Las líneas de discado se pueden acondicionar. ¿Cierto o falso?

13-54. ¿Qué tipo de línea tiene el acondicionamiento mínimo?

13-55. Se puede alcanzar el acondicionamiento tipo-D agregando inductores en serie con la línea. ¿Cierto o falso?

13-56. El acondicionamiento tipo-D es obligatorio si la rapidez de transmisión es _____ bps.

13-57. ¿Cuál frecuencia se usa como referencia para mediciones de distorsión por atenuación, cuando se determina el tipo de acondicionamiento?

13-58. ¿Cuál frecuencia se usa como referencia para mediciones de retardo de envolvente, cuando se determina el tipo de acondicionamiento?

13-59. Todas las frecuencias tardan lo mismo en propagarse por un cable dado de longitud fija. ¿Cierto o falso?

13-60. Para determinar la cantidad de distorsión de fase (*¿relativa*? o *¿absoluta*?), el parámetro más importante es el retardo de fase.

13-61. Una línea acondicionada C4 debe cumplir con requisitos (*¿más*? o *¿menos*?) estrictos que una línea acondicionada C2.

13-62. La distorsión de fase se debe a que las distintas frecuencias se tardan distinto en propagarse por una línea. ¿Cierto o falso?

13-63. La distorsión de fase es sinónimo de distorsión por retardo. ¿Cierto o falso?

13-64. Distorsión por retardo es sinónimo de EDD. ¿Cierto o falso?

13-65. La potencia máxima permisible transmitida para un circuito de marcado es 2 mW. ¿Cierto o falso?

13-66. La potencia máxima permisible de una señal transmitida en circuito de línea privada es _____ mW.

13-67. Todas las frecuencias de la banda de voz sufren la misma atenuación cuando pasan por un filtro de mensaje-C. ¿Cierto o falso?

13-68. Los filtros de mensaje-C se ponen en los circuitos para mejorar la relación de señal a ruido. ¿Cierto o falso?

13-69. ¿Cuál es la diferencia en medir un ruido de mensaje-C y uno angosto-C?

13-70. Para una línea privada, ¿cuál es la relación mínima de señal a angosto-C, cuando?:

(a) ¿Se transmite a menos de 9600 bps?
(b) ¿Se transmite a 9600 bps?

13-71. Las determinaciones empíricas resultan de demostraciones matemáticas de la teoría. ¿Cierto o falso?

13-72. Tres inestabilidades repentinas en un bloque de mensaje necesitarían acción diferente que una sola inestabilidad repentina en el mismo bloque. ¿Cierto o falso?

13-73. ¿Cuál es la diferencia entre una ganancia repentina y una inestabilidad de impulso?

13-74. Describa dos tipos de distorsión no lineal.

13-75. ¿Qué unidad de medida indica la intensidad real de la señal, respecto a la que debiera ser esa intensidad?

13-76. ¿Cuál es la magnitud del ruido, en dBm, que se usa como referencia?

13-77. ¿Cuál es la unidad de medida de ruido que se corrige a 0 TLP?

13-78. ¿La transmisión de datos relaciona una intensidad de una señal DLP que siempre es ____ dB menor que el TLP?

13-79. ¿Cuál es la unidad de ruido con que se mide el ruido de baja frecuencia?

PROBLEMAS

13-1. Determine la LRC y la VRC para el siguiente mensaje. Use paridad par para con LRC y paridad impar con VRC.

D A T A sp C O M M U N I C A T I O N S

13-2. Determine la BCS para los siguientes polinomios generadores de datos y comprobación CRC:

$$G(x) = x^7 + x^4 + x^2 + x^0 = 1\,0\,0\,1\,0\,1\,0\,1$$
$$P(x) = x^5 + x^4 + x^1 + x^0 = 1\,1\,0\,0\,1\,1$$

13-3. ¿Cuántos bits de Hamming se requieren para un solo carácter ASCII?

13-4. Determine los bits de Hamming para el carácter "B" en ASCII. Inserte los bits de Hamming en lugares uno sí y uno no, comenzando por la izquierda.

13-5. Determine la LRC y la VRC para el siguiente mensaje. Use paridad par con LRC y paridad impar con VRC.

P S S S E B P
A Y Y T A S C I I sp C O D E T C A
D N N X X C D

13-6. Determine la BCS para los siguientes polinomios generadores de datos y de CRC:

$$G(x) = x^8 + x^5 + x^2 + x^0$$
$$P(x) = x^5 + x^4 + x^1 + x^0$$

13-7. ¿Cuántos bits de Hamming se requieren para un carácter de ASCII (1 bit de arranque, 7 bits de datos, 1 bit de paridad y 2 bits de paro, en total 11 bits)?

C A P Í

Imagen protegida por derechos de autor

Protocolos de comunicación de datos y configuraciones de red

INTRODUCCIÓN

El objetivo primario de la *arquitectura de una red* es proporcionar a sus usuarios los medios necesarios para establecer la red y efectuar el control de flujo de datos. Una arquitectura de red describe la forma en que se arregla o estructura una red de comunicación de datos y, en general, incluye el concepto de *niveles* o *capas* dentro de la arquitectura. Cada nivel dentro de la red consiste en *protocolos* específicos, es decir, reglas de comunicación que desempeñan un conjunto dado de funciones.

Los protocolos son arreglos entre personas o procesos. En esencia, un protocolo es un conjunto de reglamentos acerca de la formalidad o precedencia, como por ejemplo un protocolo militar o diplomático. Un *protocolo de red de comunicación de datos* es un conjunto de reglas que gobierna el intercambio ordenado de datos dentro de la red.

Como se dijo antes, la función de una unidad de control de línea es controlar el flujo de datos entre el programa de aplicaciones y las terminales remotas. Por consiguiente, debe haber un conjunto de reglas que determinen cómo reacciona una LCU (unidad de control de línea) a distintas clases de transmisión. Este conjunto de reglas se llama *protocolo de enlace de datos*, que en esencia es a su vez, un conjunto de procedimientos que incluyen secuencias precisas de caracteres, para asegurar un intercambio ordenado de datos entre dos unidades de control de línea.

En un circuito de comunicación de datos, la estación que transmite en el momento se llama *estación maestra*, y la estación que recibe se llama *esclava*. En una red centralizada, la estación primaria controla cuándo puede transmitir cada estación secundaria. Cuando transmite una estación secundaria se convierte en estación maestra, y la estación primaria es ahora la esclava. El papel de estación maestra es temporal, y la estación primaria determina cuál estación es maestra. Al principio, la estación primaria es maestra. La estación primaria solicita por turno a cada estación secundaria *interrogándola*. Una interrogación es una invitación de la primaria a una secundaria para que transmita un mensaje. Las estaciones secundarias no pueden interrogar a una primaria. Cuando una primaria interroga a una secundaria, inicia un *cambio de dirección de línea*; la secundaria interrogada ha sido designada como maestra y debe responder. Si la primaria *selecciona* una secundaria, ésta se identifica como receptora. Una selección es una

interrogación, por parte de una primaria o una secundaria, para determinar el estado de la secundaria (es decir, lista para recibir o no lista para recibir un mensaje). Las estaciones secundarias no pueden seleccionar a la primaria. Las transmisiones de la primaria van a todas las secundarias, y depende de las estaciones secundarias la decodificación individual de cada transmisión, y la determinación de si es para ellas. Cuando una secundaria transmite, sólo manda a la primaria.

Los protocolos de enlace de datos se clasifican en general como asíncronos o síncronos. Por regla, los protocolos asíncronos usan un formato de datos asíncronos y módems asíncronos, mientras que los protocolos síncronos usan un formato de datos síncronos y módems síncronos.

INTERCONEXIÓN DE SISTEMAS ABIERTOS

El término *interconexión de sistemas abiertos* (OSI, de *open systems interconnection*) es el nombre de un conjunto de normas para comunicaciones entre computadoras. El objetivo principal de las normas OSI es contar con un lineamiento estructural para intercambiar información entre computadoras, terminales y redes. El OSI está patrocinado por ISO y también por CCITT, que trabajaron en conjunto para establecer un grupo de normas ISO y de recomendaciones CCITT que en esencia son idénticas. En 1983, ISO y CCITT adoptaron un modelo de referencia con arquitectura de comunicaciones de siete capas. Cada capa consiste en protocolos específicos para comunicación.

La jerarquía de protocolos ISO

El modelo ISO, de interconexión de sistemas abiertos con siete capas, se ve en la fig. 14-1. Esta jerarquía se desarrolló para facilitar las comunicaciones del equipo procesador de datos, separando las responsabilidades en la red en siete capas distintas. El concepto básico de estratificar las responsabilidades es que cada capa agregue valores a los servicios suministrados por los conjuntos de las capas inferiores. De esta manera, el nivel más alto cuenta con el conjunto completo de servicios necesarios para hacer funcionar una aplicación de datos distribuidos.

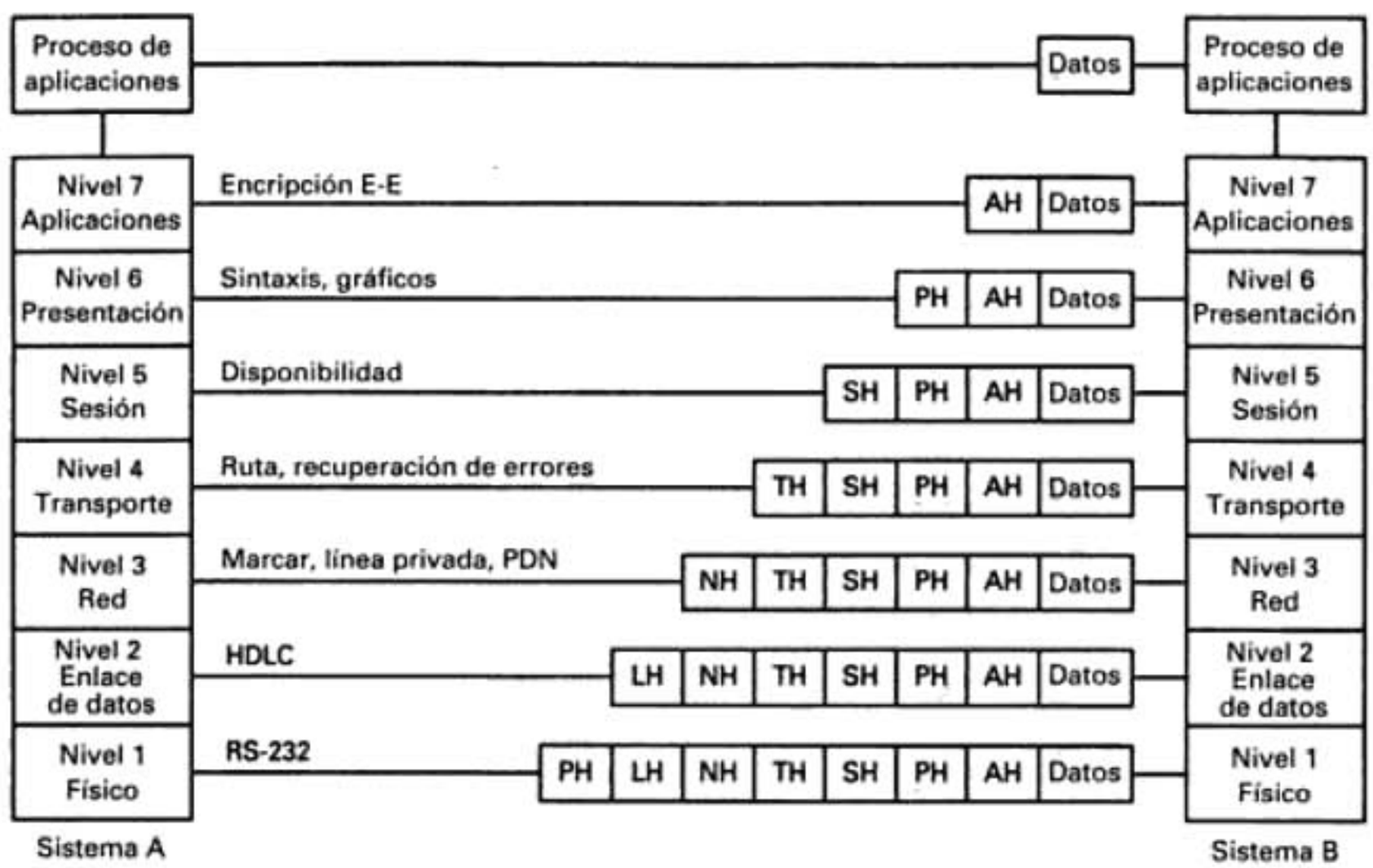


FIGURA 14-1 Jerarquías del protocolo internacional ISO. AH = encabezado de aplicaciones; PH = encabezado de presentación; SH = encabezado de sesión; TH = encabezado de transporte; NH = encabezado de red; LH = encabezado de enlace; PH = encabezado físico

Hay varias ventajas por usar una arquitectura estratificada en el modelo OSI. Las diversas capas permiten que se comuniquen diversas computadoras en distintos niveles. Además, al avanzar la tecnología, es más fácil modificar el protocolo de una capa sin tener que modificar todas las demás. Cada capa es, en esencia, independiente de las demás. Por consiguiente, muchas de las funciones que se encuentran en las capas inferiores se han eliminado por completo de las tareas de programas (*software*), y se han reemplazado por componentes (*hardware*). En la fig. 14-1 se ven algunos ejemplos de estas funciones. La desventaja principal de la arquitectura de siete niveles es la cantidad tan tremenda de indirectos necesarios para agregar encabezados a la información que se transmite entre las diversas capas. De hecho, si se tienen en cuenta las siete capas, menos de 15% del mensaje transmitido es información de la fuente; el resto es indirecto. En la fig. 14-1 se ve el resultado de agregar encabezados a cada capa.

Los niveles 4, 5, 6 y 7 permiten comunicarse a dos computadoras anfitrión en forma directa. Las tres capas inferiores tienen que ver con la mecánica real de pasar datos (a nivel de bit) de una máquina a otra. A continuación se resumen los servicios básicos suministrados por cada capa de la jerarquía.

1. *Capa física.* La capa física es el nivel más bajo de la jerarquía, y especifica las normas físicas, elécricas, funcionales y de procedimiento para entrar a la red de comunicación de datos. En esta capa se hacen definiciones como por ejemplo de valores máximos y mínimos de voltaje y de la impedancia del circuito. Las especificaciones para la capa física se parecen a las especificadas por la norma RS-232 de EIA para interfaces serie.

2. *Capa de enlace de datos.* Esta capa es responsable de las comunicaciones entre nodos primarios y secundarios de la red. La capa de enlace de datos proporciona un medio para activar, mantener y desactivar el enlace de datos. También proporciona la trama final de la envolvente de información, facilita el flujo ordenado de datos entre nodos, y permite la detección y corrección de errores. Como ejemplo de protocolos de enlace de datos está el control de enlace de comunicaciones bisíncronas (Bisync) y síncronas (SDLC, de *synchronous data link control*) de IBM.

3. *Capa de red.* La capa de red determina cuál configuración de red (marcar, rentada o de paquete) es más adecuada para la función que proporciona la red. También esa capa define el mecanismo con el que los mensajes se dividen en paquetes de datos, y se conducen de un nodo de transmisión a uno de recepción, dentro de una red de comunicaciones.

4. *Capa de transporte.* Esta capa controla la integridad del mensaje, de principio a fin, y en eso se incluye la ruta, la segmentación y la recuperación de errores para el mensaje. La capa de transporte es la más alta, en lo que se refiere a comunicaciones. Las capas superiores a la de transporte no intervienen en los aspectos tecnológicos de la red. Las tres capas superiores manejan los aspectos de aplicaciones de la red, mientras que las tres inferiores manejan la transferencia de mensajes. Así, la capa de transporte funciona como interfaz entre las capas de red y de sesión.

5. *Capa de sesión.* Es la responsable de la disponibilidad de la red (es decir, de capacidad de almacenamiento y del procesador). Entre las responsabilidades de sesión están los procedimientos de entrada y salida a la red, y la verificación de usuarios. Una sesión es una condición temporal que existe cuando los datos están transmitiéndose en realidad, y no incluye procedimientos como establecer llamada, preparar o desconectar. La capa de sesión determina la clase de diálogo disponible (es decir, símplex, semidúplex o dúplex).

6. *Capa de presentación.* Esta capa maneja toda conversión de código o de sintaxis necesaria para presentar los datos a la red, en un formato común para las comunicaciones. Entre las presentaciones se incluye el dar formato, codificar (ASCII, EBCDIC, etc.), de archivos de datos, cifrado y descifrado de mensajes, procedimientos de diálogo, compresión de datos, sincronización, interrupción y terminación. La capa de presentación hace la traducción del código y del conjunto de caracteres, y determina el mecanismo de presentación de los mensajes.

7. *Capa de aplicación.* Es la máxima en la jerarquía, y es análoga al administrador general de la red. La capa de aplicación controla la secuencia de actividades dentro de una aplicación, y también la secuencia de eventos entre la aplicación de cómputo y el usuario de otra aplicación. La capa de aplicación se comunica en forma directa con el programa de aplicación del usuario.

Los *modos de transmisión de datos* describen cómo transmiten caracteres alfanuméricos de datos las terminales operadas por humanos a las computadoras anfitrión. Básicamente, sólo hay dos modos disponibles de transmisión: modo de carácter y modo de bloque.

Modo de carácter

Cuando se trabaja en el modo de carácter, los códigos de caracteres se transmiten a la computadora anfitrión inmediatamente después de que un operador haya oprimido una tecla. El carácter se manda asíncronamente, porque las transmisiones no están sincronizadas con la velocidad de tecleo del operador. Cuando un operador no está tecleando, la terminal está en *estado inactivo*. Los caracteres de datos que se transmiten del anfitrión a la terminal remota se muestran en la pantalla, en la posición actual del cursor. Los caracteres no de datos (timbre, CR, LF, etc.) se procesan según se necesite.

Modo de bloque

En el modo de bloque en transmisión de datos, los caracteres no se transmiten de inmediato a medida que se teclean. En lugar de ello, el operador los teclea en su terminal, donde se almacenan en memorias búfer y se muestran en la pantalla. Cuando el operador está listo para transmitir la información que aparece en la pantalla, oprime la tecla INTRO, o ↵, que transmite todos los caracteres de datos que haya en la terminal local. Los caracteres transmitidos en grupo se llaman *bloque* de datos. El formato que se usa dentro del bloque depende del protocolo seleccionado de sistema. La mayoría de las terminales modernas e inteligentes son capaces de funcionar en cualquiera de los dos modos, bloque o carácter. El modo de transmisión de carácter es más común cuando las terminales se comunican en forma directa con la computadora anfitrión mediante una conexión de comunicaciones directa, como por ejemplo, líneas telefónicas de marcar. El modo de transmisión en bloque es más adecuado para circuitos de comunicaciones de datos con varias terminales, que funcionan en un ambiente de interrogación.

PROTOCOLOS ASÍNCRONOS

Dos protocolos de datos asíncronos, de los que más se usan, son el *sistema de llamada selectiva*, de Western Electric (8A1/8B1) y el *protocolo de enlace de datos asíncronos*, de IBM (83B). En esencia, estos dos protocolos son el mismo conjunto de procedimientos.

Los protocolos asíncronos están *orientados a caracteres*. Esto es, los caracteres únicos de control de enlace de datos, como por ejemplo el fin de transmisión (EOT) e inicio de texto (STX) garantizan la misma acción o hacen la misma función, sin importar dónde estén en una transmisión. Por ejemplo, el carácter de fin de transmisión que se usa en ASCII es 04H. No importa cuándo se reciba el 04H en un secundario, se borra la unidad de control de línea (LCU) y se pone en el modo de vigilancia de línea. En consecuencia, se debe tener cuidado para asegurar que las secuencias de bits para los caracteres de control de enlace de datos, no se presenten dentro de un mensaje, a menos que sean para ejecutar sus funciones asignadas de enlace de datos. El único tipo de detección de errores que usan los protocolos asíncronos es la verificación de redundancia vertical, y para corregir errores se usan sustitución de símbolos y ARQ (retransmisión). En los protocolos asíncronos, en general cada estación secundaria se limita a un solo par terminal/impresora. A este arreglo de estación se le llama *independiente*. Con la configuración independiente, todos los mensajes que se transmiten o se reciben en la pantalla de la terminal también se imprimen. Así, la impresora sólo genera una copia en papel de todas las transmisiones.

Además del modo de vigilancia de línea, una estación remota puede estar en cualquiera de los tres modos de operación: *transmisión, recepción* y *local*. Una estación secundaria está en el modo de transmisión siempre que haya sido designada maestra. En el modo de transmisión la secundaria puede mandar mensajes o reconocimientos formateados. Una secundaria está en el modo de recepción cuando haya sido seleccionada por la primaria. En el modo de recepción, la secundaria puede recibir mensajes formateados de la primaria. Para que un operador de terminal capture información en su terminal de cómputo, ésta debe estar en el modo local. Una terminal se puede poner en modo local mediante órdenes de programa, mandados de la primaria, o el operador lo puede hacer en forma manual desde el teclado.

La secuencia de interrogación en la mayoría de los protocolos asíncronos es bastante sencilla, y suele consistir en mandar uno o dos caracteres de control de enlace de datos, después una *dirección de interrogación de estación*. Una secuencia típica de interrogación es

```
E D
O C A
T 3
```

El carácter EOT es el de *borrado*, y siempre antecede a la secuencia de interrogación. Con EOT todas las secundarias se ponen en modo de vigilancia de línea. Cuando una estación está en modo de vigilancia (*monitoreo*) de línea, escucha a la línea para conocer su dirección de interrogación o selección. Cuando inmediatamente después de EOT sigue DC3, quiere decir que el siguiente carácter es de interrogación de estación. Para este ejemplo, la dirección de interrogación de estación es el carácter único "A" en ASCII. Se ha seleccionado la estación A como maestra, y debe responder con un mensaje formateado o con un reconocimiento. Hay dos secuencias de reconocimiento que se pueden transmitir como respuesta a una interrogación. Se muestran a continuación, junto con sus funciones.

Reconocimiento		Función
\	A C K	Sin mensaje por transmitir, listo para recibir
\	\	Sin mensaje por transmitir, no listo para recibir

La secuencia de selección, que se parece mucho a la de interrogación, es

```
E
O X Y
T
```

De nuevo, el carácter EOT se transmite primero para asegurar que todas las estaciones secundarias estén en el modo de vigilancia de línea. Después de EOT está una dirección de selección de dos caracteres, "XY". Se ha seleccionado la estación XY por la primaria, y se ha designado como receptor. Una vez seleccionada, una estación secundaria debe responder con una de las tres secuencias de reconocimiento, que indican su estado. Se ven a continuación, junto con sus funciones.

Reconocimiento		Función
\	A C K	Listo para recibir
\	\	No listo para recibir, terminal en local o impresora sin papel
*	*	No listo para recibir, hay mensaje formateado para transmitir

Se puede seleccionar más de una estación al mismo tiempo, con direcciones *de grupo* y direcciones *globales*. Las direcciones de grupo se usan cuando la primaria desea seleccionar más de una, pero no a todas las estaciones remotas. Hay una sola dirección global, que se usa para seleccionar en forma simultánea a todas las estaciones remotas. En los protocolos asíncronos rara vez se usan procedimientos de reconocimiento de selecciones de grupo y globales, porque son algo complicados.

Los mensajes transmitidos de la primaria y la secundaria usan exactamente el mismo formato de datos, que es el siguiente

```
S                E
T message data   O
X                T
```

Solución (a) Según la tabla 14-1, la SPA para la estación 8 es H; por consiguiente, la secuencia para una interrogación general es

```
P S S E P S S           E P
A Y Y O A Y Y H H " "   N A
D N N T D N N           Q D
```

(b) De acuerdo con la tabla 14-1, la DA del dispositivo 6 es F; por consiguiente, la secuencia para una interrogación específica es

```
P S S E P S S           E P
A Y Y O A Y Y H H F F   N A
D N N T D N N           Q D
```

(c) De la tabla 14-1, la SSA para la estación 8 es Y; por consiguiente, la secuencia para una selección es

```
P S S E P S S           E P
A Y Y O A Y Y Y Y F F   N A
D N N T D N N           Q D
```

Con bisync, sólo hay dos formas para que una secundaria responda a una interrogación: con un mensaje formateado o con un *acuerdo*. Un acuerdo no es más que una respuesta de la secundaria que indica que no tiene mensajes formateados para transmitir; es decir, un acuerdo es un reconocimiento negativo de una pregunta. La secuencia de caracteres para un acuerdo es

```
P S S E P
A Y Y O A
D N N T D
```

Una secundaria puede responder a una selección con un acuerdo positivo o negativo. Un reconocimiento positivo a una selección indica que el dispositivo seleccionado está listo para recibir. La secuencia de caracteres para un reconocimiento positivo es

```
P S S D   P
A Y Y L O A
D N N E   D
```

Un reconocimiento negativo a una selección indica que el dispositivo seleccionado no está listo para recibir. Un reconocimiento negativo se llama *interrupción inversa* (RVI, de *reverse interruption*). La secuencia de caracteres de una RVI es

```
P S S D   P
A Y Y L < A
D N N E   D
```

Con bisync, los mensajes formateados se mandan de una secundaria a la primaria, en respuesta a una interrogación, y se mandan de la primaria a la secundaria, después de haber seleccionado a ésta. Los mensajes usan el siguiente formato

```
P S S S           S           E B P
A Y Y O heading T message T   C A
D N N H           X           X C D
```

Nota: Si se usa CRC-16 para detección de errores, hay dos caracteres de verificación de bloque.

La comprobación de redundancia longitudinal (LRC, de *longitudinal redundancy checking*) se usa para detectar errores con mensajes codificados en ASCII, y la verificación de redundancia cíclica (CRC, de *cyclic redundancy checking*) se usa para EBCDIC. Se calcula el carácter de comprobación de bloque (BCC, de *block check character*) comenzando con el primer carácter después de SOH (*start of heading*, inicio del encabezado) y continuando, incluyendo ETX, el fin del texto. (Si no hay encabezado, el BCC se calcula comenzando con el primer carácter después de STX, el inicio del texto.) Con protocolos síncronos, los datos se transmiten en bloques. En general, los bloques de datos se limitan a 256 caracteres. Se usa ETX para terminar el último bloque de un mensaje. Se usa ETB (fin de bloque de texto) con mensajes de varios bloques, para terminar todos los bloques excepto el último. El último mensaje de un bloque siempre termina

comienzan con sus contadores nr y nm restablecidos en 000. La primaria manda tres tramas numeradas de información (nm = 0, 1 y 2). Al mismo tiempo, la primaria manda nr = 0, porque la siguiente trama que espera recibir es la trama 0, que es igual al nm actual de la secundaria. La secundaria responde con dos tramas de información (nm = 0 y 1). La secundaria recibió las tres tramas de la primaria sin errores, por lo que el nr transmitido en el campo de control de la secundaria es 3 (que es el número de la siguiente trama que mandará la primaria). Ahora la primaria manda las tramas de información 3 y 4 con un nr = 2. El nr = 2 confirma la recepción correcta de las tramas 0 y 1. La secundaria responde con las tramas nm = 2, 3 y 4 con un nr = 4. El nr = 4 confirma sólo la recepción de la trama 3 de la primaria (nr − 1) y, en consecuencia, la primaria debe retransmitir la trama 4. Ésta se retransmite con cuatro tramas adicionales (nm = 5, 6, 7 y 0). El nr de la primaria es 5, que confirma las tramas 2, 3 y 4 de la secundaria. Por último, la secundaria marca la trama de información 5 con un nr = 1. El nr = 1 confirma las tramas 4, 5, 6, 7 y 0 de la primaria. En este punto, se han confirmado todas las tramas transmitidas, a excepción de la trama 5 de la secundaria.

nm de la primaria:	0 1 2	3 4	4 5 6 7 0
nr de la primaria:	0 0 0	2 2	5 5 5 5 5
nm de la secundaria:	0 1	2 3 4	5
nr de la secundaria:	3 3	4 4 4	1

En SDLC una estación nunca puede mandar más de siete tramas numeradas sin recibir una confirmación. Por ejemplo, si la primaria manda ocho tramas (nm = 0, 1, 2, 3, 4, 5, 6 y 7) y la secundaria responde con un nr = 0, queda ambiguo cuáles tramas se confirman. El nr = 0 ¿indica que las ocho tramas se recibieron en forma correcta, o que la trama 0 tuvo un error y se deben retransmitir las ocho tramas? (En SDLC todas las tramas transmitidas antes que comienzan con la trama nr − 1 se deben retransmitir.)

El bit b_3 es el de *interrogación* (P de *poll*) o de *no interrogación* ($\bar{P}$) cuando lo manda la primaria, y el *final (F)* o *no final* ($\bar{F}$) cuando lo manda una secundaria. En una trama mandada por la primaria, si ésta desea interrogar a la secundaria, se activa el bit P (1). Si la primaria no desea interrogar a la secundaria, el bit P se restablece (0). Una secundaria no puede transmitir a menos que reciba una trama dirigida a ella, con el bit P activado. En una trama que manda una secundaria, si es la última (final) del mensaje, se activa el bit F (1). Si no es la trama final, se restablece el bit F (0). En las tramas I, la primaria puede seleccionar una estación secundaria, mandarle información formateada, confirmar las tramas I recibidas antes, e interrogar, con una sola transmisión.

Ejemplo 14-2

Determinar la secuencia de bits para el campo de control de una trama mandada de la estación primaria a una secundaria para las siguientes condiciones: la primaria está mandando la trama de información 3, es una interrogación, y la primaria está confirmando la recepción correcta de las tramas 2, 3 y 4 de la secundaria.

Solución

$b_7 = 0$ porque es una trama de información.
b_4, b_5 y b_6 son 011 (3 binario, por nm = 3).
$b_3 = 1$, es una trama de interrogación.
b_0, b_1 y b_2 son 101 (5 binario, por nr = 5).
campo de control = B6H

b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1	0	1	1	0	1	1	0

Trama de supervisión. No se permite que un campo de información esté en una trama de supervisión. En consecuencia, las tramas de supervisión no se pueden usar para transmitir información; se usan para ayudar en la transferencia de la información. Las tramas de supervisión

son para confirmar las tramas de información recibidas antes, o para informar condiciones de listo o de ocupado, y para informar errores de numeración de tramas. La secuencia de bits del campo de control de una trama de supervisión es la siguiente

Bit:	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
Función:	←	nr	→	P o F $\overline{P}$ o $\overline{F}$	X	X	0	1

Una trama de supervisión se identifica por un 01 en los bits b_6 y b_7, respectivamente, del campo de control. En el formato de supervisión el bit b_3 es de nuevo de interrogación o no interrogación, o final o no fnal, y b_0, b_1 y b_2 son los bits de nr. Sin embargo, en el formato de supervisión se usan los bits b_4 y b_5 para indicar el estado de recepción de la estación que transmite la trama, o para pedir transmisión o retransmisión de las tramas secuenciadas de información. Con dos bits hay cuatro combinaciones posibles, que se indican con sus funciones respectivas

b_4	b_5	Estado del receptor
0	0	Listo para recibir (RR)
0	1	No listo para recibir (RNR)
1	0	Rechazo (REJ)
1	1	No se usa en SDLC

Cuando la primaria manda una trama de supervisión con el bit P activado y un estado de lista para recibir, equivale a una interrogación general en bisync. Las tramas de supervisión las usa la primaria para interrogar y para confirmar las tramas de información recibidas antes, cuando no hay información por mandar. Una secundaria usa el formato de supervisión para confirmar las tramas de información recibidas antes, y para informar su estado de recepción a la primaria. Si la secundaria manda una trama de supervisión con estado RNR, la primaria no le puede mandar tramas numeradas de información hasta que desaparezca ese estado. Se borra RNR cuando una secundaria marca una trama de información con el bit F = 1, o una trama RR o REJ con el bit F = 0. La orden/respuesta REJ se usa para confirmar tramas de información hasta la nr − 1, y para pedir la retransmisión de tramas numeradas de información que comienzan con el número de trama identificado en la trama REJ. En una trama de supervisión se prohíbe tener un campo de información, y sólo se usa la orden o respuesta REJ en la operación dúplex.

Ejemplo 14-3

Determinar la secuencia de bits en el campo de control de una trama de supervisión que manda una estación secundaria a la primaria, para las siguientes condiciones: la secundaria está lista para recibir, es la trama final y la estación secundaria confirma las tramas 3, 4 y 5.

Solución

b_6 y b_7 = 01, porque es una trama de supervisión.
b_4 y b_5 = 00 (listo para recibir).
b_3 = 1 porque es la trama final.
b_0, b_1 y b_2 = 110 (6 en binario, por nr = 6).
campo de control = D1H

b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
1	1	0	1	0	0	0	1

Trama no numerada. Una trama no numerada se identifica haciendo que los bits b_6 y b_7 en el campo de control sean 11. La secuencia de bits en el campo de control de una trama no numerada es

Bit:	b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7
Función:	X	X	X	P o F $\overline{P}$ o $\overline{F}$	X	X	1	1

TABLA 14-2 Órdenes y respuestas no numeradas

Configuración binaria							
b_0		b_7	Acrónimo	Orden	Respuesta	I campo prohibido	Restablece nm y nr
000	P/F	0011	UI	Sí	Sí	No	No
000	F	0111	RIM	No	Sí	Sí	No
000	P	0111	SIM	Sí	No	Sí	Sí
100	P	0011	SNRM	Sí	No	Sí	Sí
000	F	1111	DM	No	Sí	Sí	No
010	P	0011	DISC	Sí	No	Sí	No
011	F	0011	UA	No	Sí	Sí	No
100	F	0111	FRMR	No	Sí	No	No
111	F	1111	BCN	No	Sí	Sí	No
110	P/F	0111	CFGR	Sí	Sí	No	No
010	F	0011	RD	No	Sí	Sí	No
101	P/F	1111	XID	Sí	Sí	No	No
001	P	0011	UP	Sí	No	Sí	No
111	P/F	0011	TEST	Sí	Sí	No	No

En una trama no numerada el bit b_3 vuelve a ser el P/$\overline{P}$ o el F/$\overline{F}$. Los bits b_0, b_1, b_2, b_4 y b_5 se usan para diversas órdenes y respuestas no numeradas. Como hay cinco bits disponibles es posible tener 32 órdenes y respuestas no numeradas. El campo de control de una trama no numerada que manda la primaria es una orden. El campo de control de una trama no numerada que manda una secundaria es una respuesta. En las tramas no numeradas no hay bits nm o nr. Por consiguiente, las tramas numeradas de información no se pueden mandar ni confirmar con el formato no numerado. Las tramas no numeradas son para mandar información sobre control de red y sobre estado. Dos ejemplos de las funciones de control son (1) poner en línea y sacar de línea las estaciones secundarias, y (2) inicializar la unidad de control de línea. La tabla 14-2 muestra varias de las órdenes y respuestas que se usan con más frecuencia. Se prohíbe un campo de información en todas las órdenes y respuestas no numeradas, excepto UI, FRMR, CFGR, TEST y XID.

Una estación secundaria debe estar en uno de los tres modos siguientes: inicialización, respuesta normal o desconexión normal. Los procedimientos del *modo de inicialización* los especifica el sistema, y varían mucho. Una secundaria en el *modo de respuesta normal* no puede iniciar transmisiones no solicitadas; sólo puede transmitir en respuesta a una trama recibida con el bit P activado. Cuando está en el *modo de desconexión normal*, una secundaria está fuera de línea. En este modo, una secundaria sólo puede recibir las órdenes TEST, XID, CFGR, SNRM o SIM de la primaria, y sólo puede responder cuando el bit P está activado.

Las órdenes y respuestas no numeradas se resumen a continuación:

Información no numerada (*UI*, de *unnumbered information*). Es una orden o respuesta para mandar información no numerada. La información no numerada que se transmite en el campo I no se confirma.

Establecer modo de inicialización (SIM, de *set initialization mode*). Es una orden que pone la estación secundaria en el modo de inicialización. El procedimiento de inicialización es especificado por el sistema, y varía desde una autoprueba sencilla del controlador de estación, hasta la ejecución completa de un programa IPL (de *initial program logic*, lógica inicial de programa). La orden SIM restablece los contadores nm y nr en las estaciones primaria y secundaria. Se espera que una secundaria responda con UIA a una orden SIM.

Pedir modo de inicialización (RIM, de *request initialization mode*). Es una respuesta que manda una estación secundaria a la petición de la primaria de mandar una orden SIM.

Establecer modo de respuesta normal (SNRM, de *set normal response mode*). Es una orden que pone a una estación secundaria en el modo de respuesta normal (NRM, de *normal*

response mode). Una estación secundaria no puede mandar ni recibir tramas numeradas de información, a menos que esté en el modo de respuesta normal. En esencia, la orden SNRM pone en línea a una estación secundaria, y restablece los contadores nm y nr en las estaciones primaria y secundaria. UA es la respuesta normal a una orden SNRM. No se permiten las respuestas no solicitadas cuando la secundaria está en el modo NRM. Una secundaria permanece en NRM hasta que recibe una orden DISC o SIM.

Modo de desconexión (DM, de *disconnect mode*). Es una respuesta que manda una estación secundaria, si la primaria trata de mandarle tramas no numeradas de información cuando la secundaria está en el modo de desconexión normal.

Petición de desconexión (RD, de *request disconnect*). Es una respuesta que manda una secundaria cuando quiere ponerse en el modo de desconexión.

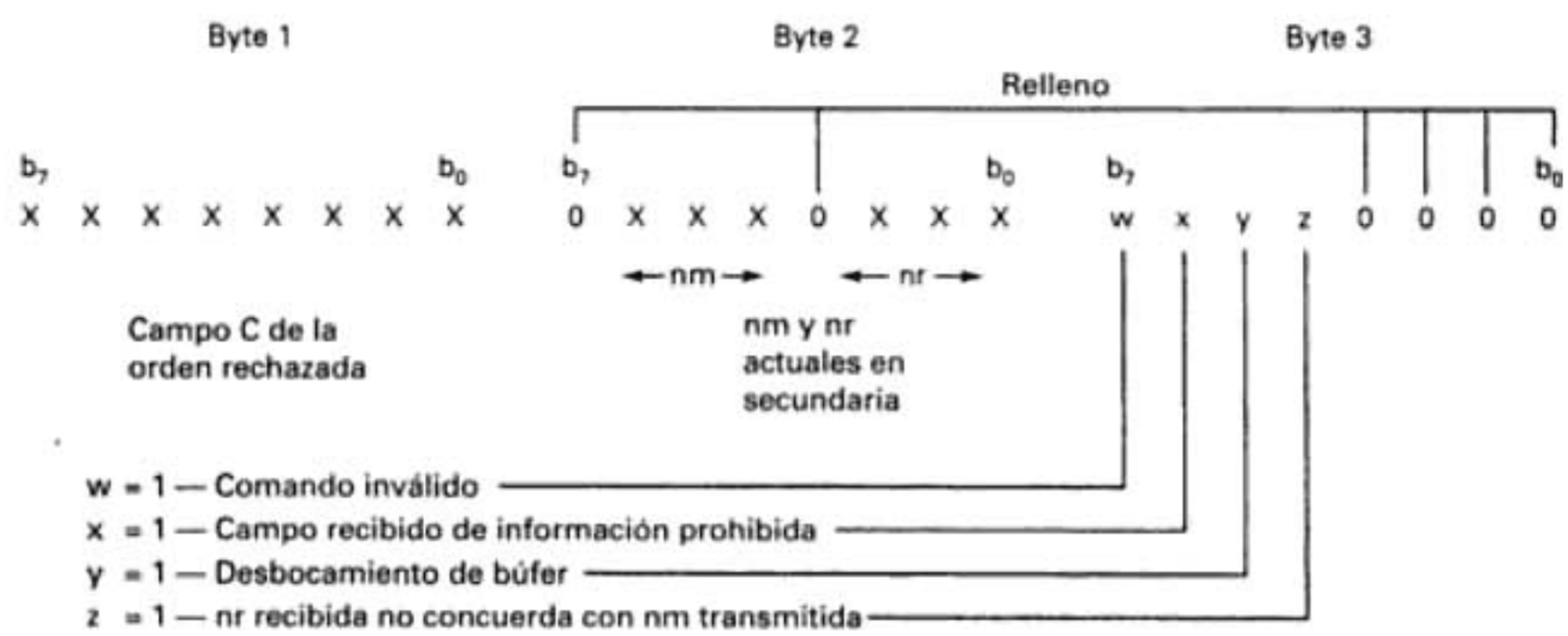
Desconexión (DISC, de *disconnect*). Es una orden que pone a una estación secundaria en el modo de desconexión normal (NDM, de *normal disconnect mode*). Una secundaria no puede mandar ni recibir tramas numeradas de información cuando está en el modo de desconexión normal. Cuando está en ese modo, una secundaria sólo puede recibir una orden SIM o SNRM y sólo puede transmitir una respuesta DM. La respuesta esperada a una orden DISC es UA.

Reconocimiento no numerado (UA, de *unnumbered acknowledgement*). Es una respuesta afirmativa que indica el apego a una orden SIM, SNRM o DISC. También se usa UA para reconocer tramas no numeradas de información.

Rechazo de trama (FRMR, de *frame reject*). Es para informar errores de procedimiento. la secuencia FRMR es una respuesta que transmite la secundaria cuando ha recibido una trama inválida de la primaria. Una trama recibida puede ser inválida por cualquiera de las siguientes razones:

1. El campo de control contiene una orden inválida o no asignada.
2. La cantidad de datos en el campo de información es mayor que el espacio de memoria en la secundaria.
3. Un campo de información se recibe en una trama que no permite información.
4. El nr es incongruente con el nm de la secundaria. Por ejemplo, si la secundaria transmitió las tramas con nm 2, 3 y 4 y la secundaria respondió con un nr igual a 7.

Una secundaria no se puede salir de la condición FRMR ni actúa sobre la trama que causó esa condición. La secundaria repite la respuesta FRMR hasta que recibe una de las siguientes órdenes de *establecer modo*: SNRM, DISC o SIM. El campo de información de una respuesta FRMR siempre contiene tres bytes (24 bits) y tiene el siguiente formato



Probar (TEST). Es una orden que se puede mandar en cualquier modo para solicitar una respuesta TEST. Si con la orden se incluye un campo de información, la secundaria la regresa con la respuesta. La orden o respuesta TEST se intercambia para fines de prueba de enlace.

Intercambiar identificación de estación (XID, de *exchange identification*). Como orden, XID solicita la identificación de la estación secundaria. Se puede incluir una trama de información en esta trama, para transmitir los datos de identificación de la estación primaria o de la secundaria. Para los circuitos de marcar, con frecuencia es necesario que se identifique la estación secundaria para que la primaria intercambie con ella tramas de información, aunque XID no se restringe sólo a circuitos de datos de marcar.

Campo de comprobación de trama. Este campo FCS (de *frame check secuence*) contiene el mecanismo de detección de errores para SDLC. El FCS equivale al BCC que se usa en el bisync. SDLC usa CRC-16 y el siguiente polinomio generador

$$x^{16} + x^{12} + x^{5} + x^{1}$$

Operación del lazo SDLC

Un *lazo* SDLC funciona en el modo semidúplex. La diferencia principal entre las configuraciones de lazo y de bus es que en un lazo, todas las transmisiones tienen la misma dirección en el canal de comunicaciones. En una configuración de lazo sólo una estación transmite a un tiempo. Primero transmite la primaria y a continuación cada estación secundaria responde en secuencia. En un lazo SDLC, el puerto de transmisión del controlador de la estación primaria se conecta con una o más estaciones secundarias en modo serie; a continuación el lazo se cierra en el puerto de recepción de la primaria. La fig. 14-4 muestra una configuración de lazo SDLC.

En un lazo SDLC, la primaria transmite tramas que se dirigen a cualquiera de las estaciones secundarias, o a todas. Cada trama transmitida por la primaria contiene la dirección de la estación secundaria a la que se dirige la trama. A su vez, cada estación secundaria decodifica el campo de dirección de cada trama, y a continuación funciona como repetidora para todas las estaciones que le siguen en el lazo. Si una secundaria detecta una trama con su dirección, la acepta y a continuación la pasa a la siguiente estación después de ella. Todas las tramas transmitidas por la primaria regresan a ella. Cuando la primaria termina de transmitir, pone ocho ceros consecutivos después de la última bandera. Una bandera seguida de ocho ceros consecutivos se llama secuencia de *cambio de dirección*, que indica el final de la transmisión de la primaria. De inmediato después de una secuencia de cambio de dirección, la estación primaria transmite unos (1) continuos, con lo cual se genera una secuencia de *proseguir* (01111111). Una secundaria no puede transmitir sino hasta que haya recibido una trama dirigida a ella con el bit P activado, una secuencia de cambio de dirección y después una secuencia de proseguir. Una vez que la primaria comenzó a transmitir unos, pasa al modo de recepción.

La primera estación secundaria que sigue en el lazo, que haya recibido una trama dirigida a ella con el bit P activado, cambia el séptimo bit 1 en la secuencia de proseguir por un 0, creando así una bandera. Esta bandera se transforma en la bandera inicial de la trama o tramas de la respuesta de la secundaria. Después de que la secundaria transmita su última trama, se transforma de nuevo en repetidora para los unos de línea inactiva de la primaria. Estos unos de línea inactiva son la secuencia de proseguir para la siguiente estación secundaria. La estación secundaria que sigue en el lazo, que haya recibido una trama dirigida a ella con el bit P activado, detecta la secuencia de cambio de dirección, todas las tramas transmitidas por las secundarias anteriores en el lazo y a continuación la secuencia de proseguir. Cada estación secundaria inserta sus tramas de respuesta de inmediato, después de la última trama repetida. El ciclo se termina cuando la primaria recibe su propia secuencia de cambio de dirección, una serie de tramas de respuesta y, a continuación, la secuencia de proseguir.

Orden/respuesta de configuración. La orden y respuesta de configuración (CFGR) es no numerada que sólo se usa en configuración de lazo. CFGR contiene un *descriptor de función* de un byte (en esencia, una suborden) en el campo de información. Una orden CFGR se reconoce con una respuesta CFGR. Si el bit menos significativo del descriptor de la función está activado, se inicia una función especificada. Si está restablecido, o desactivado, se borra la función especificada. Hay seis subórdenes que pueden aparecer en el campo de función de la orden configurar.

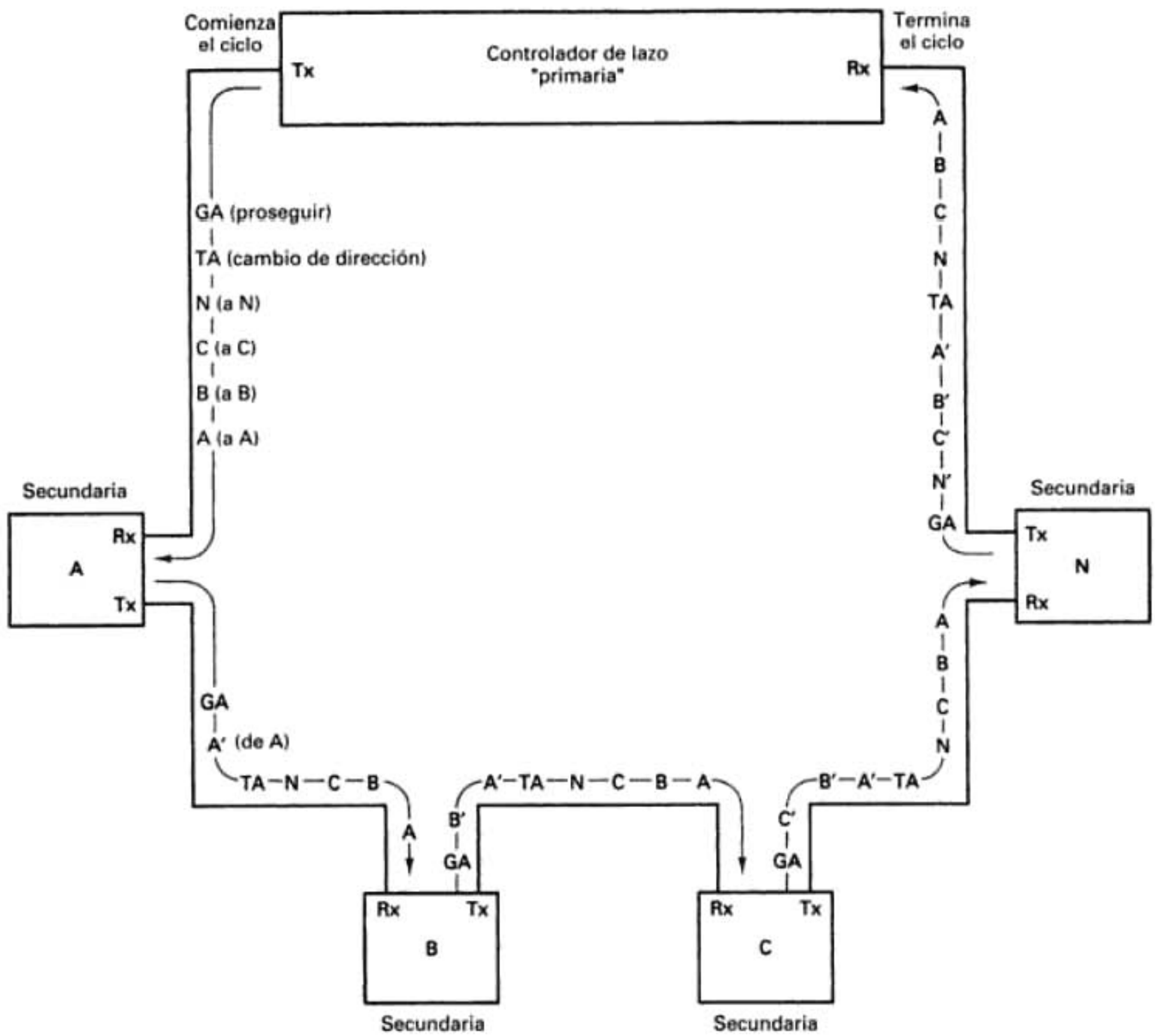


FIGURA 14-4 Configuración de lazo SDLC

1. *Clear (borrar)—00000000.* Una suborden borrar hace que la secundaria borre todas las funciones activadas con anterioridad. La respuesta de la secundaria a una suborden clear es otra suborden clear, 00000000.

2. *Prueba de portadora (BCN, de beacon)—0000000X.* La prueba de portadora hace que la secundaria que reciba esa orden encienda o apague su portadora. Si el bit X está activo, la secundaria suprime la transmisión de la portadora. Si el bit X está desactivado, la secundaria reasume la transmisión de la portadora. La prueba de la portadora se usa para aislar un problema de lazo abierto. También, siempre que una secundaria detecta la pérdida de una portadora de recepción, comienza en forma automática a transmitir su respuesta de portadora. La secundaria continuará transmitiendo la portadora hasta que el lazo reasuma su estado normal.

3. *Modo de monitor—0000010X.* La orden monitor hace que la secundaria a quien va dirigida se ponga en modo monitor (sólo recepción). Una vez en el modo monitor, una secundaria no puede transmitir sino hasta que reciba una suborden de borrar modo monitor (00000100), o una suborden borrar (00000000).

4. *Salida de línea—0000100X.* La orden salir de línea hace que la estación secundaria pase sus transmisiones en forma directa a su entrada de recepción. Esta orden pone de hecho fuera de línea a la secundaria, durante la prueba. Una estación secundaria no manda a la primaria los resultados de una prueba de salida de línea.

5. *Autoprueba—0000101X.* La orden de autoprueba hace que la secundaria a la que va dirigida inicie una serie de pruebas internas de diagnóstico. Cuando se terminan las pruebas, la secundaria responde. Si el bit P está activado en la orden configurar, la secundaria responderá después de terminar la autoprueba, en la primera oportunidad. Si el bit P está restablecido, la secundaria responderá después de la prueba en la siguiente trama del tipo de interrogación que reciba. Todas las demás transmisiones se pasan por alto por parte de la secundaria, mientras está haciendo las autopruebas. La secundaria indica los resultados de ella activando o restableciendo el bit menos significativo (X) de su respuesta de autoprueba. Un 1 indica que las pruebas no tuvieron éxito, y un 0 indica que sí tuvieron éxito.

6. *Prueba modificada de enlace—0000110X.* Si está activada la función de prueba modificada de enlace (está puesto el bit X), la estación secundaria responderá a una orden TEST con una respuesta TEST que tiene un campo de información que contiene el primer byte del campo de información de la orden TEST repetido *n* veces. El número *n* depende de la implementación del sistema. Si se restablece el bit X, la estación secundaria responderá a una orden TEST, con o sin campo de información, con una respuesta TEST con campo de información de longitud cero. La prueba modificada de enlace es una suborden opcional y sólo se usa para contar con una forma alternativa de prueba de enlace a la que se describió antes, para la orden TEST.

Transparencia

El mecanismo de transparencia que se usa en SDLC se llama *inserción de bit cero* o *inserción de cero*. La secuencia (01111110) de bits de bandera puede presentarse en una trama en donde no se pretende tener una bandera. Por ejemplo, siempre que haya 7EH en el campo de dirección, control, información o FCS se interpretaría como una bandera e interrumpiría la sincronización de caracteres. Por consiguiente, se debe prohibir la ocurrencia de un 7EH, excepto cuando se deba usar como una bandera. Para evitar que se presente la secuencia 7EH, se introduce en forma automática un cero después de cada presentación de cinco unos consecutivos, excepto en una secuencia de bandera designada (es decir, no se insertan ceros en las banderas). Cuando se reciben cinco unos consecutivos y el siguiente bit es un 0, se elimina el cero. Si el siguiente bit es un 1, debe ser una bandera válida. A continuación se ve un ejemplo de inserción y eliminación de cero.

Bits en la trama original en la estación transmisora

01111110	01101111	11010011	1110001100110101	01111110
bandera	dirección	control	FCS	bandera

Después de la inserción de cero, pero antes de la transmisión

01111110	01101111	101010011	1110000110011010 1	01111110
bandera	dirección	control	FCS	bandera

ceros insertados

Después de eliminar el cero en el receptor

01111110	01101111	11010011	1110001100110101	01111110
bandera	dirección	control	FCS	bandera

Abortar mensaje

La orden *abortar mensaje* es para terminar una trama en forma prematura. En general, sólo se hace para acomodar mensajes de alta prioridad, como por ejemplo, los procedimientos de recuperación de emergencia de enlace, etcétera. Aborto de mensaje es toda ocurrencia de 7 a 14 unos consecutivos. En una secuencia de aborto no se insertan ceros. Un aborto de mensaje termina una trama presente, y de inmediato comienza la trama de mayor prioridad. Si hay más de 14 unos consecutivos, se considera como condición de línea inactiva. Por consiguiente, con 15 o más unos consecutivos el circuito se pone en estado inactivo.

Codificación por inversión en cero

Una transmisión síncrona binaria, como SDLC está sincronizada respecto al tiempo para permitir la identificación de dígitos binarios secuenciales. En comunicaciones síncronas de datos se

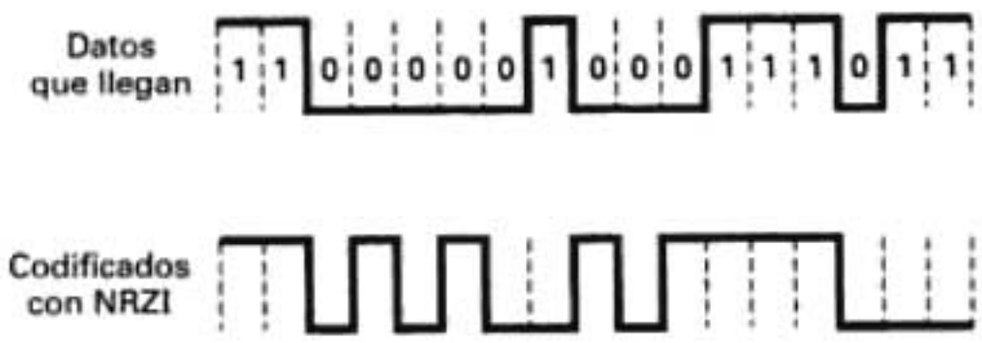


FIGURA 14-5 Codificación NRZI

supone que el DCE o el DTE proporcionan la sincronización de bits. En las transmisiones síncronas, un receptor muestrea los datos que llegan, a la misma velocidad con la que se transmitieron. Aunque pueden existir variaciones pequeñas de sincronización, los módems síncronos proporcionan la recuperación del reloj de datos y la sincronización ajustada dinámicamente con muestras, para mantener los tiempos de muestra a medio intervalo entre bits. Para que un DTE o un DCE recuperen el reloj, es necesario que en los datos haya transiciones. La *codificación por inversión en cero*, es un esquema que garantiza al menos una transición en el campo de datos por cada siete bits transmitidos. También, a la codificación por inversión en cero se le llama NRZI (de *nonreturn-to-zero inverted*, no regreso a cero invertida).

En la codificación NRZI, los datos se codifican en el transmisor y a continuación se decodifican en el receptor. La fig. 14-5 muestra un ejemplo de esta codificación. La forma de onda codificada no cambia con los unos en el codificador NRZI. Sin embargo, los ceros invierten el nivel de la transmisión codificada. En consecuencia, los ceros consecutivos generan una secuencia de alto y bajo alternativos. En SDLC no puede haber más de seis unos en sucesión (una bandera). Por consiguiente, se garantiza que haya una transición de alto a bajo al menos una vez por cada siete bits transmitidos, excepto durante un aborto de mensaje o una condición de línea inactiva. En un decodificador NRZI, siempre que hay una transición de alto a bajo en los datos recibidos, se genera un 0. La ausencia de una transición simplemente genera un 1. En la fig. 14-5 se supone nivel alto antes de codificar los datos que entran.

Se pretendió que la codificación NRZI se usara con módems asíncronos que no tienen funciones de recuperación de reloj. En consecuencia, el DTE debe proporcionar la sincronización, lo cual se promueve usando datos codificados con NRZI. Los módems síncronos tienen desorganizador y organizador incorporados, que aseguran que haya transiciones en los datos y así es innecesaria la codificación NRZI. El codificador o decodificador NRZI está entre el DTE y el DCE.

Control de alto nivel para enlace de datos

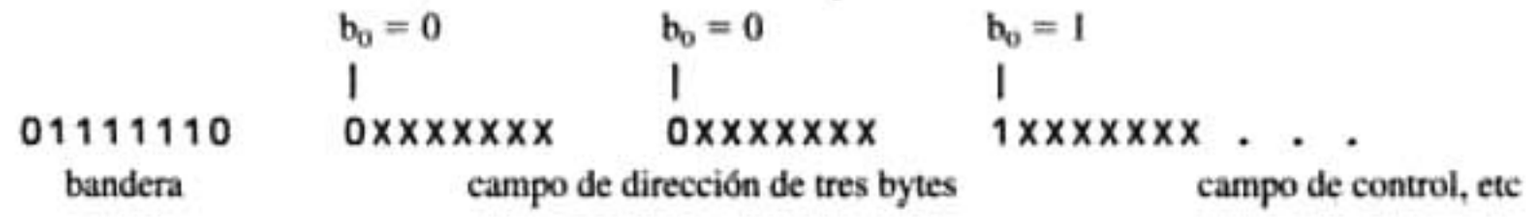
En 1975, la Organización Internacional de Normas (ISO) definió varios conjuntos de subnormas que, cuando se combinan, se llaman *control de alto nivel para enlace de datos* (HDLC, por *high-level data link control*). Como HDLC es un superconjunto de SDLC, sólo se explicarán las funciones adicionales.

HDLC abarca tres normas (subdivisiones) que, cuando se combinan, describen la estructura de trama, las normas de control y la clase de operación para un control de enlace de datos orientado a bits (DLC, de *data link control*).

ISO 3309-1976(E). Esta norma define la estructura, secuencia delimitadora y mecanismo de transparencia de trama que se usa en HDLC. En esencia, son los mismos que en SDLC, pero en HDLC tiene mayores posibilidades de direccionamiento, y comprueba la secuencia de verificación de trama (FCS) en una forma un poco distinta. La secuencia delimitadora que se usa con HDLC es idéntica a la de SDLC: es 01111110.

HDLC puede usar un campo de dirección *básico* de ocho bits, o un formato *extendido* de direccionamiento. En el direccionamiento ampliado, el campo de dirección se puede aumentar en forma recursiva. Si en el byte de dirección b_0 es un 1 lógico, los siete bits restantes son la dirección del secundario (ISO define el bits menos significativo como b_0, mientras que para SDLC es el bit más significativo). Si b_0 es un 0 lógico, el siguiente bit también es parte de la dirección. Si b_0 en el segundo byte es un 0, sigue un tercer byte de dirección y, así sucesivamente, hasta que se encuentre un byte de dirección con un 1 lógico en el bit menos significativo. En esencia, hay siete bits no disponibles en cada byte de dirección para codificar dirección. A continuación se

muestra un ejemplo de un esquema de direccionamiento ampliado, con tres bytes, con b_0 en los primeros dos bytes de dirección igual a cero, indicando que siguen más bytes de dirección, y b_0 en el tercer bit de dirección es 1, el cual termina el campo de dirección.



En HDLC se usa CRC-16 con un polinomio generador especificado por CCITT V.41. En la estación transmisora se calcula la CRC de tal modo que se incluya en el cálculo de la FCS en el receptor y el resto, cuando la transmisión no tiene errores, es F0BBH.

ISO 4334-1979(E). Esta norma Combina a las normas anteriores 6159(E) (desbalanceado) y 6256(E) (balanceado) y describe la clase de operación necesaria para establecer el protocolo de nivel de enlace.

Campo de control. Con HDLC, se puede ampliar el campo de control a 16 bits. Siete bits son para ns y 7 para nr. En consecuencia, con el formato de control extendido puede haber un máximo de 127 tramas no confirmadas en determinado momento.

Campo de información. HDLC permite cualquier cantidad de bits en el campo de información de un comando o respuesta de información (SDLC se limita a bytes de ocho bits). Con HDLC se puede usar cualquier cantidad de bits para un carácter en el campo I, siempre que todos los carateres tengan la misma cantidad de bits.

Formato de supervisión. Con HDLC, el formato de supervisión incluye una cuarta condición de estado: rechazo selectivo (SREJ, de *selective reject*). SREJ se identifica por un 11 en las posiciones de bit b_4 y b_5 de un campo de control de supervisión. Con un SREJ, se puede rechazar una sola trama. Un SREJ pide la retransmisión de sólo la trama identificada por nr, mientras un REJ pide la retransmisión de todas las tramas que comiencen con nr. Por ejemplo, el primario manda I tramas ns = 2, 3, 4 y 5. Se recibió la trama 3 con error. Un REJ pediría la retransmisión de las tramas 3, 4 y 5; un SREJ sólo pediría la retransmisión de la trama 3. Se puede usar SREJ para pedir la retransmisión de cualquier cantidad de tramas, pero sólo se identifica una por una.

Modos de operación. HDLC tiene dos modos de operación que no se especifican en SDLC: modo de respuesta asíncrona y modo de desconexión asíncrona:

1. *Modo de respuesta asíncrona (ARM).* Con el modo de respuesta asíncrona, se permite que las estaciones secundarias manden respuestas no solicitadas. Para transmitir, una secundaria no necesita haber recibido una trama de la primaria con el bit P puesto. Sin embargo, si una secundaria recibe una trama con el bit P puesto, debe responder con una trama con el bit F puesto.
2. *Modo de desconexión asíncrona (ADM).* Un ADM es idéntico al modo de desconexión normal, a excepción de que la secundaria puede iniciar una respuesta DM o RIM en cualquier momento.

ISO 7809–1985(E). Esta norma combina las anteriores 6159(E) (desbalanceado) y 6256(E) (balanceado) y describe la clase de operación necesaria para establecer el protocolo de nivel de enlace.

Operación desbalanceada. Esta clase de operación es lógicamente equivalente a un circuito multipunto de línea privada con ambiente de interrogación. Hay una sola estación primaria responsable del control central de la red. La transmisión de datos puede ser semidúplex o dúplex.

Operación balanceada. Esta clase de operación equivale lógicamente a un circuito de dos puntos de línea privada. Cada estación tiene iguales responsabilidades de enlace de datos, y el acceso al canal es a través de contención, con el modo asíncrono de respuesta. La transmisión de datos puede ser semidúplex o dúplex.

RED PÚBLICA DE DATOS

Una *red pública de datos* (PDN, de *public data network*) es una red conmutada de comunicación de datos, parecida a la red telefónica pública, excepto que una PDN está diseñada sólo para transferir datos. Las redes públicas de datos combinan los conceptos tanto de *redes con valor agregado* (VAN, de *value-added network*) y *redes de conmutación de paquetes*.

Red con valor agregado

Una red con valor agregado "agrega valor a los servicios o instalaciones proporcionados por una portadora común, para proporcionar nuevas clases de servicios de comunicaciones. Como ejemplo de valores agregados están el control de errores, mayor confiabilidad de conexión, direccionamiento dinámico, protección contra falla, multiplexado lógico y conversiones de formato de datos. Una VAN comprende una organización que renta las líneas de comunicaciones de portadoras comunes, como AT&T y MCI, y agrega nuevos servicios de comunicaciones a esas líneas. Como ejemplos de redes con valor agregado están GTE Telnet, DATAPAC, TRANSPAC y Tymnet Inc.

Red de conmutación de paquetes

La conmutación de paquetes implica dividir los mensajes de datos en pequeños grupos de información y transmitirlos a través de redes de comunicación a sus destinos con conmutadores controlados por computadoras. En las redes públicas para datos se usan tres técnicas normales: *conmutación de circuitos*, *conmutación de mensajes* y *conmutación de paquetes*.

Conmutación de circuitos. La conmutación de circuitos se usa para hacer una llamada telefónica normal en la red telefónica pública. Se establece la llamada, se transfiere la información y a continuación se desconecta la llamada. El tiempo necesario para establecer la llamada se llama tiempo *de preparación*. Una vez establecida la llamada, los circuitos interconectados por las estaciones de la red se asignan a un solo usuario durante la llamada. Después de haber establecido ésta, la información se transfiere en *tiempo real*. Cuando termina la llamada, los circuitos y las estaciones están disponibles de nuevo para otro usuario. Como hay una cantidad limitada de circuitos y rutas de conmutación disponibles, puede presentarse un *bloqueo*, que es la incapacidad de completar una llamada, por no haber instalaciones o rutas de estaciones disponibles entre la fuente y el destino. Cuando se usa el circuito para transferencia de datos, el equipo terminal en la fuente y el destino debe ser compatible; los equipos deben usar módems compatibles, con la misma frecuencia de bits, conjunto de caracteres y protocolo.

Un conmutador de circuito es *transparente*. El conmutador es transparente a los datos; no hace más que interconectar los equipos terminales de fuente y destino. Un conmutador o estación de circuito no agrega valor al mismo.

Conmutación de mensajes. La conmutación de mensajes es una forma de *sistema/red de almacenar y enviar*. Los datos, incluyendo los códigos de identificación de fuente y destino, se transmiten a la red y se almacenan en una estación. Cada estación de la red tiene posibilidades de almacenamiento de mensajes. La red transfiere los datos de una estación a otra cuando es conveniente hacerlo. En consecuencia, los datos no se transfieren en tiempo real, puede haber un retardo en cada estación. En la conmutación de mensajes no puede haber bloqueo. Sin embargo, el tiempo de retardo desde la transmisión hasta la recepción del mensaje varía de una a otra llamada, y puede ser muy largo (hasta quizá unas 24 horas). En la conmutación de mensajes, una vez que la información ha entrado a la red, se convierte a un formato más adecuado para su transmisión. En la terminal de recepción, los datos se convierten a un formato compatible con el equipo de la terminal receptora de datos. Por lo anterior, en la conmutación de mensajes, no necesitan ser compatibles los equipos de las terminales de datos de fuente y destino. La conmutación de mensajes es más eficiente que la conmutación de circuitos, porque los datos que entran a la red durante las horas pico se pueden guardar para transmitirlos después, cuando ha disminuido la carga.

Una conmutación de mensaje es *transaccional*, porque no hace más que tan sólo transferir los datos de la fuente al destino. Una conmutación de mensajes puede guardar datos o cambiar

TABLA 14-3 Resumen de las técnicas de conmutación

Conmutación de circuitos	Conmutación de mensajes	Conmutación de paquetes
Ruta dedicada de transmisión	Ruta no dedicada de transmisión	Ruta no dedicada de transmisión
Transmisión continua de datos	Transmisión de mensajes	Transmisión de paquetes
Trabaja en tiempo real	No en tiempo real	Casi en tiempo real
No se guardan los mensajes	Se guardan los mensajes	Se detienen los mensajes corto tiempo
Se establece ruta para todo el mensaje	Se establece ruta para cada mensaje	Se establece ruta para cada paquete
Retardo de preparación de llamada	Retardo en transmisión de mensaje	Retardo en transmisión de paquete
Señal de ocupado si está ocupada la parte llamada	No hay señal de ocupado	No hay señal de ocupado
Puede suceder un bloqueo	No puede suceder un bloqueo	No puede suceder un bloqueo
Usuario responsable por protección contra pérdida de mensaje	Red responsable por mensajes perdidos	La red puede ser responsable por cada paquete, pero no por todo el mensaje
Sin conversión de velocidad o de código	Conversión de velocidad y código	Conversión de velocidad y código
Transmisión con ancho de banda fijo (es decir, capacidad fija de información)	Uso dinámico del ancho de banda	Uso dinámico del ancho de banda
Sin bits de indirectos después del retardo inicial de preparación	Bits de indirectos en cada mensaje	Bits de indirectos en cada paquete

su formato y frecuencia de bit, para a continuación regresar los datos a su forma original o a otra completamente distinta, en el receptor. La conmutación de mensajes multiplexa datos de distintas fuentes a una instalación común.

Conmutación de paquetes. En la conmutación de paquetes, los datos se dividen en segmentos más pequeños, llamados *paquetes*, antes de transmitirlos por la red. Como un paquete se puede guardar en la memoria en una estación durante un corto tiempo, a veces se llama a la conmutación de paquetes red *para retener y enviar*. En este método, un mensaje se divide en paquetes, y cada paquete puede tomar distinto camino por la red. En consecuencia, todos los paquetes no llegan necesariamente al receptor al mismo tiempo o en el mismo orden con los que se transmitieron. Como los paquetes son pequeños, el tiempo de retención suele ser bastante corto, y la transferencia de mensajes es casi en tiempo real, y no puede presentarse el bloqueo. Sin embargo, las redes de conmutación de paquetes necesitan arreglos complicados y costosos, y protocolos complicados. Una estación de conmutación de paquetes también es un conmutador transaccional. En la tabla 14-3 se resumen las técnicas de conmutación de circuitos, mensajes y paquetes.

Clase de servicio CCITT X.1 para usuario internacional

La norma CCITT X.1 divide las diversas clases de servicio en tres modos básicos de transmisión para una red pública de datos. Los tres modos son: *arranque/paro, síncrono* y *de paquetes*.

Modo de arranque y paro. En este modo, los datos se transfieren de la fuente a la red, y de la red al destino, en formato de datos asíncronos, es decir, cada carácter tiene su trama dentro de un bit de arranque y uno de paro. La señalización de control de llamada se hace en el alfabeto internacional No. 5 (ASCII-77). Dos protocolos que se usan con frecuencia en transmisión de arranque y paro son el 83B de IBM y el 8A 1/B 1 de AT&T, de arreglo de llamada selectiva.

Modo síncrono. En el modo síncrono, los datos pasan de la fuente a la red y de la red al destino en formato de datos síncronos, es decir, cada mensaje está precedido por un carácter único de sincronización. La señalización de control de llamada es idéntica a la que se usa en los circuitos de datos de línea privada, y los protocolos de uso frecuente para esa transmisión síncrona son el 3270 bisync de IBM, BASIC de Burrough's y UNISCOPE de UNIVAC.

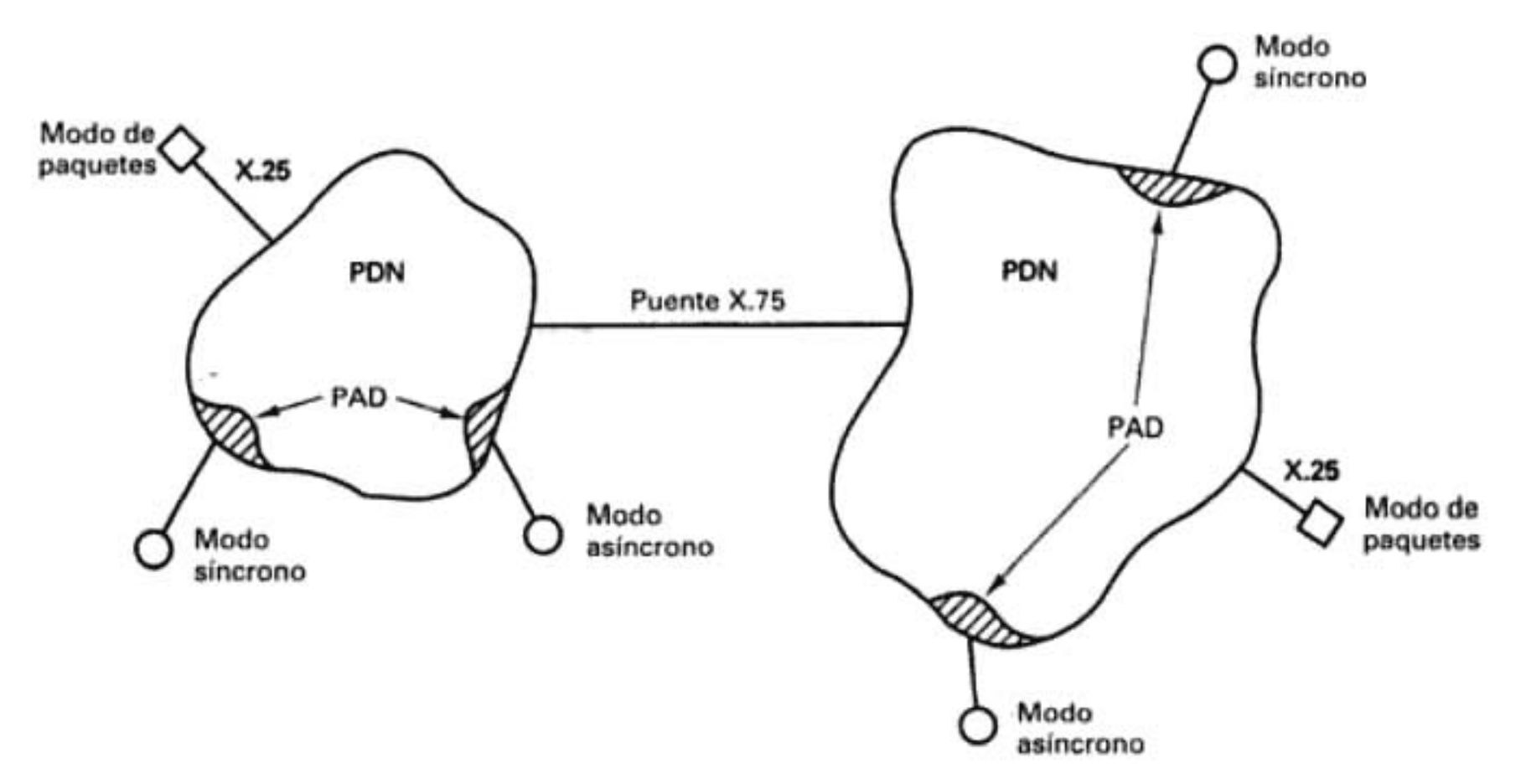


FIGURA 14-6 Red pública de datos

Modo de paquetes. En el modo de paquetes, los datos se transfieren de la fuente a la red y de la red al destino en un formato de trama. El protocolo patrón que se usa en el modo de paquetes es el formato de trama HDLC de ISO. Dentro de la red, los datos se dividen en paquetes más pequeños y se transfieren de acuerdo con el protocolo de interfaz de usuario a red CCITT X.25.

La fig. 14-6 ilustra un arreglo característico de una red pública de datos, que muestra cada uno de los tres modos de operación. El ensamblador/desensamblador de paquetes (PAD) es la interfaz del usuario con el formato X.25, cuando los datos del usuario están en modo asíncrono o síncrono. Es innecesario un PAD cuando el usuario opera en el modo de paquetes. La CCITT recomienda X.75 como protocolo de puente. Los puentes se usan para interconectar dos redes públicas de datos.

PROTOCOLO DE USUARIO Y RED CCITT X.25

En 1976 la CCITT designó a la interfaz X.25 con el usuario como norma internacional para acceso de paquetes a red. Téngase en cuenta que la X.25 es estrictamente una interfaz de *usuario a red*, y sólo se refiere a las capas física, de enlace de datos y de red en el modelo ISO de siete capas. La X.25 usa las normas existentes siempre que sea posible. Por ejemplo, expecifica las normas X.21, X.26 y X.27 como de interfaz física, que corresponden a las RS-232, RS-423A y RS-422A de EIA, respectivamente. En X.25 se define que HDLC es la norma internacional para la capa de enlace de datos, y que la norma en Estados Unidos es la 3.66 de ANSI (American National Standards Institute), *Advanced Datta Communications Control Procedures* o ADCCP. Fueron designadas la HDLC de ISO y la 3.66 de ANSI para circuitos de datos de línea privada con ambiente de interrogación. En consecuencia, los procedimientos de direccionamiento y de control que describen no son adecuados para redes de paquetes de datos. Se seleccionaron la 3.66 de ANSI y la HDLC para la capa de enlace de datos por su formato de trama, secuencia delimitadora, mecanismo de transparencia y método de detección de errores.

Al nivel de enlace, el protocolo especificado por la X.25 es un subconjunto de HDLC, llamado *Procedimiento balanceado de acceso a enlace* (LAPB, de *Link Access Procedure Balanced*). LAPB permite comunicaciones dúplex en dos sentidos entre el DTE y el DCE en el puente de paquetes de la red. Sólo puede aparecer la dirección del DTE o el DCE en el campo de dirección de la trama LAPB. El campo de dirección se refiere a una dirección de enlace y no a una de red. La dirección de red, de la terminal destino, está incrustada en el encabezado del paquete, que es parte del campo de información.

TABLA 14-4 Órdenes LAPB

	Número del bit			
Nombre de la orden	8 7 6	5	4 3 2	1
I (información)	nr	P	ns	0
RR (receptor listo)	nr	P	0 0 0	1
RNR (receptor no listo)	nr	P	0 1 0	1
REJ (rechazo)	nr	P	1 0 0	1
SABM (establecer modo asíncrono balanceado)	0 0 1	P	1 1 1	1
DISC (desconectar)	0 1 0	P	0 0 1	1

TABLA 14-5 Respuestas LAPB

	Número del bit			
Nombre de la orden	8 7 6	5	4 3 2	1
RR (receptor listo)	nr	F	0 0 0	1
RNR (receptor no listo)	nr	F	0 1 0	1
REJ (rechazo)	nr	F	1 0 0	1
UA (reconocimiento no numerado)	0 1 1	F	0 0 1	1
DM (modo de desconexión)	0 0 0	F	1 1 1	1
FRMR (trama rechazada)	1 0 0	F	0 1 1	1

Las tablas 14-4 y 14-5 muestran los comandos y respuestas, respectivamente, para una trama LAPB. Durante el funcionamiento de LAPB, la mayoría de las tramas son órdenes. Una trama de respuesta sólo se pide cuando se recibe una trama de orden que contiene una interrogación (bit P) = 1. Las letras SABM/UA representan un par de orden y respuesta que se usa para inicializar todos los contadores y relojes al principio de una sesión. De igual modo, DISC/DM es un par de orden/respuesta que se usa al final de una sesión. FRMR es una respuesta a cualquier orden ilegal para la que no hay indicación de errores de transmisión, según el campo de secuencia de verificación de trama.

Las órdenes de información (órdenes I) se usan para transmitir paquetes. Los paquetes nunca se mandan como respuestas. Se reconocen los paquetes usando nm y nr así como se hace en SDLC. Una estación manda RR cuando necesita responder (reconocer) algo, pero no tiene paquetes de información por mandar. Una respuesta a una orden de información podría ser RR con F = 1. Este procedimiento se llama *de punto de comprobación*.

Otra forma de pedir transmisión de tramas es con REJ. Para que el flujo de control indique una condición de ocupado se usa RNR, y evita toda transmisión hasta que se borra con un RR.

La capa de red de X.25 especifica tres servicios de conmutación por ofrecer en una red conmutada de datos: circuito virtual permanente, llamada virtual y diagrama de datos.

Circuito virtual permanente

Un *circuito virtual permanente* (PVC, de *permanent virtual circuit*) equivale lógicamente a un circuito de dos puntos, de línea privada dedicada, pero es más lento. Un PVC es más lento porque no se proporciona una conexión permanente (con conductores) de terminal a terminal. La primera vez que se pide una conexión, se deben establecer las conexiones y circuitos adecuados por la red para proporcionar la interconexión. Un PVC identifica la ruta entre dos suscriptores determinados de la red que se usa para todos los mensajes que siguen. Con un PVC no son necesarias las direcciones de la fuente y del destinatario, porque los dos usuarios son fijos.

Llamada virtual

Una *llamada virtual* (VC, de *virtual call*) es lógicamente equivalente a una llamada telefónica por la red DDD, excepto que no se hace conexión directa de terminal a terminal. Una llamada virtual

	Bandera	Campo de dirección de enlace	Campo de control de enlace	Identificador de formato	Identificador de canal lógico	Número progresivo de mandar paquete P(s)	0	Número progresivo de recibir paquete P(r)	0	Datos del usuario	Secuencia de verificación de trama	Bandera
Bits:	8	8	8	4	12	3/7	5/1	3/7	5/1	Hasta 1024	16	8

FIGURA 14-8 Formato de transferencia de paquete de datos

Campo de instalaciones. Este campo contiene hasta 512 bits de información opcional sobre instalación de red, como por ejemplo, información inversa de facturación, grupos cerrados de usuario, y si es una conexión símplex de transmisión o de recepción.

Identificador de protocolo. Este campo de 32 bits se reserva para que el suscriptor inserte funciones de protocolo a nivel de usuario, como por ejemplo, procedimientos para entrar y prácticas de identificación de usuario.

Campo de datos de usuario. Se pueden transmitir hasta 96 bits de datos del usuario con un paquete de llamada de petición. Estos son datos sin numerar que no se han confirmado. Por lo general, este campo se utiliza para las contraseñas de usuario.

Paquete de transferencia de datos. La fig. 14–8 muestra el formato de los campos de un paquete de transferencia de datos. Un paquete de transferencia de datos se parece a uno de llamada de petición, pero en el de transferencia de datos hay mucho menos indirectos y puede contener un campo de datos de usuario mucho mayor. El paquete de transferencia de datos contiene un campo de secuencia de paquetes que no se incluyó en el formato de llamada de petición.

Los campos de bandera, dirección de enlace, control de enlace, identificador de formato, LCI y FCS son idénticos a los que se usan en el paquete de llamada de petición. Los campos de secuencia de paquete transmitido y recibido se describen a continuación.

Campo de secuencia de paquete transmitido. El campo P(s) (de *packet send*) se usa en la misma forma que las secuencias nm y nr con SDLC y HDLC. El campo P(s) es análogo al nm, y el P(r) es análogo al nr. Cada paquete sucesivo de transferencia de datos se asigna al siguiente número progresivo P(s). Éste puede ser un número binario de 3 o de 7 bits y en consecuencia, numera a los paquetes del 0 al 7 o del 0 al 127. La secuencia de numeración se identifica en el identificador de formato. El paquete de transmitir siempre contiene ocho bits, y los que no se usan se restablecen.

Campo de secuencia de paquete recibido. P(r) se usa para confirmar los paquetes recibidos, y pedir la retransmisión de los que se hayan recibido con errores (ARQ). El campo I en un paquete de transferencia de datos puede contener bastante más información de la fuente que un campo I en un paquete de llamada de petición.

Las normas recomendadas de la serie X

La X.25 es parte de la serie X de normas recomendadas por CCITT, para redes públicas de datos. Esta serie X se divide en dos categorías: de X.1 a X.39, acerca de servicios e instalaciones, terminales e interfaces; y la X.40 a la X.199, acerca de arquitectura, transmisión, señalización, conmutación, mantenimiento y arreglos administrativos de la red. La tabla 14.6 es una lista de las normas X más importantes, con sus títulos y descripciones.

Modo de transferencia asíncrona

El modo de transferencia asíncrona (ATM, de *asynchronous transfer mode*) es una norma de comunicaciones donde se usa una forma de red de alta velocidad de conmutación de paquetes, como medio de transmisión. El ATM fue desarrollado como parte de la *red digital de servicios integrados de banda ancha* (BISDN, de *Broadband Integrated Services Digital Network*), que se describirá más adelante en este capítulo. ATM pretende utilizar la *red óptica síncrona* (SONET, de *Synchronous Optical Network*), que también se describirá después en este capítulo, para adaptarse a las necesidades especializadas de comunicaciones de datos que tienen las redes privadas corporativas. Algunos expertos dicen que puede ser que ATM llegue a reemplazar tanto a los sistemas de portadora digital T1, privados y rentados, como al equipo de conmutación en las instalaciones. En la actualidad, las máquinas electrónicas convencionales de conmutación usan un procesador central para establecer rutas de conmutación, y conducir el tráfico a través de una red. Sin embargo, las estaciones ATM tendrán procedimientos de autodireccionamiento, donde unas *celdas*

TABLA 14-6 Normas CCITT serie X

X.1	Clases de servicio para usuario internacional en redes públicas de datos. Asigna designaciones numéricas de clase a distintas velocidades y tipos de terminal.
X.2	Servicios e instalaciones para usuarios internacionales en redes públicas de datos. Especifica servicios e instalaciones esenciales y adicionales.
X.3	Instalación de ensamble y desensamble de paquetes (PAD) en una red pública de datos. Describe el ensamblador y desensamblador de paquetes, que se usa normalmente en un puente a la red, para permitir conexiones de una terminal de arranque y paro a una red de paquetes.
X.20bis	Uso de DTE en redes públicas de datos, diseñado para interconectar con módems dúplex asíncronos serie V. Permite usar V.24 y V.28 (en esencia igual que la EIA RS-232).
X.21bis	Uso de DTE, en redes públicas de datos, diseñado para interconectarse con módems dúplex síncronos serie V. Permite usar V.24 y V.28 (en esencia igual que EIA RS-232) o V.35.
X.25	Interfaz entre DTE y DCE para terminales que trabajan en modo de paquetes, en redes públicas de datos. Define la arquitectura de tres niveles de protocolos, existentes en el cable de interfaz serie entre una terminal en modo de paquetes y un puente a una red de paquetes.
X.28	Interfaz DTE/DCE para un DTE en modo de arranque y paro que entra a PAD en una red pública de datos situada en el mismo país. Define la arquitectura de los protocolos existentes en un cable de interfaz serie entre una terminal de arranque y paro y un PAD X.3.
X.29	Procedimientos para intercambiar información de control y datos de usuario entre un PAD y un DTE en modo de paquetes, o con otro PAD. Define la arquitectura de los protocolos para PAD X.3, sea entre dos PAD o entre un PAD y una terminal en modo de paquetes al otro lado de la red.
X.75	Procedimientos de control de llamada a terminal y de tránsito, y de sistema de transferencia de datos por circuitos internacionales, entre redes de datos conmutadas por paquetes. Define la arquitectura de los protocolos entre dos redes públicas de paquetes.
X.121	Plan de numeración internacional para redes públicas de datos. Define un plan de numeración que incluye asignaciones de código para cada nación.

individuales que contienen los datos del suscriptor dirigirán su propia ruta por la red de conmutación ATM en tiempo real, usando su propia dirección en lugar de confiar en un proceso externo que establezca la ruta de conmutación. Una celda es un paquete de datos corto y de longitud fija.

En ATM se usan *canales virtuales* (VC, de *virtual channel*) y *rutas virtuales* (VP, de *virtual path*) para conducir las celdas a través de una red. En esencia, un canal virtual es sólo una conexión entre un lugar de fuente y uno de destino, y eso puede implicar el establecimiento de varios enlaces ATM entre centros locales de conmutación. En ATM, todas las comunicaciones se hacen en el canal virtual, que preserva la secuencia de celdas. Por otro lado, una ruta virtual es un grupo de canales virtuales conectados entre dos puntos, que podrían implicar varios enlaces ATM.

Una celda ATM contiene toda la información de red necesaria para mandar celdas individuales de nodo a nodo por una conexión ATM preestablecida. La fig. 14-9 muestra la estructura de celdas ATM, que tiene 53 bytes de longitud, incluidos un campo de encabezado de 5 bytes y uno de información, de 48 bytes. El campo de información consiste sólo de datos del usuario. Las celdas individuales se pueden entremezclar y ser mandadas a su destino por la red de comunicaciones. El campo de encabezado es para fines de conexión de red, y contiene toda la información de dirección y control necesaria para controlar direcciones y flujos.

Campo de dirección ATM. La fig. 14-10 muestra la estructura del campo de dirección ATM, de 5 bytes, que incluye lo siguiente: campo de control de flujo genérico, identificador de trayectoria virtual, identificador de canal virtual, identificador del tipo de contenido, prioridad de pérdida de celda y control de error de encabezado.

Campo de control de flujo genérico (GFC, de *Generic Flow Control Field*). El campo GFC usa los cuatro primeros bits del primer byte en el campo de encabezado. El GFC controla el flujo del tráfico a través de la interfaz del usuario con la red y la entrada a la red.

Identificador de trayectoria virtual (VPI, de *Virtual Path Identifier*) e ***identificador de canal virtual*** (VCI, de *Virtual Channel Identifier*). Los 24 bits que siguen de inmediato al GFC se usan para la dirección ATM.

Identificador de tipo de contenido (PT, de *Payload Type*). Los primeros tres bits de la segunda mitad del byte 4 especifican el tipo de mensaje en la celda. Como hay tres bits, hay ocho

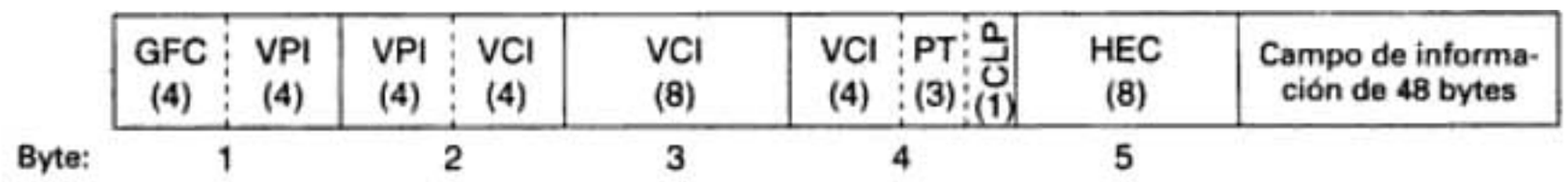


FIGURA 14-9 Estructura del campo de encabezado ATM con cinco bytes

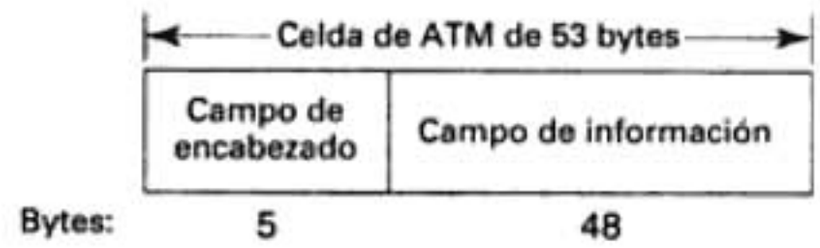


FIGURA 14-10 Estructura de celda de ATM

tipos distintos de contenido posibles. Sin embargo, en la actualidad los tipos 0 a 3 se usan para identificar la clase de datos del usuario, los tipos 4 y 5 indican información administrativa y los tipos 6 y 7 se reservan para usos futuros.

Prioridad de pérdida de celda (CLP, de *Cell Loss Priority*). El último bit del byte 4 es para indicar si una celda es elegible para ser desechada por la red durante periodos de tráfico congestionado. El usuario activa o desactiva el bit CLP. Si está puesto, la red puede desechar la celda durante horas de gran tráfico.

Control de error de encabezado (HEC, de *Header Error Control*). El último byte del campo de encabezado es para control de errores, y se usa para detectar y corregir errores de un bit que sólo sucedan en el campo del encabezado; el HEC no pretende ser un carácter de verificación de toda la celda. El valor que se pone en el HEC se calcula con los cuatro bytes anteriores del campo de encabezado. HEC proporciona algo de protección contra la entrega de celdas a direcciones equivocadas de recepción.

Campo de información ATM. El campo de información de 48 bytes se reserva para datos del usuario. La inserción de datos al campo de información de una celda es una función de la mitad superior de la capa dos en la jerarquía de protocolo ISO-OSI de siete capas. En forma específica, esta capa se llama capa de adaptación ATM (AAL). La AAL permite que ATM tenga la versatilidad necesaria para facilitar, en un solo formato, una gran variedad de distintas clases de servicios, que van desde señales de proceso continuo, como transmisión de voz, hasta mensajes que lleven impulsos de datos muy fragmentados, como los que producen las redes de área local. Como la mayoría de los datos del usuario ocupan más de 48 bytes, la AAL divide la información en grupos de 48 bytes y los pone en una serie de segmentos. Los cinco tipos de AAL son:

1. *Frecuencia constante de bits* (CBR, de *constant bit rate*). Los campos de información CBR son para acomodar el tráfico PCM-TDM, que permite que la red ATM emule los servicios de voz o DSN.
2. *Servicios sensibles a sincronización de frecuencia variable de bits* (VBR, de *variable bit rate*). Esta clase de AAL no está definida en la actualidad; sin embargo, se reserva para servicios de datos en el futuro, que requieran la transferencia de información de reloj entre los puntos terminales, así como de datos (es decir, video de paquetes).
3. *Transferencia de datos VBR orientada a conexión*. Los campos de información tipo 3 transfieren datos VBR, como por ejemplo, los datos impulsivos que se generan a intervalos irregulares entre dos suscriptores en un enlace preestablecido de datos. El enlace de datos se establece con procedimientos de señalización de red que se parecen mucho a los que se usan en la red telefónica pública conmutada. Esta clase de servicios es para transferencias de datos grandes y de gran duración, como por ejemplo, transferencias y respaldos de archivos.
4. *Transferencia de datos VBR sin conexión*. Este tipo de AAL permite la transmisión de datos VBR que no tengan una conexión preestablecida. Los campos de información tipo 4 son para usarse en transmisiones cortas y de muchos impulsos, como por ejemplo, las que se generan en una red de área local.

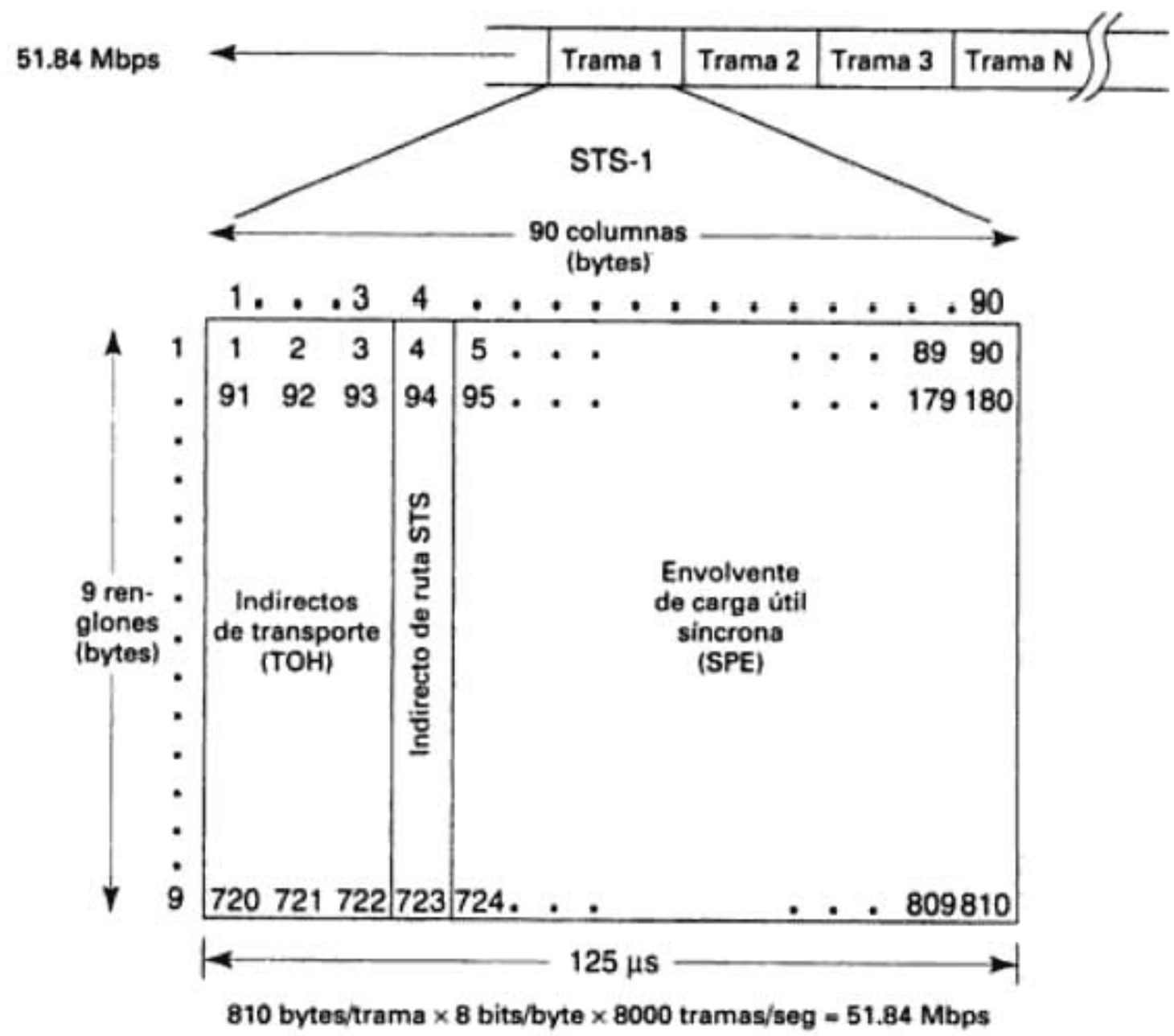


FIGURA 14-11 Formato de trama SONET STS-1

byte por byte, comenzando con el byte 1, continuando de izquierda a derecha y de arriba a abajo, para alcanzar un total de 810 bytes (9 × 90), o de 6480 bits (810 × 8). La frecuencia de transmisión de bits es

$$\text{bps} = \frac{\text{bits}}{\text{tiempo}} = \frac{6480 \text{ bits}}{125\ \mu\text{s}} = 51.84 \text{ Mbps}$$

Las tramas STS-1 se dividen en dos áreas generales: la *indirecta de transporte* (TOH, de *transport overhead*) y la *envolvente de carga útil síncrona* (SPE, de *synchronous payload envelope*). La TOH ocupa las tres primeras columnas de las 90 de la matriz. Las 87 columnas restantes son la SPE. La columna SPE inicial se llama *indirecto de trayectoria* STS, y contiene datos específicos del contenido, y no cambia a medida que SPE avanza por la red. Las 86 columnas restantes en la SPE llevan el contenido, o tráfico de red.

La parte TOH de la trama STS-1 es para vigilancia de alarmas, vigilancia de errores de bits y otros indirectos de comunicación de datos, necesarios para asegurar una transmisión confiable de la envolvente de contenido síncrono entre nodos, dentro de la red síncrona. La SPE tiene por objeto transportar datos del usuario a través de la red síncrona, de la fuente al destino. Se ensambla y desensambla sólo una vez en su ruta por la red, independientemente de cuántas veces cambia los sistemas de transporte. Sin embargo, en la mayoría de los casos, la SPE se ensambla al entrar a la red y se desensambla al salir de ella.

La *carga útil* es el tráfico que produce utilidades, transportado por la red SONET. Una vez ensamblado, la carga útil se puede dirigir a su destino a través de la red. La carga útil STS-1 tiene la capacidad de transportar las diversas clases de señales digitales enlistadas en la tabla 14-8.

Las 86 columnas de la SPE, designadas para llevar la carga útil, se arreglan de acuerdo con normas de mapas, que son una función del servicio de datos que se suscribe a la red (DS1, DS2, etc.). El arreglo específico que se usa para determinado servicio se llama *tributario virtual* (VT, de *virtual tributary*). En SONET se implementan VT de distintos tamaños. Por ejemplo,

TABLA 14-8 Capacidad de carga útil en STS-1

Capacidad	Tipo de señal	Frecuencia de señal (Mbps)	Circuitos de voz	T1
28	DS1	1.544	24	1
21	CEPT1	2.048	30	—
14	DS1C	3.152	48	2
7	DS2	6.312	96	4
1	DS3	44.736	672	28

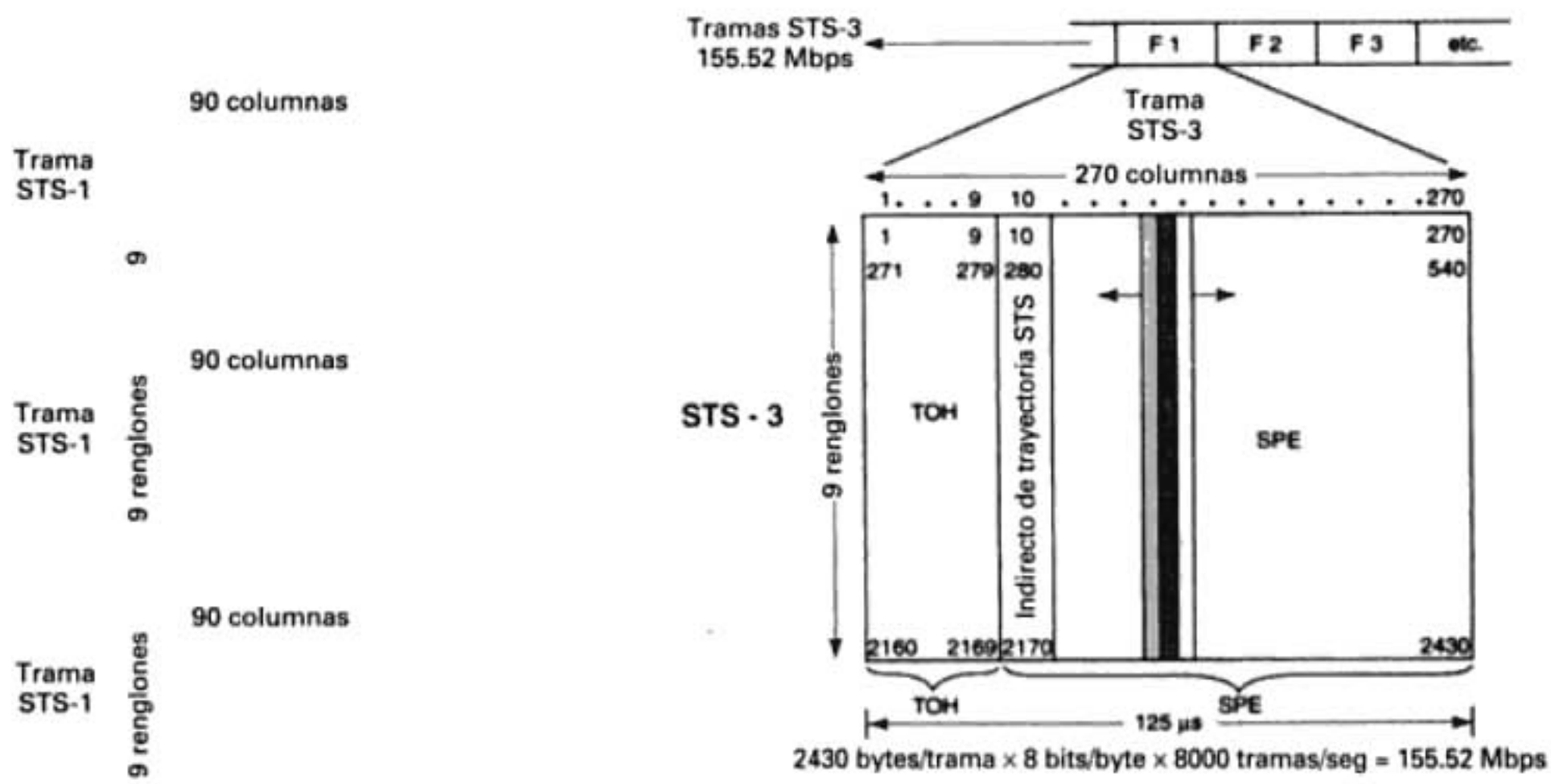


FIGURA 14-12 Multiplexado SONET STS-3 por intercalación de bytes

VT1.5 es una trama formada de 27 bytes formateados en tres columnas de nueve bytes cada una. Con un tiempo de trama de 125 µs, es decir, una frecuencia de 8 kHz, los bytes proporcionan una capacidad de transporte de 1.728 Mbps, que pueden contener el mapa de una señal compatible DS1 de 1.544 Mbps de frecuencia. Se pueden multiplexar 28 VT1.5 en la SPE STS-1.

Multiplexado SONET. En SONET se usa una técnica llamada de multiplexado por intercalación de bytes. Se pueden multiplexar N circuitos STS-1 en un solo circuito STS-*N*. La fig. 14-12 muestra cómo se multiplexan tres tramas en una sola trama STS-3. Esta trama STS-3 tiene 270 columnas y nueve renglones, para tener una capacidad total de 2430 bytes. Como la trama STS-3 también se transmite en 125 µs, la velocidad de transmisión es el triple de una sola trama STS-1, o sean, 155.2 Mbps. Nótese, en la fig. 14-12, que el proceso de intercalación de bytes produce una matriz con renglones y columnas multiplexados. Las primeras nueve columnas de la trama STS-3 comprenden los TOH (indirectos de transporte) de las tres señales STS-1, y las 261 columnas restantes se ocupan con las SPE de las tres señales STS-1.

RED DIGITAL DE SERVICIOS INTEGRADOS (ISDN)

Introducción a la ISDN

La industria de comunicaciones de datos y telefónica cambia en forma continua para satisfacer las demandas de los sistemas contemporáneos de comunicaciones por teléfono, video y compu-

tadora. Hoy son cada vez más las personas que necesitan comunicarse entre sí, que nunca antes. Para satisfacer esas necesidades se actualizan las normas anteriores, y se desarrollan e implementan nuevas, casi diariamente.

La *red digital de servicios integrados* (ISDN, de *Integrated Services Digital Network*) es un diseño de red que proponen las principales compañías telefónicas, en conjunto con el CCITT, para tratar de proporcionar telecomunicaciones mundiales que admitan información de voz, datos, video y facsímil en la misma red. En esencia, ISDN es la integración de una amplia gama de servicios en una sola red de función múltiple. La ISDN es una red que propone interconectar a una cantidad ilimitada de usuarios independientes, a través de una red común de comunicaciones.

Hasta la fecha sólo se ha desarrollado una cantidad limitada de instalaciones ISDN; sin embargo, y también en la actualidad, la industria telefónica está implementando un sistema ISDN tal, que en el futuro cercano los suscriptores entren a él usando las redes actuales públicas de teléfonos y de datos. Los principios básicos y la evolución del ISDN se describieron, por parte del CCITT en su recomendación CCITT 1.120 (1984). En esa recomendación se mencionan los siguientes principios y evolución de la ISDN:

Principios de la ISDN

1. La función principal del concepto ISDN es admitir una amplia variedad de aplicaciones de voz (telefónicas) y no de voz (datos digitales) en la misma red, usando una cantidad limitada de instalaciones normalizadas.

2. Los sistemas ISDN admiten una amplia variedad de aplicaciones, que abarcan conexiones tanto conmutadas como no conmutadas (dedicadas). Entre las conexiones conmutadas están las de circuitos y las de paquetes, y sus concatenaciones.

3. Cuando sea práctico, los nuevos servicios introducidos en una ISDN deben ser compatibles con las conexiones digitales conmutadas de 64 kbps. La conexión digital de 64 kbps es la piedra constructiva básica de la ISDN.

4. Una ISDN contendrá circuitos inteligentes con el fin de proporcionar funciones de servicio, mantenimiento y administración de red. En otras palabras, se espera que la ISDN proporcione servicios más allá de la sola preparación de las llamadas por circuito conmutado.

5. Se debe usar una estructura de protocolo en capas, para especificar los procedimientos de acceso a una ISDN, y para localizar en el modelo de interconexión de sistema abierto (OSI, de *open system interconnection*). Las normas ya desarrolladas para aplicaciones relacionadas con OSI se pueden usar en ISDN, por ejemplo, la X.25 nivel 3, para entrar a los servicios de conmutación de paquetes.

Evolución de las ISDN

1. Las ISDN se basarán en los conceptos desarrollados para las ISDN telefónicas, y podrán evolucionar incorporando, en forma progresiva, otras funciones y propiedades de red, incluyendo las de cualquier otra red dedicada, como por ejemplo conmutación de circuitos y de paquetes para datos, para proporcionar los servicios actuales y nuevos.

2. La transición de una red actual a una ISDN extensa podrá necesitar un periodo que dure una o más décadas. Durante este periodo se deben desarrollar los arreglos para la interconexión en red de las ISDN y los servicios en otras redes.

3. En la evolución hacia una ISDN, se obtendrá conectividad digital de terminal con terminal a través de plantas y equipos que se usan en las redes actuales, como por ejemplo, de transmisión digital, multiplexado por división de tiempo y/o conmutación de multiplexado por división de espacio. Las recomendaciones relevantes actuales, para esos elementos constitutivos de una ISDN, están en la serie correspondiente de recomendaciones de CCITT y CCIR.

4. En las primeras etapas de la evolución de las ISDN, se necesitará adoptar varios arreglos temporales de la red del usuario, en algunos países, para facilitar la penetración temprana de las posibilidades de servicio digital.

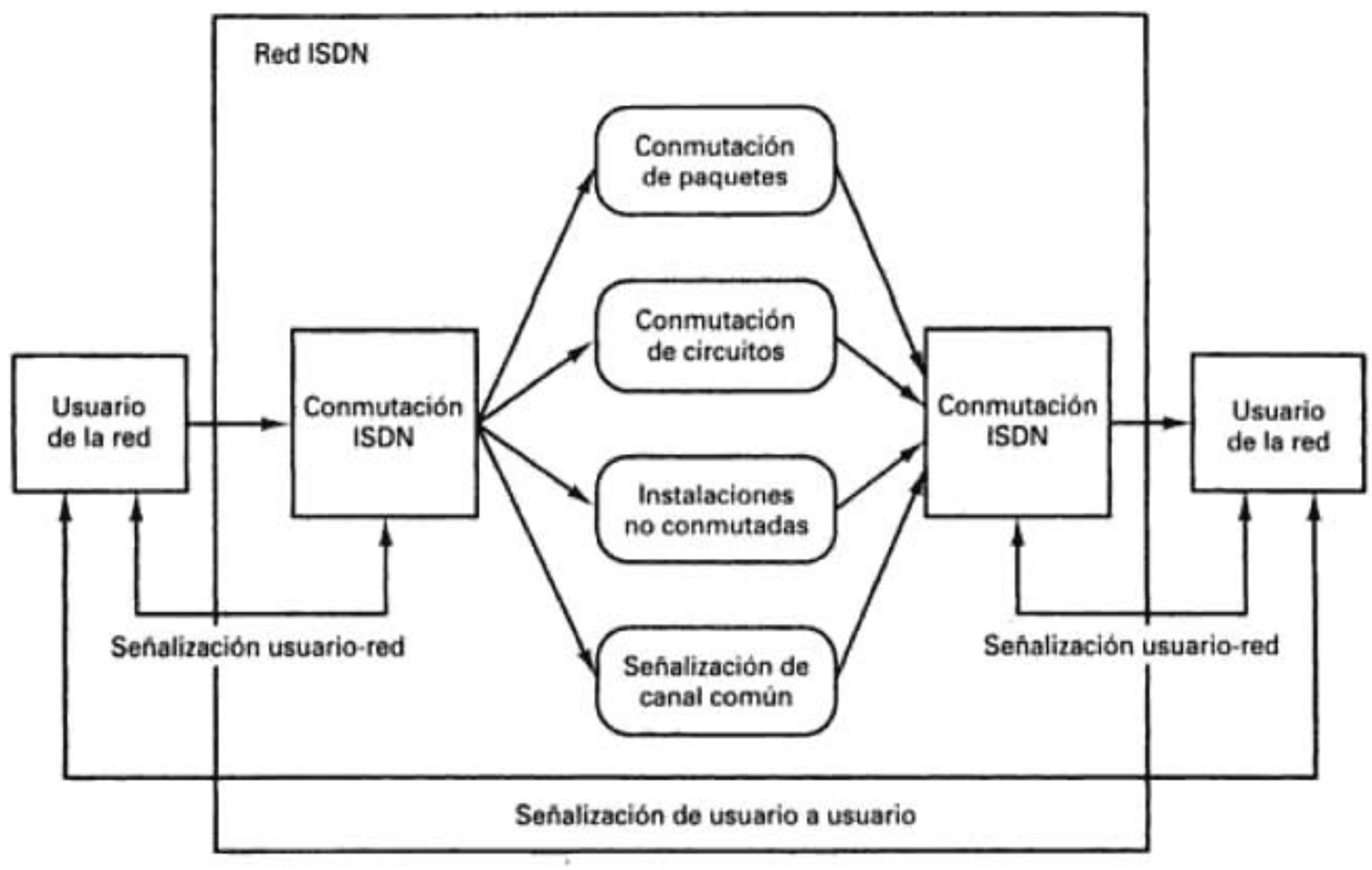


FIGURA 14-13 Arquitectura para las funciones ISDN

5. Una ISDN en evolución también podrá incluir, en etapas posteriores, conexiones conmutadas a frecuencias de bits mayores y menores que 64 kbps.

Arquitectura de la ISDN

En la fig. 14-13 se muestra un diagrama de bloques que indica la arquitectura propuesta para las funciones de ISDN. Ésta está diseñada para admitir una conexión física totalmente nueva para el usuario, un lazo digital de suscriptor y diversos servicios de transmisión.

Se definirá una *interfaz física común* entre los DTE y los DCE. Una sola interfaz se usará en teléfonos, terminales de cómputo y equipo de video. Por consiguiente, se necesitarán varios protocolos para permitir el control de la información que se intercambia entre el dispositivo del usuario y la ISDN. Son los siguientes

Canal B: 64 kbps
Canal D: 16 o 64 kbps
Canal H: 384, 1536 o 1920 kbps

Las normas ISDN especifican que a los usuarios residenciales de la red (es decir, los suscriptores) se les proporcione un *acceso básico*, consistente en tres canales digitales, dúplex, multiplexados por división de tiempo: dos trabajando a 64 kbps (designados canales B de *bearer* o portador) y uno a 16 kbps (canal D, de *datos*). Las frecuencias de bits B y D se seleccionaron para ser compatibles con los sistemas actuales de portadora digital DS1-DS4. El canal D se usa para conducir informacion de señalización y para intercambiar información de control de red. Un canal B se usa para voz codificada digital, y el otro para aplicaciones, como transmisión de datos, voz digitalizada codificada en PCM y videotex. A veces, al servicio 2B + D se le llama *interfaz de frecuencia básica* (BRI, de *basic rate interface*). Los sistemas BRI requieren anchos de banda que puedan contener dos canales B de 64 kbps y un canal D de 16 kbps más bits de trama, sincronización y otros indirectos, para tener una frecuencia total de bits de 192 kbps. Los canales H se usan para proporcionar mayores frecuencias de bits para servicios especiales, como fax rápido, video, datos de alta velocidad y audio de alta calidad.

Hay otro servicio, llamado *servicio primario*, *acceso primario* o *interfaz de frecuencia primaria* (PRI, de *primary rate interface*) que proporcionarán canales múltiples de 64 kbps, para uso de los suscriptores de alto volumen de la red. En Estados Unidos, Canadá, Japón y

Identificador de canal virtual	Carácter de detección de error en encabezado	Indefinido	Datos de canal

FIGURA 14-17 Formato ATM de encabezado de celda

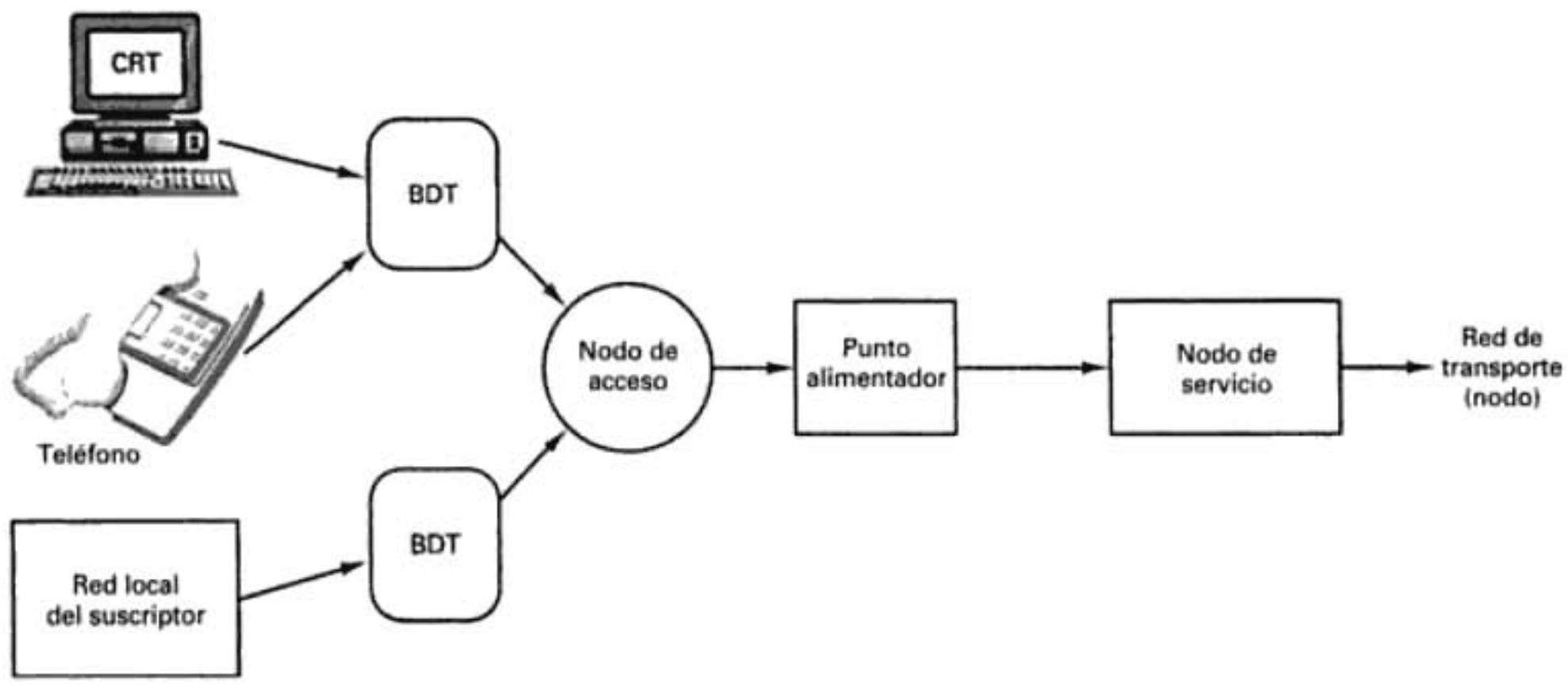


FIGURA 14-18 Acceso a BISDN

un encabezado de celda se ve en la fig. 14-17. El *identificador de canal virtual* indica la fuente de nodo y el destino del paquete. El canal es virtual y no específico, que permite determinar, el direccionamiento físico real del paquete y los tiempos de entrada y salida a la red, de acuerdo con la disponibilidad de la red y los derechos de acceso. Inmediatamente después del identificador del canal virtual está el *carácter de detección de error en encabezado*, que puede ser un CRC o cualquier otra forma de detección de errores. El CRC sólo es para la etiqueta de encabezado, y para el campo de datos se usan otros métodos de detección de errores. La siguiente sección a la celda de encabezado es no identificada, y se reserva para uso futuro.

Configuración BISDN. La fig. 14-18 muestra cómo se logra el acceso a la red BISDN. Cada periférico se interconecta al *nodo de acceso* de una red BISDN a través de una *terminal distante de banda ancha* (BDT, de *broadband distant terminal*). La BDT debe hacer la conversión eléctrica a óptica, el multiplexado de periféricos y el mantenimiento del sistema local del suscriptor. En los nodos de acceso se concentran varias BDT en líneas de fibra óptica de alta velocidad, que van desde un *punto de alimentación* a un *nodo de servicio*. La mayoría de las funciones de control para acceso al sistema, como procesamiento de llamada, funciones administrativas y de conmutación y de mantenimiento, son administradas por el nodo de servicio. Los módulos funcionales se interconectan en configuración de estrella, y pueden ser de conmutación, administrativos, de puente y de mantenimiento. La interconexión de los módulos de función se ve en la fig. 14-19. El cubo central de control funciona como interfaz con el usuario final, para señalización de control y mantenimiento de tráfico de datos. En esencia, supervisa la operación de los módulos.

Las terminales de los suscriptores cercanos a la oficina central pueden omitir por completo los nodos de acceso, y conectarse directamente con la red BISDN, a través de un nodo de servicio. Las redes BISDN que usan cables de fibra óptica pueden usar anchos de banda mucho mayores y, en consecuencia, pueden tener mayores velocidades de transmisión y ofrecer más capacidad de manejo de canales que los sistemas ISDN.

Frecuencias de canal de banda ancha. El CCITT publicó las definiciones preliminares de las nuevas frecuencias de canal de banda ancha, que se agregarán a las actuales, de ISDN de banda angosta. Las nuevas velocidades de canal son:

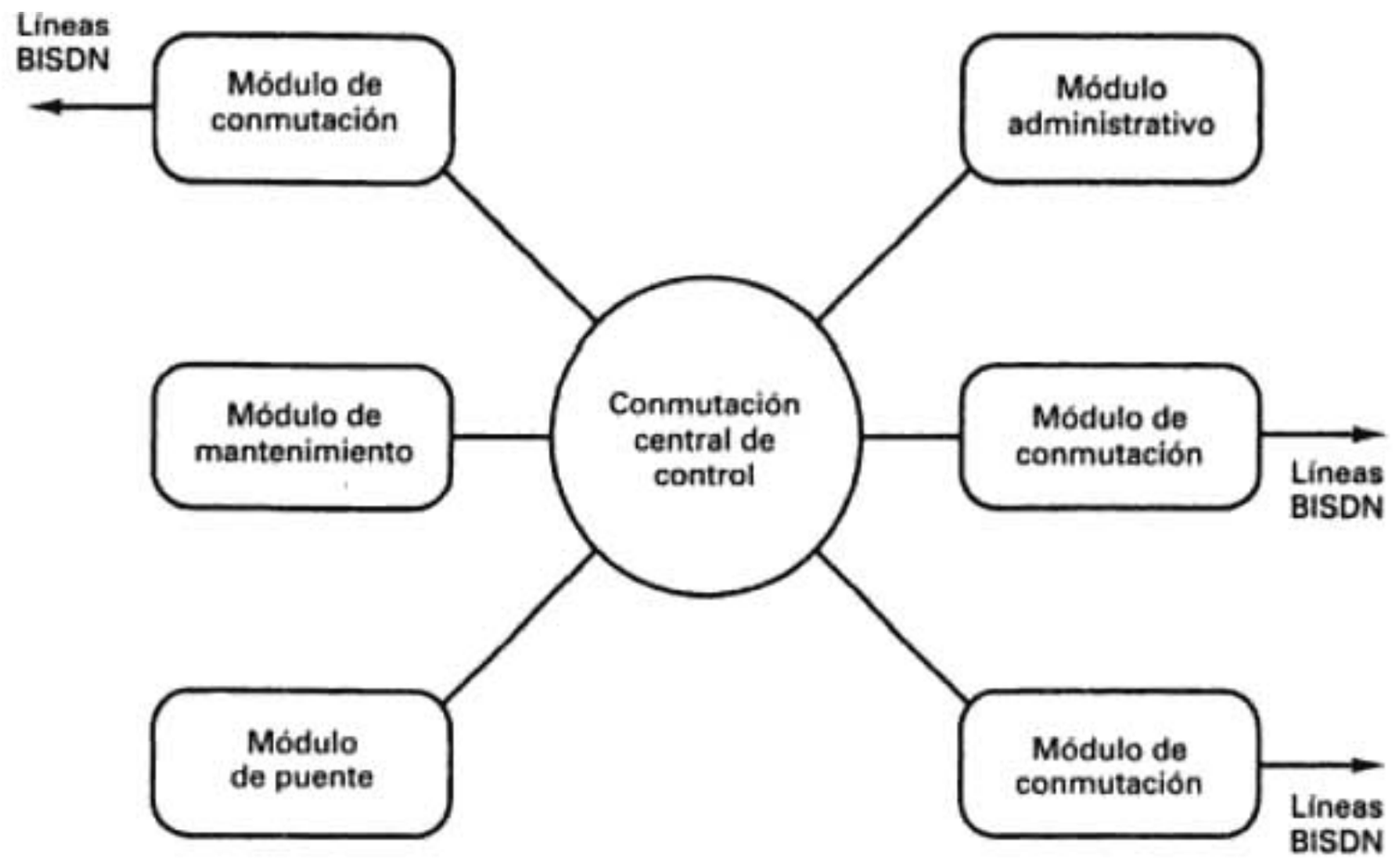


FIGURA 14-19 Interconexiones de módulos funcionales

1. H21: 32.768 Mbps.
2. H22: 43 Mbps a 45 Mbps.
3. H4: 132 Mbps a 138.24 Mbps.

Las velocidades de datos H21 y H22 pretenden usarse en transmisiones de video de movimientos para videoconferencias, videoteléfono y videomensajes. La velocidad de datos H4 es para transferencia de datos a granel, de texto, facsímil e información ampliada de video. La velocidad de datos H21 equivale a 512 canales de 64 kbps. Las velocidades H22 y H4 deben ser múltiplos de la velocidad básica de transmisión de 64 kbps.

REDES DE ÁREA LOCAL

Se ha demostrado con estudios, que la mayoría (80%) de las comunicaciones entre terminales de datos y demás equipos de datos se hace dentro de un ambiente local, relativamente pequeño. Una *red de área local* (LAN, de *local area network*) proporciona el medio más económico y eficaz de manejar las necesidades locales de comunicación de datos. Una red de área local es, normalmente, un sistema de comunicaciones de datos de propiedad privada, en el que los usuarios comparten recursos, incluyendo programas de cómputo. Las LAN proporcionan comunicaciones en dos sentidos, entre una gran variedad de terminales de comunicación de datos, dentro de un área geográfica limitada, como por ejemplo, dentro del mismo recinto, edificio o complejo de edificios. La mayor parte del equipo de enlace de las LAN está separado por unas pocas millas.

La fig. 14-20 muestra cómo se pueden conectar varias PC a una LAN para compartir recursos comunes, como por ejemplo, un módem, una impresora o un servidor. El servidor puede ser una computadora más poderosa que las demás que comparten la red, o simplemente puede tener más espacio de almacenamiento en disco. El servidor "sirve" la información a las demás PC de la red, en forma de archivos de programa y de información de datos. Una PC de servidor es análoga a una computadora central, pero es mucho más pequeña.

Las LAN permiten que un recinto o más, llenos de computadoras, compartan recursos comunes, como impresoras o módems. La PC promedio sólo usa esos dispositivos un pequeño porcentaje del tiempo, por lo que no hay necesidad de dedicar impresoras y módems individuales para cada PC. Para imprimir un documento o archivo, una PC sólo manda la información al servidor, por la red. El servidor organiza y da prioridades a los documentos y a continuación los

PC

FIGURA 14-20 Componentes de una red de área local típica

manda, uno por uno, a la impresora de uso común. Mientras tanto, las PC quedan libres para continuar efectuando otras tareas útiles. Cuando una PC necesita un módem, la red establece una *conexión virtual* entre ella y el módem. La red es transparente a la conexión virtual, que permite que la PC se comunique con el módem como si estuvieran conectados directamente entre sí.

Las LAN permiten que las personas manden y reciban mensajes y documentos por la red, mucho más rápidamente que lo de tardarían en mandarse por un sistema de correo por papelería. El *correo electrónico* (e-mail) es un sistema de comunicaciones que permite a los usuarios mandarse mensajes entre sí, a través de sus computadoras. Con el correo electrónico, cualquier PC de la red puede mandar o recibir información a o de cualquier otra PC de la red, siempre y cuando las PC y el servidor usen el mismo programa, o uno compatible. También, con el correo electrónico se pueden interconectar usuarios de distintas redes en distintas ciudades, estados, países o hasta continentes. Para mandar un mensaje por correo electrónico, el usuario en una PC manda su dirección y su mensaje junto con la dirección del destinatario, al servidor. El servidor "retransmite" el mensaje a la PC de destino, si son suscriptores de la misma red. Si la PC destinataria está ocupada o no está disponible sea cual fuere la razón, el servidor guarda el mensaje y lo vuelve a mandar después. El servidor es la única computadora que debe mantener registro del lugar y la dirección de todas las demás PC de la red. Para mandar correo electrónico a suscriptores de otras redes, el servidor retransmite el mensaje al servidor de la red del usuario destinatario, que a su vez lo retransmite a la PC del destinatario. Se puede usar el correo electrónico para mandar información en texto (letras) y también archivos de programa, archivos gráficos, de audio y hasta de video. A esto se le llama comunicaciones multimedia.

Las LAN se usan mucho para interconectar una amplia gama de servicios, entre ellos los siguientes

Terminales de datos
Impresoras láser
Graficadores
Dispositivos de memoria de disco y de cinta, de gran capacidad
Máquinas de facsímil
Computadoras personales
Computadoras principales
Módems de datos
Bases de datos
Procesadores de palabra
Redes telefónicas públicas conmutadas

Sistemas de portadora digital (portadoras T)
Servidores de correo electrónico

Consideraciones sobre sistemas de red de área local

Las posibilidades de una red de área local se determinan principalmente con tres factores: *topología, medio de transmisión* y *protocolo de control de acceso*. Estos tres factores juntos determinan el tipo de datos, la velocidad de transmisión, la eficiencia y las aplicaciones que puede admitir una red en forma efectiva.

Topologías LAN. La topología o arquitectura física de una LAN identifica la forma de interconectar las estaciones (terminales, impresoras, módems, etc.). Los medios de transmisión que se usan en las LAN incluyen *pares trenzados de alambre*, *cable coaxial* y *cables de fibra óptica*. En la actualidad, la mayoría de las LAN usa cable coaxial; sin embargo, en muchas redes nuevas se están instalando cables de fibra óptica. Los sistemas de fibra pueden funcionar con mayores frecuencias de bits en transmisión, y tienen mayor capacidad para transferir información que los cables coaxiales.

Las topologías más comunes en las LAN son de estrella, bus, árbol de buses y anillo, y se ilustran en la fig. 14-21.

Topología de estrella. La propiedad más importante de la topología de estrella es que cada estación se enlaza en forma radial a un *nodo central* a través de una conexión directa de punto a punto, como se ve en la fig. 14-21a. En la configuración de estrella, una transmisión de una estación entra al nodo central, de donde se retransmite a todos los enlaces de salida. Por consiguiente, aunque el arreglo físico del circuito se asemeja a una estrella, se configura lógicamente como un bus, es decir, las transmisiones desde cualquiera de las estaciones las reciben todas las demás estaciones.

Los nodos centrales permiten que el sistema o la estación tengan un lugar cómodo para localizar sus fallas, porque todo el tráfico entre los nodos externos debe pasar por el nodo central. A veces, al nodo central se le llama *control central*, *acoplador de estrella* o *conmutador central*, y suele ser una computadora. La configuración de estrella se adapta mejor a aplicaciones en las que la mayoría de las comunicaciones se hace entre el nodo central y los externos. También se adapta bien a sistemas en los que hay una gran demanda de comunicación sólo con unas pocas terminales remotas. Los sistemas de tiempo compartido se configuran, en general, con topología de estrella. También, una configuración en estrella es adecuada para aplicaciones de procesamiento de palabra y administración de bases de datos.

Se pueden implementar los acopladores de estrella en forma pasiva o activa. Cuando se usan acopladores pasivos con un medio metálico de transmisión, los transformadores del acoplador proporcionan enlace electromagnético, que pasa las señales que llegan a los enlaces de salida. Si se usan cables de fibra óptica como medio de transmisión, se puede hacer el acoplamiento fundiendo las fibras entre sí. En los acopladores activos, los circuitos digitales del nodo central funcionan como repetidora. Los datos que llegan sólo se regeneran y se repiten a todas las líneas de salida.

Una desventaja de la topología en estrella es que la red sólo es tan confiable como el nodo central. Cuando falla el nodo central, falla el sistema. Sin embargo, si falla uno o más de los nodos externos, el resto de los usuarios puede continuar usando el resto de la red. Cuando es crítica la falla de cualquier entidad sencilla dentro de una red, hasta el grado de interrumpir el servicio de toda la red, a esa entidad se la llama *recurso crítico*. Así, el nodo central en una configuración de estrella es un recurso crítico.

Topología de bus. En esencia, la topología de bus es una configuración multipunto o multiterminal, en la que los nodos individuales se interconectan con un canal común de comunicaciones, compartido, como se ve en la fig. 14-21b. En la topología de bus, todas las estaciones se conectan usando los componentes adecuados de interconexión, en forma directa a un medio lineal de transmisión, que se suele llamar bus. En una configuración de bus, el control de la red no está centralizado en un nodo en particular. De hecho, la propiedad más característica de una LAN de bus es que el control se distribuye entre todos los nodos conectados con la LAN. Las transmisiones de datos por una red de bus tienen, en general, la forma de pequeños paquetes que

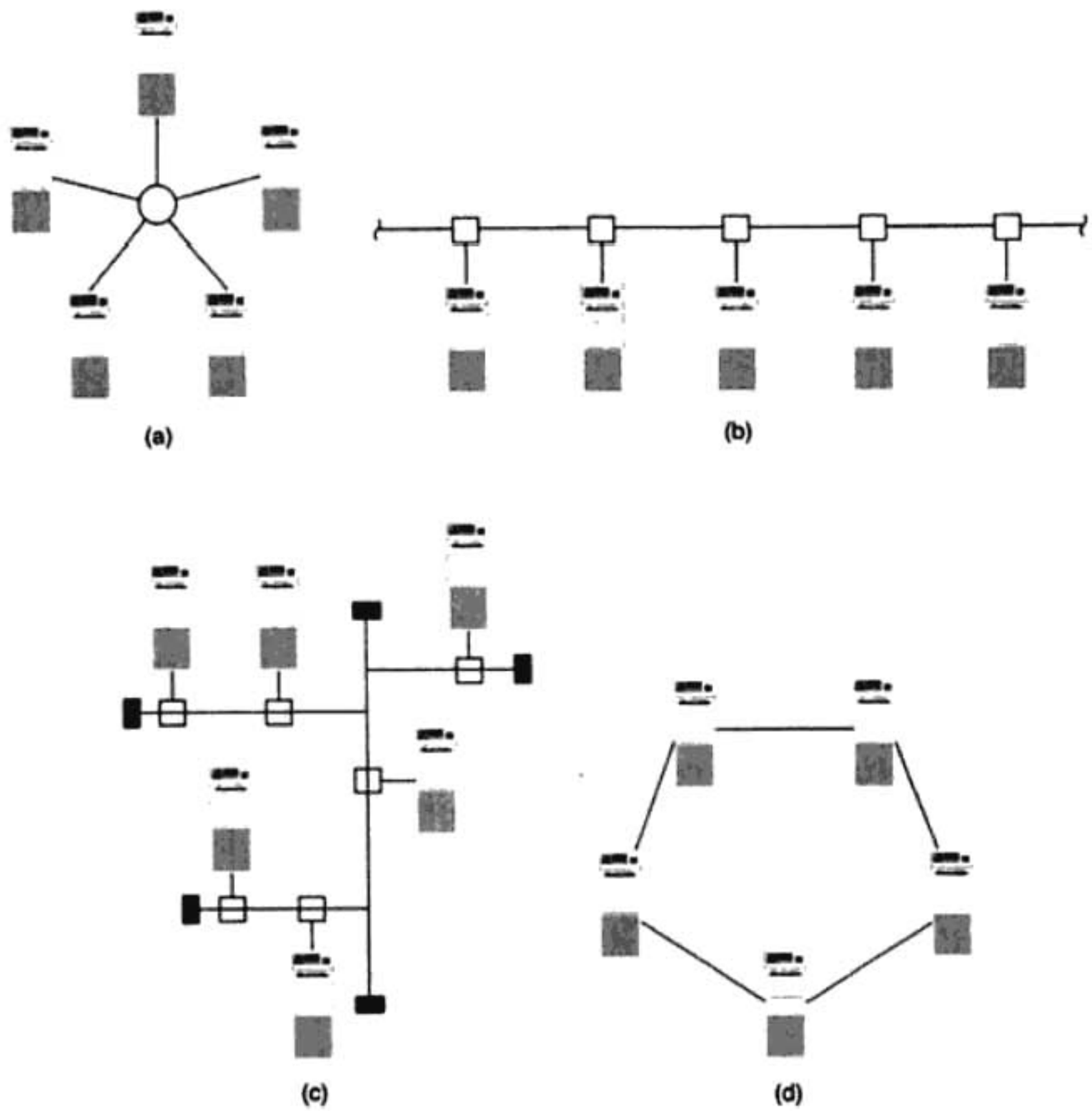


FIGURA 14-21 Topologías de LAN: (a) estrella; (b) bus; (c) árbol de buses; (d) anillo o lazo

contienen direcciones de usuario y datos. Cuando una estación desea transmitir datos a otra, vigila el bus para determinar si está siendo usada en ese momento. Si no hay otras estaciones comunicándose por la red (es decir, si la red está libre), la estación monitora puede comenzar a transmitir sus datos. Cuando una estación comienza a transmitir, todas las demás estaciones se vuelven receptoras. Cada receptora debe vigilar todas las transmisiones por la red y determinar cuáles son para ella. Cuando una estación identifica su dirección en el mensaje de datos que recibe, actúa en consecuencia o ignora la transmisión.

Una ventaja de la topología de bus es que no se requiere direccionamiento especial ni conmutación de circuito y, en consecuencia, no es necesario guardar y retransmitir mensajes dirigidos a otros nodos. Esta ventaja elimina una gran cantidad de indirectos por identificación de mensajes y tiempo de procesamiento. Sin embargo, cuando los sistemas se usan mucho, hay una gran probabilidad de que en cualquier momento alguna estación quiera transmitir al mismo tiempo. Cuando se presentan en forma simultánea transmisiones de dos o más estaciones, sucede una colisión de datos y se interrumpen las comunicaciones en toda la red. Es obvio que se necesita un esquema de prioridades para manejar las colisiones de datos. A ese esquema de prioridades se le llama *sentido de portadora, acceso múltiple con detección de colisión* (CSMA/CD), y se describirá en una sección posterior de este capítulo.

Como el control de la red no está centralizado en la configuración de bus, una falla de un nodo no interrumpirá el flujo de datos por toda la LAN. En este caso, el recurso crítico no es

un nodo, sino el bus mismo. Una falla en cualquier punto del bus abre la red y, de acuerdo con la versatilidad del canal de comunicaciones, puede interrumpir las comunicaciones en toda la red.

La adición de nuevos nodos a un bus puede ser, a veces, un problema, porque el tener acceso al cable del bus puede ser muy tedioso, en especial si está dentro de un muro, piso o techo. Una forma de reducir los problemas de instalación es agregar buses secundarios al canal primario de comunicaciones. Si se ramifican a otras bases, se forma una estructura de bus múltiple, llamada *bus de árbol*. La fig. 14-21c muestra una configuración de bus de árbol.

Topología de anillo. En una topología de anillo, las estaciones adyacentes se interconectan por repetidoras en una configuración de lazo cerrado, como muestra la fig. 14-21d. Cada nodo funciona como repetidora entre dos enlaces adyacentes dentro del anillo. Las repetidoras son dispositivos relativamente sencillos, capaces de recibir datos de un enlace y retransmitirlos a otro enlace. Los mensajes por lo general, en forma de paquetes, se propagan en modo símplex (en un sentido solamente), de nodo a nodo en torno del anillo, hasta recorrer todo el lazo y regresar al nodo de origen, donde se comprueba que los datos del mensaje de retorno sean idénticos a los que se transmitieron originalmente. Por lo anterior, la configuración de anillo tiene su mecanismo inherente de detección de errores. La o las estaciones de destino pueden reconocer la recepción de datos, activando o desactivando los bits adecuados dentro del segmento de control del paquete del mensaje. Los paquetes contienen campos de dirección para fuente y destino, así como otros datos para control de información y para el usuario en la red. Cada nodo examina los paquetes de datos que le llegan, copia los que le son dirigidos y funciona como repetidora para todos los paquetes, retransmitiéndolos, bit por bit, a la siguiente repetidora. Una repetidora no debe alterar el contenido de los paquetes recibidos, ni debe cambiar la velocidad de transmisión.

En la topología de anillo se puede usar prácticamente cualquier medio físico de transmisión. Los pares de alambre trenzado tienen bajo costo, pero las velocidades de transmisión son muy limitadas. Los cables coaxiales permiten mayor capacidad que los pares de alambre, prácticamente al mismo costo. Sin embargo, las máximas velocidades de datos se logran con cables de fibra óptica, pero tienen costo de instalación bastante mayor.

Formatos de transmisión en LAN. En las redes de área local se usan dos formatos, o técnicas, de transmisión múltiplex, desde una multitud de estaciones, a través de un solo medio de transmisión: banda base y banda ancha.

Formato de transmisión en banda base. Estos formatos se definen como formatos de transmisión que usan señalización digital. Además, los formatos de banda base usan el medio de transmisión como dispositivo de monocanal. Sólo puede transmitir una estación a la vez, y todas las estaciones deben transmitir y recibir las mismas clases de señales (esquemas de codificación, frecuencias de bits, etc.). Los formatos de transmisión de banda base multiplexan las señales por división de tiempo y las ponen en el medio de transmisión. Todas las estaciones pueden usar los medios, pero sólo una por una. Todo el espectro de frecuencias (ancho de banda) lo usa (o al menos queda disponible para) la estación que esté transmitiendo en el momento. En un formato de banda base, las transmisiones son bidireccionales. Una señal insertada en cualquier punto del medio de transmisión se propaga en ambos direcciones hasta los extremos, donde se absorbe. La señalización digital requiere una topología de bus, porque las señales digitales no se pueden propagar con facilidad a través de los divisores y acopladores necesarios en una topología de bus de árbol. Debido a las pérdidas en línea de transmisión, las LAN de banda base se limitan a una distancia no mayor que un par de millas.

Formatos de transmisión en banda ancha. En estos formatos se usan los medios de conexión como dispositivo multicanal. Cada canal ocupa una banda distinta de frecuencias, dentro del ancho de banda total asignado; es decir, es multiplexado por división de frecuencia. En consecuencia, cada canal puede contener distintos esquemas de modulación y codificación, y puede trabajar a distintas velocidades de transmisión. Una red de banda ancha permite transmitir, en forma simultánea, voz, datos digitales y video por el mismo medio de transmisión. Sin embargo, los sistemas de banda ancha son unidireccionales y necesitan módems, amplificadores y transceptores de RF, más complicados que para los sistemas de banda base. Por esta razón

estación tiene que transmitir un mensaje, pero la línea está ocupada, espera a que haya la condición de inactiva, para transmitirlo. Si dos estaciones transmiten al mismo tiempo, sucede una *colisión*. En este caso, la estación que primero siente la colisión manda una señal especial de atascamiento a las demás estaciones de la red. Todas las estaciones suspenden la transmisión (*regresan*) y esperan un período aleatorio para intentar una retransmisión. El tiempo aleatorio de retardo es distinto para cada estación y, en consecuencia, permite priorizar las estaciones de la red. Si se presentan colisiones sucesivas, se duplica el período de regreso para cada estación.

En CSMA/CD, las estaciones deben contender para entrar a la red. No se garantiza que una estación entre a la red. Para detectar la ocurrencia de una colisión, una estación debe poder transmitir y recibir en forma simultánea. La mayoría de las LAN configuradas en topología de bus usa CSMA/CD. Un ejemplo de una LAN que usa CSMA/CD es la *Ethernet*, que se describirá más adelante en este capítulo.

Otro factor que pudiera causar colisiones en CSMA/CD es el *retardo de propagación*. Es el tiempo que dura una señal en ir de una fuente a un destino. A causa del retardo de propagación, es posible que la línea parezca inactiva cuando, de hecho, otra estación está transmitiendo una señal que todavía no ha llegado a la estación monitora.

Paso de testigo. El *paso de testigo* es un método de acceso a una red, que se usa principalmente en las LAN configuradas con topología de anillo, y que usan formatos de banda base o de banda ancha. Cuando se usa el acceso por *paso de testigo*, los nodos no compiten para tener derecho a transmitir datos; un paquete específico de datos, llamado *testigo* se circula por el anillo, de estación a estación y siempre en la misma dirección. El testigo lo genera una estación designada, llamada *monitor activo*. Para que se permita transmitir a una estación, primero debe poseer el testigo. Cada estación, por turno, adquiere el testigo y examina la trama de datos para determinar si lleva un paquete dirigido a ella. Si la trama contiene un paquete con la dirección de la estación receptora, ésta lo copia en su memoria, le agrega todos los mensajes que deba mandar, y a continuación cede el testigo, retransmitiendo todos los paquetes de datos y el testigo al siguiente nodo de la red. Con el paso de testigo, cada estación tiene acceso igual al medio de transmisión. Así como en CSMA/CD, cada paquete transmitido contiene campos de dirección de la fuente y del destino. La estación destinataria confirma la buena entrega de una trama de datos, poniendo *banderas de estado de trama* y mandando a continuación la trama por el anillo hasta la estación transmisora original. A continuación, el paquete se saca de la trama antes de transmitir el testigo. No se puede usar dos veces un testigo, y hay un límite de tiempo de conservación del testigo. Así se evita que una estación interrumpa transmisiones de datos en la red, si guarda el testigo hasta que tenga un paquete por transmitir. Cuando una estación no tenga el testigo, sólo puede recibir y transferir otros paquetes destinados a otras estaciones.

Algunas redes de anillo de testigo, de 16 Mbps, usan una forma modificada de paso de testigo, en donde se cede el testigo tan pronto como se ha transmitido una trama, en lugar de esperar hasta que haya regresado la trama transmitida. A esto se le llama *mecanismo de liberación temprana de ficha*.

ANILLO DE PASO DE TESTIGO

Se acredita a Olaf Soderblum, el desarrollo de la primera arquitectura de red de *anillo de paso de testigo* en 1969. Sin embargo, la IBM fue la impulsora para la normalización y la adopción del anillo de testigo, y un prototipo desarrollado por IBM en Zurich, Suiza, sirvió como modelo para la Norma 802.5 de IEEE, de Anillo de testigo. Esta norma no incluye una especificación de rapidez de transmisión (bps) para anillos de testigo, aunque la IBM ha especificado redes de anillo de testigo que funcionan tanto a 4 Mbps como a 16 Mbps.

La fig. 14-22a muestra la distribución del anillo de pase para la trama especificada por la IEEE 802.5. El testigo consiste en tres *octetos*, siendo un octeto ocho bits de datos. Con frecuencia se usan en forma indistinta las palabras *octeto* y *byte*, porque la diferencia entre ellas es muy sutil, y está sujeta a la opinión. A veces se concibe a un byte como una palabra de ocho bits que representa determinado carácter, número o función, mientras que un octeto no es más que ocho bits de datos, que podrían representar voz digitalizada, video, imagen u otra información digital,

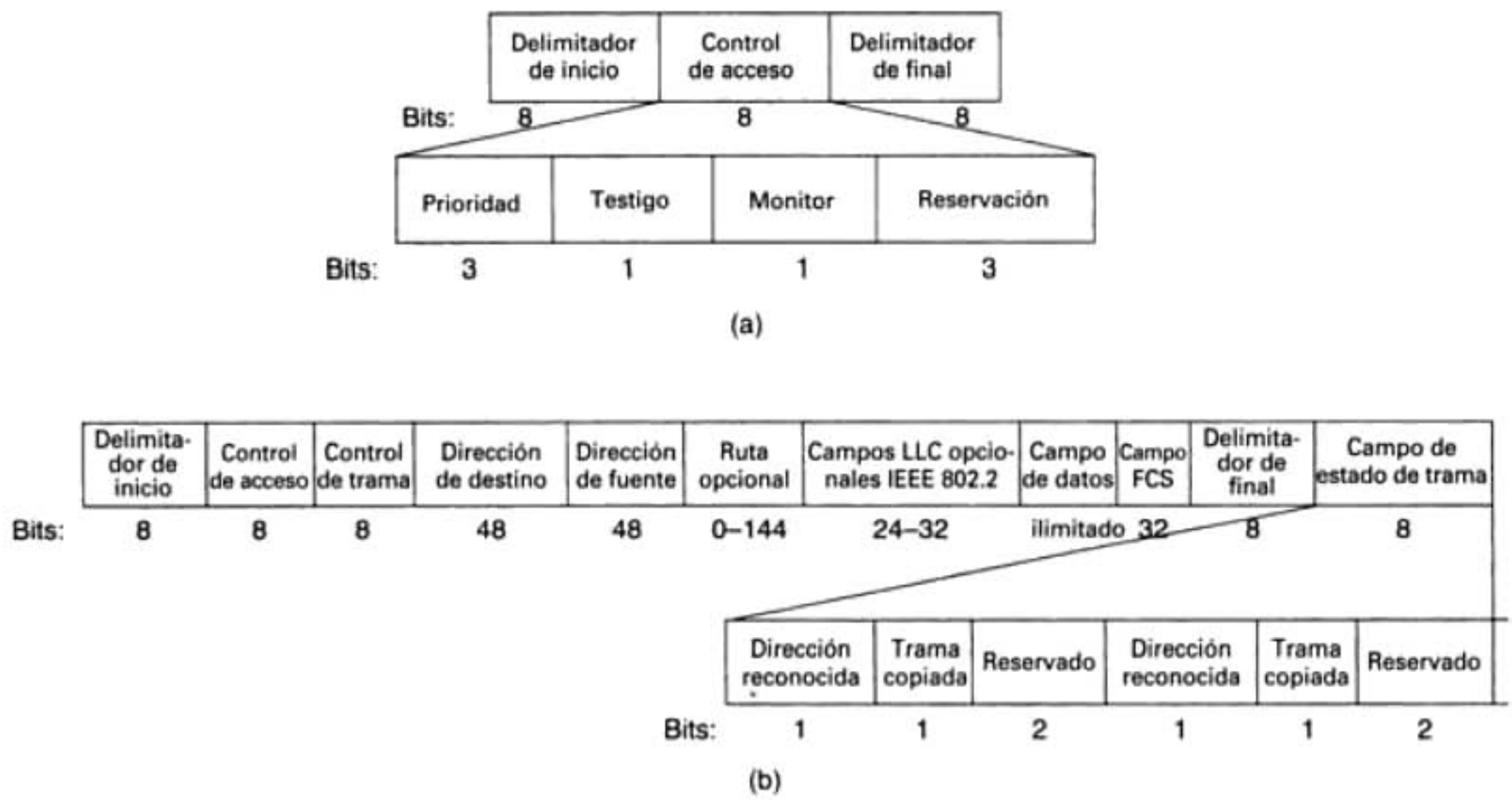


FIGURA 14-22 Formato de testigo y trama de subcapa MAC IEEE 802.5: (a) formato de trama de testigo; (b) formato de trama de subcapa MAC

o que quizá hasta se pueda descomponer en grupos más pequeños. Por otra parte, un byte de datos se concibe a veces como cualquier agrupamiento significativo de bits, independientemente de la longitud. Las funciones básicas de los campos dentro de las tramas de testigo y de subcapa según la IEEE 802.5 se describen en forma breve a continuación.

Campo delimitador inicial. Este campo alerta a las estaciones receptoras de un anillo de paso de testigo que se acerca una trama. Las tramas de testigo y de subcapa MAC comienzan con un delimitador inicial.

Campo de control de acceso. Este campo diferencia entre las tramas de testigo y de subcapa MAC (control de acceso a los medios, de *media access control*). Si el bit de testigo (T) está borrado, la trama recibida es un testigo libre, y el campo de control es seguido de inmediato por un delimitador final. La estación receptora puede obtener el testigo activando el bit de testigo. El campo delimitador inicial, más el campo de control de acceso con el bit T activado, forman los dos primeros campos de la trama de datos de subcapa MAC IEEEE 802.5 (se ven en la fig. 14-22b), y permite que la estación agregue información de dirección, datos y los campos restantes en la distribución de trama de datos; a continuación transmite la trama a la red.

El bit monitor (M) en el campo de control de acceso lo usa la estación designada como monitor activo de testigo, para administrar condiciones de error, como por ejemplo, cuando no circula testigo y cuando hay testigo persistentemente ocupado. La estación monitora detecta una estación de testigo perdido mediante un tiempo de salida mayor que el tiempo necesario para que la trama más larga recorra todo el anillo. Si durante este tiempo no se detecta un testigo, se supone que se perdió. Para recuperarlo, la estación monitora purga al anillo de todos los datos residuales y emite un testigo libre. Para detectar un testigo ocupado circulante, la estación monitora pone en 1 lógico en cualquier testigo ocupado que pase. Si detecta un testigo ocupado con el bit M ya activado, sabe que la estación de origen no pudo purgar su paquete. A continuación la estación monitora cambia el testigo ocupado por uno libre.

Los tres bits de prioridad (P) en el campo de control de acceso son para dar prioridades a las estaciones que esperan en la red. Hay ocho niveles de prioridad. Los tres bits

de reservación (R) son para reservar testigos futuros. Una estación con mayor prioridad que el testigo ocupado en el momento, puede reservar el siguiente testigo libre para su nivel de prioridad, cuando pase el testigo ocupado. Después de que la estación transmisora del momento haya terminado su transmisión, emite un testigo libre al nivel de la mayor prioridad. Las estaciones con un menor nivel de prioridad no pueden capturar al testigo y así lo dejan pasar a la estación solicitante, que tenga prioridad igual o mayor.

Campo de control de trama. Si la trama recibida contiene un bit T activado en el campo de control de acceso, la trama es la de datos de subcapa MAC, y el siguiente campo es el de control de trama, que indica si ésta contiene datos o es una especial de administración de la red.

Campos de dirección de fuente y destino. Los campos de dirección de fuente y destino contienen seis octetos cada uno, que identifican a las estaciones transmisora y receptora, respectivamente.

Campo de información de ruta (RIF, de *routing information field*). Este campo es opcional y se usa con dispositivos llamados puentes de ruta de fuente, para enlazar varias LAN de anillo de testigo. El RIF puede tener hasta 18 octetos de longitud.

Campos de control de enlace lógico IEEE 802.2 (LLC, de *logical link control*). Contienen tres o cuatro octetos opcionales para el campo de punto de acceso al servicio de destino (DSAP, de *destination service access point*), el campo de punto de acceso al servicio de fuente (SSAP, de *source service access point*) y el campo de control. Los campos DSAP y SSAP son para identificar las clases de protocolos incrustados en el campo de datos. Los códigos de punto de acceso al servicio los asigna el IEEE para identificar los protocolos particulares. Por ejemplo, el código EO de SAP identifica un protocolo Novell, y uno SAP de 06 identifica un protocolo TCP/IP.

Campo de datos. El campo de datos contiene datos de protocolo incrustado de nivel superior, si la trama es de datos, e información de administración de red si la trama es de administración de red. La longitud del campo de datos es ilimitada, siempre y cuando la estación transmisora no rebase el límite de tiempo de posesión de testigo. Por ejemplo, un límite de tiempo de 10 ms es un límite práctico de longitud del campo de datos, que para una red de anillo de testigo de 4 Mbps es de unos 4500 bytes, y de 16,000 a 18,000 bytes en una red de 16 Mbps.

Campo de secuencia de comprobación de trama (FCS, de *frame check sequence*). La IEEE 802.5 especifica una prueba de redundancia cíclica de 32 bits, idéntica a la que se usa en la IEEE 802.3.

Campo delimitador final. Este campo notifica a la estación receptora cuando llega el final de una trama, y también cuando la trama es intermedia, y que hay más datos que siguen a continuación. El delimitador final también puede indicar que una estación distinta de la de fuente o destino ha detectado un error en una trama. Cuando se detecta un error, la trama se debe pasar por alto y circular por el anillo hasta regresar a la estación transmisora para que la elimine.

Campo de estado de trama. Este campo contiene ocho bits que se usan para que la estación fuente sepa si se entregó bien la trama. Si la estación de destino reconoce su dirección, se activan los bits de reconocimiento de dirección. Si la trama no se copió bien en la memoria de la estación de destino, se activan los bits de trama copiada. Hay dos bits de reconocimiento de dirección y de trama copiada, por redundancia, porque este campo no es comprobado por el campo FCS.

Además de las funciones descritas en las secciones anteriores, la estación designada como monitor activo de la red, debe hacer las siguientes tareas:

Eliminar tramas que no las haya eliminado la estación transmisora.

Regenerar testigos perdidos o dañados.

Proporcionar una memoria especial de 24 bits si la longitud física del anillo no es la suficiente para proporcionar el suficiente retardo (latencia) para mantener el testigo de 24 bits.

Controlar el reloj maestro del sistema.

Asegurar que sólo haya una estación designada como monitor activo.

Anillo con muescas

El *anillo con muescas* es una variante del anillo de pase de testigo. Por el anillo con muescas circula una cantidad limitada de muescas de tiempo contiguas. Cada muesca es de tamaño fijo, y contiene posiciones en ella para las direcciones de fuente y destino, y para paquetes de datos. Se incluye un bit de *ocupado-desocupado* al principio de cada muesca circulante, para indicar si está disponible. Una muesca puede estar *llena* o *vacía*, dependiendo de la condición lógica del bit de ocupado: ocupado/desocupado (0 lógico = vacío; 1 lógico = lleno). Si una estación (nodo) desea transmitir datos, debe encontrar una muesca de tiempo vacía. La estación transmisora inserta sus datos en la muesca vacía, en el lugar apropiado, activa el bit ocupado y pasa la muesca a la siguiente estación. Cuando una estación identifica su dirección en un paquete recibido, copia los datos en sus memorias y restablece el bit de ocupado, para indicar que ahora la muesca está vacía y disponible para que la usen los siguientes nodos.

Bus de testigo

El pase de testigo se usa a veces en topología de bus o de árbol. En esas configuraciones, las estaciones en el bus deben formar un anillo lógico. Sin embargo, como la distribución física de un bus no se apega a un formato de anillo, se ha inventado una clase de secuencias ordenadas y lógicas para pasar el testigo de nodo a nodo. El testigo, con su dirección de destino, pasa por todo el bus, en configuración de anillo virtual. Esto es, cada estación recibe y transmite a estaciones predeterminadas, como si estuvieran antes o después que ella en un anillo. También aquí, un paquete de control, llamado testigo, regula el derecho de acceso. Cuando una estación recibe el testigo, se le asigna el control del medio de transmisión durante una muesca predeterminada de tiempo. La estación puede transmitir uno o más paquetes, y puede interrogar a otras estaciones o recibir paquetes. Cuando una estación termina con el paquete, o cuando transcurre el tiempo prescrito, la estación pasa el testigo a la siguiente estación lógica. Se permite que haya estaciones sin testigo en el bus; sin embargo, sólo pueden responder a interrogaciones de otras estaciones, o a peticiones de reconocimiento.

ETHERNET

Ethernet es un sistema de transmisión de datos en banda base, diseñado por Xerox Corporation, a mediados de la década de 1970. Se considera que Robert Metcalfe, quien después fue fundador de 3COM Corporation, fue el inventor de Ethernet. Para 1980, Xerox se unió a Intel Corporation y a Digital Equipment Corporation (DEC) para tratar de hacer que Ethernet fuera una norma de la industria. En 1981 se desarrolló una versión de Ethernet llamada DIX 1.0, o Ethernet I. Este sistema fue sustituido en 1982 por DIX 2.0, o Ethernet II, que sigue siendo la norma actual. Para 1983, el consorcio había desarrollado e introducido un controlador de un solo CI para Ethernet. Después, Mostek Corporation, en conjunto con DEC y Advanced Micro Devices, anunció un conjunto de dos CI para este objeto: uno en el controlador de red de área lógica para Ethernet (LANCE, por *logical area network controller for Ethernet*), y el segundo, como interfaz adaptadora. Después de agregar funciones de seguridad y de señalización, el Comité de Normas del IEEE adoptó a Ethernet y publicó la norma inicial de Ethernet, llamada 10BASE-5. Ese término se refiere a las especificaciones para la capa física y las capas de enlace de datos de la jerarquía de red de ISO. El "10" significa que esta adaptación de Ethernet funciona a 10 Mbps, y la palabra "BASE" indica que la LAN conduce datos de banda base (es decir, sin portadora). El "5" indica que la longitud máxima entre los extremos es de 500 m, a menos que se use una repetidora para amplificar las señales. En Ethernet se usa CSMA/CD para entrar a la red.

El subcomité de ANSI/IEEE 802.3 combinados ha definido medios de transmisión alternativos para el sistema 10BASE-5. Hoy, hay cuatro especificaciones para Ethernet: 10BASE-5 (Ethernet gruesa), 10BASE-2 (Ethernet de dos alambres), 1BASE-5 (StarLAN) y 10BASE-T (Ethernet de par trenzado). En la actualidad se trata de desarrollar métodos de transporte de señales Ethernet por cables de fibra óptica [10BASE-FL (FOIRL), fibra óptica, enlace interrepetidor] por radiodifusión terrestre y satelital, y por sistemas de luz infrarroja, en velocidades de transmisión mayores que 100 Mbps. El sistema Ethernet 1BASE-5 nunca tu-

Dirección de fuente. Esta dirección consiste en seis bytes (48 bits), y corresponden a la dirección de la estación que manda la trama.

Campo de tipo. En Ethernet no se usa el campo de tipo de 16 bits. Se pone en la trama para que se pueda usar en capas superiores de la jerarquía de protocolos OSI.

Campo de datos. El campo de datos contiene la información, y puede tener de 46 a 1500 bytes de longitud. El campo de datos es transparente. No se usan manipulaciones con caracteres de control de enlace de datos ni de bit cero. La transparencia se logra contando en reversa, a partir del carácter FCS (de secuencia de verificación de trama).

Campo de secuencia de verificación de trama. El campo CRC (de *frame check sequence*) contiene 32 bits para detección de errores, y se calcula a partir de los campos de encabezado y de datos.

Estructura estándar de trama IEEE 802.3

Preámbulo. El preámbulo consiste en siete bytes, para establecer la sincronización de los relojes. El último byte del preámbulo se usa como delimitador de la trama de arranque.

Delimitador de la trama de arranque. No es más que una serie de dos unos lógicos agregada al final del preámbulo, cuyo objetivo es marcar el final del preámbulo y el principio de la trama de datos.

Direcciones de destino y de fuente. Estas direcciones se definen de igual manera que en Ethernet II.

Campo de longitud. El campo de longitud de 2 bytes en la trama IEEE 802.3 reemplaza el campo de tipo en la trama Ethernet. El campo de longitud indica la longitud del campo de datos de control de enlace lógico (LLC, de *logical link control*), que es de longitud variable y contiene incrustados todos los protocolos de capa superior.

Control de enlace lógico (LLC). El campo LLC contiene la información, y puede tener de 46 a 1500 bytes de longitud. El campo LLC definido en la IEEE 802.3 es idéntico al definido para las redes de anillo con testigo.

Campo de secuencia de verificación de trama. El campo CRC (de *cyclic redundance check*) se define igual que en Ethernet II.

Delimitador de fin de trama. El delimitador de fin de trama es un periodo de 9.6 μs en el que no se transmiten bits. En la codificación Manchester, cuando no hay transiciones de longitud mayor que el tiempo de 1 bit, se indica el final de la trama.

Operación del sistema

Transmisión. El control de enlace de datos para Ethernet es CSMA/CD. Las estaciones ganan acceso al sistema de transmisión mediante contención. No se les interroga, ni tienen muescas específicas de tiempo para transmitir. Una estación que desea transmitir determina primero si hay otra estación que esté usando el sistema de transmisión. El controlador hace lo anterior a través del transceptor, detectando la presencia de una portadora en la línea. El controlador puede ser un componente con conexión permanente, o una función de programa, dependiendo de la complejidad de la estación.

La presencia de una portadora se representa por las transiciones de señal en la línea que produce el código Manchester. Si se detecta una portadora, la estación retrasa la transmisión hasta que la línea esté inactiva. Después del retardo requerido, la estación manda datos digitales al controlador. El controlador convierte esos datos a código Manchester, inserta la CRC, añade el preámbulo y pone el paquete en línea. Todavía no queda asegurada la transmisión del paquete completo. Puede ser que una estación diferente haya detectado línea inactiva y comenzado a transmitir su propio paquete. La primera estación vigila la línea durante un periodo llamado *ventana de colisión*, o *intervalo de colisión*. Este intervalo es función del retardo de propagación de un extremo a otro de la línea. El retardo, incluyendo aquel que producen las repetidoras que haya y medido en distancia, no puede ser mayor que 2.5 km. Si no ha sucedido una colisión de datos en este intervalo, se dice que la estación tiene la adquisición de línea, y continuará transmitiendo todo el paquete. Si se detectara una colisión, ambas estaciones transmisoras abortan de

inmediato sus transmisiones, durante un tiempo aleatorio, y a continuación intentan la retransmisión. El transceptor puede detectar una colisión de datos comparando la señal recibida con la transmitida. Para hacer la comparación, una estación debe estar transmitiendo todavía su paquete mientras que una señal transmitida antes se haya propagado hasta el extremo de la línea y haya regresado. Esto indica que un paquete debe tener cierto tamaño mínimo. Si se detecta una colisión, el controlador debe emprender la acción necesaria. La interfaz de controlador y transceptor contiene una línea para notificar presencia de colisión (onda cuadrada de 10 MHz). Las tres líneas restantes de esta interfaz son las de transmitir datos, recibir datos y la corriente para el transceptor. Como una colisión se manifiesta como alguna forma de violación de fase, el controlador sólo puede detectarla. En Ethernet, el transceptor es el que principalmente detecta las colisiones de datos. Siempre que es posible, esto se complementa con una instalación detectora de colisión en el controlador. Para asegurar que todas las estaciones estén enteradas de la colisión, se invoca un *procedimiento de consenso para forzamiento de colisión*. Cuando el controlador detecta una colisión, transmite de cuatro a seis bytes de datos aleatorios. A estos bytes se les llama *secuencia de atoramiento*. Si sucedió una colisión, se determina el tiempo aleatorio de espera de la estación, para que vuelva a transmitir, mediante un *algoritmo binario exponencial de regreso*. Se suele establecer la muesca de tiempo de modo que sea un poco más larga que el tiempo de viaje redondo de la señal en el canal. El tiempo del retardo por espera se selecciona al azar de este intervalo.

Ejemplo 14-4

Muesca de tiempo de transmisión = tiempo de 512 bits

$$\text{Tiempo máximo del intervalo} = \frac{512 \text{ bits}}{10 \text{ Mbps}} = 51.2\ \mu\text{s}$$

$$\text{Intervalo de tiempo de retardo} = 0 \text{ a } 51.2\ \mu\text{s}$$

Por cada colisión siguiente que encuentra el mismo paquete, el intervalo de tiempo se duplica, hasta llegar a un intervalo máximo. El intervalo máximo se especifica como $2^{10} \times$ muesca de tiempo de transmisión. Pasados 15 intentos sin éxito para transmitir un paquete, ya no se hacen intentos y se informa el error a la estación. Ésta es la desventaja principal de Ethernet: no puede garantizar la entrega del paquete cuando hay un tráfico intenso de transmisiones.

Recepción. La línea se vigila hasta que se detecta la dirección de la estación. El controlador desprende el preámbulo, comprueba el CRC y regresa del código Manchester al formato digital. Si el paquete contiene errores, se desecha. El final del paquete se reconoce por la ausencia de una portadora en la línea de transmisión. Eso indica que no se detectaron transiciones en el periodo de 75 a 125 ns a partir del centro de la última celda de bit. La decodificación se hace por un lazo amarrado a fase. Este lazo se inicializa mediante la secuencia conocida del preámbulo.

La tabla 14-13 muestra una comparación de la norma IEEE 802.3 y las especificaciones físicas 10BASE-5, 10BASE-2 y 10BASE-T de Ethernet.

INTERFAZ DE DATOS DISTRIBUIDOS EN FIBRA

Las normas LAN de primera generación, como Ethernet y anillo de testigo, no tienen el ancho de banda necesario para manejar las altas velocidades de transmisión requeridas en los sistemas de comunicación de alta capacidad por multiplexado, que conducen sonido, fotos fijas y video comprimido digitalizados, en especial a través de grandes distancias. En consecuencia, el American National Standards Institute (ANSI) publicó en 1984 un conjunto de normas (X3T9.5) llamado *Interfaz de datos distribuidos en fibra* (FDDI, por *fiber data distributed interface*). En el conjunto se especifican normas para una topología de anillo con cables de fibra óptica y velocidades de transmisión hasta de 100 Mbps, que permiten distancias hasta de 2 km entre estaciones. La FDDI no es una norma IEEE, sin embargo, admite los protocolos IEEE 802.2, de control de enlace lógico, y permite tener una interoperabilidad transparente para las capas de tercera a séptima, con los protocolos que se apegan a IEEE.

TABLA 14-13 Comparación de la norma 802.3 de IEEE, Especificaciones físicas para Ethernet 10BASE-5, 10BASE-2 y 10BASE-T

	10BASE-5 (Ethernet)	10BASE-2 (CheaperNet)	10BASE-T (UTP)
Control de acceso	CSMA/CD	CSMA/CD	CSMA/CD
Topología	Bus	Bus	Estrella
Protocolo de mensaje	Tamaño variable de paquetes	Tamaño variable de paquetes	Tamaño variable de paquetes
Frecuencia de señalización	10 Mbps	10 Mbps	10 Mbps
Tipo de señalización	Banda base	Banda base	Banda base
Tipo de cable	IEEE 802.3 Coaxial grueso con doble blindaje	RG-58 Coaxial	UTP nivel 3 o nivel 4
Impedancia de cable	50 Ω	50 Ω	No se aplica
Separación máxima entre nodos	2.5 m	0.5 m	No se aplica
Longitud máxima de segmento	500 m	185 m	100 m (PC a cubo)
Nodos máximos por segmento	100	30	1024 (por red)
Cantidad máxima de segmentos	5	2	No se aplica
Separación máxima entre estaciones	2500 m	925 m	No se aplica

Además de tener mayores anchos de banda y mayores velocidades de operación, la FDDI tiene también mayor confiabilidad y seguridad que Ethernet y los anillos de testigo. La mayor confiabilidad se puede atribuir al medio de transmisión. Las fibras ópticas son inmunes tanto a la interferencia electromagnética como a la interferencia por radiofrecuencia. La topología física de las FDDI se presta para tener un sistema más confiable, porque está compuesta de dos anillos separados: uno primario y uno de respaldo. La topología que se usa con la FDDI es configuración de anillo doble (contrarrotatorio doble) como se ve en la fig. 14-26. Uno de los anillos es el *anillo primario* y el otro es el *anillo secundario*. Los datos se propagan en torno a los dos anillos, en direcciones contrarias. Bajo las condiciones normales de funcionamiento, sólo el anillo primario conduce datos, y el secundario se reserva como reserva, y sólo se usa si falla el primario. Como en la norma IEEE 802, la FDDI usa pase de testigo como método de acceso. Sin embargo, hay varias diferencias entre las dos técnicas, y se pretende que la FDDI aproveche la ventaja de su anillo de alta velocidad para maximizar la eficiencia.

Nodos FDDI

Las redes LAN con FDDI pueden tener hasta 500 nodos separados por 2 km, y todo el medio puede abarcar una distancia total hasta de 200 km (125 millas) si se usan repetidoras al menos cada 2 km. En FDDI se pueden usar dos tipos de nodos: de *conexión única* y de *conexión dual*, como muestra la fig. 14-26. Allí se ve que los nodos de conexión dual son estaciones conectadas a los anillos primario y secundario, a través de un dispositivo de acceso parecido a un cubo de Ethernet, que se llama *concentrador*. Las estaciones de conexión única no pueden usar como respaldo el anillo secundario, si falla el primario. Así, si falla o se interrumpe la conexión entre el concentrador y el anillo primario, todas las estaciones de conexión única se desconectan de la red. Las estaciones de conexión dual deben tener dos transmisores ópticos y dos receptores ópticos, por lo cual es obvio que sean más costosas. En consecuencia, en muchas redes FDDI sólo tienen conexión dual los nodos más críticos, como el servidor de red, por ejemplo.

Igual que en los cubos de anillo de testigo en Ethernet, los concentradores FDDI pueden sentir cuándo está apagado o desconectado uno de los nodos de conexión única. Al suceder esto, el concentrador desconecta de la red el nodo desactivado. También, el concentrador tiene capacidad de saltarse los nodos defectuosos de conexión única, para restablecer la continuidad en el resto del anillo primario.

La fig. 14-27 ilustra la función de autocicatrización de una red FDDI. Siempre que hay una ruptura de cable, los nodos de conexión doble a ambos lados de la interrupción conectan a los anillos primario y secundario, formando un solo anillo, más largo. La capacidad que tiene

	Preámbulo	Delimitador de inicio	Control de trama	Delimitador de final
Bytes:	8	1	1	1

(a)

	Preámbulo	Delimitador de inicio	Control de trama	Dirección de destino	Dirección de fuente	Datos hasta	Secuencia de comprobación de trama	Delimitador de final	Estado de trama
Bytes:	8	1	1	6	6	4500	4 (32 bit CRC)	.5 (4 bits)	1.5 (12 bits)

(b)

FIGURA 14-28 Formatos de trama de testigo y de datos: (a) formato FDDI de testigo; (b) estructura FDDI de trama de datos

una red FDDI para reconfigurarse cerca de defectos de cable y de circuito, estableciendo lo que se llama *salida de anillo*, hace que sea más confiable que las redes Ethernet o de anillo de testigo.

En FDDI se usa una forma modificada de pase de testigo, como metodología de acceso. La metodología FDDI es distinta de la norma IEEE 802.5 en dos aspectos. En primer lugar, debido a las mayores longitudes posibles en las LAN FDDI, no es práctico convertir los testigos libres en testigos ocupados y permitir que una estación monopolice la red hasta recibir la confirmación de que se recibió bien su mensaje en la estación de destino. En lugar de sólo cambiar la condición lógica del bit T en el byte de control de acceso, en FDDI se elimina físicamente el testigo del anillo, y se transmite una trama completa de datos. Después de terminar la transmisión se emite un nuevo testigo. Se evitan las colisiones porque sólo una estación puede poseer, en cualquier momento, el testigo libre.

La segunda diferencia en FDDI es que una sola estación puede mandar muchos mensajes sucesivos antes de que ceda el testigo. Recuérdese que la IEEE 802.5 sólo permite que haya un mensaje por testigo y por usuario. Las tramas transmitidas en forma de corriente continua se llaman tramas síncronas, y se les da prioridad de acuerdo con una metodología llamada *asignación de ancho de banda síncrona* (SBA, de *synchronous bandwidth allocation*). La fig. 14-28 muestra el formato de un testigo FDDI y una estructura de trama de datos FDDI.

Red FDDI con líneas de transmisión de par de conductores metálicos trenzados

La FDDI se diseñó originalmente para usarse con una línea de transmisión de fibra óptica, y no con hilos de cobre. Los cables de fibra óptica pueden admitir mayores velocidades de transmisión de datos, son mucho más inmunes al ruido de interferencia y no irradian energía electromagnética como sus contrapartes metálicas. Sin embargo, los transductores que se usan con los cables de fibra óptica, que convierten la energía luminosa en energía eléctrica (los fotodetectores) y viceversa (los diodos emisores de luz) son relativamente costosos y pueden añadir varios cientos de dólares al costo de cada nodo de una red FDDI.

Para reducir el costo inicial de una red FDDI, la ANSI publicó una versión de la norma FDDI que especifican conductores metálicos en par trenzado, y no un cable de fibra óptica. En la norma se especifica una velocidad de transmisión de 100 Mbps, con el mismo protocolo que en la versión de FDDI para fibra óptica. En vista de que la mayoría de los cables telefónicos no pueden permitir una frecuencia de bits de datos de 100 Mbps, se debe instalar un cable especial, *grado datos*. Sin embargo, aun con este cable, que es más costoso, las redes metálicas FDDI siguen siendo menos costosas que los sistemas de cable óptico.

Canal Fibre

Aunque las normas de una red de alta velocidad, como la FDDI, permiten tener comunicaciones digitales poco costosas y fidedignas a velocidades de datos hasta de 100 Mbps, muchos ambientes comerciales, educativos, industriales y científicos siguen investigando sistemas que permitan velocidades de datos aún mayores. En fecha reciente, anunciaron Hewlett-Packard Co., IBM

14-65. Explique el sistema CSMA/CD.

14-66. Explique el pase de testigo.

14-67. Describa qué es una ISDN, y quién propuso este concepto.

14-68. ¿Cuáles son los principios primarios de una ISDN?

14-69. ¿Cuáles fueron las evoluciones de ISDN?

14-70. Describa la arquitectura propuesta para ISDN.

14-71. Describa un canal D de ISDN. Describa un canal B de ISDN.

14-72. Describa los siguientes términos respecto a ISDN: *equipo de terminal 1, equipo de terminal 2, adaptador de terminal, punto de referencia S, punto de referencia T, punto de referencia R, terminación 1 de red, terminación 2 de red, terminación 1, 2 de red; unidad de terminación de línea, terminación de intercambio* y *punto de referencia V.*

14-73. Describa el formato LAP-D en ISDN.

14-74. ¿Cuáles son las diferencias entre los formatos LAP-D y LAP-B en ISDN?

14-75. ¿Qué es un identificador de punto de servicio ISDN y cuándo se usa?

14-76. ¿Qué es un identificador de terminal extrema y cuándo se usa?

14-77. Describa los conceptos básicos de BISDN, y en qué difiere de ISDN de banda angosta.

14-78. ¿Qué quiere decir modo de transferencia asíncrona?

14-79. ¿Cuáles son las nuevas velocidades de canal de datos propuestas para BISDN?

14-80. ¿Cuáles son los medios de transmisión que se usan con la norma de comunicaciones con modo de transferencia asíncrona?

14-81. Describa los canales virtuales y las trayectorias virtuales que se usan con ATM.

14-82. Identifique y explique los diversos subcampos que se usan en un campo de encabezado en el modo ATM.

14-83. Identifique y explique los diversos subcampos que se usan dentro de un campo de información en el ATM.

14-84. ¿Cuál es el principal medio de transmisión que se usa en SONET?

14-85. Describa el formato de trama que se usa en SONET.

14-86. ¿Qué técnica de multiplexado se usa en SONET para lograr mayores niveles de transporte síncrono?

14-87. Describa las topologías que se usan con frecuencia en las LAN.

14-88. Haga una lista de los formatos de transmisión que se usan en las LAN, y descríbalos.

14-89. Haga una lista de las metodologías de control de acceso que se usan en las LAN, y descríbalas.

14-90. Describa los diversos campos y tramas de subcapa que se usan en una red de anillo de pase de testigo.

14-91. Describa la operación de una LAN de anillo con muescas.

14-92. ¿Qué es Ethernet?

14-93. Describa en forma breve los siguientes sistemas Ethernet: 10BASE-5, 10BASE-2 y 10BASE-T.

14-94. Identifique y explique las diversas secciones de un formato de trama en Ethernet II.

PROBLEMAS

14-1. Determine el código hexadecimal en el campo de control de una trama SDLC, para las siguientes condiciones: trama de información, interrogación, transmisión de trama 4 y confirmación de recepción de tramas 2, 3 y 4.

14-2. Determine el código hexadecimal en el campo de control de una trama SDLC para las siguientes condiciones: trama de supervisión, listo para recibir, final, confirmar recepción de las tramas 6, 7 y 0.

14-3. Inserte ceros en la siguiente corriente de datos SDLC.

111 001 000 011 111 111 100 111 110 100 111 101 011 111 111 111 001 011

14-4. Elimine ceros de la siguiente corriente de datos SDLC.

010 111 110 100 011 011 111 011 101 110 101 111 101 011 100 011 111 00

14-5. Trace la forma de onda NRZI para la siguiente corriente de datos; comience con una condición alta.

1 0 0 1 1 1 0 0 1 0 1 0

14-6. Determine el código hexadecimal en el campo de control en una trama SDLC para las siguientes condiciones: trama de información, no es pregunta, transmitiendo la trama número 5 y confirmando la recepción de las tramas 0, 1, 2 y 3.

14-7. Determine el código hexadecimal del campo de control en una trama SDLC para las siguientes condiciones: trama de supervisión, no listo para recibir, no es final, confirmando recepción de tramas 7, 0, 1 y 2.

14-8. Inserte ceros en la siguiente corriente de datos SDLC.

0110111111101100001111100101110001011111111011111001

14-9. Elimine ceros de la siguiente corriente de datos SDLC.

0010111110011111011111011000100011111011101011000101

14-10. Trace los niveles NRZI para la siguiente corriente de datos; comience con una condición alta.

1 1 0 1 0 0 0 1 1 0 1

CAPÍT

Imagen protegida por derechos de autor

Transmisión digital

INTRODUCCIÓN

Como se dijo en capítulos anteriores, la *transmisión digital* es el transporte de señales digitales entre dos o más puntos en un sistema de comunicaciones. Las señales pueden ser binarias o cualquier otra forma de pulsos digitales de valores discretos. La información de la fuente original puede estar en forma digital, o podrían ser señales analógicas convertidas en impulsos digitales antes de su transmisión, para reconvertirlas en señales analógicas en el receptor. En los sistemas de transmisión digital se requiere una instalación física, como cable metálico, cable coaxial o un enlace de fibra óptica, para interconectar los diversos puntos del sistema. Los pulsos están contenidos en la instalación y se propagan por ella.

Ventajas de la transmisión digital

1. La ventaja principal de la transmisión digital respecto a la analógica es su inmunidad al ruido. Los impulsos digitales son menos susceptibles a variaciones causadas por ruido, que las señales analógicas. En la transmisión digital no es necesario evaluar las características de amplitud, frecuencia y fase con tanta precisión como en la transmisión analógica. En lugar de ello, los pulsos recibidos se evalúan durante un intervalo preciso de muestreo, y se hace una determinación simple para ver si el pulso está arriba o abajo de un nivel de umbral. No es importante la amplitud, frecuencia o fase exactas de la señal recibida.

2. Las señales digitales se prestan mejor a su procesamiento y multiplexado que las señales analógicas. El *procesamiento digital de la señal* (DSP, de *digital signal processing*) es el procesamiento de las señales analógicas aplicando métodos digitales. En el procesamiento digital se incluyen el filtrado, igualación y desplazamiento de fase. Los pulsos digitales se pueden guardar con más facilidad que las señales analógicas. También, la rapidez de transmisión de un sistema digital se puede cambiar con facilidad para adaptarse a ambientes distintos, y para interconectar distintas clases de equipo.

3. Los sistemas digitales de transmisión son más resistentes al ruido que sus contrapartes analógicas. Los sistemas digitales usan regeneración de señal, y no usan amplificación de señal. El ruido producido en los circuitos de amplificadores electrónicos es aditivo y, en consecuencia, la relación de señal a ruido se deteriora cada vez que se amplifica una señal analógica. Así, la

distancia total a la que se pueden transportar las señales analógicas está limitada por la cantidad de amplificadores. Por otra parte, los regeneradores digitales muestrean la señal de entrada con ruido y a continuación reproducen una señal digital enteramente nueva, con la misma relación de señal a ruido que la señal original transmitida. En consecuencia, las señales digitales se pueden transportar a distancias mayores que las señales analógicas.

4. Es más fácil medir y evaluar las señales digitales. En consecuencia, es más fácil comparar la eficiencia de sistemas digitales alternativos con capacidades distintas de señalización e información que en sistemas equiparables analógicos.

5. Los sistemas digitales se adaptan más para evaluar el funcionamiento con errores. Se pueden detectar y corregir los errores de transmisión en señales digitales, con más facilidad y más exactitud que las que son posibles en los sistemas analógicos.

Desventajas de la transmisión digital

1. La transmisión de señales analógicas codificadas digitalmente requiere un ancho de banda bastante mayor que la simple transmisión de la señal analógica original. Es importante el ancho de banda por ser costoso, y porque con frecuencia es muy limitado.

2. Las señales analógicas se deben convertir en códigos digitales antes de su transmisión, y reconvertirse a la forma analógica en el receptor, necesitando, por consiguiente, circuitos adicionales de codificación y decodificación.

3. La transmisión digital requiere una sincronización precisa, respecto al tiempo, entre los relojes del transmisor y del receptor. Por consiguiente, los sistemas digitales requieren costosos circuitos de recuperación de reloj en todos los receptores.

4. Los sistemas de transmisión digital son incompatibles con las instalaciones anteriores, de transmisión analógica.

MODULACIÓN DE PULSO

En la *modulación de pulso* se incluyen muchos métodos distintos para convertir información a la forma de pulso, para transferir éste de una fuente a un destino. Los cuatro métodos principales son *modulación por ancho de pulso* (PWM, de *pulse width modulation*), *modulación por posición de pulso* (PPM, de *pulse position modulation*), *modulación por amplitud de pulso (PAM,* de *pulse amplitude modulation*) y *modulación por código de pulso* (PCM, de *pulse code modulation*). Estos métodos se describen en resumen a continuación y se muestran en la fig. 15-1.

1. *PWM.* A este método se le llama a veces modulación por duración de pulso (PDM, de *pulse duration modulation*) o modulación por longitud de pulso (PLM, de *pulse length modulation*). El ancho del pulso (la parte activa del ciclo de trabajo) es proporcional a la amplitud de la señal analógica.

2. *PPM.* Se varía la posición de un pulso de ancho constante, dentro de una muesca predeterminada de tiempo, de acuerdo con la amplitud de la señal analógica.

3. *PAM.* Se varía la amplitud de un pulso de ancho constante y posición constante, de acuerdo con la amplitud de la señal analógica.

4. *PCM.* Se muestrea la señal analógica y se convierte en un número binario en serie, de longitud fija, para su transmisión. El número binario varía de acuerdo con la amplitud de la señal analógica.

La PAM se usa como forma intermedia de modulación en PSK, QAM y PCM, aunque casi nunca se usa por sí misma. Las modulaciones PWM y PPM se usan en sistemas de comunicaciones de propósito especial, en especial los militares, pero casi nunca se usan en los sistemas comerciales. La PCM es, con mucho, el método más frecuente de modulación de pulso y, en consecuencia, será el tema de las descripciones en lo que resta del capítulo.

MODULACIÓN POR CÓDIGO DE PULSO

La *modulación por código de pulso* (PCM, de *pulse code modulation*) fue desarrollada en 1937 por la AT&T en sus laboratorios de París. Se acredita a Alex H. Reeves haberla inventado. Aunque se

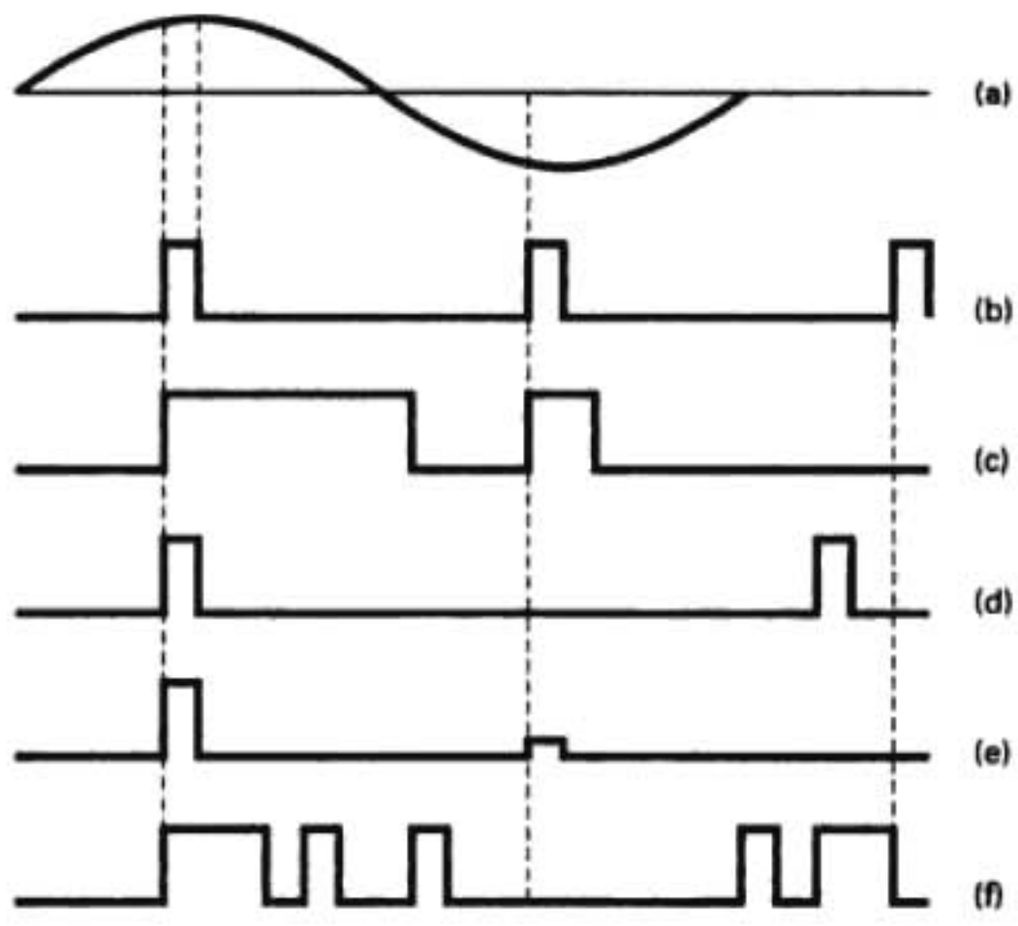


FIGURA 15-1 Modulación de pulso: (a) señal analógica; (b) pulsos de muestra; (c) PWM; (d) PPM; (e) PAM; (f) PCM

reconocieron pronto sus méritos al desarrollarla, no fue sino hasta mediados de la década de 1960, al difundirse la electrónica de estado sólido, que la PCM se hizo dominante. Hoy, en Estados Unidos, el método preferido de comunicaciones dentro de la red telefónica pública conmutada es la PCM.

La modulación por código de pulso es la única de las técnicas de modulación por codificación digital de la fig. 15-1 que se usa para transmisión digital. El término "modificación por código de pulso" es algo equívoco, porque en realidad no se trata de una forma de modulación, sino más bien una forma de codificación de fuente. En la PCM, los pulsos son de longitud y amplitud fijas. Es un sistema binario en el que la presencia o ausencia de un pulso dentro de una muesca predeterminada de tiempo representa una condición de uno lógico o de cero lógico. Las técnicas PWM, PPM y PAM son digitales, pero casi nunca son binarias, porque un pulso no representa a un solo dígito binario (bit).

La fig. 15-2 muestra un diagrama de bloques simplificado de un sistema PCM de un solo canal o *símplex* (sólo de un sentido). El filtro pasabanda limita la frecuencia de la señal analógica de entrada al intervalo de frecuencias de banda de voz, de 300 a 3000 Hz. El *circuito de muestreo y retención* muestra, en forma periódica, la señal analógica de entrada, y convierte esas muestras en una señal PAM de varios niveles. El *convertidor analógico a digital* (ADC, de *analog-to-digital converter*) convierte las muestras PAM en códigos PCM paralelos, que se convierten a su vez en datos en serie, en el convertidor de *paralelo a serie*, y a continuación salen a la línea de transmisión. Las repetidoras en la línea de transmisión regeneran en forma periódica los códigos PCM.

En el receptor, el *convertidor serie a paralelo* convierte los datos serie de la línea de transmisión en códigos PCM paralelos. El *convertidor digital a analógico* (DAC, de *digital-to-analog converter*) convierte el código paralelo PCM en señales PAM de varios niveles. El circuito de *retención* y el *filtro pasabajas* regresan la señal PAM a su forma analógica original.

La figura 15-2 muestra también varios pulsos de reloj y de muestreo que se explicarán en secciones posteriores de este capítulo. Un circuito integrado que ejecuta las funciones de codificación y decodificación PCM se llama *codec* (codificador/decodificador) y se detalla en el capítulo 16.

Muestreo PCM

La función de un circuito de muestreo en un transmisor PCM es tomar una muestra periódica de la señal analógica de entrada, que varía en forma continua, y convertir esas muestras en una serie de pulsos que se puedan convertir con más facilidad a un código PCM binario. Para que el ADC convierta fielmente una señal en código binario, la señal debe ser relativamente constante. Si no lo es, antes de que el ADC termine la conversión, la señal cambiaría y el ADC trataría en forma continua de seguir los cambios analógicos, y nunca se estabilizaría en algún código PCM.

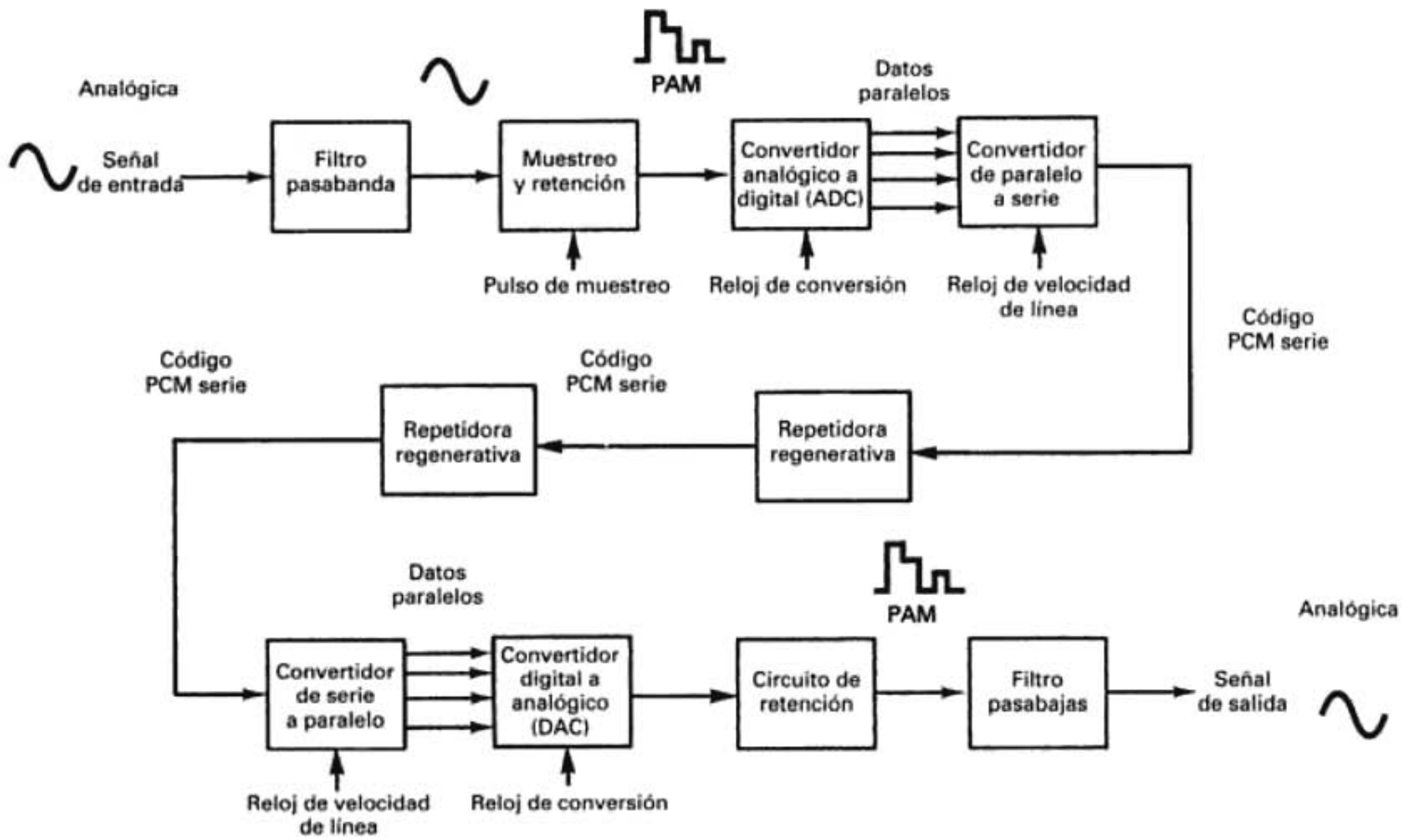


FIGURA 15-2 Diagrama de bloques simplificado de un sistema de transmisión PCM simplex, de un solo canal

Hay dos técnicas básicas para llevar a cabo la función de muestreo y retención: muestreo natural y muestreo de parte plana. El *muestreo natural* se ilustra en la fig. 15-3. Es cuando se muestrean las partes superiores de la forma de onda analógica que se muestrea y conservan su forma natural. En la fig. 15-3a, el interruptor analógico de FET tan sólo conecta a tierra la forma de onda de entrada, cuando el pulso de muestreo es alto. Sin embargo, cuando es bajo, se permite que la señal de entrada pase inalterada por el amplificador de salida, y llegue a la entrada del convertidor analógico a digital. La forma de onda de una señal con muestreo natural se ve en la fig. 15-3b.

En el muestreo natural, el espectro de frecuencias de la salida muestreada es distinto al de una muestra ideal. La amplitud de los componentes de frecuencia, obtenidos con pulsos angostos y de ancho finito es menor para las armónicas altas de acuerdo con la función (sen x)/x. Esto altera el espectro de frecuencias de información, y hace necesario usar igualadores de frecuencia (filtros de compensación) antes de la recuperación mediante un filtro de paso bajo.

El método más común para muestrear señales de voz en sistema PCM es *muestreo de parte plana*, que se logra en un *circuito de muestreo y retención*. El objeto de ese circuito es muestrear en forma periódica la señal analógica de entrada, que cambia en forma continua, y convertir esas muestras en una serie de niveles de PAM de amplitud constante. El muestreo de parte plana altera el espectro de frecuencias, e introduce un error llamado *error de abertura*, que evita que el circuito de recuperación, en el receptor PCM, reproduzca con exactitud la señal analógica original. La magnitud del error depende de cuánto cambia la señal analógica durante la toma de la muestra.

La fig. 15-4a muestra el esquema de un circuito de muestreo y retención. El FET funciona como un interruptor analógico sencillo. Cuando se enciende, Q_1 proporciona una trayectoria de baja impedancia para depositar el voltaje de la muestra analógica a través del capacitor C_1. El tiempo durante el cual Q_1 está activo o encendido se llama *tiempo de abertura* o *tiempo de adquisición*. En esencia, C_1 es el circuito de retención. Cuando Q_1 está apagado, C_1 no tiene tra-

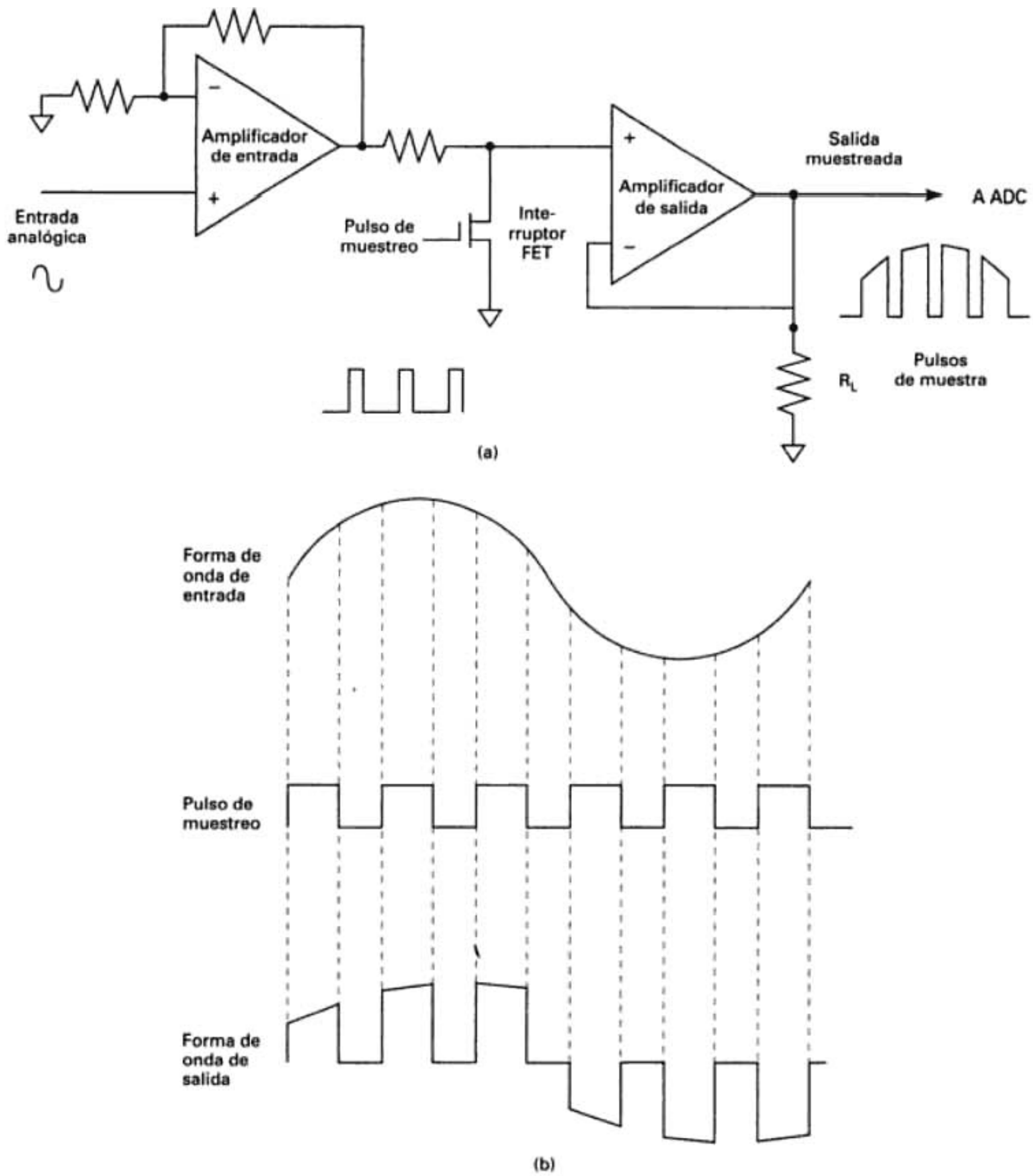


FIGURA 15-3 (a) Circuito de muestreo natural; (b) formas de onda de entrada y salida

yectoria completa para descargarse y, en consecuencia, guarda el voltaje muestreado. El *tiempo de almacenamiento* del capacitor se llama *tiempo de conversión* de analógico a digital, o A/D, porque es durante este tiempo cuando el ADC convierte el voltaje de muestra en código PCM. El tiempo de adquisición debería ser muy corto, para asegurar que el cambio sea mínimo en la señal analógica mientras se deposita en C_1. Si la entrada al ADC cambia mientras se está haciendo la conversión, resulta una *distorsión por abertura*. Así, al tener un tiempo corto de abertura y mantener relativamente constante la entrada al ADC, el circuito de muestreo y retención puede reducir la distorsión por abertura. El muestreo de parte plana introduce menor distorsión de abertura que el muestreo natural, y requiere un convertidor analógico a digital más lento.

siendo C= capacitancia máxima (faradios)

i = corriente máxima de salida de Z_1, 10 mA

dv = cambio máximo de voltaje a través de C_1, igual a 10 V

dt = tiempo de carga, igual al tiempo de abertura, 10 μs

Por consiguiente,
$$C_{máx} = \frac{(10\text{ mA})(10\ \mu\text{s})}{10\text{ V}} = 10\text{ nF}$$

La constante de tiempo de carga para C, cuando Q_1 está activado, es
$$\tau = RC$$

en donde τ = una constante de tiempo de carga (segundos)

R = impedancia de salida de Z_1, más la resistencia de Q_1 activado (ohms)

C = valor de la capacitancia de C_1 (faradios)

Esta ecuación se rearregla y de ella se despeja C
$$C_{máx} = \frac{\tau}{R}$$

El tiempo de carga del capacitor C_1 también depende de la exactitud que se desea en el dispositivo. La exactitud porcentual y su constante RC requerida se puede resumir como sigue

Exactitud (%)	Tiempo de carga
10	2.3τ
1	4.6τ
0.1	6.9τ
0.01	9.2τ

Para la exactitud de 1%,
$$C = \frac{10\ \mu\text{s}}{4.6(20)} = 108.7\text{ nF}$$

Para satisfacer las limitaciones de corriente de salida de Z_1, se requirió una capacitancia máxima de 10 nF. Para satisfacer los requisitos de exactitud se requirieron 108.7 nF. Para satisfacer ambos requisitos se debe usar la capacitancia de valor menor. En consecuencia, C_1 no puede ser mayor que 10 nF.

Frecuencia de muestreo

El teorema de muestreo de Nyquist establece la *frecuencia mínima de muestreo* (f_s) que se puede usar en determinado sistema PCM. Para que una muestra se reproduzca con exactitud en el receptor, se debe muestrear cuando menos dos veces cada ciclo de la señal analógica de entrada (f_a). En consecuencia, la frecuencia mínima de muestreo es igual al doble de la frecuencia máxima de la entrada de audio. Si f_s es menor que f_a, se producirá distorsión. A esta distorsión se le llama *distorsión por alias*, o *por doblez en la imagen*. La frecuencia de muestreo mínima de Nyquist es
$$f_s \geq 2f_a \tag{15-1}$$

en donde f_s = frecuencia mínima de muestreo de Nyquist (hertz)

f_a = máxima frecuencia que se debe muestrear (hertz)

En esencia, un circuito de muestreo y retención es un modulador de AM. El interruptor es un dispositivo no lineal que tiene dos entradas: el pulso de muestreo y la señal analógica de entrada. En consecuencia, entre esas dos señales hay *mezclado no lineal* o *heterodinado*. La fig. 15-5a muestra la representación del espectro de salida de un circuito de muestreo y retención en el dominio de frecuencias. En la salida se incluyen las dos entradas originales (de audio y la frecuencia fundamental del pulso de muestreo), sus frecuencias de suma y de diferencia ($f_s \pm f_a$), todas las armónicas de f_s y f_a ($2f_s$, $2f_a$, $3f_s$, $3f_a$, etc.) y sus productos cruzados correspondientes ($2f_s \pm f_a$, $3f_s \pm f_a$, etc.).

Como el pulso de muestreo es una forma repetitiva de onda, está formado por una serie de ondas senoidales relacionadas armónicamente. Cada una de estas ondas senoides tiene ampli-

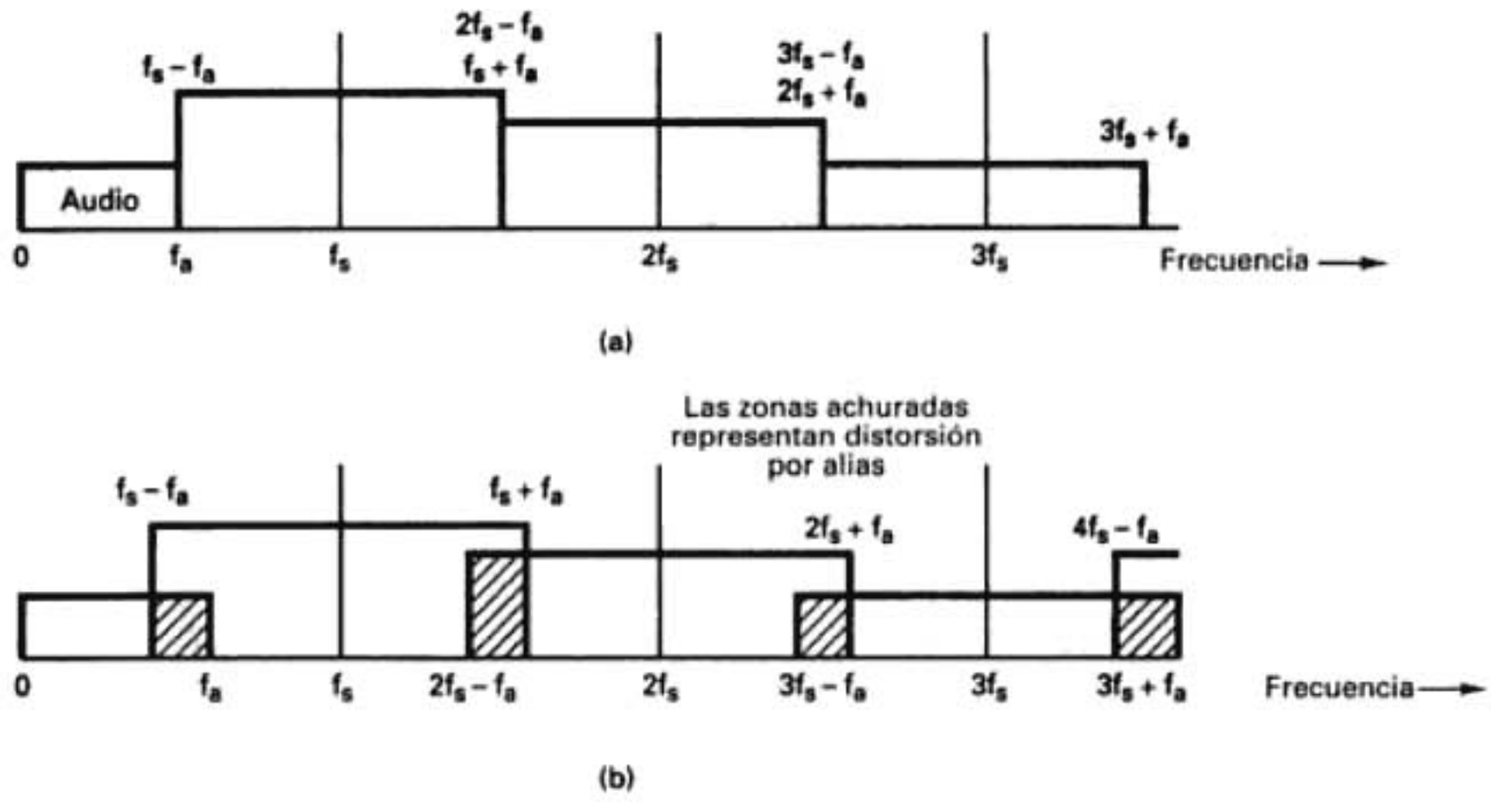


FIGURA 15-5 Espectro de salida para un circuito de muestreo y retención: (a) sin distorsión por alias; (b) con distorsión por alias

tud modulada por la señal analógica, y produce frecuencias de suma y diferencia simétrica respecto a cada una de las armónicas de f_s. Cada frecuencia de suma y diferencia que se genera está separada de su respectiva frecuencia central por f_a. Mientras f_s sea al menos el doble de f_a, ninguna de las frecuencias laterales de alguna armónica entrará en las bandas laterales de otra armónica y no sucede el (o *la*) alias. La fig. 15-5b muestra el resultado de una frecuencia analógica de entrada mayor que $f_s/2$ module a f_s. Las frecuencias laterales de una armónica se pasan o despliegan hacia la banda lateral de otra armónica. La frecuencia que se pasa es un alias o reproducción de la señal de entrada (de ahí el nombre de "distorsión por alias"). Si una frecuencia lateral alias de la primera armónica se pasa al espectro de audio, no se puede eliminar por filtrado, ni por cualquier otra técnica.

Ejemplo 15-2

Determinar la frecuencia mínima de muestreo y la frecuencia de alias producida, en un sistema PCM con frecuencia máxima de entrada de audio de 4 kHz, si se permitiera que una señal de audio de 5 kHz entrara al circuito de muestreo y retención.

Solución De acuerdo con el teorema de muestreo de Nyquist (ecuación 15-1),

$$f_s \geq 2f_a, \quad \text{y, en consecuencia,} \quad f_s \geq 8 \text{ kHz}$$

Si entrara una frecuencia de audio de 5 kHz al circuito de muestreo y retención, se produciría el espectro de salida de la fig. 15-6. Se puede ver que la señal de 5 kHz produce una frecuencia alias de 3 kHz que se ha introducido en el espectro de audio original.

El filtro pasabanda de la entrada, en la fig. 15-2, se llama *filtro antialias*. Su frecuencia superior de corte se escoge de tal modo que no pueda entrar una frecuencia mayor que la mitad de la frecuencia de muestreo al circuito de muestreo y retención, eliminando así la posibilidad de que ocurra la distorsión por alias.

En la PCM se muestrea la señal analógica de entrada, a continuación se convierte en un código binario en serie. El código binario se transmite al receptor, donde se reconvierte a la señal analógica original. Los códigos binarios que se usan para PCM son de n bits, donde n puede ser cualquier entero positivo mayor que 1. En la actualidad se usan *códigos de signo* y *magnitud* para PCM, en donde el *bit más significativo*) es el bit del signo, y los demás bits se usan para la magnitud. La tabla 15-1 muestra un código PCM de n bits, en el que $n = 3$. Se usa el bit más significativo para representar el signo de la muestra (1 lógico = positivo, y 0 lógico = negativo). Los dos bits restantes representan la magnitud. Con dos bits para la magnitud, hay cuatro códigos

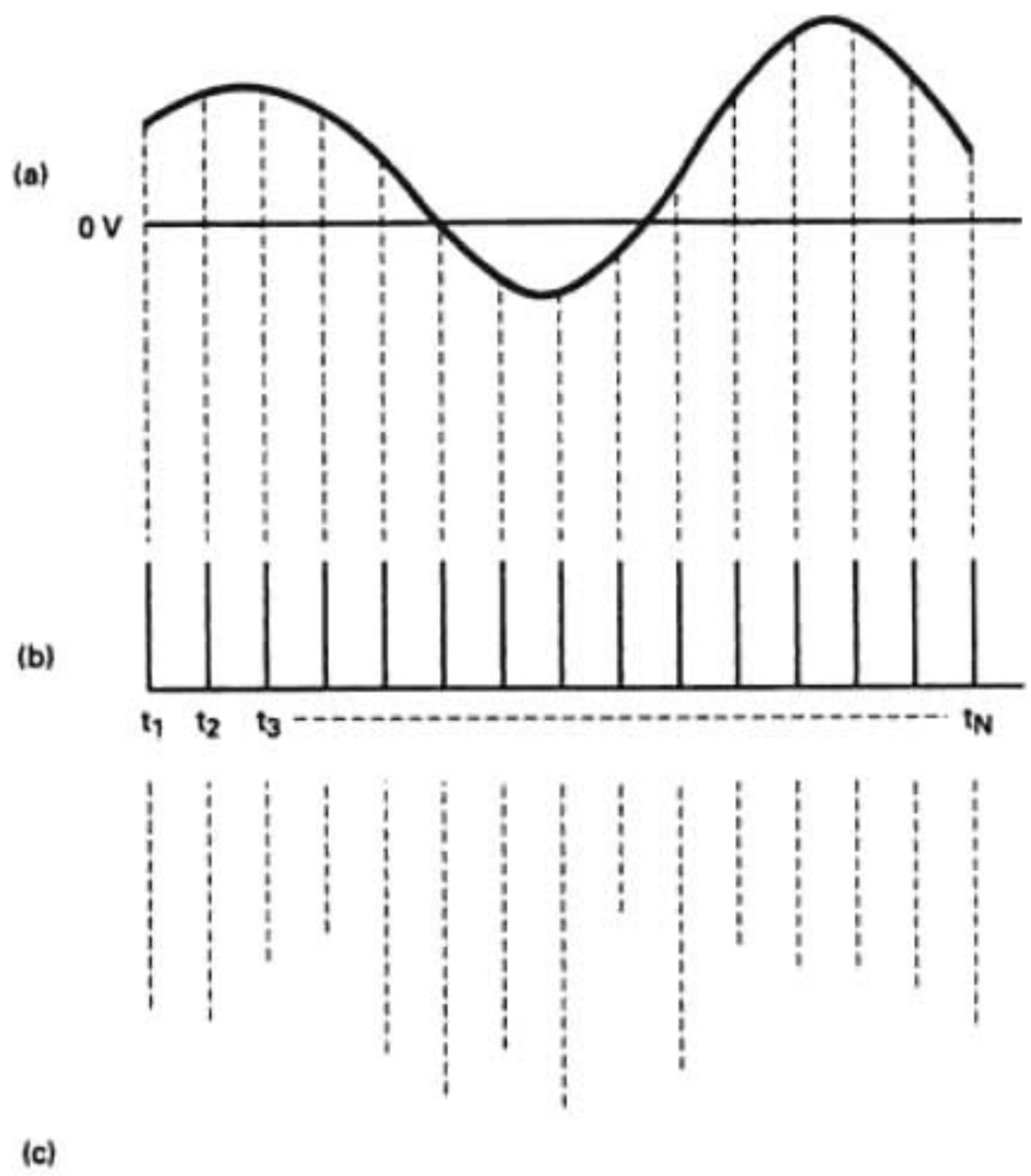


FIGURA 15-9 PAM: (a) señal de entrada; (b) impulsos de muestreo; (c) señal PAM

Para el sistema de la tabla 15-1,

$$DR = \frac{3\ V}{1\ V} = 3$$

Se acostumbra representar el intervalo dinámico (DR, de *dynamic range*) en decibelios y, en consecuencia,

$$DR = 20 \log \frac{V_{máx}}{V_{mín}} = 20 \log \frac{3}{1} = 9.54\ dB$$

Un intervalo dinámico igual a 3 indica que la relación de voltajes máximo a mínimo de la señal decodificada es 3.

Si se desea una resolución menor, como por ejemplo 0.5 V, que mantenga el intervalo dinámico de 3, se debe reducir el voltaje máximo admisible en la entrada, en el mismo factor; la mitad.

$$DR = \frac{1.5}{0.5} = 3$$

Por consiguiente, $V_{máx}$ se reduce en un factor de 2, y el intervalo dinámico es independiente de la resolución. Si la resolución se redujera en un factor de 2 (0.25 V), para mantener la misma amplitud máxima de entrada, habría que duplicar el intervalo dinámico

$$DR = \frac{1.5}{0.25} = 6$$

La cantidad de bits que se usan en un código PCM depende del intervalo dinámico. Con un código PCM de dos bits, la magnitud mínima decodificable tiene clave binaria 01. La magnitud máxima es 11. La relación del códigos binarios máximo a mínimo es 3, igual que el inter-

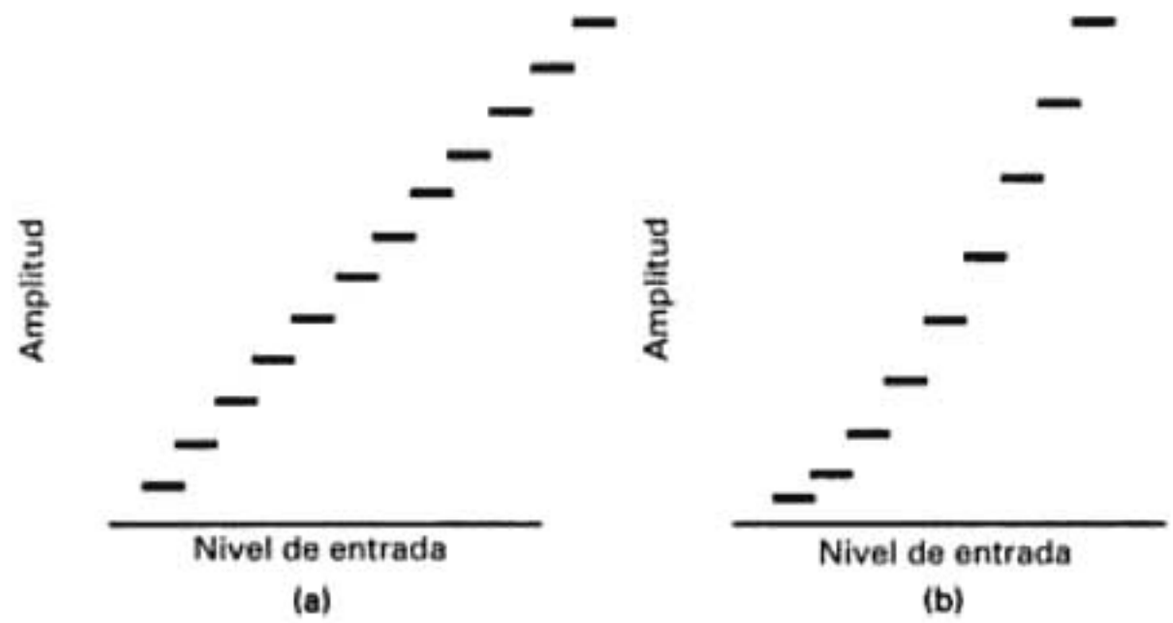


FIGURA 15-10 Codificación: (a) lineal; (b) no lineal

Nótese que con codificación no lineal hay más códigos en la parte inferior de la escala que en la parte superior, y así se aumenta la exactitud para las señales más pequeñas. También nótese que la distancia entre códigos sucesivos es mayor para las señales de mayor amplitud, y aumenta el error por cuantización y se reduce la relación WQR. También, debido a que la señal de $V_{máx}$ a $V_{mín}$ aumenta con la codificación no lineal, el intervalo dinámico es mayor que con un código uniforme. Es evidente que la codificación no lineal es un término medio; se sacrifica la relación SQR para las señales de gran amplitud, para lograr más exactitud para las de baja amplitud, y para lograr un mayor intervalo dinámico. Es difícil fabricar convertidores ADC no lineales y, en consecuencia, se han inventado métodos alternativos de alcanzar los mismos resultados, que se describirán más adelante en este capítulo.

Ruido de canal inactivo

Mientras no hay señal analógica de entrada, la única entrada al muestreador PAM es el ruido aleatorio térmico. Este ruido se llama *de canal inactivo* y se convierte en una muestra PAM de la misma manera que si fuera una señal. En consecuencia, el ADC cuantiza hasta el ruido de entrada. La fig. 15-11 muestra una forma de reducir el ruido de canal inactivo, con un método llamado *cuantización con paso inicial mayor*. En este método, se hace más grande la amplitud del primer intervalo de cuantización que el resto de los incrementos. En consecuencia, el ruido de entrada puede ser muy grande, y seguir siendo cuantizado como código cero, negativo o positivo. El resultado es que se suprime el ruido durante el proceso de codificación.

En los códigos PCM que se han descrito hasta ahora, los códigos de menor magnitud, positivo o negativo, tienen el mismo intervalo de voltaje que los demás códigos (+ o − la mitad de la resolución). A esto se le llama *cuantización con paso homogéneo*. La fig. 15-11 muestra la diferencia entre el ruido de canal inactivo transmitido con un código PCM con paso homogéneo y el transmitido con cuantización con paso inicial mayor. La ventaja de esta última cuantización es menor ruido de canal inactivo. La desventaja es una posible magnitud mayor de Q_e en el intervalo mínimo de cuantización.

Con un código binario reflejado PCM, el ruido residual que fluctúa un poco arriba o abajo de 0 V se convierte en un código PCM cero + o − y, en consecuencia, se elimina. En los sistemas que no usan las dos asignaciones de 0 V, el ruido residual podría hacer que el codificador PCM alternara entre el código 0 y el mínimo + o −. Entonces, el decodificador reproduciría el ruido codificado. Con un código binario reflejado, la mayor parte del ruido residual lo elimina el codificador, de manera inherente.

Métodos de codificación

Hay varios métodos de codificación para cuantizar señales PAM en 2^n niveles. Esos métodos se clasifican de acuerdo con si la operación de codificación se hace con un nivel cada vez, un dígito cada vez o una palabra cada vez.

Codificación de nivel por nivel. Este tipo de codificación compara la señal PAM con una forma de onda de rampa, mientras un contador binario avanza a una velocidad uniforme. Cuando la onda en rampa es igual o mayor que la muestra PAM, el contador contiene el código

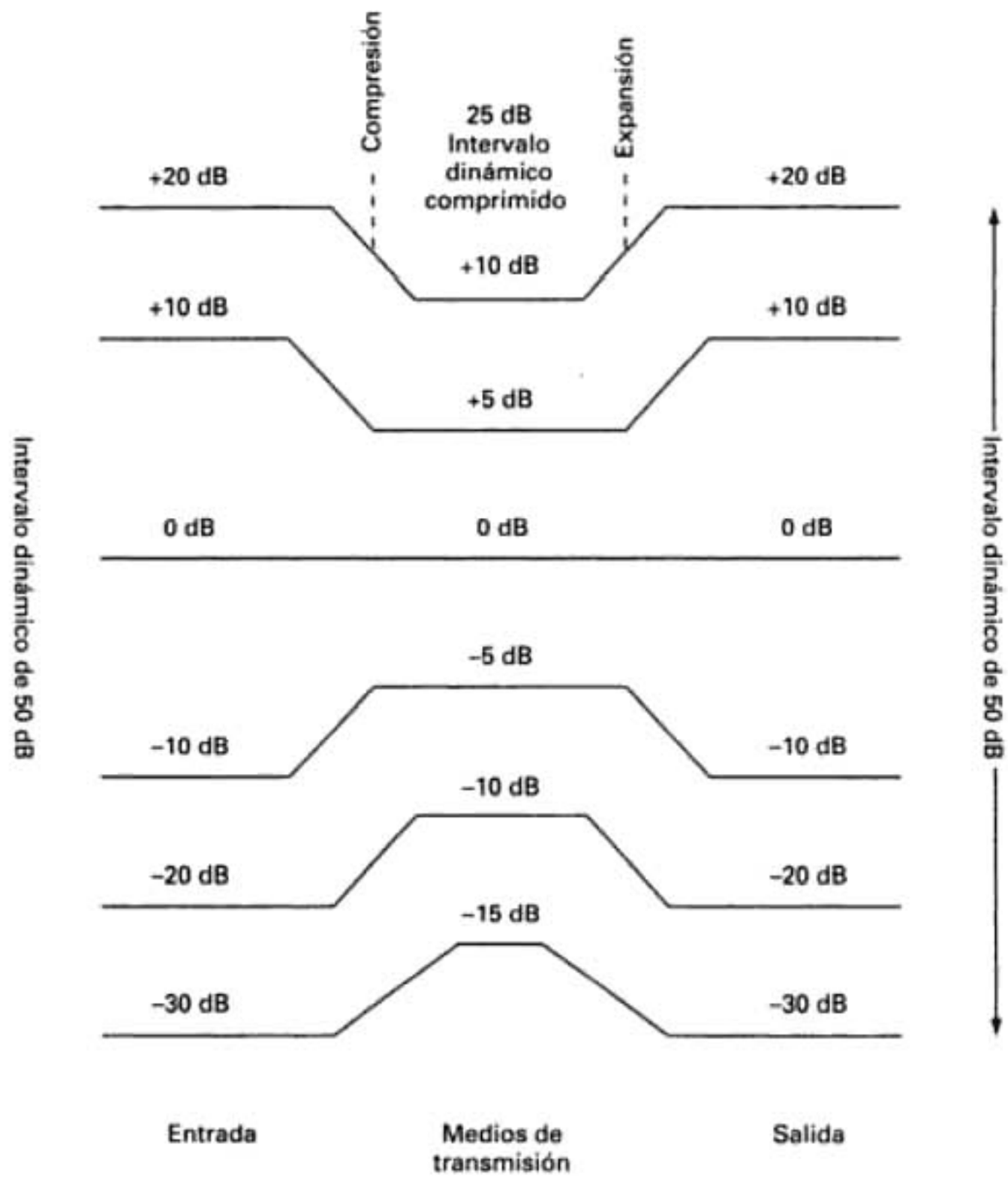


FIGURA 15-12 Proceso básico de compresión-expansión

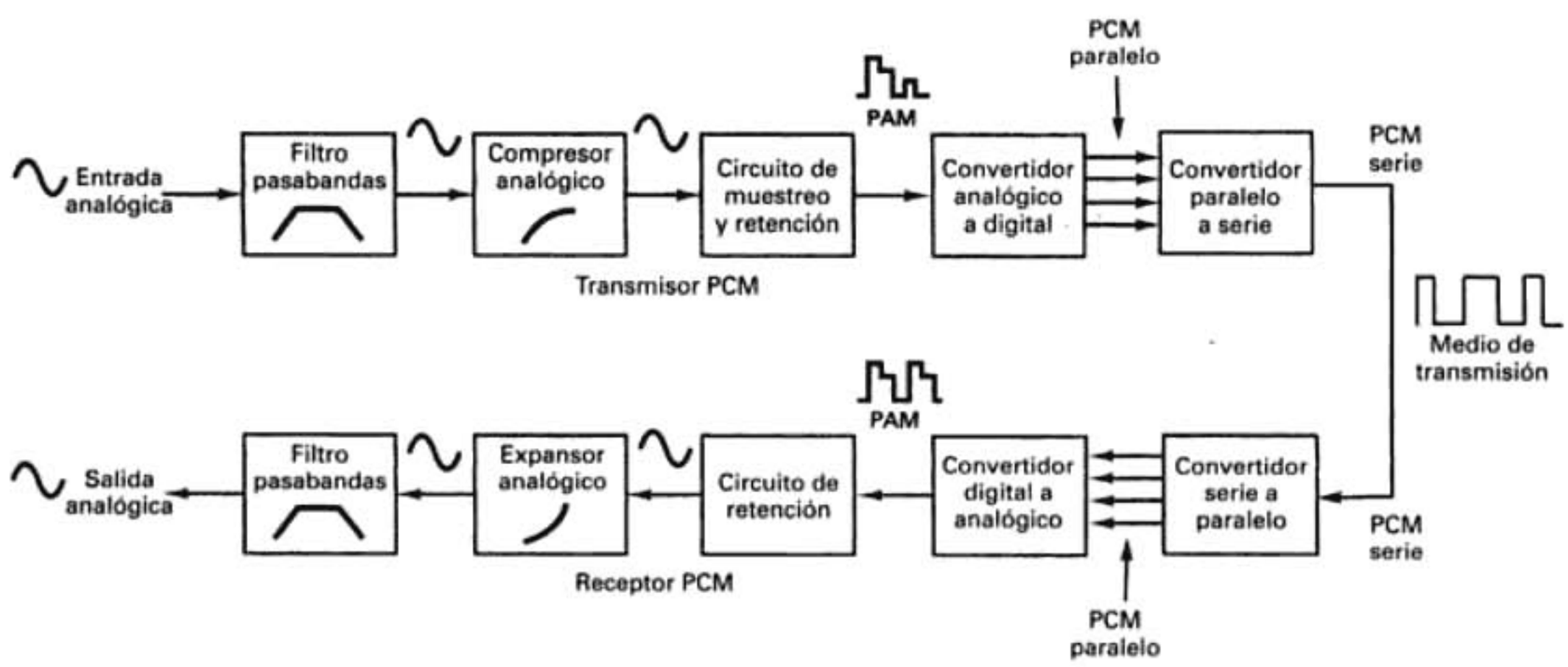


FIGURA 15-13 Sistema PCM con compresión y expansión analógicas

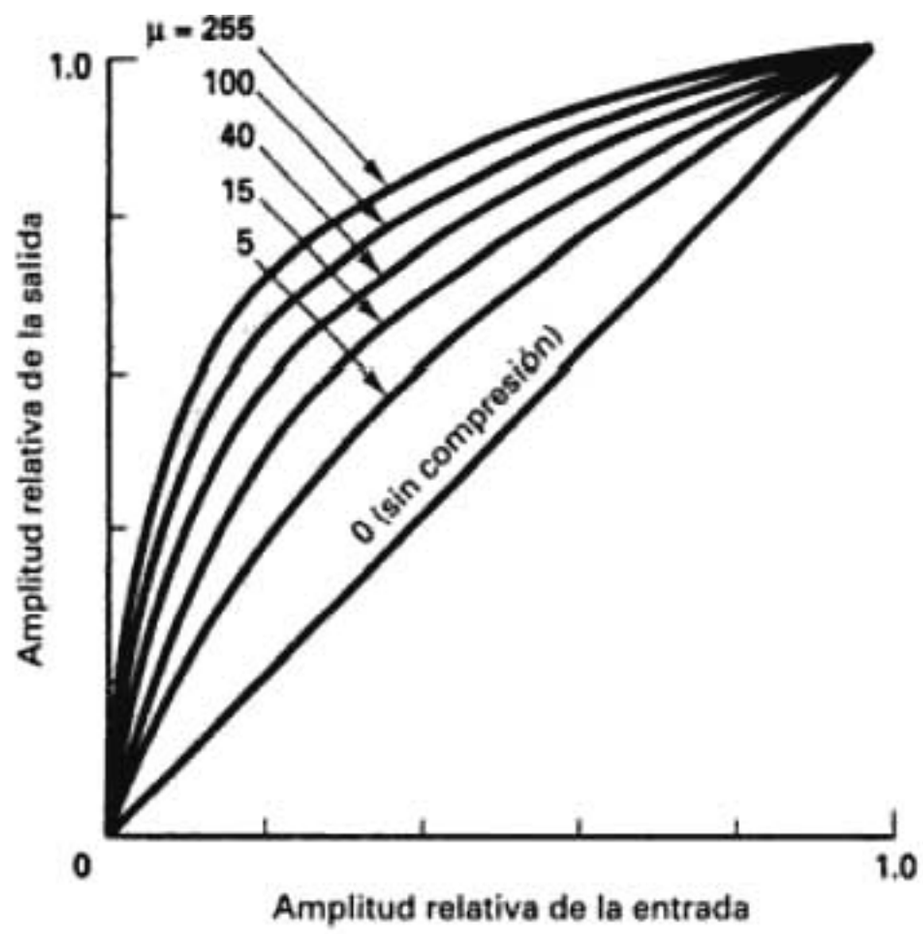


FIGURA 15-14 Características de compresión con ley μ

vierte a un código lineal PCM. En el receptor, el código PCM se convierte en señal PAM, se filtra y a continuación se expande, regresando a sus características de amplitud original en la entrada.

Para distribuciones diferentes de señal se requieren diferentes características de compresión-expansión. Por ejemplo, las señales de voz requieren desempeño SQR relativamente constante dentro de un intervalo dinámico muy ancho, lo que significa que la distorsión debe ser proporcional a la amplitud de la señal para cualquier nivel de señal de entrada. Para esto se requiere una relación de compresión logarítmica. Un código de asignación verdaderamente logarítmico requiere un intervalo dinámico infinito, y una cantidad infinita de códigos PCM, lo cual es imposible. Hay dos métodos de compresión-expansión analógica de uso actual, que aproximan mucho una función logarítmica, y que con frecuencia se les llama códigos *log-PCM*, o PCM logarítmicos. Son *compresión-expansión de ley* μ, y *de ley* A.

Compresión-expansión de ley μ. En Estados Unidos y en Japón se usa compresión-expansión de ley μ. Para esa ley, la característica de compresión es

$$V_{sal} = \frac{V_{máx} \times \ln(1 + \mu V_{ent}/V_{máx})}{\ln(1 + \mu)} \tag{15-8}$$

en donde $V_{máx}$ = amplitud máxima de la entrada analógica no comprimida (volts)
V_{ent} = amplitud de la señal de entrada en determinado momento (volts)
μ = parámetro para definir la cantidad de compresión (adimensional)
V_{sal} = amplitud de la salida comprimida (volts)

La fig. 15-14 muestra la compresión para varios valores de μ. Nótese que mientras mayor es μ, la compresión es mayor. También, nótese que para μ = 0, la curva es lineal (no hay compresión).

El parámetro μ determina el intervalo de potencia en el cual la relación SQR es relativamente constante. La transmisión de voz requiere un intervalo dinámico mínimo de 40 dB, y un código PCM de siete bits. Para una relación SQR relativamente constante y un intervalo dinámico de 40 dB, se requiere que μ = 100 o más. Los primeros sistemas de transmisión digital de Bell System usaban un código PCM de siete bits con μ = 100. Los sistemas más recientes de transmisión digital usan códigos PCM de ocho bits y μ = 255.

Ejemplo 15-4

Determinar, para un compresor con μ = 255:

(a) La ganancia de voltaje para los siguientes valores relativos de V_{ent}: $V_{máx}$, $0.75V_{máx}$, $0.5V_{máx}$ y $0.25V_{máx}$.

(b) El voltaje comprimido de salida para un voltaje máximo de entrada de 4 V.

(c) Los intervalos dinámicos y la compresión en la entrada y la salida.

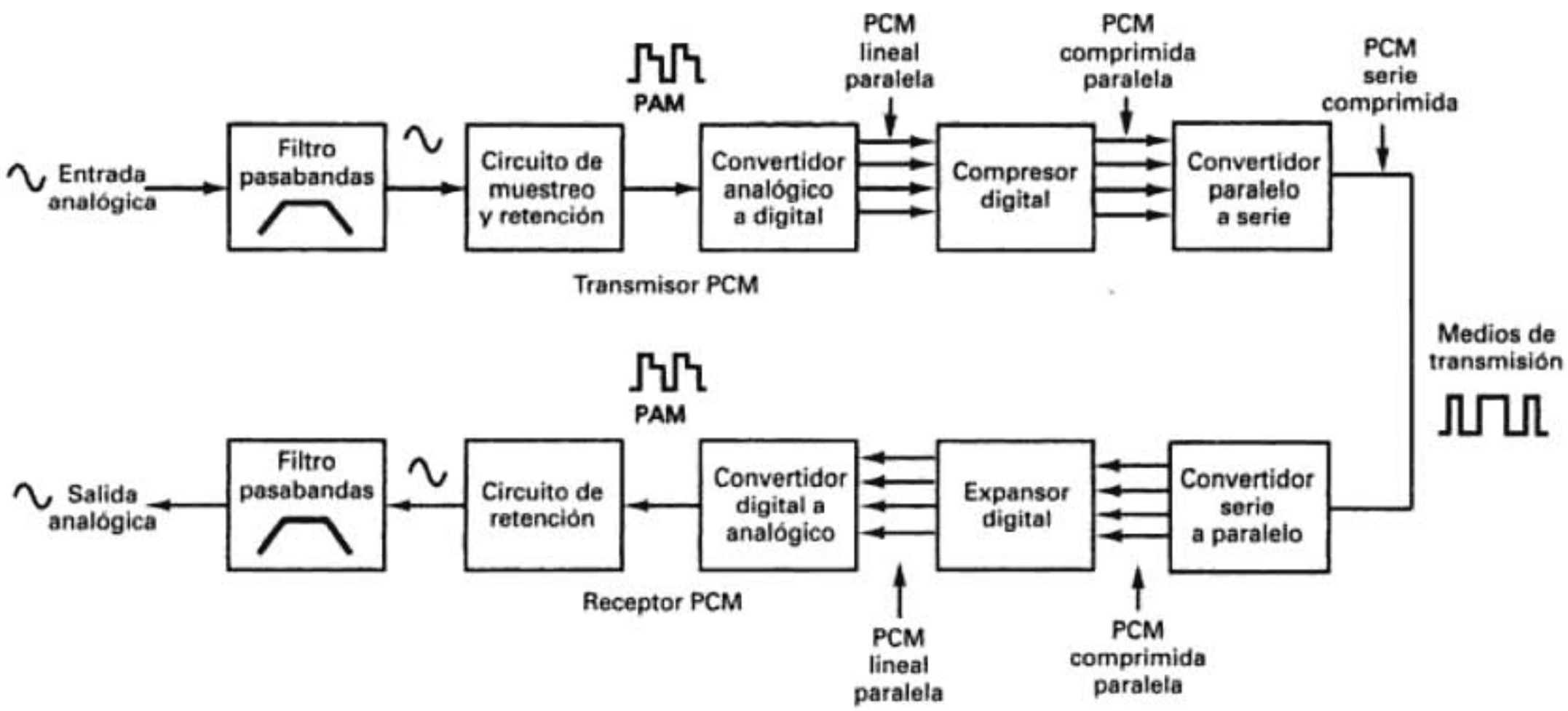


FIGURA 15-15 Sistema PCM con compresión y expansión digital

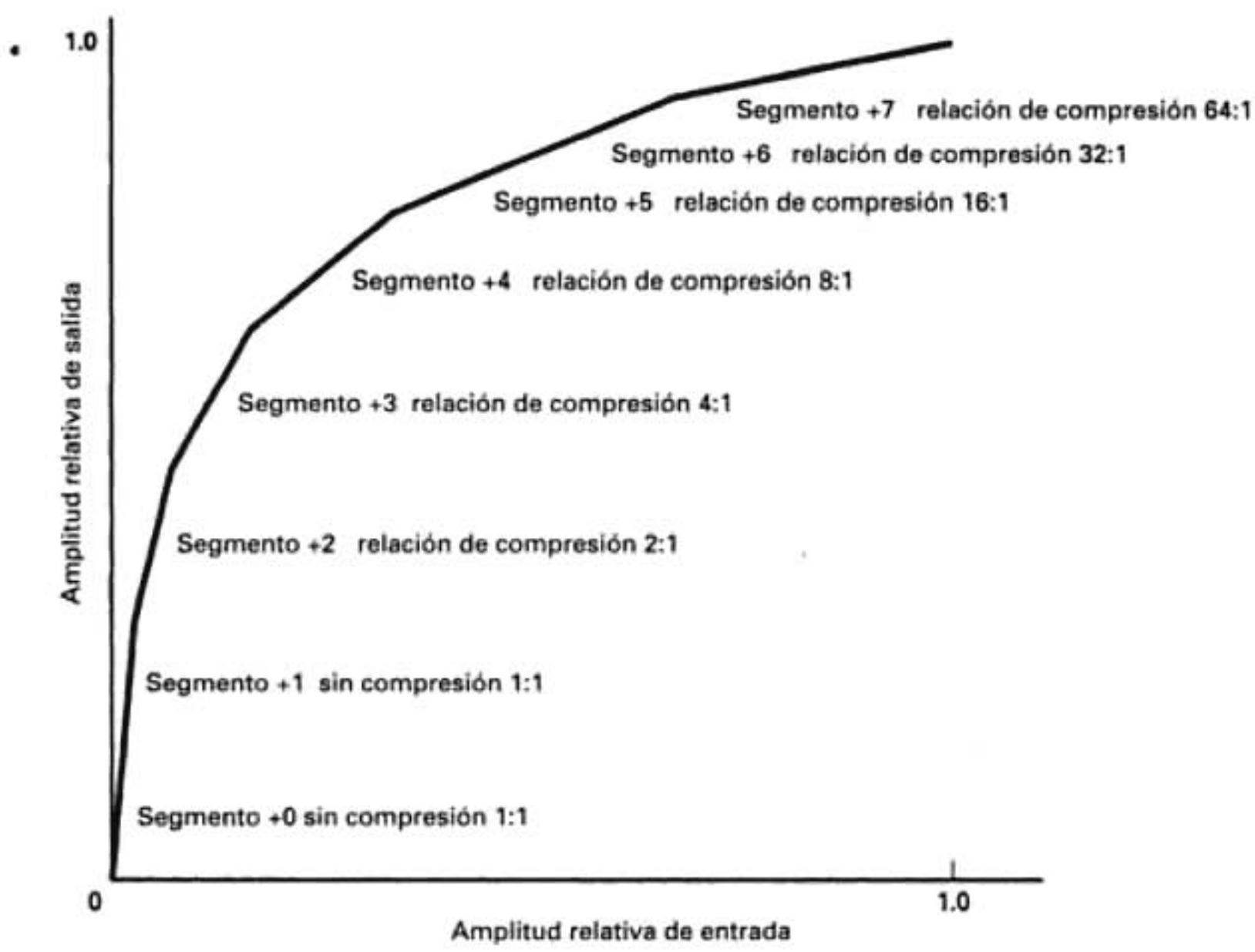


FIGURA 15-16 Características de la compresión μ255 (sólo valores positivos)

hay 16 segmentos, 8 positivos y 8 negativos, a este esquema se le suele llamar *compresión de 13 segmentos*, porque la curva de los segmentos +0, +1, −0 y −1 es una recta con pendiente constante, y con frecuencia se considera como un solo segmento.

El algoritmo de compresión-expansión digital para un código comprimido de 12 bits lineal a 8 bits es, en realidad, bastante simple. El código comprimido de 8 bits consiste en un bit de signo, un identificador de segmento de 3 bits y un código de magnitud de 4 bits que identifica el *intervalo de cuantización* dentro del segmento especificado (véase la fig. 15-17a).

Bit de signo 1 = + 0 = −	Identificador de segmento de 3 bits	Intervalo de cuantización de 4 bits
		A B C D
	000 a 111	0000 a 1111

(a)

Medios de transmisión

PCM codificada — Transmitir

Segmento	Código lineal de 12 bits	Código comprimido de 8 bits
0	s0000000ABCD	s000ABCD
1	s0000001ABCD	s001ABCD
2	s000001ABCDX	s010ABCD
3	s00001ABCDXX	s011ABCD
4	s0001ABCDXXX	s100ABCD
5	s001ABCDXXXX	s101ABCD
6	s01ABCDXXXXX	s110ABCD
7	s1ABCDXXXXXX	s111ABCD

(b)

Recibir — PCM decodificada

Código comprimido de 8 bits	Código recuperado de 12 bits	Segmento
s000ABCD	s0000000ABCD	0
s001ABCD	s0000001ABCD	1
s010ABCD	s000001ABCD1	2
s011ABCD	s00001ABCD10	3
s100ABCD	s0001ABCD100	4
s101ABCD	s001ABCD1000	5
s110ABCD	s01ABCD10000	6
s111ABCD	s1ABCD100000	7

(c)

FIGURA 15-17 Compresión-expansión digital de 12 a 8 bits: (a) formato de código comprimido μ255 de 8 bits; (b) tabla de codificación μ255; (c) tabla de decodificación μ255

En la tabla de codificación μ255 (es decir, con $\mu = 255$) que muestra la fig. 15-17b, se truncan las posiciones de bit designadas con X durante la compresión y, en consecuencia, se pierden. Los bits designados con A, B, C y D se transmiten tal como son. También el bit de signo se transmite tal cual es. Nótese que para los segmentos 0 y 1, se duplican exactamente los 12 bits en la salida del decodificador (fig. 15-17c), mientras que para el segmento 7 sólo se recuperan los seis bits más significativos. Como hay 11 bits de magnitud, hay 2048 códigos posibles. Hay en el segmento 0 y 1 16 códigos. En el segmento 2 hay 32 códigos; el segmento 3 tiene 64. Cada segmento sucesivo que comienza con el segmento 3 tiene el doble de códigos que el anterior. En cada uno de los ocho segmentos sólo se pueden recuperar 16 códigos de 12 bits. En consecuencia, en los segmentos 0 y 1 no hay compresión (de los 16 códigos posibles, los 16 se pueden recuperar). En el segmento 2 hay una relación de compresión de 2 a 1: 32 códigos posibles de transmisión y 16 códigos posibles recuperado. En el segmento 3 hay una relación de compresión de 4 a 1: 64 códigos posibles de transmisión y 16 códigos posibles recuperados. La relación de compresión se duplica en cada segmento sucesivo. En el segmento 7, la relación de compresión es 1024/16, o sea 64 a 1.

El proceso de compresión se hace como sigue: se muestrea la señal analógica y se convierte en un código de signo y magnitud de 12 bits, lineal. El bit de signo pasa en forma directa al código de 8 bits. El segmento se determina contando la cantidad de ceros delanteros, o ceros a la izquierda, en la parte del código correspondiente a la magnitud, de 11 bits, que comienza con el bit más significativo. Se resta de 7 la cantidad de ceros a la izquierda (que no debe ser mayor de 7). El resultado es el número de segmento, que se convierte a un número binario de 3 bits y se sustituye como identificador de segmento en el código de 8 bits. Los cuatro bits de magnitud, A, B, C y D, son el intervalo de cuantización, y se sustituyen en los cuatro bits menos significativos del código comprimido de 8 bits.

En esencia, los segmentos 2 a 7 se subdividen en subsegmentos más pequeños. Cada segmento tiene 16 subsegmentos, que corresponden a las 16 condiciones posibles de los bits A, B,

Segmento	Código lineal de 12 bits		Código expandido de 12 bits	Subsegmento
5	s00111111111 . . . . . . s00111110000	16 : 1	s00111111000	15
5	s00111101111 . . . . . . s00111100000	16 : 1	s00111101000	14
5	s00111011111 . . . . . . s00111010000	16 : 1	s00111011000	13
5	s00111001111 . . . . . . s00111000000	16 : 1	s001110010000	12
5	s00110111111 . . . . . . s00110110000	16 : 1	s00110111000	11
5	s00110101111 . . . . . . s00110100000	16 : 1	s00110101000	10
5	s00110011111 . . . . . . s00110010000	16 : 1	s00110011000	9
5	s00110001111 . . . . . . s00110000000	16 : 1	s00110001000	8
5	s00101111111 . . . . . . s00101110000	16 : 1	s00101111000	7
5	s00101101111 . . . . . . s00101100000	16 : 1	s00101101000	6
5	s00101011111 . . . . . . s00101010000	16 : 1	s00101011000	5
5	s00101001111 . . . . . . s00101000000	16 : 1	s00101001000	4
5	s00100111111 . . . . . . s00100110000	16 : 1	s00100111000	3
5	s00100101111 . . . . . . s00100100000	16 : 1	s00100101000	2
5	s00100011111 . . . . . . s00100010000	16 : 1	s00100011000	1
5	s00100001111 . . . . . . s00100000000	16 : 1	s00100001000	0
	s001ABCD----			

(b)

FIGURA 15-18 (Continuación) b) segmento 5

can durante la compresión. En el receptor, el decodificador debe determinar cuáles eran esos bits. Las posibilidades son cualquier código de 00000 a 11111. El tanteo lógico es 10000, que es aproximadamente la mitad de la magnitud máxima. En consecuencia, el error máximo de compresión es un poco mayor que la mitad de la magnitud de este segmento.

Ejemplo 15-5

Determinar el código lineal de 12 bits, el código comprimido de 8 bits y el código recuperado de 12 bits para tener una resolución de 0.01 V con voltajes analógicos de muestra de:

(a) 0.05 V.
(b) 0.32 V.
(c) 10.23 V.

Solución a) Para determinar el código lineal de 12 bits para 0.05 V, simplemente se divide el voltaje de muestra entre la resolución, y se convierte el resultado en un número binario para signo y magnitud de 12 bits.

Segmento	Código lineal de 12 bits		Código expandido de 12 bits	Subsegmento
2	s00000111111 ... s00000111110	2 : 1	s00000111111	15
2	s00000111101 ... s00000111100	2 : 1	s00000111101	14
2	s00000111011 ... s00000111010	2 : 1	s00000111011	13
2	s00000111001 ... s00000111000	2 : 1	s00000111001	12
2	s00000110111 ... s00000110110	2 : 1	s00000110111	11
2	s00000110101 ... s00000110100	2 : 1	s00000110101	10
2	s00000110011 ... s00000110010	2 : 1	s00000110011	9
2	s00000110001 ... s00000110000	2 : 1	s00000110001	8
2	s00000101111 ... s00000101110	2 : 1	s00000101111	7
2	s00000101101 ... s00000101100	2 : 1	s00000101101	6
2	s00000101011 ... s00000101010	2 : 1	s00000101011	5
2	s00000101001 ... s00000101000	2 : 1	s00000101001	4
2	s000000100111 ... s00000100110	2 : 1	s00000100111	3
2	s00000100101 ... s00000100100	2 : 1	s00000100101	2
2	s000000100011 ... s00000100010	2 : 1	s00000100011	1
2	s00000100001 ... s00000100000	2 : 1	s00000100001	0
	s000001ABCD-			

(c)

FIGURA 15-18 (Continuación) (c) segmento 2

Código lineal de 12 bits

$$\frac{0.05\ \text{V}}{0.01\ \text{V}} = 5 = \underset{s}{1}\ \underbrace{0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0}_{\text{magnitud (número binario de 11 bits)}}\ \overset{A}{1}\ \overset{B}{0}\ \overset{C}{1}\ \overset{D}{0}$$

Código comprimido de 8 bits

1	0 0 0 0 0 0 0 0	0 1 0 1
s	(7 − 7 = 0 o 000)	A B C D
1	0 0 0	0 1 0 1
↑		
bit de signo (+)	identificador de unidad (segmento 0)	intervalo de cuantización (5)

Código recuperado de 12 bits

```
1              0 0 0   0 1 0 1
s (7 − 0 = 7 ceros a la izquierda) A  B  C  D
1 0 0 0 0 0 0 0 0 1 0 1
↑
bit de signo  identificador de      intervalo de
              segmento; determina   cuantización
              la cantidad de ceros
              a la izquierda
```

Como se puede ver, el código recuperado de 12 bits es exactamente igual al código lineal original, de 12 bits. Esto es cierto para todos los códigos en los segmentos 0 y 1. En consecuencia, en estos dos segmentos no hay error de compresión.

b) Para la muestra de 0.32 V.

Código lineal de 12 bits

$$\frac{0.32\ \text{V}}{0.01\ \text{V}} = 32 = \underset{\text{s}}{1}\ \underbrace{0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0}_{\text{magnitud}}$$

Código comprimido de 8 bits

```
1 0 0 0 0 0 1     0 0 0 0 0
s     (7 − 5 = 2 o 010)  A B C D X
1           0 1 0     0 0 0 0 ↑
(+)         (segmento 2)        truncado
```

Código recuperado de 12 bits

```
1          0 1 0          0 0 0 0
s     (7 − 2 = 5 ceros a la izquierda)    A B C D X
1 0 0 0 0 0 1          0 0 0 0 1
                ↑                    ↑
              insertado            insertado
```

Nótense los dos unos insertados en el código decodificado de 12 bits. El bit menos significativo se determina con la tabla de decodificación, en la fig. 15-17. El uno intercalado en la posición 6 de bit se eliminó durante la conversión de 12 bits a 8 bits. La transmisión de este bit es redundante, porque si no fuera un uno, la muestra no estaría en el segmento 2. En consecuencia, en todos los segmentos a excepción del 0 se inserta un uno en forma automática, después de los ceros reinsertados. En esta muestra hay un error en el voltaje recibido, igual a la resolución, 0.01 V. En el segmento 2, por cada 2 códigos posibles de 12 bits, sólo hay un código recuperado de 12 bits. En esta forma se realizó una compresión de codificación de 2 a 1.

c) Para determinar los códigos para 10.23 V, el proceso es igual.

Código lineal de 12 bits

```
1   0   1 1 1 1 1   1 1 1 1 1
↑       ↑
signo   |           ABCD      truncados
        omitido
      (7 − 1 = 6)
```

Código comprimido de 8 bits

```
1    1 1 0   1 1 1 1
↑
signo  segmento  A B C D
         6
```

Código recuperado de 12 bits

```
1   0   1   1 1 1 1     1 0 0 0 0
↑       ↑                 ↑
signo   |     ABCD        insertado
        insertado
```

de corto plazo, y con frecuencia las formas de onda decodificadas, en función del tiempo, se asemejan en forma vaga a la señal original de entrada. Los codificadores de voz no se pueden usar en aplicaciones donde haya otras señales además de la voz, como por ejemplo las señales de salida de módems de datos en banda de voz. Los codificadores de voz suelen producir habla de sonido *no natural* y, en consecuencia, se usan en general para información grabada, como por ejemplo mensajes de "número equivocado", transmisión en clave de voz por circuitos telefónicos analógicos, señales de salida de computadora y juegos educativos.

El objeto de un codificador de voz es codificar la cantidad mínima de información hablada necesaria para reproducir un mensaje inteligible, con menos bits que los necesarios para un codificador y decodificador convencional. Se usan principalmente en aplicaciones de ancho de banda limitado. En esencia, hay tres técnicas disponibles para codificar voz: *codificador de voz de canal*, el *codificador formante de voz* y el *codificador predictivo lineal*.

Codificadores de voz de canal. El primer codificador de voz de canal fue desarrollado por Homer Dudley en 1928. Este aparato comprimía las formas de onda de la conversación convencional en una señal analógica con un ancho aproximado de banda de 300 Hz. Los codificadores de voz digitales actuales funcionan a menos de 2 kbps. Los codificadores de voz digitales de canal usan filtros pasabanda para separar la forma de onda de voz en *subbandas* más angostas. Cada subbanda se rectifica de onda completa, se filtra y a continuación se codifica digitalmente. La señal codificada se transmite al receptor de destino, donde se decodifica. Hablando en general, la calidad de la señal a la salida de uno de estos codificadores de voz es bastante mala. Sin embargo, algunos de los más avanzados trabajan a 2400 bps y pueden producir una conversación muy inteligible, aunque suena un poco a "sonido sintético".

Codificadores formantes de voz. Un codificador formante de voz aprovecha que la densidad espectral de corto plazo de las señales de voz casi nunca se distribuye uniformemente por todo el espectro de banda de voz, de 300 a 3000 Hz. En lugar de ello, el espectro de potencias de la mayor parte de la energía de voz se concentra en 3 o 4 frecuencias máximas, llamadas *formantes*. Un codificador formante de voz tan sólo determina el lugar de esos máximos y codifica y transmite sólo la información con los componentes a corto plazo más importantes. Por consiguiente, los codificadores formantes de voz pueden trabajar a menores frecuencias de bits y requieren entonces menores anchos de banda. A veces, los codificadores formantes de voz tienen problemas de seguimiento de cambios en las formantes. Sin embargo, una vez identificadas las formantes, uno de estos codificadores de voz pueden transferir habla inteligible a menos de 1000 bps.

Codificadores predictivos lineales. Un codificador predictivo lineal extrae las partes más importantes de la información de voz, en forma directa de la forma de onda en función del tiempo, más que del espectro de frecuencias, como sucede con los codificadores de voz de canal y formantes. Un codificador predictivo lineal produce un modelo variable en el tiempo, de la *excitación del tracto vocal* y la función de transferencia, directamente de la forma de onda de la voz. En el extremo receptor hay un *sintetizador* que reproduce el habla pasando la excitación especificada a través de un modelo matemático del tracto vocal. Los codificadores predictivos lineales permiten obtener un habla de sonido más natural que los codificadores de voz de canal o formantes. Los codificadores predictivos lineales suelen codificar y transmitir la voz entre 1.2 y 2.4 kbps.

Velocidad de PCM de la línea

La velocidad de la línea es tan sólo la frecuencia con que los bits PCM serie salen sincronizados del transmisor a la línea de transmisión, o la frecuencia con la que los bits PCM se sacan sincronizados de la línea de transmisión y pasan al receptor. La velocidad de la línea es directamente proporcional a la frecuencia de muestreo y a la cantidad de bits en el código PCM comprimido. La ecuación de definición es

$$\text{velocidad de la línea} = \frac{\text{muestras}}{\text{segundo}} \times \frac{\text{bits}}{\text{muestra}} \tag{15-11}$$

en donde la velocidad de la línea es la velocidad de transmisión de bits (bps). Por ejemplo, un sistema PCM de monocanal, con una frecuencia de muestreo f_s 8000 muestras/segundo, y un código PCM comprimido de ocho bits necesitaría lo siguiente

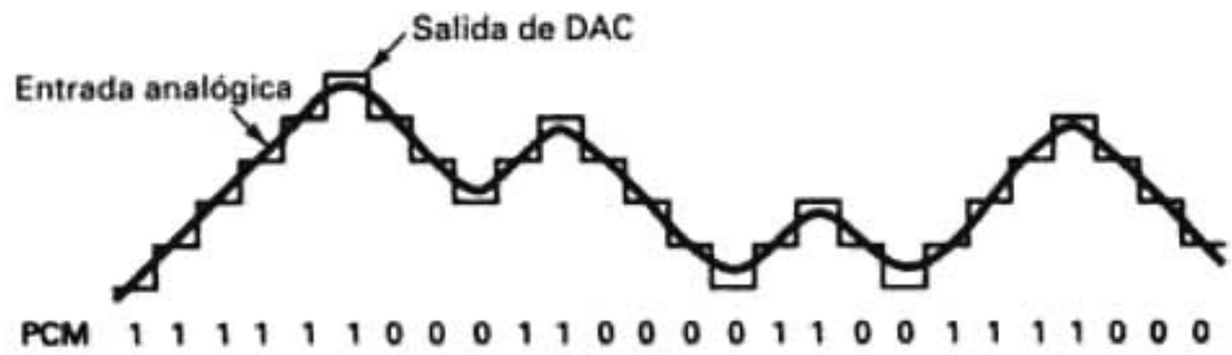


FIGURA 15-20 Funcionamiento ideal de un codificador por modulación delta

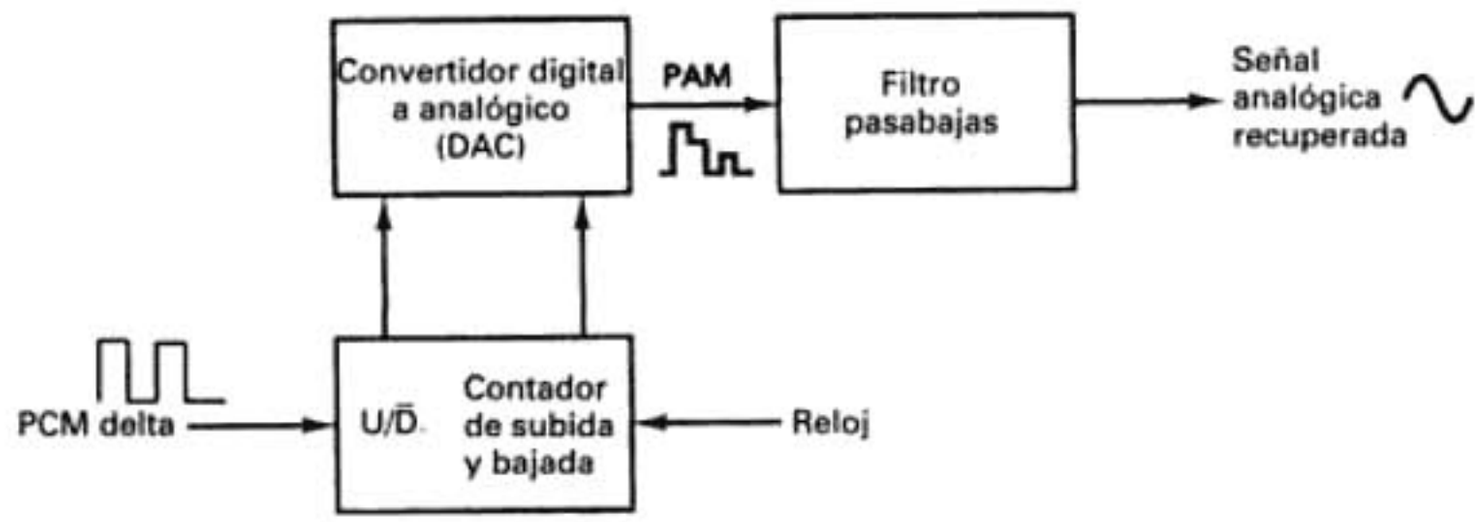


FIGURA 15-21 Receptor con modulación delta

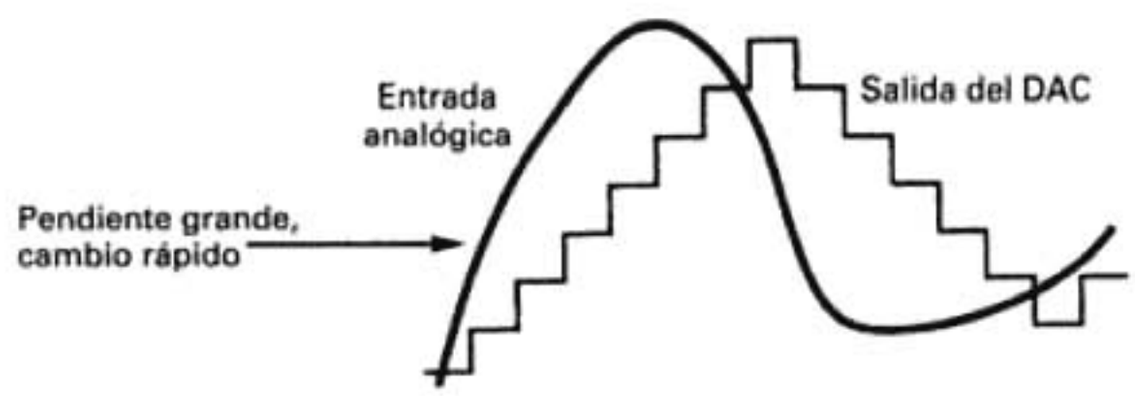


FIGURA 15-22 Distorsión por sobrecarga de pendiente

salida del DAC es mayor que la muestra analógica; a continuación el contador comenzará a contar hacia abajo, hasta que la salida del DAC sea menor que la amplitud de la muestra. En el caso idealizado de la fig. 15-20, la salida DAC sigue a la señal de entrada. Cada vez que la cuenta de subida y bajada se incrementa, se transmite un 1 lógico, y cada vez que se decrementa, se transmite un 0 lógico.

Receptor con modulación delta

La fig. 15-21 muestra el diagrama de bloques de un receptor con modulación delta. Se puede ver que es casi idéntico al transmisor, con la excepción del comparador. Al recibir los unos y ceros lógicos, el contador se incrementa o decrementa en forma correspondiente. En consecuencia, la salida del DAC del decodificador es idéntica a la del DAC en el transmisor.

En la modulación delta, cada muestra sólo requiere la transmisión de un bit y, por consiguiente, las frecuencias de bits asociadas con esta modulación son menores que las de los sistemas PCM convencionales. Sin embargo, la modulación delta tiene dos problemas que no se presentan con la PCM convencional: sobrecarga de pendiente y ruido granular.

Sobrecarga de pendiente. La fig. 15-22 muestra lo que sucede cuando una señal de entrada analógica cambia con una rapidez mayor que la que puede sostener el DAC. La pendiente de la señal analógica es mayor que la que puede mantener el modulador delta, a lo cual se llama *sobrecarga de pendiente*. Si se aumenta la frecuencia del reloj se reduce la probabilidad

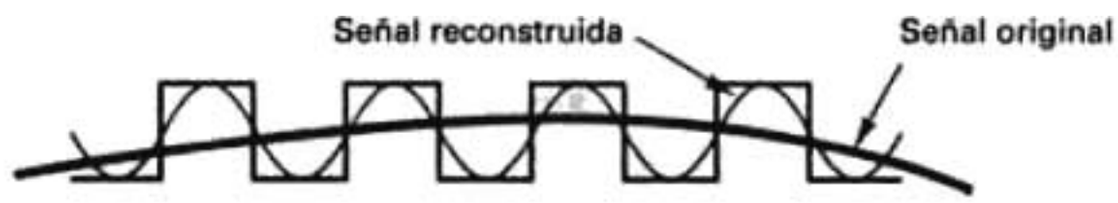


FIGURA 15-23 Ruido granular

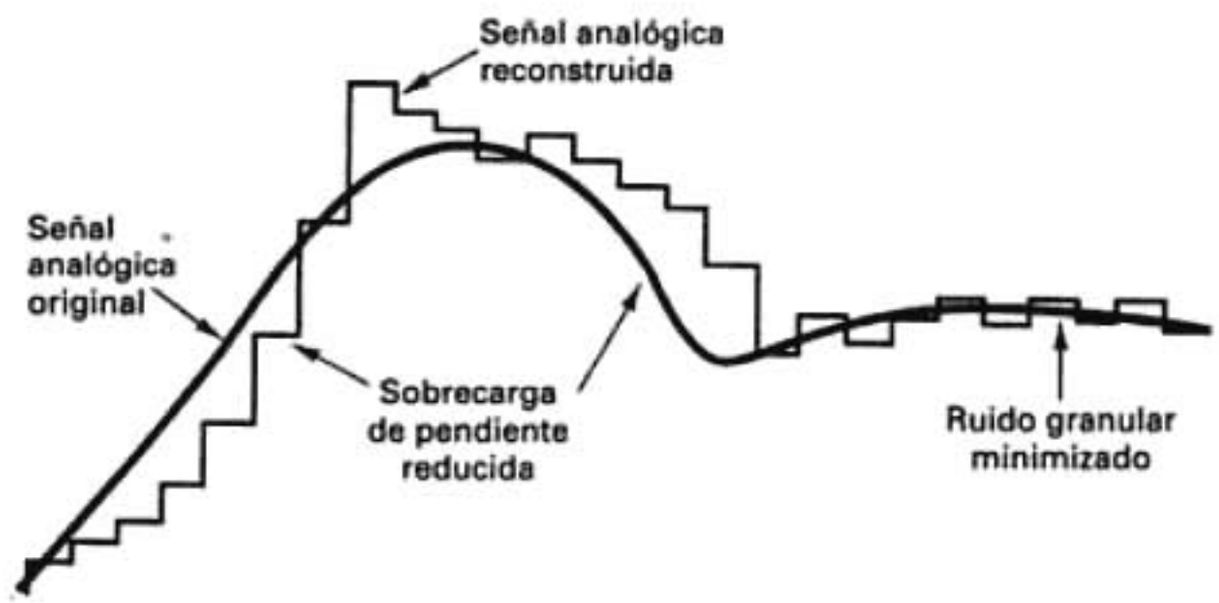


FIGURA 15-24 Modulación delta adaptativa

que haya sobrecarga de pendiente. Otra forma de evitarla es aumentar el tamaño del incremento mínimo.

Ruido granular. La fig. 15-23 muestra la diferencia entre las señales original y reconstruida, asociada con un sistema de modulación delta. Se aprecia que cuando la señal analógica original de entrada tiene una amplitud relativamente constante, la señal reconstruida tiene variaciones que no había en la señal original. A esto se le llama *ruido granular*. En la modulación delta, el ruido granular es análogo al ruido de cuantización en la PCM convencional.

El ruido granular se puede reducir disminuyendo el tamaño de escalón. Así, para reducirlo, se necesita una resolución pequeña, y para reducir la posibilidad de que se presente una sobrecarga de pendiente, se necesita una resolución grande. Es obvio que se necesita un compromiso entre las dos resoluciones.

El ruido granular domina más en señales analógicas que tienen pendientes graduales, y cuyas amplitudes sólo varían en una pequeña cantidad. La sobrecarga de pendiente es más frecuente en señales analógicas que tienen grandes pendientes, o cuyas amplitudes varían con rapidez.

PCM DE MODULACIÓN DELTA ADAPTATIVA

La *modulación delta adaptativa* es un sistema de modulación delta en el que el tamaño de escalón del DAC se varía en forma automática, dependiendo de las características de amplitud de la señal analógica de entrada. La fig. 15-24 muestra cómo funciona un modulador delta adaptativo. Cuando la salida del transmisor es una cadena de unos o ceros consecutivos, indica que la pendiente de la salida DAC es menor que la de la señal analógica, sea en dirección positiva o negativa. En esencia, el DAC ha perdido el lugar exacto de las muestras analógicas, y la posibilidad de que haya sobrecarga de pendiente es alta. En un modulador delta adaptativo, después de una cantidad predeterminada de unos o ceros consecutivos, aumenta en forma automática el tamaño del incremento. Después de la siguiente muestra, si la amplitud de la salida del DAC sigue siendo menor que la de la muestra, se aumenta el siguiente escalón todavía más, hasta que al final el DAC alcance a la señal analógica. Cuando se presenta una sucesión de unos y ceros alternativos, quiere decir que es alta la posibilidad de tener ruido granular. En consecuencia, el DAC se regresa en forma automática a su tamaño mínimo de escalón y así reduce la magnitud el ruido de error.

Un algoritmo frecuente para un modulador por delta adaptativo es cuando se presentan tres unos o ceros consecutivos, y se aumenta o disminuye el tamaño del escalón para el DAC, en un factor de 1.5. Se pueden usar varios algoritmos, que dependen de los requisitos particulares del sistema.

MODULACIÓN POR CÓDIGO DE IMPULSOS DIFERENCIAL

En una forma de onda de voz codificada por PCM, con frecuencia hay muestras sucesivas en las que hay poca diferencia entre sus amplitudes. En este caso se necesita transmitir varios códigos PCM idénticos, lo cual es redundante. La modulación por código de impulsos diferencial (DPCM, de *differential pulse code modulation*) se diseñó en forma específica para aprovechar las redundancias entre muestra y muestra en las formas de onda características de la voz. En la DPCM, la diferencia en amplitudes de dos muestras sucesivas es la que se transmite, y no la muestra real. Como el intervalo de diferencias entre muestras suele ser menor que el de las muestras individuales, se requieren menos bits para la DPCM que en la PCM convencional.

La fig. 15-25 muestra un diagrama simplificado de bloques de un transmisor DPCM. La señal analógica de entrada se limita en su banda a la mitad de la frecuencia de muestreo, después se compara con la señal acumulada anterior, en el diferenciador. La salida de éste es la diferencia entre las dos señales. Esta diferencia se codifica por PCM y se transmite. El convertidor A/D funciona igual que en un sistema PCM convencional, pero suele usar menos bits por muestra.

La fig. 15-26 representa un diagrama simplificado de bloques de un receptor DPCM. Cada muestra recibida se reconvierte a forma analógica, se guarda y a continuación se suma a la siguiente muestra recibida. En el receptor de esta figura, la integración se hace sobre las señales analógicas, aunque también se podría efectuar en forma digital.

TRANSMISIÓN DE PULSOS

Todos los sistemas de portadora digital implican la transmisión de pulsos a través de un medio, con ancho de banda finita. Un sistema muy selectivo necesitaría una gran cantidad de secciones filtrantes, lo cual es impráctico. En consecuencia, en los sistemas digitales prácticos se usan en general filtros con anchos de banda aproximado de 30% o más, en exceso del ancho de banda de Nyquist. La fig. 15-27a muestra la forma de onda de salida característica de un canal de comunicaciones de *banda limitada*, cuando se aplica un pulso angosto a su entrada. Se ve que al limitar la banda de un pulso, la energía de éste se reparte en un tiempo bastante mayor, en forma de *lóbulos* secundarios. Estos lóbulos secundarios se llaman *oscilación en los extremos*. El

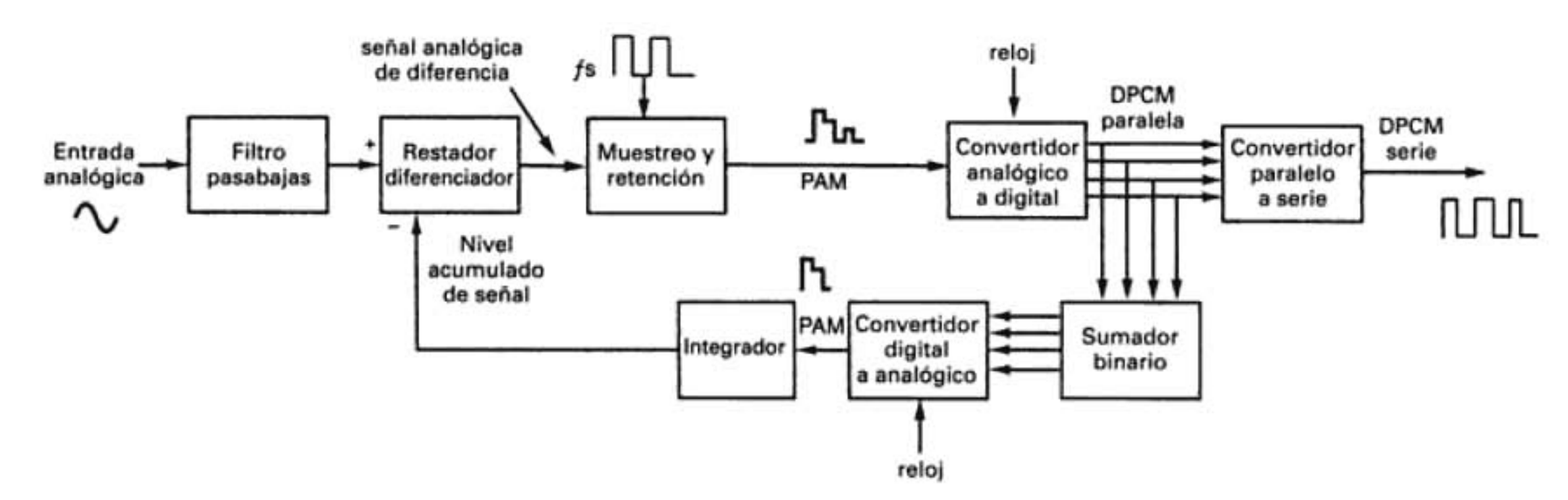


FIGURA 15-25 Transmisor DPCM

espectro de frecuencias de salida que corresponde a un pulso rectangular se representa con una respuesta (sen x)/x, y se calcula como sigue

$$f(\omega) = (T)\frac{\text{sen}(\omega T/2)}{\omega T/2} \tag{15-12}$$

donde $\omega = 2\pi f$ (radianes)
T = ancho del pulso (segundos)

La fig. 15-27b muestra la distribución del espectro total de potencias. En forma aproximada, el 90% de la potencia de la señal está contenida dentro el primer *cero espectral*, es decir, $f = 1/T$. Por consiguiente, se puede confinar la señal a un ancho de banda $B = 1/T$, y seguir pasando la mayor parte de la energía de la forma de onda original. En teoría, sólo se necesita conservar la amplitud en la mitad de cada intervalo de pulso. Por lo tanto, si el ancho de banda se confina en $B = 1/2T$, la frecuencia máxima de señalización alcanzable en un filtro de paso bajo, con un ancho especificado de banda, sin causar demasiada distorsión, se determina con la frecuencia de Nyquist, y es igual a dos veces el ancho de banda. La ecuación de la frecuencia de Nyquist es

$$R = 2B \tag{15-13}$$

siendo R = frecuencia de señalización = $1/T$
B = ancho de banda especificado

Interferencia entre símbolos

En la fig. 15-28 se ve la señal de entrada a un filtro pasabajas ideal, de ancho de banda mínimo. Esta señal es una secuencia aleatoria, binaria, sin regreso a cero (NRZ). La fig. 15-28b muestra la salida de un filtro pasabajas que no introduce distorsión alguna de fase o de amplitud. Nótese que la señal de salida llega a su valor total, para cada pulso transmitido, exactamente en el centro de cada intervalo de muestreo. Sin embargo, si es imperfecto el filtro de paso bajo (es el caso real), la respuesta de salida se asemejará más a la que se ve en la fig. 15-28c. En los momentos del muestreo, es decir, al centro de los pulsos, la señal no siempre llega a su valor máximo. Las oscilaciones en los extremos de varios pulsos se han *traslapado*, interfiriendo con el *lóbulo principal de pulso*. Suponiendo que no hay retardos en el sistema, aparece energía en forma de respuestas espurias del tercero y cuarto pulsos debidas a un pulso, durante el momento de muestreo ($T = 0$) de otro pulso. A esta interferencia se le suele llamar *interferencia entre símbolos* o ISI (de *intersymbol interference*). Es un aspecto importante en la transmisión de pulsos por circuitos con ancho limitado de banda y respuesta de fase no lineal. Dicho en forma simple, los pulsos rectangulares no permanecerán rectangulares cuando el ancho de banda sea menor que infinito. Mientras menor sea el ancho de banda, los pulsos serán más redondeados. Si es excesiva la distorsión de fase, el pulso se *inclinará* y, en consecuencia, afectará al siguiente pulso. Cuando se multiplexan pulsos de más de una fuente, las amplitudes, frecuencias y fases de las respuestas se vuelven todavía más críticas. La ISI causa *diafonía* entre canales que ocupan muescas de tiempo adyacentes en un sistema de portadora multiplexado por división de tiempo. Unos filtros especializados, llamados *igualadores*, se intercalan en la ruta de transmisión para "igualar" la distorsión para todas las frecuencias, creando un medio uniforme de transmisión y reduciendo los deterioros de ésta. Las cuatro causas principales de la interferencia entre símbolos son:

1. *Imprecisión en sintonía.* En los sistemas de transmisión digital, las inexactitudes de sincronización en el transmisor causan interferencia entre símbolos, si la frecuencia de la transmisión no se apega a la *frecuencia de llamada* o frecuencia de señalización diseñada para el canal de comunicaciones. En general, las inexactitudes de sincronización de este tipo son insignificantes. Como la información del reloj del receptor se toma de las señales recibidas, que están contaminadas con ruido, es más probable que en los receptores haya sincronización inexacta en el muestreo, que en los transmisores.

2. *Ancho de banda insuficiente.* Los errores de sincronización se presentan con menos probabilidades cuando la velocidad de transmisión es bastante menor que el ancho de banda del canal (es decir, el ancho de banda de Nyquist es bastante menor que el ancho de banda del

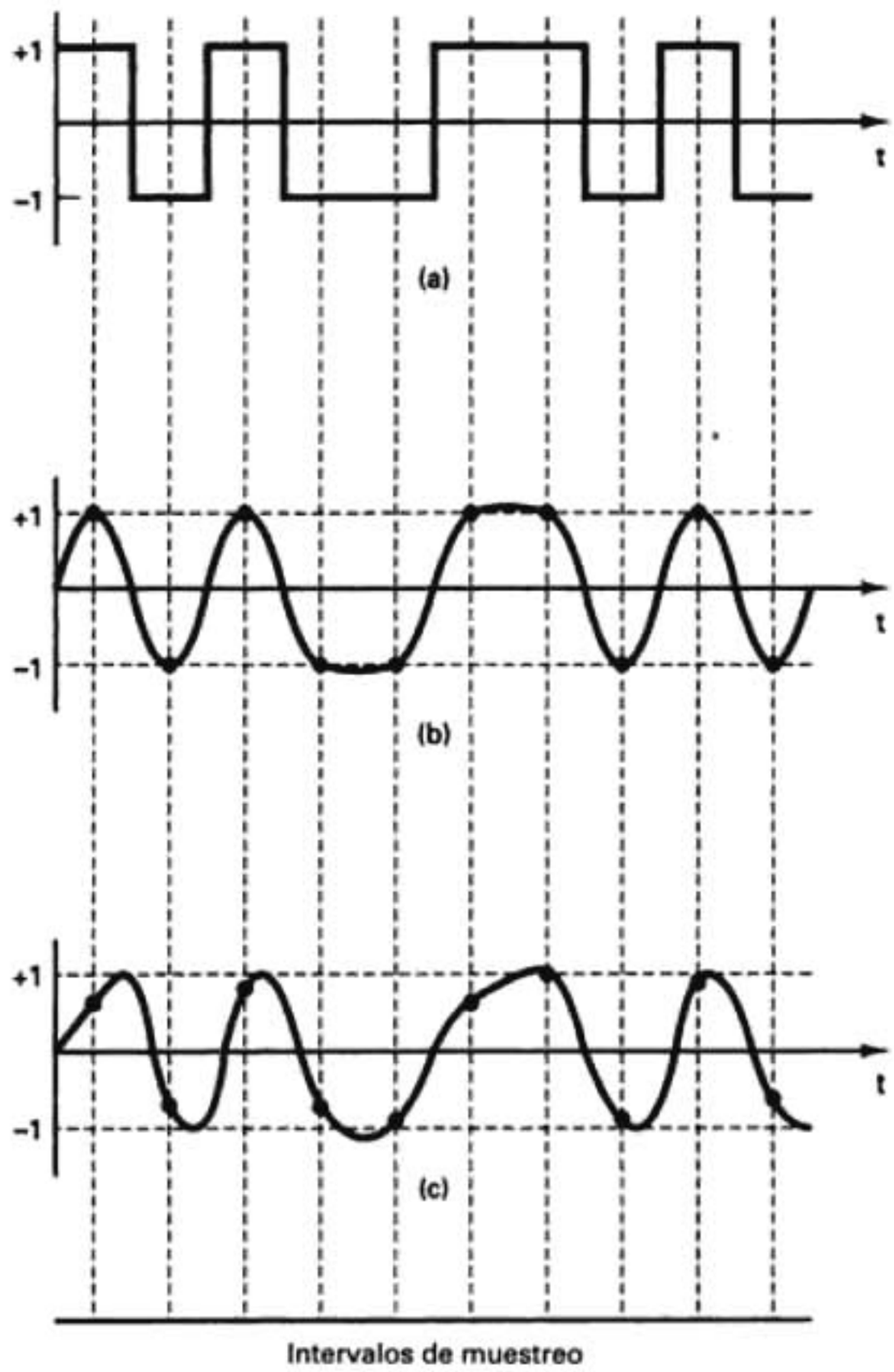


FIGURA 15-28 Respuestas al impulso: (a) señal de entrada NRZ; (b) salida de un filtro perfecto; (c) salida de un filtro imperfecto

canal). Cuando se reduce el ancho de banda de un canal de comunicaciones, la frecuencia de llamada se reduce y es más probable que haya interferencia entre símbolos.

3. *Distorsión de amplitud.* Los filtros se instalan en un canal de comunicaciones para limitar la banda de las señales y reducir o eliminar el ruido y la interferencia esperados. También se usan los filtros para producir un pulso específico de respuesta. Sin embargo, la respuesta de frecuencia de un canal no siempre se puede pronosticar en forma absoluta. Cuando las características de frecuencia de un canal de comunicaciones se alejan de los valores normales, o esperados, se produce la *distorsión de pulso*. Esta distorsión se presenta cuando se reducen los máximos de los pulsos, causando frecuencias incorrectas de llamada en el dominio del tiempo. A la compensación de esos deterioros se le llama igualación de amplitud.

4. *Distorsión de fase.* Un pulso es una superposición de una serie de ondas senoidales relacionadas armónicamente, con relaciones específicas de fase y amplitud. Por consiguiente, si se alteran las relaciones de fase entre las ondas senoidales individuales, se causa la distorsión de fase. Esta distorsión se presenta cuando los componentes de frecuencia sufren distintas cantidades de retardo al propagarse por el medio de transmisión. En la trayectoria de transmisión se instalan igualadores especiales de demora, para compensar los retardos variables y reducir la distorsión de fase. Los igualadores de fase se pueden ajustar en forma manual, o se diseñan para ajustarse a sí mismos, en forma automática, a diversas características de la transmisión.

Patrones de ojo

La eficiencia de un sistema de transmisión digital depende, en parte, de la capacidad que tenga una repetidora para regenerar los pulsos originales. De igual modo, la calidad del proceso de regeneración depende del circuito de decisiones dentro de la repetidora, y de la calidad de la señal que entra a ese circuito. Por lo anterior, se puede medir la eficiencia de un sistema de transmisión digital mostrando la señal recibida en un osciloscopio, y ajustando la base de tiempo a la velocidad de los datos. Así, todas las combinaciones de formas de onda se sobreponen en intervalos de señalización adyacentes. A esta presentación se le llama *patrón de ojo*, *figura de ojo* o *diagrama de ojo*. Un patrón de ojo es una técnica cómoda para determinar los efectos de las degradaciones introducidas en los pulsos, cuando viajan al regenerador. El aparato de prueba para mostrar un patrón de ojo se ve en la fig. 15-29. La corriente de impulsos recibida se alimenta a la entrada vertical del osciloscopio, y el reloj de símbolos se conecta con la entrada de disparo externo, mientras que la frecuencia de barrido se ajusta, aproximadamente, a la frecuencia de símbolos.

La fig. 15-30 muestra un patrón de ojo generado por una forma de onda simétrica, para señales *ternarias*, en las que los pulsos individuales a la entrada del regenerador tienen una forma de coseno cuadrado. En un sistema de m niveles habrá $m - 1$ ojos separados. Las líneas horizontales, identificadas con $+1$, 0 y -1, corresponden a las amplitudes recibidas ideales. Las líneas verticales, separadas por T, el intervalo de señalización, corresponden a los *tiempos de*

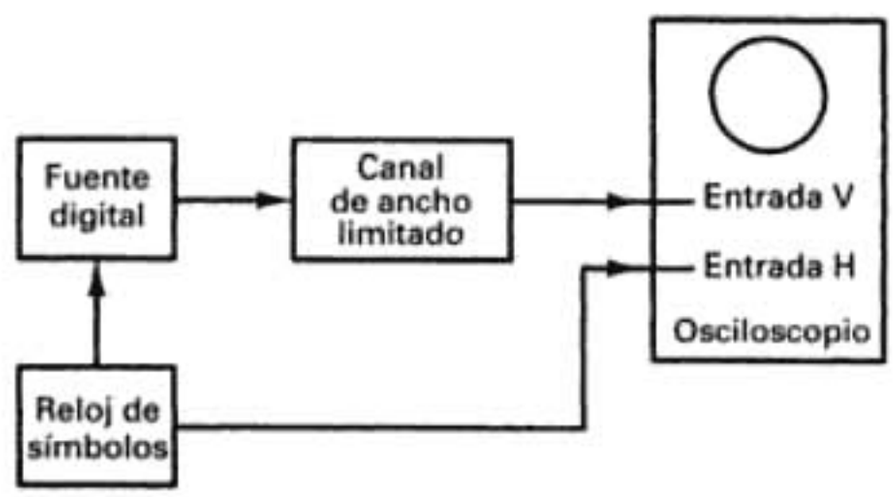


FIGURA 15-29 Conjunto de medición para diagrama de ojo

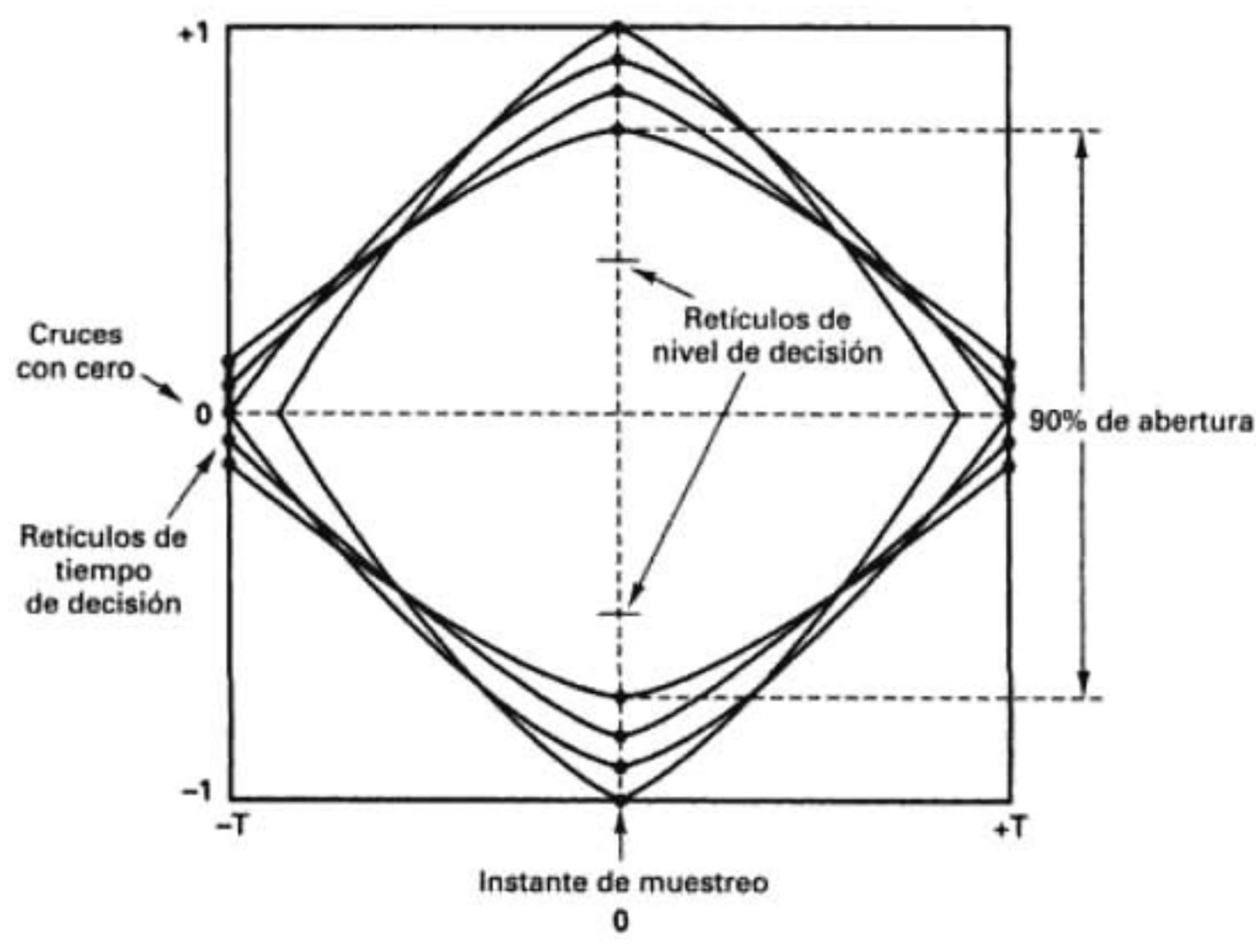


FIGURA 15-30 Diagrama de ojo

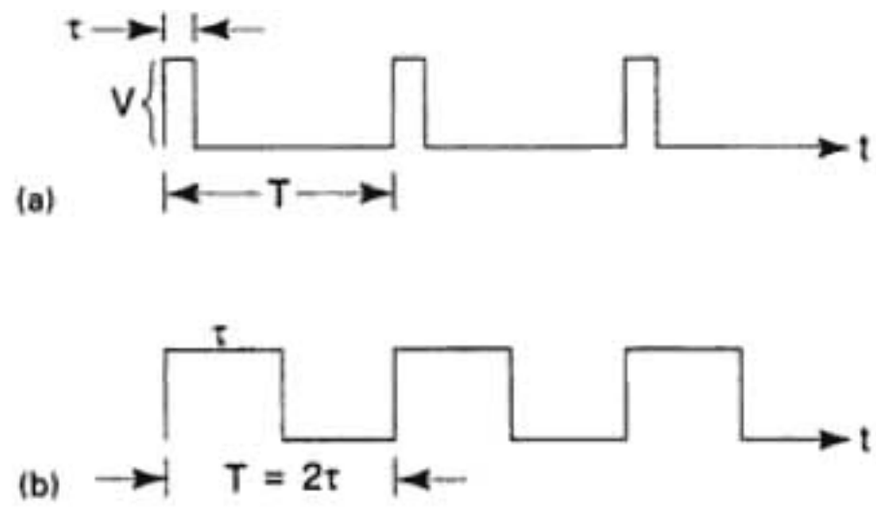


FIGURA 15-31 Señales digitales binarias: (a) $\tau/T < 0.5$; (b) $\tau/T = 0.5$

Si los pulsos rectangulares de amplitud V y relación τ/T comienzan en $t = 0$, entonces

$$v(t) = \begin{cases} V & 0 \le t \le \tau \\ 0 & \tau < t \le T \end{cases} \tag{15-17}$$

Así, de acuerdo con la ecuación 15-16,

$$\overline{P} = \frac{1}{T_0}\int_0^T (V)^2 dt = \frac{1}{T_0} V^2 t|_0^\tau \tag{15-18}$$

$$= \frac{\tau}{T_0} V^2$$

y

$$\overline{P} = \left(\frac{\tau}{T}\right)\frac{V^2}{R}$$

Ya que el valor rms o efectivo de una onda periódica se calcula con $P = (V_{rms})^2/R$, el voltaje rms de un pulso rectangular es

$$V_{rms} = \sqrt{\frac{\tau}{T}} V \tag{15-19}$$

Como $\overline{P} = (V_{rms})^2/R$, entonces $\overline{P} = (\sqrt{\tau/T}\, V)^2/R = (\tau V^2)/(TR)$.

Para la onda cuadrada de la fig. 15-31, $\tau/T = 0.5$ y, por consiguiente, $\overline{P} = V^2/2R$. Así, el voltaje rms para la onda cuadrada es igual que para las ondas senoidales, $V_{rms} = V/\sqrt{2}$.

PREGUNTAS

15-1. Describa las ventajas y desventajas de la transmisión digital.

15-2. ¿Cuáles son los cuatro métodos más comunes de modulación de pulsos?

15-3. ¿Cuál método de la pregunta anterior es la única forma de modulación de pulso que se usa en un sistema de transmisión digital? Explique por qué.

15-4. ¿Cuál es el objeto del circuito de muestreo y retención?

15-5. Defina *abertura* y *tiempo de adquisición*.

15-6. ¿Cuál es la diferencia entre muestreo natural y de parte plana?

15-7. Defina qué es *pérdida*. ¿Qué la provoca?

15-8. ¿Qué es la frecuencia de muestreo de Nyquist?

15-9. Defina y describa las causas de la distorsión por alias.

15-10. Explique la diferencia entre un código sólo de magnitud y uno de signo y magnitud.

15-11. Explique qué es distorsión por sobrecarga.

15-12. Explique qué es el cuantizado o la cuantización.

15-13. ¿Qué es intervalo de cuantización? ¿Qué es error por cuantización?

15-14. Defina el *intervalo dinámico*.

15-15. Explique la relación entre el intervalo dinámico, la resolución y la cantidad de bits en un código PCM.

15-16. Explique qué es eficiencia de codificación.

15-17. ¿Qué es SQR? ¿Cuál es la relación entre SQR, resolución, intervalo dinámico y cantidad de bits en un código PCM?

15-18. Describa la diferencia entre códigos PCM lineales y no lineales.

15-19. Explique el ruido de canal inactivo.

15-20. Describa la diferencia entre cuantización con paso inicial mayor y cuantización con paso homogéneo.

15-21. Defina qué es *compresión-expansión*.

15-22. ¿Qué determina el parámetro μ?

15-23. Explique en forma breve el proceso de compresión y expansión digital.

15-24. ¿Cuál es el efecto de la compresión digital sobre la SQR, la resolución, el intervalo de cuantización y el ruido de cuantización?

15-25. Describa la diferencia entre PCM de modulación delta y PCM normal.

15-26. Defina qué es *sobrecarga de pendiente* y *ruido granular*.

15-27. ¿Cuál es la diferencia entre la modulación delta adaptativa y la modulación delta convencional?

15-28. Describa la diferencia entre PCM diferencial y PCM convencional.

PROBLEMAS

15-1. Determine la frecuencia de muestreo de Nyquist para una frecuencia de entrada analógica máxima de:

(a) 4 kHz.
(b) 10 kHz.

15-2. Para el circuito de muestreo y retención de la fig. 15-4a, determine el capacitor de valor máximo que se puede usar. Use los siguientes parámetros: impedancia de salida de $Z_1 = 20\ \Omega$, y resistencia de Q_1 encendido de 20 Ω, tiempo de adquisición 10 μs, corriente máxima de salida de $Z_1 = 20$ mA, y exactitud de 1%.

15-3. Calcule la frecuencia máxima de entrada analógica para una frecuencia de muestreo de 20 kHz.

15-4. Calcule la frecuencia alias para una frecuencia de muestreo de 14 kHz y una frecuencia de entrada analógica de 8 kHz.

15-5. Calcule el intervalo dinámico para un código PCM de signo y magnitud, de 10 bits.

15-6. Calcule la cantidad mínima de bits necesarios en un código PCM, para un intervalo dinámico de 80 dB. ¿Cuál es la eficiencia de codificación?

15-7. Para tener una resolución de 0.04 V, calcule los voltajes para los siguientes códigos PCM lineales, de signo y magnitud, de 7 bits:

(a) 0 1 1 0 1 0 1
(b) 0 0 0 0 0 1 1
(c) 1 0 0 0 0 0 1
(d) 0 1 1 1 1 1 1
(e) 1 0 0 0 0 0 0

15-8. Determine la SQR para una señal de $2v_{rms}$, y un intervalo de cuantización de 0.2 V.

15-9. Determine la resolución y el error de cuantización para un código PCM lineal, de signo y magnitud, de 8 bits, para un voltaje máximo decodificado de 1.27 V.

15-10. Se comprime digitalmente un código PCM lineal de 12 bits, a 8 bits. La resolución es 0.03 V. Calcule lo siguiente, para un voltaje analógico de entrada de 1.465 V:

a) Código PCM lineal de 12 bits.
b) Código comprimido de 8 bits.
c) Código decodificado de 12 bits.
d) Voltaje decodificado.
e) Error porcentual.

15-11. Para un código lineal PCM de 12 bits, con resolución de 0.02 V, determine el intervalo de voltajes que se podría convertir a los siguientes códigos PCM:

15-21. Calcule la SQR para las siguientes magnitudes de señal de entrada y ruido de cuantización:

V_s	V_n (V)
1 v_{rms}	0.01
2 v_{rms}	0.02
3 v_{rms}	0.01
4 v_{rms}	0.2

15-22. Calcule la resolución y el ruido de cuantización para un código PCM lineal, de signo y magnitud y de 8 bits, para los siguientes voltajes máximos decodificados: $V_{máx}$ = 3.06 V_p, 3.57 V_p, 4.08 V_p y 4.59 V_p.

15-23. Un código PCM lineal, de signo y magnitud, con 12 bits, se comprime digitalmente a 8 bits. Para tener una resolución de 0.016 bits, calcule las siguientes cantidades, con los voltajes de entrada indicados: código lineal PCM de 12 bits, código comprimido de 8 bits, código decodificado de 12 bits, voltaje decodificado y error porcentual. V_{ent} = −6.592 V, +12.992 V y −3.36 V.

15-24. Calcule el intervalo de voltajes a los que se convertirían los siguientes códigos PCM lineales de 12 bits:

Código lineal de 12 bits	Resolución (V)
100011110010	0.12
000001000000	0.10
000111111000	0.14
111111110000	0.12

15-25. Para los siguientes códigos PCM lineales de 12 bits, determine el código comprimido de 8 bits al que se convertirían:

Código lineal de 12 bits
100011110010
000001000000
000111111000
111111110010
000000100000

15-26. Determine el código expandido de 12 bits al que se convertirían los siguientes códigos comprimidos de 8 bits:

Código de 8 bits
11001010
00010010
10101010
01010101
11110000
11011011

CAPÍT

Imagen protegida por derechos de autor

Multiplexado

INTRODUCCIÓN

Multiplexado es la transmisión de información (en cualquier forma) de más de una fuente a más de un destino a través del mismo medio (instalación) de transmisión. Aunque las transmisiones sucedan en la misma instalación, no necesariamente suceden al mismo tiempo. El medio de transmisión puede ser un par de alambres metálicos, un cable coaxial, un teléfono móvil PCS, un sistema de microondas terrestres de radio, uno de microondas satelitales o un cable de fibra óptica.

Hay varios dominios en los que puede hacerse el multiplexado, incluyendo espacio, fase, tiempo, frecuencia y longitud de onda. El *multiplexado por división de espacio* (SDM, de *space-division multiplexing*) es una forma bastante sencilla de multiplexar, que sólo consiste en propagar señales de distintas fuentes por distintos cables, que están contenidos en la misma cepa. Se considera que la cepa es el medio de transmisión. La manipulación QPSK es una forma de *multiplexado por división de fase* (PDM, de *phase-division multiplexing*) en la que dos canales de datos (el I y el Q) modulan la misma frecuencia de portadora, que se ha desplazado 90° en fase. Así, los bits del canal I modulan una onda senoidal portadora, mientras que los bits del canal Q modulan una onda portadora cosenoidal. Después de haberse hecho la modulación, se combinan linealmente las portadoras I y Q y se propagan al mismo tiempo por el medio de transmisión, que puede ser un cable o el espacio libre.

Los tres métodos más predominantes para multiplexar señales son el de *multiplexado por división de tiempo* (TDM, de *time-division multiplexing*), por *división de frecuencia* (FDM, de *frequency-division multiplexing*) y el desarrollado en fecha más reciente, por *división de longitud de onda* (WDM, de *wavelength-division multiplexing*). El resto de este capítulo se dedicará al multiplexado por división de tiempo, frecuencia y longitud de onda.

MULTIPLEXADO POR DIVISIÓN DE TIEMPO

En el TDM, las transmisiones de varias fuentes se hacen por la misma instalación, pero no al mismo tiempo. Las transmisiones procedentes de distintas fuentes se *intercalan* en el dominio del tiempo. La clase más común de modulación que se usa en TDM es la PCM (modulación por código de pulso). En un sistema PCM-TDM, se muestrean dos o más canales de banda de voz, se convierten a códigos PCM y a continuación se multiplexan por división de tiempo a través de un solo par en cable metálico, o en un cable de fibra óptica.

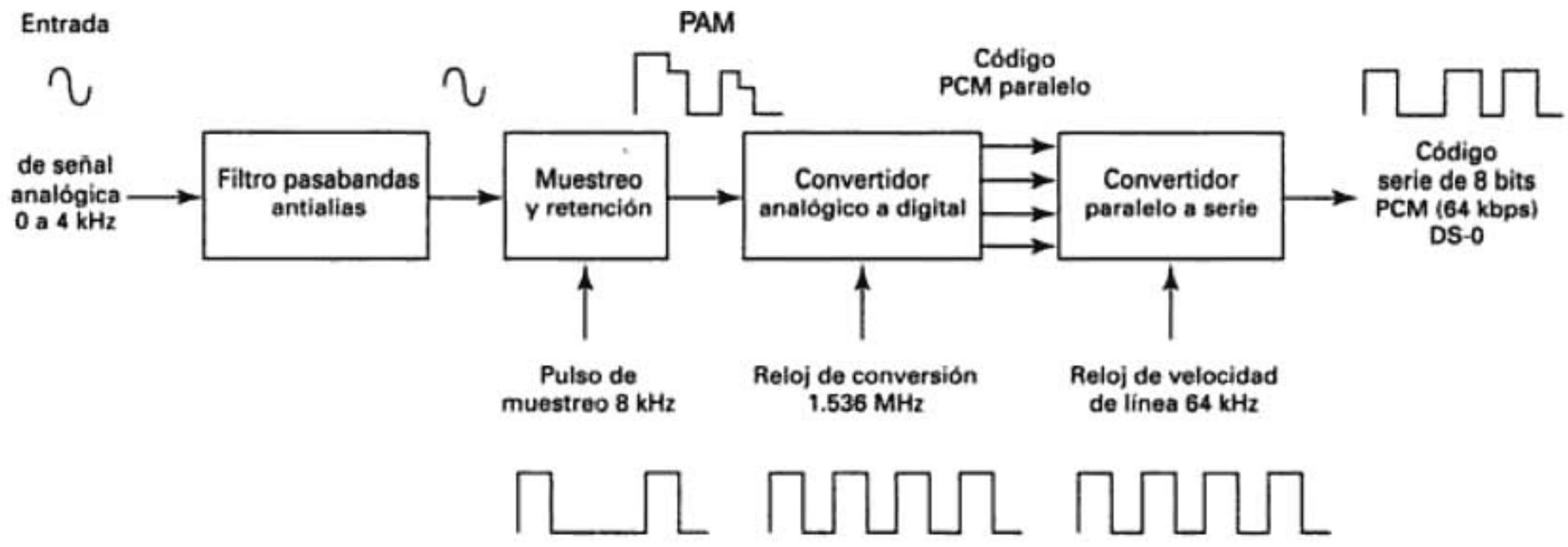


FIGURA 16-1 Sistema de transmisión PCM de un solo canal (DS nivel 0)

La piedra constructiva fundamental de cualquier sistema TDM comienza con un canal DS-0 (señal digital-nivel 0). La fig. 16-1 muestra el diagrama simplificado de bloques de un sistema PCM monocanal DS-0. Como allí se ve, los canales DS-0 usan una frecuencia de muestreo de 8 kHz y un código PCM de ocho bits, que produce una señal PCM de 64 kbps en su salida

$$\frac{8000 \text{ muestras}}{\text{segundo}} \times \frac{8 \text{ bits}}{\text{muestra}} = 64 \text{ kbps}$$

La fig. 16-2a muestra el diagrama simplificado de bloques para un sistema multiplexado de portadora PCM-TDM formado por dos canales DS-0. Se muestrea en forma alternada cada canal de entrada, con una frecuencia de 8 kHz, y se convierte a código PCM. Mientras se transmite el código PCM para el canal 1, se muestrea el canal 2 y se convierte a cógigo PCM. Mientras se transmite el código PCM del canal 2, se toma la siguiente muestra del canal 1 y se convierte a código PCM. Este proceso continúa y se toman muestras en forma alternada de cada canal, se convierten a código PCM y se transmiten. El multiplexor no es más que un interruptor digital controlado electrónicamente con dos entradas y una salida. Se seleccionan en forma alternada el canal 1 y el canal 2, y se conectan con la salida del multiplexor. El tiempo que tarda en transmitir una muestra de cada canal se llama *tiempo de trama*. Este tiempo es igual a la recíproca de la frecuencia de muestreo ($1/f_s$, o $1/8000 = 125\ \mu s$). La fig. 16-2b muestra la asignación de tramas TDM para un sistema PCM de dos canales, con una frecuencia de muestreo de 8 kHz.

El código PCM para cada canal ocupa una muesca fija de tiempo (época) dentro de la trama TDM total. Con un sistema de dos canales, se toma una muestra de cada canal durante cada trama, y el tiempo asignado para transmitir los bits PCM de cada canal es igual a la mitad del tiempo total de trama. Así, se deben transmitir ocho bits de cada canal durante cada trama (un total de 16 bits por trama). Por lo tanto, la velocidad de línea a la salida del multiplexor es

$$\frac{2 \text{ canales}}{\text{trama}} \times \frac{8000 \text{ tramas}}{\text{segundo}} \times \frac{8 \text{ bits}}{\text{canal}} = 128 \text{ kbps}$$

Aunque cada canal produce y transmite sólo 64 kbps, los bits deben salir sincronizados a la línea a una frecuencia de 128 kHz para permitir la transmisión de ocho bits por canal en cada muesca de tiempo de 125 μs.

SISTEMA DE PORTADORA DIGITAL T1

Un sistema de portadora digital es de comunicaciones, que usa pulsos digitales y no señales analógicas, para codificar la información. La fig. 16-3a muestra el diagrama de bloques del sistema

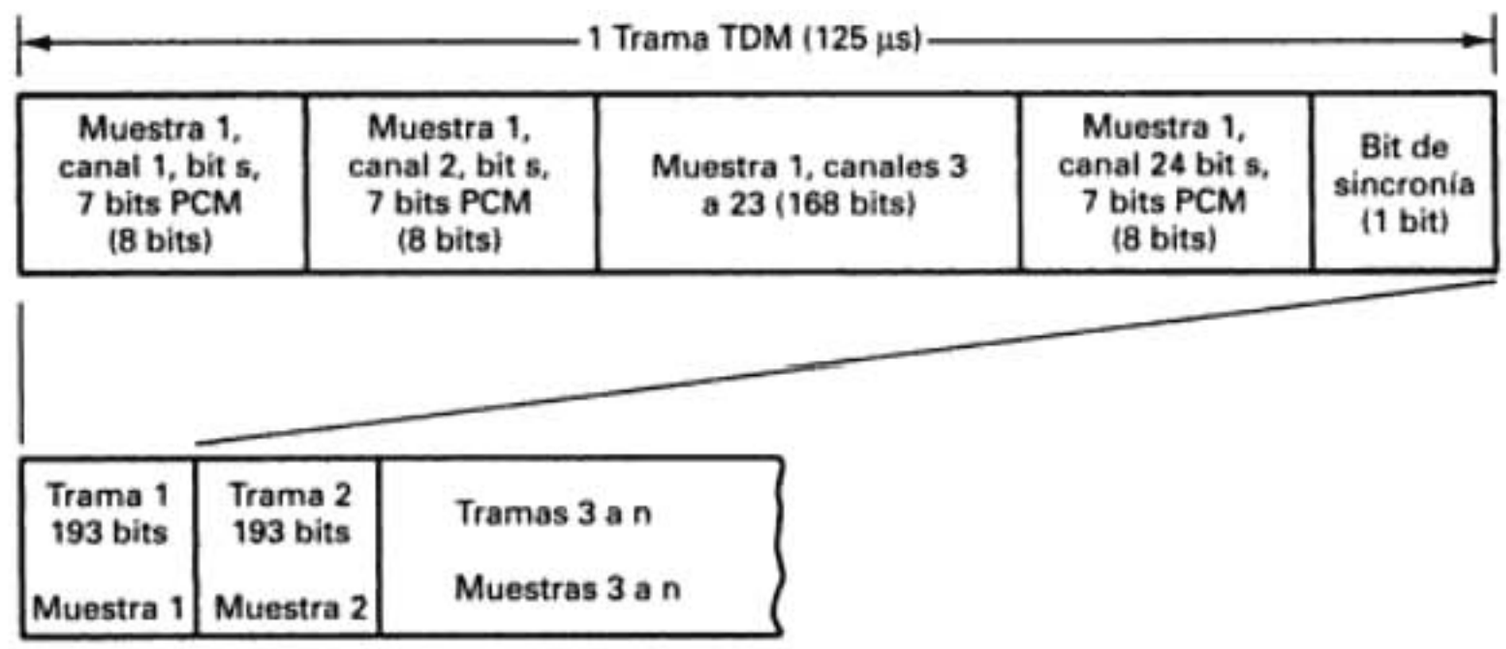


FIGURA 16-4 Alineación de trama y muestra en sistema de portadora T1, usando bancos de canales D1A

Después se agrega un bit adicional, el *bit de trama* a la señal transmitida de cada trama. El bit de trama se presenta una vez por trama (velocidad de 8000 bps) y se recupera en los circuitos del receptor, donde se usa para mantener la sincronización de trama y muestra entre el transmisor y el receptor TDM. En consecuencia, cada trama contiene 193 bits, y la velocidad de línea para un sistema de portadora digital T1 es

$$\frac{193 \text{ bits}}{\text{trama}} \times \frac{8000 \text{ tramas}}{\text{segundo}} = 1.544 \text{ Mbps}$$

Es importante notar que cada canal se muestrea con la misma frecuencia, pero no necesariamente al mismo tiempo. La fig. 16-3b muestra la secuencia de muestreo de canales para un sistema de portadora digital T1. Allí se ve que cada canal se muestrea una vez por cada trama, pero no al mismo tiempo. La muestra de cada canal está desplazada respecto a la anterior la 24ª parte del tiempo total de la trama.

Bancos de canales tipo D

Los primeros sistemas de portadora T1 tenían bancos de canales D1A, que usan un código PCM de 7 bits, sólo de magnitud, con compresión y expansión analógica y $\mu = 100$. Una versión posterior del banco de canal D1 (la D1D) usaba un código PCM de ocho bits, con signo y magnitud. En los bancos de canales D1D se agrega un octavo bit, el de signo (s) a cada palabra en código PCM para *señalización* (supervisión: colgado, descolgado, marcación, etcétera). En consecuencia, la frecuencia de señalización para los bancos de canales D1D es 8 kbps. También, en esos bancos, la secuencia de bit de trama simplemente es una serie de 1/0 alternados. La fig. 16-4 muestra el alineamiento de trama y muestra para el sistema de portadora T1 con bancos de canales D1D.

En forma genérica, el sistema de portadora T1 ha evolucionado pasando por los bancos de canales D2, D3, D4, D5 y D6. Los D4, D5 y D6 usan un código PCM comprimido de ocho bits, de signo y magnitud, con $\mu = 255$. En el banco de canales D1, las características de compresión y expansión se implementaban en circuitos, aparte del codificador y del decodificador. En los bancos D2, D3, D4 y D5 se incorporan las funciones de compresión y expansión en forma directa, en los codificadores y decodificadores. Aunque los bancos de canales D2 y D3 se parecen funcionalmente, los bancos D3 fueron los primeros en tener circuitos integrados LSI (en gran escala) codec (codificador-decodificador) para cada canal de banda de voz. En los bancos de canales D1, D2 y D3, el equipo común hace las funciones de codificación y decodificación. En consecuencia, un solo desperfecto del equipo provoca una falla total del sistema.

Los bancos de canales D1A usan un código sólo de magnitudes y, en consecuencia, un error en el bit más significativo (MSB) de una muestra de canal produce siempre un error decodificado igual a la mitad de todo el intervalo de cuantización, $V_{máx}$. Como los bancos de canal D1D, D2, D3, D4 y D5 usan código con signo y magnitud, un error en el MSB, que es el bit de signo, causa un error decodificado igual al doble de la magnitud de la muestra (de +V a −V o

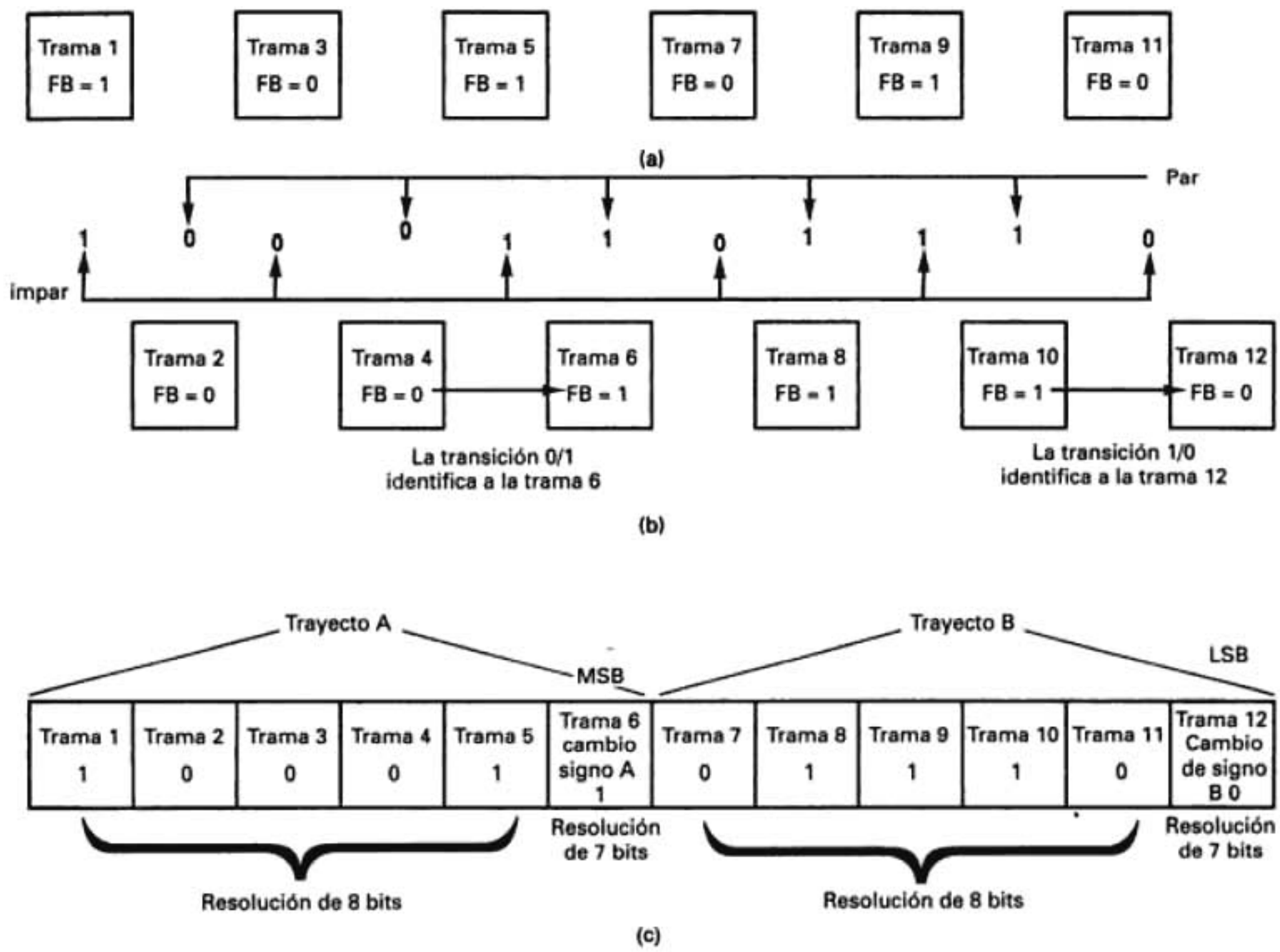


FIGURA 16-5 Secuencia de bits de entramado para el formato de supertrama T1, usando bancos de canales D2 o D3: (a) bits de sincronización de trama (tramas impares); (b) bits de señalización de alineación de trama (tramas pares); (c) alineación de tramas combinadas

viceversa). El error en el peor de los casos es igual a dos veces el intervalo total de cuantización. Sin embargo, las muestras con amplitud máxima se presentan rara vez, y la mayoría de los errores en la codificación D1D, D2, D3, D4 y D5 es menor que la mitad del intervalo de codificación. En promedio, la eficiencia contra errores en un código de signo y magnitud es mejor que en un código sólo de magnitud.

Formato de supertrama

La rapidez de señalización de 8 kbps que se usa en los bancos de canales D1 es demasiada para la transmisión de voz. Por consiguiente, en los bancos de canales D2 y D3 sólo se sustituye un bit de señalización en el bit menos significativo (LSB) de cada sexta trama. Así, cinco de cada seis tramas tienen resolución de 8 bits, mientras que una de cada seis tramas (la trama de señalización) sólo tiene 7 bits de resolución. En consecuencia, la rapidez de señalización en cada canal es 1.333 kbps (8000 bps/6), y la cantidad efectiva de bits por muestra es en realidad 7-5/6 bits, y no ocho.

Ya que sólo cada seis tramas incluyen un bit de señalización, es necesario numerar todas las tramas, para que el receptor sepa cuándo extraer la información de señalización. También, como la señalización se hace con una palabra binaria de dos bits, es necesario identificar el bit más significativo y el menos significativo (MSB y LSB) de la palabra de señalización. Por lo anterior, se inventó el formato de *supertrama*, que se ve en la fig. 16-5. En cada supertrama hay 12 tramas con número consecutivo, del 1 al 12. Los bits de señalización se sustituyen en las tramas 6 y 12, el MSB en la trama 6 y el LSB en la trama 12. Las tramas 1 a 5 se llaman trayecto A, y la trama 6 se designa trama de señalización de canal A. Las tramas 7 a 12 se llaman trayecto B, y la trama 12 se designa como la de señalización de canal B. Así, además de identificar a las tramas de señalización, se deben identificar en forma positiva las tramas sexta y décima segunda.

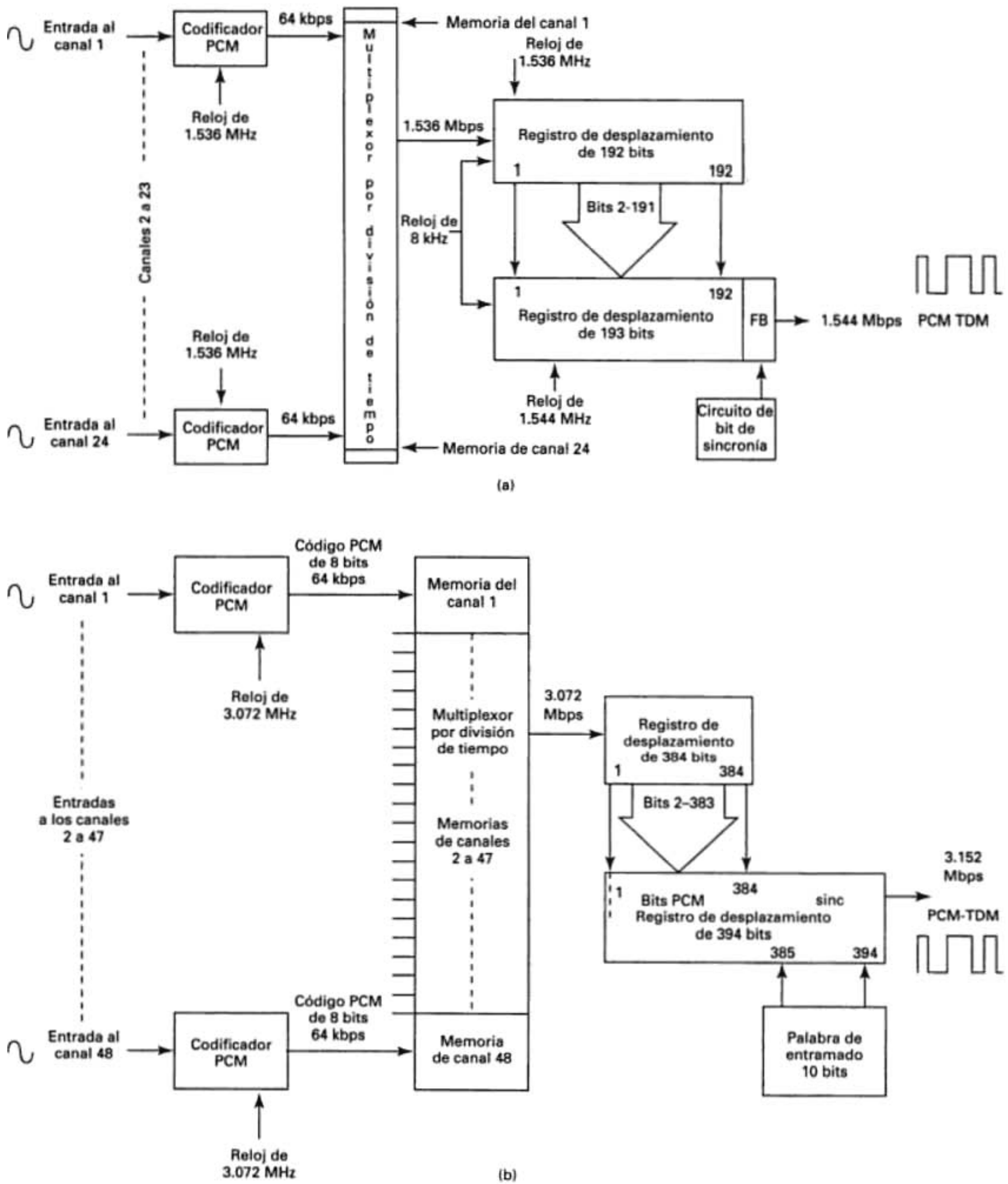


FIGURA 16-7 Circuitos de bit de sincronía en sistema de portadora T1: (a) DS-1; (b) DS-1C

Muesca de tiempo 0	Muesca de tiempo 1	Muesca de tiempo 2 a 16	Muesca de tiempo 17	Muesca de tiempo 18 a 30	Muesca de tiempo 31
Canal de entramado y alarma 8 bits	Canal de voz 1 8 bits	Canales de voz 2 a 15 112 bits	Canal común de señalización 8 bits	Canales de voz 16 a 29 112 bits	Canal de voz 30 8 bits

(a)

Muesca de tiempo 17

Trama	Bits 1234	Bits 5678
0	0000	xyxx
1	C 1	C 16
2	C 2	C 17
3	C 3	C 18
4	C 4	C 19
5	C 5	C 20
6	C 6	C 21
7	C 7	C 22
8	C 8	C 23
9	C 9	C 24
10	C 10	C 25
11	C 11	C 26
12	C 12	C 27
13	C 13	C 28
14	C 14	C 29
15	C 15	C 30

16 tramas equivalen a una multitrama; 500 multitramas se transmiten cada segundo

x = reserva

y = pérdida de alineación de multitrama si es 1

Se transmiten 4 bits por canal una vez cada 16 tramas, resultando en una rapidez de señalización de 500 palabras por segundo (2000 bps) para cada canal

(b)

FIGURA 16-8 Alineación de trama CCITT TDM y alineación de canal común de señalización: (a) trama CCITT TDM (125 μs, 256 bits, 2.048 Mbps); (b) canal común de señalización

porque combinan las funciones de codec y de filtro en el mismo paquete LSI. El filtro de entrada y salida hace las siguientes funciones: limitación de banda, rechazo de ruido, antialias y reconstrucción de las ondas analógicas de audio después de decodificar. El codec hace las siguientes funciones: muestreo analógico, codificación y decodificación (conversiones de analógico a digital y de digital a analógico) y compresión y expansión digital.

CHIPS COMBINADOS

Un chip combinado puede hacer las conversiones analógico a digital y digital a analógico, y el filtrado de transmisión y recepción necesario para interconectar un circuito telefónico de voz dúplex (de 4 alambres) al trayecto PCM de un sistema de portadora TDM. En esencia, un chip combinado reemplaza a la anterior combinación de chips codec y de filtro.

La tabla 16-2 muestra algunos de los chips combinados disponibles, junto con sus funciones principales.

Funcionamiento general

Un chip combinado hace las siguientes funciones principales:

1. Filtrado pasabanda de las señales analógicas antes de codificar y después de decodificar.
2. Codificación y decodificación de señales de voz y progreso de llamada.
3. Codificación y decodificación de información de señalización y supervisión.
4. Compresión y expansión digitales.

TABLA 16-2 Propiedades de varios chips combinados Codec/Filtro

2916 (16 terminales)	2917 (16 terminales)	2913 (20 terminales)	2914 (24 terminales)
Sólo compresión y expansión con ley μ	Sólo compresión y expansión con ley A	Compresión y expansión con leyes μ y A	Compresión y expansión con leyes μ y A
Reloj maestro sólo de 2.048 MHz	Reloj maestro sólo de 2.048 MHz	Reloj maestro de 1.536, 1.544 o 2.048 MHz	Reloj maestro de 1.536, 1.544 o 2.048 MHz
Velocidad fija de datos	Velocidad fija de datos	Velocidad fija de datos	Velocidad fija de datos
Velocidad variable de datos, 64 kbps a 2.048 Mbps	Velocidad variable de datos 64 kbps a 4.096 Mbps	Velocidad variable de datos 64kbps a 4.096 Mbps	Velocidad variable de datos, 64 kbps a 4.096 Mbps
Intervalo dinámico de 78 dB	Intervalo dinámico de 78 dB	Intervalo dinámico de 78 dB	Intervalo dinámico de 78 dB
Compatible con ATT D3/4	Compatible con ATT D3/4	Compatible con ATT D3/4	Compatible con ATT D3/4
Entrada unilateral	Entrada unilateral	Entrada diferencial	Entrada diferencial
Salida unilateral	Salida unilateral	Salida diferencial	Salida diferencial
Sólo ajuste de ganancia de transmisión	Sólo ajuste de ganancia de transmisión	Ajuste de ganancia de transmisión y recepción	Ajuste de ganancia de transmisión y recepción
Relojes síncronos	Relojes síncronos	Relojes síncronos	Relojes síncronos
			Relojes asíncronos
			Regreso analógico
			Señalización

La fig. 16-9 muestra el diagrama de bloques de un chip combinado típico. La fig. 16-9b muestra la curva de respuesta en frecuencia del filtro pasabanda de transmisión, y en la fig. 16-9c se ve la respuesta en frecuencia del filtro pasabajas de recepción.

Modo de velocidad fija de datos

En el modo de *velocidad fija de datos*, o *frecuencia fija de datos*, los *relojes maestros de transmisión* y *recepción* en un chip combinado (CLKX y CLKR) hacen las siguientes funciones:

1. Proporcionar el reloj maestro para el filtro incorporado (*de a bordo*) de capacitor conmutado.
2. Proporcionar el reloj para los convertidores analógico a digital y digital a analógico.
3. Determinar las frecuencias de datos de entrada y salida entre el codec y el trayecto PCM.

Por lo anterior, en el modo de velocidad fija de datos, las frecuencias de datos en transmisión y recepción deben ser 1.536, 1.54 o 2.048 Mbps, igual que la frecuencia del reloj maestro.

Los pulsos de sincronización de trama en transmisión y recepción (FSX y FSR) son entradas de 8 kHz, que establecen las frecuencias de muestreo de transmisión y recepción, y diferencian las tramas de *señalización* y *no de señalización*. La salida $\overline{\text{TSX}}$ es de *activar búfer estroboscópico de ranura de tiempo* que se usa para conducir la palabra PCM al trayecto PCM, cuando se usa el búfer externo para activar la línea. También, $\overline{\text{TSX}}$ se usa como pulso externo de control para un multiplexor por división de tiempo (véase la fig. 16-10).

Los datos se transmiten al trayecto PCM desde DX en las ocho primeras transiciones positivas de CLKX después del lado creciente de FSX. En el canal de recepción se reciben los datos del trayecto PCM desde DR en los primeros ocho lados decrecientes de CLKR, después de la ocurrencia de FSR. Por consiguiente, se deben sincronizar las ocurrencias de FSX y FSR entre los codecs, en un sistema de canal múltiple, para asegurar que sólo transmita o reciba un codec a o del trayecto PCM en determinado momento.

La fig. 16-10 muestra el diagrama de bloques y la secuencia de sincronización para un sistema PCM de un solo canal, que usa un chip combinado en el modo de velocidad fija de datos, y que funciona con una frecuencia de reloj maestro de 1.536 MHz. En el modo de velocidad

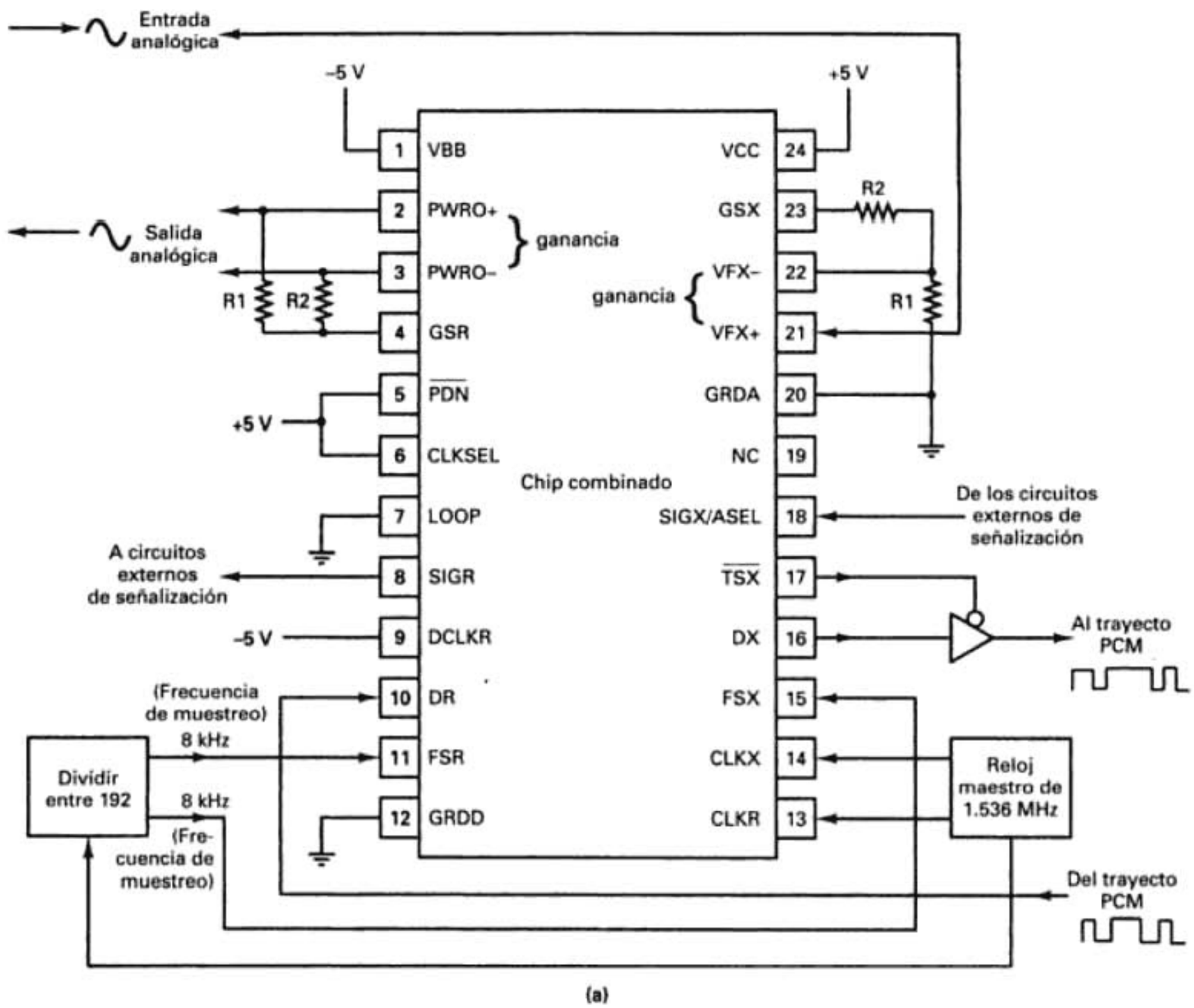


FIGURA 16-10 Sistema PCM de un solo canal con un chip combinado en el modo de velocidad fija de datos: (a) diagrama de bloques (*continúa*)

fija de datos, los datos entran y salen en un mismo canal en ráfagas cortas. A este modo de operación se le llama a veces *modo de ráfaga*. Cuando sólo hay un canal, el trayecto PCM sólo está activo 1/24 parte del tiempo total de trama. Se pueden agregar más canales al sistema, siempre y cuando se sincronicen sus transmisiones de tal modo que al mismo tiempo no haya transmisiones de ningún otro canal.

En la fig. 16-10 se puede observar lo siguiente:

1. Las frecuencias de bits de entrada y salida del codec son igual a la frecuencia del reloj maestro, 1.536 Mbps.
2. Las entradas y salidas del codec son de 64,000 bits PCM por segundo.
3. La salida de datos (DX) y la entrada de datos (DR) sólo están activadas la 24ª parte del tiempo total de trama, 125 μs.

Para agregar canales al sistema de la Fig. 16-10, se debe sincronizar la ocurrencia de las señales FSX, FSR y $\overline{\text{TSX}}$ para cada canal adicional, para que sigan una secuencia exacta, y no permitan que transmita o reciba más de un codec al mismo tiempo. La fig. 6-11 muestra el diagrama de bloques y la secuencia de sincronización para un sistema PCM-TDM de 24 canales, funcionando con una frecuencia de reloj maestro de 1.536 MHz.

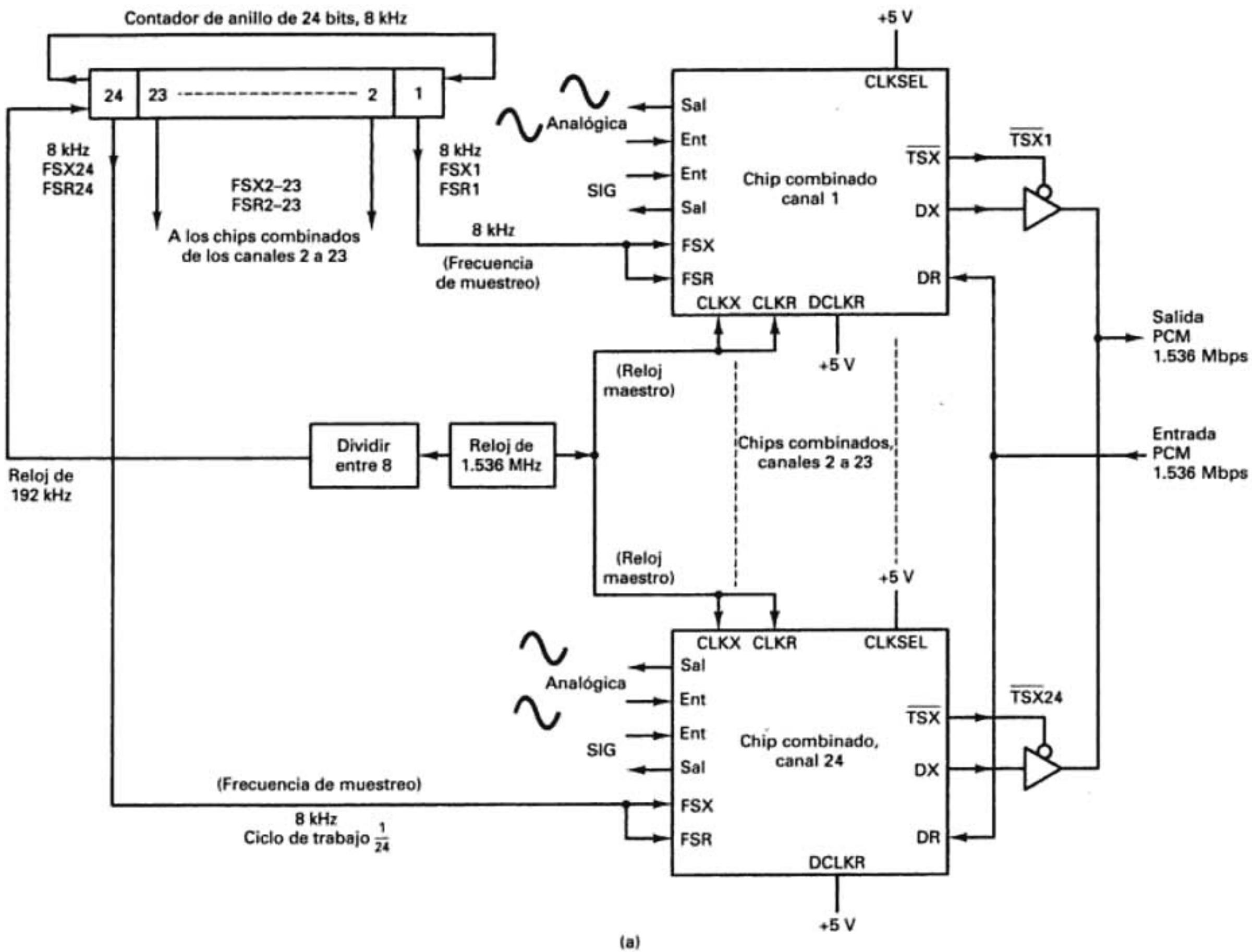


FIGURA 16-11 Sistema PCM-TEM de 24 canales con un chip combo en el modo de velocidad fija de datos, trabajando con una frecuencia de reloj maestro igual a 1.536 MHz: (a) diagrama de bloques (*continúa*)

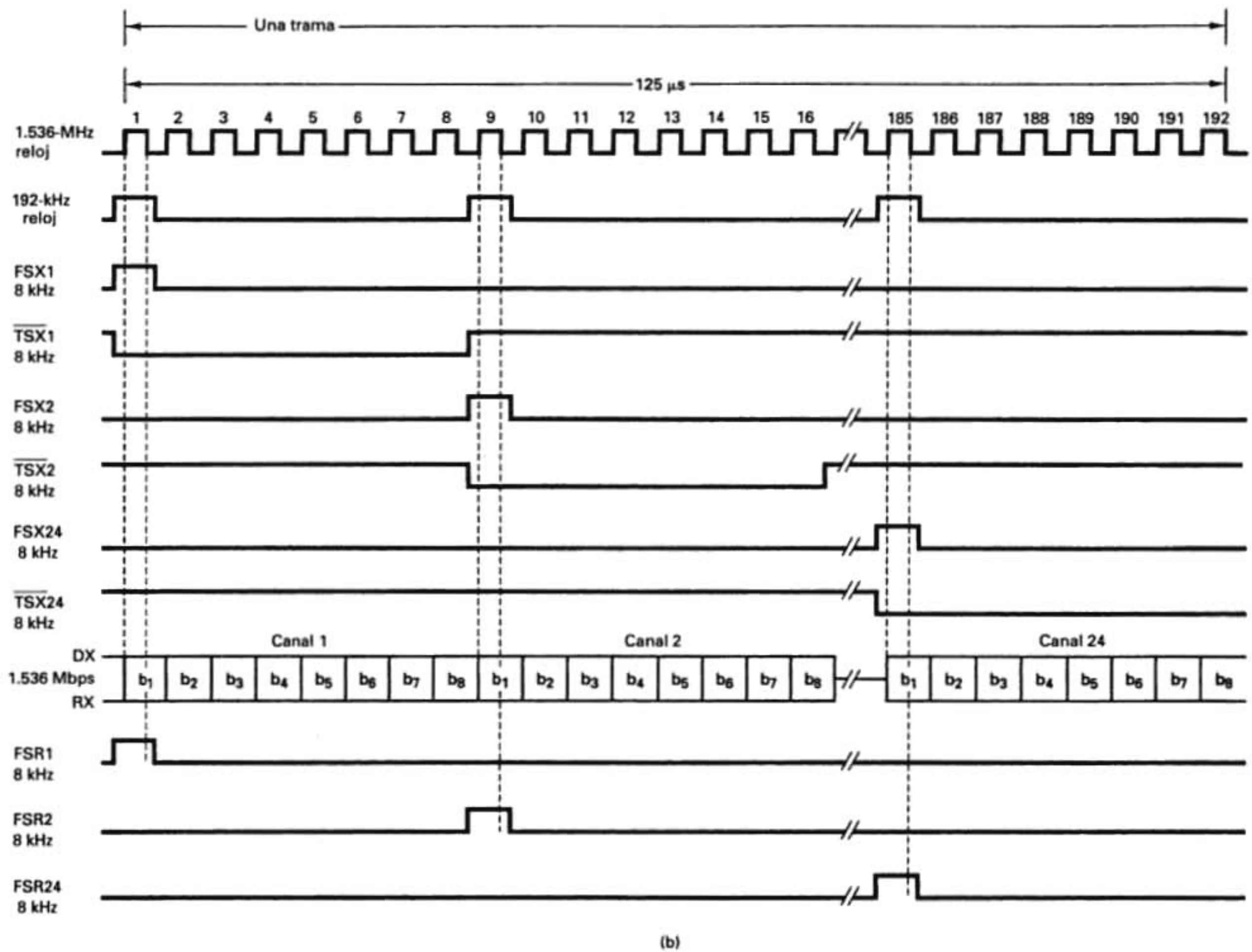


FIGURA 16-11 (*Continuación*) (b) diagrama de sincronización

Cuando la frecuencia de muestreo es de 8 kHz, el tiempo de trama es 125 μs. Por consiguiente, una palabra PCM de 8 bits de cada canal se transmite y/o recibe durante cada trama de 125 μs. Para que haya 16 bits en 125 μs, se requiere un reloj de transmisión y recepción de datos de 128 kHz.

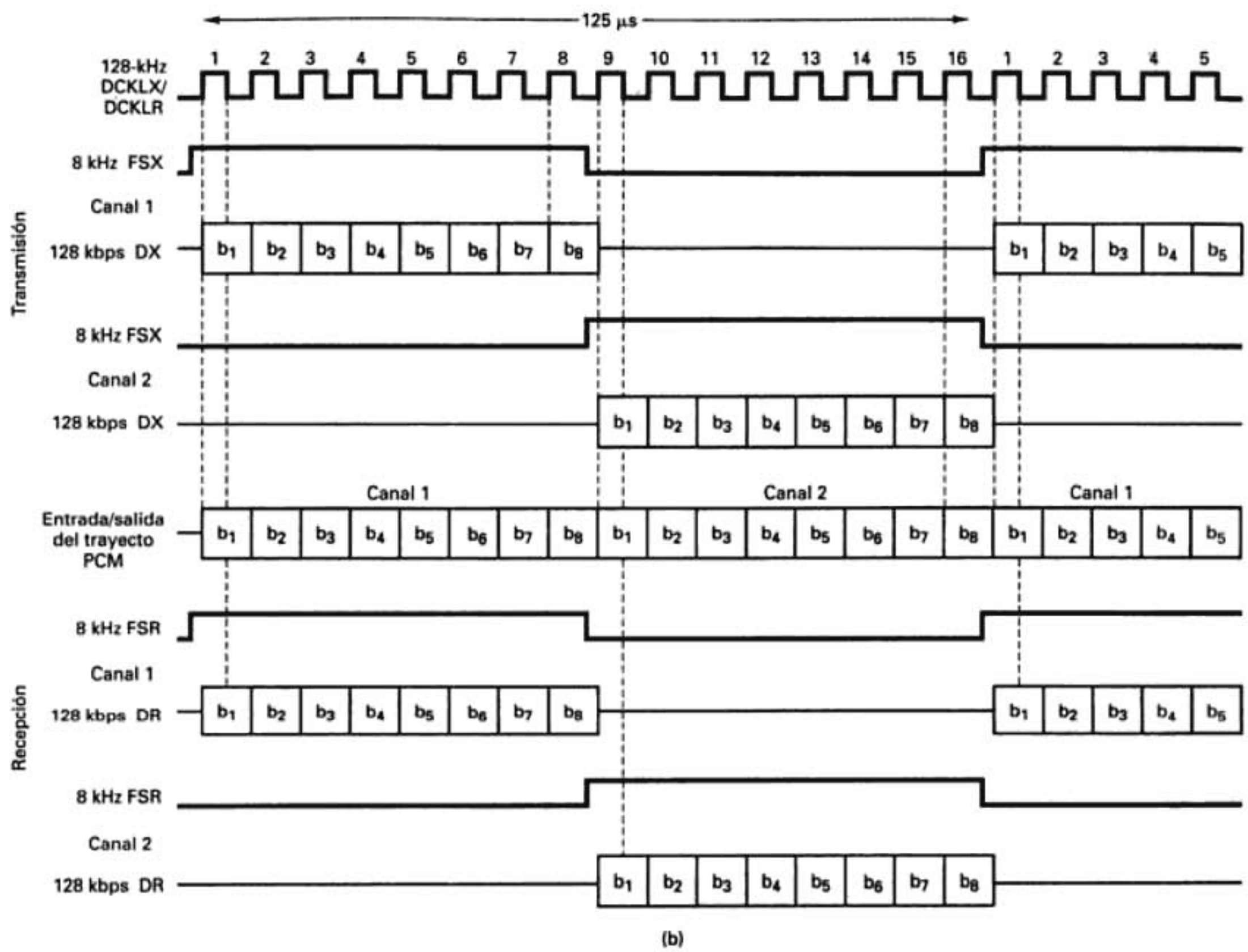
$$t_b = \frac{1 \text{ canal}}{8 \text{ bits}} \times \frac{1 \text{ trama}}{2 \text{ canales}} \times \frac{125\mu\text{s}}{\text{trama}} = \frac{125\ \mu\text{s}}{16 \text{ bits}} = \frac{7.8125\ \mu\text{s}}{\text{bit}}$$

$$\text{frecuencia de bits} = \frac{1}{t_b} = \frac{1}{7.8125\ \mu\text{s}} = 128 \text{ kbps}$$

o sea

$$\frac{8 \text{ bits}}{\text{canal}} \times \frac{2 \text{ canales}}{\text{trama}} \times \frac{8000 \text{ tramas}}{\text{segundo}} = 128 \text{ kbps}$$

Las señales de activar transmisión y recepción, FSX y FSR respectivamente, para cada codec están activas durante la mitad del tiempo total de trama. En consecuencia, a un codec se alimentan señales de activar transmisión y recepción de datos (FSX y FXR) de 8 kHz, con ciclo de trabajo de 50%, y se alimentan al otro codec desfasadas 180° (invertidas), activando así sólo a un codec al mismo tiempo.



(b)

FIGURA 16-12 (*Continuación*) (b) diagrama de sincronización

TABLA 16-3 Resumen de la jerarquía digital norteamericana

Tipo de línea	Señal digital	Frecuencia de bits (Mbps)	Capacidades de canal	Servicios ofrecidos
T1	DS-1	1.544	24	Teléfono con banda de voz
T1C	DS-1C	3.152	48	Teléfono con banda de voz
T2	DS-2	6.312	96	Teléfono con banda de voz y teléfono visual
T3	DS-3	46.304	672	Teléfono con banda de voz, teléfono visual y televisión de calidad comercial
T4M	DS-4	274.176	4032	Igual que T3, pero con más capacidad
T5	DS-5	560.160	8064	Igual que T4, pero con más capacidad

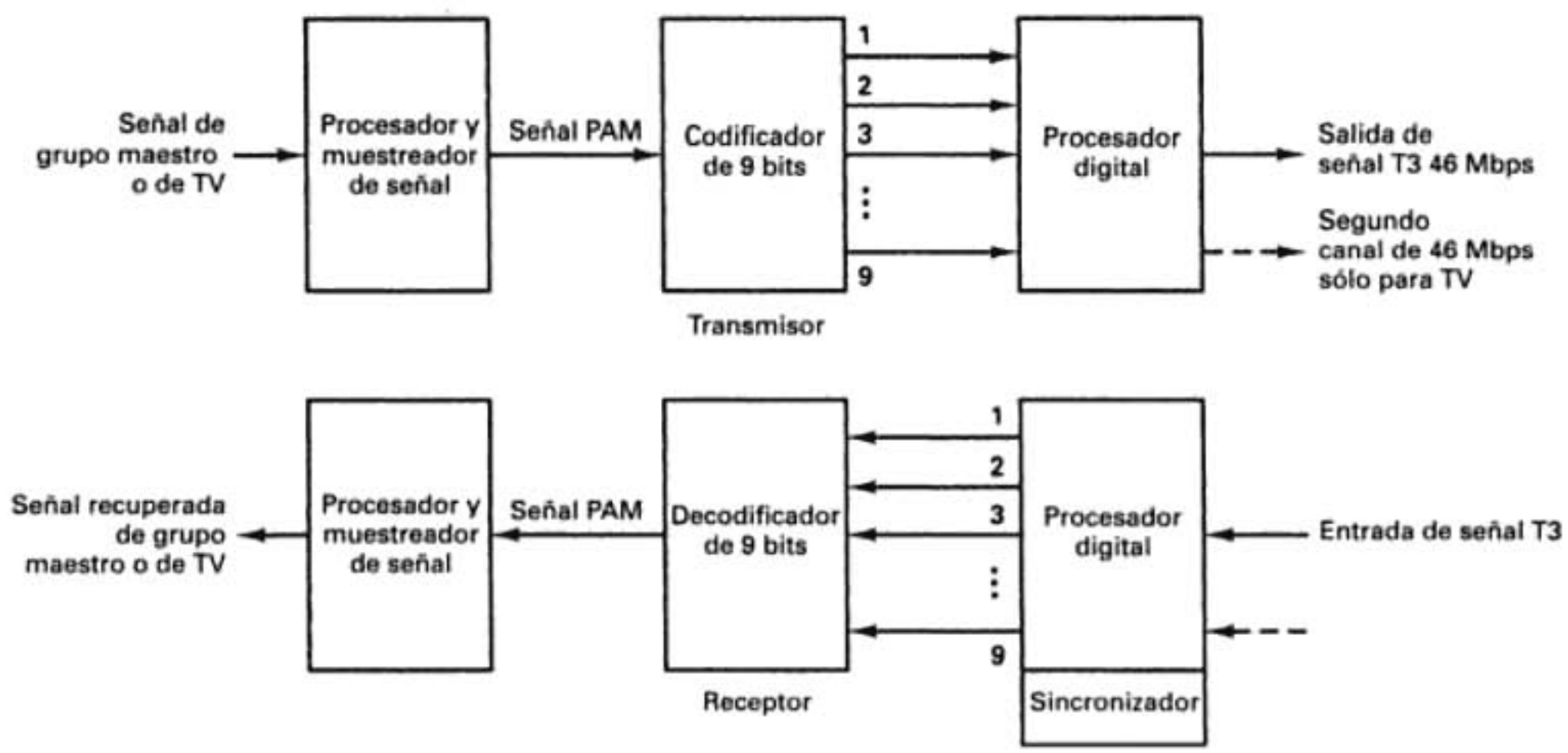


FIGURA 16-14 Diagrama de bloques de una terminal digital de grupo maestro o de televisión comercial

de señal proporciona desplazamiento de frecuencia para las señales de grupo maestro (las desplaza de una banda de 564 a 3084 kHz a una de 0 a 2520 kHz) y la restauración de la señal de televisión a cd. Al desplazar la banda de grupo maestro, es posible muestrearla a una frecuencia de 5.1 MHz. El muestreo de la televisión comercial se hace al doble de esa frecuencia, es decir, a 10.2 MHz.

Para cumplir con los requisitos de transmisión, se usa un código PCM de 9 bits para digitalizar cada muestra del grupo maestro o señal de televisión. La salida digital de la terminal es, entonces, de unos 46 Mbps para el grupo maestro y 92 Mbps (el doble) para la señal de televisión.

La terminal digital de la fig. 16-14 tiene tres funciones específicas: 1) convierte los datos paralelos de la salida del codificador a datos serie; 2) inserta bits de sincronización de trama y 3) convierte la señal binaria en serie en una forma más adecuada para su transmisión. Además, para la terminal de televisión comercial, la señal digital de 92 Mbps se debe dividir en dos señales digitales de 46 Mbps, porque no hay velocidad de línea de 92 Mbps en la jerarquía digital.

Terminal de teléfono visual

En esencia, un *teléfono visual* es una transmisión de video de baja calidad entre suscriptores no dedicados. Por razones económicas, se prefiere codificar una señal de teléfono visual en la capacidad T2 de 6.312 Mbps, bastante menor que las señales de transmisión comercial. De esta

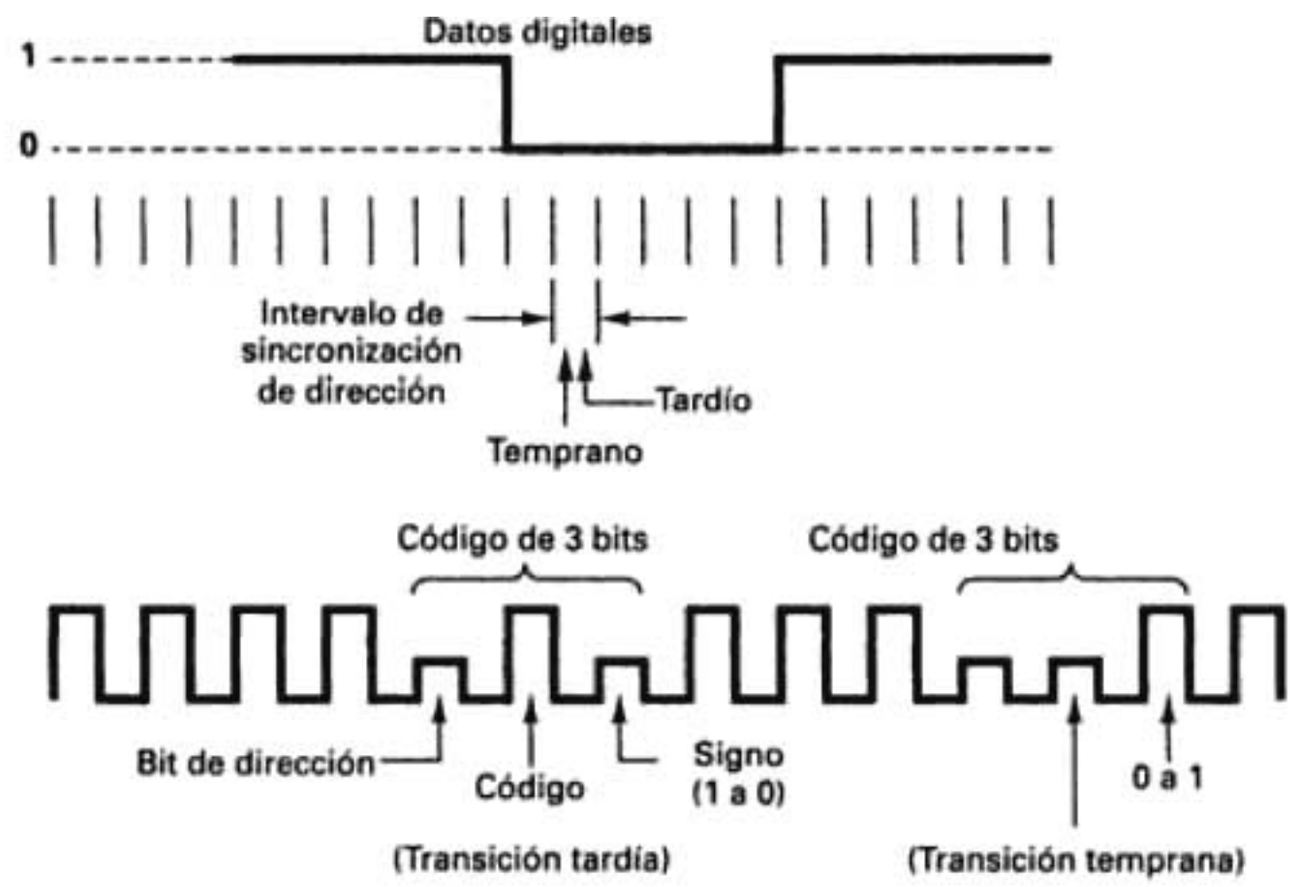


FIGURA 16-15 Formato de codificación de datos

forma se reduce en forma apreciable el costo, y el servicio es más asequible. Al mismo tiempo, permite la transmisión de un detalle, contraste y resolución adecuados, que satisfacen al suscriptor promedio del teléfono visual. El servicio de teléfono visual se adapta en forma ideal a un código PCM diferencial. La PCM diferencial se parece a la PCM convencional, pero no se transmite la magnitud exacta de una muestra. En su lugar, sólo se codifica y se transmite la diferencia entre esa muestra y la anterior. Para codificar la diferencia entre las muestras se requieren bastante menos bits que para codificar toda la muestra real.

Terminal de datos

La parte del tráfico de comunicaciones donde intervienen los datos (señales distintas de la voz) se encuentra en crecimiento exponencial. También, en la mayoría de los casos, las frecuencias de datos generadas por cada suscriptor individual son bastante menores que las capacidades de las líneas digitales. Por lo anterior, es lógico que se diseñen terminales que transmitan señales de datos de varias fuentes, por la misma línea digital.

Las señales de datos se podrían muestrear en forma directa; sin embargo, para eso se necesitarían frecuencias excesivas de muestreo, y se ocasionarían frecuencias de transmisión de bits demasiado altas, en especial para secuencias de datos con pocas o ninguna transición. Un método más eficiente es uno que codifica los tiempos de transición. Ese método se muestra en la fig. 16-15. En el formato de codificación que se ilustra, se usa un código de tres bits para identificar cuándo suceden transiciones en los datos, y si la transición es de un 1 a un 0 o viceversa. El primer bit del código se llama bit de dirección. Cuando este bit es un 1 lógico, indica que no hubo transición; un 0 lógico indica que sí hubo transición. El segundo bit indica si la transición ocurrió durante la primera mitad (0) o la segunda mitad (1) del intervalo de muestra. El tercer bit indica el signo o la dirección de la transición: un 1 para este bit indica una transición de 0 a 1, y un 0 indica una transición de 1 a 0. En consecuencia, cuando no hay transiciones en los datos, se transmite una señal con sólo unos. Sería suficiente sólo la transmisión del bit de dirección; sin embargo, el bit de signo permite contar con cierto grado de protección contra errores, y limita la propagación de los errores (es cuando un error provoca un segundo error, y así sucesivamente). La eficiencia aproximada de este formato es 33%; hay tres bits de código por cada bit de datos. La ventaja de usar un formato codificado y no los datos originales, es que los datos codificados sustituyen a la voz, con más eficiencia, en los sistemas analógicos. Sin este formato de codificación, para transmitir una señal de datos de 250 kbps se requiere el mismo ancho de banda que el necesario para transmitir 60 canales de voz con multiplexado analógico. En este formato codificado, una señal de datos de 50 kbps desplaza a tres canales codificados PCM de 64 kbps, y una corriente de datos de 250 kbps sólo desplaza 12 canales de banda de voz.

CODIFICACIÓN DE LÍNEA

La *codificación de línea* consiste en convertir niveles lógicos normalizados (TTL, CMOS y semejantes) a una forma más adecuada para su transmisión por línea telefónica. En esencia, se deben tener en cuenta seis factores principales al seleccionar un formato de codificación de línea:

1. Voltajes de transmisión y componente de CD.
2. Ciclo de trabajo.
3. Ancho de banda.
4. Recuperación de reloj.
5. Detección de errores.
6. Facilidad de detección y decodificación.

Voltajes de transmisión y componente de CD

Los voltajes o niveles de transmisión se pueden clasificar como *unipolares* (UP) o *bipolares* (BP). La transmisión unipolar de datos binarios implica transmitir sólo un nivel de voltaje distinto de cero (por ejemplo, +V para 1 lógico y 0 V o tierra para un 0 lógico). En la transmisión bipolar intervienen dos niveles de voltaje distintos de cero (por ejemplo, +V para un 1 lógico y −V para un 0 lógico).

En una línea de transmisión digital, es más eficiente, en cuando a potencia, codificar datos binarios con voltajes cuya magnitud sea igual, pero de polaridad opuesta y balanceada simétricamente respecto a 0 V. Por ejemplo, suponiendo una resistencia de 1 Ω y un 1 lógico para +5 V y un 0 lógico para 0 V, la potencia promedio requerida es 12.5 W (suponiendo probabilidades iguales de ocurrencia de 1 y 0). Si el nivel 1 lógico es para +2.5 V y el 0 lógico es para −2.5 V, la potencia promedio sólo es 6.25 W. Así, al usar voltajes bipolares simétricos, la potencia promedio se reduce en un 50%.

Ciclo de trabajo

También se puede usar el *ciclo de trabajo* de un pulso binario para clasificar el tipo de transmisión. Si el pulso binario se mantiene durante todo el tiempo del bit, se llama *sin regreso a cero* o *no regreso a cero* (NRZ, de *nonreturn to zero*). Si el tiempo activo del pulso binario ocupa menos que el 100% del tiempo del bit, se trata de un *regreso* o *retorno* a cero (RZ), de *return to zero*).

Los voltajes de transmisión unipolar y bipolar, y la codificación con regreso a cero y sin regreso a cero se pueden combinar de varias maneras, para obtener determinado esquema de codificación de línea. La fig. 16-16 muestra cinco posibilidades de codificación de línea.

En la fig. 16-16a, sólo hay un nivel de voltaje distinto de cero (+V = 1 lógico); un voltaje cero simplemente implica un 0 binario. También, cada 1 lógico mantiene el voltaje positivo durante todo el tiempo del bit (ciclo de trabajo de 100%). En consecuencia, la fig. 16-16a representa una señal unipolar sin regreso a cero (UPNRZ). En la fig. 16-16b hay dos voltajes distintos de cero (+V = 1 lógico, y −V = 0 lógico), y se usa un ciclo de trabajo de 100%. La fig. 16-16b representa una señal bipolar sin regreso a cero (BPNRZ). En la fig. 16-16c sólo se usa un voltaje distinto de cero, pero cada pulso sólo está activo durante el 50% del tiempo de bit. En consecuencia, la fig. 16-16c representa una señal unipolar con regreso a cero (UPRZ). En la fig. 16-16d, hay dos voltajes distintos de cero (+V = 1 lógico y −V = 0 lógico). También, cada pulso sólo está activo el 50% del tiempo total de bit. Por lo tanto, la fig. 16-16d representa una señal bipolar con regreso a cero (BPRZ). En la fig. 16-16e de nuevo hay dos niveles de voltaje distintos de cero (–V y +V), pero aquí ambas polaridades representan un 1 lógico, y 0 V representa un 0 lógico. A este método de codificación se le llama *inversión alterna de marca* (AMI, de *alternate mark inversion*). En las transmisiones AMI, cada 1 lógico sucesivo se invierte en polaridad respecto al 1 lógico anterior. Como se usa el retorno a cero, a esta técnica de codificación se le llama *bipolar con regreso a cero e inversión alterna de marca* (BPRZ-AMI).

En la codificación NRZ, una cadena larga, de unos o ceros, produce una condición en la que el receptor puede perder su referencia de amplitud, para tener una discriminación óptima entre los 1 y 0 recibidos. A esta condición se le llama *variación de CD*. El problema también se puede presentar cuando hay un desequilibrio importante en la cantidad de unos y ceros transmitidos. La fig. 16-17 muestra cómo se produce la variación de cd debido a una larga cadena de unos lógicos sucesivos. Se ve que después de esa cadena larga, son más probables los errores de 1 a 0

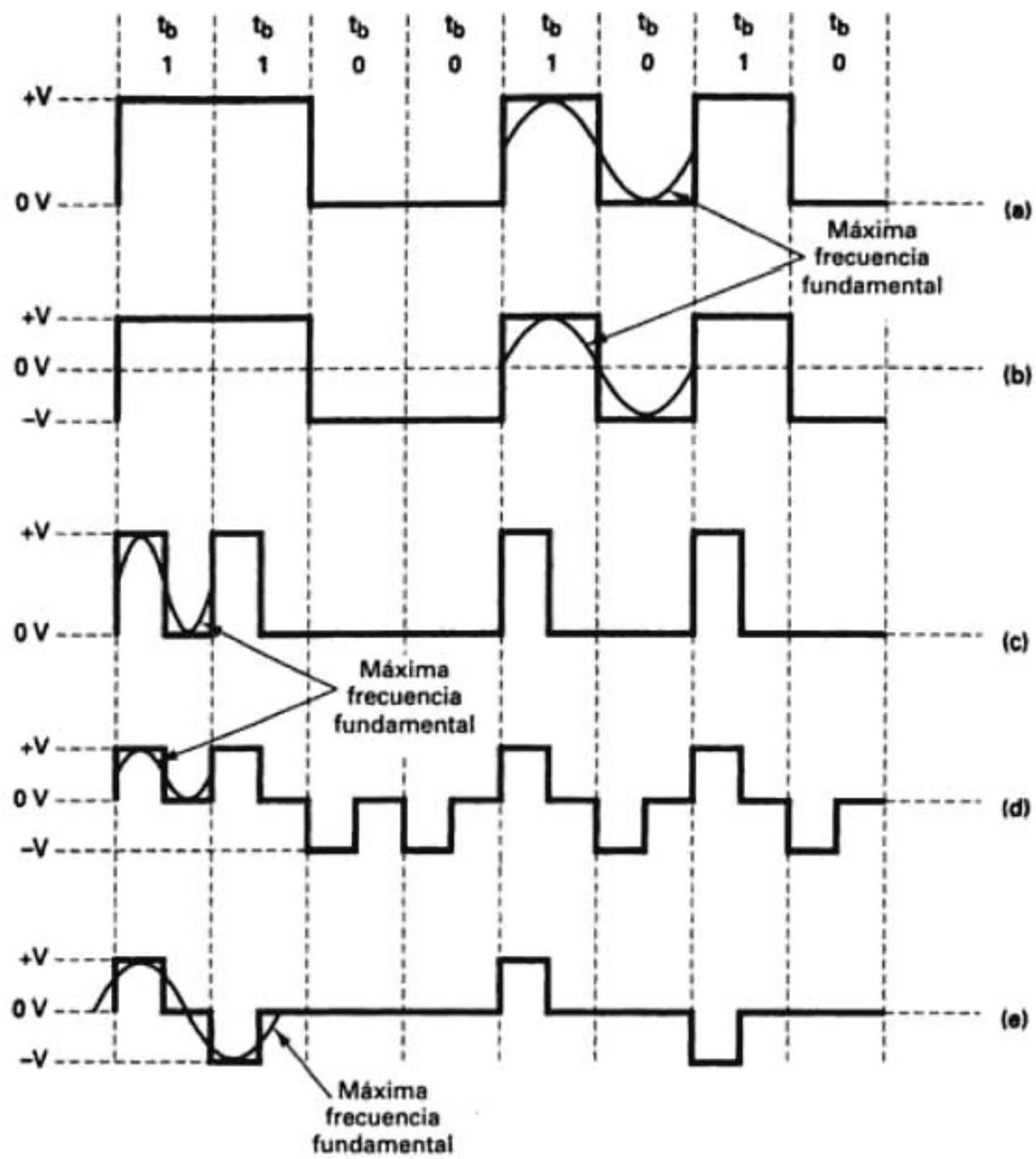


FIGURA 16-16 Formatos de codificación de línea: (a) UPNRZ; (b) BPNRZ; (c) UPRZ; (d) BPRZ; (e) BPRZ-AMI

FIGURA 16-17 Variación de CD

que de 0 a 1. De igual modo, las cadenas largas de ceros aumentan la probabilidad de un error de 0 a 1.

El método de codificación de línea que se use determina el ancho mínimo de banda necesario para la transmisión, la facilidad con la que se puede extraer un reloj de ella, la facilidad de decodificarla, el nivel promedio de cd, y si cuenta con un método conveniente para extraer errores.

Consideraciones de ancho de banda

Para determinar el ancho mínimo de banda necesario para propagar una señal codificada en línea, se debe determinar la mayor frecuencia fundamental asociada con ella (véase la fig. 16-16). La frecuencia fundamental máxima se determina con el peor de los casos (transición más rápida) de secuencia de bits. En UPNRZ la condición del peor de los casos es una frecuencia de 1/0

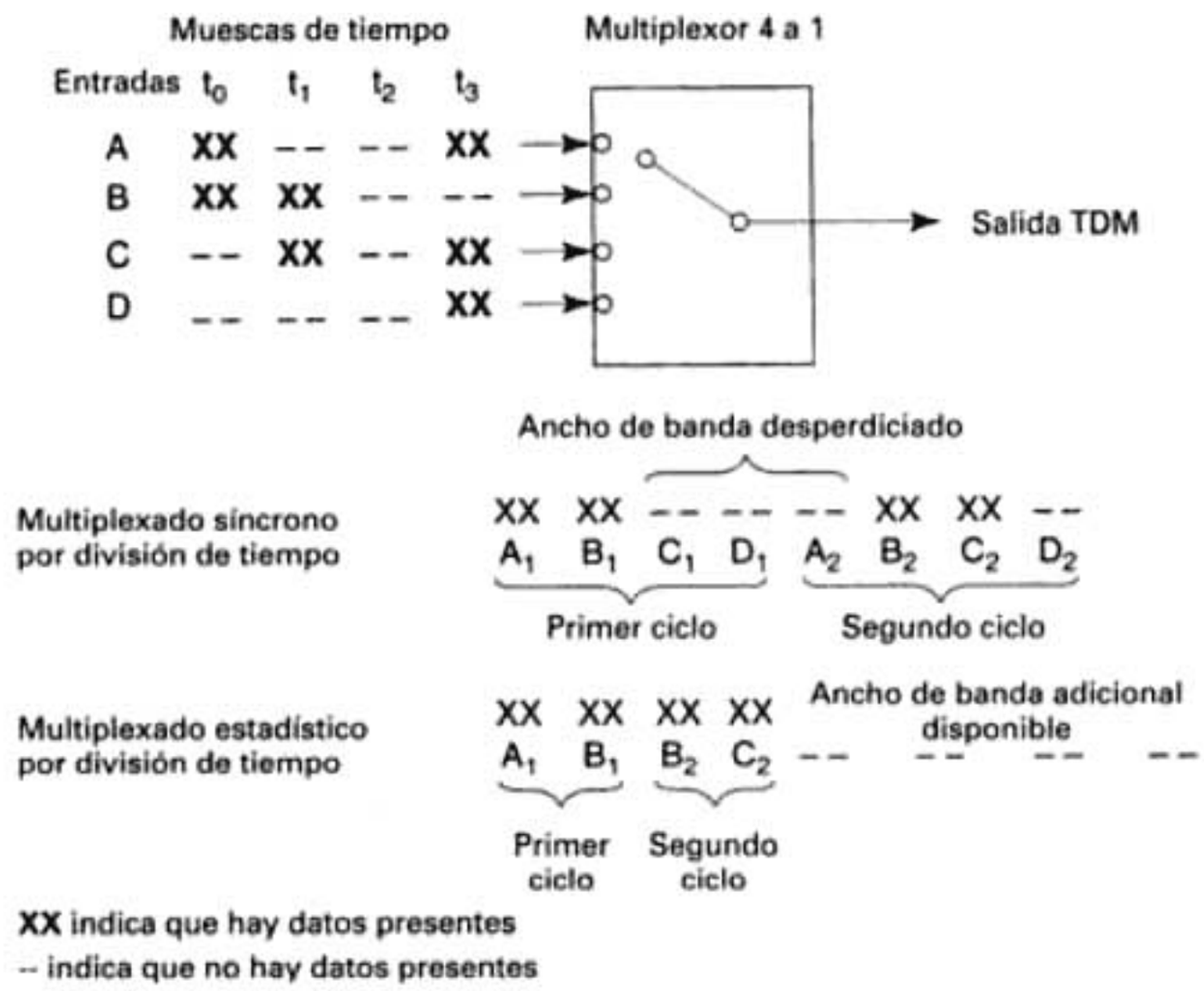


FIGURA 16-22 Comparación entre TDM síncrona y TDM estadística

cuando todas las terminales estén activas. Lo mismo sucede en los sistemas PCM-TDM que conducen conversaciones telefónicas en grado de voz. La conversación normal implica, generalmente, que la conversación sólo se transfiera en una dirección al mismo tiempo, y que haya pausas importantes en las secuencias normales de conversación. Así, hay mucho tiempo que se desperdicia dentro de cada trama TDM. Sin embargo, hay una alternativa eficiente para TDM síncrona, llamada TDM estadística, que a veces se llama TDM asíncrona o inteligente.

Un multiplexor TDM estadístico aprovecha las interrupciones naturales en las transmisiones asignando, en forma dinámica, muescas de tiempo según la demanda. Así como en el multiplexor de un sistema TDM síncrono, un multiplexor estadístico tiene una cantidad finita de líneas de datos de baja velocidad y una línea de salida de alta velocidad, de datos multiplexados, y cada línea de entrada tiene su propia memoria o búfer. En el multiplexor estadístico hay n líneas de entrada, pero sólo hay k muescas de tiempo disponibles dentro de la trama TDM (siendo $k < n$). El multiplexor recorre las memorias de entrada y colecta datos hasta llenar una trama, y en ese momento se transmite la trama. En el extremo receptor sucede lo mismo: hay más líneas de salida que muescas de tiempo dentro de la trama TDM. El demultiplexor saca los datos de las muescas de tiempo y los distribuye a sus memorias correspondientes de salida.

La TDM estadística aprovecha que los dispositivos conectados a las entradas y salidas no transmiten ni reciben todas todo el tiempo, y que la frecuencia de datos de la línea multiplexada es menor que las frecuencias combinadas de datos de los dispositivos conectados. En otras palabras, un multiplexor estadístico TDM requiere menor velocidad de datos que la que necesita un multiplexor síncrono para dar servicio a la misma cantidad de dispositivos de entrada. También, un multiplexor estadístico TDM que funciona a la misma velocidad de transmisión que uno TDM puede dar servicio a más dispositivos.

La fig. 16-22 muestra una comparación entre TDM estadística y síncrona. Se muestran cuatro fuentes de datos, A, B, C y D, y cuatro muescas de tiempo o épocas, t_0, t_1, t_2 y t_3. El multiplexor síncrono tiene una frecuencia de salida de datos igual a cuatro veces la frecuencia de datos de cada uno de los canales de entrada. Durante cada trama, los datos se colectan de las cuatro fuentes, y se transmiten sin importar si hay o no datos de entrada. Como se ve en la figura, durante la muesca de tiempo t_0, los canales C y D no tenían datos de entrada, y se produjo una trama TDM transmitida sin información, en las muescas de tiempo C y D. Sin embargo, en el multiplexor estadístico no se transmiten las muescas de tiempo vacías. Así, durante la primera muesca de

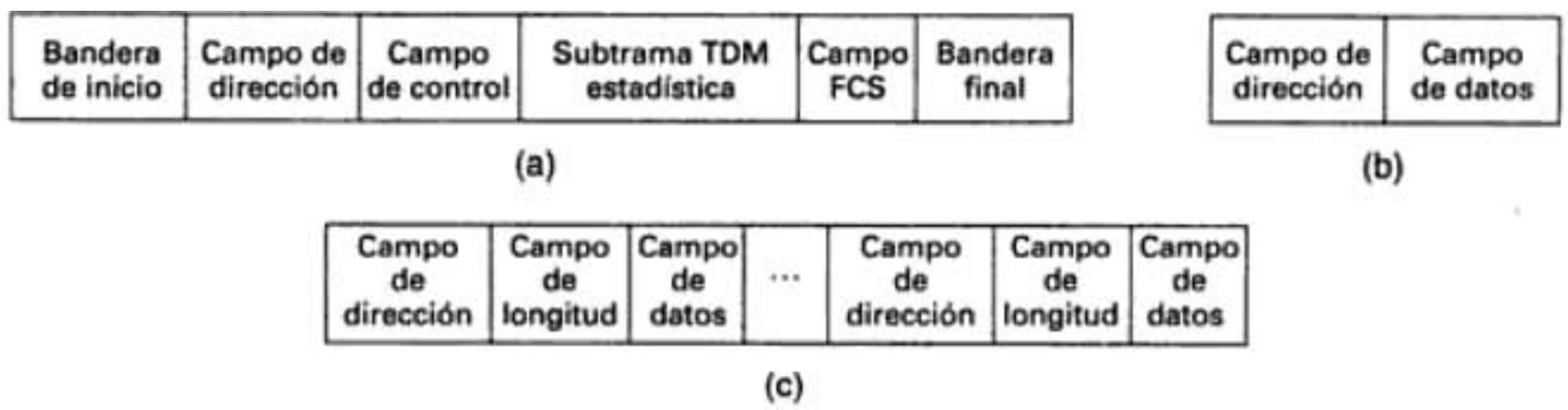


FIGURA 16-23 Formato de trama en TDM estadística: a) trama general TDM estadística; (b) una fuente por trama; (c) varias fuentes por trama

tiempo sólo se transmite información de los canales A y B. Sin embargo, una desventaja del formato estadístico es que se pierde el significado posicional de las muescas de tiempo. No hay forma de conocer por adelantado cuáles canales de datos estarán en cuál muesca de tiempo. Como los datos llegan y son distribuidos en forma impredecible en las memorias de recepción, es necesaria la información de dirección para asegurar la entrega adecuada. Para eso se necesitan más indirectos por muesca de tiempo con TDM estadística, porque cada ranura debe contener una dirección, además de los datos.

El formato de trama que usa un multiplexor estadístico tiene un impacto directo sobre la eficiencia del sistema. Es obvio que se desea minimizar los indirectos para mejorar la capacidad de datos. En el caso normal, un sistema TDM estadístico usa un protocolo síncrono, como el HDLC. Se deben incluir los bits de control para el multiplexado en la trama HDLC. La fig. 16-23a muestra el formato general de trama para un multiplexor estadístico TDM. La fig. 16-23b muestra la trama cuando sólo transmite una fuente de datos. El dispositivo transmisor se identifica con una dirección. La longitud del campo de datos es variable, y sólo se limita por la longitud de la trama. Este esquema funciona bien cuando la carga es ligera, pero es muy ineficiente cuando las cargas son grandes. La fig. 16-23c muestra una forma de mejorar la eficiencia, permitiendo incluir más de una fuente de datos dentro de una sola trama. Sin embargo, cuando hay varias fuentes, es necesario algún medio de especificar la longitud de la corriente de datos de cada fuente. En consecuencia, la trama estadística consiste en secuencias de campos de datos identificados con una dirección y una cuenta de bits. Hay varias técnicas con las que se puede mejorar la eficiencia. El campo de dirección se puede acortar usando un esquema de direccionamiento relativo, en el que cada dirección especifica la posición de la fuente actual, en relación con la fuente transmitida anteriormente, módulo la cantidad total de fuentes. Con direccionamiento relativo, un campo de dirección de ocho bits se puede reemplazar por uno de cuatro bits.

Otro método de refinar la trama es usar una etiqueta de dos bits con el campo de longitud. Los valores binarios 00, 01, 10 y 11 corresponden a un campo de datos de 1, 2 o 3 bits, y no se necesita campo de longitud. El código 11 indica que se incluye un campo de longitud.

MULTIPLEXADO POR DIVISIÓN DE FRECUENCIA

En el *multiplexado por división de frecuencia* (FDM, de *frequency-division multiplexing*), se convierte cada fuente de varias que originalmente ocupaban el mismo espectro de frecuencias, a una banda distinta de frecuencias, y se transmite en forma simultánea por un solo medio de transmisión. Así se pueden transmitir muchos canales de banda relativamente angosta por un solo sistema de transmisión de banda ancha.

El FDM es un esquema análogo de multiplexado; la información que entra a un sistema FDM es analógica y permanece analógica durante su transmisión. Un ejemplo de FDM es la banda comercial de AM, que ocupa un espectro de frecuencias de 535 a 1605 kHz. Cada estación tiene una señal de información con un ancho de banda de 0 a 5 kHz. Si se transmitiera el audio de cada estación con el espectro original de frecuencias, sería imposible separar una estación de las demás. En lugar de ello, cada estación modula por amplitud una frecuencia distinta de portadora, y produce una señal de doble banda lateral de 10 kHz. Como las frecuencias de las

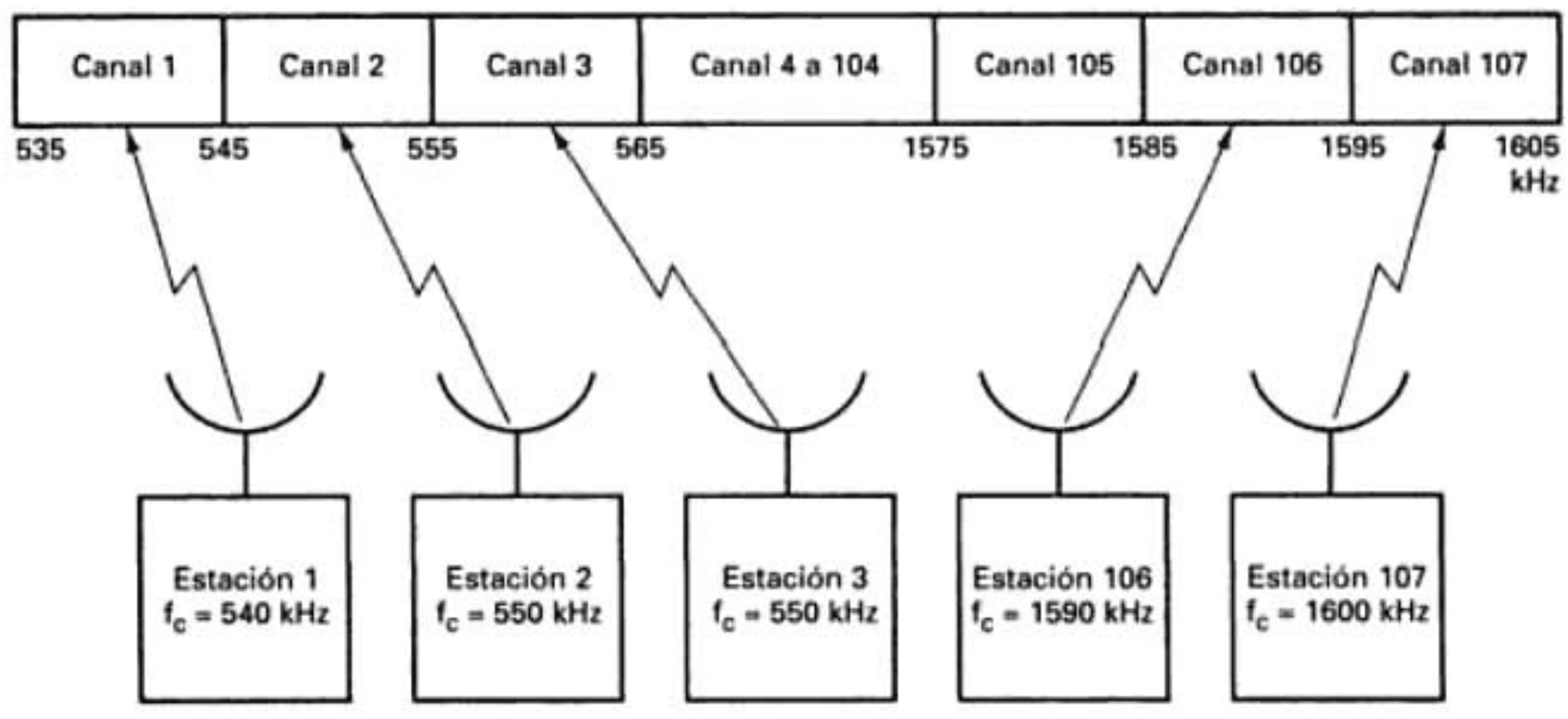


FIGURA 16-24 Multiplexado por división de frecuencia en estaciones comerciales de AM

portadoras de estaciones adyacentes están separadas por 10 kHz, la banda comercial total de AM se divide en 107 muescas de frecuencia de 10 kHz, apiladas una sobre otra en el dominio de la frecuencia. Para recibir determinada estación, sólo se sintoniza el receptor con la banda de frecuencias asociada con las transmisiones de esa estación. La fig. 16-24 muestra cómo están multiplexadas las estaciones de radio comerciales de AM, por división de frecuencias, y se transmiten por un solo medio de transmisión: el espacio libre.

Hay muchas otras aplicaciones de la FDM, por ejemplo, la FM comercial y las emisoras de televisión, así como los sistemas de telecomunicaciones de alto volumen. Dentro de cualquiera de las bandas de transmisión comercial, las transmisiones de cada estación son independientes de las demás. En consecuencia, el proceso de multiplexado (apilado) se hace sin que haya sincronización entre las estaciones. En los sistemas de comunicaciones telefónicas de alto volumen, muchos canales telefónicos de banda de voz se pueden originar en una fuente común y terminar en un destino común. El equipo de terminal de fuente y destino se parece más a un *sistema electrónico de conmutación* (ESS, de *electronic switching system*) de alta capacidad. Debido a la posibilidad de que haya una gran cantidad de canales de banda angosta que se originen y terminen en el mismo lugar, todas las operaciones de multiplexado y demultiplexado se deben sincronizar.

JERARQUÍA FDM DE AT&T

Aunque AT&T ya no es el único proveedor de portadora común de larga distancia en Estados Unidos, sigue proporcionando la mayor parte de los servicios de larga distancia y, aunque no sea más que por su tamaño desmesurado, se ha convertido en la organización normalizadora para la industria telefónica en Norteamérica.

La red de comunicaciones de AT&T en Estados Unidos se subdivide en dos grupos: *corta distancia* y *larga distancia*. La portadora T1 explicada antes en este capítulo es un ejemplo de sistema de comunicaciones con portadora de corta distancia.

Comunicaciones de larga distancia con FDM

La fig. 16-25 muestra la jerarquía norteamericana FDM de AT&T para comunicaciones de larga distancia. Sólo se ve una terminal transmisora, aunque en la terminal receptora se debe hacer el conjunto completo de funciones inversas.

Canal de mensaje

El *canal de mensaje* es la piedra constructiva de la jerarquía FDM. El canal básico de mensaje era, en un principio, para transmisión de voz, aunque ahora incluye todas las transmisiones que

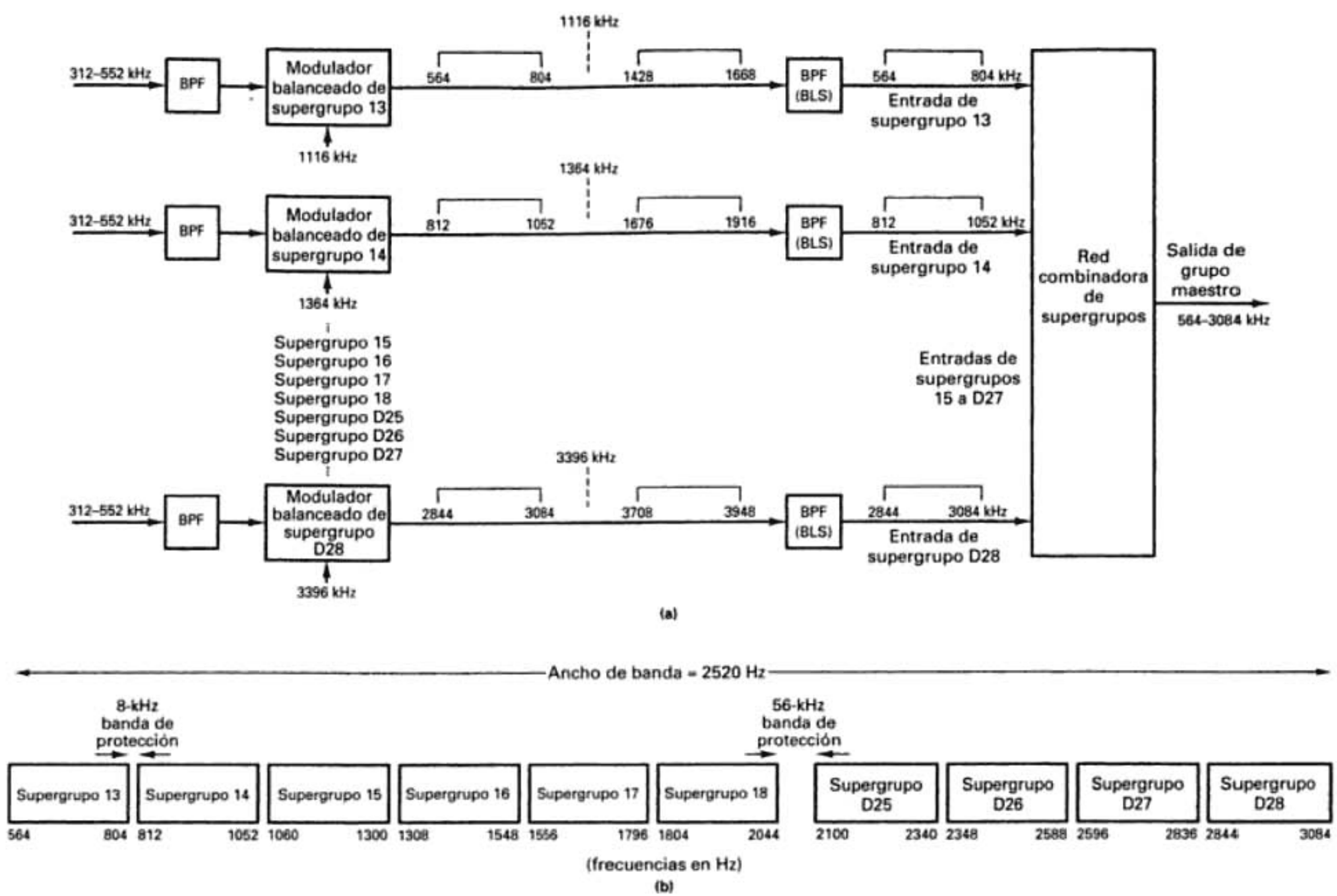


FIGURA 16-28 Formación de un grupo maestro U600: (a) diagrama de bloques de banco de supergrupos y red combinadora; (b) espectro de salida

TABLA 16-8 Frecuencias de portadora de supergrupo para un grupo maestro U600

Supergrupo	Frecuencia de portadora (kHz)
13	1116
14	1364
15	1612
16	1860
17	2108
18	2356
D25	2652
D26	2900
D27	3148
D28	3396

Para el supergrupo 13,

$$f_{sal} = 1116 \text{ kHz} - (312 \text{ kHz a } 552 \text{ kHz}) = 564 \text{ kHz a } 804 \text{ kHz}$$

Para el supergrupo 14,

$$f_{sal} = 1364 \text{ kHz} - (312 \text{ kHz a } 552 \text{ kHz}) = 812 \text{ kHz a } 1052 \text{ kHz}$$

Para el supergrupo D28,

$$f_{sal} = 3396 \text{ kHz} - (312 \text{ kHz a } 552 \text{ kHz}) = 2844 \text{ kHz a } 3084 \text{ kHz}$$

Las salidas de los 10 moduladores de supergrupo se suman en el sumador lineal para obtener el espectro total de grupo maestro de la fig. 16-28b (de 564 a 3084 kHz). Nótese que entre dos supergrupos adyacentes cualquiera hay una banda vacía de frecuencias, que no se incluye dentro de banda alguna de supergrupo. Estos huecos se llaman *bandas de protección*. Son necesarios, porque el proceso de desmultiplexado se hace por filtrado y conversión descendente. Sin las bandas de protección sería difícil separar un supergrupo de otro adyacente. Las bandas de protección reducen el *factor de calidad* (Q) necesario para hacer el filtrado necesario. La banda de protección es de 8 kHz entre todos los supergrupos, excepto el 18 y el D25, donde es de 56 kHz. En consecuencia, el ancho de banda de un grupo maestro U600 es 2520 kHz (de 564 a 3084 kHz), mayor que el necesario para apilar 600 canales de banda de voz (600 × 4 kHz = 2400 kHz).

Las bandas de protección no fueron necesarias entre grupos adyacentes, porque las frecuencias de grupo son bastante bajas, y es relativamente fácil fabricar filtros pasabandas para separar un grupo de los demás.

En el banco de canales, el filtro antialias en la entrada del canal pasa una banda de 0.3 a 3 kHz. La separación entre las frecuencias de canales adyacentes es 4 kHz. Por consiguiente, hay una banda de protección entre los canales adyacentes. Esto se ve en la fig. 16-29.

Grupo maestro L600

En un grupo maestro L600, se combinan 10 supergrupos como en el grupo maestro U600, pero las frecuencias de portadora de supergrupo son menores. La tabla 16-9 es una lista de las frecuencias de portadora de supergrupo para un grupo maestro L600. En estos grupos maestros, el espectro de banda de base compuesto ocupa una banda de frecuencias inferiores a las del grupo maestro tipo U (fig. 16-30). Un grupo maestro L600 ya no se sigue multiplexando y, entonces, la capacidad máxima de canal para un sistema de microondas o de cable coaxial que usa un solo grupo maestro L600 es 600 canales de banda de voz.

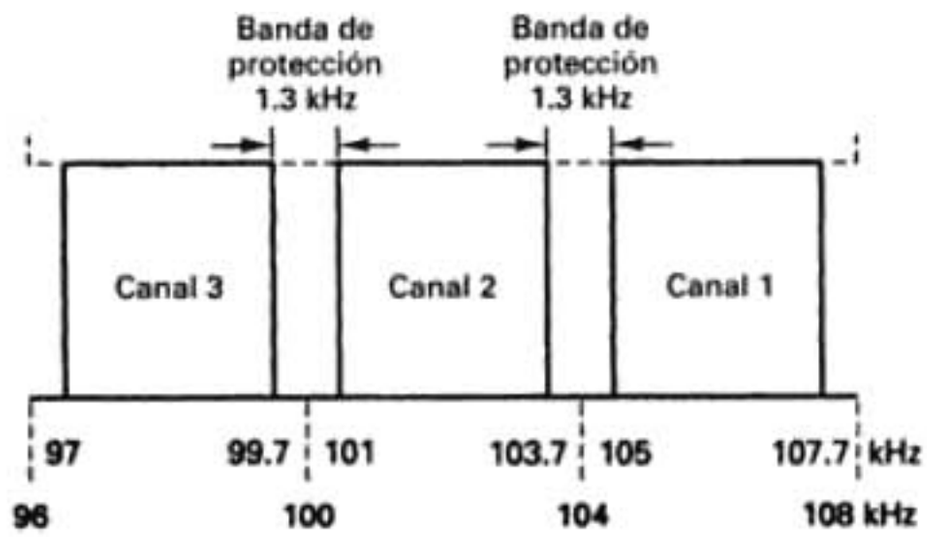


FIGURA 16-29 Bandas de protección de canales

TABLA 16-9 Frecuencias de portadora de supergrupo para un grupo maestro L600

Supergrupo	Frecuencia de portadora (kHz)
1	612
2	Directa
3	1116
4	1364
5	1612
6	1860
7	2108
8	2356
9	2724
10	3100

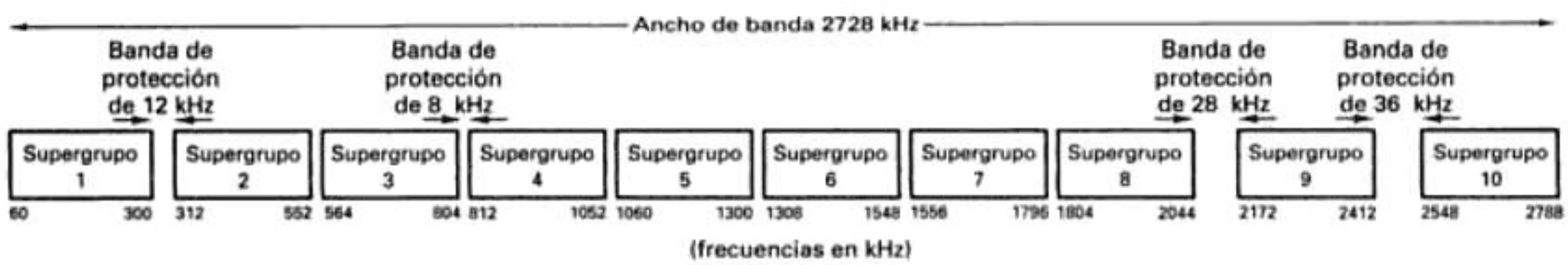


FIGURA 16-30 Grupo maestro L600

Formación de un canal de radio

Un *canal de radio* abarca ya sea un solo grupo maestro L600, o hasta tres grupos maestros U600 (1800 canales de banda de voz). La fig. 16-31a muestra cómo se forma una señal de banda de base FDM compuesta de 1800 canales, para transmitir por un solo canal de radio de microondas. El grupo maestro 1 se transmite en forma directa tal cual, mientras que los grupos maestros 2 y 3 pasan por un paso adicional de multiplexado. Los tres grupos maestros se suman en una red combinadora de grupo maestro para producir el espectro de salida que se ve en la fig. 16-31b. Nótese la banda de protección de 80 kHz entre los grupos maestros adyacentes.

El sistema de la fig. 16-31 se puede aumentar de 1800 a 1860 canales de banda de voz, agregando un supergrupo adicional, el 12, en forma directa al grupo maestro 1. El supergrupo adicional de 312 kHz a 552 kHz amplía el espectro compuesto de salida, que es de 312 a 8284 kHz.

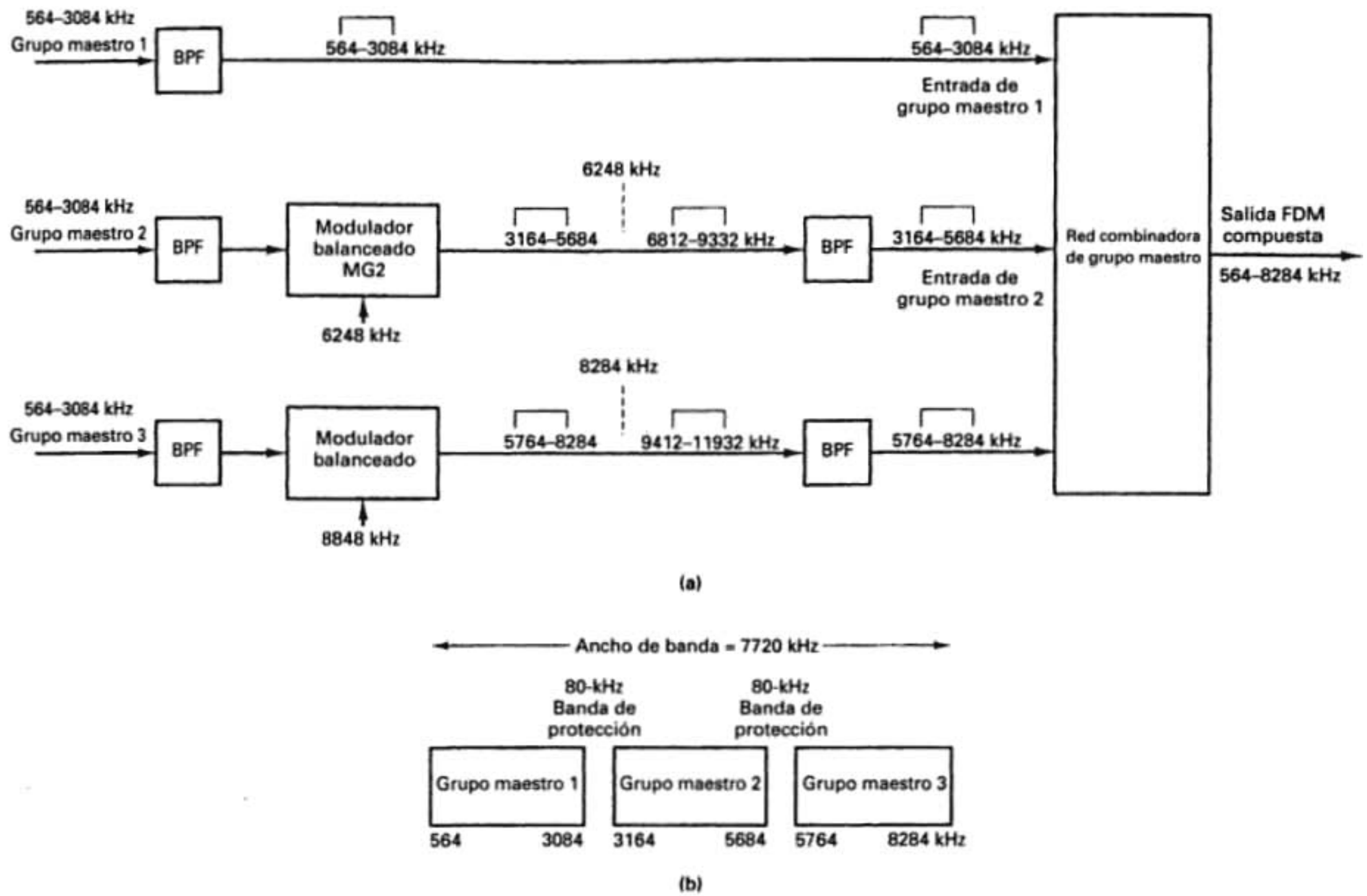


FIGURA 16-31 Canal de radio de tres grupos maestros: (a) diagrama de bloques; (b) espectro de salida

MULTIPLEXADO POR DIVISIÓN DE LONGITUD DE ONDA

El *multiplexado por división de longitud de onda* (WDM, de *wavelength-division multiplexing*) se llama a veces *multiplexado por división de onda*. Como la longitud de onda y la frecuencia se relacionan en forma estrecha, el multiplexado por división de longitud de onda se parece al de división de frecuencia (FDM). Se parecen porque se trata de mandar señales de información que ocupen la misma banda de frecuencias, a través de la misma fibra y al mismo tiempo, sin que interfieran entre sí. Esto se logra modulando diodos láser de inyección que transmiten ondas luminosas muy concentradas a distintas longitudes de onda (es decir, a distintas frecuencias). Por consiguiente, el WDM es acoplar la luz de dos o más longitudes de onda discretas para que entre y salga de una fibra óptica. Cada longitud de onda es capaz de llevar grandes cantidades de información, sea analógica o digital, y la información puede estar ya multiplexada por división de tiempo o de frecuencia. Aunque la información que se usa con los láseres casi siempre consiste en señales digitales multiplexadas por división de tiempo, en la separación de longitudes de onda que se usa con el WDM es análoga a los canales analógicos de radio que funcionan a distintas frecuencias de portadora.

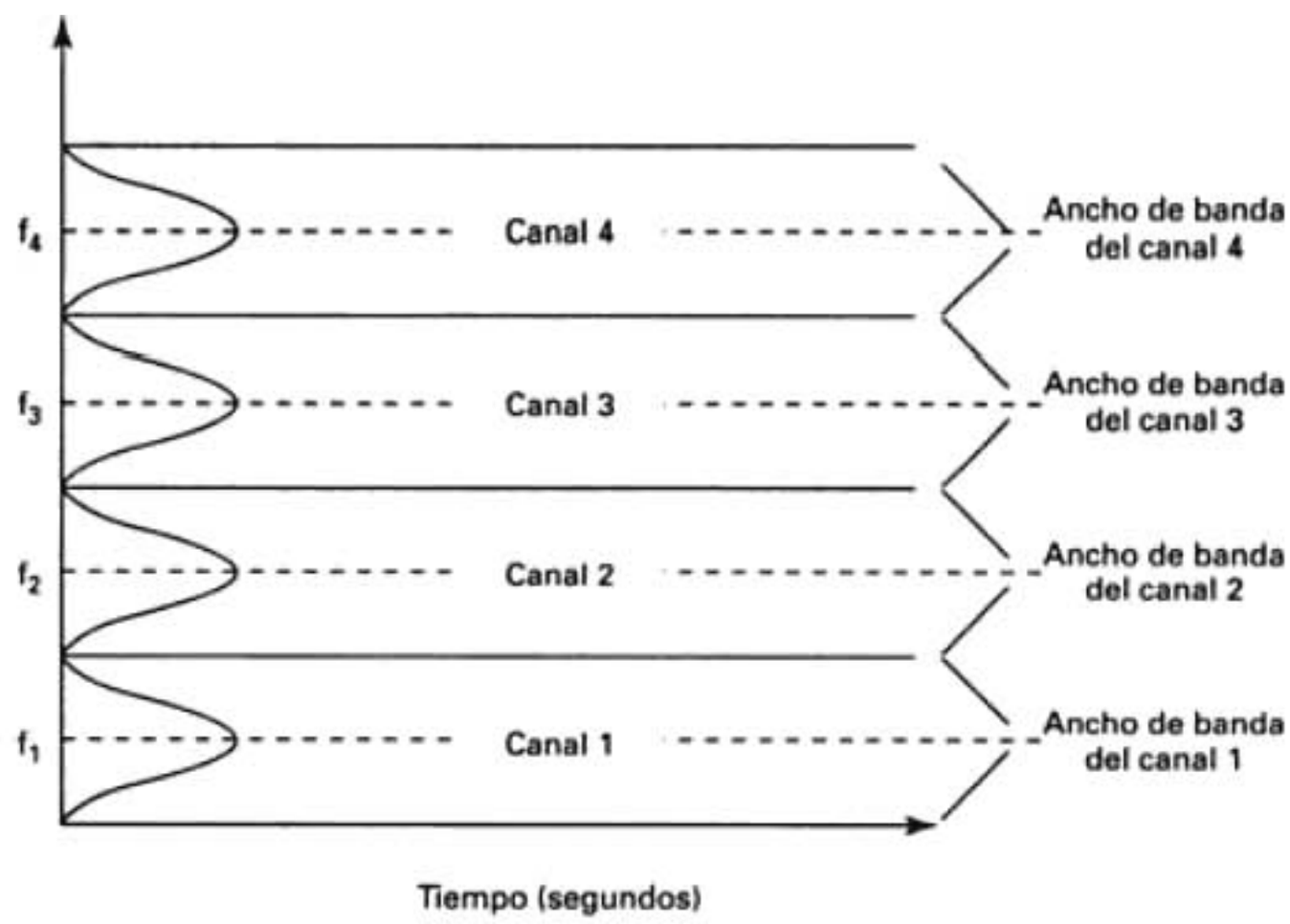
Comparación de multiplexado por división de longitud de onda y por división de frecuencia

Aunque el multiplexado por división de frecuencia y de longitud de onda se basan en los mismos principios, no son lo mismo. La diferencia más obvia es que las frecuencias ópticas (en THz) son mucho mayores que las radiofrecuencias (en MHz y GHz). Sin embargo, es posible que la mayor diferencia sea la forma en que se propagan las dos señales por sus medios respectivos de transmisión. En el FDM, las señales de información con el mismo ancho de banda procedentes de varias fuentes modulan distintas frecuencias, y cada frecuencia tiene su propio circuito modulador, y cada señal de información tiene su propia tasa de modulación. Las señales FDM se propagan al mismo tiempo, por el mismo medio y siguen el mismo camino de transmisión.

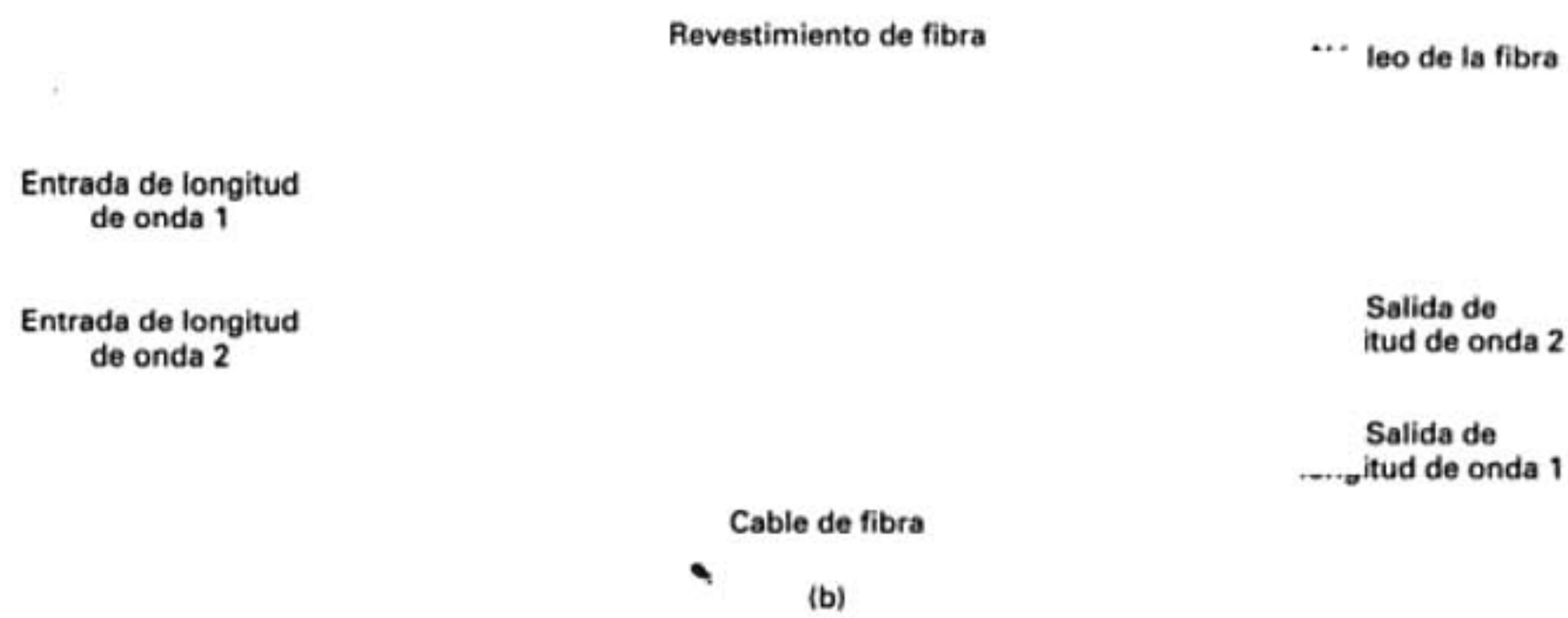
Sin embargo, el principio básico del WDM es algo distinto. Las longitudes de onda distintas de un pulso de luz viajan a través de una fibra óptica a velocidades diferentes; la luz azul se propaga con más lentitud que la luz roja. En los sistemas normales de comunicaciones con fibra óptica, al propagarse la luz por el cable, la dispersión de longitudes de onda hace que las ondas luminosas se repartan y distribuyan su energía por un tiempo mayor. Así, en los sistemas normales de fibra óptica, la dispersión de longitudes de onda origina problemas que establecen limitaciones para la eficiencia del sistema. Sin embargo, en el WDM, la dispersión de longitudes de onda es la esencia de la operación del sistema. En él, las señales de información procedentes de varias fuentes que ocupan el mismo ancho de banda modulan láseres que funcionan a distintas longitudes de onda. Por lo mismo, las señales entran a la fibra al mismo tiempo, y viajan por el mismo medio. Sin embargo, no toman la misma trayectoria por la fibra; como cada longitud de onda toma una ruta de transmisión distinta, cada una de ellas llega al receptor en momentos ligeramente distintos. Así, se pueden codificar datos en forma secuencial en el canal WDM con un solo modulador de datos. El resultado es una serie de arcoiris formados de distintos colores (longitudes de onda), cada uno de unas 20 mil millonésimas de segundo de duración, y se propagan en forma simultánea por el cable. La fig. 16-32 muestra los principios básicos de la propagación de señales FDM y WDM a través de su respectivo medio de transmisión.

D-WDM, longitudes de onda y canales de longitudes de onda

El WDM se logra en general a las longitudes de ondas aproximadas de 1550 nm (1.55 μm) y las frecuencias sucesivas separadas por múltiplos de 100 GHz; es decir, son 100 GHz, 200 GHz, 300 GHz. En 1550 nm y una separación de frecuencias de 100 GHz, la separación de longitudes de onda aproximada es 0.8 nm. Por ejemplo, tres longitudes de onda adyacentes, separadas entre sí por 100 GHz, corresponden a longitudes de onda de 1550.0 nm, 1549.2 nm y 1548.4 nm. Si se usa una técnica de multiplexado llamada *multiplexado por división de onda densa* (D-WDM, de *dense-wave-division multiplexing*), la distancia entre las frecuencias adyacentes es bastante menor. Sin embargo, parece no haber definición oficial exacta de qué significa D-WDM, pero en general, se consideran como WDM normales a los sistemas ópticos que conducen varias señales



(a)



(b)

FIGURA 16-32 (a) Multiplexado por división de frecuencia; (b) multiplexado por división de longitud de onda

ópticas distanciadas más de 200 GHz o 1.6 nm, en la proximidad de 1550 nm, y a los sistemas WDM que conducen varias señales ópticas cerca de 1550 nm con separación menor que 200 GHz se les considera D-WDM. Es obvio que cuanto más longitudes de onda se usen en un sistema WDM, más se acercarán entre sí, y el espectro de longitudes de onda será mayor.

Las ondas luminosas tienen muchas frecuencias (longitudes de onda) y cada frecuencia corresponde a un color diferente. Se han desarrollado transmisores y receptores, para fibras ópticas, que sólo transmiten y reciben un color específico, es decir, una longitud de onda específica, con una frecuencia específica y ancho de banda fijo. El WDM es un proceso en el que distintas fuentes de información (canales) se propagan por una fibra óptica a distintas longitudes de onda, que no interfieren entre sí. En esencia, cada longitud de onda forma un "carril" óptico de la supercarretera de transmisión, y mientras más carriles haya, se puede conducir más tráfico (de voz, datos, video, etc.) por un solo cable de fibra óptica. En contraste, los sistemas convencionales de fibra óptica sólo tienen un canal por cable, que se usa para llevar información que ocupa un ancho de banda relativamente pequeño. Un equipo de investigación de Bell Laboratories construyó un transmisor D-WDM con un solo láser de anillo de fibra, de erbio dopado, de un femtosegundo de duración, que puede transmitir en forma simultánea 206 longitudes de onda

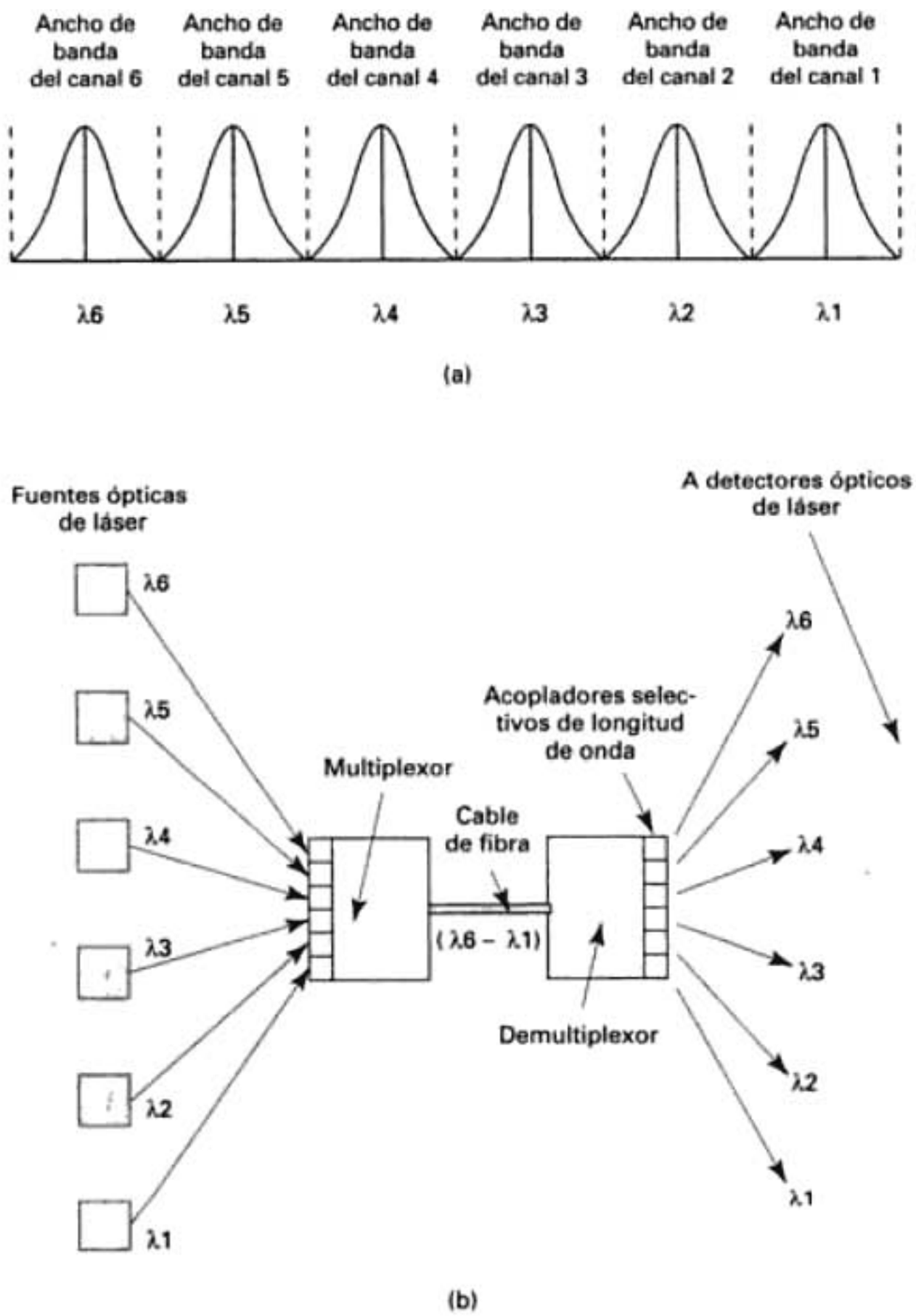


FIGURA 16-33 (a) Espectro de longitudes de onda de un sistema WDM que usa seis longitudes de onda; (b) multiplexado y desmultiplexado de 6 láseres

moduladas digitalmente a través de un solo cable de fibra óptica. Cada longitud de onda (canal) tiene una frecuencia de bits de 36.7 Mbps, y la distancia aproximada entre canales es 36 GHz.

La fig. 6-33a muestra el espectro de longitudes de onda de un sistema WDM que usa seis longitudes de onda, cada una modulada con señales de información de igual ancho de banda. La fig. 16-33b muestra cómo se combinan (multiplexan) las longitudes de onda de seis láseres y a continuación se propagan por un solo cable de fibra óptica, para después separarse (demultiplexarse) en el receptor con acopladores selectivos de longitud de onda. Aunque se ha demostrado que una sola fuente luminosa ultrarrápida puede generar cientos de canales individuales de comunicaciones, en general los sistemas de comunicaciones WDM se limitan entre 2 y 16 canales.

El multiplexado por división de longitud de onda amplía el desempeño de la fibra óptica añadiendo canales a los cables existentes. Cada longitud de onda agregada corresponde a agregar un canal distinto, con su propia fuente de información y su propia rapidez de transmisión de bits. Así, el WDM puede ampliar la capacidad de conducción de información de una fibra hasta miles de gigabits por segundo o más.

Ventajas y desventajas del WDM

Una ventaja obvia del WDM es su mayor capacidad, y con ese multiplexado, también es posible la transmisión dúplex con una sola fibra. Además, en las redes de comunicaciones se usan componentes ópticos que son más sencillos, más fiables y con frecuencia menos costosos que sus contrapartes electrónicas. El WDM tiene la ventaja de ser, en forma inherente, más fácil de reconfigurar, es decir, agregar o quitar canales. Por ejemplo, se han instalado redes de área local con WDM que permiten a los usuarios entrar a la red sólo sintonizando cierta longitud de onda.

También el WDM tiene sus limitaciones. No se pueden poner las señales tan cercanas en el espectro de longitudes de onda, porque se interfieren. Su proximidad depende de los parámetros de diseño del sistema, como por ejemplo si se usa amplificación óptica, y qué técnica de amplificación se usa para combinar y separar las señales de distintas longitudes de onda. La Unión Internacional de Telecomunicaciones adoptó una retícula patrón de frecuencias para D-WDM, con una separación de 100 GHz o de múltiplos enteros de 100 GHz, lo cual, a 1550 nm, corresponde a una separación aproximada de longitudes de onda de 0.8 nm.

En el WDM la intensidad general de la señal debería ser aproximadamente igual para cada longitud de onda. Esta intensidad se ve influida por las características de atenuación de la fibra, y por el grado de amplificación; estos dos parámetros dependen de la longitud de onda. Bajo condiciones normales, las longitudes de onda seleccionadas para un sistema se separan tan poco entre sí que la atenuación varía muy poco entre ellas.

Una diferencia entre FDM y WDM es que el multiplexado WDM se hace a frecuencias ópticas extremadamente altas, mientras que el FDM se hace a las frecuencias relativamente bajas de radio y de banda base. Por consiguiente, las señales de radio que conducen FDM no se limitan a propagarse por un medio de transmisión físico contenido, por ejemplo, a través de un cable óptico. Las señales de radio se pueden propagar a través de casi cualquier medio de transmisión, incluyendo el espacio libre. Por lo anterior, se pueden transmitir en forma simultánea señales de radio a muchos destinos, mientras que las ondas luminosas que conducen el WDM se limitan a un circuito de dos puntos, o a una combinación de muchos circuitos de dos puntos, que sólo pueden estar donde van los cables.

La capacidad de información de una sola fibra óptica se puede aumentar n veces, donde n representa cuántas longitudes de onda distintas se propagan por la fibra al mismo tiempo. Cada longitud de onda en un sistema WDM está modulada por señales de información de distintas fuentes. Por consiguiente, un sistema de comunicaciones que use un solo cable de fibra óptica que propague n longitudes de onda distintas debe utilizar n moduladores y n demoduladores.

Componentes de los circuitos WDM

Los componentes de circuito que se usan en WDM son similares a los de los sistemas convencionales de transmisión por ondas de radio y por conductor metálico; sin embargo, algunos de los nombres que se usan para los acopladores WDM causan confusión.

Multiplexores y demultiplexores por división de longitud de onda. Los *multiplexores* o *combinadores* mezclan o combinan las señales ópticas con distintas longitudes de onda, en una forma que les permite a todas ellas pasar a través de una sola fibra óptica sin interferirse entre sí. Los *demultiplexores* o *divisores* separan las señales de distintas longitudes de onda, en forma parecida a como los filtros separan las señales eléctricas de distintas frecuencias. Los demultiplexores por longitud de onda tienen tantas salidas como longitudes de onda manejen, y cada salida (longitud de onda) va a un destino distinto. Los multiplexores y demultiplexores son los extremos terminales de los sistemas de comunicaciones de fibra óptica.

Multiplexores y demultiplexores por división de longitud de onda, de agregar y quitar. Estos *multiplexores* y *demultiplexores* son parecidos a los normales, pero se ubican en puntos intermedios del sistema. Los multiplexores y demultiplexores de agregar y quitar son dispositivos que separan una longitud de onda de un cable de fibra, y la pasan a otra fibra que va en dirección distinta. Una vez quitada una longitud de onda, se puede reemplazar por una nueva señal de esa misma longitud. En esencia, los multiplexores de agregar y quitar se usan para reconfigurar cables de fibra óptica.

C A P Í T

Imagen protegida por derechos de autor

Radiocomunicaciones por microondas y ganancia del sistema

INTRODUCCIÓN

Se suelen describir las microondas como ondas electromagnéticas cuyas frecuencias van desde unos 500 MHz hasta 300 GHz o más. Por consiguiente, las señales de microondas, a causa de sus altas frecuencias inherentes, tienen longitudes de onda relativamente cortas, de ahí el nombre "micro" ondas. Por ejemplo, la longitud de onda de una señal de microondas de 100 GHz es de 0.3 cm, mientras que una señal de 100 MHz, como las de la banda comercial de FM, tiene una longitud de 3 m. Las longitudes de las frecuencias de microondas van de 1 a 60 cm, un poco mayores que la energía infrarroja. La tabla 17-1 es una lista de algunas de las bandas de radiofrecuencia en microondas disponibles en Estados Unidos. Para la operación dúplex (en dos sentidos) que se requiere en general en los sistemas de comunicaciones por microondas, cada banda de frecuencias se divide a la mitad, y la mitad inferior se llama *banda baja* y la superior es la *banda alta*. En cualquier estación de radio dada, los transmisores funcionan normalmente en la banda baja o en la alta, mientras que los receptores funcionan en la otra.

La gran mayoría de los sistemas de comunicaciones establecidos desde mediados de la década de 1980 es de naturaleza digital y por consiguiente transporta información en forma digital. Sin embargo, los sistemas terrestres (basados en la Tierra) de *radio repetidoras de microondas* que usan portadoras moduladas en frecuencia (FM) o moduladas digitalmente (PSK o QAM) siguen constituyendo 35% del total de los circuitos de transporte de información en Estados Unidos. Hay muchos tipos distintos de sistemas de microondas funcionando a distancias que varían de 15 a 4000 millas. Los sistemas de microondas de servicio *intraestatal* o *alimentador* se consideran en general de *corto alcance*, porque se usan para llevar información a distancias relativamente cortas, por ejemplo, entre ciudades de un mismo estado. Los sistemas de microondas de *largo alcance* son los que se usan para llevar información a distancias relativamente largas, por ejemplo, en aplicaciones de ruta *interestatal* y de *red primaria*. Las capacidades de los sistemas de radio de microondas van desde menos de 12 canales de banda de voz hasta más de 22,000. Los primeros sistemas tenían circuitos de banda de voz multiplexados por división de frecuencia, y usaban técnicas convencionales, de modulación en frecuencia no coherentes. Los más modernos tienen circuitos de banda de voz modulados por codificación de pulsos y multiplexados por división de tiempo; usan técnicas de modulación digital más

TABLA 17-1 Asignaciones de radiofrecuencias de microondas

Servicio	Frecuencia (MHz)	Banda
Militar	1710–1850	L
Operación fija	1850–1990	L
Enlace de transmisor de estudio	1990–2110	L
Portadora común	2110–2130	S
Operación fija	2130–2150	S
Portadora operacional	2160–2180	S
Operación fija	2180–2200	S
Operación fija de televisión	2500–2690	S
Portadora común y enlace satelital descendente	3700–4200	S
Militar	4400–4990	C
Militar	5250–5350	C
Portadora común y enlace ascendente	5925–6425	C
Operación fija	6575–6875	C
Enlace de transmisor de estudio	6875–7125	C
Portadora común y enlace satelital descendente	7250–7750	C
Portadora común y enlace satelital ascendente	7900–8400	X
Portadora común	10700–11700	X
Operación fija	12200–12700	X
Enlace de estudio de televisión por cable (CATV)	12700–12950	Ku
Enlace de transmisor de estudio	12950–13200	Ku
Militar	14400–15250	Ka
Portadora común	17700–19300	Ka
Enlace satelital ascendente	26000–32000	K
Enlace satelital descendente	39000–42000	Q
Enlace satelital cruzado	50000–51000	V
Enlace satelital cruzado	54000–62000	V

modernas, como la modulación por conmutación de fase (PSK) o por amplitud en cuadratura (QAM).

VENTAJAS DE LAS RADIOCOMUNICACIONES POR MICROONDAS

Los radios de microondas propagan señales a través de la atmósfera terrestre, entre transmisores y receptores que con frecuencia están en la punta de torres a distancias de 15 a 30 millas. Así, los sistemas de radio de microondas tienen la ventaja obvia de contar con capacidad de llevar miles de canales individuales de información entre dos puntos, sin necesidad de instalaciones físicas, como cables coaxiales o fibras ópticas. Así, claro está, se evita la necesidad de adquirir derechos de vía a través de propiedades privadas. Además, las ondas de radio se adaptan mejor para salvar grandes extensiones de agua, montañas altas o terrenos muy boscosos que constituyen formidables barreras para los sistemas de cable. Entre las ventajas del radio de microondas están las siguientes:

1. Los sistemas de radio no necesitan adquisiciones de derecho de vía entre estaciones.
2. Cada estación requiere la compra o alquiler de sólo una pequeña extensión de terreno.
3. Por sus grandes frecuencias de operación, los sistemas de radio de microondas pueden llevar grandes cantidades de información.
4. Las frecuencias altas equivalen a longitudes cortas de onda, que requieren antenas relativamente pequeñas.
5. Las señales de radio se propagan con más facilidad en torno a obstáculos físicos, por ejemplo, a través del agua o las montañas altas.
6. Para la amplificación se requieren menos repetidoras.
7. Las distancias entre los centros de conmutación son menores.

Los generadores de microondas consisten en un oscilador de cristal seguido por una serie de multiplicadores de frecuencia. Por ejemplo, un oscilador de cristal de 125 MHz seguido por una serie de multiplicadores, con factor combinado de multiplicación igual a 48, se podría usar para una frecuencia de portadora de microondas de 6 GHz. La red combinadora de canales proporciona un medio de conectar más de un transmisor de microondas a una sola línea de transmisión que alimente a la antena.

Radiorreceptor de microondas de FM

En el radiorreceptor de microondas de FM que se ve en la fig. 17-1b, la red separadora de canales proporciona el aislamiento y el filtrado necesarios para separar canales de microondas individuales, y dirigirlos hacia sus respectivos receptores. El filtro pasabandas, el mezclador AM y el oscilador de microondas bajan las frecuencias desde las RF de microondas hasta las FI, y las pasan al desmodulador FM. Este desmodulador es un detector convencional, *no coherente* de FM (es decir, un discriminador o un desmodulador PLL). A la salida del detector de FM, una red de deénfasis restaura la señal de banda base a sus características originales de amplitud en función de la frecuencia.

RADIO REPETIDORAS DE MICROONDAS DE FM

La distancia admisible entre un transmisor de microondas de FM y su receptor asociado depende de muchas variables del sistema, por ejemplo, de la potencia de salida del transmisor, umbral de ruido del receptor, terreno, condiciones atmosféricas, capacidad del sistema, objetivos de confiabilidad y expectativas de eficiencia. Normalmente, esa distancia es de 15 a 40 millas (23 a 64 km). Los sistemas de microondas de larga distancia salvan distancias mucho mayores. En consecuencia, un sistema de microondas de un solo salto, como el de la fig. 17-1, es inadecuado en la mayoría de las aplicaciones prácticas. En sistemas con más de 40 millas, o cuando hay obstrucciones geográficas, como una montaña, en la trayectoria de transmisión, se necesitan *repetidoras*. Una repetidora de microondas es un receptor y un transmisor instalados espalda con espalda, o en *tándem* con el sistema. En la fig. 17-2 se muestra un diagrama simplificado de bloques de una repetidora de microondas. La estación repetidora recibe una señal, la amplifica y la reconforma, y a continuación la retransmite hacia la siguiente repetidora o estación receptora.

El lugar de las repetidoras intermedias depende mucho de la naturaleza del terreno entre ellas y el que les rodea. En la planeación preliminar de la ruta se suele suponer que las áreas son relativamente planas, y que las longitudes de los tramos (saltos) tendrán un promedio de 25 a 35 millas, entre estaciones. En un terreno relativamente plano, si se aumenta la longitud del tramo será necesario aumentar las alturas de las torres de antena. De igual modo entran en el proceso de selección la potencia de salida del transmisor y de la antena. La distancia exacta se determina principalmente por la falta de obstáculos en la línea de vista y por la intensidad de la señal recibida. Cuando las frecuencias son mayores de 10 GHz, la pluviosidad local también podría tener una gran influencia sobre la longitud de trayectoria. Sin embargo, en todos los casos se deben usar trayectorias tan horizontales como sea posible. Además, se debe tener en cuenta la posibilidad de interferencia, sea interna o externa.

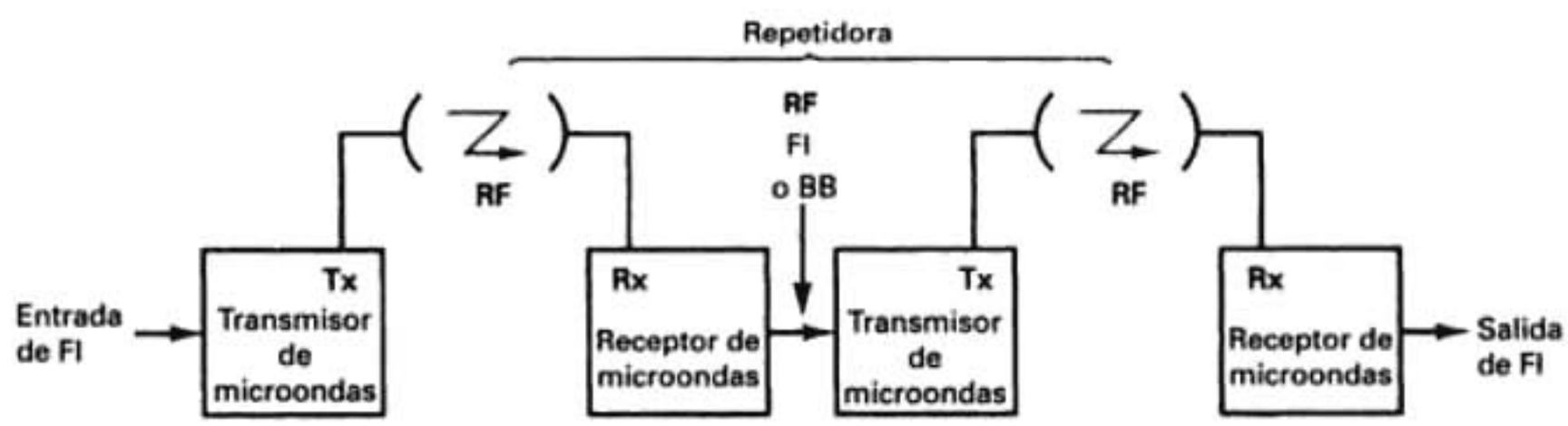
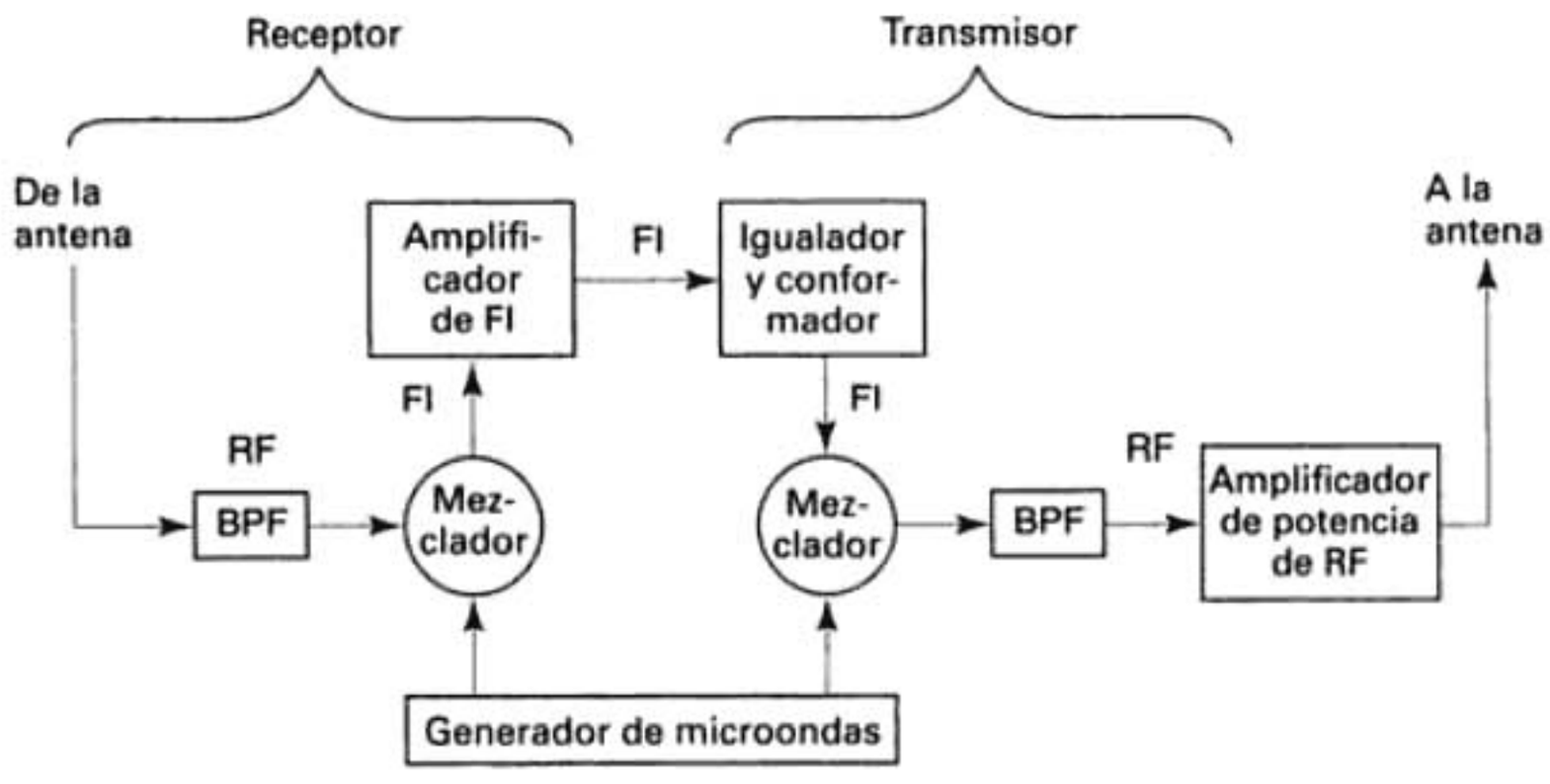
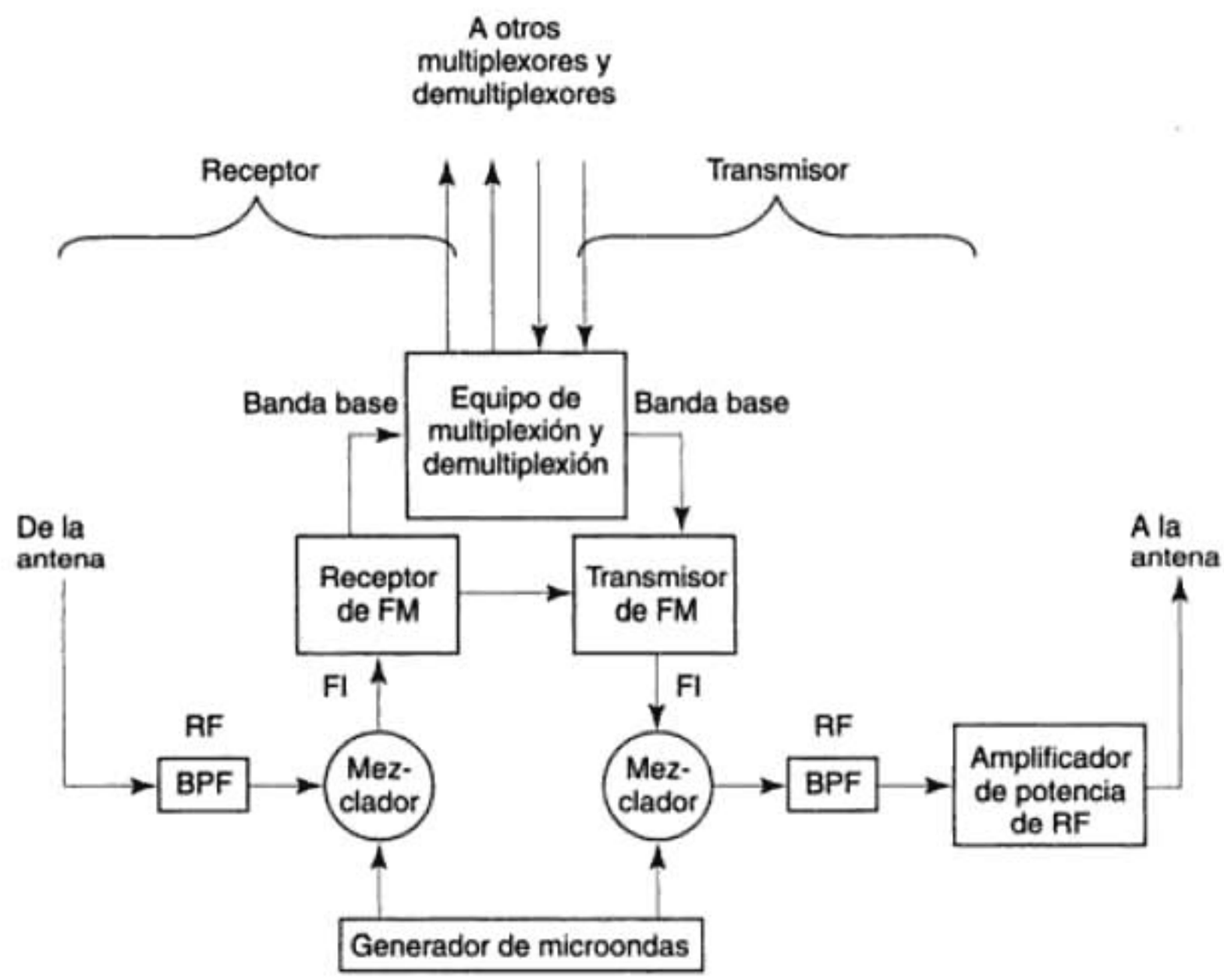


FIGURA 17-2 Repetidora de microondas



(a)



(b)

FIGURA 17-3 Repetidoras de microondas: (a) FI; (b) banda base; (*continúa*)

Hay tres tipos básicos de repetidoras de microondas: FI, banda base y RF (véase la fig. 17-3). Las repetidoras de FI también se llaman *heterodinas*. En una repetidora de FI (fig. 17-3a), la portadora de RF recibida se reduce hasta una FI, se amplifica, reconforma y se sube a una RF, para retransmitirse a continuación. Nunca se demodula la señal a menos de la FI. En consecuencia, la repetidora no modifica información de banda base. En una repetidora de banda base (fig. 17-3b), la portadora recibida de RF se baja hasta una FI, se amplifica, filtra y a continuación se demodula más hasta la banda base. La señal de banda base, que suele ser de canales de banda de voz multiplexados por división de frecuencia, se sigue demodulando hasta los niveles

La fig. 17-3d muestra una repetidora de RF a RF. En estas repetidoras, la señal recibida de microondas no se baja a la FI o a la banda base; tan sólo se mezcla (heterodina) con una frecuencia de oscilador local en un mezclador no lineal. La salida del mezclador se sintoniza a la suma o diferencia de la RF que llega y la frecuencia del oscilador local, dependiendo de si se desea conversión elevadora o reductora de frecuencia. A veces, al oscilador local se le llama oscilador de desplazamiento, y su frecuencia es bastante menor que las de radio que se reciben o se transmiten. Por ejemplo, una RF de llegada de 6.2 GHz se mezcla con una frecuencia de oscilador local de 0.2 GHz, y se producen las frecuencias de suma y diferencia de 6.4 GHz y 6.0 GHz. Para una conversión elevadora de frecuencia, la salida del mezclador se sintonizaría a 6.4 GHz, y para una conversión reductora de frecuencia, a 6.0 GHz. En las repetidoras de RF a RF, la señal de radio sólo se convierte en frecuencia, para a continuación amplificarse y retransmitirse hacia la siguiente repetidora o estación receptora. En las repetidoras RF a RF no son posibles la reconfiguración ni la reforma.

DIVERSIDAD

Los sistemas de microondas usan transmisión en *línea de vista*, y en consecuencia debe existir una trayectoria directa de señal entre las antenas de transmisión y de recepción. Así, si esa trayectoria de señal sufre un gran deterioro, se interrumpirá el servicio. A través del tiempo, las pérdidas en la trayectoria de radio varían con las condiciones atmosféricas, que pueden variar mucho y causar una reducción correspondiente en la intensidad de señal recibida, de 20, 30 o 40 dB o más. Esta reducción de intensidad de señal es temporal, y se llama *desvanecimiento de radio*. El desvanecimiento puede durar desde unos milisegundos (corto plazo) o varias horas o hasta días (largo plazo). Los circuitos de control automático de ganancia, incorporados en los radiorreceptores, pueden compensar desvanecimientos de 25 a 40 dB, dependiendo del diseño del sistema; sin embargo, los desvanecimientos mayores de 40 dB pueden causar una pérdida total de la señal recibida. Cuando esto sucede se pierde la continuidad del servicio.

Diversidad sugiere que hay más de una ruta de transmisión, o método de transmisión disponibles entre un transmisor y un receptor. En un sistema de microondas, el objetivo de usar diversidad es aumentar la confiabilidad del sistema, aumentando su disponibilidad. La tabla 17-2 muestra un medio relativamente sencillo para interpretar determinado porcentaje de confiabilidad del sistema a términos que se puedan relacionar fácilmente con la experiencia. Por ejemplo, un porcentaje de confiabilidad de 99.99% corresponde a unos 53 minutos de tiempo fuera de servicio al año, mientras que si el porcentaje de confiabilidad es 99.9999%, equivale sólo a 32 segundos de tiempo interrumpido por año.

Cuando hay más de una trayectoria o método de transmisión disponible, el sistema puede seleccionar la ruta o método que produzca la máxima calidad en la señal recibida. En general, la máxima calidad se determina evaluando la relación de portadora a ruido (C/N, de *carrier-to-noise*) en la entrada del receptor, o tan sólo midiendo la potencia de la portadora recibida. Aunque

TABLA 17-2 Confiabilidad y tiempo de interrupción

Confiabilidad (%)	Tiempo de interrupción (%)	Tiempo de interrupción Año (horas)	por mes (horas)	Día (horas)
0	100	8760	720	24
50	50	4380	360	12
80	20	1752	144	4.8
90	10	876	72	2.4
95	5	438	36	1.2
98	2	175	14	29 minutos
99	1	88	7	14.4 minutos
99.9	0.1	8.8	43 minutos	1.44 minutos
99.99	0.01	53 minutos	4.3 minutos	8.6 segundos
99.999	0.001	5.3 minutos	26 segundos	0.86 segundos
99.9999	0.0001	32 segundos	2.6 segundos	0.086 segundos

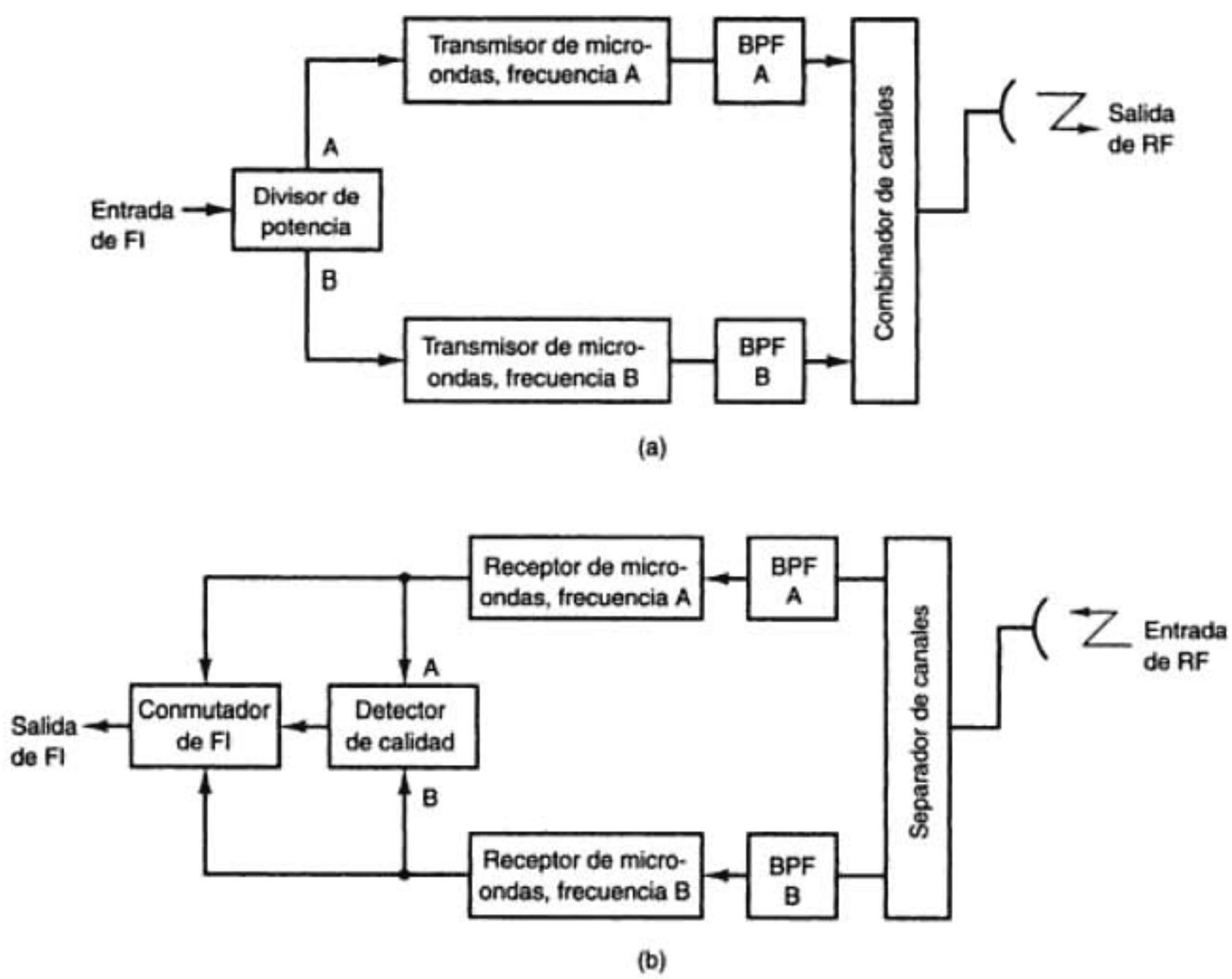


FIGURA 17-4 Sistema de microondas con diversidad de frecuencia: (a) transmisor; (b) receptor

hay muchas formas de obtener la diversidad, los métodos utilizados más comunes son de frecuencia, espacial, polarización, híbrido o cuádruple.

Diversidad de frecuencia

La diversidad de frecuencia sólo consiste en modular dos RF distintas de portadora con la misma información de FI, y transmitir entonces ambas señales de RF a un destino dado. En el destino, se demodulan ambas portadoras y la que produzca la señal de FI de mejor calidad es la que se selecciona. La fig. 17-4 muestra un sistema de microondas de un solo canal con diversidad de frecuencia.

En la fig. 17-4a, la señal de entrada de FI se alimenta a un divisor de potencia, que la dirige a los transmisores A y B de microondas. Las salidas de RF de los dos transmisores se combinan en la red combinadora de canales y se alimentan a la antena de transmisión. En el lado receptor (fig. 17-4b), el separador de canales dirige las portadoras A y B de RF a sus receptores respectivos de microondas, donde se bajan a FI. El circuito detector de calidad determina cuál canal, el A o el B, es el de mejor calidad, y dirige ese canal por el conmutador de FI para seguirlo demodulando hasta la banda base. Muchas de las condiciones atmosféricas adversas que degradan una señal de RF son de frecuencia selectiva; pueden degradar más una frecuencia que otra. En consecuencia, durante determinado tiempo, el conmutador de FI puede alternar muchas veces del receptor A al B y viceversa.

Los arreglos de diversidad de frecuencia constituyen una redundancia de equipo, completa y sencilla, y tienen la ventaja adicional de proporcionar dos trayectorias eléctricas del transmisor al receptor. Su desventaja obvia es que duplica el espectro de frecuencia y el equipo necesarios.

Diversidad espacial

En la diversidad espacial, la salida de un transmisor se alimenta a dos o más antenas, físicamente separadas por una cantidad apreciable de longitudes de onda. De igual manera, en el receptor, puede haber más de una antena que proporcione la señal de entrada al receptor. Si se usan

Diversidad de polarización

En la *diversidad de polarización*, una sola portadora de RF se propaga con dos polarizaciones electromagnéticas diferentes, vertical y horizontal. Las ondas electromagnéticas de distintas polarizaciones no necesariamente están sometidas a las mismas degradaciones de transmisión. La diversidad de polarización se usa en general junto con la diversidad espacial. Un par de antenas de transmisión y recepción se polariza en sentido vertical, y el otro en sentido horizontal. También es posible usar en forma simultánea la diversidad de frecuencia, espacial y de polarización.

Diversidad híbrida

La *diversidad híbrida* es una forma algo especializada de diversidad, que consiste en una trayectoria normal de diversidad de frecuencia, en la que los dos pares de transmisor y receptor en un extremo de la trayectoria están separados entre sí y conectados a distintas antenas, separadas verticalmente como en la diversidad espacial. El arreglo proporciona un efecto de diversidad espacial en ambas direcciones; en una porque los receptores están separados verticalmente, y en la otra porque están separados horizontalmente. Este arreglo combina las ventajas operativas de la diversidad de frecuencia con la mejor protección de la diversidad espacial. Sin embargo, la diversidad híbrida tiene la desventaja de requerir dos radiofrecuencias para obtener un canal en funcionamiento.

Diversidad cuádruple

La *diversidad cuádruple* es otra forma de diversidad híbrida, e indudablemente proporciona la transmisión más confiable; sin embargo, también es la más costosa. El concepto básico de diversidad cuádruple es bastante sencillo: es una combinación de diversidad de frecuencia, espacial, de polarización y de recepción en un solo sistema. Su desventaja obvia es que necesita equipo electrónico, frecuencias, antenas y guías de onda redundantes, que son cargas económicas.

ARREGLOS DE CONMUTACIÓN DE PROTECCIÓN

Para evitar una interrupción de servicio durante periodos de gran desvanecimiento o de fallas de equipo, se ponen a disposición, en forma temporal, otras instalaciones en un arreglo de *conmutación de protección*. Los conceptos generales de conmutación de protección y de diversidad son bastante parecidos. Ambos proporcionan protección contra fallas de equipo y desvanecimientos atmosféricos. La diferencia principal entre ellos es, simplemente, que los sistemas de diversidad sólo proporcionan una ruta de transmisión alternativa para un solo enlace de microondas, es decir, entre un transmisor y un receptor, dentro del sistema general de comunicaciones. Por otro lado, los arreglos de conmutación de protección proporcionan protección para una parte mucho mayor del sistema de comunicaciones, que en general comprende varias repetidoras que abarcan distancias de 100 millas (160 km) o más. También, los sistemas de diversidad proporcionan en general 100% de protección a un solo canal de radio, mientras que los arreglos de conmutación de protección se suelen compartir entre varios canales de radio.

En esencia, hay dos tipos de arreglos de conmutación de protección: *reserva continua* y *por diversidad*. En la protección por reserva continua, cada canal funcional de radio tiene un canal de respaldo dedicado, o de reserva. En la protección por diversidad, un solo canal de respaldo está a la disposición hasta para 11 canales funcionales. Los sistemas de reserva continua ofrecen protección de 100% para cada canal funcional de radio. Un sistema por diversidad ofrece una protección de 100% sólo al primer canal funcional que falle. Si dos canales de radio fallan al mismo tiempo, se produce una interrupción del servicio.

Reserva continua

La fig. 17-6a muestra un arreglo de conmutación de protección por reserva continua de un solo canal. En el transmisor, la FI entra en un *puente de inicio*, que reparte la potencia de la señal y la dirige a los canales de microondas activo y de reserva, en forma simultánea. En consecuencia, tanto el canal funcional como el de reserva portan la misma información de banda base. En el receptor, el conmutador de FI pasa la señal de FI del canal funcional al equipo de terminal de FM. El conmutador de FI vigila en forma continua la señal recibida por el canal funcional y, si

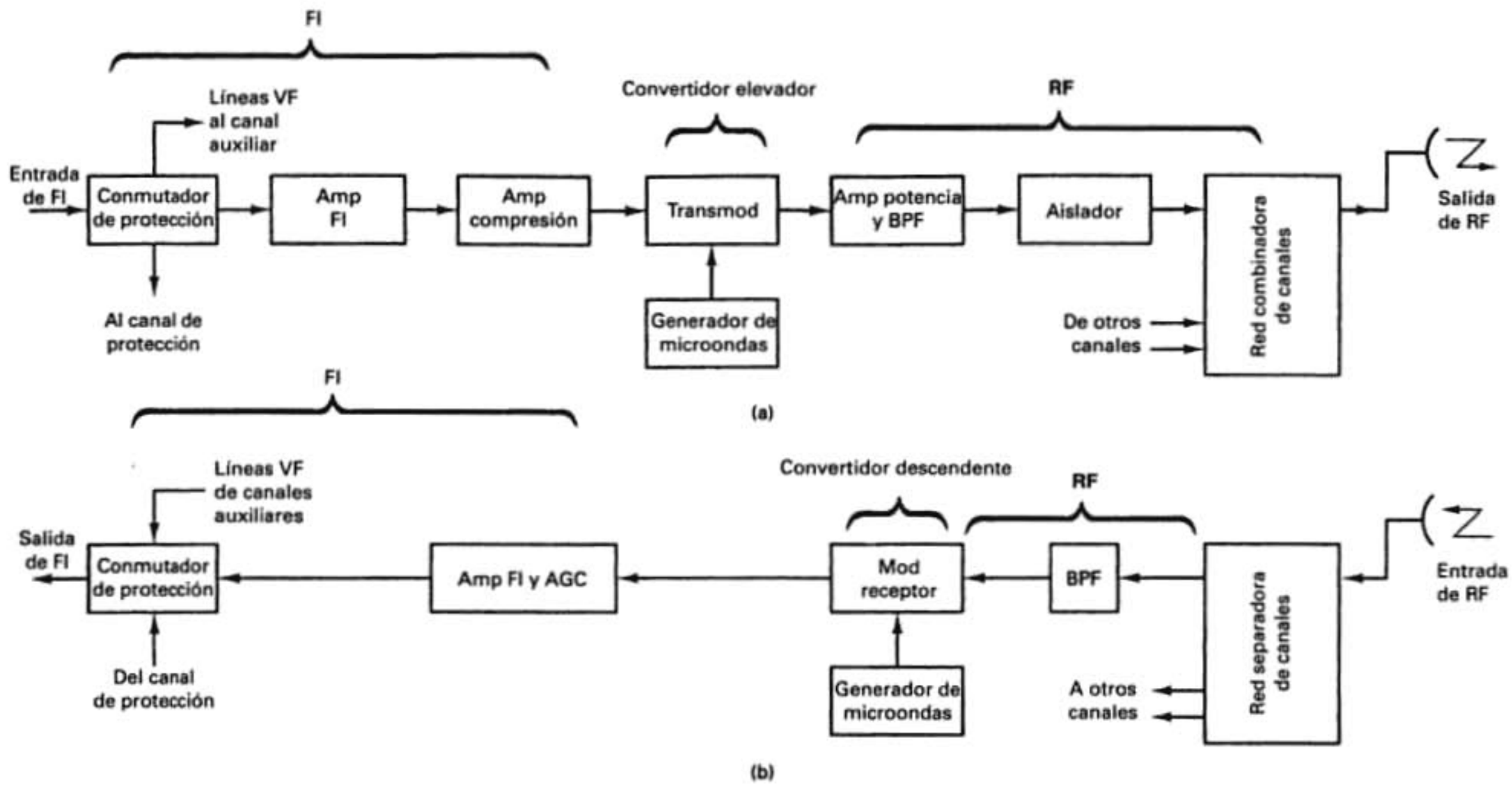


FIGURA 17-8 Estación terminal de microondas: (a) transmisor; (b) receptor

los amplificadores de salida del transmisor. Las ganancias normales en las antenas de microondas van de 10 a 40 dB, y las potencias normales de salida del transmisor son de 0.5 a 10 W.

Un *generador de microondas* proporciona la entrada de portadora de RF al convertidor de subida de frecuencia. Se llama generador de microondas, y no oscilador, porque es difícil construir un circuito estable que oscile en la zona de los gigahertz. En lugar de ello se usa un oscilador controlado por cristal, que funciona en el intervalo de 5 a 25 MHz para proporcionar una frecuencia básica que se multiplique hasta la frecuencia RF deseada en la portadora.

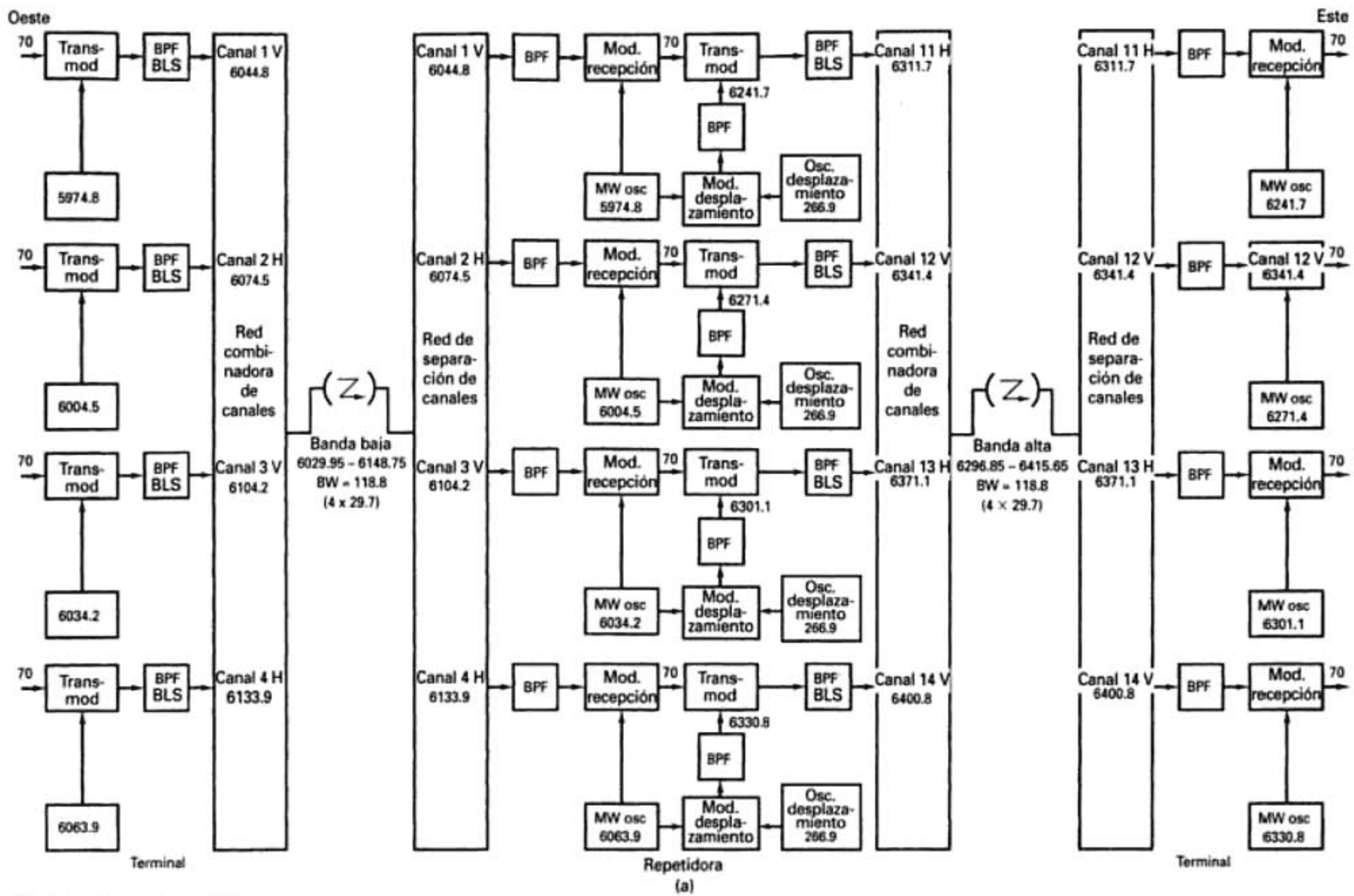
Un *aislador* es un dispositivo unidireccional, fabricado a menudo con ferrita. El aislador se usa junto con una red combinadora de canales para evitar que la salida de un transmisor interfiera con la de otro transmisor.

El receptor de RF (fig. 17-8b) es, en esencia, igual que el transmisor, pero funciona en dirección contraria. Sin embargo, una diferencia es la presencia de un amplificador de FI en el receptor. Este amplificador de FI tiene un circuito de *control automático de ganancia* (AGC). También, con mucha frecuencia, no hay amplificadores de RF en el receptor. En el caso normal, se usa un desmodulador balanceado de bajo ruido y sensible, como desmodulador de recepción. Así se elimina la necesidad de un amplificador de RF y se mejora la relación general de señal a ruido. Cuando se requieren amplificadores de RF, se usan de alta calidad y de *bajo ruido* (LNA, de *low-noise amplifier*). Como ejemplos de LNA que se usan con frecuencia están los diodos de efecto túnel y los amplificadores paramétricos.

Estación repetidora

La fig. 17-9 muestra el diagrama de bloques de una repetidora de FI de microondas. La señal de RF recibida entra al receptor a través de la red de separación de canales y el filtro pasabandas. El desmodulador de recepción convierte la portadora de RF, bajándola a FI. Los circuitos AMP/AGC y el igualador de FI amplifican y reconforman la señal de FI. El igualador compensa las *no linealidades de ganancia en función de frecuencia* y la *distorsión por retardo de envolvente* que se introducen en el sistema. De nuevo, el modulador de transmisión convierte la IF en RF para su retransmisión. Sin embargo, en una estación repetidora es un poco distinto el método que se usa para generar las frecuencias de portadora de microondas RF, respecto al que se usa en una estación terminal. En la repetidora de FI sólo se requiere un generador de microondas para suministrar la señal de portadora de RF, tanto al modulador de transmisión como al de recepción. El generador de microondas, el oscilador de desplazamiento y el modulador de desplazamiento permiten que la repetidora reciba una frecuencia de portadora RF, le baje la frecuencia a una FI, para después convertir la FI a una portadora RF de distinta frecuencia. Es posible que la estación C reciba las portadoras tanto de la estación A como la B en forma simultánea (a esto se le llama *interferencia por salto múltiple*, y se muestra en la fig. 17-10a). Esto puede suceder sólo cuando tres estaciones están instaladas en una recta geográfica en el sistema. Para evitar que ocurra, el ancho de banda asignado al sistema se divide a la mitad, y se crea una banda de baja frecuencia y una de alta frecuencia. Cada estación, una a la vez, pasa de una portadora de transmisión de banda baja a una de banda alta (fig. 17-10b). Si se recibe en la estación C una transmisión de la estación A, será rechazada en la red separadora de canales y no causa interferencia. A este arreglo se le llama sistema de repetidoras de microondas alta/baja. Las reglas son sencillas: si una estación repetidora recibe una portadora de RF de la banda baja, retransmite una portadora de RF en la banda alta, y viceversa. La única vez que se pueden recibir varias portadoras de la misma frecuencia es cuando se recibe una transmisión de una estación, en otra que esté a tres saltos de distancia. Casi no es probable que suceda eso.

Otra razón para usar un esquema de frecuencia alta/baja es evitar la potencia que "se sale" por las partes trasera y laterales de una antena de transmisión, que puede interferir con la señal que entre a una antena receptora cercana. A esto se le llama *radiación de pérdida*. Todas las antenas, sin importar lo alto de su ganancia o su gran direccionalidad de radiación, irradian un pequeño porcentaje de su potencia por sus lados trasero y laterales, determinando una *eficiencia direccional* de la antena. Aunque la eficiencia direccional de una antena típica de microondas es bastante alta, la cantidad relativamente baja de la potencia que se irradia por su parte trasera puede ser bastante apreciable, en comparación con la potencia de la portadora normal recibida en el sistema. Si son distintas las frecuencias de transmisión y recepción, los filtros de la red separadora del receptor evitan que perjudique la radiación de pérdida.



Todas las fecuencias en MHz.

FIGURA 17-11 Plan de frecuencia alta/baja con ocho canales: (a) oeste a este; (*continúa*)

1 de banda baja corresponde al canal 11 de banda alta, el canal 2 al canal 12, etcétera. La dirección este a oeste (fig. 17-11b) propaga las portadoras de banda alta y baja en el orden contrario al del sistema oeste a este. También están invertidas las polarizaciones. Si algo de la potencia del canal 1 de la terminal oeste se propagara en forma directa hacia el receptor de la terminal este, tendría distinta frecuencia y polarización que las transmisiones del canal 11. En consecuencia, no interferiría con la recepción del canal 11; no habría interferencia por salto múltiple. También, nótese que ninguno de los canales de transmisión o de recepción en la repetidora tiene igual frecuencia y polarización al mismo tiempo. En consecuencia, la interferencia de los transmisores hacia los receptores, debida a la radiación de pérdida, es insignificante.

CARACTERÍSTICAS DE LA TRAYECTORIA

Las *trayectorias de propagación* normales entre dos antenas de radio en un sistema de microondas se ven en la fig. 17-12. La *trayectoria en espacio libre* es la *trayectoria de línea de vista*, directamente entre las antenas transmisora y receptora (también se llama *onda directa*). La *onda reflejada en tierra* es la porción de la señal transmitida que se refleja en la superficie terrestre y la captura la antena receptora. La *onda superficial* consiste en los campos eléctrico y magnético asociados con las corrientes inducidas en la superficie terrestre. La magnitud de la onda superficial depende de las características de la superficie terrestre y de la polarización electromagnética de la onda. La suma de esas tres trayectorias, tomando en cuenta su amplitud y su fase, se llama *onda terrestre*. La *onda reflejada* u *onda celeste* es la parte de la señal transmitida que regresa (se refleja) hacia la superficie terrestre, debido a las capas ionizadas de la atmósfera terrestre.

Todas las trayectorias de la fig. 17-12 existen en cualquier sistema de radio de microondas, pero algunas son despreciables en ciertos intervalos de frecuencia. En frecuencias menores que 1.5 MHz, la onda superficial proporciona la cobertura primaria, y la onda celeste contribuye a aumentar esta cobertura por las noches, cuando es mínima la absorción de la ionosfera. En frecuencias mayores que unos 30 a 50 MHz, por lo general las trayectorias de importancia son las de espacio libre y de reflejo en tierra. A esas frecuencias también se puede despreciar la onda superficial, siempre y cuando las alturas de las antenas no sean muy pequeñas. La onda reflejada es sólo una fuente de interferencia ocasional de larga distancia, y no es una señal confiable para fines de comunicaciones por microondas. En este capítulo se desprecian las propagaciones superficial y reflejada, y se centra la atención en los fenómenos que afectan a las ondas directas y reflejadas.

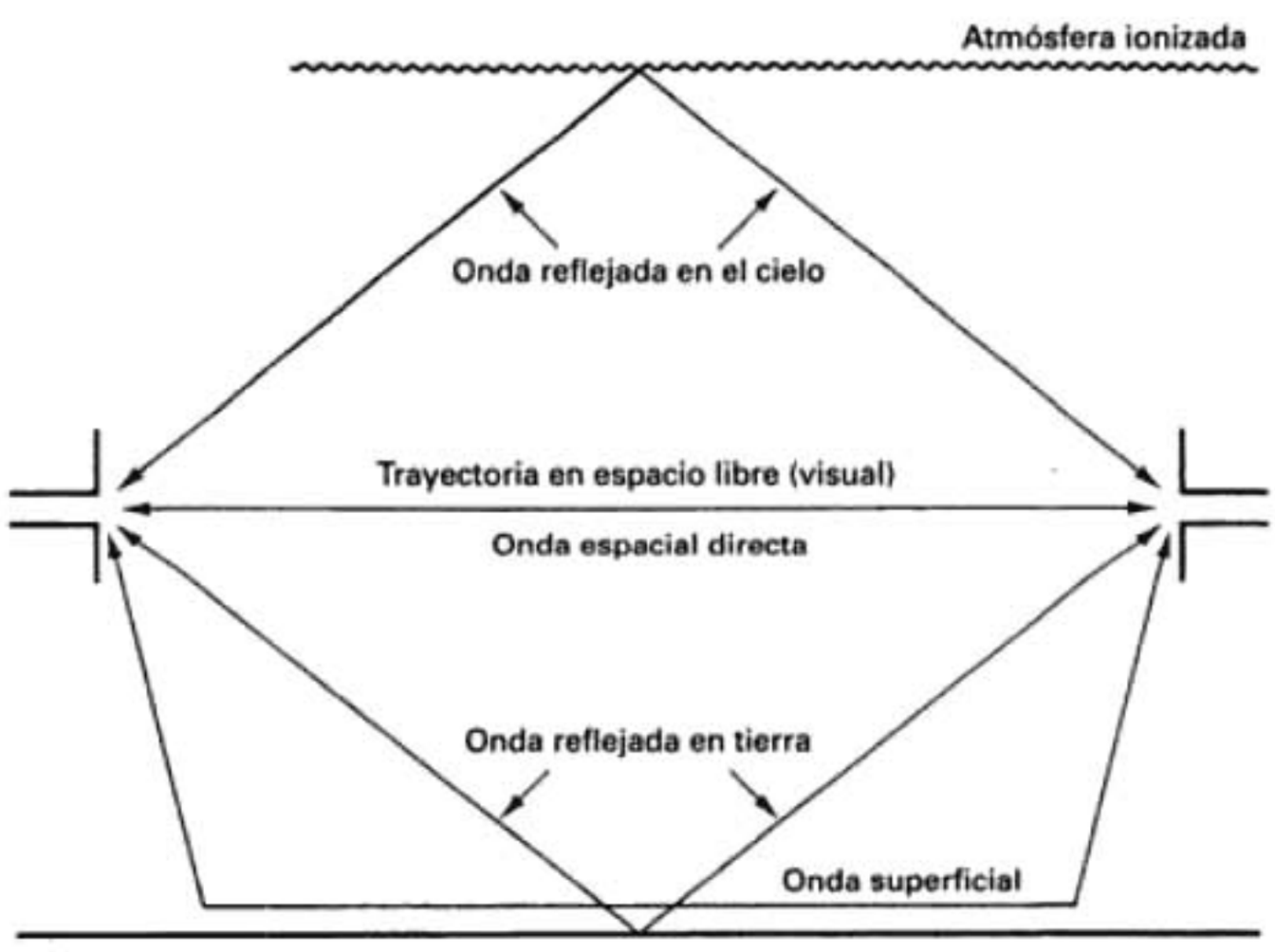


FIGURA 17-12 Trayectorias de propagación

Desvanecimiento

Desvanecimiento es un término general que se aplica a la reducción de intensidad de señal en la entrada a un receptor. Este término se aplica a variables de propagación en la trayectoria física de la radiación, que afectan a los cambios en la pérdida de trayectoria entre el transmisor en una estación, y su receptor normal en otra estación. Los cambios de trayectoria se asocian tanto a condiciones atmosféricas, como a la geometría de la trayectoria misma, es decir, a la posición relativa de la antena con respecto al suelo y a otro terreno que la rodee. La refracción atmosférica subnormal puede transformar una trayectoria de línea de vista clara en una muy obstruida. El desvanecimiento se puede presentar bajo condiciones de densas nieblas rastreras, o cuando el aire extremadamente frío se mueve sobre un terreno cálido. El resultado en cada caso es un aumento sustancial en la pérdida de trayectoria dentro de una amplia banda de frecuencias. La magnitud y la rapidez de este tipo de desvanecimiento lento y plano se puede reducir, en general, usando mayores alturas de antena.

Una forma más común de desvanecimiento es uno relativamente rápido y selectivo en frecuencias, causado por la interferencia entre uno o más rayos en la atmósfera. Las trayectorias separadas entre el transmisor y el receptor se deben a irregularidades en las variaciones de permitividad eléctrica con la altura. Los márgenes de transmisión que se deben proporcionar para contrarrestar ambos tipos de desvanecimiento son consideraciones importantes al determinar los parámetros generales del sistema y la confiabilidad del funcionamiento.

Un tipo de desvanecimiento por interferencia puede presentarse con cualquier intensidad, pero por fortuna mientras más intenso es el desvanecimiento, es menos probable que suceda, y cuando sucede, su duración será más corta. Tanto la cantidad de desvanecimientos como el porcentaje del tiempo por debajo de determinado valor, tienden a aumentar a medida que aumenta la distancia entre repetidoras o la frecuencia de operación. Las trayectorias múltiples suelen ser aéreas, aunque a veces pueden deberse en parte a reflexiones en tierra. Se pueden minimizar los efectos de desvanecimiento por múltiples trayectorias recurriendo a la diversidad espacial o de frecuencia.

GANANCIA DEL SISTEMA

En su forma más sencilla, la *ganancia del sistema* es la diferencia entre la potencia nominal de salida de un transmisor, y la potencia de entrada mínima requerida por un receptor. La ganancia del sistema debe ser mayor o igual a la suma de todas las ganancias y pérdidas incurridas por una señal al propagarse de un transmisor a un receptor. En esencia, representa la pérdida neta de un sistema de radio. La ganancia del sistema se usa para calcular la confiabilidad de un sistema para determinados parámetros del mismo. La ecuación de la ganancia del sistema es

$$G_s = P_t - C_{\text{mín}}$$

donde G_s = ganancia del sistema (dB)
P_t = potencia de salida del transmisor (dBm)
$C_{\text{mín}}$ = potencia mínima de entrada al receptor para determinado objetivo de calidad (dBm)

y donde

$$P_t - C_{\text{mín}} \geq \text{pérdidas} - \text{ganancias}$$

Ganancias: A_t = ganancia de la antena de transmisión (dB) en relación con un radiador isotrópico
A_r = ganancia de la antena de recepción (dB) en relación con un radiador isotrópico

Pérdidas: L_p = pérdida en trayectoria por espacio libre entre las antenas (dB)
L_f = pérdida en alimentador de guía de onda (dB) entre la red de distribución (la red combinadora de canales o separadora de canales) y su antena respectiva (véase la tabla 17-3)
L_b = pérdida total por acoplamiento o por ramificación (dB) en los circuladores, filtros y red de distribución, entre la salida de un transmisor o la entrada a un receptor, y su guía de onda de alimentación respectiva (véase tabla 17-3)
F_m = margen de desvanecimiento para determinado objetivo de confiabilidad

TABLA 17-3 Parámetros de ganancia del sistema

			Pérdida por ramificación, L_b (dB)			
	Pérdida por ramificación, L_f		Diversidad		Ganancia de antena, A_t o A_r	
Frecuencia (GHz)	Tipo	Pérdida (db/100 m)	Frecuencia	Espacial	Tamaño (m)	Ganancia (dB)
1.8	Cable coaxial lleno de aire	5.4	5	2	1.2	25.2
					2.4	31.2
					3.0	33.2
					3.7	34.7
7.4	Guía de onda elíptica EWP 64	4.7	3	2	1.5	38.8
					2.4	43.1
					3.0	44.8
					3.7	46.5
8.0	Guía de onda elíptica EWP 69	6.5	3	2	2.4	43.8
					3.0	45.6
					3.7	47.3
					4.8	49.8

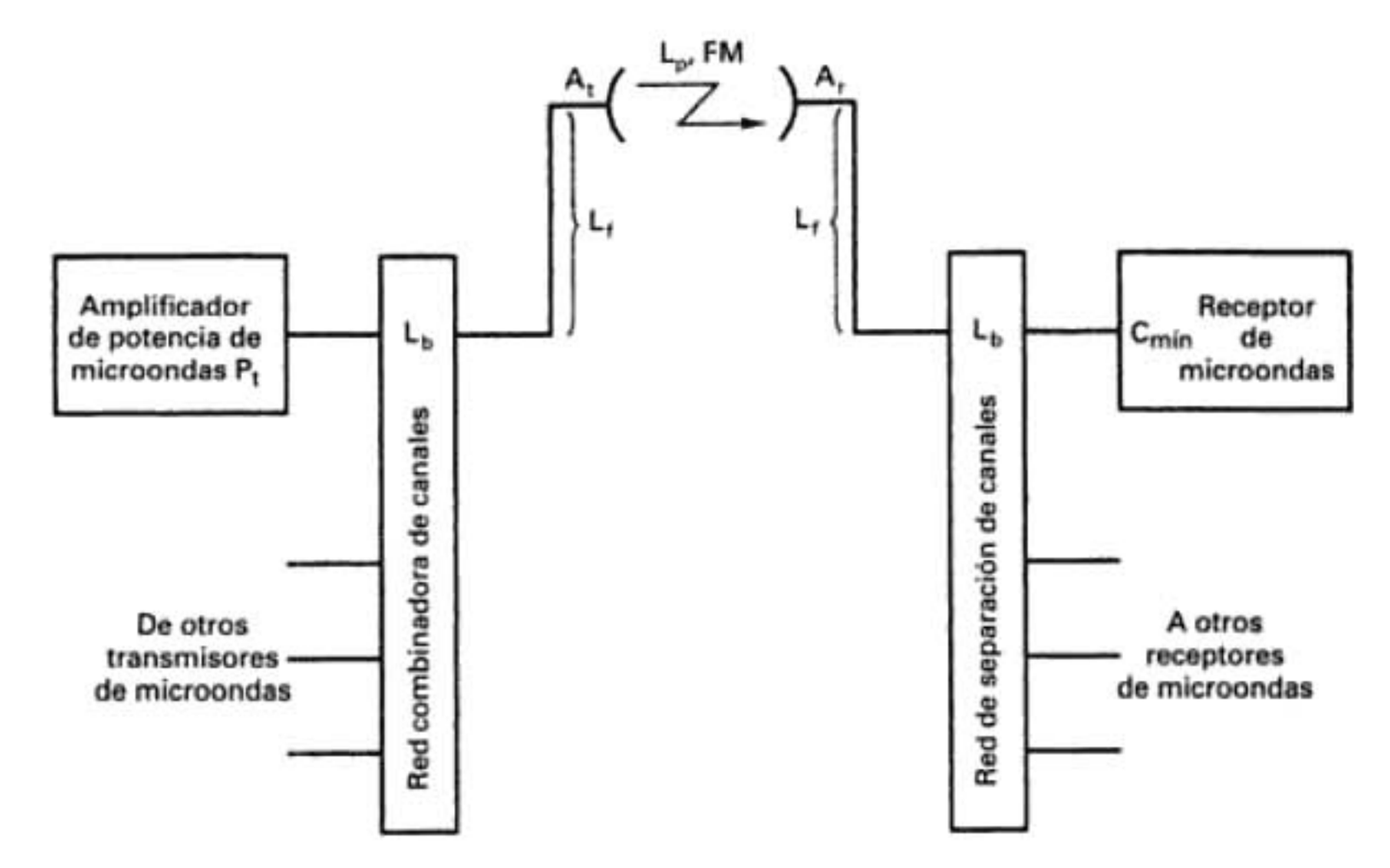


FIGURA 17-13 Ganancias y pérdidas de un sistema

La descripción matemática de la ganancia del sistema es

$$G_s = P_t - C_{\min} \geq F_m + L_p + L_f + L_b - A_t - A_r \tag{17-1}$$

en donde todos los valores se expresan en dB o dBm. Como la ganancia del sistema es indicativa de una pérdida neta, las pérdidas se representan con valores positivos de dB, y las ganancias con valores negativos de dB. La fig. 17-13 muestra un diagrama general de un sistema de microondas, e indica dónde suceden las respectivas pérdidas y ganancias.

Pérdidas en la trayectoria en espacio libre

Se define a las pérdidas en la trayectoria en espacio libre (a veces llamadas *pérdidas por dispersión*) como la pérdida incurrida por una onda electromagnética al propagarse en línea recta a

través del vacío, sin energías de absorción o reflexión debidas a objetos cercanos. Las pérdidas en la trayectoria en espacio libre dependen de la frecuencia, y aumentan con la distancia. La ecuación para determinar estas pérdidas es la siguiente

$$L_p = \left(\frac{4\pi D}{\lambda}\right)^2 = \left(\frac{4\pi f D}{c}\right)^2$$

donde L_p = pérdidas en la trayectoria en espacio libre (adimensional)
D = distancia (metros)
f = frecuencia (hertz)
λ = longitud de onda (metros)
c = velocidad de la luz en el espacio libre (3×10^8 m/s)

Al pasar a dB se obtiene

$$\begin{aligned} L_{p(\mathrm{dB})} &= 10 \log \left(\frac{4\pi f D}{c}\right)^2 \\ &= 20 \log \frac{4\pi f D}{c} \\ &= 20 \log \frac{4\pi}{c} + 20 \log f + 20 \log D \end{aligned} \tag{17-2}$$

Cuando la frecuencia está en MHz y la distancia en km,

$$\begin{aligned} L_{p(\mathrm{dB})} &= 20 \log \frac{4\pi(10)\,(10)^3}{3 \times 10^8} + 20 \log f_{(\mathrm{MHz})} + 20 \log D_{(\mathrm{km})} \\ &= 32.4 + 20 \log f_{(\mathrm{MHz})} + 20 \log D_{(\mathrm{km})} \end{aligned} \tag{17-3}$$

Cuando la frecuencia está en GHz y la distancia en km,

$$L_{p(\mathrm{dB})} = 92.4 + 20 \log f_{(\mathrm{GHz})} + 20 \log D_{(\mathrm{km})} \tag{17-4}$$

Se pueden hacer conversiones parecidas para la distancia en millas, frecuencia en kHz, etc.

Ejemplo 17-2

Calcular las pérdidas en la trayectoria en espacio libre, para una frecuencia de portadora de 6 GH y 50 km de distancia.

Solución

$$\begin{aligned} L_p &= 32.4 + 20 \log 6000 + 20 \log 50 \\ &= 32.4 + 75.6 + 34 = 142 \text{ dB} \\ L_p &= 92.4 + 20 \log 6 + 20 \log 50 \\ &= 92.4 + 15.6 + 34 = 142 \text{ dB} \end{aligned}$$

Margen de desvanecimiento

En esencia el *margen de desvanecimiento* es un "factor ficticio" que se incluye en la ecuación de ganancia del sistema, para tener en cuenta las características no ideales y menos predecibles de la propagación de las ondas de radio, por ejemplo, la *propagación por múltiples trayectorias* (*pérdida por múltiples trayectorias*) y la *sensibilidad del terreno*. Estas características son causa de condiciones atmosféricas temporales y anormales que alteran las pérdidas en la trayectoria en espacio libre y, por lo general, son perjudiciales para la eficiencia general del sistema. El margen de desvanecimiento también tiene en cuenta los objetivos de confiabilidad de un sistema. Por lo anterior, el margen de desvanecimiento se incluye como una pérdida en la ecuación de ganancia del sistema.

W. T. Barnett, de Bell Telephone Laboratories, describió en abril de 1969, formas de calcular el tiempo de interrupción debido al desvanecimiento en una trayectoria sin diversidad, en función del terreno, el clima, la longitud de la trayectoria y el margen de desvanecimiento. Arvids Vignant, también de Bell Laboratories, dedujo en junio de 1970 fórmulas para calcular el mejoramiento efectivo alcanzable mediante diversidad espacial vertical, en función de la distancia de separación, longitud de trayectoria y frecuencia.

Al resolver las ecuaciones de confiabilidad de Barnett-Vignant para determinada disponibilidad anual de un sistema no protegido y sin diversidad, se obtiene la siguiente ecuación

$$F_m = \underbrace{30 \log D}_{\text{efecto de trayectoria múltiple}} + \underbrace{10 \log (6ABf)}_{\text{sensibilidad del terreno}} - \underbrace{10 \log (1 - R)}_{\text{objetivos de confiabilidad}} - \underbrace{70}_{\text{constante}} \qquad (17\text{-}5)$$

siendo F_m = margen de desvanecimiento (dB)
D = distancia (kilómetros)
f = frecuencia (gigahertz)
R = confiabilidad en decimales (es decir, 99.99% = 0.9999 de confiabilidad)
$1 - R$ = objetivo de confiabilidad para una ruta de 400 km en un sentido
A = factor de aspereza
= 4 sobre agua o un terreno muy liso
= 1 sobre terreno promedio
= 0.25 sobre un terreno muy áspero y montañoso
B = factor para convertir la probabilidad del peor de los meses en probabilidad anual
= 1 para convertir una disponibilidad anual a la base del peor de los meses
= 0.5 para áreas cálidas o húmedas
= 0.25 para áreas continentales promedio
= 0.125 para áreas muy secas o montañosas

Ejemplo 17-3

Un sistema de radio de microondas con diversidad espacial funciona con frecuencia de portadora RF igual a 1.8 GHz. Cada estación tiene una antena parabólica de 2.4 m de diámetro, que se alimenta con 100 m de cable coaxial lleno de aire. El terreno es liso, y la zona tiene clima húmedo. La distancia entre estaciones es 40 km. Se desea un objetivo de confiabilidad de 99.99%. Calcular la ganancia del sistema.

Solución Se sustituye en la ecuación 17-5, y se ve que el margen de desvanecimiento es

$$\begin{aligned} F_m &= 30 \log 40 + 10 \log [(6)(4)(0.5)(1.8)] - 10 \log (1 - 0.9999) - 70 \\ &= 48.06 + 13.34 - (-40) - 70 \\ &= 48.06 + 13.34 + 40 - 70 = 31.4 \text{ dB} \end{aligned}$$

Al sustituir en la ecuación 17-4 se obtienen las pérdidas en la trayectoria

$$\begin{aligned} L_p &= 92.4 + 20 \log 1.8 + 20 \log 40 \\ &= 92.4 + 5.11 + 32.04 = 129.55 \text{ dB} \end{aligned}$$

De acuerdo con la tabla 17-3,

$$L_b = 4 \text{ dB } (2 + 2 = 4)$$
$$L_f = 10.8 \text{ dB } (100 \text{ m} + 100 \text{ m} = 200 \text{ m})$$
$$A_t = A_r = 31.2 \text{ dB}$$

Se sustituyen los valores en la ecuación 17-3, para obtener la ganancia del sistema

$$G_s = 31.4 + 129.55 + 10.8 + 4 - 31.2 - 31.2 = 113.35 \text{ dB}$$

Este resultado indica que para que este sistema funcione con confiabilidad de 99.99%, dadas las condiciones del terreno, redes de distribución, líneas de transmisión y antenas, la potencia de salida del transmisor debe ser, como mínimo, 113.35 dB mayor que la potencia de la señal recibida.

Umbral del receptor

La relación de *portadora a ruido* (C/N, de *carrier-to-noise*) es, probablemente, el parámetro más importante que se considera al evaluar el funcionamiento de un sistema de comunicaciones por microondas. La potencia mínima de portadora de banda ancha ($C_{mín}$) a la entrada de un receptor que produzca una salida útil de banda base se llama *umbral* del receptor, o a veces *sensibilidad* del receptor. Este umbral del receptor depende de la potencia del ruido de banda ancha presente en la entrada de un receptor, el ruido introducido dentro del receptor y de la sensibilidad del detector de banda base al ruido. Antes de poder calcular $C_{mín}$, se debe determinar la potencia de ruido en la entrada. Esa potencia es

$$N = KTB$$

$$F = \frac{\text{relación de señal a ruido en la entrada}}{\text{relación de señal a ruido en la salida}} \text{ (como relación adimensional)} \quad (17\text{-}7a)$$

El índice de ruido es sólo el factor de ruido expresado en dB, y es un parámetro de uso frecuente para indicar la calidad de un receptor. La ecuación de definición del índice de ruido es

$$NF = 10 \log \frac{\text{relación de señal a ruido en la entrada}}{\text{relación de señal a ruido en la salida}} \text{ (dB)} \quad (17\text{-}7b)$$

o sea

$$NF = 10 \log F \quad (17\text{-}7c)$$

En esencia, el índice de ruido indica cuánto se deteriora la relación de señal a ruido al propagarse una onda desde la entrada hasta la salida de un circuito. Por ejemplo, un amplificador con índice de ruido de 6 dB quiere decir que la relación de señal a ruido en su salida es 6 dB menor que la que era en su entrada. Si un circuito fuera perfecto y no agregara más ruido a la señal, la relación de señal a ruido a la salida sería igual a la de la entrada. Para un circuito perfecto y sin ruido, el factor de ruido es 1 y el índice de ruido es 0 dB.

Un circuito electrónico amplifica las señales y el ruido por igual, dentro de su banda pasante. Por lo tanto, si el amplificador es ideal y sin ruido, la señal y el ruido en la entrada se amplifican igual, y la relación de señal a ruido en la salida será igual a la de la entrada. Sin embargo, en realidad los amplificadores no son ideales. Por consiguiente, el amplificador agrega ruido generado internamente a la forma de onda, y reduce la relación general de señal a ruido. El ruido más predominante es el ruido térmico, que se genera en todos los componentes eléctricos. Por consiguiente, todas las redes, amplificadores y sistemas agregan ruido a la señal y así reducen la relación general de señal a ruido a medida que la señal pasa por ellos.

Cuando se conectan en cascada dos o más amplificadores, como se ve en la fig. 17-14, el factor total de ruido es la acumulación de los factores individuales. Para calcular el factor de ruido total de varios amplificadores se usa la *fórmula de Friiss*, que es la siguiente

$$F_T = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1 A_2} + \frac{F_n - 1}{A_1 A_2 A_3} \quad (17\text{-}8)$$

donde F_T = factor total de ruido para n amplificadores en cascada
F_1 = factor de ruido, amplificador 1
F_2 = factor de ruido, amplificador 2
F_3 = factor de ruido, amplificador 3
F_n = factor de ruido, amplificador n
A_1 = ganancia de potencia, amplificador 1
A_2 = ganancia de potencia, amplificador 2
A_3 = ganancia de potencia, amplificador 3

Nótese que para usar la fórmula de Friiss, se deben convertir los índices de ruido a factores de ruido. El índice total de ruido es

$$NF_{T(dB)} = 10 \log F_T \quad (17\text{-}9)$$

Se puede ver que el factor de ruido del primer amplificador (F_1) contribuye más al índice general de ruido. El ruido que se introduce en la primera etapa se amplifica en cada uno de los amplificadores siguientes. En consecuencia, cuando se compara con el ruido introducido en la

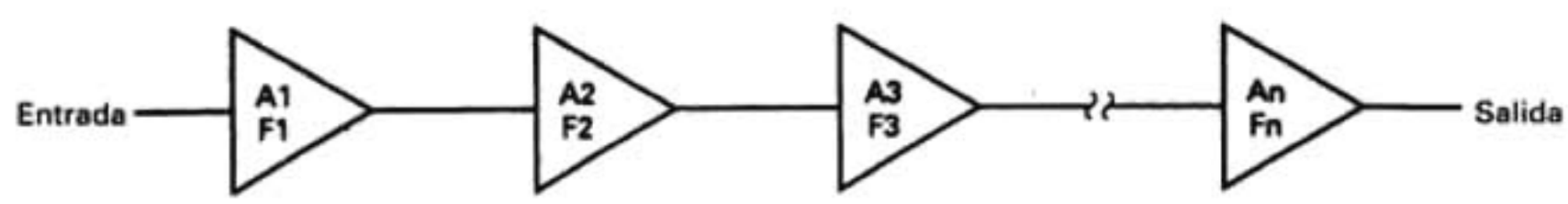


FIGURA 17-14 Índice total de ruido

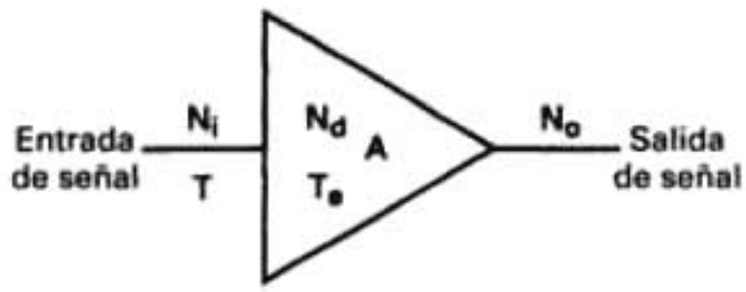


FIGURA 17-15 Índice de ruido en función de la temperatura

primera etapa, el que se agrega con cada amplificador siguiente se reduce en un factor igual al producto de las ganancias de potencia de los amplificadores anteriores.

Cuando son necesarios cálculos precisos (de 0.1 dB o menos) de ruido, en general es más cómodo expresar el índice de ruido en función de la temperatura de ruido, o la temperatura equivalente de ruido, y no en potencia absoluta. Como la potencia del ruido (N) es proporcional a la temperatura, el ruido presente en la entrada a un dispositivo se puede expresar en función de la temperatura ambiente del dispositivo (T) y de su temperatura equivalente de ruido (T_e). El factor de ruido se puede convertir en un término que sólo depende de la temperatura, del siguiente modo (véase la fig. 17-15).

Sea

N_d = potencia de ruido aportada por un solo amplificador, referida a su entrada

Entonces
$$N_d = KT_eB \tag{17-10}$$

donde T_e es la temperatura equivalente de ruido. Sean

N_o = potencia total de ruido en la salida de un amplificador (watts)
N_i = potencia total de ruido en la entrada del amplificador (watts)
A = ganancia de potencia del amplificador (adimensional)

Entonces, N_o se puede expresar en la siguiente forma

$$N_o = AN_i + AN_d$$

y
$$N_o = AKTB + AKT_eB$$

Al simplificar se obtiene
$$N_o = AKB\,(T + T_e)$$

y el factor general de ruido (F_T) es igual a

$$F_T = \frac{(S/N)_{ent}}{(S/N)_{sal}} = \frac{S/N_i}{AS/N_o} = \frac{N_o}{AN_i} = \frac{AKB\,(T + T_e)}{AKTB} \tag{17-11}$$

$$F_T = \frac{T + T_e}{T} = 1 + \frac{T_e}{T}$$

Ejemplo 17-5

Sean $NF_1 = NF_2 = NF_3 = 3$ dB y $A_1 = A_2 = A_3 = 10$ db en la figura 17-14. Encontrar el índice de ruido total.

Solución Al sustituir en la ecuación 17-8 se obtiene: (nótese que todas las ganancias e índices de ruido están dados en valores absolutos)

$$F_T = F_1 + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1A_2}$$

$$F_T = 2 + \frac{2 - 1}{10} + \frac{2 - 1}{10} = 2.11$$

$$NF_T = 10 \log 2.11 = 3.24 \text{ dB}$$

17-2. Calcule la pérdida en la trayectoria para una señal de 3.4 GHz que se propague 20,000 m.

17-3. Determine el margen de desvanecimiento para un salto de microondas de 60 km. La RF de portadora es 6 GHz, el terreno es muy liso y seco, y el objetivo de confiabilidad es 99.95%.

17-4. Calcule la potencia de ruido para un ancho de banda de 20 MHz en la entrada del receptor, con una temperatura de ruido en la entrada de 290° C.

17-5. Calcule la potencia mínima de transmisión, P_t, para una ganancia de sistema de 120 dB, una C/N mínima en la entrada de 30 dB y una potencia de ruido de −115 dBm en la entrada.

17-6. Calcule la cantidad de pérdida atribuida a un objetivo de confiabilidad de 99.98%.

17-7. Determine la pérdida por sensibilidad al terreno para una portadora de 4 GHz que se propaga sobre un área montañosa muy seca.

17-8. Un sistema de microondas con diversidad de frecuencia funciona con una RF de portadora igual a 7.4 GHz. La FI es una subportadora de frecuencia modulada y bajo índice. La señal de banda base es el sistema FDM de 1800 canales, descrito en el capítulo 16: de 564 kHz a 8284 kHz. Las antenas son platos parabólicos de 4.8 m de diámetro. Las longitudes de alimentador son 150 m en una estación, y 50 m en la otra. El objetivo de confiabilidad es 99.999%. El sistema se propaga sobre un terreno promedio, con clima muy seco. La distancia entre las estaciones es 50 km. La relación mínima de portadora a ruido en la entrada del receptor es 30 dB. Calcule lo siguiente: margen de desvanecimiento, ganancia de antena, pérdidas en la trayectoria en espacio libre, pérdidas totales por ramificación y en alimentadores, potencia de ruido en la entrada del receptor ($C_{mín}$), potencia mínima de transmisión y ganancia del sistema.

17-9. Determine el índice general de ruido para un receptor que tiene dos amplificadores de RF, cada uno con índice de ruido de 6 dB y 10 dB de ganancia, un convertidor mezclador de bajada con índice de ruido de 10 dB y una ganancia de conversión de −6 dB, y ganancia de FI de 40 dB con índice de ruido de 6 dB.

17-10. Un receptor de microondas tiene una potencia total de ruido en la entrada de −102 dBm, y un índice general de ruido de 4 dB. Para tener una relación mínima de C/N de 20 dB a la entrada del detector de FM, calcule la potencia mínima de portadora en la entrada del receptor.

17-11. Calcule la potencia de ruido en la entrada de un receptor, con las siguientes temperaturas y anchos de banda:

T (°C)	B (kHz)
0	10
20	40
200	20
500	50

17-12. Calcule las pérdidas en la trayectoria para las siguientes frecuencias y distancias:

f (MHz)	D (km)
200	0.5
800	0.8
3000	5
5000	10
8000	25
18000	10

17-13. Calcule el margen de desvanecimiento para un salto de microondas de 30 km. La frecuencia RF es 4 GHz, el terreno es agua, y el objetivo de confiabilidad es 99.995%.

17-14. Determine la potencia de ruido para un ancho de banda de 40 MHz en la entrada de un receptor, cuya temperatura de entrada es $T = 400°$ C.

17-15. Para una ganancia de sistema de 114 dB, C/N mínima de entrada = 34 dB y potencia de ruido en la entrada de −111 dBm, calcule la potencia mínima de transmisión (P_t).

17-16. Calcule la cantidad de pérdida que contribuye a un objetivo de confiabilidad de 99.9995%.

17-17. Calcule la pérdida por sensibilidad al terreno, para una portadora de 8 GHz que se propaga por un terreno muy liso y seco.

17-18. Un sistema de microondas con diversidad de frecuencia funciona con una RF = 7.4 GHz. La FI es una subportadora de bajo índice, con frecuencia modulada. La señal de banda base es un sistema DFM de un solo grupo maestro. Las antenas son platos parabólicos de 2.4 m de diámetro. Las longitudes de alimentador son 120 m en una estación, y 80 m en la otra. El objetivo de confiabilidad es 99.995%. El sistema se propaga sobre un terreno promedio, que tiene clima muy seco. La distancia entre estaciones es 40 km. La relación mínima de portadora a ruido en la entrada del receptor es 28 dB. Calcule lo siguiente: margen de desvanecimiento, ganancia de antena, pérdidas en la trayectoria en espacio libre, pérdidas totales en ramificaciones y alimentadores, potencia de entrada al receptor ($C_{mín}$), potencia mínima de transmisión y ganancia del sistema.

17-19. Calcule el índice general de ruido para un receptor que tiene dos amplificadores de RF, cada uno con índice de ruido igual a 8 dB, y una ganancia de 13 dB; un mezclador convertidor descendente con índice de ruido igual a 6 dB y una ganancia de conversión de −6 dB, y una ganancia de FI de 36 dB con índice de ruido de 10 dB.

17-20. Un receptor de microondas tiene potencia de ruido total en la entrada igual a −108 dBm, e índice general de ruido de 5 dB. Para tener una relación C/N mínima de 18 dB en la entrada del detector de FM, calcule la potencia mínima de portadora recibida.

A partir de los proyectos *Syncom*, varias naciones y empresas privadas han lanzado con éxito satélites que hoy se usan para proporcionar comunicaciones globales nacionales, regionales e internacionales. Hoy, hay varios cientos de sistemas satelitales de comunicaciones que funcionan casi en cualquier rincón del mundo. Estas empresas proporcionan circuitos mundiales, telefónicos y de datos, de portadora común fija, emisión de televisión de punto a punto, distribución de red de televisión, emisión de música, servicio de teléfono móvil, servicio de navegación y redes privadas de comunicaciones para grandes corporaciones, agencias gubernamentales y aplicaciones militares.

El primer satélite comercial de telecomunicaciones fue *Intelsat I* (apodado *pájaro madrugador*). Fue lanzado de Cabo Kennedy en 1965, y usó dos transpondedores y un ancho de banda de 25 MHz, para conducir en forma simultánea una señal de televisión y 480 canales de voz. Intelsat quiere decir Organización *In*ternacional de *Tel*ecomunicaciones *Sat*elitales. Es una red comercial global que se manifestó en 1964, desde el interior de las Naciones Unidas. Es un consorcio de más de 120 naciones, que tiene el compromiso de proporcionar comunicaciones satelitales a nivel mundial, sin discriminación, usando cuatro categorías básicas de servicio: telefonía internacional conmutada pública, emisiones de radio y TV, redes de líneas privadas y comerciales y comunicaciones domésticas y regionales. Entre 1966 y 1987, Intelsat lanzó una serie de satélites llamados *Intelsat II, III, IV, V* y *VI*. El *Intelsat VI* tiene una capacidad de 80,000 canales de voz. Entre los lanzamientos más recientes de satélites de Intelsat están los vehículos espaciales de las series 500, 600, 700 y 800.

La antigua Unión Soviética lanzó el primer conjunto de *satélites domésticos* (Domsats) o *nacionales* en 1966, llamándolos *Molniya*, que quiere decir "relámpago". Los Domsats son satélites que son propiedad de un solo país, que los opera y usa. En 1972, Canadá lanzó su primer satélite comercial llamado *Anik*, palabra Inuit que quiere decir "hermanito". La Western Union lanzó su primer satélite Westar en 1984, y la Radio Corporation of America (RCA) lanzó sus primeros Satcom (*Sat*ellite *Com*munications) en 1975. Hoy, en Estados Unidos, una empresa de propiedad pública, llamada *Com*munications *Sat*ellite Corporation (Comsat) reglamenta el uso y funcionamiento de los satélites estadounidenses, y también establece sus tarifas. Aunque una empresa o gobierno puede poseer un satélite, por lo general, sus instalaciones se ponen a la disposición de quien quiera pagarlas. En Estados Unidos se usa hoy la mayor parte del tiempo mundial disponible de satélites: el 24%; el segundo lugar es de Gran Bretaña con 13% y les sigue Francia con 6%.

LEYES DE KEPLER

Un satélite permanece en órbita porque las fuerzas centrífugas causadas por su rotación en torno a la Tierra se equilibran con la atracción gravitacional de ésta. A principios del siglo XVII, al investigar las leyes del movimiento planetario (es decir, el movimiento de los planetas y sus cuerpos asociados llamados lunas), Johannes Kepler (1571-1630), astrónomo alemán, descubrió las leyes que gobiernan el movimiento de los satélites. Las leyes del movimiento planetario describen la forma de la órbita, las velocidades del planeta y la distancia de un planeta con respecto al Sol. Las *leyes de Kepler* se pueden enunciar en forma sencilla como sigue: 1) los planetas describen elipses con el Sol en uno de los focos, 2) la línea que une al Sol con un planeta barre áreas iguales en intervalos iguales de tiempo, y 3) el cuadrado del tiempo de revolución de un planeta, dividido entre el cubo de su distancia promedio al Sol es un número igual para todos los planetas. Las leyes de Kepler se pueden aplicar a dos cuerpos cualesquiera en el espacio que interaccionen por gravitación. El mayor de los dos cuerpos se llama *primario*, y el menor es el *secundario*, o *satélite*.

La primera ley de Kepler establece que un satélite describe una órbita alrededor de un cuerpo primario (como la Tierra) siguiendo una trayectoria elíptica. Una elipse tiene dos *focos*, como se ve en la fig. 18-1a: F_1 y F_2, y el centro de masa, llamado baricentro, de un sistema de dos cuerpos siempre está en uno de los focos. Como la masa de la Tierra es mucho mayor que la del satélite, el centro de masa siempre coincide con el centro de la Tierra. Las propiedades geométricas de la elipse se suelen referir a uno de los focos, que lógicamente se selecciona como el que está en el centro de la Tierra.

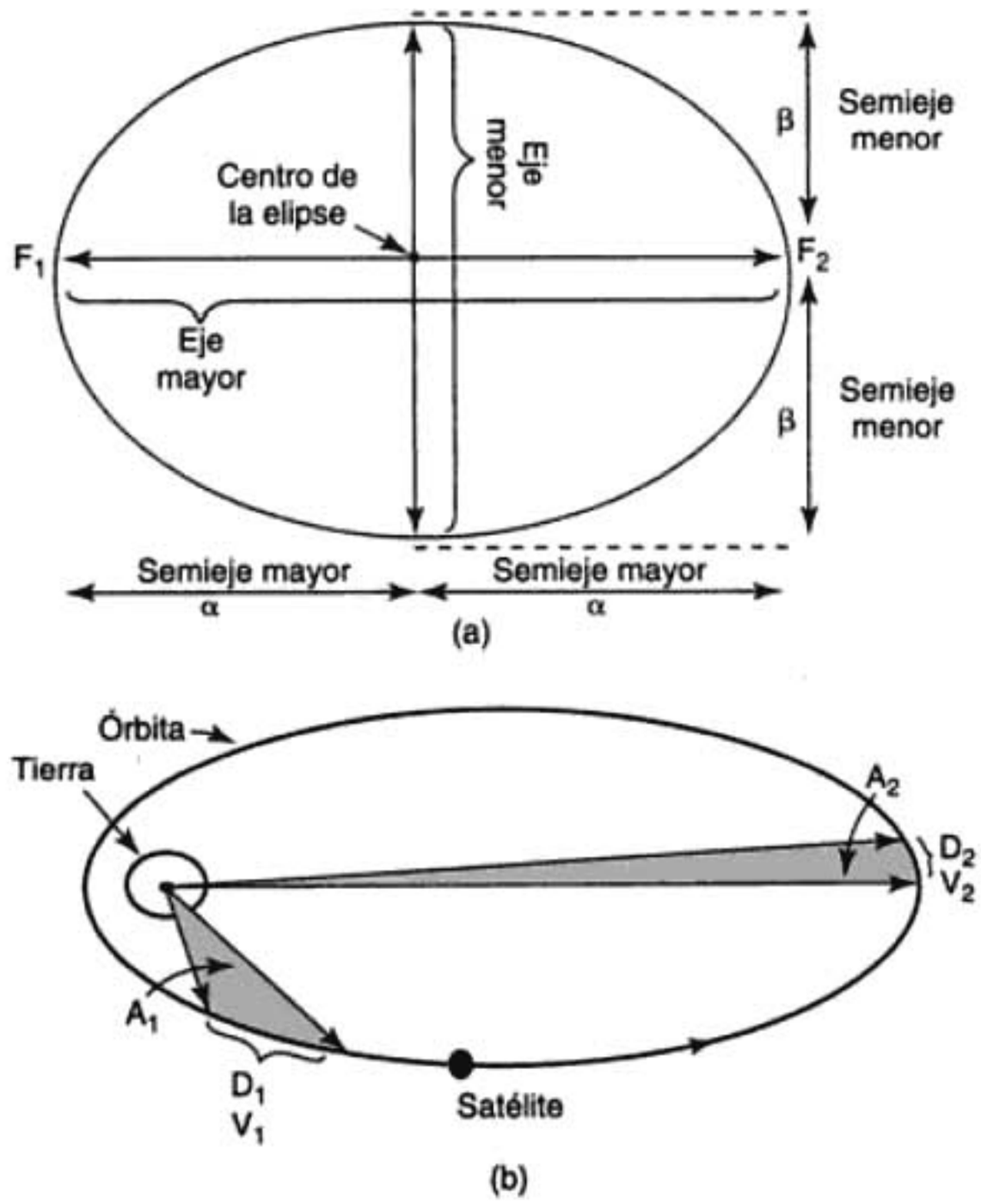


FIGURA 18-1 (a) Focos F_1 y F_2; semieje mayor a y semieje menor b de una elipse; (b) segunda ley de Kepler

Para el semieje mayor (α) y el semieje menor (β) de la fig. 18-1a, la *excentricidad* de la elipse se define como sigue

$$\epsilon = \frac{\sqrt{\alpha^2 - \beta^2}}{\alpha} \tag{18-1}$$

siendo ϵ la excentricidad.

La segunda ley de Kepler, enunciada en 1609 con la primera, se conoce como la *ley de las áreas*, y establece que para intervalos iguales de tiempo, un satélite barre áreas iguales en el plano de su órbita, con foco en el baricentro. Como se ve en la fig. 18-1b, para un satélite que recorre las distancias D_1 y D_2 metros en 1 segundo, las áreas A_1 y A_2 son iguales. Debido a la ley de áreas iguales, la distancia D_1 debe ser mayor que D_2 y, en consecuencia, la velocidad V_1 debe ser mayor que la velocidad V_2. La velocidad es máxima en el punto de máxima aproximación a la Tierra (llamado *perigeo*), y la velocidad es mínima en el punto más alejado de la Tierra (llamado *apogeo*).

La tercera ley de Kepler, publicada en 1619, se llama a veces la *ley armónica*. Establece que el cuadrado del periodo (el tiempo en recorrer la órbita) es proporcional al cubo de la distancia promedio entre el primario y el satélite. Esta distancia promedio es igual al semieje mayor; por lo anterior, se puede enunciar matemáticamente la tercera ley de Kepler como sigue

$$\alpha = AP^{2/3} \tag{18-2}$$

donde A = constante (kilómetros)
α = semieje mayor (kilómetros)
P = días solares medios en la Tierra

y P es la relación del tiempo de un día sideral (t_s = 23 horas y 56 minutos) entre el tiempo de una revolución de la Tierra en torno a su eje (t_e = 24 horas)

Por lo anterior,

$$P = \frac{t_s}{t_e}$$

$$= \frac{1436 \text{ minutos}}{1440 \text{ minutos}}$$

$$= 0.9972$$

Al rearreglar la ecuación 18-2 y despejar la constante A en el caso de la Tierra, se obtiene

$$A = 42241.0979$$

Las ecuaciones 18-1 y 18-2 se aplican al caso ideal, cuando un satélite gira en torno a un cuerpo perfectamente esférico, sin fuerzas exteriores. En realidad, el agrandamiento ecuatorial de la Tierra y las fuerzas perturbadoras externas producen desviaciones del movimiento ideal del satélite. Sin embargo, y por fortuna, se pueden calcular y compensar las desviaciones principales. Los satélites con órbita cercana a la Tierra quedan afectados por la resistencia atmosférica y por el campo magnético terrestre. Sin embargo, para los satélites más alejados, las principales fuerzas perturbadoras son los campos gravitacionales del Sol y de la Luna.

ÓRBITAS DE SATÉLITES

La mayoría de los satélites que se han mencionado hasta ahora son *orbitales*, que son *asíncronos*. Los satélites asíncronos giran en torno a la Tierra en órbitas elípticas o circulares, como se ve en las figs. 18-2a y b. En una órbita circular, la velocidad de rotación es constante; sin embargo, en órbitas elípticas la velocidad depende de la altura del satélite sobre la Tierra. Es mayor cuando el satélite está cerca de la Tierra que cuando está más lejos.

Si el satélite describe su órbita en la misma dirección que la de rotación de la Tierra (en el sentido de las manecillas del reloj) y su velocidad angular es mayor que la de la Tierra ($\omega_s > \omega_e$), la órbita se llama prógrada o posígrada. Si su órbita tiene dirección contraria a la de rotación de la Tierra, o la misma dirección, pero con una velocidad angular menor que la de la Tierra ($\omega_s < \omega_e$), la órbita del satélite es *retrógrada*. La mayoría de los satélites asíncronos gira en torno a la Tierra con órbitas prógradas. Por consiguiente, su posición cambia en forma continua con respecto a un punto fijo de la Tierra. Por lo mismo, los satélites asíncronos se deben usar cuando están disponibles, lo cual puede ser sólo de 15 minutos por órbita. Otra desventaja de estos satélites es la necesidad de equipo complicado y costoso de rastreo en las estaciones terrestres, para poder localizar al satélite cuando se avista en cada órbita, y entonces enganchar la antena al satélite y seguirlo a medida que va pasando. Sin embargo, una gran ventaja de los satélites orbitales es que no se requieren a bordo cohetes de propulsión para mantenerlos en sus órbitas respectivas.

Categorías de elevación de satélites

En general, los satélites se clasifican por tener ya sea una *órbita terrestre baja* (LEO, por *low Earth orbit*), órbita terrestre intermedia (MEO, por *medium Earth orbit*) o una órbita terrestre geosíncrona (GEO, por *geosynchronous Earth orbit*). La mayoría de los satélites LEO trabaja en el intervalo de frecuencias de 1.0 a 2.5 GHz. *Iridium*, el sistema telefónico satelital de Motorola, es un sistema LEO que usa una constelación de 66 satélites que orbitan a unas 480 millas sobre la superficie terrestre. La principal ventaja de estos satélites es que la pérdida de trayectoria entre las estaciones terrestres y los vehículos espaciales es mucho menor que para satélites que giran en órbitas de altura intermedia o grande. La menor pérdida de trayectoria equivale a menores potencias de transmisión, menores antenas y menor peso.

Los satélites MEO trabajan en la banda de frecuencias de 1.2 a 1.66 GHz, y giran entre 6000 y 12,000 millas sobre la Tierra. El sistema satelital *NAVSTAR*, del Departamento de

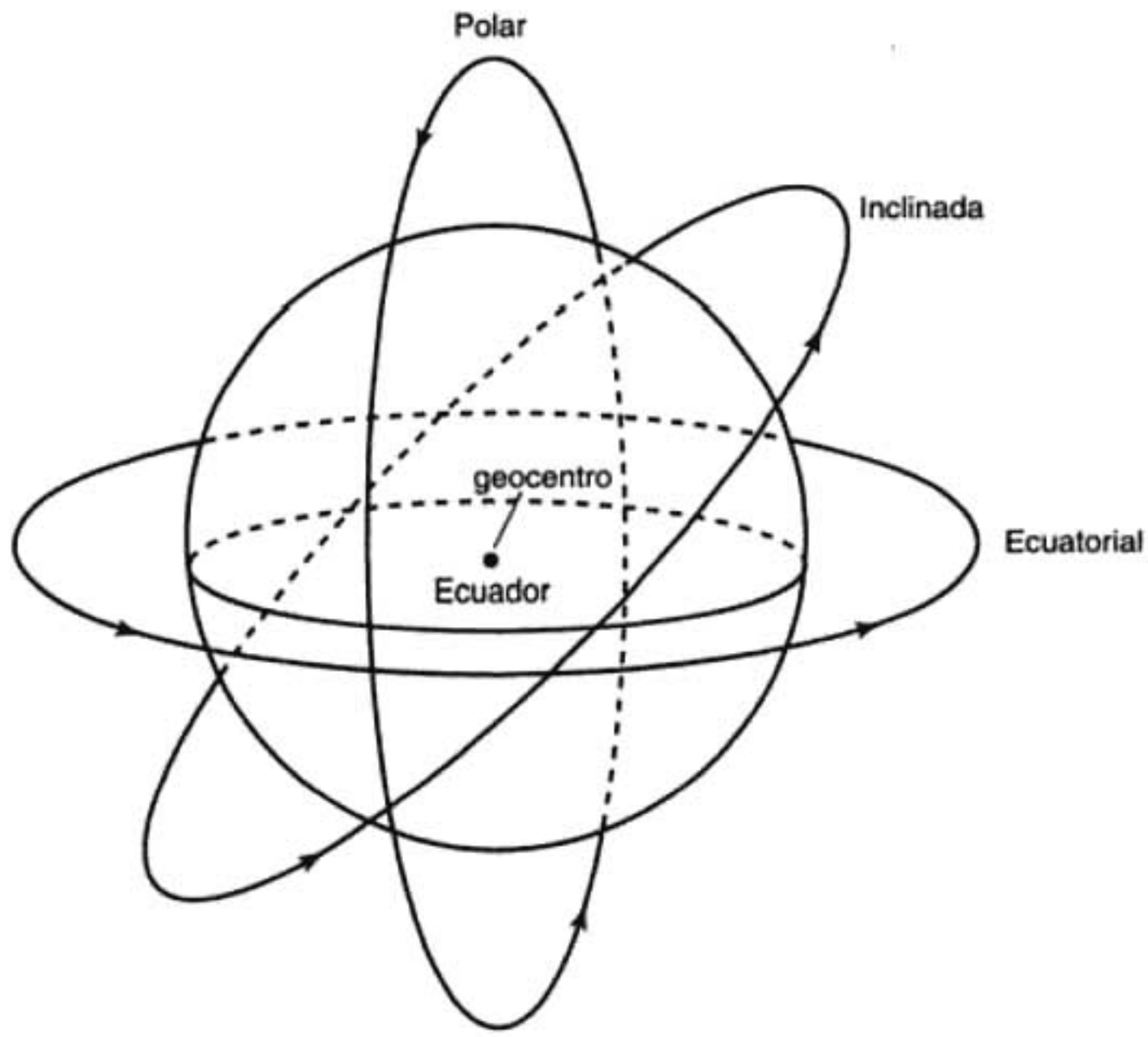


FIGURA 18-4 Clases de órbitas de satélite

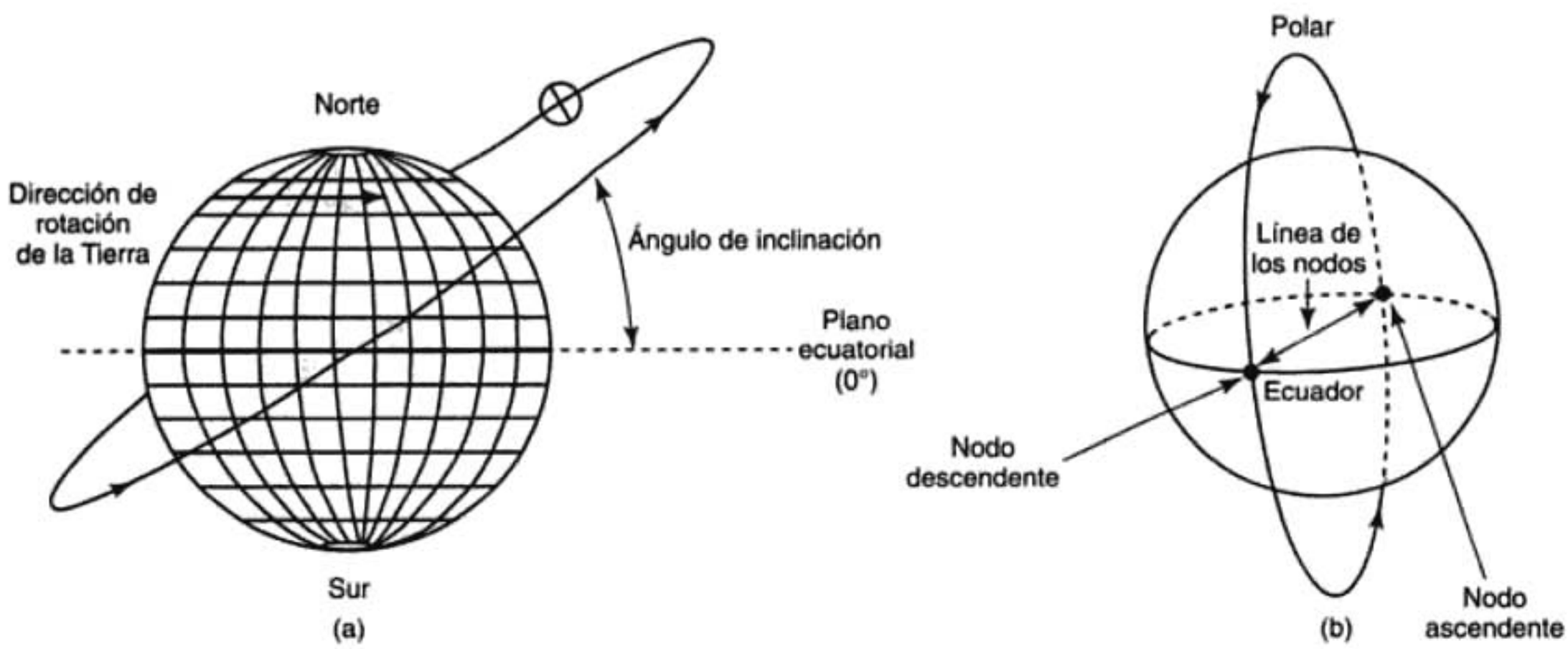


FIGURA 18-5 (a) Ángulo de inclinación; (b) nodo ascendente, nodo descendente y línea de los nodos

La Tierra no es una esfera perfecta, porque se ensancha en el ecuador. De hecho, se creía en forma errónea, hasta principios del siglo XIX, que el Volcán Chimborazo, de 6310 msnm, era el punto más alto del planeta. Sin embargo, a causa del agrandamiento ecuatorial, sucede que el Chimborazo es el punto más alejado del centro de la Tierra. Un efecto importante del ensanchamiento ecuatorial es que causa que las órbitas elípticas giren de tal manera que el apogeo y el perigeo giren en torno a la tierra. A este fenómeno se le llama *rotación de la línea de los ápsides*; sin embargo, si el ángulo de inclinación es 63.4°, la rotación es cero. Entonces los satélites que

2. Los satélites geosíncronos están a gran altura e introducen retardos de propagación mucho mayores. El retardo de propagación de ida y vuelta entre dos estaciones terrestres, pasando por un satélite geosíncrono, es de 500 a 600 ms.
3. Los satélites geosíncronos requieren mayores potencias de transmisión y receptores más sensibles, por las mayores distancias y mayores pérdidas en la trayectoria.
4. Se requieren artificios espaciales de gran precisión para poner en órbita un satélite geosíncrono, y para mantenerlo en ella. También, se requieren motores de propulsión a bordo del satélite, para mantenerlo en su órbita respectiva.

ÁNGULOS VISUALES DE UNA ANTENA

Para optimizar el funcionamiento de un sistema de comunicaciones por satélite, la dirección de ganancia máxima de una antena de estación terrestre (que a veces se llama la *mira*) se debe apuntar directamente al satélite. Para asegurar que esté alineada la antena de la estación terrestre, se deben determinar dos ángulos: el *azimut*, y la *elevación*. El ángulo de azimut y el de elevación se llaman, en conjunto, *ángulos de visual* de la antena. Con los satélites geosíncronos, los ángulos de visual de las antenas terrestres sólo se deben ajustar una vez, porque el satélite permanece siempre en determinada posición, salvo pequeñas variaciones ocasionales.

La ubicación de un satélite se suele especificar en términos de latitud y longitud, en forma parecida a como se ubica un punto sobre la Tierra; sin embargo, como un satélite está en órbita a muchas millas sobre la superficie terrestre, no tiene latitud ni longitud. En consecuencia, su lugar se identifica con un punto en la superficie terrestre directamente abajo del satélite. Este punto se llama *punto subsatelital* (SSP, de *subsatellite point*), y para los satélites geosíncronos el SSP debe caer en el ecuador. Estos puntos subsatelitales, y las ubicaciones de las estaciones terrestres se especifican con coordenadas normales de latitud y longitud. La convención normal especifica los ángulos de longitud entre 0° y 180° al este o al oeste del meridiano de Greenwich. Las latitudes del Hemisferio Norte son ángulos de 0° a 90° N, y las del Hemisferio Sur son de 0° a 90° S. Como los satélites geosíncronos están directamente arriba del ecuador, todos tienen latitud 0°. Por consiguiente, las ubicaciones de los satélites geosíncronos se suelen especificar en grados de longitud al este o al oeste del meridiano de Greenwich, por ejemplo, 122° Oeste o 78° E. La fig. 18-9 muestra la posición de un vehículo satelital geosíncrono hipotético (GSV, de *geosynchronous satellite vehicle*, su respectivo punto subsatelital (SSP) y una estación terrestre (ES) arbitraria, todos ellos con relación al geocentro. El SSP del satélite de la fig. 18-9 es 30° E de longitud y 0° de latitud. La estación terrestre tiene una ubicación de 30° O de longitud y 20° N de latitud.

Ángulo de elevación

El *ángulo de elevación* es el ángulo vertical que se forma entre la dirección de movimiento de una onda electromagnética irradiada por una antena de estación terrestre que apunta directamente hacia un satélite, y el plano horizontal. Mientras menor es el ángulo de elevación, la distancia que debe recorrer una onda propagada a través de la atmósfera terrestre es mayor. Como en el caso de cualquier onda propagada por la atmósfera terrestre, sufre absorción, y también se puede contaminar mucho con ruido. En consecuencia, si el ángulo de elevación es muy pequeño y la distancia que la onda viaja por la atmósfera terrestre es demasiado grande, la onda se puede deteriorar hasta el grado de ya no proporcionar una calidad aceptable de transmisión. En general, se considera que 5° es el ángulo de elevación mínimo aceptable. La fig. 8-10 muestra la manera en que el ángulo de elevación afecta la intensidad de señal de una onda electromagnética propagada, debido a la absorción atmosférica normal, la cual es causada por la niebla espesa y la lluvia intensa. Se puede ver que la banda de 14/12 GHz de la fig. 18-10b se afecta más que la banda de 6/4 GHz de la fig. 18-10a, debido a las menores longitudes de onda asociadas con las mayores frecuencias. También se ve en la figura que en ángulos de elevación menores que 5°, la cantidad de potencia de señal perdida aumenta mucho. La fig. 18-11a ilustra el ángulo de elevación de una antena de estación terrestre con respecto a un plano horizontal.

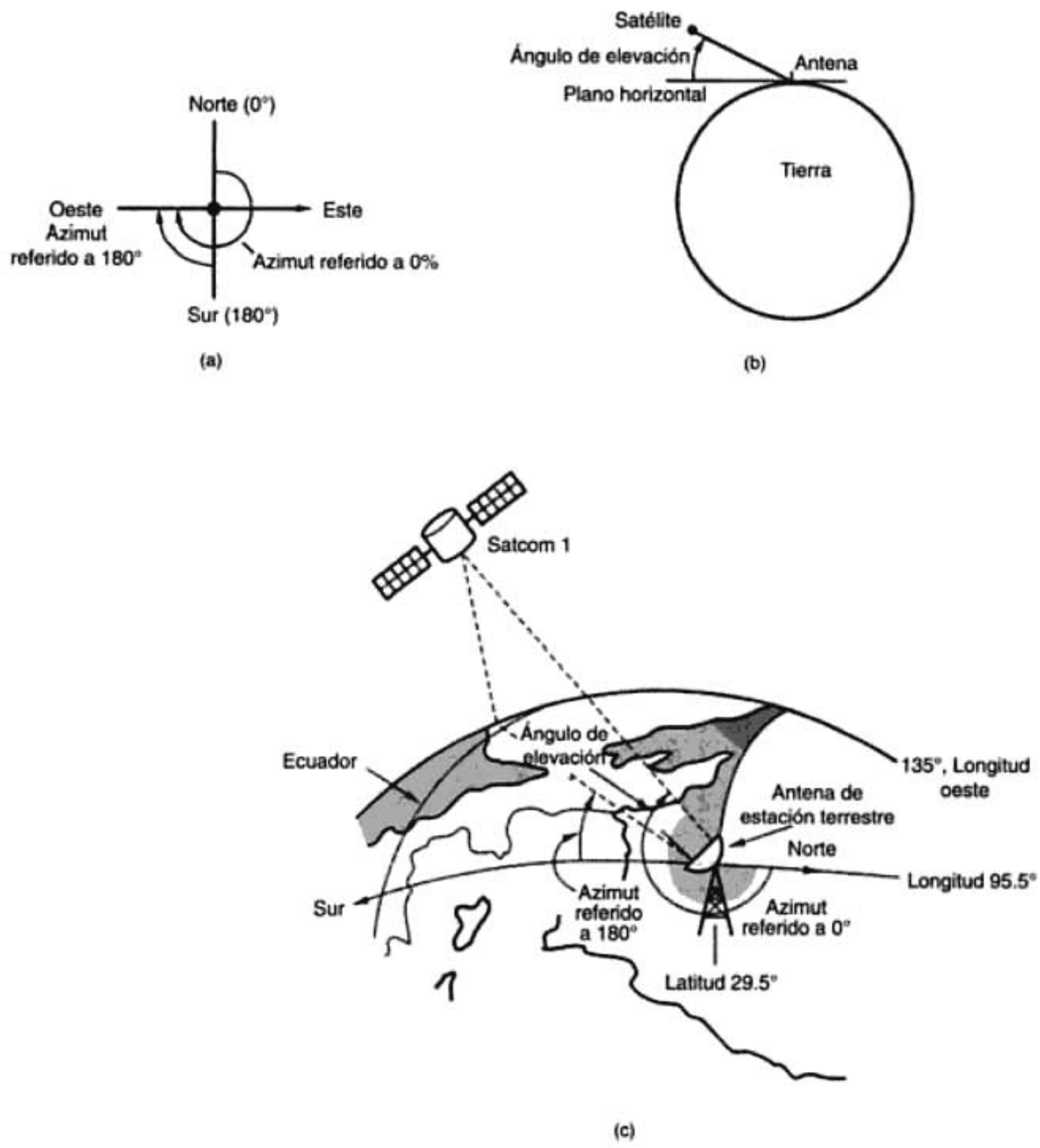


FIGURA 18-11 "Ángulos de visión": azimut y ángulo de elevación

Ángulo de azimut

El *azimut* es la distancia angular horizontal a una dirección de referencia, que puede ser el punto sur o el norte del horizonte. El *ángulo de azimut* se define como el ángulo horizontal de apuntamiento de una antena de estación terrestre. Para fines de navegación, el ángulo de azimut se suele medir en grados a partir del norte verdadero, en el sentido de las manecillas del reloj. Sin embargo, para las estaciones terrestres del Hemisferio Norte con satélites en órbitas geosíncronas, la referencia en general del ángulo de azimut es la dirección del sur verdadero (es decir, 180°). La fig. 18-11b muestra el ángulo de azimut referido a la dirección norte (0°) y a la dirección sur (180°), y la fig. 18-11c muestra los ángulos de elevación y de azimut de una antena de estación terrestre que apunta a un satélite.

Los ángulos de elevación y de azimut dependen de la latitud y la longitud de la estación terrestre, y del satélite en órbita. Para un satélite geosíncrono en una órbita ecuatorial, el procedimiento para determinar los ángulos de elevación y de azimut es el siguiente: en un buen mapa, determinar la longitud y la latitud de la estación terrestre. Con la tabla 18-1 determinar la longitud del satélite de interés. Calcular la diferencia, en grados (ΔL), entre la longitud del satélite y la de la estación terrestre. A continuación, con la fig. 18-12, determinar el ángulo de

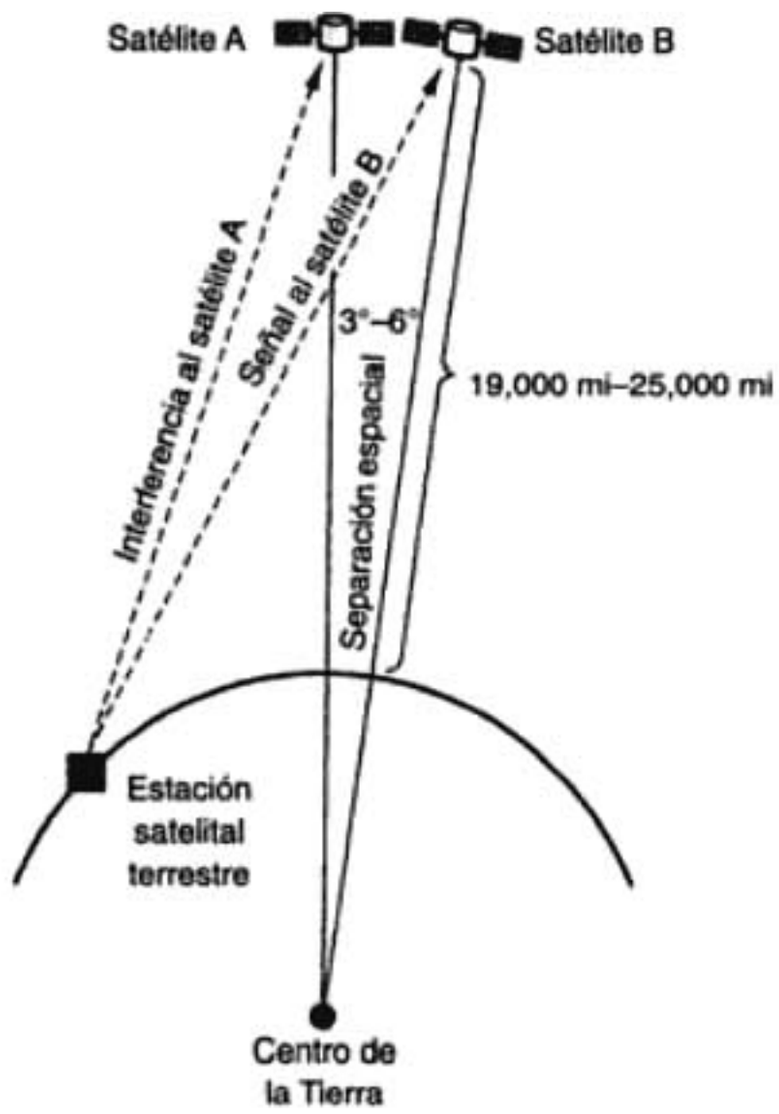


FIGURA 18-15 Separación espacial de satélites en órbita geosincrona

dor. La posición en el intervalo depende de la banda de frecuencias de comunicaciones que se use. Los satélites que trabajan la misma o casi la misma frecuencia deben tener una separación suficiente en el espacio para evitar interferir entre sí (vea la fig. 18-15). Hay un límite realista de la cantidad de satélites que pueden *estacionarse* en determinada área del espacio. La *separación espacial* requerida depende de las siguientes variables:

1. Anchos de banda y lóbulos laterales de radiación de las antenas, tanto de la estación terrestre como del satélite.
2. Frecuencia de portadora de RF.
3. Técnica de codificación que se use.
4. Límites aceptables de interferencia.
5. Potencia de la portadora de transmisión.

En general, se requiere una separación espacial de 3° a 6°, que depende de estas variables.

Las frecuencias de portadora más comunes que se usan en comunicaciones vía satélite son las bandas de 6/4 y de 14/12 GHz. El primer número es la frecuencia de enlace de subida (estación terrestre a transpondedor), y el segundo es la frecuencia de enlace de bajada (transpondedor a estación terrestre). Se usan frecuencias distintas de enlace de subida y de bajada para evitar que haya radiación de pérdida (capítulo 17). Mientras mayor sea la frecuencia de la portadora, el diámetro necesario de la antena es menor, para una ganancia dada. La mayoría de los satélites domésticos usa la banda de 6/4 GHz. Desafortunadamente, esta banda también se usa mucho en sistemas terrestres de microondas. Se debe tener cuidado al diseñar una red satelital, para evitar interferencias con otros enlaces existentes de microondas.

Algunas posiciones de órbita geosíncrona tienen mayor demanda que otras. Por ejemplo, la posición atlántica media, que se usa para interconectar Norteamérica con Europa, es de demanda excepcionalmente alta; la pacífica media también es otra de gran demanda.

Las frecuencias asignadas por la Conferencia Mundial Administrativa de Radio (WARC, de *World Administrative Radio Conference*) se resumen en la fig. 18-16. La tabla 18-2 muestra los anchos de banda disponibles para varios servicios en Estados Unidos. Entre estos servicios están *punto fijo* (entre estaciones terrestres en puntos geográficos fijos), *emisión* (cobertura de áreas amplias), *móvil* (tierra a vehículos en aire, mar o tierra) e *intersatelital* (enlaces cruzados entre satélites).

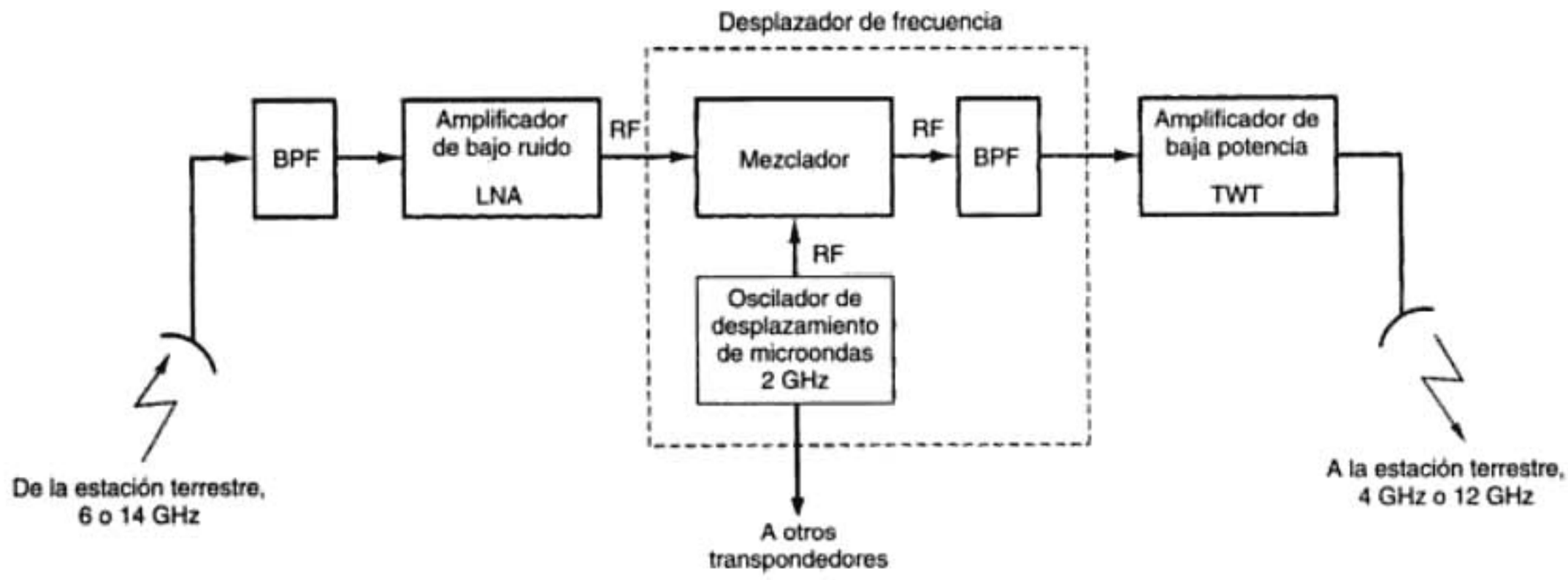


FIGURA 18-20 Transpondedor de satélite

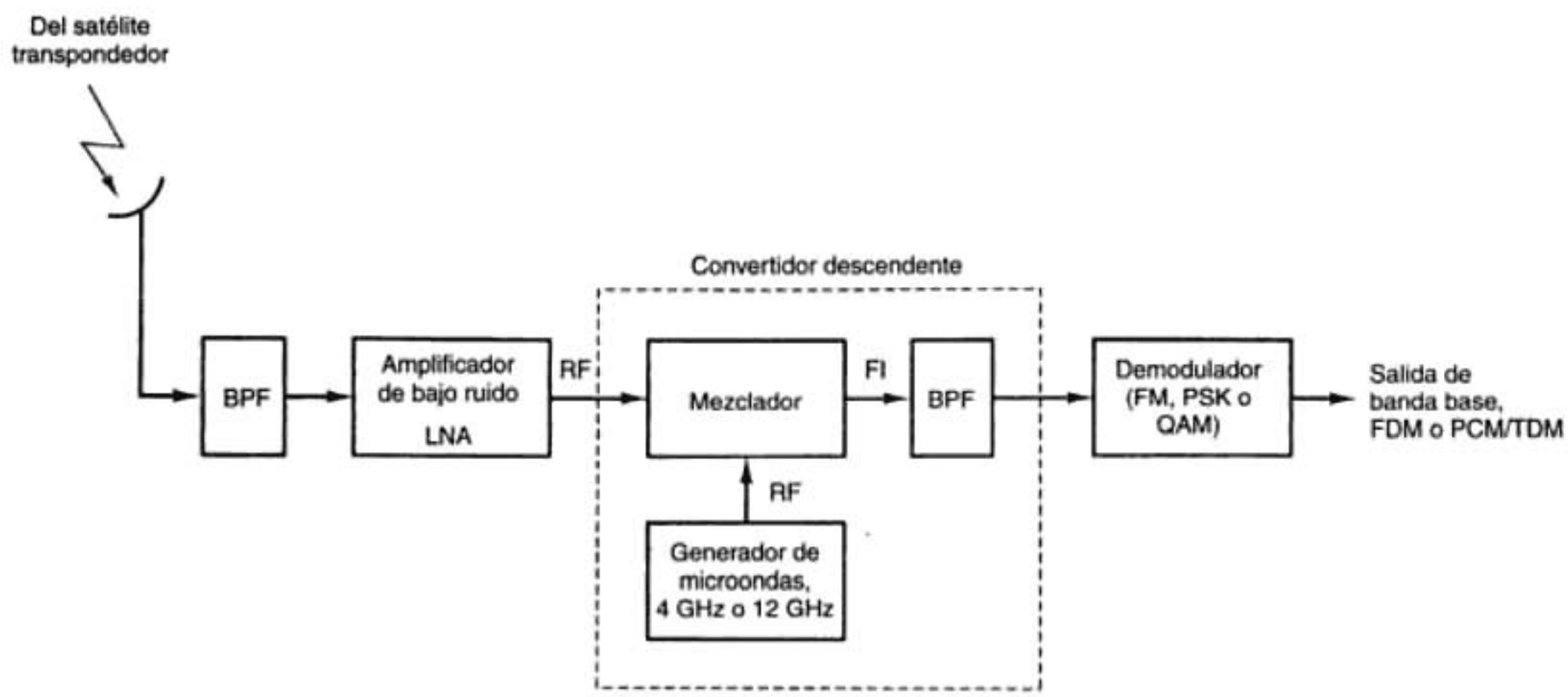


FIGURA 18-21 Modelo de enlace satelital de bajada

Modelo de enlace de bajada

Un receptor en la estación terrestre comprende un BPF de entrada, un LNA y un convertidor descendente de RF a FI. La fig. 18-21 muestra un diagrama de bloques de un receptor típico de estación terrestre. También aquí, el BPF limita la potencia de entrada de ruido al LNA. Éste es un dispositivo de gran sensibilidad y bajo ruido, como un amplificador de diodo túnel o un amplificador paramétrico. El convertidor descendente de RF a FI es una combinación de mezclador y filtro pasabandas, que convierte la señal de RF recibida a una frecuencia FI.

Enlaces cruzados

A veces se presenta una aplicación en donde es necesario comunicarse entre satélites. Esto se hace con *enlaces satelitales cruzados* o *enlaces intersatelitales* (ISL, de *intersatellite links*), como se ve en la fig. 18-22. Una desventaja de usar un ISL es que tanto el transmisor como el receptor están *acotados por espacio*. En consecuencia, tanto la potencia de salida del transmisor como la sensibilidad de entrada del receptor son limitadas.

(c) Determinar C/N si el ruido se mide en un punto antes del filtro pasabandas, donde el ancho de banda es igual al triple del ancho de banda de Nyquist.

Solución (a) La manipulación 8-PSK tiene una eficiencia de ancho de banda de 3 bps/Hz y, en consecuencia, requiere un ancho mínimo de banda de la tercera parte, es decir, 30 MHz. Según la fig. 18-24, la C/N mínima es 18.5 dB. Al sustituir en la ecuación 18-5 se obtiene

$$\frac{E_b}{N_0} = 18.5 \text{ dB} + 10 \log \frac{30 \text{ MHz}}{90 \text{ Mbps}}$$
$$= 18.5 \text{ dB} + (-4.8 \text{ dB}) = 13.7 \text{ dB}$$

(b) La ecuación 18-15 se rearregla, y se sustituye este valor de E_b/N_0, para obtener

$$\frac{C}{N} = 13.7 \text{ dB} - 10 \log \frac{60 \text{ MHz}}{90 \text{ Mbps}}$$
$$= 13.7 \text{ dB} - (-1.77 \text{ dB}) = 15.47 \text{ dB}$$

(c) También, rearreglando la ecuación 18-15 y sustituyendo E_b/N_0, se obtiene

$$\frac{C}{N} = 13.7 \text{ dB} - 10 \log \frac{90 \text{ MHz}}{90 \text{ Mbps}}$$
$$= 13.7 \text{ dB} - 0 \text{ (dB)} = 13.7 \text{ dB}$$

Es evidente, de acuerdo con los ejemplos 18-6 y 18-7, que las relaciones E_b/N_0 y C/N sólo son iguales cuando el ancho de banda de ruido es igual a la frecuencia de bits. También, a medida que aumenta el ancho de banda en el punto de medición, C/N disminuye.

Cuando son distintos el esquema de modulación, la frecuencia de bits, el ancho de banda y la relación C/N de dos sistemas digitales de radio, con frecuencia es difícil determinar cuál de ellos tiene la menor probabilidad de error. La relación E_b/N_0 es independiente del ancho de banda y del esquema de modulación, por lo que es un denominador común adecuado para comparar la probabilidad de errores en dos sistemas digitales de radio.

Relación de ganancia a temperatura equivalente de ruido

En esencia, la *relación de ganancia a temperatura equivalente de ruido*, G/T_e, es una cifra de mérito que representa la calidad de un receptor de satélite o de estación terrestre. Para un receptor, G/T_e es la relación de la ganancia de la antena receptora a la temperatura equivalente de ruido, T_e, del receptor. Debido a las potencias de portadora tan extremadamente pequeñas en recepción que hay en los sistemas satelitales, con mucha frecuencia se ubica físicamente un amplificador de bajo ruido (LNA, de *low-noise amplifier*). Cuando así sucede, G/T_e es una relación de la ganancia de la antena receptora más la ganancia del LNA, entre la temperatura equivalente de ruido. Es decir, para este caso,

$$\frac{G}{T_e} = \frac{A_r + A_{(\text{LNA})}}{T_e} \tag{18-16}$$

Expresada en logaritmos

$$\frac{G}{T_e}(\text{dBK}^{-1}) = A_{r(\text{dB})} + A_{(\text{LNA})(\text{dB})} - T_{e(\text{dBK})} \tag{18-17}$$

La relación G/T_e es un parámetro útil para determinar las relaciones E_b/N_0 y C/N en los receptores del satélite transpondedor y de la estación terrestre. En esencia, es el único parámetro que se requiere en un receptor de satélite o de estación terrestre cuando se completa el cálculo de los gastos de enlace.

Ejemplo 18-8

Calcular la figura de mérito G/T_e, para un satélite transpondedor con ganancia de antena receptora de 12 dB, ganancia de LNA de 10 dB y una temperatura equivalente de ruido de 26 dBK.

Solución Al sustituir en la ecuación 18-17 se obtiene

$$\frac{G}{T_e} = 12 \text{ dB} + 10 \text{ dB} - 26 \text{ dBK} = -4 \text{ dBK}^{-1}$$

TABLA 18-4 Parámetros de tres sistemas satelitales hipotéticos

	Sistema A: 6/4 GHz, cobertura global, modulación QPSK, 60 Mbps	Sistema B: 14/12 GHz, cobertura global, modulación 8PSK, 90 Mbps	Sistema C: 14/12 GHz, cobertura global, modulación 8 PSK, 120 Mbps
Enlace de subida			
Potencia de salida del transmisor (saturación, dBW)	35	25	33
Pérdidas por reducción en la estación terrestre (dB)	2	2	3
Pérdidas por ramificación y en alimentador de la estación terrestre (dB)	3	3	4
Otras pérdidas atmosféricas (dB)	0.6	0.4	0.6
Ganancia de la estación terrestre (dB)	55	45	64
Pérdidas en el trayecto por el espacio libre (dB)	200	208	206.5
Ganancia de la antena receptora del satélite (dB)	20	45	23.7
Pérdidas por ramificación y en alimentador, en el satélite (dB)	1	1	0
Temperatura equivalente de ruido en satélite (K)	1000	800	800
G/T_e en el satélite (dBK^{-1})	−10	16	−5.3
Enlace de bajada			
Potencia de salida del transmisor (saturación, dBW)	18	20	10
Pérdidas por reducción en el satélite (dB)	0.5	0.2	0.1
Pérdidas por ramificación y en alimentador del satélite (dB)	1	1	0.5
Otras pérdidas atmosféricas (dB)	0.8	1.4	0.4
Ganancia de la antena del satélite (dB)	16	44	30.8
Pérdidas en el trayecto por el espacio libre (dB)	197	206	205.6
Ganancia de la antena de recepción en la estación terrestre (dB)	51	44	62
Pérdidas por ramificación y en alimentador de la estación terrestre (dB)	3	3	0
Temperatura equivalente de ruido de la estación terrestre (K)	250	1000	270
G/T_e de la estación terrestre (dBK^{-1})	27	14	37.7

Enlace de bajada

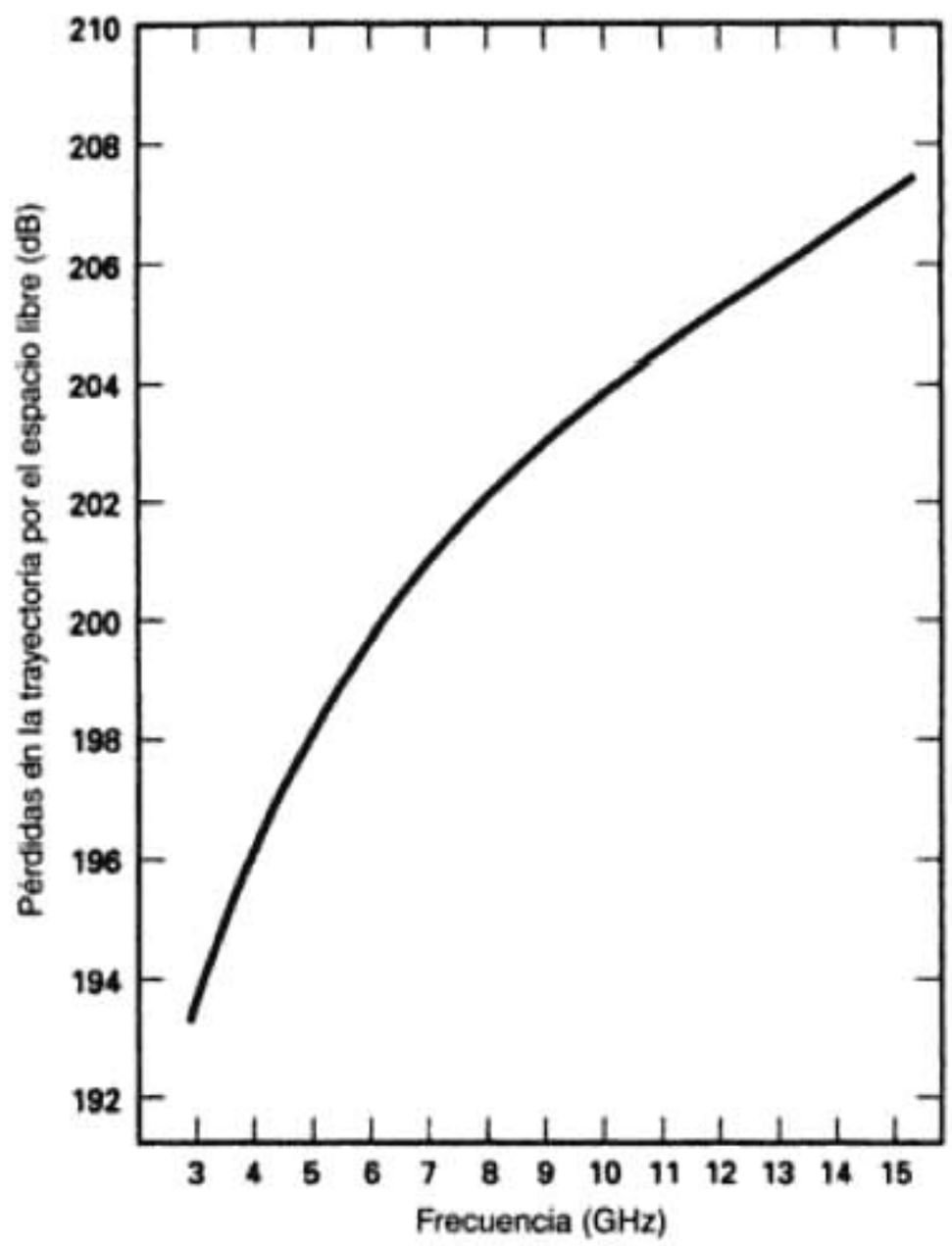
1. Potencia de salida del transmisor en satélite, a la saturación, 10 W	10	dBW
2. Pérdidas por reducción en satélite	0.1	dB
3. Pérdidas por ramificación y en alimentador de satélite	0.5	dB
4. Ganancia de la antena transmisora del satélite (de la fig. 18-27, 0.37 m a 12 GHz)	30.8	dB
5. Otras pérdidas atmosféricas de enlace de bajada	0.4	dB
6. Pérdidas en la trayectoria por el espacio libre (de la fig. 18-28, a 12 GHz)	205.6	dB
7. Ganancia de la antena receptora de la estación terrestre (15 m, 12 GHz)	62	dB
8. Pérdidas por ramificación y en alimentador de la estación terrestre	0	dB
9. Temperatura equivalente de ruido en la estación terrestre	270	K
10. Relación G/T_e de la estación terrestre	37.7	dBK^{-1}
11. Frecuencia de bits	120	Mbps
12. Esquema de modulación	8-PSK	

Solución *Gastos de enlace de subida:* Expresada en logaritmos,

$$\text{EIRP (estación terrestre)} = P_t + A_t - L_{bo} - L_{bf}$$
$$= 33 \text{ dBW} + 64 \text{ dB} - 3 \text{ dB} - 4 \text{ dB} = 90 \text{ dBW}$$

Densidad de potencia de portadora en la antena del satélite

$$C' = \text{EIRP (estación terrestre)} - L_p - L_u$$
$$= 90 \text{ dBW} - 206.5 \text{ dB} - 0.6 \text{ dB} = -117.1 \text{ dBW}$$



Corrección por ángulo de elevación:

Ángulo	+db
90°	0
45°	0.44
0°	1.33

FIGURA 18-28 Pérdidas en la trayectoria por espacio libre (Lp) calculadas con $L_p = 183.5 + 20 \log f$ (GHz), ángulo de elevación = 90° y distancia = 35,9300 km

Densidad de potencia de portadora en la antena de la estación terrestre

$$C' = \text{EIRP} - L_p - L_d$$
$$= 40.2 \text{ dBW} - 205.6 \text{ dB} - 0.4 \text{ dB} = -165.8 \text{ dBW}$$

C/N_0 en el receptor de la estación terrestre

$$\frac{C}{N_0} = \frac{C}{KT_e} = \frac{C}{T_e} \times \frac{1}{K} \quad \text{donde } \frac{C}{T_e} = C' \times \frac{G}{T_e}$$

Así,

$$\frac{C}{N_0} = C' \times \frac{G}{T_e} \times \frac{1}{K}$$

Expresada en logaritmos,

$$\frac{C}{N_0} = C' + \frac{G}{T_e} - 10 \log (1.38 \times 10^{-23})$$
$$= -165.8 \text{ dBW} + (37.7 \text{ dBK}^{-1}) - (-228.6 \text{ dBWK}) = 100.5 \text{ dB}$$

Un método alternativo para despejar C/N_0 es

$$\frac{C}{N_0} = C' + A_r - T_e - K$$
$$= -165.8 \text{ dBW} + 62 \text{ dB} - 10 \log 270 - (-228.6 \text{ dBWK})$$
$$= -165.8 \text{ dBW} + 62 \text{ dB} - 24.3 \text{ dBK}^{-1} + 228.6 \text{ dBWK} = 100.5 \text{ dB}$$

$$\frac{E_b}{N_0} = \frac{C}{N_0} - 10 \log f_b$$
$$= 100.5 \text{ dB} - 10 \log (120 \times 10^6)$$
$$= 100.5 \text{ dB} - 80.8 \text{ dB} = 19.7 \text{ dB}$$

PREGUNTAS

18-1. Describa un satélite, en forma breve.

18-2. ¿Qué es un satélite pasivo? ¿Qué es un satélite activo?

18-3. Describa las diferencias entre satélites asíncronos y geosíncronos.

18-4. Defina *prógrada* y *retrógrada*.

18-5. Defina *apogeo* y *perigeo*.

18-6. Explique en forma breve las características de las órbitas de satélite de baja altura, altura intermedia y gran altura.

18-7. Explique qué son órbitas ecuatoriales, polares e inclinadas.

18-8. Describa las ventajas y las desventajas de los satélites geosíncronos.

18-9. Defina *ángulo de visión*, *ángulo de elevación* y *azimut*.

18-10. Defina *separación espacial de satélites* y sus restricciones.

18-11. Describa qué es una "huella".

18-12. Describa las distribuciones de radiación localizada, zonal y de cobertura terrestre.

18-13. Explique qué es *reuso*.

18-14. Describa en forma breve las características de funcionamiento de un modelo de enlace de subida, un transpondedor y un enlace de bajada, para un sistema satelital.

18-15. Defina la *pérdida por reducción* y su relación con el amplificador saturado y la potencia de transmisión.

18-16. Defina la *energía de bit*.

18-17. Defina la *potencia efectiva isotrópica irradiada*.

18-18. Defina la *temperatura equivalente de ruido*.

18-19. Defina la *densidad de ruido*.

18-20. Defina la *relación de portadora a densidad de ruido* y la *relación de energía de bit a densidad de ruido*.

18-21. Defina la *relación de ganancia a temperatura equivalente de ruido*.

18-22. Describa qué son los gastos de enlace de satélite y cómo se usan.

PROBLEMAS

18-1. Una estación terrestre está en Houston, Texas, y sus coordenadas son 99.5° de longitud y 29.5° de latitud norte. El satélite que interesa es *Satcom V*. Calcule los ángulos de visión para la antena de la estación terrestre.

18-2. Un sistema de satélite funciona con 14 GHz en enlace de subida y 11 GHz en enlace de bajada, y tiene una $P(e)$ proyectada de 10^{-7}. El esquema de modulación es 8-PSK, y el sistema manejará 20 Mbps. La temperatura equivalente de ruido del receptor es 400 K, y el ancho de banda de ruido del receptor es igual a la frecuencia mínima de Nyquist. Calcule los siguientes parámetros: Relación *C/N* teórica mínima, relación E_b/N_0 teórica mínima, densidad de ruido, ruido total en la entrada del receptor, potencia de portadora mínima en el receptor y la energía mínima por bit en la entrada del receptor.

18-3. Un sistema satelital funciona con 6 GHz en enlace de subida y 4 GHz en enlace de bajada, y tiene una $P(e)$ proyectada de 10^{-6}. El esquema de modulación es QPSK, y el sistema conducirá 100 Mbps. La temperatura equivalente de ruido es 290 K, y el ancho de banda de ruido del receptor es igual a la frecuencia mínima de Nyquist. Calcule la relación *C/N* que se medirá en un punto del receptor antes del BPF donde el ancho de banda es igual a a) 1½ por la frecuencia mínima de Nyquist y b) 3 × frecuencia mínima de Nyquist.

18-4. ¿Cuál de los siguientes sistemas tiene la mejor BER proyectada?

(a) 8-QAM, C/N = 15 dB, $B = 2f_N$, $f_b = 60$ Mbps.

(b) QPSK, C/N = 16 dB, $B = f_N$, $f_b = 40$ Mbps.

18-5. Un transmisor de estación terrestre tiene un HPA cuya potencia de salida saturada es 10,000 W. La relación de reducción es 6 dB, la pérdida por ramificación es 2 dB, la pérdida en alimentador es 4 dB y la ganancia de antena es 40 dB. Calcule la potencia real irradiada y la EIRP.

A los satélites de comunicaciones que funcionan en la banda C se les asigna un ancho total de banda de 500 MHz, simétrico en torno a la frecuencia central del satélite. A esto se le suele llamar un canal satelital, que se subdivide en canales de radio. La mayoría de los satélites de comunicaciones porta 12 transpondedores (pares de transmisor-receptor de canal de radio), cada uno con 36 MHz de ancho de banda. Las portadoras de los 12 transpondedores se multiplexan por división de frecuencia con una banda de protección de 4 MHz entre cada una de ellas, y una de protección de 10 MHz en ambos extremos del espectro asignado de frecuencia de 500 MHz.

Si los transpondedores adyacentes del espectro de 500 MHz se alimentan con una antena polarizada en cuadratura, la cantidad de transpondedores (canales de radio) disponibles en un satélite se puede duplicar, hasta 24. Doce transpondedores de número impar transmiten y reciben con una antena polarizada verticalmente, y 12 transpondedores de número par transmiten y reciben con una antena polarizada horizontalmente. Las frecuencias de portadora de los canales pares se desplazan 20 MHz respecto a las de los transpondedores impares, para reducir la diafonía entre los transpondedores adyacentes. A este método de asignar distintas polarizaciones electromagnéticas a canales adyacentes se llama *reutilización de frecuencia*, y es posible con la polarización ortogonal y distanciando 20 MHz los canales adyacentes. La reutilización de la frecuencia es una técnica para lograr mejor utilización del espectro disponible de frecuencias.

Satélite de comunicaciones Anik-E

Anik es una palabra esquimal que significa "hermanito". Los satélites de comunicaciones Anik-E son Domsat (satélites domésticos) operados por Telsat Canada. La fig. 19-2 muestra el plan de frecuencias y polarización del sistema satelital Anik-E. Un grupo de 12 canales de radio (grupo A) usa polarización horizontal, y otro grupo de 12 canales de radio (grupo B) usa polarización vertical, haciendo un total de 24 canales de radio, cada uno con 36 MHz de ancho de banda. Hay un ancho de banda de 4 MHz entre los canales adyacentes de radio, y uno de 10 MHz en cada extremo del espectro, haciendo un total de 500 MHz de ancho de banda del canal del satélite. Hay 12 canales de radio primarios y 12 canales de radio de reserva o de prioridad.

ACCESO MÚLTIPLE

El *acceso múltiple* al satélite (que a veces se llama *destino múltiple*) implica que más de un usuario tiene acceso a uno o más canales de radio (transpondedores) dentro de un canal de comunicaciones vía satélite. En forma característica, una empresa u operadora común renta los transpondedores con el fin de proporcionar transmisión de voz o de datos a una multitud de usuarios. El método por el cual se usa o *accede* el ancho de banda de un transpondedor depende del método de acceso múltiple que se use.

La fig. 19-3 muestra los tres *arreglos de acceso múltiple* de uso más común: acceso múltiple por división de frecuencia (FDMA, de *frequency-division multiple accessing*), acceso múltiple por división de tiempo (TDMA, de *time-division multiple accesing*) y acceso múltiple por división de código (CDMA, de *code-division multiple accessing*). En el FDMA, a las transmisiones de cada estación terrestre se les asignan bandas de frecuencia específicas de enlace de subida y de bajada, dentro del ancho de banda asignado al satélite; pueden ser *preasignadas* o *asignadas por demanda*. En consecuencia, las transmisiones FDMA se separan en el dominio de la frecuencia y, por consiguiente, deben compartir el ancho de banda total del transpondedor así como la potencia total del transpondedor. En el TDMA, cada estación transmite una corta ráfaga de información durante una ranura (*época*) específica de tiempo dentro de una trama TDMA. Las ráfagas se deben sincronizar de modo que la de cada estación llegue al satélite en un momento distinto. En consecuencia, las transmisiones TDMA están separadas en el dominio del tiempo, y en TDMA, todo el ancho de banda y la potencia del transpondedor se usan para cada transmisión, pero sólo durante un intervalo preestablecido de tiempo. En el sistema CDMA, todas las estaciones terrestres transmiten dentro de la misma banda de frecuencias y, para todo fin práctico, no tienen limitaciones de cuándo deben transmitir o en cuál frecuencia de portadora. Así, en CDMA, todo el ancho de banda del satélite transpondedor lo usan todas las estaciones en forma continua. La separación de señales se hace con técnicas de cifrado y descifrado de envolvente.

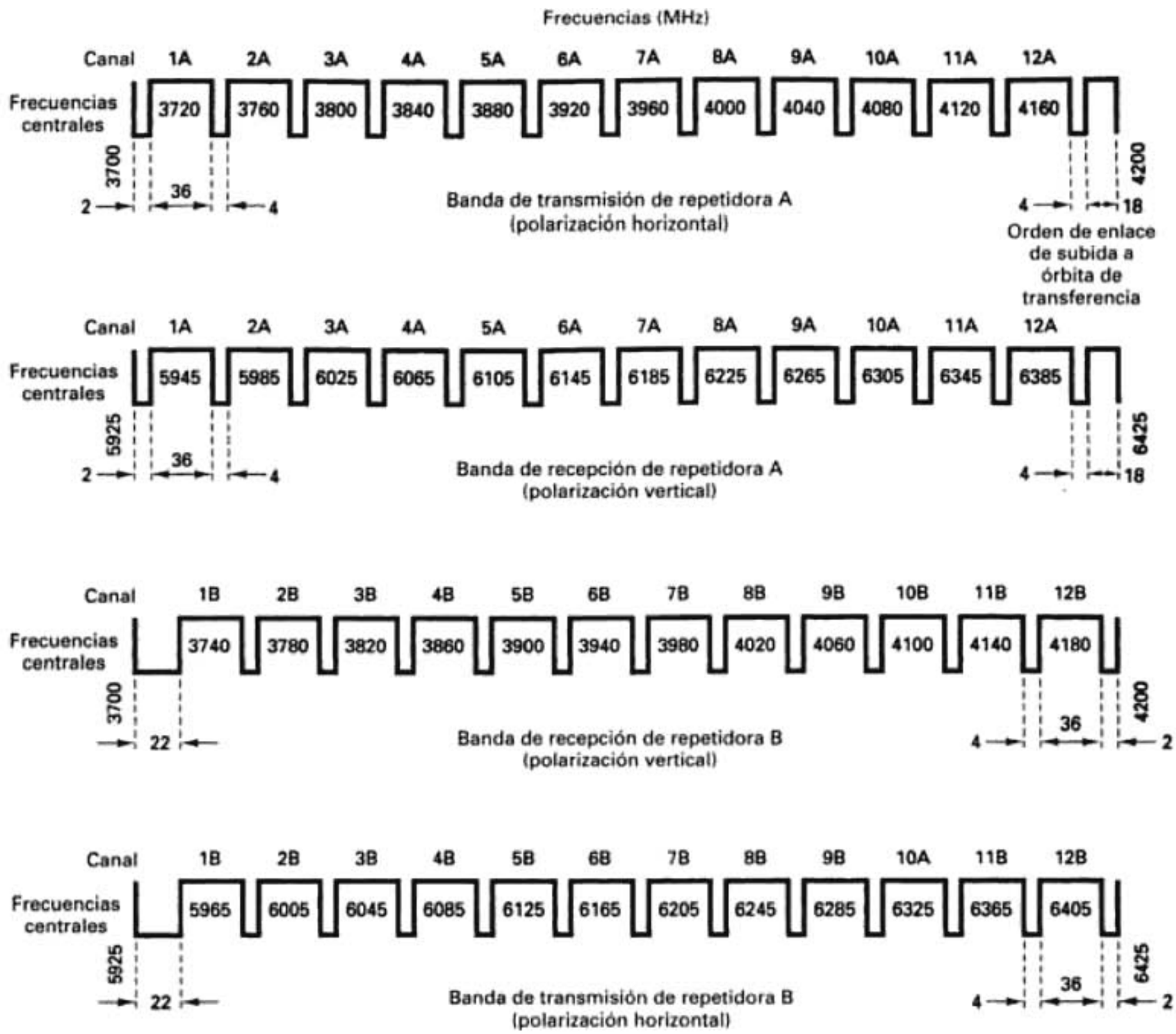


FIGURA 19-2 Plan de polarización y frecuencias del *Anik-E*

Acceso múltiple por división de frecuencia

El *acceso múltiple por división de frecuencia* (FDMA, de *frequency-division multiple access*) es un método de acceso múltiple en el que determinado ancho de banda de RF se divide en bandas menores de frecuencia, llamadas *subdivisiones*. Cada subdivisión tiene su propia FI de portadora. Se usa un mecanismo de control para asegurar que dos o más estaciones terrestres no transmitan en la misma subdivisión y al mismo tiempo. En esencia, el mecanismo de control designa una estación receptora para cada una de las subdivisiones. En sistemas de asignación por demanda, el mecanismo de control también se usa para establecer o terminar los enlaces de banda de voz entre las estaciones terrestres de origen y de destino. En consecuencia, cualquiera de las estaciones terrestres participantes puede usar cualquiera de las subdivisiones en cualquier momento. Si cada subdivisión sólo porta un canal de banda de voz de 4 kHz, a esto se le llama sistema de *un canal por portadora* (SCPC, de *single-channel per carrier*). Cuando se multiplexan varios canales de banda de voz por división de frecuencia, para formar una señal compuesta de banda base formada por grupos, supergrupos o hasta grupos maestros, se asigna una subdivisión más ancha. A esto se le llama *múltiples canales por portadora* (MCPC, de *multiple-channel per carrier*).

Las frecuencias y anchos de banda para los sistemas satelitales FDM/FM que usan formatos de varios canales por portadora se suelen asignar y permanecen fijos durante un largo tiempo.

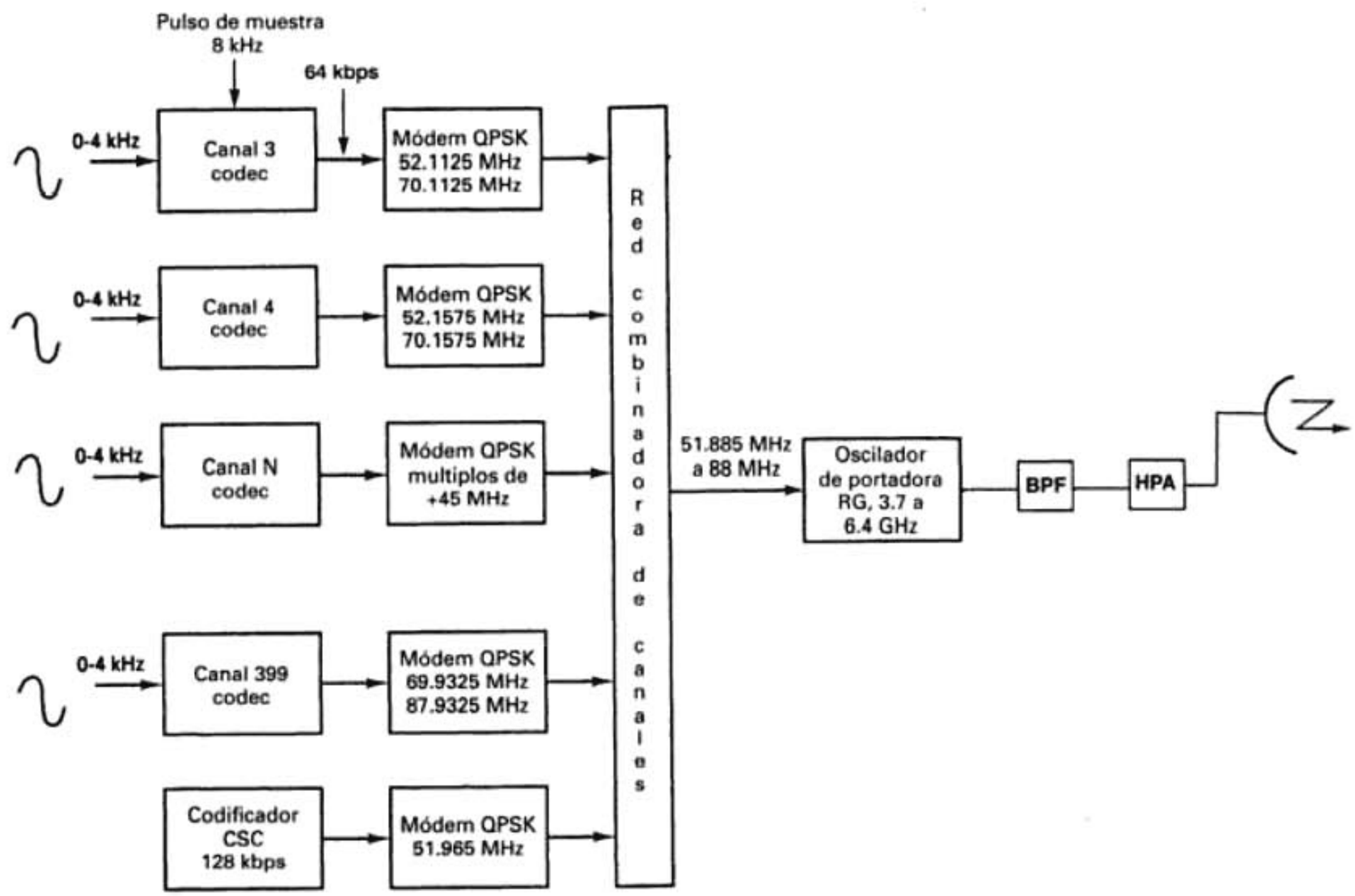


FIGURA 19-4 Estación terrestre transmisora FDMA-SPADE

se divide a la mitad, y se obtienen dos bandas de 400 canales, una baja y una alta. Para operación dúplex, se usan 400 canales para transmitir en una dirección y 400 para la dirección opuesta. También, los canales 1, 2 y 400 de cada banda permanecen siempre vacíos. Así se reduce la cantidad de canales útiles dúplex de banda de voz a 397. La banda C de 6 GHz va de 5.725 GHz a 6.425 GHz (700 MHz). Esto permite unos 36 canales de RF de 36 MHz por sistema. Cada canal de RF tiene una capacidad de 397 canales dúplex de banda de voz.

Cada canal de RF (fig. 19-5) tiene un *canal común de señalización* (CSC, de *common signaling channel*) de 160 kHz. El CSC es una transmisión multiplexada por división de tiempo, que se multiplexa por división de frecuencia en el espectro de FI abajo de los canales de banda de voz codificados por QPSK. La fig. 19-6 muestra la estructura de trama TDM para el CSC. El tiempo total de trama es 50 ms, que se subdivide en 50 ranuras de 1 ms. Cada estación terrestre transmite en el canal CSC sólo durante su ranura preasignada de tiempo de 1 ms. La señal CSC está en código binario de 128 bits. Para transmitir un código binario de 128 bits en 1 ms se requiere una velocidad de transmisión de 128 kbps. El código CSC se usa para establecer y desconectar enlaces de banda de voz entre dos estaciones terrestres usuarias, cuando se usa asignación de canal por demanda.

El canal CSC ocupa un ancho de banda de 160 kHz, que incluye los 45 kHz para el canal 1 de banda baja. En consecuencia, el canal CSC va de 51.885 a 52.045 MHz. El código binario QPSC del canal CSC de 28 kbps modula una portadora de 51.965 MHz. El ancho mínimo de banda requerido por el canal CSC es 64 kHz; esto da como consecuencia una banda de protección de 48 kHz en cada lado de la señal CSC.

En el FDMA, cada estación terrestre puede transmitir en forma simultánea dentro del mismo espectro de RF de 36 MHz, pero en dos canales distintos de voz. En consecuencia, las transmisiones simultáneas de canales de banda de voz de todas las estaciones terrestres de la red

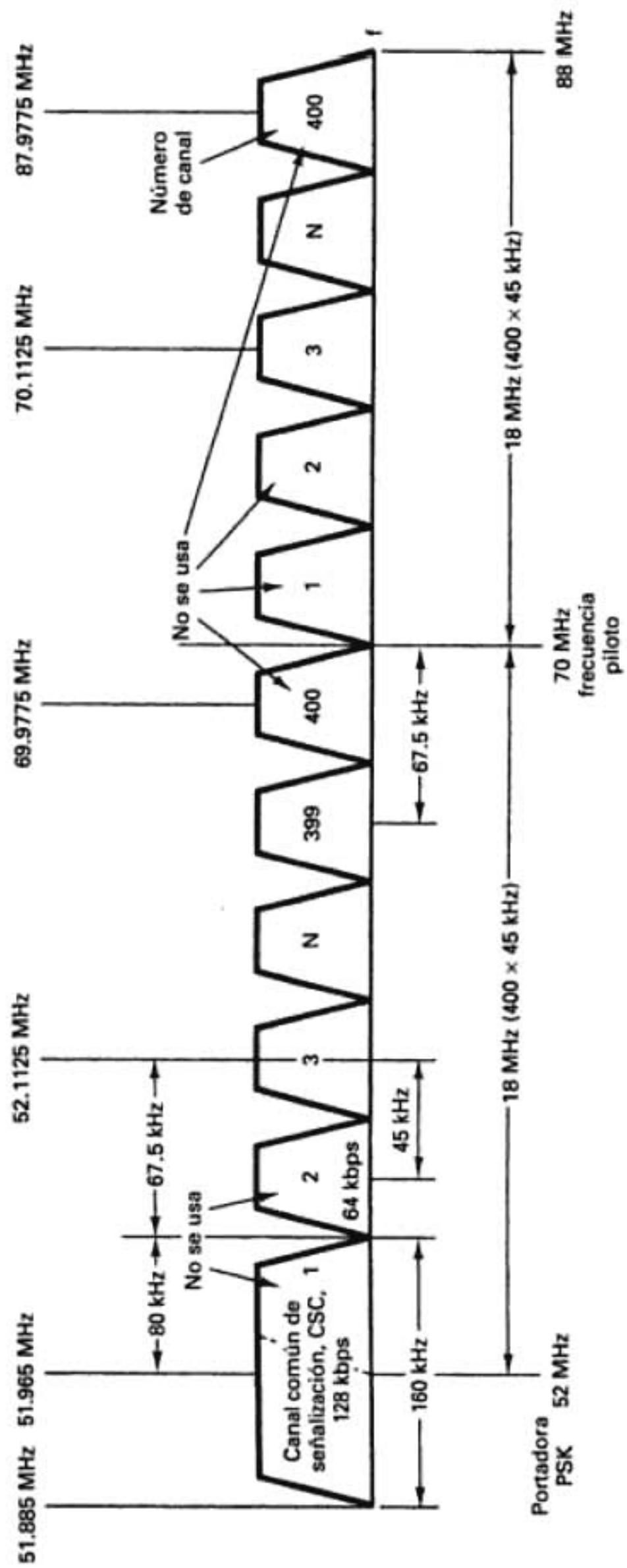


FIGURA 19-5 Asignaciones de frecuencia de portadora para el equipo de *Intelsat*, de asignación por demanda, acceso múltiple PCM y una portadora por canal (SPADE)

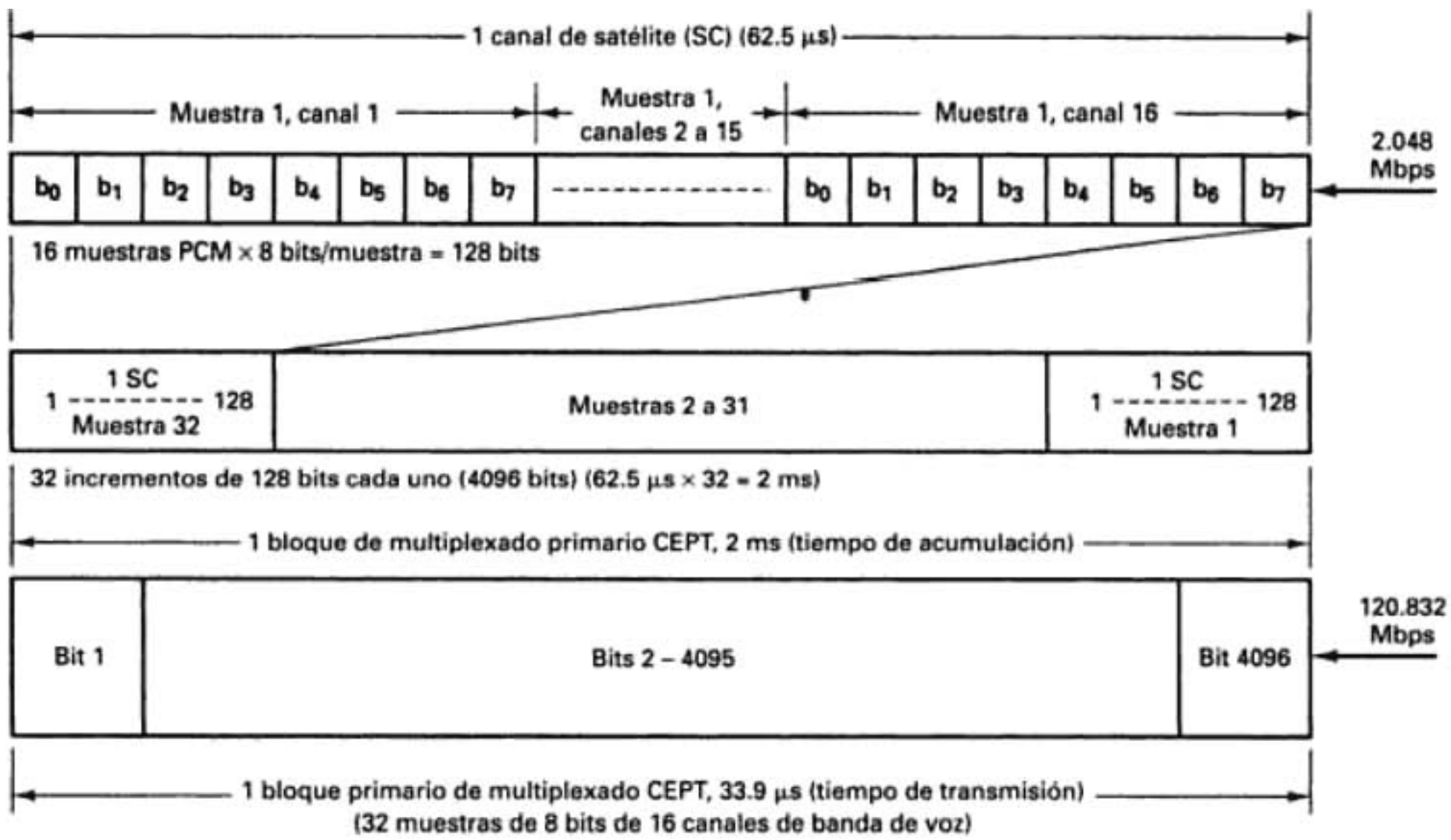


FIGURA 19-10 Trama multiplexada primaria CEPT, TDMA

periodo de 2 ms cada una de las estaciones terrestres participantes deben transmitir en momentos distintos, las transmisiones individuales de cada estación se deben hacer en tiempos bastante menores. En la trama CEPT se usa una frecuencia de transmisión de 120.832 Mbps. Esta velocidad es el quincuagésimo noveno múltiplo de 2.048 Mbps. En consecuencia, la transmisión real de los 4096 bits acumulados dura unos 33.9 μs. En los receptores de estación terrestre, los 4096 bits se guardan en un registro y se desplazan a una frecuencia de 2.048 Mbps. Como todas las frecuencias de reloj (500 Hz, 16 kHz, 128 kHz, 2.048 MHz y 120.832 MHz) están sincronizadas, los códigos PCM se acumulan, guardan, transmiten, reciben y decodifican en sincronía perfecta. Para los usuarios, parece que la transmisión de voz es un proceso continuo.

El sistema TDMA tiene varias ventajas sobre el FDMA. La primera, y quizá la más importante, es que en el TDMA sólo existe la portadora de una estación terrestre en el satélite transpondedor en cualquier momento, por lo que se reduce la distorsión por intermodulación. En segundo lugar, en FDMA cada estación debe poder transmitir y recibir en una multitud de frecuencias de portadora, para lograr funciones de acceso múltiple. En tercer lugar, el acceso TDMA se adapta mucho mejor a la transmisión de información digital que el FDMA. Las señales digitales se aclimatan con más naturalidad para su almacenamiento, conversiones de frecuencia y procesamiento en dominio del tiempo, comparadas con sus contrapartes analógicas.

La desventaja principal del TDMA respecto a FDMA, es que en el primero se requiere una sincronización precisa. Las transmisiones de cada estación terrestre se deben hacer durante una ranura exacta de tiempo. También, en TDMA se deben lograr y mantener las sincronizaciones de bits y de tramas.

Acceso múltiple por división de código

En el sistema FDMA, las estaciones terrestres se limitan a determinado ancho de banda dentro de un canal o sistema satelital, pero no tienen restricción en cuanto al momento de transmitir. En TDMA, las transmisiones de una estación terrestre se restringen a una ranura precisa de tiempo, pero no tienen restricción sobre la frecuencia o ancho de banda que pueden usar, dentro de una asignación de sistema o canal satelital especificada. En el *acceso múltiple por división de código* (CDMA, de *code-division multiple access*) no hay restricciones de tiempo ni de ancho de banda. Cada transmisor de estación terrestre puede transmitir cuando quiera, y puede usar cualquiera de las bandas o todo el ancho de banda asignado a un sistema o canal satelital determinado. Como no

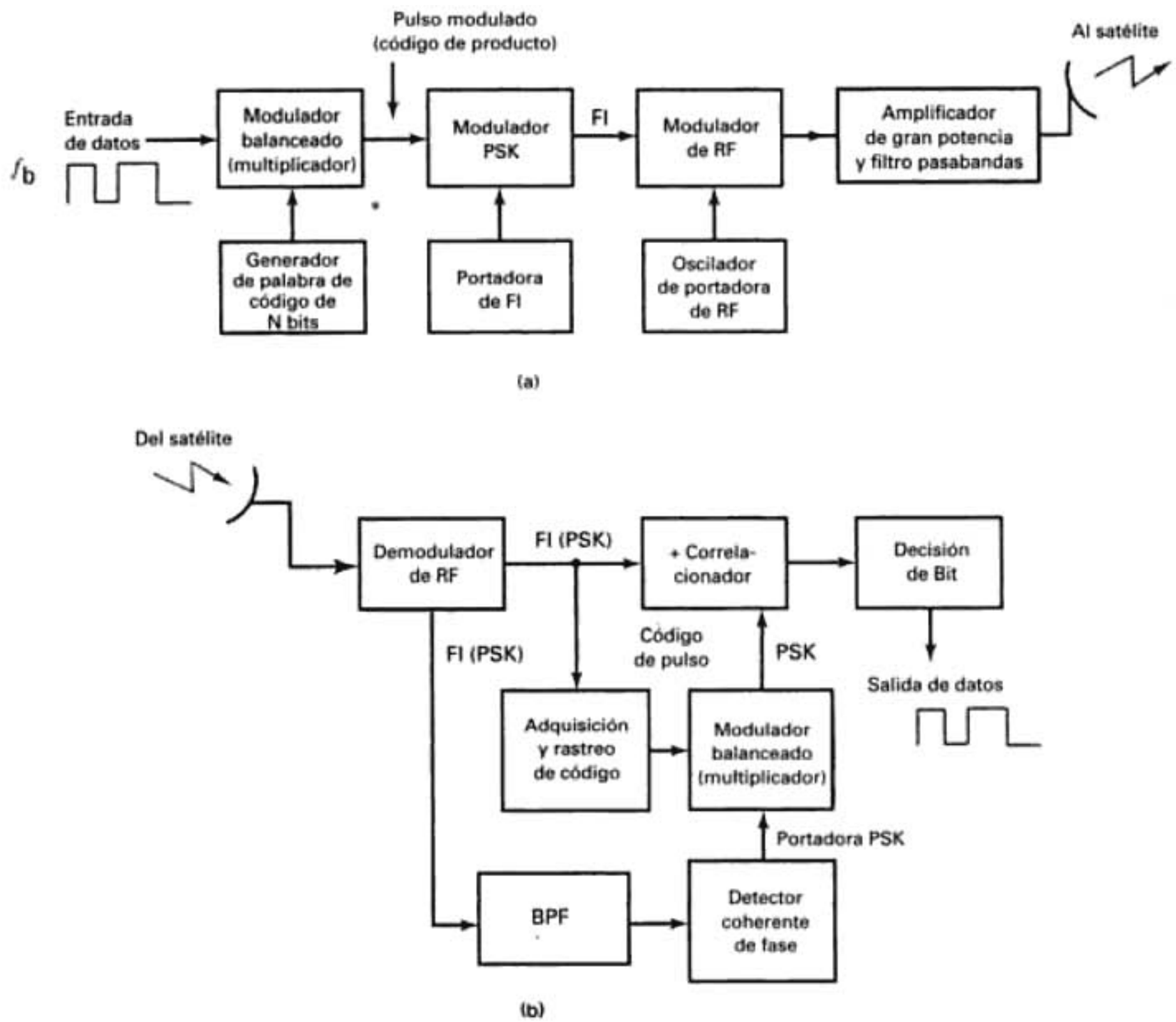


FIGURA 19-11 Acceso múltiple por división de código (CDMA): (a) codificador; (b) decodificador

hay limitación para el ancho de banda, a veces se llama *acceso múltiple por dispersión de espectro* al acceso CDMA; las transmisiones se pueden repartir por todo el ancho de banda asignado. Las transmisiones se separan mediante técnicas de cifrado y descifrado de envolvente. Esto es, las estaciones de cada estación terrestre se codifican con una palabra binaria única, llamada *código de pulso*. Cada estación tiene su propio código de pulso. Para recibir la transmisión de determinada estación terrestre, otra estación terrestre debe conocer el código de pulso de la primera.

La fig. 19-11 muestra el diagrama de bloques de un codificador y decodificador CDMA. En el codificador (fig. 19-11a), los datos de entrada, que pueden ser señales de banda de voz codificadas por PCM, o datos digitales en bruto, se multiplica por un código de pulso único. El código de producto modula en PSK una portadora de FI, que sufre una conversión elevadora hasta la RF de transmisión. En el receptor (fig. 19-11b), la RF tiene conversión reductora hasta la FI. De la FI se recupera una portadora PSK coherente. También, se adquiere y usa el código de pulso para sincronizar al generador de código de la estación receptora. Téngase en cuenta que la estación receptora conoce el código de pulso, pero debe generar un código de pulso que esté sincronizado en el tiempo con el código de recibir. El código sincrónico de pulso recuperado multiplica a la portadora PSK recuperada y genera una señal modulada PSK que contiene la portadora PSK y el código de pulso. La señal recibida en FI que contiene el código de pulso, la portadora PSK y la información de datos se compara con la señal de FI recibida, en el *correlacionador*. La función del correlacionador es comparar las dos señales y recuperar los datos originales. En esencia, el correlacionador resta la portadora PSK recuperada + el código de pulso, de la portadora PSK recibida + el código de pulso + los datos. El resultado son los datos.

siempre sucede. Para sistemas que no tienen transmisiones sincrónicas en el tiempo, se deben desarrollar códigos donde no haya correlación entre el código de una estación y cualquier fase de código de otra estación.

La diferencia principal entre los transmisores PSK de dispersión de espectro y otras clases de transmisores PSK es el modulador adicional donde se multiplica la palabra de código por los datos que entran. Debido a la naturaleza pseudoaleatoria de la palabra de código se le llama *ruido pseudoaleatorio* (PRN, de *pseudorandom noise*). Este ruido debe tener una propiedad de gran autocorrelación (consigo mismo) y una propiedad de baja correlación con los códigos pseudoaleatorios de otros transmisores. La rapidez de palabras de código (R_{cw}) debe ser mayor que la rapidez de datos que llegan (R_d) en varios órdenes de magnitud. Además, la rapidez del código debe ser estadísticamente independiente de la señal de datos. Cuando se satisfacen esas dos condiciones, el espectro final de señal de salida aumentará (se dispersará) en un factor llamado *ganancia de procesamiento*, que se expresa como sigue

$$G = \frac{R_{cw}}{R_d} \tag{19-1}$$

donde G es la ganancia de procesamiento, y $R_{cw} >> R_d$.

Una señal de dispersión de espectro no se puede demodular con exactitud si el receptor no posee un circuito que se acople con el generador de palabras de código en el transmisor. Tres de las técnicas más usadas para producir la función de dispersión son *secuencia directa, salto de frecuencia* y una combinación de las anteriores, llamada *híbrido de salto de frecuencia con secuencia directa* (DS/FH híbrido, de *direct-sequence frequency hopping*).

Secuencia directa (DS). La *dispersión de espectro por secuencia directa* (DS-SS, de *direct-sequence spread spectrum*) se produce cuando una señal bipolar modulada de datos se multiplica linealmente por la señal de dispersión en un modulador balanceado especial llamado *correlacionador de dispersión*. La rapidez de código de dispersión es $R_{cw} = 1/T_c$, donde T_c es la duración de un solo pulso bipolar (es decir, el *pulso*). Las frecuencias de pulso son de 100 a 1000 veces mayores que los datos del mensaje y, en consecuencia, los tiempos de pulso son de 100 a 1000 veces menores que el tiempo de un solo bit de datos. Por lo anterior, el espectro de frecuencias de salida transmitida con dispersión de espectro es de 100 a 1000 veces más ancho que el ancho de banda de la señal inicial modulada de datos PSK. El diagrama de bloques de un sistema de dispersión de espectro por secuencia directa se ve en la fig. 19-13. Como allí se ve, la fuente de datos modula en forma directa a la señal de portadora, que a continuación es modulada a su vez por la palabra en código de dispersión, en el correlacionador de dispersión.

Los códigos de dispersión (de pulso) que se usan en sistemas de dispersión de espectro pueden ser *códigos de secuencia de longitud máxima*, que a veces se llaman *códigos de secuencia m*, o *códigos dorados*. Los códigos dorados son combinaciones de códigos de longitud máxima inventados por Magnavox Corporation en 1967, especialmente para aplicaciones de acceso múltiple CDMA. Hay un conjunto relativamente grande de códigos dorados, disponible con correlación mínima entre códigos de pulso. Cuando hay una cantidad razonable de usuarios de satélite es imposible lograr códigos perfectamente ortogonales. Sólo se puede diseñar para tener una correlación cruzada mínima entre pulsos.

Una de las ventajas del CDMA era que todo el ancho de banda de un canal o sistema satelital se puede usar para cada transmisión desde cada estación terrestre. Para nuestro ejemplo, la rapidez de pulsos era seis veces la frecuencia original de bits. En consecuencia, la velocidad real de transmisión de la información sólo era la sexta parte de la frecuencia de modulación PSK, y el ancho de banda necesario es seis veces mayor que el que se requiere para sólo transmitir los datos originales en binario. A causa de la ineficiencia de codificación causada por transmitir pulsos en vez de bits, se contrarresta la ventaja de más ancho de banda y, por consiguiente, ya no es tanta ventaja. Además, si se debe sincronizar la transmisión de pulsos desde diversas estaciones, se requiere una sincronización perfecta para que trabaje el sistema. Por lo anterior, la desventaja de necesitar sincronización de tiempo en sistemas TDMA también existe con el CDMA. En resumen, el acceso CDMA no es todo lo que se decía de él. La ventaja más importante del CDMA es su inmunidad a la interferencia, que lo hace ideal para aplicaciones militares.

al que se usa en el sistema SPADE. La DSI es una forma de asignación por demanda; se asignan SC en forma aleatoria según se necesiten.

En DSI se ve que hay una *compresión de canal*; pueden haber más TC asignados que la cantidad de SC. En general, se usa una relación TC:SC de 2:1. Para un circuito de comunicaciones dúplex (dos sentidos simultáneos) hay voz en cada dirección 40% del tiempo, y el circuito está inactivo 20% del tiempo en ambas direcciones. Por lo anterior, se obtiene una ganancia DSI un poco mayor que 2. La ganancia DSI se ve afectada por un fenómeno llamado *recorte competitivo*. Es cuando se detecta energía de voz en un TC y no hay SC qué asignarle. Durante el tiempo de *espera*, se pierde la información de voz. El suscriptor no nota el recorte competitivo si su duración es menor que 50 ms.

Para aumentar todavía más la capacidad del canal se usa una técnica llamada *robo de bits*. Con ella se pueden agregar canales a sistemas totalmente cargados robando bits de los canales en uso. En general, se genera un canal de sobrecarga robando el bit menos significativo de otros 7 canales del satélite. El robo de bits da como resultado 8 canales con resolución de 7 bits, mientras se usa el *canal de sobrecarga*. En consecuencia, el robo de bits da como resultado una SQR menor que la normal.

Interpolación de voz por asignación de tiempo

La *interpolación de voz por asignación de tiempo* (TASI, de *time-assignment speech interpolation*) es una forma de compresión analógica de canal que se ha usado durante muchos años en cables submarinos. TASI se parece mucho a DSI, pero las señales interpoladas son analógicas, y no digitales. En TASI también se usa una relación de compresión de 2:1. TASI fue el primer método que se usó para desorganizar señales de voz por seguridad militar. Se parece a una red de paquetes de datos; el mensaje de voz se recorta en pequeños segmentos, hechos de sonidos o de partes de sonidos. Los sonidos se mandan por la red en forma de paquetes separados de energía, y a continuación se vuelven a armar en el receptor, para reconformar el mensaje original de voz.

RADIONAVEGACIÓN POR SATÉLITE

Se puede definir a la *navegación* como el arte o la ciencia de trazar, definir o dirigir el curso de movimientos; en otras palabras, es *conocer dónde está uno, y poder encontrar el camino*. El método más antiguo y rudimentario de navegación es el de *vagabundear*, que no es más que continuar viajando hasta llegar al destino, suponiendo, naturalmente, que haya destino. Todd Ferguson, mi buen amigo y viajero mundial, dijo una vez "el viaje verdadero no tiene destino". El vagabundeo es la técnica de navegación tan frecuente que usan muchos alumnos durante su primera semana de clases en todos los colegios y universidades. Es probable que el primer método efectivo o útil de navegación haya sido la *navegación celeste*. En ella, se determinan la dirección y la distancia mediante el seguimiento de cuerpos celestes, incluyendo a las estrellas y a la Luna, cronometrados con gran precisión. Es una técnica primitiva que data de hace miles de años. Una desventaja obvia de la navegación celeste es que funciona mejor por la noche y de preferencia con cielo despejado.

Otro método bastante rudimentario de navegación es el *pilotaje*, que es fijar posición y dirección con respecto a señas importantes, como vías de ferrocarril, tanques elevados, graneros, picos de montaña y cuerpos de agua. El pilotaje deriva su nombre de los primeros pilotos de aviones, que usaron este método de navegación.

La *navegación a estima* es una técnica donde se determina la posición extrapolando una serie de incrementos de velocidad medidos. En inglés se llama "*dead* (o "*ded*") *reckoning*", y lo de "ded" se deriva de "deducido". Charles Lindbergh usó con bastante éxito la navegación a estima en 1927, durante su histórico viaje trasatlántico de 33 horas, y Amelia Earhart lo usó con bastante fracaso en 1937, durante su intento de volar por primera vez alrededor del mundo.

Aunque cada uno de los métodos de navegación que se han descrito hasta ahora tienen su lugar en el tiempo, sin duda la técnica de navegación más exacta hasta la fecha es la *radionavegación*, o *navegación electrónica*. En ella, se determina la posición midiendo el tiempo de recorrido

de una onda electromagnética cuando va de un transmisor a un receptor. Hay más o menos 100 tipos distintos de sistemas domésticos de radionavegación que se usan en la actualidad. Algunos usan transmisores terrestres y otros usan transmisores de satélite. Entre los sistemas de radionavegación más exactos y útiles se incluyen los siguientes

Decca (emisión en superficie terrestre)
Omega (emisión en superficie terrestre)*
Loran (emisión en superficie terrestre)
Sistema de posicionamiento global *Navy Transit* (emisión de satélite en órbita baja)*
Sistema de posicionamiento global *Navstar* (emisión de satélite en órbita intermedia)*

Los sistemas de posicionamiento global (GPS, de *global positioning system*) Loran y Navstar son los dos de radionavegación que más se usan hoy.

Navegación Loran

Hasta hace poco el *Loran* (de *Long Range Navigation*, navegación a larga distancia) era el método más efectivo, confiable y exacto de radionavegación. El Loran A se desarrolló durante la Segunda Guerra Mundial, y el Loran C, su versión más reciente, salió en 1980. Hoy se usa principalmente en los aviones y barcos recreativos.

En el Loran, los receptores adquieren señales codificadas especialmente de dos pares de transmisores de alta potencia, en tierra, cuyas ubicaciones se conocen con precisión. Se mide el tiempo transcurrido entre la recepción de las señales codificadas, y en el receptor se convierten en distancia usando la velocidad de propagación de las ondas electromagnéticas. Al aplicar la geometría básica y la relación entre la distancia d, la velocidad v y el tiempo t (que es $d = vt$), se puede determinar la ubicación del receptor con gran exactitud. Sólo hay un conjunto de coordenadas que posee determinada relación de tiempo (distancia) según tres fuentes.

El Loran es tan preciso como la precisión con que se puedan medir los tiempos de transmisión de las señales codificadas. Los errores del sistema se deben principalmente a problemas de propagación, como el hecho de que la superficie terrestre no es lisa, ni es perfectamente redonda. Las condiciones atmosféricas y las múltiples trayectorias de transmisión también pueden afectar el funcionamiento del Loran. Sin embargo, es probable que la desventaja más notable del Loran sea que no da una cobertura global. Los transmisores en tierra sólo se pueden instalar en tierra, y la tierra es una parte relativamente pequeña de la superficie de la Tierra. En consecuencia, hay lugares donde simplemente no se pueden recibir las señales de Loran (puntos muertos). Sin embargo, no es importante lo bueno o lo malo que haya sido el Loran, porque ha emergido una técnica más nueva y mejor de radionavegación, llamada *GPS Navstar*, que usa transmisores basados en satélites.

GPS Navstar

Navstar es acrónimo de *Navigation System with Time And Ranging*, sistema de navegación con tiempo (horas) y telemetría, y *GPS* es la abreviatura de *Global Positioning System*, sistema de posicionamiento global. GPS Navstar es el sistema más reciente y exacto de radionavegación del que se dispone. Está basado en satélites y es abierto, lo cual significa que está a la disposición de quienquiera que tenga un receptor GPS. El Departamento de Defensa de Estados Unidos desarrolló Navstar para tener información continua y muy precisa sobre posición, velocidad y horas para usuarios en tierra, mar, aire y espacio. En esencia, GPS Navstar es un sistema de navegación basado en espacio, de posicionamiento tridimensional y de distribución de tiempo. El propósito del sistema es usar una combinación de estaciones terrestres, satélites en órbita y receptores especiales, para proporcionar casi a todos posibilidades de navegación, en cualquier momento y en cualquier lugar del mundo, independientemente de las condiciones del clima. El sistema satelital Navstar se terminó en 1994, y es mantenido por la Fuerza Aérea de E. U. A.

*Da cobertura global

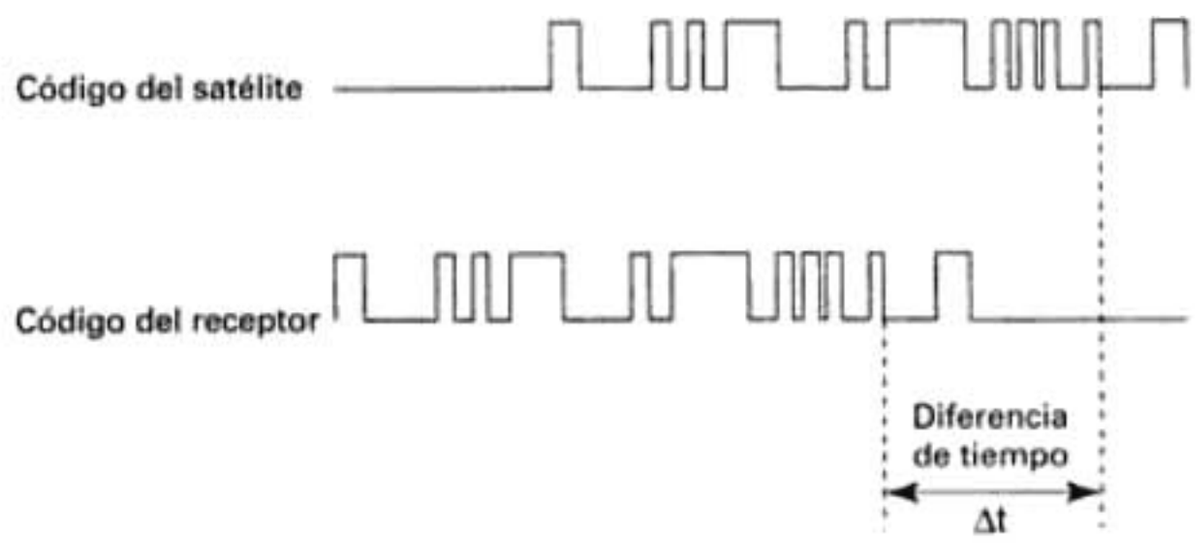


FIGURA 19-18 Código pseudoaleatorio de tiempo GPS

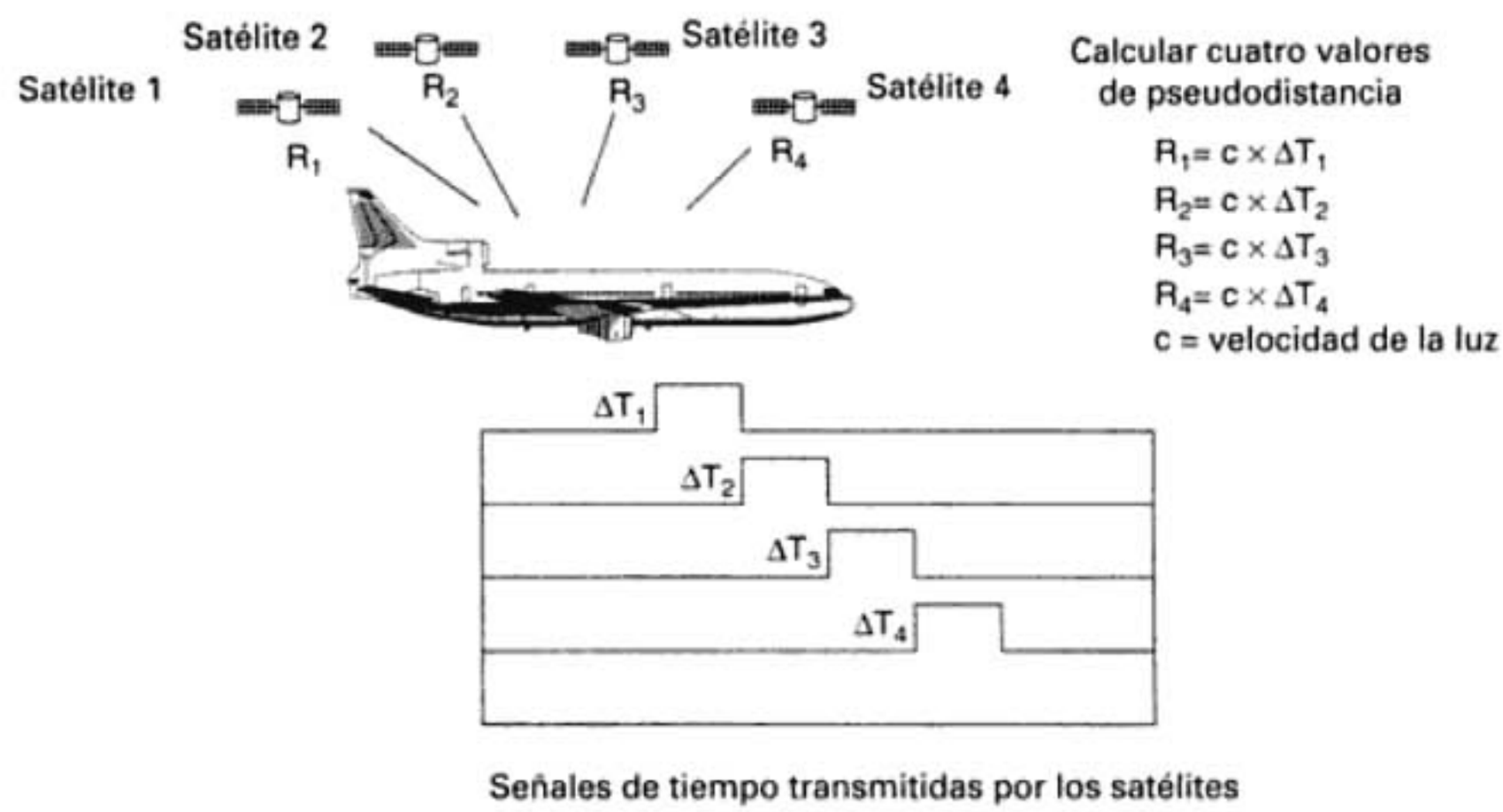


FIGURA 19-19 Solución de telemetría GPS

Claro está que el chiste es determinar con exactitud cuándo salió la señal de sincronización del satélite. Para determinarlo, el transmisor del satélite y el receptor de la estación terrestre producen códigos idénticos de sincronización (pseudoaleatorios) exactamente al mismo tiempo, como se ve en la fig. 19-18. Cada satélite transmite en forma continua su código de sincronización precisa. Después de haber recibido un código de sincronización, un receptor sólo lo compara con su propio código, producido en forma local, para determinar el tiempo de propagación. La diferencia de tiempo multiplicada por la velocidad de la señal de radio determina la distancia al satélite.

La fig. 19-19 ilustra cómo puede un avión determinar la distancia a la que se encuentra respecto a cuatro satélites distintos, tan sólo midiendo los tiempos de propagación, o de retardo, y multiplicándolos por la velocidad de la luz. También aquí se pueden usar ecuaciones simultáneas para determinar la longitud y la latitud del avión.

Para que un receptor en tierra determine su longitud y latitud, debe recibir señales de tres o más satélites que identifiquen el número del satélite o su código pseudoaleatorio (PRN) de sincronización, y la localización de cada satélite. La ubicación de un satélite se describe con un sistema tridimensional de coordenadas en relación con el centro de la Tierra, como se ve en la fig. 19-20. El centro de la Tierra es el punto de referencia, y sus coordenadas son 0, 0, 0. Así, cada satélite tiene sus coordenadas X_s, Y_s, Z_s que determinan su lugar con respecto al geocentro. Sin

Subtrama #	Una subtrama = 300 bits, 6 segundos			
1	TLM	HOW	Datos de corrección del satélite	30 segundos
2	TLM	HOW	Datos de efemérides del satélite (I)	
3	TLM	HOW	Datos de efemérides del satélite (II)	
4	TLM	HOW	Otros datos del sistema	
5	TLM	HOW	Datos de almanaque para todos los satélites	

TLM – palabra de telemetría

Preámbulo, 8 bits	Datos, 16	Paridad, 6

HOW – palabra de transición

Hora de la semana, 17 bits	Datos, 7	Paridad, 6

FIGURA 19-24 Formato de trama de datos de navegación

para transmitir distintas páginas de datos del sistema, incluyendo datos de almanaque para todos los sistemas. El mensaje total de navegación está formado de 25 tramas (125 subtramas), que se manda durante un periodo de 12.5 min.

Segmento de control. El segmento de control Navstar, llamado *sistema de control de operación* (OCS, de *operational control system*) incluye todas las *estaciones monitoras* terrestres fijas, ubicadas en todo el mundo, una *Estación Maestra de Control* (MCS, de *Master Control Station*) y transmisores de enlace de subida. Hay estaciones monitoras pasivas en California, Hawaii, Alaska, Isla de Ascensión (frente a África Occidental), Diego García (Océano Índico), Kwajalein (Océano Pacífico), entre otras. Las estaciones monitor no son más que receptores GPS que rastrean los satélites cuando pasan sobre ellas, y acumulan datos de telemetría y efemérides (orbitales) de ellos. Esta información se transmite a la Estación de control maestro en la Base Falcon de la Fuerza Aérea, a 12 millas al este de Colorado Springs, Colorado, E.U.A., donde se procesa y determina si la posición real del satélite se compara con su posición calculada por GPS. La Estación de control maestro está administrada por el 2do. Escuadrón de operaciones espaciales de la Fuerza Aérea de E.U.A.

La MCS recibe datos de las estaciones monitoras en tiempo real, 24 horas por día, y con esa información determina si los satélites sufren cambios de reloj o de efemérides, y detecta el mal funcionamiento del equipo. Se calcula la nueva información de navegación y efemérides a partir de las señales monitoreadas, y se carga en los satélites una o dos veces al día. La información calculada por la MCS, junto con órdenes de mantenimiento rutinario, se mandan a los satélites a través de antenas terrestres de enlace de subida. Las instalaciones de antena transmiten a los satélites por un radioenlace de *banda S*. Además de su función principal, la MCS mantiene las 24 horas un sistema computarizado de tabla de boletines, con las últimas novedades y estados del sistema.

Segmento del usuario. El segmento del usuario de GPS consiste en todos los receptores de GPS y la comunidad de usuarios. Los receptores GPS convierten las señales recibidas de los vehículos espaciales en estimaciones de posición, velocidad y tiempo. Para calcular las cuatro dimensiones de x, y y z (posición) y t (tiempo), se requieren cuatro satélites. Los receptores GPS se usan para navegación, posicionamiento, diseminación de la hora, cartografía, sistemas

CAPÍT

Imagen protegida por derechos de autor

Servicio telefónico móvil

INTRODUCCIÓN

Aunque los conceptos básicos de las comunicaciones radiotelefónicas en dos sentidos son bastante sencillos, el *servicio telefónico móvil* ha evolucionado hasta llegar a ser un monstruo de las comunicaciones, que es bastante complicado en su conjunto. Los teléfonos móviles implican redes de comunicaciones intrincadas y algo complejas, formadas con metodologías tanto analógicas como digitales, complicados centros y procedimientos de conmutación controlada por computadora, y varios métodos de acceso múltiple. El objeto de este capítulo es presentar los conceptos fundamentales del servicio telefónico móvil y presentar al lector varios sistemas de uso actual para proporcionar el servicio, incluyendo los *servicios* estandar *de teléfonía celular* (CTS, de *cellular telephone services*), *sistemas de comunicación personal* (PCS, de *personal communications satellite systems*) y *sistemas satelitales de comunicación personal* (PCSS, de *personal communications satellite systems*). El servicio telefónico móvil comenzó a principios de la década de 1940, pero a causa de su gran costo, disponibilidad limitada y angosta asignación de frecuencias, no se usaron mucho los primeros sistemas. Sin embargo, en años recientes los avances tecnológicos aunados a un espectro más ancho de frecuencias, mayor disponibilidad y mejor confiabilidad han provocado un aumento fenomenal en los deseos de las personas de hablar por teléfono, en cualquier lugar y a cualquier momento, independientemente de que sea productivo o necesario.

La estación móvil suele ser, en la actualidad, un aparato manual pequeño. En la terminología de radio de antes, el término *móvil* sugería cualquier radiotransmisor, receptor o transceptor que podía cambiar de lugar mientras funcionaba. Por otro lado, el término *portátil* describía un transceptor de radio relativamente pequeño manual, de baterías y que una persona caminando a paso normal podía llevar. Sin embargo, la definición actual de *móvil* ha llegado a indicar moverse a gran velocidad, como en un bote o un automóvil, o a baja velocidad, como en la bolsa de una persona a pie. Por consiguiente, la definición moderna e incluyente de *teléfono móvil* significa cualquier radioteléfono capaz de funcionar mientras está en movimiento a cualquier velocidad, que es de baterías y que es lo bastante pequeño como para que una persona lo pueda llevar.

El teléfono móvil es parecido al radio móvil de dos sentidos, porque la mayor parte de la comunicación se efectúa entre una estación de base y una estación móvil, o entre dos o más

Imagen protegida por derechos de autor

FIGURA 20-2 Retícula de células hexagonales sobrepuesta sobre un área metropolitana

De acuerdo con las ecuaciones 20-2a y b, la capacidad de canales de un sistema telefónico celular es directamente proporcional a la cantidad de veces que se duplica o reproduce un grupo en determinada área de servicio. Al factor N se le llama *tamaño de grupo*, y suele ser de 3, 7 o 12. Cuando se reduce el tamaño de grupo manteniendo constante el tamaño de la célula, se requieren más unidades de asignación para cubrir determinada área y, por consiguiente, la capacidad es mayor. El factor de reutilización de frecuencia de un sistema telefónico celular es inversamente proporcional a la cantidad de células en un grupo, es decir, a $1/N$. Por consiguiente, a cada célula dentro de un grupo se le asigna una $1/N$-ésima parte de los canales totales disponibles en el grupo.

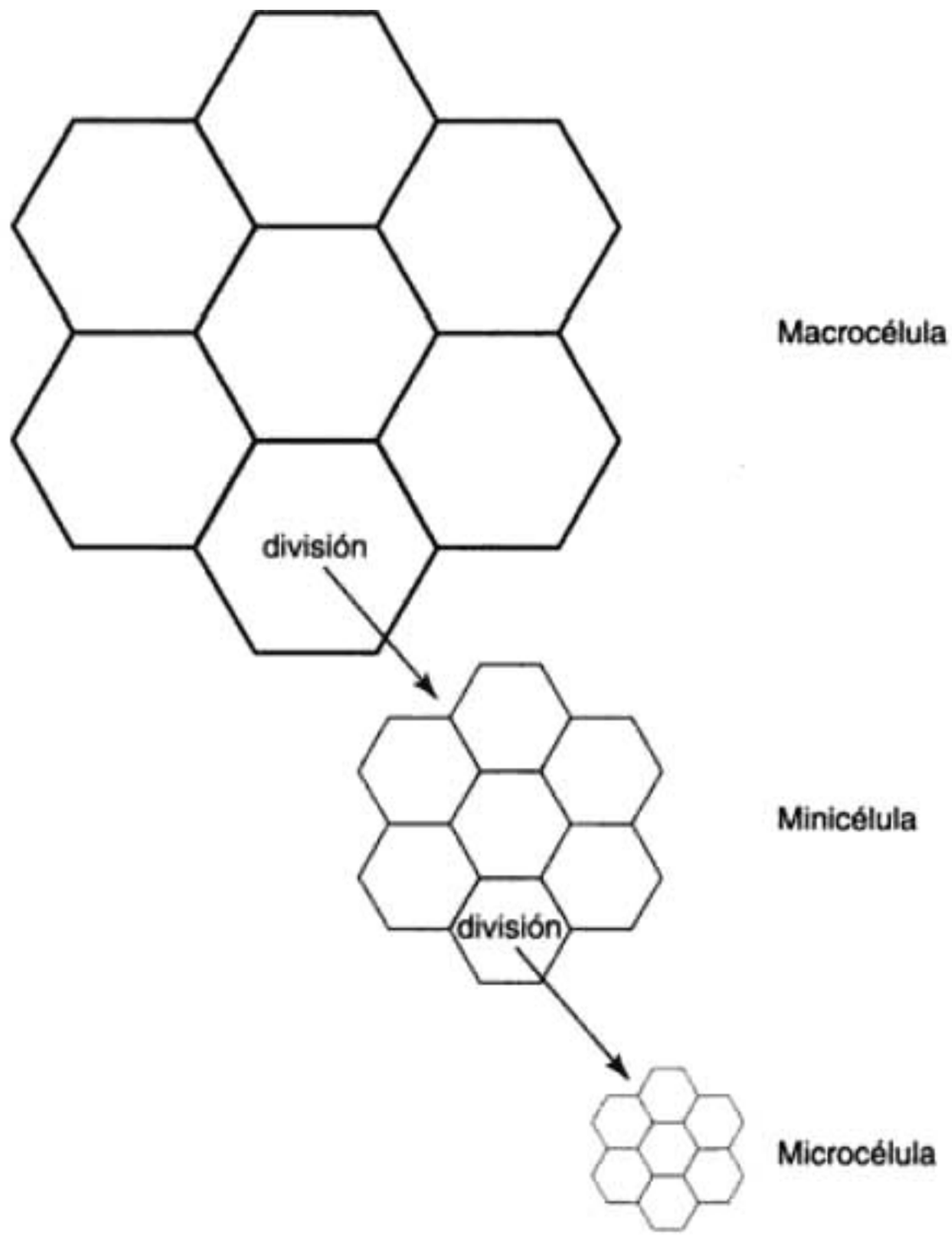


FIGURA 20-5 División de célula

Por consiguiente, la división de célula disminuye el radio celular a la vez que mantiene la misma relación de reutilización de canal compartido (*D/R*).

Sectorización

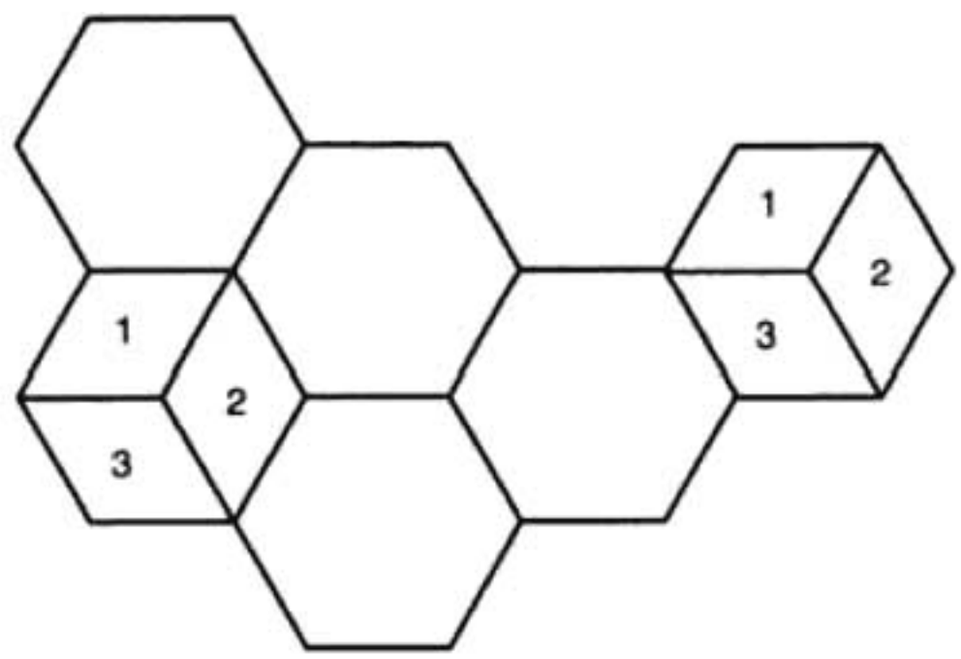
Otro método para aumentar la capacidad de canales es disminuir la relación *D/R* manteniendo al mismo tiempo el mismo radio de célula. Se puede lograr el mejoramiento de la capacidad si se reduce la cantidad de células en un grupo, aumentando así la reutilización de frecuencia. Para lograr lo anterior se debe reducir la interferencia relativa sin disminuir la potencia de transmisión.

En un sistema telefónico celular se puede disminuir la interferencia por canal compartido si se reemplaza una sola antena omnidireccional con varias antenas direccionales, irradiando cada una hacia un área específica. A estas áreas se les conoce como sectores, y a la disminución de la interferencia por canal compartido aumentando al mismo tiempo su capacidad mediante antenas direccionales se le llama *sectorización*. El grado de reducción de la interferencia por canal compartido depende de la cantidad de sectorización que se use. En el caso normal, una célula se divide ya sea en tres sectores de 120°, o en seis de 60°, como se ve en la fig. 20-6.

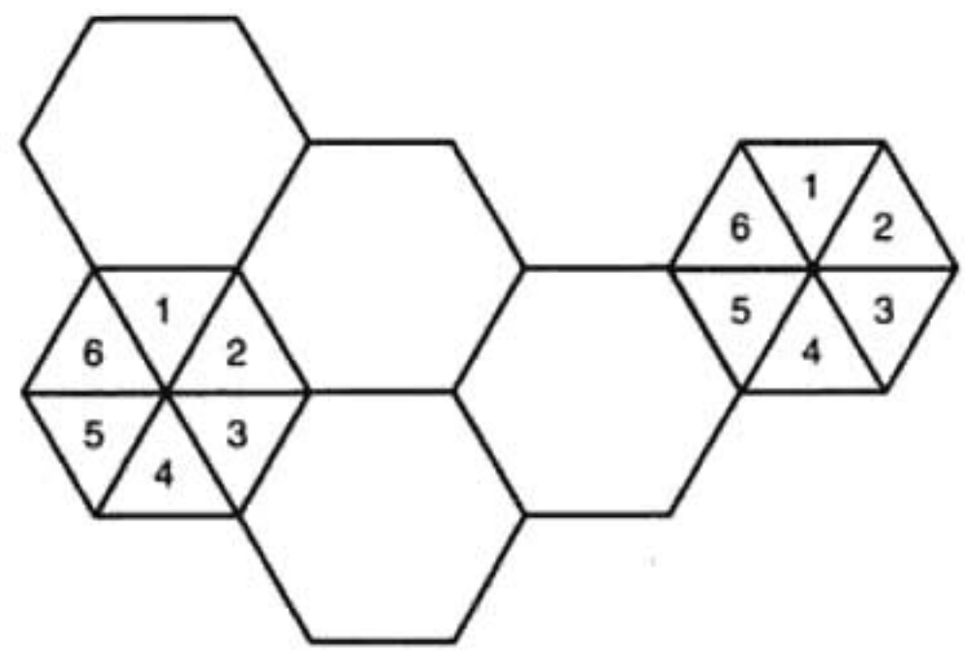
Cuando se usa la sectorización, los canales que se usan en determinado sector se dividen en grupos sectorizados que sólo se usan dentro de determinado sector. En la reutilización de siete células y sectores de 120°, la cantidad de células que se interfieren en la hilera más cercana se reduce de seis a dos. La sectorización mejora la relación de señal a interferencia y, con ello, aumenta la capacidad.

Segmentación y dualización

La *segmentación* y la *dualización* son técnicas que se incorporan cuando se requieren más células dentro de la distancia de reutilización. En la segmentación se divide un grupo de canales en



(a)



(b)

FIGURA 20-6 Sectorización: (a) sectores de 120 grados; (b) sectores de 60 grados

agrupamientos o segmentos de frecuencias mutuamente excluyentes, y a los sitios de célula, que están dentro de la distancia de reutilización, se les asigna su propio segmento del grupo de canales. La segmentación es un método para evitar la interferencia por canal compartido, aunque baja la capacidad de una célula al permitir la reutilización dentro de la distancia de reutilización, lo cual se suele prohibir.

La dualización es un método para evitar la división completa de célula, cuando de otro modo se necesitaría segmentar toda el área en células más pequeñas. Cuando se establece una nueva célula que requiere el mismo grupo de canales que una célula existente (célula 1) y una segunda célula (célula 2) no está a la distancia suficiente de la célula 1 para la reutilización normal, la parte ocupada de la célula 1 (el centro) se convierte en una célula primaria, y se pueden asignar las mismas frecuencias de canal a la nueva célula competidora (célula 2). Si se necesita usar todos los canales disponibles en la célula 2 surgiría un problema, porque la mayor célula secundaria en la célula 1 usa algunos de ellos, y habría interferencia. Sin embargo, en la práctica se asignan distintos canales a las células, por lo que en general esto no es un problema. Un inconveniente de la dualización es que requiere una estación base adicional, en la mitad de la célula 1. Ahora se tienen dos estaciones base en la célula 1: una de gran potencia que abarca toda la célula secundaria, y una de baja potencia que cubre la célula primaria, más pequeña.

Estructura del sistema de células

La fig. 20-1 muestra un sistema telefónico celular simplificado, donde se indican todos los componentes básicos necesarios en las comunicaciones por teléfono celular. Hay una red de radio que cubre un conjunto de áreas geográficas (células), en cuyo interior se pueden comunicar unidades

tenga la mayor intensidad de señal, y se sincroniza con los datos de control transmitidos por el controlador de sitio. La unidad móvil interpreta los datos y continúa vigilando el o los canales de control. Esta unidad barre en forma periódica los canales, para asegurarse que usa el mejor.

En un sistema celular, las llamadas se pueden hacer entre un teléfono de cable y uno móvil, o entre dos teléfonos móviles.

Llamadas de teléfono por cable a móvil. El centro de conmutación del sistema celular recibe una llamada de un teléfono por cable, a través de una línea dedicada de interconexión de la red telefónica pública. La central traslada los dígitos de marcación recibidos y determina si la unidad a la que se destina la llamada está desocupada u ocupada. Si está disponible, la central *localiza* al suscriptor móvil. A continuación de una *respuesta de voceo* de la unidad móvil, la central asigna un canal inactivo e instruye a la unidad móvil para que se sintonice a ese canal. La unidad móvil manda una verificación de sintonización de canal a través del controlador de sitio, que responde con un *tono de progreso de llamada* al teléfono móvil del suscriptor, haciéndolo que suene. La central termina los tonos de progreso de llamada cuando recibe una indicación positiva de que el suscriptor ha contestado el teléfono, y que ha comenzado la conversación entre las dos partes.

Llamada de teléfono móvil a teléfono por cable. Un suscriptor móvil que desee llamar a un teléfono por cable teclea primero al número que llama en la memoria de la unidad, usando botones de tono, o un disco en el teléfono. El suscriptor oprime entonces una *tecla de mandar*, que transmite el número llamado, así como el número de identificación del suscriptor móvil, a la central. Si el número de identificación es válido, la central conduce la llamada por una interconexión de línea de cable rentada, a la red telefónica pública, que completa la conexión al teléfono por cable llamado. La central, usando el controlador de sitio, asigna a la unidad móvil un canal desocupado de usuario y la instruye para sintonizarse a ese canal. Después que la central recibe la verificación de que la unidad móvil está sintonizada al canal asignado, el suscriptor móvil recibe un *tono de llamada en progreso*, audible, de la central. Cuando la parte llamada descuelga, la central termina los tonos de llamada y puede comenzar la conversación.

Llamadas de móvil a móvil. También se pueden hacer llamadas entre dos unidades móviles en el sistema de radio celular. Para iniciar una llamada a otra unidad móvil, la parte que llama ingresa el número a que llama en la memoria de la unidad, mediante el teclado, y a continuación oprime el botón de mandar. La central recibe el número de identificación de quien llama, y el número llamado, y a continuación determina si la unidad llamada está desocupada para recibir una llamada. La central manda una *orden de voceo* a todos los controladores de sitio, y la parte llamada, que puede estar en cualquier lugar del área de cobertura, recibe esa señal. Después de una localización positiva por la parte llamada, la central asigna a cada parte un canal desocupado de usuario, y las instruye para sintonizarse a su canal respectivo de usuario. A continuación suena el teléfono llamado. Cuando el sistema recibe el aviso de que la parte llamada ha contestado el teléfono, la central interrumpe el tono de llamada en progreso y puede comenzar la conversación entre dos unidades móviles.

Si un suscriptor móvil desea iniciar una llamada, y están ocupados todos los canales de usuario, la central manda una *orden dirigida de reintentar*, que instruye al suscriptor para reintentar la llamada por una célula vecina. Si el sistema no puede asignar un canal de usuario a través de una célula vecina, la central transmite un *mensaje de intercepción* a la unidad móvil que llama, por el canal de control. Siempre que la parte llamada esté *descolgada*, la parte que llama recibe una señal de ocupado. También, si el número llamado es inválido, el sistema manda un mensaje grabado a través del canal de control, o anuncia que no puede procesar la llamada.

Función de transferencia de llamada. Una de las propiedades más importantes de un sistema celular es su capacidad de transferir llamadas que ya están efectuándose, de uno a otro controlador de sitio a medida que las unidades móviles pasan de una a otra célula de la red. A

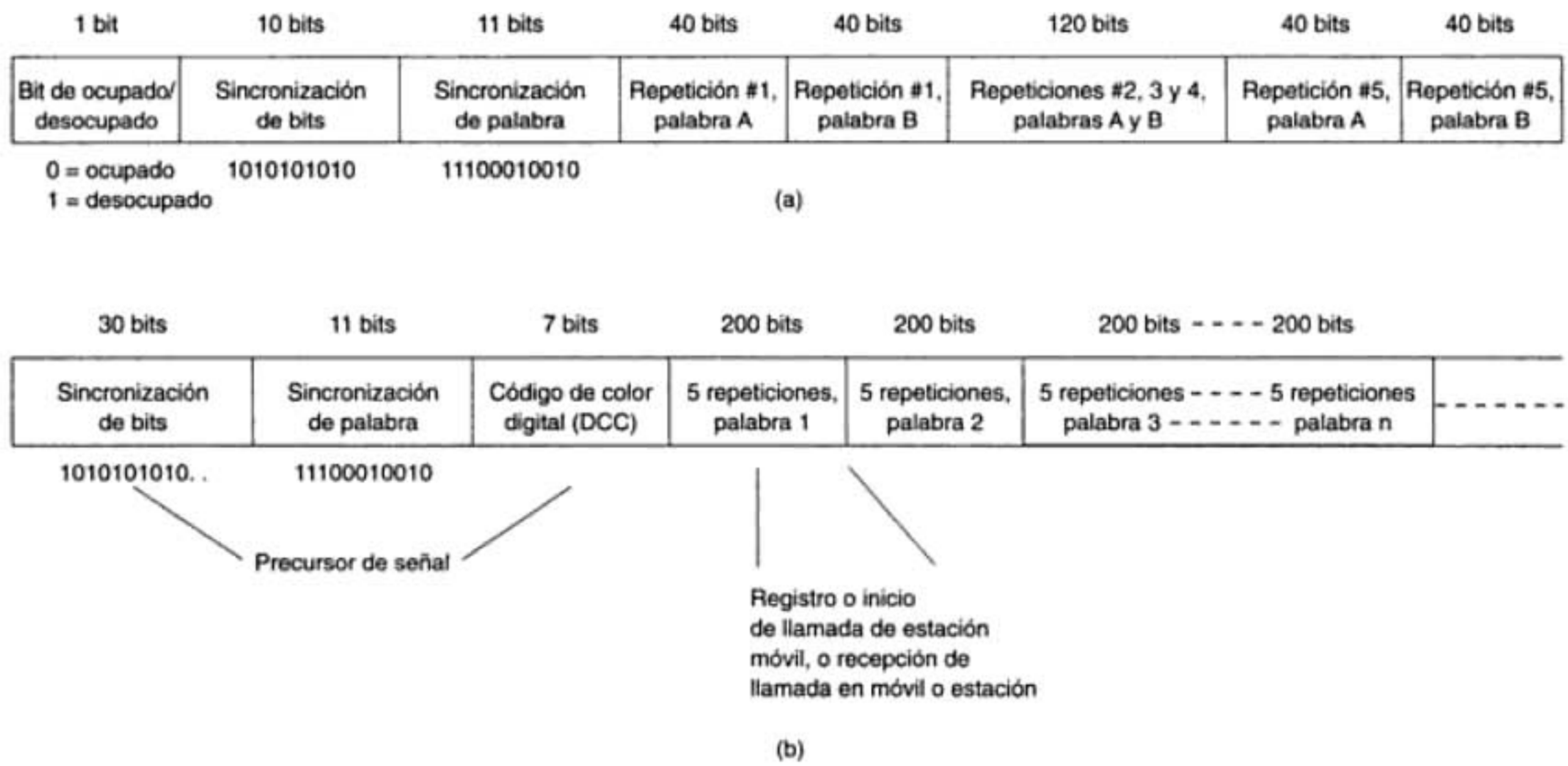


FIGURA 20-8 Formato de canal de control: (a) canal de control en sentido directo; (b) canal de control en sentido inverso

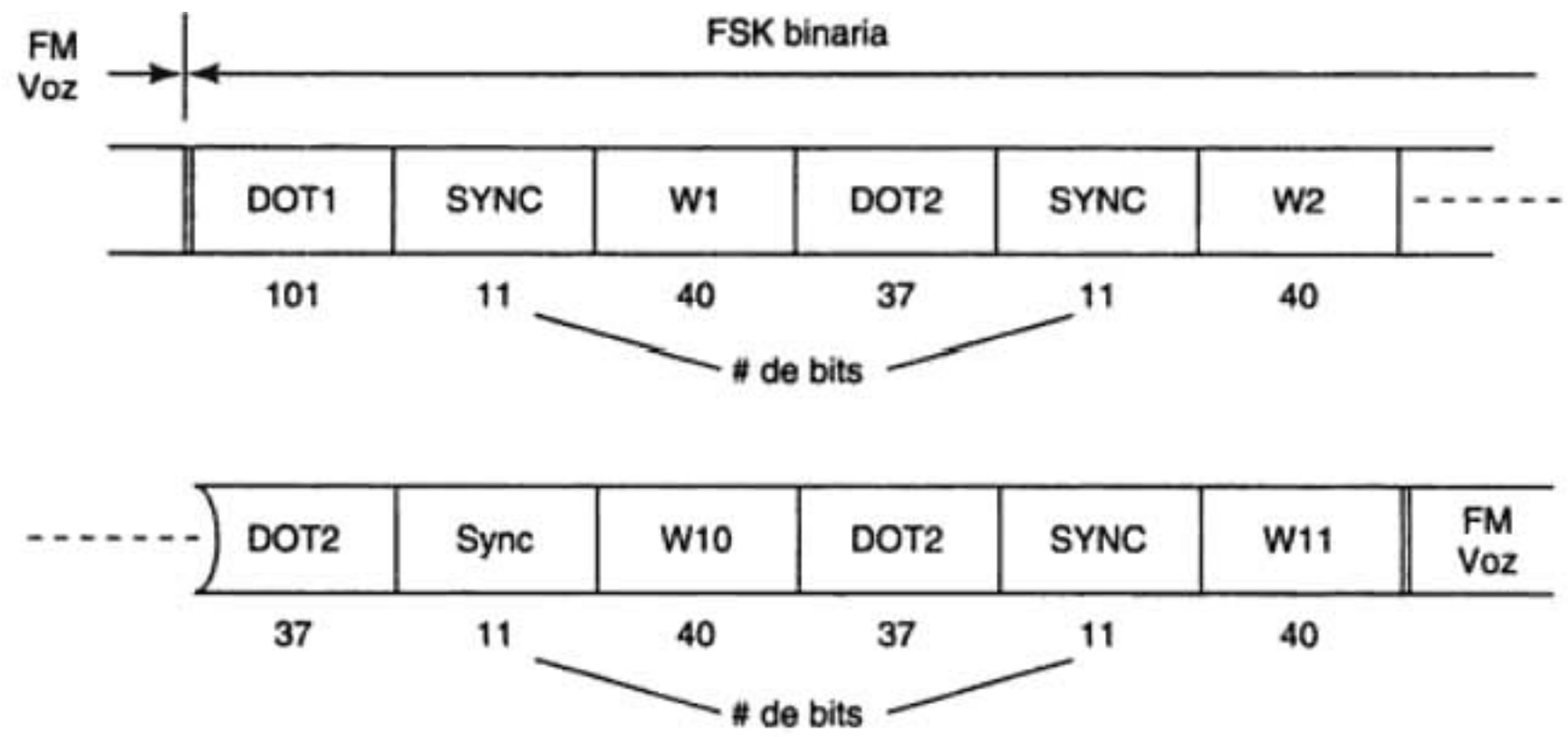
La fig. 20-8b muestra el formato del canal de control en sentido inverso, que se transmite de la unidad móvil a la estación base. Los datos de control se transmiten a una velocidad de 10 kbps, e incluyen *respuestas de voceo*, *peticiones de acceso* y *peticiones de registro*. Todos los mensajes RCC comienzan con el precursor de enganche RCC, formado por una *secuencia de puntos* de 30 bits, una *palabra de sincronización* de 11 bits y el *código de color digital* (DCC, de *digital color code*) que se agrega para que el control de canal no se confunda con un canal de control de una célula no adyacente que esté reutilizando la misma frecuencia. El teléfono móvil lee el DCC de la estación base y a continuación regresa una versión codificada del mismo, para verificar que la unidad está amarrada a la señal correcta. Cuando se termina la llamada, se transmite una *señal de señalización de tiempo fuera* de 1.8 segundos. Cada palabra de mensaje contiene 40 bits, y se repite cinco veces para hacer un total de 200 bits.

Señalización de canal de voz

Los canales celulares analógicos llevan información tanto de voz, con FM, como digital, usando FSK binaria. Al transmitir información digital de señalización, se inhiben las transmisiones de voz. A esto se le llama *silencio y ráfaga*: la voz se silencia y se transmiten los datos en una ráfaga corta. La frecuencia de bits de la información digital es 10 kbps. La fig. 20-9a muestra el formato de señalización de canal de voz para un canal de voz en sentido directo, y la fig. 20-9b muestra el formato para el canal de sentido inverso. La secuencia digital de señalización comienza con una secuencia de 101 bits de puntos que prepara al receptor a recibir información digital. Después de la secuencia de puntos se manda una palabra de sincronización para indicar el inicio del mensaje. En el canal de voz en sentido directo los mensajes digitales de señalización se repiten 11 veces, para asegurar su integridad, y en el canal de recepción se repiten 5 veces. El canal de sentido directo usa palabras de 40 bits, y el canal de sentido inverso usa palabras de 48 bits.

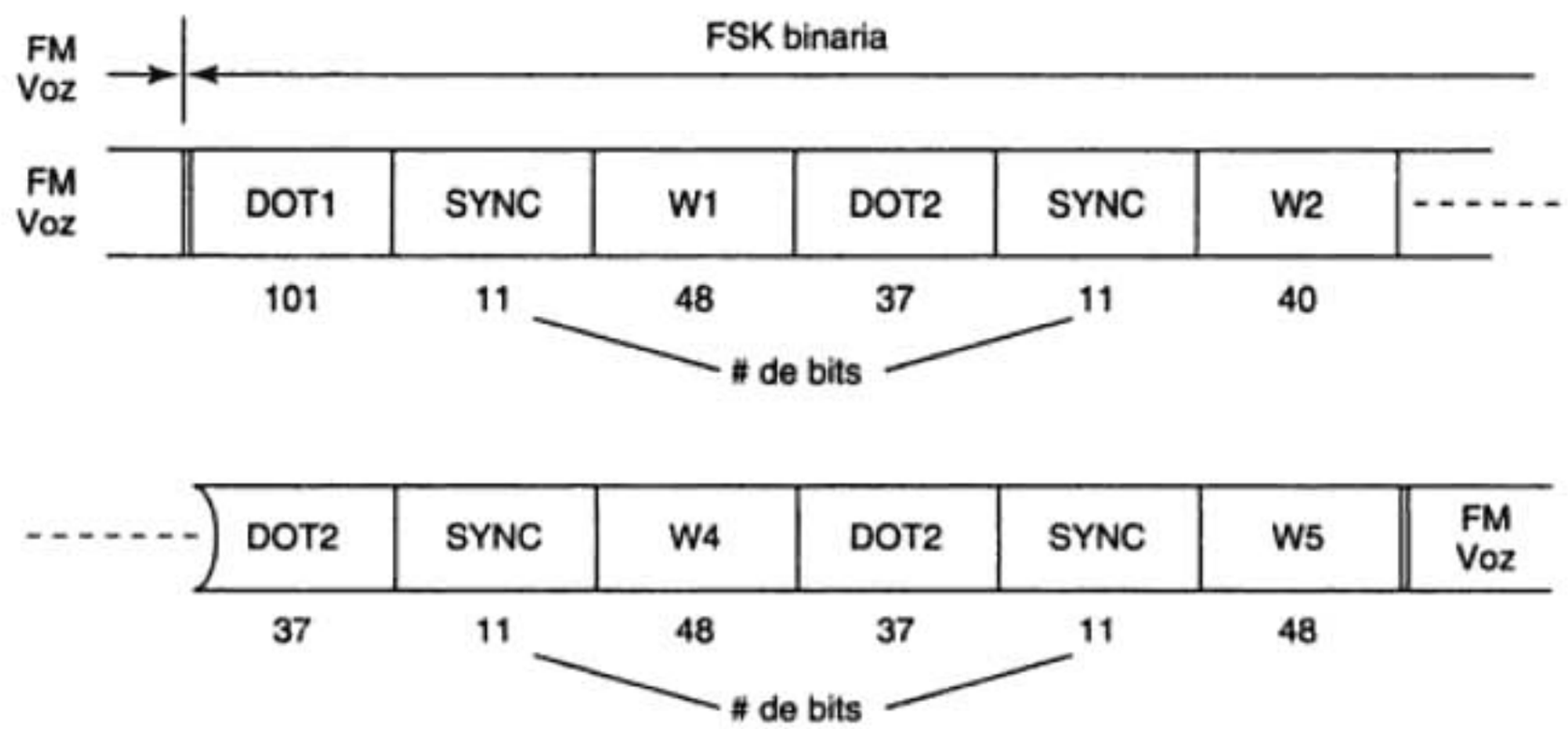
Tonos de audio de supervisión y señalización

Mientras se están usando los canales de voz hay dos procedimientos adicionales que se usan para mantener la supervisión entre la unidad móvil y la estación base: *tono de audio de supervisión* (SAT, de *supervisory audio tone*) y el *tono de señalización* (ST, de *signaling tone*). Los SAT se usan para asegurar la confiabilidad de las comunicaciones de voz. Los transmite la estación base y la unidad móvil los manda de regreso. Si el SAT coincide con el SAT transmitido, la unidad



DOT1 = Secuencia de puntos de 101 bits
DOT2 = Secuencia de puntos de 37 bits
SYNC = Palabra de sincronización
WN = Palabra de mensaje (N)
N = Cantidad de palabras de mensaje repetidas

(a)



DOT1 = Secuencia de puntos de 101 bits
DOT2 = Secuencia de puntos de 37 bits
SYNC = Palabra de sincronización
WN = Palabra de mensaje (N)
N = Cantidad de palabras de mensaje repetidas

(b)

FIGURA 20-9 Formato de canal de voz: (a) canal de sentido directo; (b) canal de sentido inverso

móvil sabe que está sintonizada a la estación base correcta. Las señales SAT son tonos analógicos de 5970, 6000 o 6030 Hz. La estación base transmite en forma continua uno de los tres SAT en cada canal activo de voz. Los SAT se sobreponen a la señal de voz en los enlaces de sentido directo e inverso. En las células que contienen más de una estación base, los SAT identifican el lugar de la estación base específica para determinado canal, y son asignados por el centro de conmutación para cada llamada.

El tono de señalización es una ráfaga de datos de 10 kbps, que inicia una terminación de llamada por parte de la unidad móvil. El tono de señal de *mensaje de fin de llamada* consiste en

El receptor de la fig. 20-10 es de tipo superheterodino de doble conversión. Usa inyección por el lado de alta y un circuito de oscilador local formado por un sintetizador de frecuencia de un solo chip con un preescalador y filtro de lazo a bordo, y un VCO (oscilador controlado por voltaje) de 915 a 937 MHz (45 MHz arriba de la RF de llegada). La primera FI (45 MHz) es suficientemente alta como para impulsar la frecuencia de imagen mucho más allá de la banda de paso del filtro preamplificador, y la segunda FI (455 kHz) es lo bastante baja como para amplificarse lo suficiente y activar el detector de FM. En el transmisor, la señal moduladora se aplica en forma directa al VCO del transmisor para producir una señal de portadora directa de FM de relativamente baja frecuencia, que se multiplica y después se amplifica en forma apreciable, antes de llegar a la antena de transmisión.

Nótese que el microprocesador tiene líneas de control a cada bloque funcional principal del transistor. El microprocesador determina las frecuencias en las salidas del sintetizador de frecuencia. También vigila la salida del detector de FM para determinar la calidad de la señal recibida, y para detectar la información recibida de preparación. También, el microprocesador proporciona información de preparación al transmisor y controla la ganancia del amplificador final de potencia.

TELÉFONO CELULAR DIGITAL

Las empresas telefónicas encaran el problema de una base de clientes que se amplía con rapidez, y al mismo tiempo el espectro asignado de frecuencias permanece sin cambios. Como se describió, la capacidad de usuarios se puede ampliar subdividiendo las células existentes en subcélulas menores (es decir, por división de células) y modificando la distribución de la radiación de las antenas (es decir, por sectorización). Sin embargo, el grado de subdivisión y redireccionamiento está limitado por la complejidad y la cantidad de sobrecarga necesaria para procesar las transferencias de llamada entre células. Otra restricción grave es el costo de compra de terrenos para sitios de célula en las áreas de máxima densidad de tráfico.

AMPS fue un sistema telefónico celular analógico de primera generación que no estaba diseñado para manejar las demandas de gran capacidad del mundo moderno, en especial en áreas metropolitanas de gran densidad. A finales de la década de 1980 varios de los fabricantes principales de equipo celular determinaron que los sistemas de teléfonos celulares digitales podrían permitir mejoras sustanciales, tanto en capacidad como en funcionamiento. En consecuencia, se diseñó el sistema *Celular Digital de Estados Unidos* (USDC, de *United States Digital Cellular*), que se desarrolló para tratar de sostener una mayor densidad de usuarios dentro de un espectro fijo de frecuencias.

El sistema USDC es compatible con el esquema anterior de asignación de frecuencias AMPS, y se diseñó para usar las mismas frecuencias de portadora, el mismo plan de reutilización de frecuencias y las mismas estaciones base. En consecuencia, las estaciones base y las unidades móviles se pueden equipar con canales tanto AMPS como USDC dentro del mismo equipo telefónico. Para manejar ambos sistemas, las portadoras celulares pueden proporcionar teléfonos USDC a los nuevos clientes, y al mismo tiempo pueden proporcionar servicio a los clientes actuales con teléfonos AMPS. El sistema USDC mantiene la compatibilidad con el AMPS de varias maneras. Por esta causa, USDC también se conoce como AMPS digital (D-AMPS).

En USDC se usa acceso múltiple por división de tiempo (TDMA) que, como el FDMA divide el espectro total de radiofrecuencias disponibles en canales individuales. Sin embargo, en TDMA se sigue dividiendo cada canal de radio en ranuras de tiempo, una para cada usuario. En los sistemas FDMA se asigna un canal a un suscriptor durante su llamada. Sin embargo, en TDMA los suscriptores con unidades móviles sólo pueden "mantener" un canal mientras lo están usando. Durante las pausas u otras interrupciones normales de una conversación, otras unidades móviles pueden usar el canal. Esta técnica de canales de *tiempo compartido* aumenta en forma importante la capacidad de un sistema, permitiendo que lo usen más suscriptores con unidades móviles en determinada área geográfica.

Las ventajas de los sistemas digitales TDMA de acceso múltiple sobre los FDMA normales son las siguientes:

1. El intercalamiento de muestras en el dominio del tiempo permite un aumento de la cantidad de suscriptores que usan un solo canal, al triple. El tiempo compartido se realiza

el sistema AMPS. La CDVCC es un mensaje de 12 bits transmitido en cada ranura de tiempo, y consiste en un número de código de color de voz digital de 8 bits, de 1 a 255, seguido por 4 bits adicionales de codificación obtenidos de un código de Hamming acortado. La estación base transmite un número CDVCC en el canal de voz de sentido directo, y toda unidad móvil que usa el canal TDMA debe recibir, decodificar y retransmitir el mismo código CDVCC (de reconocimiento) de regreso a la estación base, por el canal de voz de sentido inverso. Si los dos valores CDVCC no son iguales, se cede la ranura de tiempo a otros usuarios, y el transmisor de la unidad móvil se apaga en forma automática.

Canal lento para control asociado. Este SACCH es un canal de señalización para transmitir mensajes de control y supervisión entre la unidad digital móvil y la estación base, mientras la unidad móvil está ocupada con una llamada. El SACCH usa 12 bits codificados por ráfaga TDMA, y se transmite en cada ranura de tiempo, proporcionando así un canal de señalización en paralelo con la información digitalizada de voz. Por consiguiente, los mensajes SACCH se pueden transmitir sin interferir con el procesamiento de señales digitalizadas de voz. Como el SACCH sólo consiste en 12 bits por trama, se pueden necesitar hasta 22 tramas para transmitir un solo mensaje SACCH. Este canal conduce diversa información de control y de supervisión entre la unidad móvil y la estación base, por ejemplo, los cambios de nivel de potencia de comunicación y peticiones de transferencia. También se usa el SACCH en la unidad móvil para informar mediciones de intensidad de señal de las estaciones base vecinas para que, cuando sea necesario, la estación base pueda iniciar una *transferencia móvil asistida* (MAHO, de *mobile-assisted handoff*).

Canal rápido para control asociado. El FACCH es un segundo canal de señalización para transmitir mensajes de control y supervisión especializada, y de tráfico, entre la estación base y las unidades móviles. A diferencia del CDVCC y el SACCH, este último no tiene una ranura dedicada de tiempo. Es una transmisión del tipo de *silencio y ráfaga* que, cuando se transmite, sustituye la información digitalizada de voz por mensajes de control y supervisión dentro de una ranura de tiempo del suscriptor. No hay límite para la cantidad de tramas de voz que se pueden reemplazar con datos FACCH. Sin embargo, la información digitalizada de voz se protege en cierto modo, evitando que toda una transmisión de voz digitalizada sea sustituida por datos FACCH. La velocidad neta de transmisión digitalizada de voz de 13 kbps no se puede reducir a menos de 3250 bps en determinada ranura de tiempo. No hay campos dentro de una ranura normal de tiempo, que la identifiquen como voz digitalizada o como mensaje FACCH. Para determinar si se recibe un mensaje FACCH, la unidad móvil debe tratar de decodificar los datos como voz. Si decodifica en forma errónea, decodifica los datos como mensaje FACCH. Si el carácter de redundancia cíclica (CRC) se calcula en forma correcta, se supone que el mensaje es de FACCH. Este canal soporta transmisiones de *tonos de toque de frecuencia múltiple y tono dual* (DSTMF, de *dual-tone multiple-frequency*), pide instrucciones de liberar, instrucciones instantáneas de colgar y transferencia móvil asistida, o peticiones de estado de unidad móvil. Los datos FACCH están empaquetados e intercalados para ajustarse a una ranura de tiempo, en forma parecida a como se maneja la voz digitalizada.

Codificación de voz

La fig. 20-12 muestra el diagrama de bloques de un codificador de voz USDC de canal digital de voz. El control de errores de canal para los datos digitalizados de voz usa tres mecanismos para minimizar esos errores: 1) se usa una frecuencia igual a la mitad del código convencional para proteger los bits más vulnerables de la corriente de datos del codificador de voz; 2) los datos transmitidos se intercalan para cada trama de codificador de voz en dos ranuras de tiempo, para reducir los efectos del desvanecimiento de Rayleigh, y 3) se ejecuta una comprobación de redundancia cíclica en los bits más importantes, desde el punto de vista de percepción, de los datos digitalizados de voz.

En el sistema USDC, las señales analógicas de voz que llegan se muestrean primero, y a continuación se convierten a una PCM binaria en un *codificador de voz* especial, llamado *codificador lineal predictivo de suma vectorial excitadora* (VSELP, de *vector sum exciter linear predictive*), o *codificador lineal predictivo de excitación estocástica* (SELP, de *scochastically excited linear predictive*). Los codificadores lineales predictivos usan codificadores de voz en dominio del tiempo, que tratan de extraer las características más importantes de la forma de onda de voz

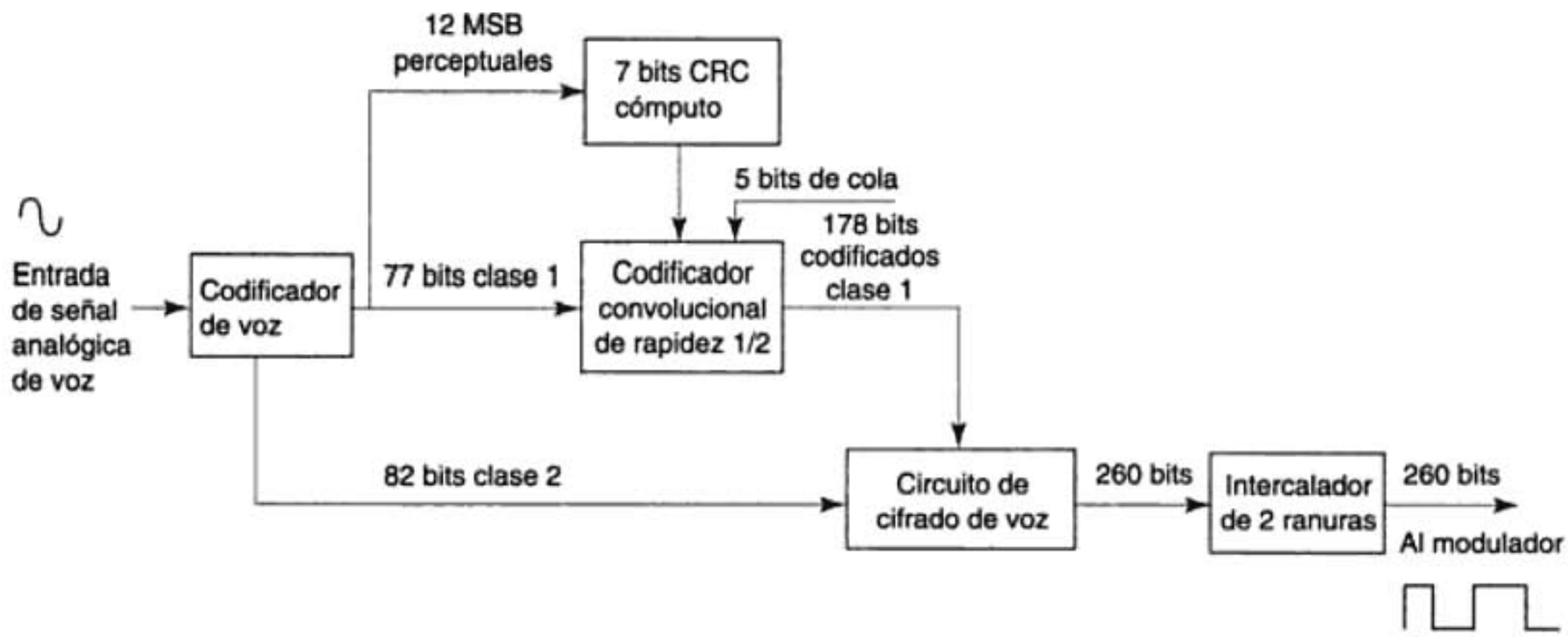


FIGURA 20-12 Codificador de voz USDC de canal digital de voz

variable en el tiempo. En estos codificadores, es posible transmitir voz, con buena calidad, a 4.8 kbps, y una voz aceptable, aunque de menor calidad, a menores frecuencias de bits.

Como hay muchos órdenes predecibles en los patrones de conversación, es posible, mediante algoritmos avanzados, comprimir las muestras binarias y transmitir la corriente de bits que resulta a una velocidad de 13 kbps. Un consorcio de empresas, que incluyó a Motorola, desarrolló el algoritmo VSELP, que a continuación fue adoptado como norma IS-54. Se agregan bits de *detección y corrección de errores* (EDC, de *error-detection and correction*) a las señales de voz comprimidas digitalmente, para reducir los efectos de la interferencia, llevando la velocidad final de datos de voz a 48.6 kbps. Las funciones de compresión y expansión y de detección y corrección de errores se realizan en el aparato telefónico mediante un microprocesador especial, llamado *procesador digital de señales* (DSP, de *digital signal processor*).

Los codificadores VSELP tienen salida de 7950 bps, y producen una trama de voz cada 20 ms, es decir,

$$\frac{7950 \text{ bits}}{\text{segundo}} \times \frac{20 \text{ ms}}{\text{trama}} = 159 \text{ bits por trama}$$

Cincuenta tramas de voz salen cada segundo, con 159 bits cada una, es decir,

$$\frac{50 \text{ tramas}}{\text{segundo}} \times \frac{159 \text{ bits}}{\text{trama}} = 7950 \text{ bps}$$

Los 159 bits incluidos en cada trama de codificador de voz se dividen en dos clases, según la importancia con que se perciben. Hay 77 bits de clase 1 y 82 bits de clase 2. Los bits de clase 1 son los más importantes y, en consecuencia, se protegen contra errores. Los 12 bits más significativos de clase 1 se codifican en bloque mediante un código de detección de errores CRC de 7 bits, para asegurar que los bits más significativos del codificador de voz se decodifiquen con baja probabilidad de error. Los bits de clase 2, menos significativos, no tienen forma de protegerse contra errores.

Después de codificar los 159 bits, cada trama de código de voz se convierte a 260 bits codificados de canal por trama, en un codificador de convolucional de rapidez 1/2, y se transmiten 50 tramas por segundo. Por consiguiente, la velocidad de transmisión de bits aumenta de 7950 bps para cada canal digital de voz, a 13 kbps,

$$\frac{260 \text{ bits}}{\text{trama}} \times \frac{50 \text{ tramas}}{\text{segundo}} = 13 \text{ kbps}$$

La fig. 20-13 muestra la ranura de tiempo y el formato de trama para los enlaces de sentido directo (estación base a unidad móvil) e inverso (unidad móvil a estación base) de un canal

G1	RSDSDVSDWSDXSDYS	G2

Donde G1 = tiempo de protección de 6 bits
R = tiempo de rampa de 6 bits de longitud
S = palabra de sincronización de 28 bits
D = código CDVCC de 12 bits
G2 = tiempo de protección de 44 bits
V = 0000
W = 00000000
X = 000000000000
Y = 0000000000000000

FIGURA 20-14 Formato de canal de voz digital con ráfaga acortada USDC

móvil transmite ranuras de ráfaga acortada, hasta que la estación base determina el desplazamiento requerido en el tiempo. El retardo predeterminado entre las ranuras de recepción y transmisión es de 44 símbolos, que da como resultado una distancia máxima de 72 millas para una célula IS-54, a la cual una estación móvil pueda operar en la célula.

Esquema de modulación digital USDC

Para lograr una velocidad de transmisión de 48.6 kbps en un canal de voz AMPS de 30 kHz, se requiere una *eficiencia de ancho banda (espectral)* de 1.62 bps/Hz, bastante lejos de las posibilidades de la FSK binaria. Los requisitos de eficiencia espectral se pueden alcanzar usando esquemas convencionales de modulación de 4 fases, de pulso conformado, como QPSK y OQPSK. Sin embargo, los canales USDC de voz y de control usan una técnica de *modulación por conmutación de fase, simétrica diferencial*, llamado DQPSK $\pi/4$, o *modulación por conmutación cuadrifásica diferencial* $\pi/4$ (DQPSK), que ofrece varias ventajas en un ambiente de radio móvil, como mejor rechazo de canal común y mejor eficiencia de ancho de banda.

Para una velocidad de datos de 48.6 kbps se requiere una velocidad de símbolos (bauds) de 24.3 kbps (24.3 kilobauds por segundo), con duración de símbolo de 41.1523 μs. El uso de conformación de pulso y de DQPSK $\pi/4$ soporta la transmisión de tres señales digitalizadas de voz, distintas, de 48.6 kbps en un ancho de banda de 30 kHz, con aislamiento de canal adyacente hasta de 50 dB. Por lo anterior, la eficiencia de ancho de banda cuando se usa DQPSK $\pi/4$ es

$$\eta = \frac{3 \times 48.6 \text{ kbps}}{30 \text{ kHz}}$$
$$= 4.86 \text{ bps/Hz}$$

siendo η la eficiencia de ancho de banda.

En un modulador DQPSK $\pi/4$, los bits de datos se dividen en dos canales paralelos que producen un desplazamiento específico de fase en la portadora analógica y, como hay cuatro pares posibles de bits, hay cuatro desplazamientos posibles de fase si se usa un modulador I/Q en cuadratura. Los cuatro cambios diferenciales posibles de fase son $\pi/4$, $-\pi/4$, $3\pi/4$ y $-3\pi/4$, y definen ocho fases posibles de portadora. La conformación de pulsos se usa para minimizar el ancho de banda y al mismo tiempo limitar la interferencia entre símbolos. En el transmisor se filtra la señal PSK con un filtro cosenoidal elevado de raíz cuadrada, con factor de atenuación progresiva de 0.35. Las señales PSK, después de la conformación de pulso, se pasan por una técnica de modulación lineal, que requiere amplificación lineal para conservar la forma del pulso. El uso de conformación de pulso con DQPSK $\pi/4$ permite la transmisión simultánea de tres señales de voz separadas, de 48.6 kbps en un ancho de banda de 30 kHz.

Canal de control digital USDC

El canal de control digital es complejo, por necesidad, y su descripción completa sale del alcance de este libro. En consecuencia, la siguiente descripción pretende ser una perspectiva general del funcionamiento de un canal de control digital USDC.

Como se dijo antes, la norma IS-54 especifica tres clases de canales: analógico de control, analógico de voz y un canal de control digital FSK binario de 10 kbps (DCCH, de *digital control channel*). La norma IS-54 Rev. C (IS-136) menciona las mismas tres clases de canales, y una cuarta: un canal de control digital con una frecuencia de señalización de 48.6 kbps, en los canales de control exclusivos de USDC. El nuevo canal de control digital pretende reemplazar, al final, al canal de control analógico. Al agregar un canal de control digital, una unidad móvil puede funcionar por completo en el dominio digital: usar el canal de control digital para seleccionar sistema y célula, y para entrada a canal, y el canal digital de voz para transmisiones digitalizadas de voz.

La IS-136 detalla la funcionalidad exacta del canal USDC de control digital. La versión inicial de la IS-136 fue la versión 0, que de desde entonces se ha actualizado por la revisión A. La versión 0 agregó muchos servicios y funciones nuevos al sistema telefónico celular digital USDC, incluyendo mayores servicios al usuario, como mensajes cortos y presentación del número telefónico de la llamada que entra, modo de reserva, que permite una mayor duración de baterías del aparato telefónico cuando no se use, servicio de sistema privado o residencial y mayor seguridad y validación contra fraudes. La versión más reciente de la IS-136 revisión A, fue desarrollada para proporcionar también muchas funciones y servicios nuevos, al introducir un codificador ampliado de voz, activación al vuelo en donde se permite que los operadores de la red programen información en los teléfonos en forma directa por el aire, identificación de nombre y número de quien llama, mayor libertad de manos y acceso prioritario a los canales de control.

En la norma IS-136 se especifican varias funciones de grupo para usuarios privados, que se adaptan bien a aplicaciones PBX inalámbrica y aplicaciones de voceo. Sin embargo, las terminales de usuario IS-136 funcionan a 48.6 kbps y, en consecuencia, no son compatibles con las terminales FSK IS-54. Entonces, los módems IS-136 son más rentables, porque sólo es necesario incluir el módem de 48.6 kbps en el equipo de terminal.

Canales lógicos. El nuevo canal digital de control incluye varios *canales lógicos* con distintas funciones, como el *canal de acceso aleatorio* (RACH, de *random access channel*); *SMS punto a punto, voceo y canal de respuesta de acceso* (SPACH, de *SMS point-to-point, paging and access response channel*); el *canal de control de emisión* (BCCH, de *broadcast control channel*) y el canal de *retroalimentación en canal compartido* (SCF, de *shared channel feedback*). La fig. 20-15 muestra los canales lógicos de la norma IS-136.

Canal de acceso aleatorio (RACH). Este canal lo usan las unidades móviles para pedir acceso al sistema telefónico celular. El RACH es unidireccional, especificado sólo para transmisiones de unidades móviles a base. Los mensajes de acceso, como origen, registro, respuestas de voceo, confirmación de auditoría, número de serie y confirmación de mensaje se transmiten por el canal RACH. También transmite mensajes que proporcionan información sobre validación, actualizaciones de parámetros de seguridad y ***servicio de mensajes cortos*** (SMS, de

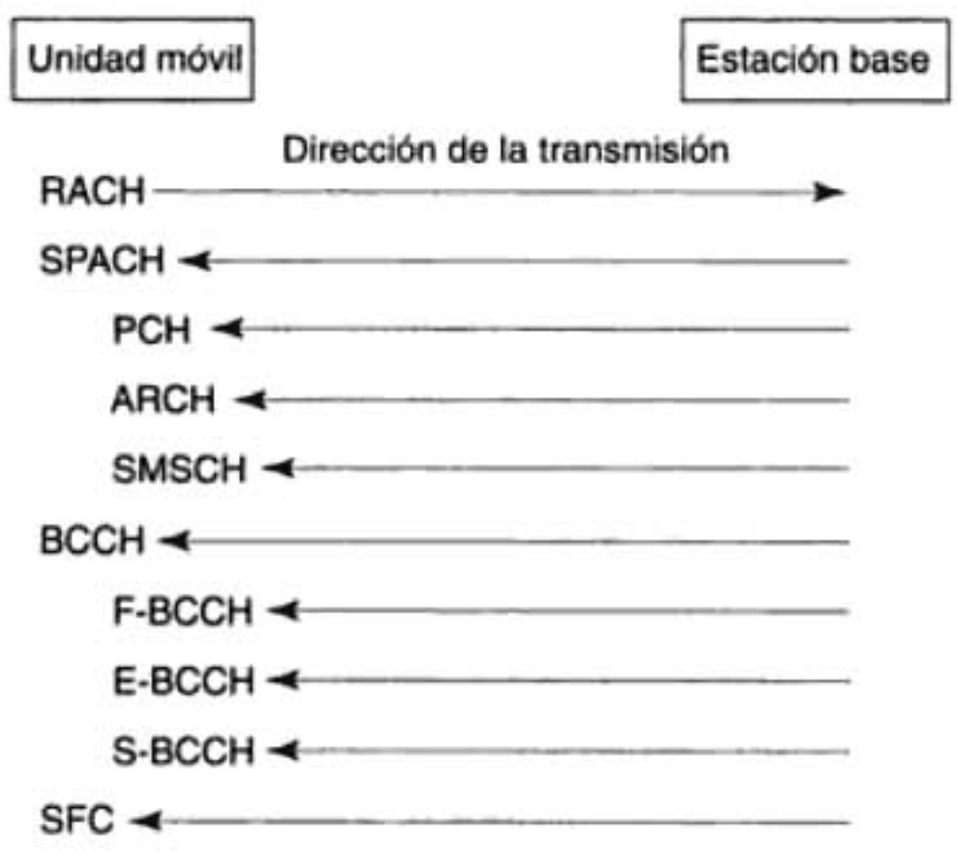


FIGURA 20-15 Canal de control digital USDC IS-136: canal lógico y subcanales lógicos

short message service) de punto a punto. Este canal puede funcionar en dos modos, con resolución de contención parecida a los canales de voz. También puede funcionar en ***modo de reservación***, para responder a una orden de la estación base.

Canal SMS punto a punto, voceo y respuesta de acceso (SPACH). Se usa para transmitir información desde las estaciones base a estaciones móviles específicas. El RACH es un canal unidireccional, especificado para transmitir sólo de estaciones base a unidades móviles, y lo comparten todas las unidades móviles. La información transmitida por el canal SPACH incluye tres subcanales lógicos separados: ***mensajes SMS punto a punto, mensajes de voceo y mensajes de respuesta de acceso***. Este canal puede llevar mensajes relacionados a una sola unidad móvil o a pequeños grupos de unidades móviles, y permite dividir mensajes mayores en bloques más pequeños para su transmisión.

El *canal de voceo* (PCH, de *paging channel*) es un subcanal del canal lógico del SPACH. El canal PCH se dedica a transmitir búsquedas de personas y órdenes. Transmite *mensajes de voceo*, *mensajes de mensaje en espera* y *mensajes de alerta al usuario*. Cada mensaje PCH puede llevar hasta cinco identificadores móviles. Los mensajes de voceo siempre se transmiten y a continuación se repiten una segunda vez. Los mensajes como *actualizaciones de conteo e historial de llamadas* y *actualizaciones de datos secretos compartidos* que se usan en procesos de validación y de cifrado también se mandan por el PCH.

El *canal de respuesta de acceso* (ARCH, de *access response channel*) también es un subcanal lógico del SPACH. Una unidad móvil se pasa en forma automática a un ARCH de inmediato, al completar el acceso basado en contención o reservación en un RACH. El canal ARCH se puede usar para llevar asignaciones a otro recurso u otras respuestas al intento de acceso de la estación móvil. Los mensajes que asignan una unidad móvil a un canal analógico de voz o a un canal digital de voz, o que redireccionan la unidad móvil a una célula distinta, también se mandan por el ARCH, junto con mensajes de registro de acceso (aceptar, rechazar o liberar).

El canal SMS o SMSCH se usa para entregar mensajes cortos de punto a punto a una estación móvil específica. Cada mensaje se limita a un máximo de 20 caracteres de texto. El SMS originado en móvil también se soporta; sin embargo, no se soporta el SMS donde una estación base puede emitir un mensaje corto dirigido a varias unidades móviles, en la norma IS-136.

Canal de control de emisión (BCCH). BCCH es un acrónimo que representa los canales lógicos F-BCCH, E-BCCH y S-BCCH. Estos canales se usan para conducir información genérica, relacionada con el sistema. El BCCH es una transmisión unidireccional de la estación base a la unidad móvil, que comparten todas las unidades móviles.

El *canal de control de emisión rápida* (F-BCCH) emite parámetros de estructura del canal digital de control (DCCH), que incluyen información acerca de la cantidad de ranuras de tiempo F-BCCH, E-BCCH y S-BCCH, en la trama DCCH. Las unidades móviles usan la información F-BCCH cuando entran por primera vez al sistema, para determinar el inicio y final de cada canal lógico en la trama DCCH. También, el F-BCCH incluye información acerca de los parámetros de acceso, con la información necesaria para validación y cifrados, e información de los intentos de acceso de la unidad móvil, como la cantidad de intentos de acceso del número, tamaño de la ráfaga de acceso, nivel de potencia en acceso inicial e indicación de si la célula está protegida o no. También, por el canal F-BCCH se manda información sobre los distintos tipos de registro, periodos de registro e identificación del sistema, incluyendo el tipo de red, código móvil de país y revisión de protocolo.

El *canal de control de emisión extendida* (E-BCCH) lleva información de radiodifusión menos crítica que el F-BCCH, dirigida a las unidades móviles. El E-BCCH conduce información acerca de las celdas vecinas analógicas y TDMA, y mensajes opcionales, como información de emergencia, mensajes de hora y fecha, y los tipos de servicios que soportan las células vecinas.

El *canal de control de emisión SMS* (S-BCCH) es un canal lógico para mandar mensajes cortos a unidades móviles individuales.

Canal de retroalimentación compartida de canal (SCF). Este canal se usa para soportar la operación de canales de acceso aleatorio, proporcionando información acerca de cuáles ranuras de tiempo de la unidad móvil se pueden usar para intentos de acceso, y también si la transmisión RACH anterior de la unidad se recibió bien.

El canal inferior CDMA en la banda A está en el canal 283 de AMPS, y la portadora inferior CDMA en la banda B está en el canal 384 AMPS. Como la banda disponible entre los canales 667 y 716 sólo es 1.5 MHz en la banda A, los operadores de banda A deben obtener autorización de las portadoras de banda B, para usar una portadora CDMA en esa parte del espectro de frecuencias. Cuando una portadora CDMA se está usando junto a una portadora no de CDMA, la distancia de portadoras debe ser 1.77 MHz. Hay hasta 9 portadoras CDMA disponibles para el operador de banda A y B en el espectro de frecuencias AMPS. Sin embargo, los operadores de banda A y B tienen un ancho de banda de 30 MHz en la banda de frecuencias PCS de 1900 MHz, donde pueden proporcionar hasta 11 canales CDMA.

En el sistema CDMA, muchos usuarios pueden compartir canales comunes de transmisión y recepción, con velocidad de transmisión de datos de 9.6 kbps. Sin embargo, con ciertas técnicas, la información del usuario se dispersa con un factor de 128 hasta una velocidad de pulsos del canal de 1.2288 Mpulsos/s, y los canales de transmisión y recepción usan distintos procesos de dispersión.

En el canal de sentido directo se codifican los datos del suscriptor con un código de 1/2 convolucional de rapidez, intercalado y difundido por 1 de 64 secuencias de difusión ortogonal, usando funciones de Walsh. Se mantiene la ortogonalidad entre todos los suscriptores en el canal celular de sentido directo dentro de determinada célula, porque todas las señales de la célula se desorganizan en forma síncrona.

Los canales de sentido inverso usan una estrategia distinta de dispersión, porque la señal recibida por cada unidad móvil toma una trayectoria distinta de transmisión y, en consecuencia, llega a la estación base en un tiempo distinto. Las corrientes de datos en el canal de sentido inverso se codifican primero en forma convolucional con un código de convolución de rapidez de 1/3. Después de intercalarlos, cada bloque de seis símbolos codificados se distribuye en una de las funciones ortogonales de Walsh disponibles, asegurando una señalización ortogonal 64-aria. Se hace una dispersión adicional de cuatro veces con códigos específicos de suscriptor y de estación base, con periodos de pulso de $2^{42}-1$ y 2^{15} pulsos, respectivamente, con lo que aumenta la velocidad de transmisión a 1.2288 Mpulsos/s. Se imponen estrictos requisitos en la potencia de transmisión del canal de sentido inverso, para evitar el problema cercano-lejano causado por diversos niveles de potencia de recepción.

A cada unidad móvil en determinada celda se le asigna una secuencia única de dispersión, con lo que se asegura una separación casi perfecta entre las señales de distintas unidades de suscriptor, y se permite diferenciar transmisiones entre usuarios. Todas las señales en determinada célula se desorganizan mediante una secuencia pseudoaleatoria de 2^{15} pulsos de longitud. Así se reduce la interferencia de radiofrecuencia entre unidades móviles en celdas vecinas, que puedan estar usando la misma secuencia de dispersión, y se proporcionan las características espectrales deseadas de banda ancha, aun cuando todos los códigos de Walsh no produzcan un espectro de potencias de banda ancha.

Hay dos técnicas que se suelen usar para dispersar el espectro: *salto* (o *cambio*) *de frecuencia* y *secuencia directa*. Las dos son características de transmisiones por un ancho de banda mucho mayor que el que normalmente se usa en los sistemas telefónicos celulares FDMA/TDMA de banda angosta, como AMPS y USDC. Para conocer una descripción más detallada del cambio de frecuencia y del secuenciamiento directo, véase el capítulo 19.

Dispersión del espectro por cambio de frecuencia. Este sistema se usó por primera vez en aplicaciones militares, para asegurar comunicaciones confiables sin interferencia en un ambiente de campo de batalla. El concepto fundamental del cambio de frecuencia es dividir un mensaje en bloques de datos de tamaño fijo, y transmitir cada bloque en secuencia, pero en distintas frecuencias de portadora. En los saltos de frecuencia se usa un código pseudoaleatorio para generar una secuencia única de saltos. La secuencia en que se seleccionan las frecuencias debe ser conocida por el transmisor y el receptor, antes de iniciar la transmisión. El transmisor manda un bloque en una portadora de radiofrecuencia, y a continuación cambia o salta a la siguiente frecuencia de la secuencia, y así sucesivamente. Después de recibir una recepción de datos en una frecuencia, el receptor cambia a la siguiente de la secuencia. Cada transmisor del sistema tiene una secuencia distinta de cambios, para evitar que un suscriptor interfiera las transmisiones de otros suscriptores que usen la misma frecuencia de canal de radio.

Dispersión de espectro por secuencia directa. En los sistemas de secuencia directa se agrega un código pseudoaleatorio con alta velocidad de bits, a una señal de información de baja velocidad de bits, para generar una señal pseudoaleatoria que se parece mucho al ruido, y contiene tanto la señal de datos originales como el código pseudoaleatorio. También aquí es necesario que el transmisor y el receptor conozcan el código pseudoaleatorio. Cuando un receptor detecta una transmisión de secuencia directa, sólo resta la señal pseudoaleatoria de la señal recibida compuesta, para extraer los datos de información. En los sistemas telefónicos celulares CDMA, el ancho de banda de radiofrecuencias total se divide en unos pocos canales de radio de banda ancha, cuyo ancho es mucho mayor que la señal de voz digitalizada. Esta señal digitalizada se suma a la señal generada de alta frecuencia de bits y se transmite en tal forma que ocupe toda la banda ancha del canal de radio. Al sumar una señal pseudoaletoria de alta frecuencia de bits a la información de voz hace que la señal sea más dominante y menos susceptible a la interferencia, permitiendo una transmisión de menor potencia y, en consecuencia, menor cantidad de transmisores y receptores menos costosos.

Canales de tráfico CDMA

Los canales de tráfico CDMA consisten en un canal de sentido inverso o enlace descendente (estación base a unidad móvil) y uno de enlace ascendente o de sentido directo (estación móvil a estación base). En la fig. 20-17a se ve un canal de tráfico de enlace descendente. Se ve que ese canal consiste en hasta 62 canales, que incluyen uno de radiodifusión, usado para control, y canales de tráfico para llevar información de suscriptor. El canal de radiodifusión consiste en un canal piloto, un canal de sincronización, hasta siete canales de voceo y hasta 63 canales de tráfico. Todos estos canales comparten la misma asignación de frecuencias CDMA de 1.25 MHz. El canal de tráfico se identifica con una secuencia distinta de código largo, y específica del usuario, y cada canal de acceso se identifica con una secuencia distinta de código largo.

El canal piloto está incluido en cada célula, con el fin de proporcionar una señal para que use el receptor en la adquisición de sincronización, y proporcionar una referencia de fase para demodulación coherente. También lo usan las unidades móviles para comparar las intensidades de señal entre estaciones base, y determinar cuándo se debe iniciar una transferencia. El canal de sincronización usa un código Walsh W32, y la misma secuencia pseudoaleatoria y desplazamiento de fase que el canal piloto, permitiéndole ser demodulado por cualquier receptor que pueda adquirir la señal piloto. El canal de sincronización difunde mensajes de sincronización a las unidades móviles y funciona a 1200 bps. Los canales de voceo conducen información de la estación base a la móvil, como por ejemplo, mensajes de parámetro de sistema o de acceso, mensajes de lista de canales CDMA y mensajes de asignación de canal. Los canales de voceo son opcionales, y su cantidad puede ir de 0 a 7. Se usan para transmitir información de control y

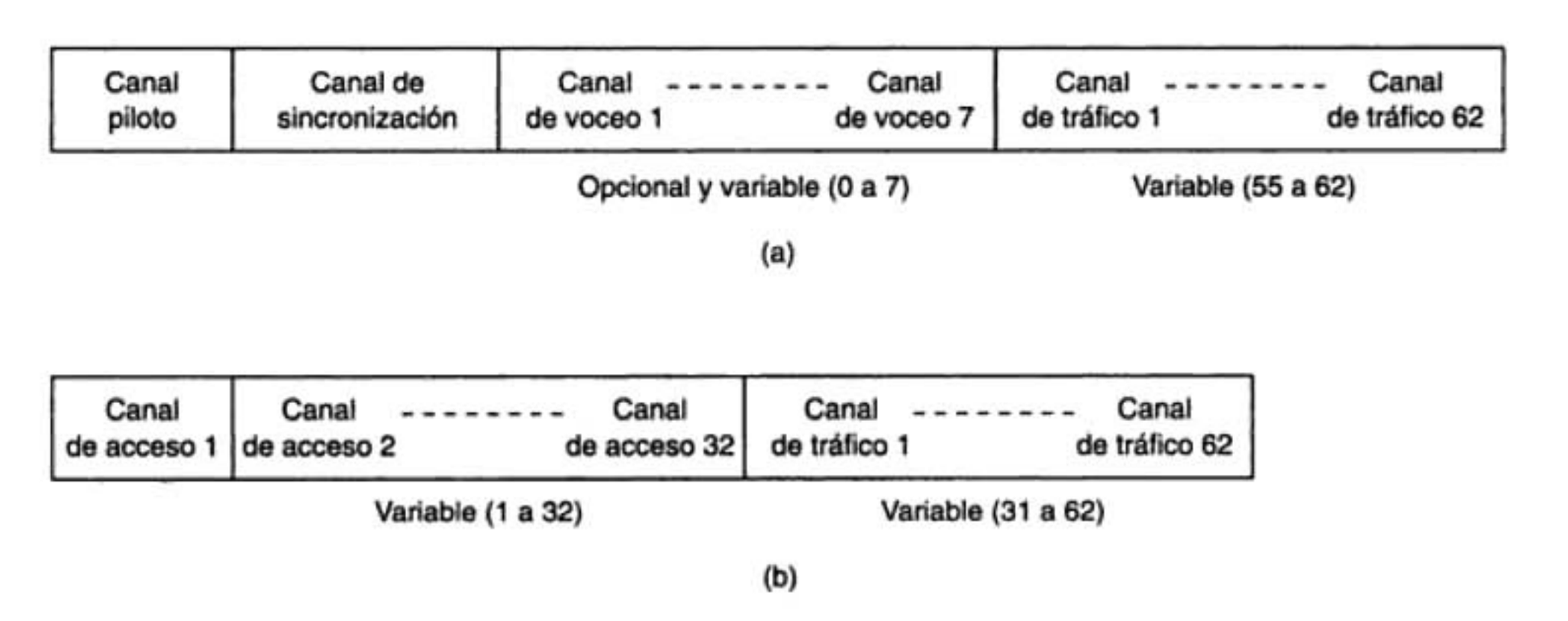


FIGURA 20-17 Canales de tráfico IS-95: (a) enlace en sentido descendente; (b) enlace en sentido ascendente

mensajes de voceo de la estación base a las unidades móviles, y puede funcionar a 9600 bps, 4800 bps o 2400 bps. Un solo canal piloto de 9600 bps puede soportar, en forma característica, unos 180 voceos por segundo, haciendo una capacidad total de 1260 voceos por segundo.

Los datos del tráfico en el canal de enlace descendente, o de sentido inverso, se agrupan en tramas de 20 ms. Primero se codifican en forma convolucional, y a continuación se formatean e intercalan para compensar las diferencias de las velocidades reales de usuario, que son variables. La señal resultante se dispersa con un código de Walsh y una secuencia pseudoaleatoria larga, a la velocidad de 1.2288 Mpulsos/s.

El canal de transmisión de enlace en sentido directo se ve en la fig. 20-17b, y está formado por canales de acceso y hasta 62 canales de tráfico en sentido directo. Los canales de acceso y de tráfico de canal de enlace ascendente usan la misma asignación de frecuencias, con técnicas CDMA de secuencia directa. Los canales de acceso son sólo canales de sentido directo, compartidos y de punto a punto, que proporcionan comunicaciones de unidades móviles a estaciones base, cuando la unidad móvil no usa un canal de tráfico. La unidad móvil usa los canales de tráfico para iniciar comunicaciones con una estación base, y para responder a mensajes del canal de voceo. Entre los mensajes normales del canal de acceso están de reconocimiento y número de secuencia, parámetros de identificación del usuario móvil y parámetros de comprobación. El canal es de acceso aleatorio, y cada canal de suscriptor está identificado en forma única por su código pseudoaleatorio. El canal CDMA de sentido directo puede contener un máximo de 32 canales de acceso por cada canal soportado de voceo. Este canal de acceso en sentido directo funciona en modo de velocidad variable de datos, y los canales de acceso funcionan a una velocidad de datos fija, de 4800 bps. Los mensajes del canal de acceso consisten en registro, orden, ráfaga de datos, origen, respuesta al voceo, respuesta al desafío de comprobación, respuesta de estado y terminación de asignación.

Los datos del suscriptor en el canal de radio de enlace en sentido directo también se agrupan en tramas de 20 ms codificadas en forma convolucional, intercaladas en bloque, con modulación ortogonal 64-aria y dispersados antes de su transmisión.

SISTEMA GLOBAL PARA COMUNICACIONES MÓVILES

A principios de la década de 1980, los sistemas telefónicos celulares analógicos tuvieron un periodo de rápido crecimiento en Europa Occidental, en especial en Escandinavia y en el Reino Unido, y en menor grado en Francia y Alemania. A continuación, cada país desarrolló su propio sistema celular, incompatible con los de los demás, tanto en equipo como en el sistema de operación. La mayoría de los sistemas existentes funcionaba en distintas frecuencias, y todos eran analógicos. En 1982, la *Conferencia de Correos y Telégrafos Europeos* (CEPT) formó un grupo de estudio, llamado *Groupe Spécial Mobile* (GSM) para estudiar y desarrollar un sistema telefónico móvil terrestre y público, paneuropeo. En 1989 fue transferida la responsabilidad de la GSM al *Instituto Europeo de Normas de Telecomunicación* (ETSI, de *European Telecommunication Standards Institute*) y en 1990 se publicó la fase I de las especificaciones GSM. Éste tuvo la ventaja de haberse diseñado desde cero, sin importar si era compatible con los sistemas telefónicos celulares analógicos existentes.

El servicio comercial GSM comenzó en 1991, y para 1993 había 36 redes GSM en 22 países. Hoy, están planeadas o ya funcionan redes GSM en más de 80 países en todo el mundo. América del Norte entró tarde al mercado GSM, con un derivado del sistema llamado PCS 1900. Hoy, existen sistemas GSM en todos los continentes, y el significado actual de GSM es *Global System for Mobile Communications*, Sistema global de comunicaciones móviles.

El GSM es un sistema telefónico celular de segunda generación, desarrollado al principio para resolver los problemas de fragmentación inherentes a los sistemas de primera generación en Europa. Antes del GSM, todos los países europeos usaban distintas normas de telefonía celular, por lo que era imposible que un suscriptor de cualquier país usara un solo teléfono en toda Europa. El GSM fue el primer sistema telefónico celular totalmente digital, diseñado para usar los servicios de la ISDN (red digital de servicios integrados) para proporcionar una amplia gama de servicios a la red. Ahora cuenta con entre 20 y 50 millones de suscriptores, y es la norma mundial más popular para equipos telefónicos y equipos personales de comunicación nuevos.

Servicios de GSM

La intención original era hacer GSM compatible con ISDN, en términos de servicios ofrecidos y formatos de señalización de control. Desafortunadamente, las limitaciones de ancho de banda de canal y de costo evitan que el GSM funcione a la velocidad básica de datos de 64 kbps de ISDN.

Los servicios telefónicos GSM se pueden clasificar en forma amplia en tres categorías: *servicios de portador*, *teleservicios* y *servicios suplementarios*. Es probable que el servicio al portador más básico (*teleservicio*) que proporciona GSM es el de telefonía. En GSM, las señales analógicas de voz se codifican digitalmente y a continuación se transmiten por la red, como corriente de datos digitales. También hay un servicio de emergencia, en el que se notifica al proveedor más cercano correspondiente marcando tres dígitos, en forma parecida a los servicios del 911 en Estados Unidos. Se ofrece una amplia gama de servicios de datos por GSM, donde los usuarios pueden mandar y recibir datos hasta con 9600 bps a suscriptores en las redes POTS (*plain old telephone service*, servicio telefónico antiguo simple), ISDN, *red pública de datos de paquetes conmutados* (PSPDN, de *Packet Switched Public Data Network*), y *red pública de datos de circuitos conmutados* (CSPDN, de *Circuit Switched Public Data Network* (CSPDN), mediante diversos métodos y protocolos de acceso, como X.25. También, como la GSM es una red digital, no se requiere módem entre el usuario y esa red.

Entre otros servicios de datos GSM están facsímil de grupo 3, según la recomendación T.30 de ITU-T. Una función única de GSM que no existe en los sistemas analógicos anteriores es el *servicio de mensajes cortos* (SMS, de *Short Message Service*), que es bidireccional para mandar mensajes alfanuméricos hasta de 160 bytes de longitud. Estos mensajes SMS son transportados por el sistema en forma de almacenar y enviar. También se pueden usar en un modo de radioemisión celular, para mandar mensajes a receptores múltiples. También se ofrecen en GSM varios servicios suplementarios, como *notificación de llamadas* e *impedimento de llamadas*.

Arquitectura del sistema GSM

La arquitectura del sistema GSM, mostrada en la fig. 20-18, consiste en tres subsistemas principales interconectados, que interaccionan entre sí y con los suscriptores a través de interfaces especificadas de red. Los tres subsistemas principales del GSM son *subsistema de estación base*

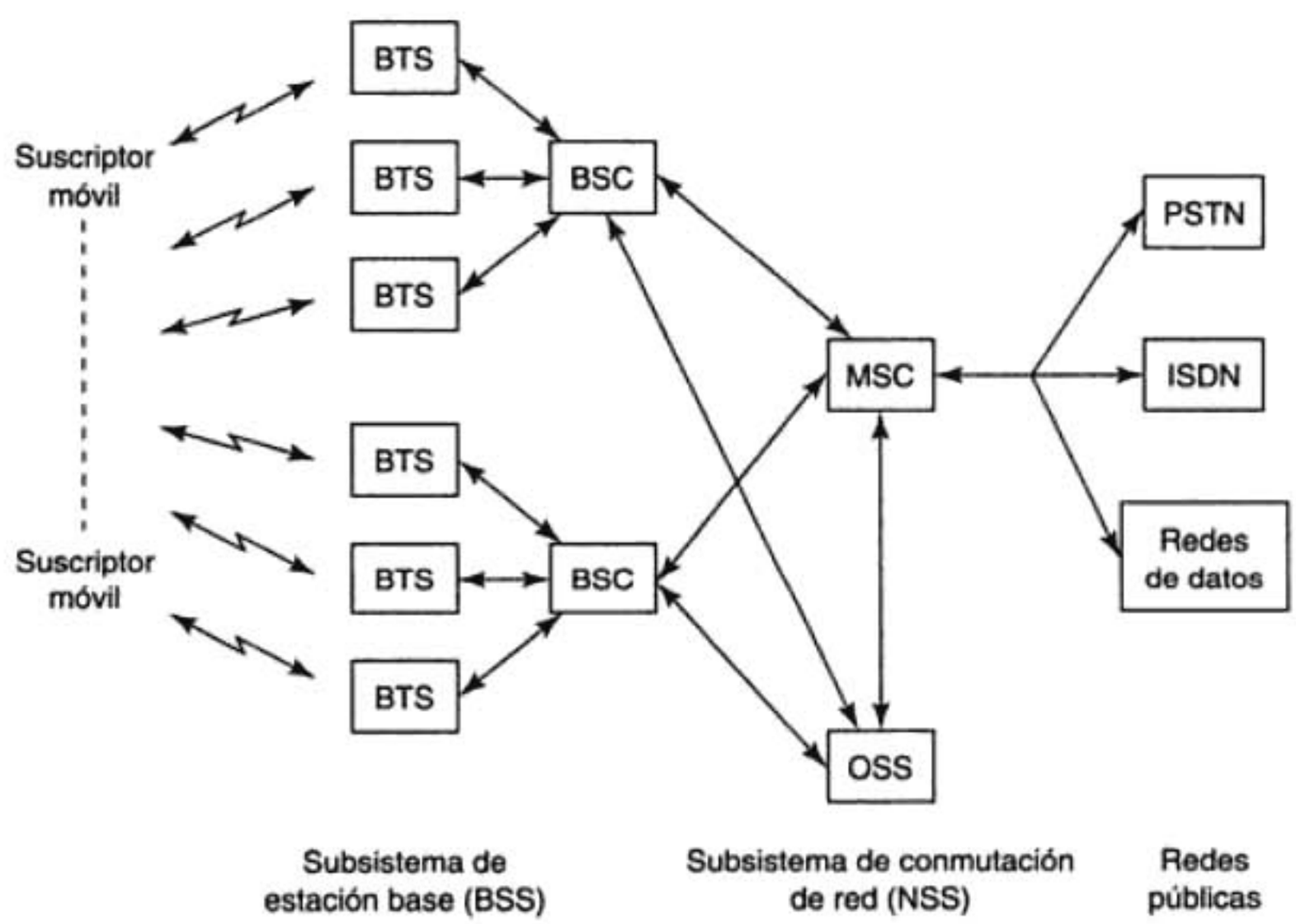


FIGURA 20-18 Arquitectura del sistema GSM

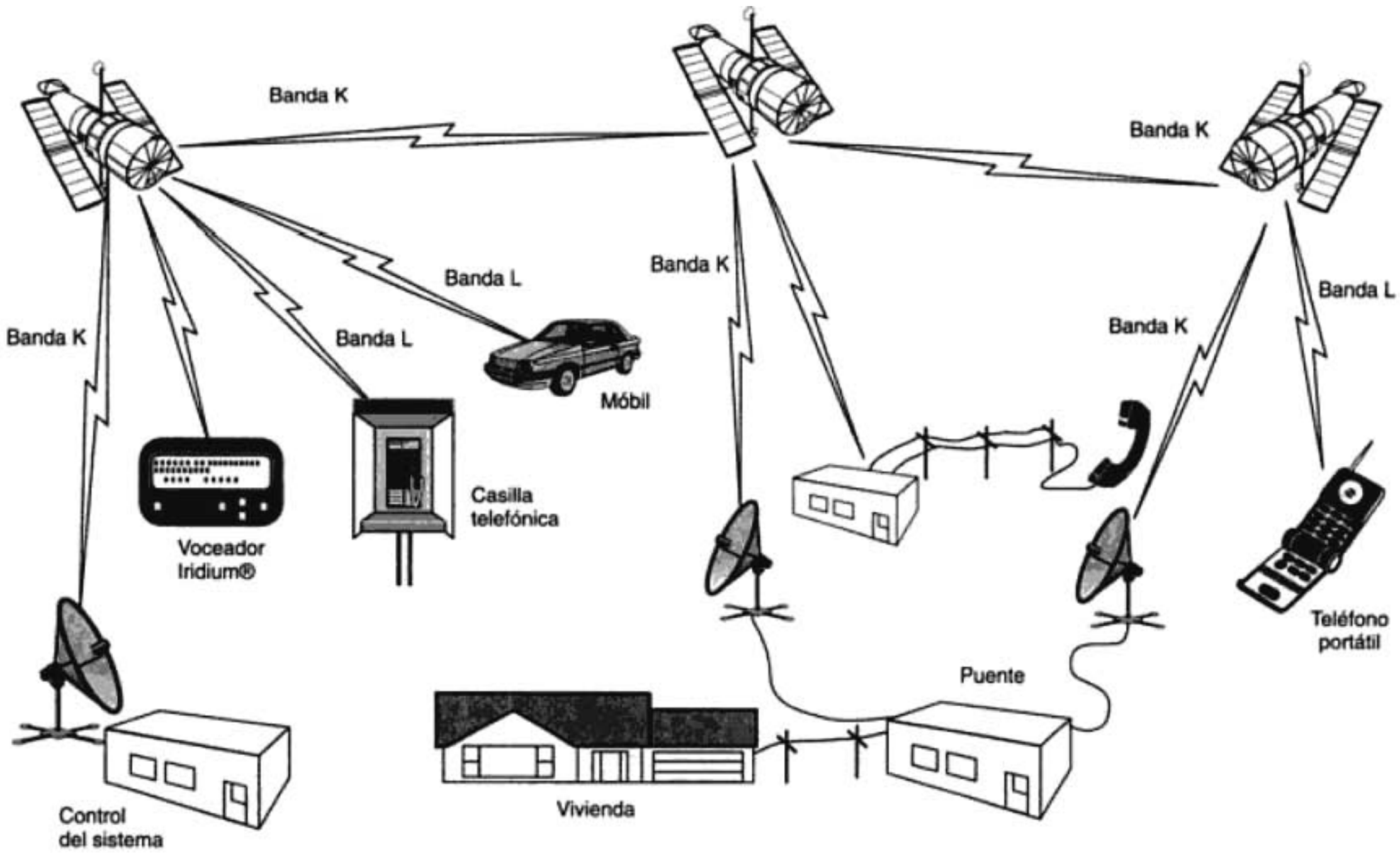


FIGURA 20-19 Panorámica del sistema telefónico móvil PCSS Iridium

3. Proporcionar servicios de mensajes, facsímil y datos a los suscriptores de Iridium.
4. Facilitar las actividades comerciales del sistema Iridium a través de un conjunto de acuerdos corporativos mutuos.

Constelación de satélites. La base del sistema de satélites Iridium es proporcionar cobertura terrestre total. El sistema usa 66 satélites funcionales (también hay algunos de reserva) configurados a una elevación media de 420 millas sobre la Tierra, en seis planos orbitales casi polares (inclinados 86.4°), en el que giran 11 satélites en torno a la Tierra en cada órbita, con un periodo de órbita de 100 min 28 s. Esto permite que Iridium cubra toda la superficie de la Tierra, y siempre que un satélite desaparezca de la vista de un suscriptor, lo sustituya otro. Los satélites se reparten en forma adecuada en collares norte-sur que forman *planos corrotatorios* de subida en un lado de la Tierra, cruzan los polos y de bajada en el otro lado. Los planos primero y último giran en direcciones opuestas, formando una *unión* virtual. Los planos corrotatorios están separados por 31.6°, y los planos de unión están a 22° entre sí.

Cada satélite tiene tres antenas de banda L, y se forma una distribución en panal formada por 48 haces individuales localizados con un total de 1628 células apuntadas directamente hacia abajo del satélite, como se ve en la fig. 20-20. Al recorrer su órbita el satélite, las huellas se mueven a través de la superficie terrestre, y las señales del suscriptor se conmutan de un haz al siguiente, o de un satélite al siguiente, en un proceso de transferencia. Cuando los satélites se acercan a los polos Norte o Sur, se traslapan sus huellas y sus haces. Entonces, los haces exteriores se apagan, para evitar ese traslape y ahorrar potencia en el satélite. Cada célula tiene 174 canales de voz dúplex, haciendo un total de 283,272 canales en todo el mundo.

La llave única del sistema Iridium es el uso de *enlaces cruzados* entre satélites, y también es la principal diferencia entre el sistema Iridio y el tradicional *sistema de tubo doblado*, en el que todas las transmisiones siguen la trayectoria de tierra-satélite-tierra. Iridium es el primer satélite

100 MHZ de ancho de banda, y los enlaces intersatelitales usan 200 MHz de ancho de banda. Las bandas de frecuencia son las siguientes:

Banda L, enlaces de voz de suscriptor a satélite = 1.616 a 1.6265 GHz
Banda Ka, enlaces de bajada por puente = 19.4 a 19.6 GHz
Banda Ka, enlaces de subida por puente = 29.1 a 29.3 GHz
Banda Ka, enlaces cruzados intersatelitales = 23.18 a 23.38 GHz

PREGUNTAS

20-1. Describa en forma breve la diferencia entre un aparato telefónico *móvil* y uno *portátil*.

20-2. Describa en forma breve la diferencia entre *radio móvil en dos sentidos* y *teléfono móvil en dos sentidos*.

20-3. ¿Qué quiere decir sistema telefónico *por cable*?

20-4. ¿Cuándo otorgó el primer permiso la FCC para operar un servicio de radio celular en desarrollo?

20-5. ¿Cuál fue el primer sistema telefónico celular en E. U. A.?

20-6. Haga una lista de los sistemas más comunes de radiocomunicación en dos sentidos, y describa en forma breve cada uno.

20-7. Describa la diferencia entre *zona de cobertura* y *célula*.

20-8. Explique por qué se usan células hexagonales y no circulares.

20-9. Describa las diferencias entre una *macrocélula* y una *microcélula*.

20-10. Explique el concepto básico de *reutilización de frecuencia*.

20-11. ¿Cuál es la diferencia entre una celda *excitada en el borde*, y una *excitada en el centro*?

20-12. ¿Cuáles son las dos formas principales de interferencia en los sistemas telefónicos celulares?

20-13. Describa en forma breve las dos formas de interferencia mencionadas en la pregunta anterior.

20-14. Describa la ventaja y desventaja de tener un alto factor de *reutilización de canal compartido*.

20-15. ¿Qué quiere decir *efecto cercano-lejano*, y cómo afecta el funcionamiento del teléfono celular?

20-16. Describa en forma breve el término *división de célula*.

20-17. ¿Cuál es el objetivo principal de la división de célula?

20-18. ¿Qué quiere decir el término *bloqueo*?

20-19. Describa en forma breve los conceptos de *sectorización, segmentación* y *dualización*.

20-20. Describa en forma breve las funciones de una *central telefónica móvil* (MTSO).

20-21. ¿Qué es una *transferencia*?

20-22. Explique la diferencia entre una *transferencia suave* y una *dura*.

20-23. Haga una lista y describa las funciones de los seis componentes principales de un sistema telefónico celular.

20-24. Describa en forma breve los procedimientos para llamadas de *cable a móvil*, de *móvil a cable* y de *móvil a móvil*.

20-25. ¿Qué quieren decir los términos *acceso múltiple por división de frecuencia* y *duplexado por división de frecuencia*?

20-26. Explique qué son los *enlaces en sentido directo* y los *enlaces en sentido inverso*.

20-27. Describa en forma breve el formato del *canal de control* AMPS.

20-28. ¿Qué quiere decir el término *silencio y ráfaga*?*¿Cuándo se utiliza?*

20-29. Describa en forma breve las funciones del *tono de audio de supervisión* (SAT) y del *tono de señalización* (ST).

20-30. Describa en forma breve las diferencias entre AMPS y N-AMPS.

20-31. ¿Qué es ETACS?

20-32. Haga una lista de las ventajas de TDMA sobre FDMA.

20-33. ¿Cuál es la diferencia entre AMPS y D-AMPS?

20-34. ¿Qué quiere decir el término *modo dual*?

20-35. ¿Cuál es la velocidad de transmisión de bits especificada por la IS-54 para transmisiones de voz digitalizada?

20-36. ¿Cuál es la diferencia entre el espectro *normal* y el *extendido* de frecuencia en los sistemas celulares USDC?

20-37. ¿Cuántos canales de control se especifican para los sistemas AMPS? ¿Y para los sistemas USDC?

20-38. ¿Qué tipo de modulación se usa en los canales de control AMPS? ¿Y en los canales de control USDC?

20-39. ¿Cuál es el ancho de banda de un canal de voz AMPS? ¿Y de un canal de voz USDC?

20-40. ¿Qué *relación de compresión de canal* se logra en los sistemas USDC?

20-41. Describa en forma breve las funciones de los siguientes términos: *verificación digital codificada por código de colores, canal lento para control asociado, y canal rápido para control asociado.*

20-42. ¿Cuándo se usaría un formato de trama de *ráfaga acortada*?

20-43. ¿Qué tipo de modulación se usa con los canales USDC de voz y de control?

20-44. ¿Cuál es la eficiencia de ancho de banda del canal de voz con USDC?

20-45. Describa en forma breve las funciones de los canales siguientes: *canal lógico, canal de acceso aleatorio; canal SMS punto a punto, voceo y de respuesta de acceso; canal de control de transmisión, y canal compartido de retroalimentación.*

20-46. ¿Qué norma especifica los sistemas celulares digitales CDMA?

20-47. ¿Qué quiere decir PCS?

20-48. Cite y describa en forma breve dos técnicas de *dispersión de espectro.*

20-49. ¿Qué quiere decir canal *de enlace ascendente*? ¿Canal *de enlace descendente*?

20-50. ¿Cuál es el ancho de banda de un canal *de tráfico* CDMA?

20-51. ¿Qué partes forman el canal de tráfico CDMA?

20-52. ¿Dónde se usan los sistemas GSM?

20-53. Diga el nombre de las tres *clasificaciones* amplias de los servicios telefónicos GSM.

20-54. Haga una lista y describa en forma breve los tres *subsistemas primarios* del GSM.

20-55. ¿Cuál es el ancho de banda de un canal de voz GSM?

20-56. ¿Cuál es la separación de frecuencias de portadora GSM de transmisión y recepción?

20-57. Describa el término PCSS.

20-58. ¿Cuál es la desventaja principal de los sistemas PCSS?

20-59. Haga una lista de las ventajas de los sistemas PCSS.

20-60. ¿Cuántos satélites se usan en la constelación *Iridium*?

20-61. ¿En cuál *arquitectura telefónica* se basa Iridium?

20-62. Describa en forma breve la constelación de satélites Iridium.

PROBLEMAS

20-1. Calcule la cantidad de canales dúplex disponibles en un grupo, y la capacidad total de un sistema celular donde hay 10 unidades de asignación, formada cada una por 20 células con 18 canales en cada célula.

20-2. ¿Cuántos canales dúplex se requieren en cada célula para formar un sistema celular de 12 unidades de asignación, de 15 células por unidad, con una capacidad total de 3960 canales dúplex?

20-3. Calcule la cantidad de células en un grupo para los siguientes valores: $j = 3$ e $i = 4$.

20-4. Calcule la relación de reutilización de canal compartido para un grupo con 20 células.

20-5. Calcule la relación de reutilización de canal compartido para un grupo, cuando la distancia al centro de la célula más cercana de canal compartido es 4 km, y el radio de la célula es 1.2 km.

20-6. Determine las frecuencias de transmisión y recepción para los siguientes canales AMPS: 33, 156, 680, 991 y 1008.

A P É N

Imagen protegida por derechos de autor

El diagrama de Smith

INTRODUCCIÓN

Las soluciones matemáticas para el cálculo de impedancias de líneas de transmisión son laboriosas. En consecuencia se acostumbra usar gráficas para resolver los problemas de impedancias de líneas de transmisión. La ecuación A-1 es la fórmula para determinar la impedancia en determinado punto de una línea de transmisión

$$Z = Z_O\left[\frac{Z_L + jZ_O \tan \beta S}{Z_O + jZ_L \tan \beta S}\right] \tag{A-1}$$

en donde Z = impedancia de la línea en determinado punto
Z_L = impedancia de carga
Z_O = impedancia característica de la línea
βS = distancia de la carga al punto donde se calcula el valor de la impedancia

Hay varias gráficas en las que se presentan las propiedades de las líneas de transmisión. Sin embargo, las más útiles son aquellas que expresan las relaciones de impedancia que existen a lo largo de una línea de transmisión sin pérdida, para diversas condiciones de carga. El *diagrama de Smith* es la calculadora más usada en líneas de transmisión de este tipo. Es un tipo especial de sistema de coordenadas de impedancia, que retrata la relación entre la impedancia en cualquier punto a lo largo de una línea de transmisión uniforme, entre la impedancia en cualquier otro punto de la línea.

Esta gráfica fue desarrollada por Philip H. Smith, en los Bell Telephone Laboratories, y se describió por primera vez en un artículo titulado "*Transmission Line Calculator*" (*Electronics*, enero de 1939). En la fig. A-1 se presenta un diagrama de Smith. Se basa en dos conjuntos de círculos *ortogonales*. Un conjunto representa la relación del componente resistivo de la impedancia de línea (R) a la impedancia característica de la línea (Z_O), que para una línea sin pérdidas también es puramente resistiva. El segundo conjunto de círculos representa la relación del componente reactivo de la impedancia de línea ($\pm jX$) entre la impedancia

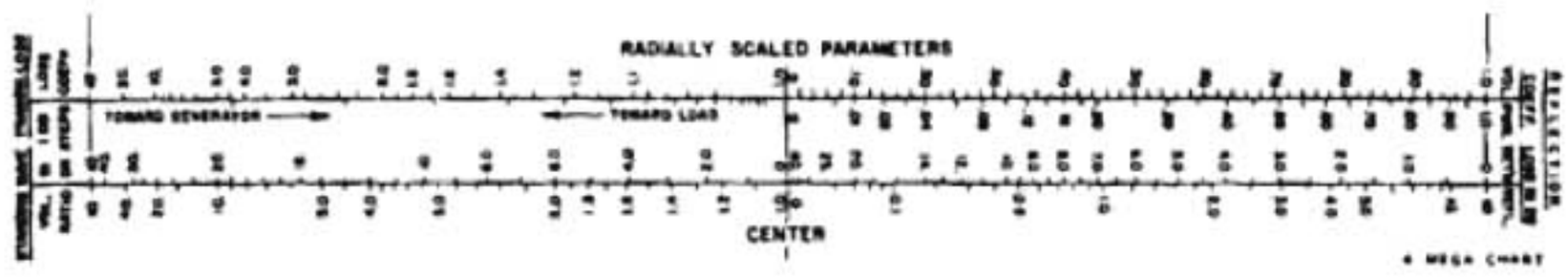


FIGURA A-1 Diagrama de Smith, para calcular líneas de transmisión

característica de la línea (Z_O). Entre los parámetros que se grafican en la gráfica de Smith están los siguientes:

1. Impedancia (o admitancia) en cualquier punto a lo largo de una línea de transmisión.
   a. Magnitud del coeficiente de reflexión (Γ).
   b. Ángulo del coeficiente de reflexión, en grados.
2. Longitud de la línea de transmisión entre dos puntos cualesquiera, en longitudes de onda.
3. Atenuación entre dos puntos cualesquiera.
   a. Coeficiente de pérdida de onda estacionaria.
   b. Pérdida por reflexión.

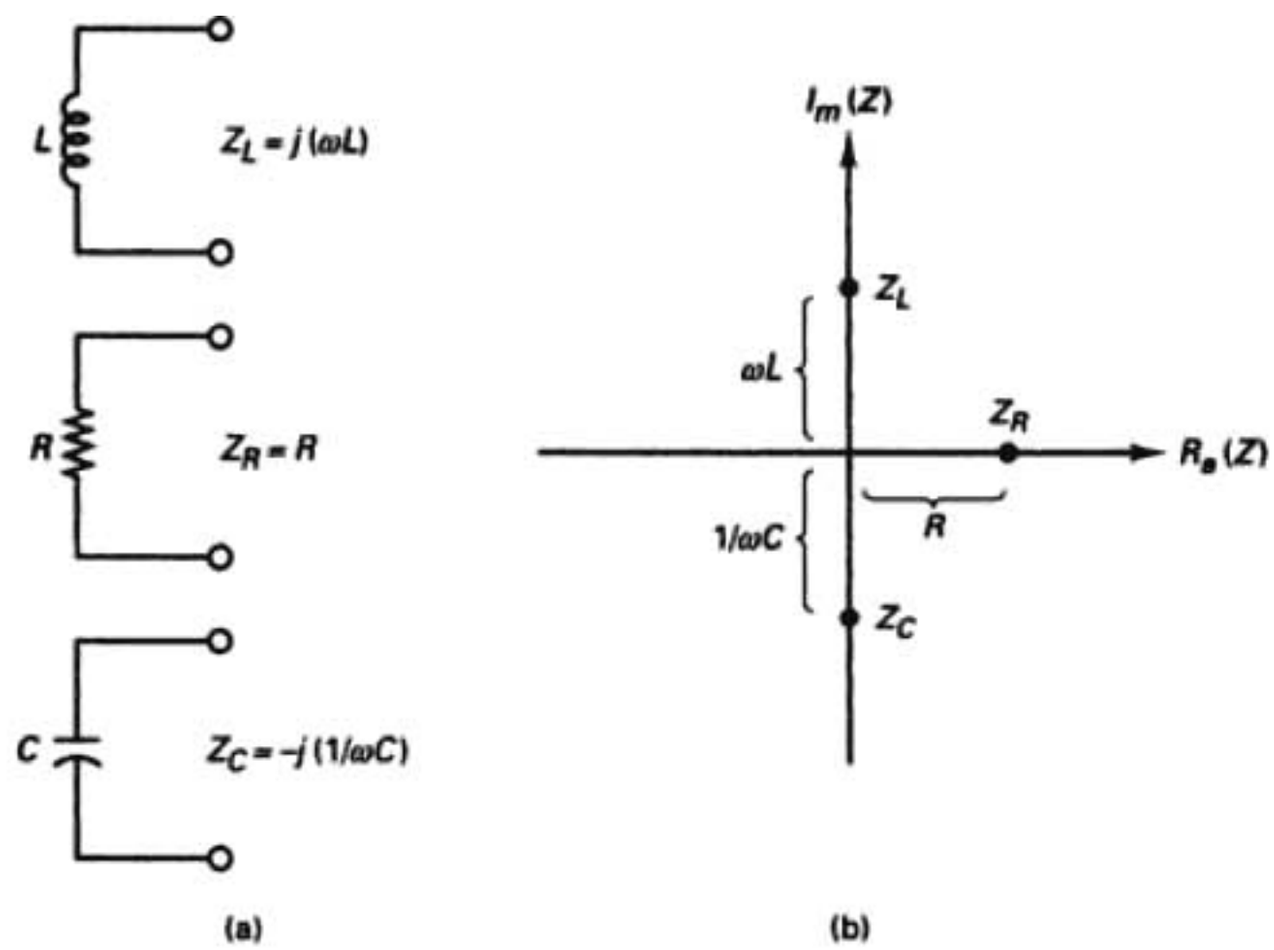


FIGURA A-2 (a) Elementos típicos de un circuito; (b) impedancias graficadas en un plano de coordenadas rectangulares; (Nota: ω es la frecuencia angular a la que se mide Z)

4. Relación de ondas estacionarias de voltaje o corriente.
   a. Relación de onda estacionaria.
   b. Límites de voltaje y corriente debidos a ondas estacionarias.

DEDUCCIÓN DEL DIAGRAMA DE SMITH

La impedancia de una línea de transmisión, Z, está formada por componentes *real* e *imaginario* de cualquier signo ($Z = R \pm jX$). La fig. A-2a muestra tres elementos típicos de circuito, y la fig. A-2b muestra su impedancia, graficada en un plano de coordenadas *rectangulares*. Todos los valores de Z que corresponden a redes pasivas se deben graficar en o hacia la derecha del eje imaginario del plano Z; esto se debe a que un componente real negativo implica que la red es capaz de suministrar energía. Para mostrar la impedancia de todas las redes pasivas posibles en una gráfica rectangular, esa gráfica se debería prolongar al infinito en tres direcciones, $+R$, $+jX$ y $-jX$. El diagrama de Smith supera esta limitación, porque se grafica el *coeficiente de reflexión* complejo,

$$\Gamma = \frac{z - 1}{z + 1} \tag{A-2}$$

en donde z es igual a la impedancia normalizada a la impedancia característica, es decir, $z = Z/Z_O$.

De acuerdo con la ecuación A-2, para todos los valores z de impedancia pasiva, la magnitud de Γ está entre 0 y 1. También, como $|\Gamma| \leq 1$, todo el lado derecho del plano z se puede trazar en un área circular en el plano Γ. El círculo que resulta tiene un radio $r = 1$ y su centro está en $\Gamma = 0$, lo cual corresponde a $z = 1$ o a $Z = Z_O$.

Líneas de $R_e(z)$ constante

La fig. A-3a muestra la gráfica rectangular de cuatro líneas de resistencia constante $R_e(z) = 0$, 0.5, 1 y 2. Por ejemplo, toda impedancia cuya parte real sea $R_e = 1$, estará en la recta $R = 1$. Las impedancias con componente reactivo positivo X_L quedarán arriba del eje real, mientras que

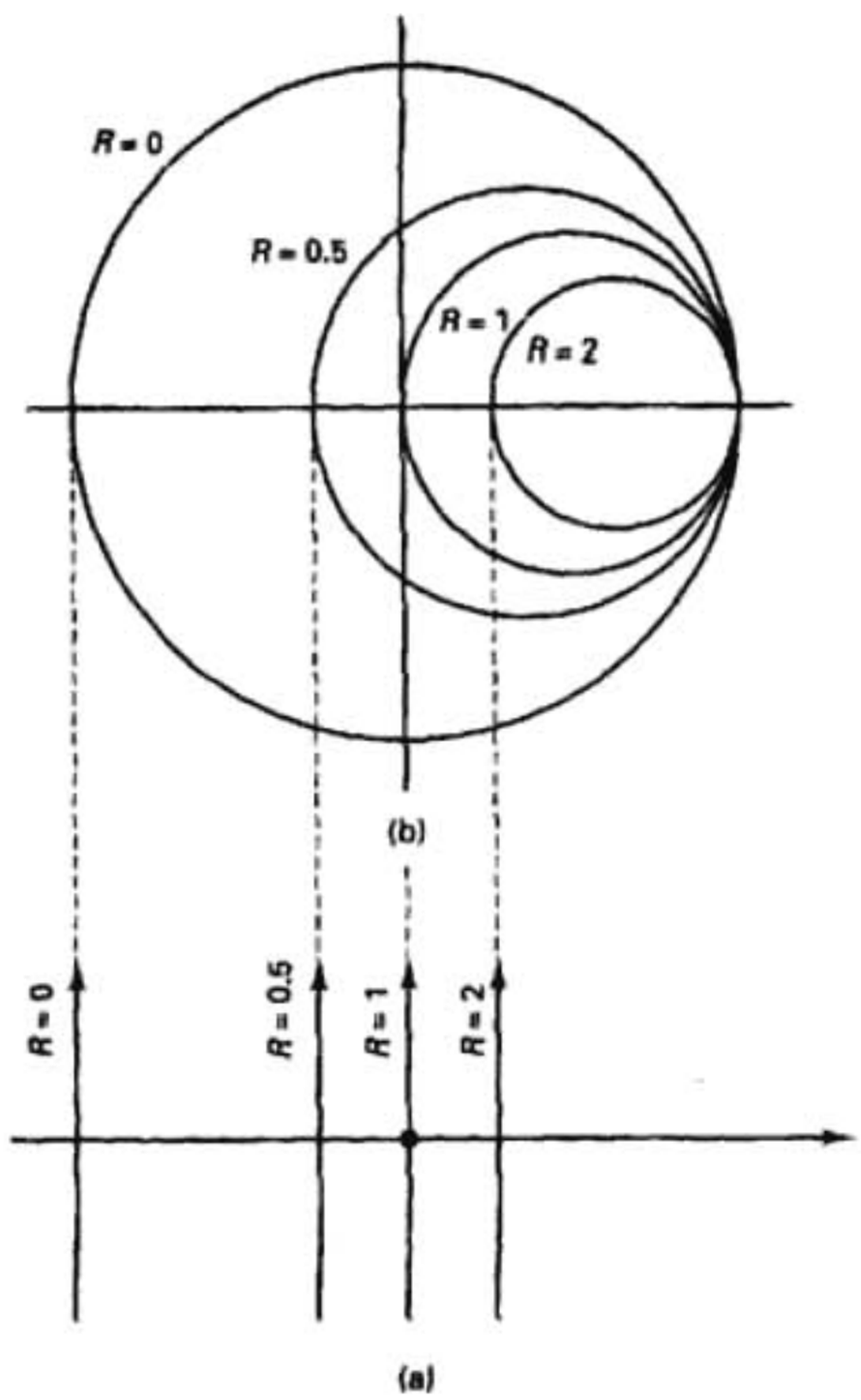


FIGURA A-3 (a) Gráfica en coordenadas rectangulares; (b) plano I

las impedancias con componente reactivo negativo X_C quedarán abajo del eje real. La fig. A-3b muestra los mismos cuatro valores trazados en el plano Γ. Ahora, $R_e(z)$ son círculos de $R_e(\Gamma)$. Sin embargo, las impedancias inductivas se siguen transfiriendo a la zona arriba del eje horizontal, y las capacitivas al área abajo del eje horizontal. La diferencia principal entre las dos gráficas es que en la gráfica circular las líneas ya no se prolongan al infinito. Todos los puntos al infinito se encuentran en el plano, a la distancia 1 a la derecha del origen. Esto implica que para $z = \infty$, sea real, inductiva o capacitiva, $\Gamma = 1$.

Líneas de *X(z)* constante

La fig. A-4a muestra la gráfica rectangular de tres líneas de reactancia inductiva constante, $X = 0.5$, 1 y 2, tres líneas de reactancia capacitiva constante, $X = -0.5$, -1 y -2, y una línea de reactancia cero, $X = 0$. La fig. A-4b muestra los mismos siete valores de jX, graficados en el plano Γ. Se aprecia que todos los valores de magnitud infinita se vuelven a encontrar en $\Gamma = 1$. Todo el plano rectangular z se enrosca hacia la derecha, y sus tres ejes, que antes se prolongaban al infinito, se encuentran en la intersección del círculo $\Gamma = 1$ con el eje horizontal.

Inversión de impedancia (admitancia)

La *admitancia*, Y, es la inversa matemática de Z, es decir, $Y = 1/Z$. Y, o cualquier número complejo, se puede determinar gráficamente con el diagrama de Smith sólo con graficar z en el plano complejo Γ, y a continuación girando 180° este punto respecto a $\Gamma = 0$. Al girar 180° cada punto de la gráfica se puede establecer un segundo conjunto de coordenadas, las coordenadas y, que es una imagen especular invertida del diagrama original. Véase la fig. A-5a. A veces se superponen las coordenadas de admitancia en la misma gráfica que las de impedancia. Véase la fig. A-5b. Si se usa la gráfica combinada se pueden leer en forma directa los valores de impedancia y admitancia, refiriéndose al conjunto adecuado de coordenadas.

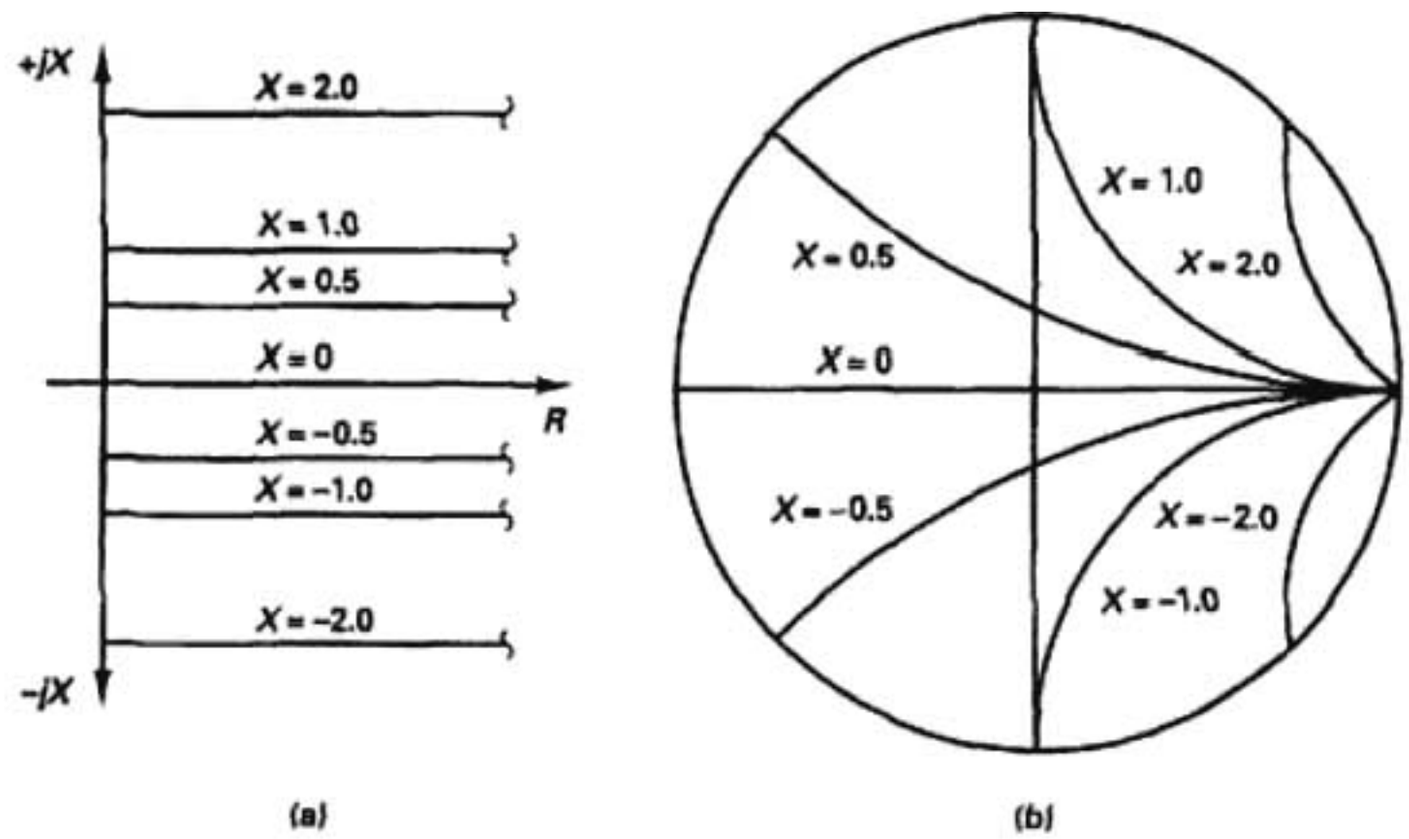


FIGURA A-4 (a) Gráfica en coordenadas rectangulares; (b) plano I

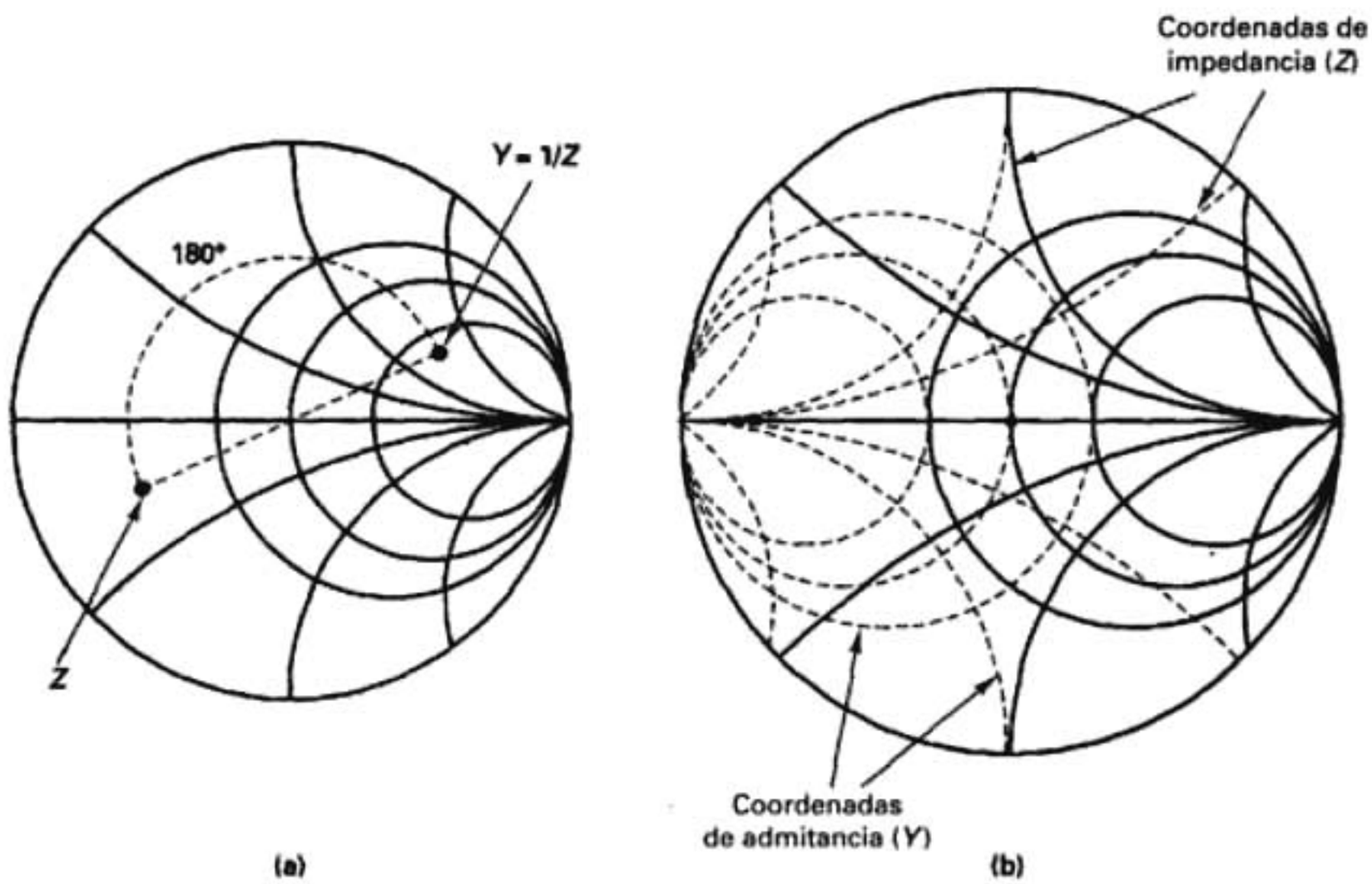


FIGURA A-5 Inversión de impedancia

Complejo conjugado

Se puede determinar el *complejo conjugado* con facilidad en el diagrama de Smith, sólo invirtiendo el signo del ángulo de Γ. En el diagrama de Smith se suele representar Γ en forma polar, y los ángulos son más negativos (retardo de fase) cuando giran en dirección de las manecillas del reloj respecto a la gráfica. Por consiguiente, 0° está en el extremo derecho del eje real, y $\pm 180°$ está en el extremo izquierdo. Por ejemplo, sea $\Gamma = 0.5 \angle +150°$. El complejo conjugado Γ^* es $0.5 \angle -150°$. En la fig. A-6 se ve que Γ^* se determina reflejando a Γ en el eje real.

GRAFICACIÓN DE LA IMPEDANCIA, ADMITANCIA Y SWR EN EL DIAGRAMA DE SMITH

Toda impedancia Z se puede graficar en el diagrama de Smith sólo *normalizando* su valor respecto a la impedancia característica, es decir, $z = Z/Z_O$, y graficando las partes real e imaginaria.

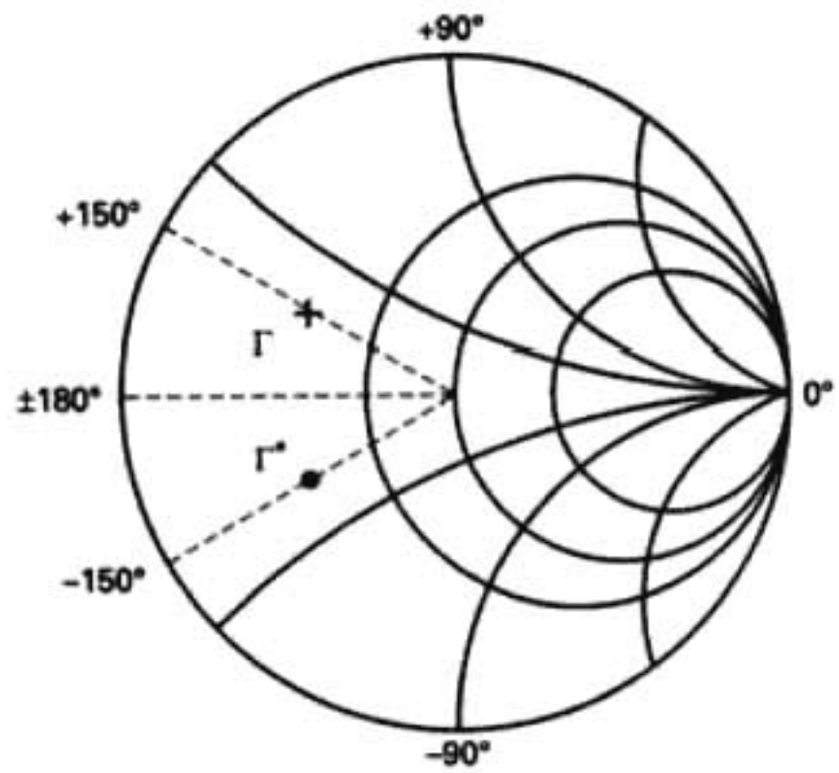


FIGURA A-6 Complejos conjugados

Por ejemplo, para una impedancia característica $Z_O = 50\ \Omega$ y una impedancia resistiva $Z = 25\ \Omega$, la impedancia normalizada z se calcula como sigue

$$z = \frac{Z}{Z_O} = \frac{25}{50} = 0.5$$

Como z es puramente resistiva, esta gráfica debe estar directamente sobre el eje horizontal, ya que $\pm jX = 0$. $Z = 25$ se grafica en la fig. A-7 en el punto A, es decir, $z = 0.5$. Al girar 180° la gráfica se obtiene el valor de admitancia normalizado, $y = 2$ (siendo $y = Y/Y_O$). En la fig. A-7 se grafica y en el punto B.

Como se dijo antes, una característica muy importante del diagrama de Smith es que cualquier línea sin pérdidas se puede representar por un círculo con origen en $1 \pm j0$, el centro de la gráfica, y con radio igual a la distancia entre el origen y la gráfica de impedancia. Por consiguiente, la *relación de onda estacionaria* (SWR, de *standing-wave ratio*) que corresponde a algún círculo determinado es igual al valor de Z/Z_O en el que el círculo cruza el eje horizontal al lado derecho de la gráfica. En consecuencia, para este ejemplo, SWR = 0.5 (ya que $Z/Z_O = 25/50 = 0.5$). También se debe notar que todo punto de impedancia o admitancia puede girar 180° tan sólo trazando una recta del punto, pasando por el centro de la gráfica, y terminando con la intersección con el círculo en el lado opuesto.

Para una impedancia característica $Z_O = 50$ y una carga inductiva $Z = +j25$, la impedancia normalizada z se determina como sigue

$$z = \frac{Z}{Z_O} = \frac{+jX}{Z_O} = \frac{+j25}{50} = +j0.5$$

Como z es puramente inductiva, su gráfica debe estar en el eje $R = 0$, que es el círculo exterior de la gráfica. El punto $z = +j0.5$ se grafica en la fig. A-8 en el punto A, y su admitancia $y = -j2$ se determina gráficamente tirando 180° en torno a la gráfica (punto B). En este ejemplo, SWR debe estar en el extremo derecho del eje horizontal, que se grafica en el punto C y corresponde a SWR = ∞, inevitable para una carga puramente reactiva. SWR se grafica en el punto C.

Para una impedancia compleja $Z = 25 + j25$, z se determina como sigue

$$z = \frac{25 + j25}{50} = 0.5 + j0.5$$

$$= 0.707\ \underline{/45^\circ}$$

En consecuencia, $$Z = 0.707\ \underline{/45^\circ} \times 50 = 35.35\ \underline{/45^\circ}$$

y $$Y = \frac{1}{35.35\ \underline{/45^\circ} = 0.02829\ \underline{/45^\circ}}$$

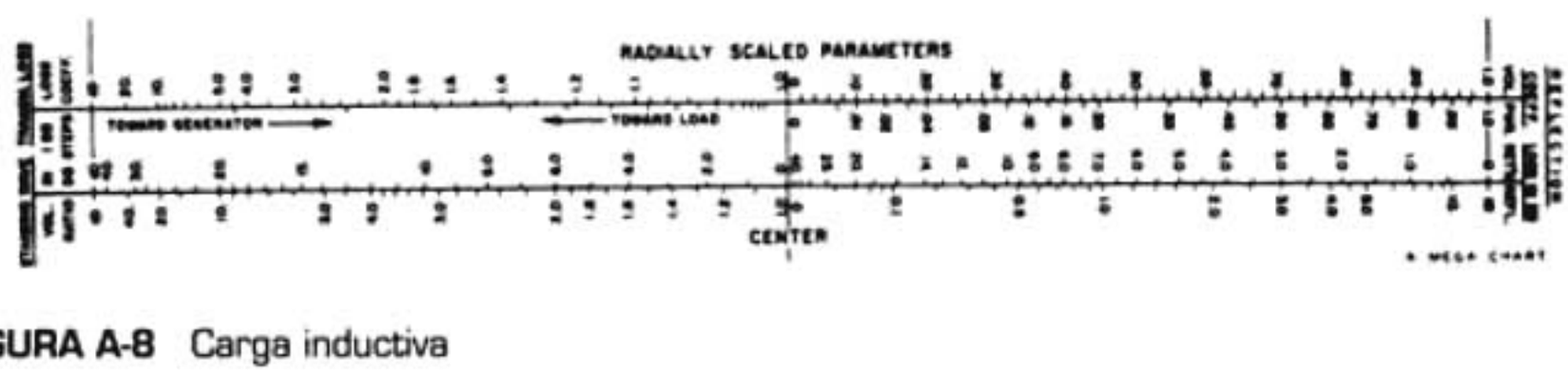


FIGURA A-8 Carga inductiva

distancia, en *longitudes de onda* (véase la fig. A-1). La escala externa muestra la distancia de la carga al generador, y aumenta en dirección de las manecillas del reloj, y la segunda escala muestra la distancia de la fuente a la carga y aumenta en dirección de las manecillas del reloj. Sin embargo, ninguna de las escalas indica necesariamente la posición de la fuente o de la carga. Una revolución completa (360°) representa una distancia de la mitad de una longitud de onda (0.5λ), y la mitad de una revolución (180°) representa una distancia de un cuarto de longitud de onda (0.25λ), etcétera.

Una línea de transmisión que termina en un circuito abierto tiene una impedancia en el extremo abierto que es puramente resistiva, e igual a infinito (capítulo 8). En el diagrama de Smith, este punto se grafica en el extremo derecho de la línea $X = 0$ (punto A, fig. A-10). Al avanzar hacia la fuente (generador), la impedancia de entrada se determina girando en torno a la gráfica en dirección de las manecillas del reloj. Al girar más y más en torno al círculo (al moverse

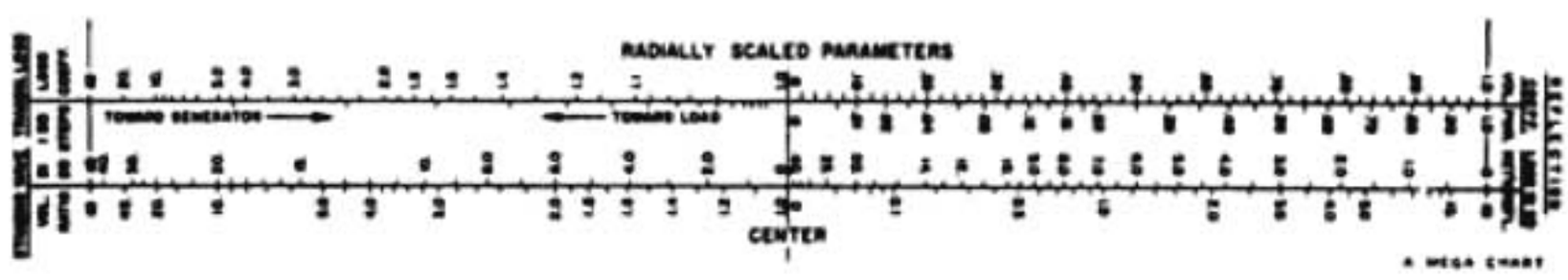


FIGURA A-9 Impedancia compleja

hacia el generador), la capacitancia disminuye hacia un valor normalizado de la unidad (es decir, $z = -j1$) a la distancia de un octavo de longitud de onda de la carga (punto C en la fig. A-10) y a un valor mínimo justo por debajo de un cuarto de longitud de onda. A la distancia de un cuarto de longitud de onda, la impedancia de entrada es puramente resistiva e igual a 0 Ω (punto B de la fig. A-10). Como se describe en el capítulo 8, hay una inversión de impedancia cada cuarto de longitud de onda, en una línea de transmisión. Al rebasar un cuarto de longitud de onda, la impedancia se transforma en inductiva y mínima; a continuación crece hasta un valor normalizado de unidad (es decir, $z = +j1$) a una distancia de tres octavos de longitud de onda de la carga (punto D en la fig. A-10) y un valor máximo justo antes de media longitud de onda. A la distancia de media longitud de onda, la impedancia de entrada de nuevo es puramente resistiva e igual a infinito (regreso al punto A de la fig. A-10). Los resultados de este análisis son idénticos a los encontrados con el análisis fasorial en el capítulo 8, graficados en la fig. 8-20.

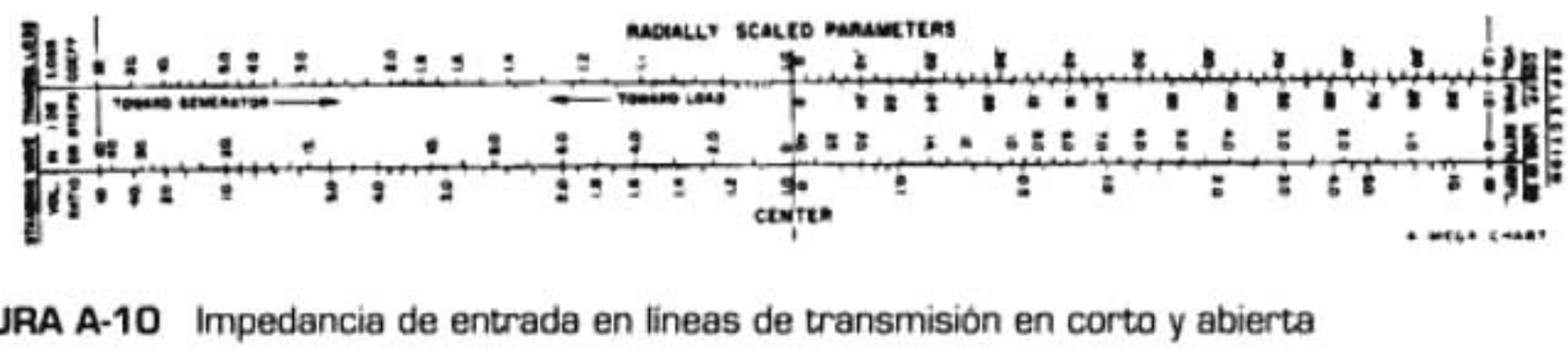


FIGURA A-10 Impedancia de entrada en líneas de transmisión en corto y abierta

Se puede hacer un análisis semejante con una línea de transmisión terminada en cortocircuito, aunque se obtienen variaciones de impedancia opuestas a las que había con un circuito abierto. En la carga, la impedancia de entrada es puramente resistiva, igual a 0. Por consiguiente, la carga está en el punto *B* de la fig. A-10, y el punto *A* representa una distancia de un cuarto de longitud de onda de la carga. El punto *D* está a la distancia de un octavo de longitud de onda de la carga, y el punto *C*, a tres octavos de longitud de onda. Los resultados de ese análisis son idénticos a los que se obtuvieron con fasores en el capítulo 8, graficados en la fig. 8-21.

Para una línea de transmisión terminada en una carga puramente resistiva distinta de Z_O, el análisis con diagrama de Smith es muy parecido al proceso descrito en la sección anterior. Por ejemplo, para una impedancia de carga $Z_L = 37.5\ \Omega$ resistiva, y una impedancia característica de línea de transmisión $Z_O = 75\ \Omega$, la impedancia de entrada a varias distancias de la carga se determina como sigue

COORDENADAS DE IMPEDANCIA O ADMITANCIA

FIGURA A-11 Cálculos de impedancia de entrada

1. La impedancia normalizada de carga z es

$$z = \frac{Z_L}{Z_O} = \frac{37.5}{75} = 0.5$$

2. Se grafica $z = 0.5$ en la gráfica de Smith (punto A de la fig. A-11). Se traza un círculo que pase por el punto A, con su centro ubicado en la intersección del círculo $R = 1$ y el arco $x = 0$.

3. Se lee SWR en forma directa en la intersección del círculo $z = 0.5$ con la recta $X = 0$ en el lado derecho (punto F). En este caso, SWR = 2. Con el círculo de impedancia se pueden describir todas las impedancias a lo largo de la línea de transmisión. Por consiguiente, la impedancia de entrada (Z_i) a la distancia de 0.125λ de la carga se determina proyectando el círculo z

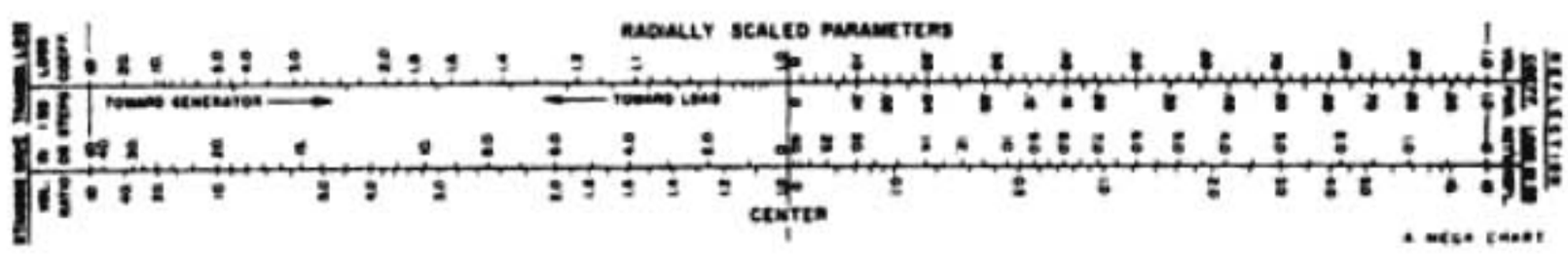


FIGURA A-12 Diagrama de Smith para el ejemplo A-1

3. Se proyecta el punto A hasta la escala externa (punto B). La impedancia característica de una línea de transmisión es puramente resistiva. En consecuencia, si el transformador de cuarto de onda está a una distancia de la carga en la que la impedancia de entrada sea puramente resistiva, el transformador puede acoplar la línea de transmisión a la carga. Hay dos puntos sobre el círculo de impedancia donde la impedancia de entrada es puramente resistiva: donde el círculo intersecta la recta $X = 0$ (puntos C y D de la fig. A-13). Por consiguiente, la distancia de la carga a un punto donde la impedancia de entrada sea puramente resistiva se determina tan sólo con calcular la distancia en longitudes de onda del punto B de la fig. A-13 a cualquiera de los puntos C o D, la que resulte menor. La distancia del punto B al punto C es

punto C	0.250λ
$-$punto B	-0.192λ
distancia	0.058λ

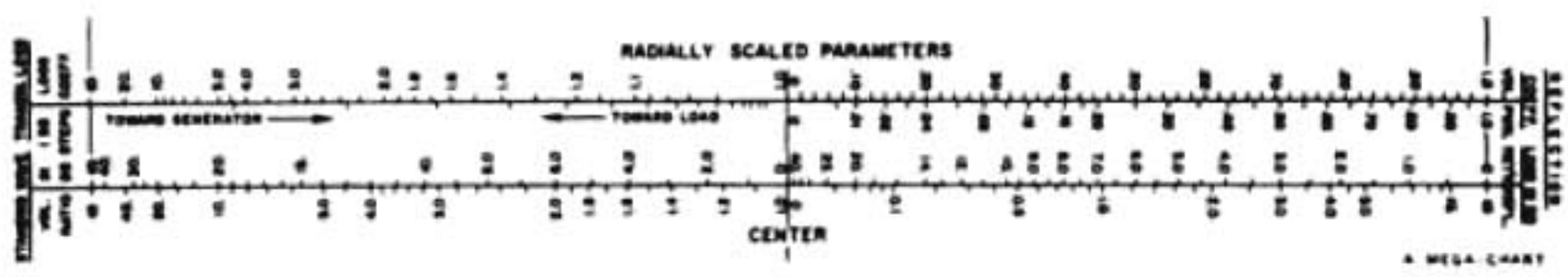


FIGURA A-13 Diagrama de Smith, transformador de un cuarto de onda

Si se instala un transformador de cuarto de longitud de onda a 0.058λ de la carga, la impedancia de entrada se lee en forma directa de la fig. A-13, $z_i = 2.4$ (punto C).

4. Se observa que 2.4 también es la relación SWR de la línea sin compensar, y se lee en forma directa en la gráfica.

5. La impedancia real de entrada es $Z_i = 50(2.4) = 120\ \Omega$. La impedancia característica del transformador de cuarto de onda se calcula con la ecuación 8-32.

$$Z'_O = \sqrt{Z_O Z_i} = \sqrt{50 \times 120} = 77.5\ \Omega$$

Así, si se inserta un cuarto de longitud de onda de una línea de transmisión de 77.5 Ω a 0.058λ de la carga, la línea queda acoplada. Se debe observar que un transformador de cuarto de onda no elimina por completo las ondas estacionarias de la línea de transmisión. Sólo las elimi-

SISTEMAS DE COMUNICACIONES ELECTRÓNICAS

CUARTA EDICIÓN

WAYNE TOMASI

Sistemas de comunicaciones electrónicas, cuarta edición, presenta una descripción detallada del campo de las comunicaciones electrónicas. Explica al lector los conceptos básicos de los sistemas analógicos convencionales de comunicaciones electrónicas y amplía sus conocimientos describiéndole los sistemas más modernos de comunicaciones digitales, por fibra óptica, por microondas, satelitales y telefónicos celulares y PCS.

Se han actualizado todos los temas para presentar la última información disponible en este campo de cambios tan rápidos, y se han agregado secciones completamente nuevas, que describen los siguientes temas de importancia:

- Envolventes de AM producidas por señales no senoidales complejas.
- Modulación de amplitud en cuadratura.
- Limitadores y supresores de ruido.
- Mediciones alternativas de señal a ruido.
- Multiplexado por banda lateral única con supresión de la portadora y multiplexado por división de frecuencia.
- Multiplexado por banda lateral doble y multiplexado por cuadratura.
- Microbanda y línea de cinta.
- Fuentes luminosas, potencia óptica, fuentes ópticas y gastos de enlace.
- Codificación Trellis.
- Recomendaciones de CCITT para módem.
- Velocidad de línea PCM.
- Formato de supertrama extendido y multiplexado por división de longitud de onda.
- Leyes de Kepler, órbitas de Clarke y límites de visibilidad.

Además, se agregó un capítulo totalmente nuevo, para explicar los sistemas telefónicos móviles (incluyendo AMPS, PSC y PCSS).

OTRAS OBRAS DE INTERÉS PUBLICADAS POR **PEARSON:**

BOYLESTAD: *Electrónica: Teoría de circuitos*, sexta edición
OPPENHEIM: *Señales y sistemas*, segunda edición
PROAKIS: *Tratamiento digital de señales*, tercera edición
RASHID: *Electrónica de potencia*, segunda edición
TOCCI y WIDMER: *Sistemas digitales*, octava edición

ISBN 970-26-0316-1
9 789702 603160 90000

www.pearsonedlatino.com

This page wasn't downloaded by GBD