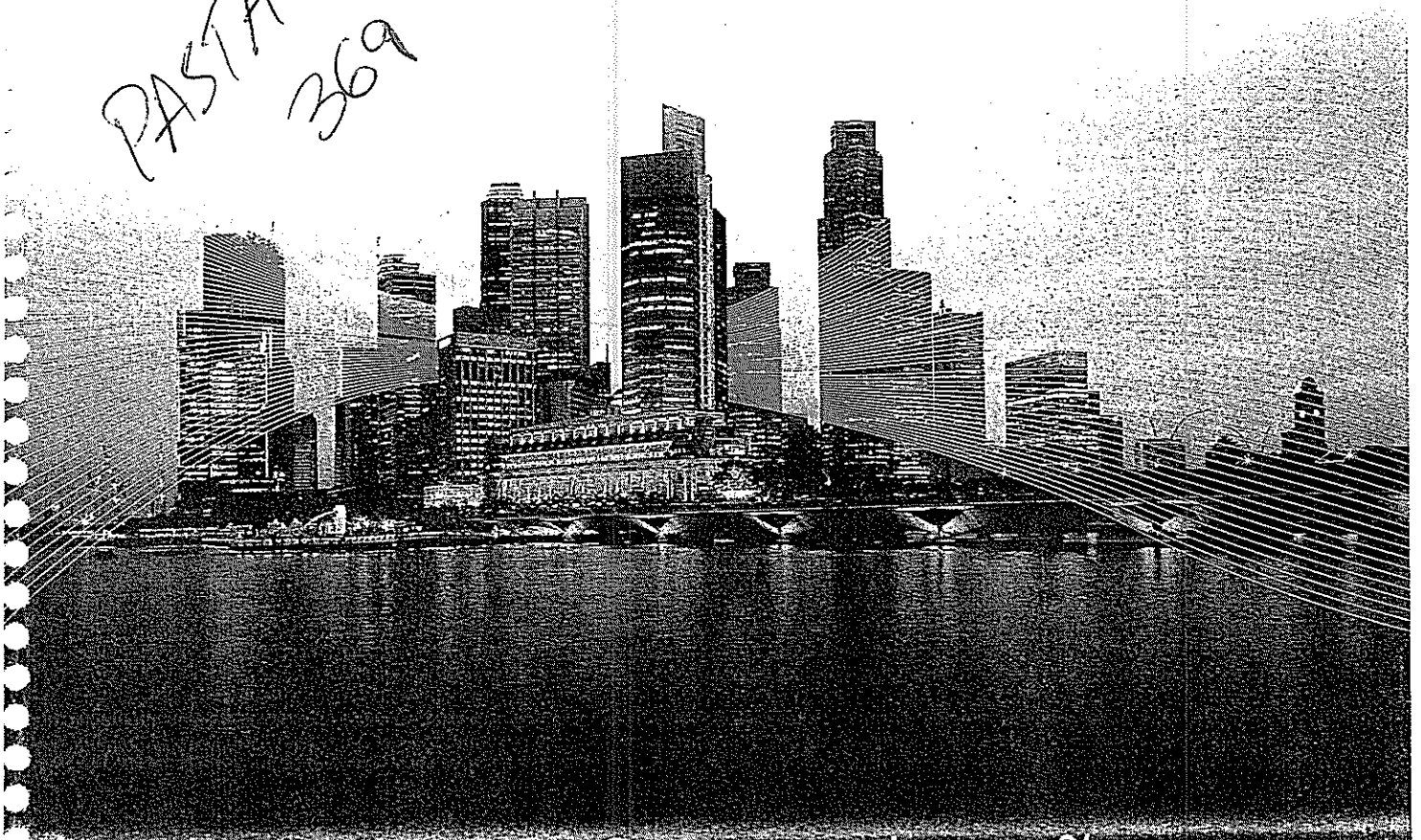
QUARTA EDIÇÃO

PASTA 369



Sistemas de Comunicações Analógicos e Digitais Modernos

B.P. Lathi • Zhi Ding

genio
GEN | Informação Online

gen | LTC

Esta quarta edição internacional de ***Sistemas de Comunicações Analógicos e Digitais Modernos*** contém os mais recentes avanços tecnológicos e suas aplicações em sistemas de comunicações, integrando conceitos matemáticos com o software MATLAB. Os estudantes e os leitores interessados podem usar novas ferramentas de ensino e realizar experimentos e práticas na tela do computador.

A obra abrange um vasto espectro de conteúdos, como sinais e sistemas, modulação analógica, modulação digital por codificação de pulsos e uma visão geral da teoria da informação e dos princípios de códigos corretores de erros, entre outros. Apresenta sistemas com múltiplas entradas e múltiplas saídas – com vasta aplicação comercial – e sistemas celulares, redes locais wireless e serviços de internet com fio. Os canais seletivos em frequência e sistemas OFDM também são abordados, o que assinala o valor das novas tecnologias, como espalhamento espectral e multiplexação por divisão em frequências ortogonais.

Ao buscar atender às crescentes demandas tecnológicas, ***Sistemas de Comunicações Analógicos e Digitais Modernos*** consolida-se como um livro ímpar na área de telecomunicações, apresentando o texto de forma objetiva, didática e atualizada.

gen | LTC

www.grupogen.com.br
http://gen-io.grupogen.com.br

ISBN 978-85-216-2027-3

9 788521 620273

SISTEMAS DE COMUNICAÇÕES ANALÓGICOS E DIGITAIS MODERNOS

Quarta Edição

abdr
ASSOCIAÇÃO BRASILEIRA DE DIREITOS REPROGRÁFICOS
Respeite o direito autoral

O GEN | Grupo Editorial Nacional reúne as editoras Guanabara Koogan, Santos, Roca, AC Farmacêutica, Forense, Método, LTC, E.P.U. e Forense Universitária, que publicam nas áreas científica, técnica e profissional.

Essas empresas, respeitadas no mercado editorial, construíram catálogos inigualáveis, com obras que têm sido decisivas na formação acadêmica e no aperfeiçoamento de várias gerações de profissionais e de estudantes de Administração, Direito, Enfermagem, Engenharia, Fisioterapia, Medicina, Odontologia, Educação Física e muitas outras ciências, tendo se tornado sinônimo de seriedade e respeito.

Nossa missão é prover o melhor conteúdo científico e distribuí-lo de maneira flexível e conveniente, a preços justos, gerando benefícios e servindo a autores, docentes, livreiros, funcionários, colaboradores e acionistas.

Nosso comportamento ético incondicional e nossa responsabilidade social e ambiental são reforçados pela natureza educacional de nossa atividade, sem comprometer o crescimento contínuo e a rentabilidade do grupo.

SISTEMAS DE COMUNICAÇÕES ANALÓGICOS E DIGITAIS MODERNOS

Quarta Edição

B. P. Lathi
Professor Emérito da California State University – Sacramento

Zhi Ding
Professor da University of California – Davis

Tradução
J. R. Souza, Ph.D.
Professor Adjunto da Universidade do Estado do Rio de Janeiro (UERJ)

Revisão Técnica
José Alexandre Nalon
Mestre em Engenharia Elétrica pela UNICAMP
Professor do Centro Universitário Salesiano de São Paulo (UNISAL)

gen | LTC

Os autores e a editora empenharam-se para citar adequadamente e dar o devido crédito a todos os detentores dos direitos autorais de qualquer material utilizado neste livro, dispondo-se a possíveis acertos caso, inadvertidamente, a identificação de algum deles tenha sido omitida.

Não é responsabilidade da editora nem dos autores a ocorrência de eventuais perdas ou danos a pessoas ou bens que tenham origem no uso desta publicação.

Apesar dos melhores esforços dos autores, do tradutor, da editora e dos revisores, é inevitável que surjam erros no texto. Assim, são bem-vindas as comunicações de usuários sobre correções ou sugestões referentes ao conteúdo ou ao nível pedagógico que auxiliem o aprimoramento de edições futuras. Os comentários dos leitores podem ser encaminhados à LTC – **Livros Técnicos e Científicos Editora** pelo e-mail ltc@grupogen.com.br.

MODERN DIGITAL AND ANALOG COMMUNICATION SYSTEMS, INTERNATIONAL FOURTH EDITION
Copyright © 1983 by CBS College Publishing; © 1989 by B. P. Lathi & Saunders College Publishing, a division of Holt, Rinehart, and Winston, Inc.; © 1995, 1998, 2010 by B. P. Lathi.

All rights reserved.

MODERN DIGITAL AND ANALOG COMMUNICATION SYSTEMS, INTERNATIONAL FOURTH EDITION was originally published in English in 2009. This translation is published by arrangement with Oxford University Press.

MODERN DIGITAL AND ANALOG COMMUNICATION SYSTEMS, INTERNATIONAL FOURTH EDITION foi editada originalmente em inglês em 2009. Esta tradução é uma publicação por acordo com a Oxford University Press.

Direitos exclusivos para a língua portuguesa
Copyright © 2012 by
LTC – Livros Técnicos e Científicos Editora Ltda.
Uma editora integrante do GEN | Grupo Editorial Nacional

Reservados todos os direitos. É proibida a duplicação ou reprodução deste volume, no todo ou em parte, sob quaisquer formas ou por quaisquer meios (eletrônico, mecânico, gravação, fotocópia, distribuição na internet ou outros), sem permissão expressa da editora.

Travessa do Ouvidor, 11
Rio de Janeiro, RJ – CEP 20040-040
Tels.: 21-3543-0770 / 11-5080-0770
Fax: 21-3543-0896
ltc@grupogen.com.br
www.ltceditora.com.br
Capa: Dan Niver Foto de Capa: © Robert Churchill / IStockphoto.com
Editoração Eletrônica: K2 design

CIP-BRASIL. CATALOGAÇÃO-NA-FONTE
SINDICATO NACIONAL DOS EDITORES DE LIVROS, RJ.

L379s

Lathi, B. P. (Bhagwandas Pannalal), 1933-
Sistemas de comunicações analógicos e digitais modernos / B. P. Lathi, Zhi Ding ; tradução J. R. Souza ; revisão técnica José Alexandre Nalon. - Rio de Janeiro : LTC, 2012.
il. ; 28 cm
Tradução de: Modern digital and analog communication systems, 4th ed
Apêndice
Inclui bibliografia e índice
ISBN 978-85-216-2027-3
1. Sistemas de telecomunicação. 2. Comunicações digitais. I. Ding, Zhi, 1962-. II. Título.

12-0450. CDD: 621.382
CDU: 621.39

SUMÁRIO GERAL

Prefácio xix

1 Introdução 1

2 Sinais e Espaço de Sinais 19

3 Análise e Transmissão de Sinais 59

4 Modulações e Demodulações em Amplitude 130

5 Modulações e Demodulações em Ângulo 187

6 Amostragem e Conversão Analógico-Digital 232

7 Princípios de Transmissão de Dados Digitais 302

8 Fundamentos da Teoria da Probabilidade 364

9 Processos Aleatórios e Análise Espectral 421

10 Análise de Desempenho de Sistemas de Comunicações Digitais 468

11 Comunicações por Espalhamento Espectral 567

12 Comunicações Digitais com Canais Sujeitos a Distorção Linear 617

13 Introdução à Teoria da Informação 681

14 Códigos Corretores de Erro 743

A Ortogonalidade de Alguns Conjuntos de Sinais 810

B Desigualdade de Cauchy-Schwarz 812

C Ortogonalização de Gram-Schmidt de um Conjunto de Vetores 814

D Propriedades Básicas Matrizes e Operações com Matrizes 817

E Miscelâneas 822

Índice 826

SUMÁRIO

PREFÁCIO xix

1 INTRODUÇÃO 1

1.1 SISTEMAS DE COMUNICAÇÃO 1

1.2 MENSAGENS ANALÓGICAS E MENSAGENS DIGITAIS 3

1.2.1 Imunidade de Sinais Digitais ao Ruído 4

1.2.2 Viabilidade de Repetidores Regenerativos sem Distorção 5

1.2.3 Conversão Analógico-Digital (A/D) 6

1.2.4 Modulação por Codificação de Pulsos — Uma Representação Digital 7

1.3 EFEITO, RELAÇÃO SINAL-RUÍDO E CAPACIDADE DO CANAL 8

1.3.1 Largura de Banda e Potência de Sinal 8

1.3.2 Capacidade do Canal e Taxa de Dados 9

1.4 MODULAÇÃO E DETECÇÃO 10

1.4.1 Facilidade de Radiação/Transmissão 11

1.4.2 Transmissão Simultânea de Múltiplos Sinais — Multiplexação 12

1.4.3 Demodulação 12

1.5 CODIFICAÇÕES DIGITAIS DE FONTE E PARA CORREÇÃO DE ERROS 12

1.6 BREVE REVISÃO HISTÓRICA DE TELECOMUNICAÇÕES MODERNAS 14

2 SINAIS E ESPAÇO DE SINAIS 19

2.1 TAMANHO DE UM SINAL 19

2.2 CLASSIFICAÇÃO DE SINAIS 21

2.2.1 Sinais em Tempo Contínuo e Sinais em Tempo Discreto 21

2.2.2 Sinais Analógicos e Sinais Digitais 22
2.2.3 Sinais Periódicos e Sinais Aperiódicos 23
2.2.4 Sinais de Energia e Sinais de Potência 24
2.2.5 Sinais Determinísticos e Sinais Aleatórios 24

2.3 SINAL IMPULSO UNITÁRIO 24

2.4 ANALOGIA ENTRE SINAIS E VETORES 26
2.4.1 Componente de um Vetor na Direção de Outro Vetor 27
2.4.2 Decomposição de um Sinal e Componentes de um Sinal 28
2.4.3 Espaço Complexo de Sinais e Ortogonalidade 30
2.4.4 Energia da Soma de Sinais Ortogonais 31

2.5 CORRELAÇÃO DE SINAIS 32
2.5.1 Funções de Correlação 33
2.5.2 Função de Autocorrelação 34

2.6 CONJUNTO ORTOGONAL DE SINAIS 34
2.6.1 Espaço Vetorial Ortogonal 34
2.6.2 Espaço Ortogonal de Sinais 35
2.6.3 Teorema de Parseval 36

2.7 SÉRIE DE FOURIER EXPONENCIAL 37

2.8 EXERCÍCIOS COM O MATLAB 43

3 ANÁLISE E TRANSMISSÃO DE SINAIS 59

3.1 REPRESENTAÇÃO DE SINAIS APERIÓDICOS ATRAVÉS DA INTEGRAL DE FOURIER 59

3.2 TRANSFORMADAS DE ALGUMAS FUNÇÕES ÚTEIS 66

3.3 ALGUMAS PROPRIEDADES DA TRANSFORMADA DE FOURIER 71
3.3.1 Dualidade Tempo-Frequência 71
3.3.2 Propriedade de Dualidade 73
3.3.3 Propriedade de Dilatação no Tempo 74
3.3.4 Propriedade de Translação no Tempo 76
3.3.5 Propriedade de Translação na Frequência 78
3.3.6 Teorema da Convolução 81
3.3.7 Diferenciação e Integração no Domínio do Tempo 83

3.4 TRANSMISSÃO DE SINAL EM UM SISTEMA LINEAR 85
3.4.1 Distorção do Sinal Durante a Transmissão 86
3.4.2 Transmissão sem Distorção 87

3.5 FILTROS IDEAIS *VERSUS* FILTROS PRÁTICOS 89

3.6 DISTORÇÃO DE SINAL EM UM CANAL DE COMUNICAÇÃO 91
3.6.1 Distorção Linear 91
3.6.2 Distorção Causada por Não Linearidades do Canal 93
3.6.3 Distorção Causada por Efeitos de Multipercurso 95
3.6.4 Canais com Desvanecimento 96
3.7 ENERGIA E DENSIDADE ESPECTRAL DE ENERGIA DE SINAIS 96
3.7.1 Teorema de Parseval 97
3.7.2 Densidade Espectral de Energia (ESD) 97
3.7.3 Largura de Banda Essencial de um Sinal 98
3.7.4 Energia de Sinais Modulados 101
3.7.5 Função de Autocorrelação Temporal e Densidade Espectral de Energia 102
3.8 POTÊNCIA E DENSIDADE ESPECTRAL DE POTÊNCIA DE SINAIS 104
3.8.1 Densidade Espectral de Potência (PSD) 104
3.8.2 Função de Autocorrelação Temporal de Sinais de Potência 105
3.8.3 Densidades Espectrais de Potência de Entrada e de Saída 109
3.8.4 PSD de Sinais Modulados 110
3.9 CÁLCULO NUMÉRICO DA TRANSFORMADA DE FOURIER: A DFT 110
3.10 EXERCÍCIOS COM O MATLAB 115

4 MODULAÇÕES E DEMODULAÇÕES EM AMPLITUDE 130
4.1 COMUNICAÇÃO EM BANDA BASE *VERSUS* COMUNICAÇÃO POR PORTADORA 130
4.2 MODULAÇÃO EM AMPLITUDE COM BANDA LATERAL DUPLA 132
4.3 MODULAÇÃO EM AMPLITUDE (AM) 141
4.4 MODULAÇÕES EM AMPLITUDE COM EFICIÊNCIA DE LARGURA DE BANDA 147
4.5 MODULAÇÕES EM AMPLITUDE: BANDA LATERAL VESTIGIAL (VSB) 155
4.6 SINCRONIZAÇÃO DA PORTADORA LOCAL 159
4.7 MULTIPLEXAÇÃO POR DIVISÃO EM FREQUÊNCIA (FDM) 160

4.8 MALHA DE CAPTURA DE FASE (*PHASE LOCKED LOOP — PLL*) E ALGUMAS APLICAÇÕES 161

4.9 EXERCÍCIOS COM O MATLAB 168

5 MODULAÇÕES E DEMODULAÇÕES EM ÂNGULO 187

5.1 MODULAÇÃO NÃO LINEAR 187

5.2 LARGURA DE BANDA DE ONDAS MODULADAS EM ÂNGULO 193

5.3 GERAÇÃO DE ONDAS FM 205

5.4 DEMODULAÇÃO DE SINAIS FM 213

5.5 EFEITOS DE DISTORÇÃO NÃO LINEAR E INTERFERÊNCIA 216

5.6 RECEPTORES ANALÓGICOS AM/FM SUPER-HETERÓDINOS 221

5.7 SISTEMAS DE DIFUSÃO FM 222

5.8 EXERCÍCIOS COM O MATLAB 224

6 AMOSTRAGEM E CONVERSÃO ANALÓGICO-DIGITAL 232

6.1 TEOREMA DA AMOSTRAGEM 232

6.1.1 Reconstrução de Sinais a partir de Amostras Uniformes 234

6.1.2 Questões Práticas Relativas à Amostragem e à Reconstrução de Sinais 239

6.1.3 Máxima Taxa de Informação: Duas Porções de Informação por Segundo por Hertz 242

6.1.4 Análise de Amostragem Prática Não Ideal 243

6.1.5 Algumas Aplicações do Teorema da Amostragem 247

6.2 MODULAÇÃO POR CODIFICAÇÃO DE PULSO (PCM) 248

6.2.1 Vantagens de Comunicação Digital 249

6.2.2 Quantização 250

6.2.3 Princípio da Taxação Progressiva: Quantização Não Uniforme 253

6.2.4 Largura de Banda de Transmissão e SNR de Saída 257

6.3 TELEFONIA DIGITAL: PCM EM SISTEMAS DE PORTADORAS 259

6.4 MULTIPLEXAÇÃO DIGITAL 263

6.4.1 Formato de Sinal 264

6.4.2 Canais Assíncronos e Preenchimento de Bits 265
6.4.3 Hierarquia Digital Plesiócrona (Quase Síncrona) 266

6.5 MODULAÇÃO POR CODIFICAÇÃO DE PULSO DIFERENCIAL (DPCM) 268

6.6 PCM DIFERENCIAL ADAPTATIVA (ADPCM) 271

6.7 MODULAÇÃO DELTA 272

6.8 VOCODERS E COMPRESSÃO DE VÍDEO 278
6.8.1 Vocoders com Codificação de Predição Linear 278

6.9 EXERCÍCIOS COM O MATLAB 287

7 PRINCÍPIOS DE TRANSMISSÃO DE DADOS DIGITAIS 302

7.1 SISTEMAS DE COMUNICAÇÃO DIGITAL 302
7.1.1 Fonte 302
7.1.2 Codificador de Linha 303
7.1.3 Multiplexador 304
7.1.4 Repetidor Regenerador 304

7.2 CODIFICAÇÃO DE LINHA 305
7.2.1 PSD de Diversos Códigos de Linha 305
7.2.2 Sinalização Polar 309
7.2.3 Construção de um Nulo DC na PSD via Formatação de Pulso 311
7.2.4 Sinalização On-Off 312
7.2.5 Sinalização Bipolar 314

7.3 FORMATAÇÃO DE PULSO 317
7.3.1 Interferências Intersimbólicas (ISI) e Efeitos 318
7.3.2 Primeiro Critério de Nyquist para ISI Nula 318
7.3.3 ISI Controlada ou Sinalização de Resposta Parcial 323
7.3.4 Exemplo de Pulso Duobinário 324
7.3.5 Relações de Pulsos entre Sinalizações com ISI Zero, Duobinária e Duobinária Modificada 325
7.3.6 Detecção de Sinalização Duobinária e Codificação Diferencial 326
7.3.7 Geração de Pulsos 328

7.4 EMBARALHAMENTO (*SCRAMBLING*) DE DADOS 329

7.5 RECEPTORES DIGITAIS E REPETIDORES REGENERATIVOS 331
7.5.1 Equalizadores 332
7.5.2 Extração de Temporização 336

7.5.3 Detecção de Erro 337

7.6 DIAGRAMAS DE OLHO: UMA FERRAMENTA ÚTIL 338

7.7 PAM: SINALIZAÇÃO *M*-ÁRIA EM BANDA BASE PARA TAXA DE DADOS MAIS ELEVADA 341

7.8 SISTEMAS DIGITAIS COM PORTADORA 344

7.8.1 Modulações Binárias Básicas com Portadora 344

7.8.2 PSD de Modulação Digital com Portadora 346

7.8.3 Relações entre Modulações Analógica e Digital com Portadora 348

7.8.4 Demodulação 348

7.9 MODULAÇÃO DIGITAL *M*-ÁRIA COM PORTADORA 351

7.10 EXERCÍCIOS COM O MATLAB 357

8 FUNDAMENTOS DA TEORIA DA PROBABILIDADE 364

8.1 CONCEITO DE PROBABILIDADE 364

8.2 VARIÁVEIS ALEATÓRIAS 378

8.3 MÉDIAS (VALORES MÉDIOS) ESTATÍSTICAS (OS) 395

8.4 CORRELAÇÃO 404

8.5 ESTIMAÇÃO QUADRÁTICA MÉDIA LINEAR 407

8.6 SOMA DE VARIÁVEIS ALEATÓRIAS 410

8.7 TEOREMA DO LIMITE CENTRAL 412

9 PROCESSOS ALEATÓRIOS E ANÁLISE ESPECTRAL 421

9.1 DE VARIÁVEL ALEATÓRIA A PROCESSO ALEATÓRIO 421

9.2 CLASSIFICAÇÃO DE PROCESSOS ALEATÓRIOS 426

9.3 DENSIDADE ESPECTRAL DE POTÊNCIA 429

9.4 PROCESSOS ALEATÓRIOS MÚLTIPLOS 442

9.5 TRANSMISSÃO DE PROCESSOS ALEATÓRIOS POR MEIO DE SISTEMAS LINEARES 443

9.6 APLICAÇÃO: FILTRAGEM ÓTIMA (FILTRO DE WIENER-HOPF) 447

9.7 APLICAÇÃO: ANÁLISE DE DESEMPENHO DE SISTEMAS ANALÓGICOS EM BANDA-BASE 449

9.8 APLICAÇÃO: SISTEMAS ÓTIMOS DE PRÉ-ÊNFASE/DEÊNFASE 451

9.9 PROCESSOS ALEATÓRIOS PASSA-FAIXA 454

10 ANÁLISE DE DESEMPENHO DE SISTEMAS DE COMUNICAÇÕES DIGITAIS 468

10.1 DETECTOR LINEAR ÓTIMO PARA SINALIZAÇÃO POLAR BINÁRIA 468

10.1.1 Limiar de Detecção Binária 469

10.1.2 Filtro Receptor Ótimo – Filtro Casado 470

10.2 SINALIZAÇÃO BINÁRIA GENÉRICA 474

10.2.1 Análise de Receptor Linear Ótimo 474

10.2.2 Análise de Desempenho de Sistemas Binários Genéricos 477

10.3 RECEPTORES COERENTES PARA MODULAÇÕES DIGITAIS COM PORTADORA 480

10.4 ANÁLISE DE DETECÇÃO ÓTIMA NO ESPAÇO DE SINAIS 485

10.4.1 Espaço Geométrico de Sinais 485

10.4.2 Espaço de Sinais e Base de Sinais 487

10.5 DECOMPOSIÇÃO VETORIAL DE PROCESSOS ALEATÓRIOS DE RUÍDO BRANCO 489

10.5.1 Determinação de Funções de Base para um Processo Aleatório 490

10.5.2 Representação Geométrica de Processos de Ruído Branco 491

10.5.3 Ruído Gaussiano Branco 492

10.5.4 Propriedades de Processos Aleatórios Gaussianos 494

10.6 RECEPTOR ÓTIMO PARA CANAIS COM RUÍDO GAUSSIANO BRANCO 495

10.6.1 Representações Geométricas 495

10.6.2 Dimensionalidade do Espaço de Sinais de Detecção 497

10.6.3 Espaço de Sinais e Procedimento de Decisão (Simplificados) 500

10.6.4 Regiões de Decisão e Probabilidade de Erro 504

10.6.5 Sinalização Multiamplitude (PAM) 508
10.6.6 Análise QAM *M*-ária 512

10.7 EXPRESSÃO GERAL PARA A PROBABILIDADE DE ERRO DE RECEPTORES ÓTIMOS 517

10.8 CONJUNTOS EQUIVALENTES DE SINAIS 525
10.8.1 Conjunto de Sinais de Mínima Energia 527
10.8.2 Conjunto de Sinais Simplex 530

10.9 RUÍDO DE CANAL NÃO BRANCO (COLORIDO) 532

10.10 OUTROS CRITÉRIOS ÚTEIS DE DESEMPENHO 532

10.11 DETECÇÃO NÃO COERENTE 536

10.12 EXERCÍCIOS DE MATLAB 543

11 COMUNICAÇÕES POR ESPALHAMENTO ESPECTRAL 567

11.1 SISTEMAS DE ESPALHAMENTO ESPECTRAL POR SALTOS EM FREQUÊNCIA (FHSS) 567

11.2 SISTEMAS FHSS MULTIUSUÁRIO E DESEMPENHO 570

11.3 APLICAÇÕES DE FHSS 573

11.4 ESPALHAMENTO ESPECTRAL POR SEQUÊNCIA DIRETA 577

11.5 CARACTERÍSTICAS DE RESILIÊNCIA DE DSSS 580

11.6 MÚLTIPLO ACESSO POR DIVISÃO DE CÓDIGO (CDMA) DE DSSS 582

11.7 DETECÇÃO MULTIUSUÁRIO (MUD) 589

11.8 SISTEMAS CDMA DSSS PRÁTICOS MODERNOS 594
11.8.1 CDMA em Redes de Telefonia Celular 595
11.8.2 CDMA no Sistema de Posicionamento Global (GPS) 598
11.8.3 Padrão IEEE 802.11b para LAN Sem Fio 601

11.9 EXERCÍCIOS COM O MATLAB 602

12 COMUNICAÇÕES DIGITAIS COM CANAIS SUJEITOS A DISTORÇÃO LINEAR 617

12.1 DISTORÇÕES LINEARES EM CANAIS SEM FIO DE MULTIPERCURSO 617

12.2 EQUALIZAÇÃO DO CANAL NO RECEPTOR 620
12.2.1 Filtro Antimascaramento *versus* Filtro Casado 621
12.2.2 Estimação da Sequência de Máxima Verossimilhança (MLSE) 623
12.3 EQUALIZAÇÃO LINEAR COM ESPAÇAMENTO *T* (TSE) 626
12.3.1 TSE com Forçamento a Zero 627
12.3.2 Projeto de TSE com Base em MMSE 629
12.4 EQUALIZADORES LINEARES FRACIONALMENTE ESPAÇADOS (FSE) 633
12.4.1 Modelo Uma Entrada, Múltiplas Saídas (SIMO) 634
12.4.2 Configurações de FSE 635
12.5 ESTIMAÇÃO DE CANAL 637
12.6 EQUALIZADOR COM REALIMENTAÇÃO DE DECISÃO 638
12.7 COMUNICAÇÕES OFDM (MULTIPORTADORA) 640
12.7.1 Princípios de OFDM 641
12.7.2 Ruído em Canal OFDM 646
12.7.3 OFDM com Preenchimento de Zeros 648
12.7.4 Redundância de Prefixo Cíclico em OFDM 649
12.7.5 Equalização OFDM 649
12.8 MODULAÇÕES MULTITONS DISCRETOS (DMT) 650
12.9 APLICAÇÕES PRÁTICAS DE OFDM E DMT 654
12.10 EQUALIZAÇÃO CEGA E IDENTIFICAÇÃO 658
12.11 DISTORÇÕES EM CANAL VARIANTE NO TEMPO DEVIDO À MOBILIDADE 659
12.12 EXERCÍCIOS COM O MATLAB 662

13 INTRODUÇÃO À TEORIA DA INFORMAÇÃO 681
13.1 MEDIDA DE INFORMAÇÃO 681
13.2 CODIFICAÇÃO DE FONTE 686
13.3 COMUNICAÇÃO SEM ERRO EM UM CANAL RUIDOSO 691
13.4 CAPACIDADE DE CANAL DE UM CANAL DISCRETO SEM MEMÓRIA 694
13.5 CAPACIDADE DE CANAL DE UM CANAL CONTÍNUO SEM MEMÓRIA 701

13.6 EQUAÇÃO DE SHANNON E SISTEMAS DE COMUNICAÇÃO PRÁTICOS 716

13.7 CAPACIDADE DE CANAL SELETIVO EM FREQUÊNCIA 719

13.8 SISTEMAS DE COMUNICAÇÃO COM MÚLTIPLAS ENTRADAS E MÚLTIPLAS SAÍDAS 723

13.8.1 Capacidade de Canais MIMO 723

13.8.2 Transmissor sem Conhecimento do Canal 725

13.8.3 Transmissor com Conhecimento do Canal 727

13.9 EXERCÍCIOS COM O MATLAB 731

14 CÓDIGOS CORRETORES DE ERRO 743

14.1 VISÃO GERAL 743

14.2 REDUNDÂNCIA PARA CORREÇÃO DE ERRO 744

14.3 CÓDIGOS DE BLOCOS LINEARES 746

14.4 CÓDIGOS CÍCLICOS 753

14.5 EFEITOS DA CORREÇÃO DE ERRO 762

14.6 CÓDIGOS CONVOLUCIONAIS 766

14.6.1 Codificador Convolucional 766

14.6.2 Decodificação de Códigos Convolucionais 770

14.7 DIAGRAMA EM TRELIÇAS DE CÓDIGOS DE BLOCOS 777

14.8 COMBINAÇÃO E ENTRELAÇAMENTO DE CÓDIGOS 778

14.9 DECODIFICAÇÃO SUAVE 780

14.10 ALGORITMO DE VITERBI COM SAÍDA SUAVE (SOVA) 783

14.11 CÓDIGOS-TURBO 784

14.12 CÓDIGOS VERIFICADORES DE PARIDADE DE BAIXA DENSIDADE (LDPC) 792

14.13 EXERCÍCIOS COM O MATLAB 798

A ORTOGONALIDADE DE ALGUNS CONJUNTOS DE SINAIS 810

A.1 Ortogonalidade do Conjunto de Sinais Trigonométricos 810

A.2 Ortogonalidade do Conjunto de Sinais Exponenciais 811

B DESIGUALDADE DE CAUCHY-SCHWARZ 812

C ORTOGONALIZAÇÃO DE GRAM-SCHMIDT DE UM CONJUNTO DE VETORES 814

D PROPRIEDADES BÁSICAS MATRIZES E OPERAÇÕES COM MATRIZES 817

D.1 NOTAÇÃO 817

D.2 PRODUTO DE MATRIZES E SUAS PROPRIEDADES 818

D.3 MATRIZES IDENTIDADE E DIAGONAL 818

D.4 DETERMINANTE DE MATRIZ QUADRADA 819

D.5 TRAÇO 820

D.6 AUTODECOMPOSIÇÃO 820

D.7 MATRIZES QUADRADAS HERMITIANAS ESPECIAIS 821

E MISCELÂNEA 822

E.1 REGRA DE L'HÔPITAL 822

E.2 SÉRIES DE TAYLOR E DE MACLAURIN 822

E.3 SÉRIES DE POTÊNCIAS 822

E.4 SOMAS 823

E.5 NÚMEROS COMPLEXOS 823

E.6 IDENTIDADES TRIGONOMÉTRICAS 823

E.7 INTEGRAIS INDEFINIDAS 824

ÍNDICE 826

PREFÁCIO

O principal objetivo desta quarta edição é atender ao enorme progresso tecnológico em sistemas de comunicação ocorrido ao longo da década desde o lançamento da terceira edição. Ao mesmo tempo, novos pacotes de *software* e ferramentas de ensino se tornaram disponíveis, facilitando a apresentação de exemplos ricos e ilustrativos, assim como mais oportunidades experimentais para os estudantes. Nesta nova edição, grandes mudanças foram implementadas para incorporar esses recentes avanços tecnológicos na área de telecomunicações. Para aguçar o interesse dos estudantes e permitir que relacionem os conteúdos de cursos a experiências com ferramentas de comunicação com maior facilidade, apresentaremos informações relevantes sobre a operação e as características de sistemas celulares, redes locais sem fio (LANs) e serviços de internet com fio (linha digital de assinantes – DSL), entre outros.

Revisão Importante

Diversas alterações importantes foram motivadas pela necessidade de enfatizar os fundamentos de sistemas de comunicação digital que permeiam nossa vida diária. Em especial, devido às disseminadas aplicações de novas tecnologias, como espalhamento espectral e multiplexação por divisão em frequências ortogonais (OFDM), acrescentamos dois novos capítulos: sobre espalhamento espectral (Capítulo 11) e sobre canais seletivos em frequência e sistemas OFDM (Capítulo 12). Como exemplos práticos desses sistemas, elaboramos um texto introdutório básico sobre os atuais padrões de comunicação sem fio, incluindo sistemas celulares e redes de área local (LANs) sem fio baseadas no padrão IEEE 802.111a/b/g/n. Adicionalmente, incluímos um resumo das mais modernas tecnologias de modems e serviços DSL. De certo modo, a teoria e a forma de codificar a informação também foram transformadas pelo progresso em diversas áreas importantes. Nesta obra, incluímos os princípios básicos da tecnologia de sistemas com múltiplas entradas e saídas (MIMO), que passou a ter amplo uso comercial. Cobrimos, ainda, diversas inovações excelentes na codificação para corrigir erros, incluindo decodificação suave, códigos-turbo e códigos de verificação de paridade de baixa densidade (LDPC).

Para tornar o aprendizado mais aprimorado e dar aos estudantes a oportunidade de treinamento experimental baseado em computador, apresentamos exemplos relevantes e exercícios em MATLAB nos capítulos, o que aumenta a qualidade do conteúdo abordado, mediante o uso de experiências práticas.

Organização

A obra inicia tradicionalmente com a revisão de fundamentos de sinais e sistemas e prossegue com os tópicos de comunicação essenciais: modulações analógica e digital por codificação de pulsos. A seguir, mencionamos as ferramentas básicas da teoria da probabilidade e de processos aleatórios, que são usados em todo o texto para análise e projeto de sistemas de comunicação digital. Após cobrir os fundamentos destes sistemas, os dois últimos capítulos apresentam uma abordagem superficial da teoria da informação e dos princípios de códigos corretores de erros.

O ideal seria que o conteúdo deste livro fosse dividido em dois cursos: um sobre a operação básica de sistemas de comunicação, e outro sobre a análise de sistemas modernos de comunicação sujeitos a ruído e a outras distorções. No primeiro, a base seria as ferramentas de análise determinística, como séries de Fourier, transformadas de Fourier e teorema da amostragem, enquanto no segundo seriam cobertas as ferramentas de probabilidade e de processos aleatórios para tratar a imprevisibilidade de sinais de mensagens e ruídos. Todavia, atualmente, dada a grande variedade de cursos, pode ser difícil incluir dois semestres básicos sobre comunicação no programa típico de engenharia. Algumas universidades exigem como pré-requisito um curso de probabilidade e processos aleatórios, o que permite que as duas áreas sejam cobertas razoavelmente bem em uma única disciplina de um semestre. Este livro foi projetado para ser adotado tanto em um curso de um semestre (no qual se enfatizam os aspectos determinísticos de sistemas de comunicação com uma abordagem superficial dada aos efeitos de ruído e interferência) da mesma forma que em um curso que estude os aspectos determinísticos e probabilísticos de sistemas de comunicação. O livro é autossuficiente e provê toda a teoria básica de probabilidade e processos aleatórios. Entretanto, como dissemos, se o objetivo for cobrir os aspectos determinísticos e probabilísticos referentes a esses sistemas em um semestre, é altamente desejável que os estudantes já tenham bom conhecimento de probabilidade.

O Capítulo 1 apresenta um panorama dos sistemas de comunicação. Explicamos de forma qualitativa e heurística todos os conceitos importantes relativos à teoria da comunicação. Isso atrai os estudantes de cursos que cobrem genericamente esse tópico. Com este estímulo, eles se entusiasmam a estudar, nos Capítulos 2 e 3, a ferramenta de análise de sinais; nesses capítulos, os estudantes aprendem a ver um sinal como um vetor e a pensar no espectro de Fourier como uma forma de representar um sinal em termos de suas componentes vetoriais. Os Capítulos 4 e 5 discutem modulações em amplitude (linear) e em ângulo (não linear), respectivamente. Na era digital, muitos professores acham que a modulação deveria ter menos relevância. Consideramos que ela é mais uma ferramenta essencial de processamento de sinais do que um método de comunicação; ela será sempre necessária, tanto na área de comunicação (digital ou analógica), como em vários outros campos da engenharia elétrica. Por conseguinte, desprezá-la pode significar a ausência de uma perspectiva. O Capítulo 6, que serve como um elo entre comunicações analógicas e digitais, descreve o processo de conversão analógico-digital (ADC). Ele detalha amostragem, modulação por codificação de pulsos (incluindo DPCM), modulação delta, codificação de voz (vocoder), codificação e compressão de imagem/vídeo. O Capítulo 7 discute os princípios e técnicas empregados em modulação digital, introduzindo o conceito de distorção de canal e apresentando a equalização como uma forma eficaz de compensar a distorção.

Os Capítulos 8 e 9 apresentam aspectos básicos das teorias de probabilidade e de processos aleatórios, que representam a segunda ferramenta necessária ao estudo de sistemas de comunicação. Todo o esforço é feito para motivar os estudantes e assim conservá-los à medida que avançam na leitura dos capítulos; para isto, sempre que possível, apresentamos aplicações para problemas de comunicação. O Capítulo 10 trata da análise de sistemas de comunicação digital na presença de ruídos, incluindo a detecção ótima de sinais. O Capítulo 11 prioriza as comunicações por espalhamento espectral. O Capítulo 12 apresenta diversas técnicas práticas que podem ser empregadas no combate a distorções reais de canais, abordando tanto a equalização de canal como a tecnologia OFDM amplamente utilizada. O Capítulo 13 oferece um tutorial sobre teoria da informação. Por fim, o Capítulo 14 discute os princípios e importantes aspectos práticos da codificação para controle de erro.

Um de nossos objetivos ao escrever este livro foi tornar prazeroso o aprendizado ou, pelo menos, torná-lo uma experiência menos intimidadora para os estudantes; nesse sentido, cuidamos para apresentar o assunto de forma clara, compreensível e logicamente organizada. Sempre que possível, esforçamo-nos para apresentar visões esclarecedoras – mais do que compreensíveis – bem como explicações heurísticas de resultados teóricos. Incluímos, também, vários exemplos que auxiliam no entendimento de resultados abstratos. Mesmo que o sucesso em atingir esse objetivo seja parcial, nosso esforço terá valido a pena.

Um Mundo Completamente Novo

Assim que publicamos a terceira edição em 1998, surgiram importantes desenvolvimentos tecnológicos. Primeiro, a telefonia celular se tornou profundamente enraizada na vida das pessoas seja nos centros urbanos ou nos subúrbios, tanto na maioria dos países desenvolvidos quanto nos em desenvolvimento. Em 1998, pouquíssimos estudantes dispunham de pagers e de telefones celulares. Hoje, praticamente todos têm um celular. Segundo, em 1998, a maioria das conexões domésticas de internet era feita em baixa velocidade (28,8 kbit/s) via modems para linhas telefônicas discadas. Atualmente, a maioria dos estudantes se conecta ao ciberespaço por meio de serviços DSL ou a cabo. Além disso, as redes LAN sem fio fizeram com que termos esotéricos como IEEE 802.11 se tornassem corriqueiros. Grande parte dos estudantes já explorou essas tecnologias.

Devido aos enormes avanços tecnológicos, a nova geração de estudantes tem grande interesse em estudá-los e aprender como implementá-los. Eles anseiam por compreender como e onde podem contribuir para essa indústria. Tanto entusiasmo deve ser proveitosamente estimulado e explorado. Esta nova edição permitirá que o próprio professor recorde alguns tópicos ou que prepare material de leitura para que os estudantes possam assimilar informações relevantes. Esses objetivos podem ser alcançados com o destaque dos aspectos digitais do texto e a incorporação das mais conhecidas tecnologias digitais com e sem fio.

Uso do Livro em Cursos

Juntos, temos mais de 55 anos de experiência de ensino e lecionamos, em grandes universidades, em cursos com duração de um trimestre ou de um semestre. De forma complementar, as experiências pessoais dos estudantes relacionadas a sistemas de comunicação se multiplicaram continuamente desde a década de 1960 ao século XXI, saindo dos simples aparelhos de rádio até o acesso fácil a redes LAN sem fio, aos dispositivos celulares, às estações de rádio via satélite e aos serviços domésticos de internet. Em consequência, mais e mais estudantes têm interesse em aprender como esses aparelhos eletrônicos tão comuns funcionam. Considerando essa importante necessidade e nossa experiência de ensino, revimos essa edição visando adequá-la apropriadamente às diferentes configurações de programas de estudo. Em todos os casos, uma abordagem básica deve ensinar os fundamentos de comunicações analógicas e digitais (Capítulos 1 a 7).

Curso de Um Semestre (sem grande conhecimento de probabilidade)

Em muitos programas de estudo, os estudantes de graduação não utilizam ferramentas simples de probabilidade até chegarem aos cursos de comunicação. Em geral, isso ocorre porque as instituições os forçam a fazer um curso introdutório de estatística, desvinculado das necessidades da área de engenharia. Este texto é adequado aos estudantes com essa formação. Os primeiros sete capítulos constituem uma cobertura abrangente de modernos sistemas de comunicação analógica e digital, tendo em vista o típico estudante de engenharia. Um curso como este pode ser lecionado em um semestre (40 a 45 horas/aula). Com base na premissa de que os estudantes tenham adquirido conhecimento consistente de análise de Fourier em um curso anterior obrigatório sobre *sinais e sistemas*, a maior parte dos três capítulos iniciais pode ser tratada como revisão em uma semana. O resto do semestre pode ser totalmente dedicado ao estudo dos Capítulos 4 a 7, com cobertura parcial dos sistemas práticos dos Capítulos 11 e 12, para aumentar o interesse dos estudantes.

Curso de Um Semestre (com grande conhecimento de probabilidade)

No caso de programas de estudo que reforçaram o estudo inicial de teoria da probabilidade, uma cobertura mais abrangente de comunicações digitais pode ser alcançada em um semestre. A teoria da probabilidade pode ser ensinada com maior rigor no contexto da análise de sinais e sistemas (cf. George R. Cooper e Clare D. McGillem, *Probabilistic Methods of Signal and System Analysis*, Oxford University Press, 1999). Nesse cenário, além dos Capítulos 1 a 7, o Capítulo 10 e a parte do Capítulo 12 sobre equalização também podem ser lecionados em um semestre, desde que os estudantes domi-

nem o conteúdo de probabilidade, permitindo que se restrinja a cobertura dos Capítulos 8 a 9 a algumas poucas horas. Os estudantes que completarem este curso estarão bem preparados para ingressar no mercado de trabalho de telecomunicações ou em cursos de pós-graduação.

Série de Dois Semestres (sem um curso separado de probabilidade)

Todo o texto pode ser integral e detalhadamente coberto em dois semestres, no caso de programas de estudo sem um curso prévio de probabilidade. Em outras palavras, para uma série de dois cursos, o objetivo é ensinar sistemas de comunicações e os fundamentos de probabilidades. Em uma era em que numerosos cursos devem ser acomodados no programa de engenharia, é difícil alocar cursos de dois semestres direcionados apenas para comunicações. Contudo, a maioria das universidades oferece um curso autônomo de probabilidade oferecido por professores sem formação em engenharia. Nesse cenário, seria preferível incluir toda a teoria da probabilidade nos dois cursos de comunicações. Assim, para cursos de dois semestres, a cobertura pode ser dividida da seguinte forma:

- Primeiro semestre: Capítulos 1 a 7 (Sinais e Sistemas de Comunicação)
- Segundo semestre: Capítulos 8 a12 (Modernos Sistemas de Comunicação Digital)

Curso de Um Trimestre (com grande conhecimento de probabilidade)

Em um sistema trimestral, os estudantes devem fazer um curso anterior de probabilidade e estatística, em um nível avançado (cf. Cooper e McGillem, *Probabilistic Methods of Signal and System Analysis*). Também devem dominar a análise de Fourier. Em um trimestre, as aulas podem tratar dos fundamentos de sistemas de comunicações analógico e digital (Capítulos 3 a 7) e, com os Capítulos 10 e 11, de análise de sistemas de comunicação digital e comunicações por espalhamento espectral, respectivamente.

Curso de Um Trimestre (sem grande conhecimento de probabilidade)

No caso raro de estudantes com pouco conhecimento de probabilidade, é importante que eles aprendam os fundamentos de sistemas de comunicação. Sugerimos que o curso não se proponha a analisar os sistemas de comunicação digital. Em vez disso, a cobertura básica sem conhecimento de probabilidade pode ser feita com o ensino da operação de sistemas analógicos e digitais (Capítulos 1 a 7) e a discussão em alto nível de sistemas sem fio por espalhamento espectral (Capítulo 11).

Série de Dois Trimestres (com conhecimento básico de probabilidade)

Ao contrário do curso de um trimestre, uma série de dois trimestres pode ser bem projetada para ensinar a maior parte do conteúdo importante sobre sistemas de comunicação e sua análise. Todo o texto pode ser ensinado em dois trimestres, caso o programa de estudo inclua alguma cobertura preliminar de análise de Fourier e probabilidades. Em essência, se os Capítulos 1 a 3 e o Capítulo 8 forem vistos como parcialmente novos e parcialmente de revisão, a cobertura pode ser feita da seguinte forma:

- Primeiro trimestre: Capítulos 1 a 9 (Sistemas de Comunicação e Análise)
- Segundo trimestre: Capítulos 10 a 14 (Sistemas de Comunicação Digital)

MATLAB e Experimentos de Laboratório

Como muitas universidades não mais dispõem de laboratórios físicos para comunicações para auxiliar o aprendizado, o livro inclui exercícios sobre sistemas de comunicação baseados em MATLAB. Utilizando a tela do computador e medidas de taxa de erro de bits, os estudantes podem projetar sistemas, modificar seus parâmetros e avaliar os efeitos sobre o desempenho dos sistemas de comunicação. Desta forma, os estudantes podem aprender de forma autônoma como projetar e realizar simulações com sistemas de comunicação.

Agradecimentos

Primeiro, os autores agradecem a todos os estudantes que tiveram em todos esses anos. Esta edição não seria possível sem os comentários e sugestões fornecidos pelos nossos estudantes e sem as discussões que tivemos com eles. Os autores agradecem a todos os revisores por fornecerem valorosa ajuda no aprimoramento do texto. Por fim, os autores também agradecem ao professor Norman Morrison da University of Cape Town por sugerir um novo exercício (E8-2.3) nesta edição.

B.P. Lathi
Zhi Ding

Material Suplementar

Este livro conta com o seguinte material suplementar:

- Ilustrações da obra em formato de apresentação (restrito a docentes)

O acesso ao material suplementar é gratuito, bastando que o leitor se cadastre em: http://gen-io.grupogen.com.br.

GEN | Informação Online

GEN-IO (GEN | Informação Online) é o repositório de material suplementar e de serviços relacionados com livros publicados pelo GEN | Grupo Editorial Nacional, o maior conglomerado brasileiro de editoras do ramo científico-técnico-profissional, composto por Guanabara Koogan, Santos, Roca, AC Farmacêutica, Forense, Método, LTC, E.P.U. e Forense Universitária.

1 INTRODUÇÃO

Ao longo da última década, a rápida expansão das tecnologias de comunicação digital foi simplesmente impressionante. A internet, uma palavra e conceito inicialmente familiares apenas a técnicos e à comunidade científica, permeou todos os aspectos da vida das pessoas. Na sociedade moderna, é difícil encontrar qualquer indivíduo que não tenha sido influenciado pelas novas tecnologias de comunicação, dos telefones celulares ao Bluetooth. Este livro examina os princípios básicos de comunicação por sinais elétricos. Antes dos tempos modernos, mensagens eram transportadas por corredores, pombos-correio, luzes e fogo. Estes esquemas eram adequados às distâncias e "taxas de dados" da época. Na maior parte do mundo, esses modos de comunicação foram substituídos por sistemas elétricos de comunicação,* capazes de transmitir sinais por distâncias muito maiores (até a planetas e galáxias distantes) e à velocidade da luz.

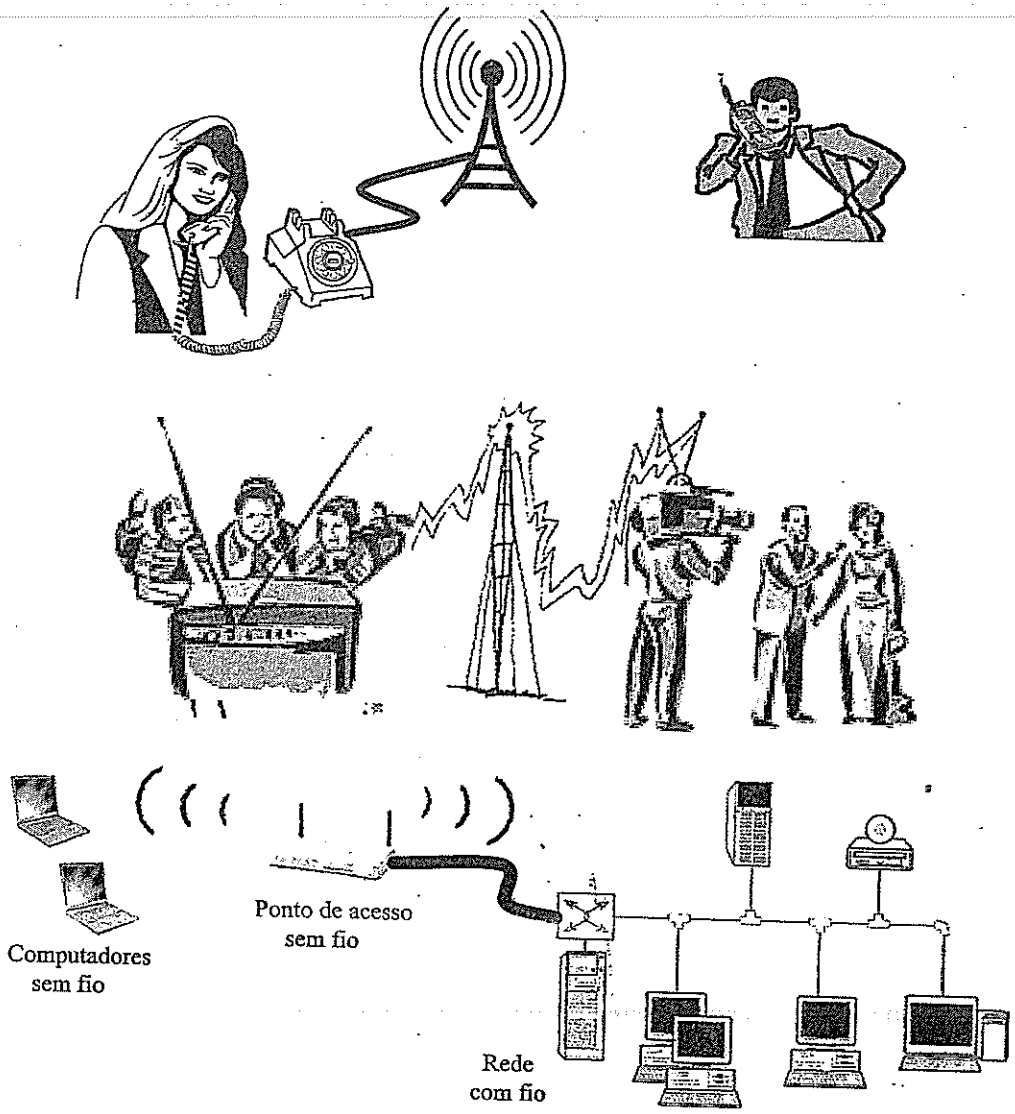
A comunicação elétrica é confiável e econômica, e as tecnologias nela empregadas aumentam a produtividade e a conservação de energia. Com crescente frequência, as reuniões de trabalho são conduzidas via teleconferência, economizando tempo e energia que seriam gastos com viagens. A comunicação ubíqua permite que o gerenciamento e a coordenação de participantes de um projeto sejam feitos em tempo real de qualquer ponto do globo. O correio eletrônico está substituindo rapidamente os mais custosos e lentos "correios-lesma". O comércio eletrônico também reduziu alguns custos e atrasos associados à venda, e os consumidores passaram a ter mais informação sobre novos produtos e suas características. As formas tradicionais de mídia, como televisão, rádio e jornais, evoluíram rapidamente nos últimos anos para se adequar às novas tecnologias de comunicação e de redes e delas tirar maior proveito. O objetivo deste livro é prover o conhecimento técnico fundamental de que necessitarão os engenheiros e técnicos de comunicações da próxima geração para que sejam capazes de projetar sistemas de comunicação ainda melhores no futuro.

1.1 SISTEMAS DE COMUNICAÇÃO

A Figura 1.1 ilustra três sistemas de comunicação típicos: uma conexão entre telefones de linha discada e celular, um sistema de difusão de TV e uma rede sem fio de computadores. Devido aos numerosos exemplos sistemas de comunicação, seria uma tolice tentar estudar, neste livro, os detalhes de todos os tipos de sistemas de comunicação. A forma mais eficiente e eficaz de aprender comunicações é estudar os principais blocos fundamentais que são comuns a praticamente todos os sistemas de comunicação. Assim, os alunos não apenas aprendem o funcionamento dos sistemas existentes que estudaram, mas, o que é mais importante, podem adquirir o conhecimento básico necessário para projetar e analisar sistemas novos e não mencionados neste livro. Para iniciar, é essencial definir um modelo para um típico sistema de comunicação, como mostrado na Figura 1.2. Os principais componentes de um sistema de comunicação são os seguintes:

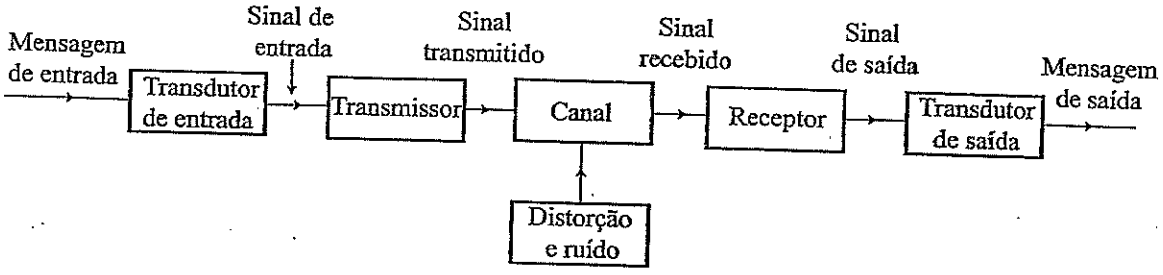
*Excetuando o serviço postal.

Figura 1.1
Alguns exemplos de sistemas de comunicação.



A **fonte** origina uma mensagem, como uma fala humana, uma imagem de televisão, uma mensagem de correio eletrônico ou algum dado. Se o dado for não elétrico (por exemplo, fala humana, texto de correio eletrônico, vídeo de televisão), deve ser convertido por um **transdutor de entrada** em uma forma de onda elétrica, referida como **sinal em banda base** ou **sinal de mensagem**; dispositivos físicos que realizam essa conversão são, por exemplo, microfone, teclado de computador, câmera fotográfica/filmadora digital.

Figura 1.2
Sistema de comunicação.



O **transmissor** modifica o sinal em banda base para transmissão eficiente. O transmissor pode consistir em um ou mais subsistemas: conversor A/D, codificador e modulador. De forma semelhante, o receptor pode ser um demodulador, um decodificador e um conversor D/A.

O **canal** é um meio de escolha no qual os sinais elétricos na saída do transmissor são transportados ao longo de uma dada distância. Um canal típico pode ser um par de fios de cobre trançados (telefone e DSL), cabo coaxial (televisão e internet), fibra óptica ou enlace de rádio. Além disso, um canal também pode ser uma conexão ponto a ponto em uma malha de canais interconectados que formam uma rede de comunicação.

O **receptor** processa o sinal recebido do canal, revertendo as modificações feitas pelo transmissor no sinal e removendo as distorções feitas pelo canal. A saída do receptor é alimentada ao **transdutor de saída**, que converte o sinal elétrico à sua forma original — a mensagem.

O **destino** é a unidade à qual a mensagem é endereçada.

Um canal é um meio físico que se comporta parcialmente como um filtro que, em geral, atenua o sinal e distorce as formas de onda transmitidas. A atenuação do sinal aumenta com o comprimento do canal, variando de uma pequena porcentagem, no caso de curtas distâncias, a ordens de magnitude, no caso de comunicação interplanetária. A distorção das formas de onda dos sinais tem origem em fenômenos físicos como ganhos dependentes da frequência, efeitos de multipercurso e deslocamento Doppler. Por exemplo, um canal *seletivo em frequência* causa diferentes valores de atenuação e de deslocamento de fase a diferentes componentes de frequência do sinal. Um pulso quadrado é arredondado ou "espalhado" durante a transmissão por um canal passa-baixas. Distorções desse tipo, denominadas **distorções lineares**, podem ser parcialmente corrigidas no receptor por um equalizador com características de ganho e fase complementares às do canal. Os canais também podem causar **distorção não linear** através de atenuação que varia com a amplitude do sinal. Tais distorções também podem ser parcialmente corrigidas no receptor, por um equalizador de características complementares. Caso as distorções do canal sejam conhecidas, também podem ser pré-compensadas por transmissores com a aplicação de pré-distorções dependentes do canal.

Em um ambiente prático, sinais que passam por canais de comunicação não apenas sofrem distorções do canal, mas também são corrompidos ao longo do percurso por interferências e perturbações indesejáveis, agrupadas no abrangente termo **ruído**. Esses sinais interferentes são aleatórios e imprevisíveis, e suas fontes são externas e internas. O ruído externo inclui sinais interferentes transmitidos em canais vizinhos, ruído gerado pelo homem por meio de comutadores defeituosos em equipamentos elétricos, radiação proveniente da ignição de automóveis, lâmpadas fluorescentes ou ruído natural advindo de raios, fornos de micro-ondas e emissões de telefones celulares, assim como tempestades elétricas e radiação intergaláctica. Um cuidado adequado no projeto do sistema pode reduzir ou, em alguns casos, até mesmo eliminar o ruído externo. Já o ruído interno resulta da agitação térmica de partículas carregadas em condutores, emissão espontânea, difusão ou recombinação de portadores de carga em dispositivos eletrônicos. Seus efeitos podem ser reduzidos com cuidado adequado, mas jamais eliminados. O ruído é um dos fatores fundamentais que limitam a taxa de comunicação.

Em sistemas de comunicação práticos, portanto, o canal distorce o sinal, e o ruído se acumula ao longo do percurso. Pior ainda, a intensidade do sinal diminui com a distância, enquanto o nível de ruído se mantém estacionário, independentemente do afastamento desde o transmissor. Em consequência, a qualidade do sinal se deteriora continuamente enquanto transpõe o comprimento do canal. A amplificação do sinal recebido para compensar a atenuação não é útil, pois o ruído será amplificado na mesma proporção, de modo que, na melhor das hipóteses, a qualidade do sinal fica inalterada.* Esses são os desafios importantes que devemos enfrentar no projeto de modernos sistemas de comunicação.

1.2 MENSAGENS ANALÓGICAS E MENSAGENS DIGITAIS

As mensagens são digitais ou analógicas. Mensagens digitais são combinações ordenadas de uma quantidade finita de símbolos ou de palavras de código. Por exemplo, o inglês (impresso) consiste

* Na verdade, a amplificação deteriorará ainda mais o sinal, devido ao ruído introduzido pelo amplificador.

em 26 letras, 10 números, um espaço e diversos caracteres de pontuação e de acentuação. Com isso, um documento de texto escrito em inglês é uma mensagem digital construída a partir de um teclado ASCII de 128 símbolos. A fala humana também é uma mensagem digital, pois é constituída de um vocabulário finito em alguma linguagem.* As notas musicais também são digitais, embora o som da música seja analógico. Da mesma forma, uma mensagem telegráfica em código Morse é uma mensagem digital construída a partir de um conjunto de apenas **dois** símbolos — traço e ponto. Consequentemente, é uma mensagem **binária**, ou seja, de apenas dois símbolos. Uma mensagem digital construída com M símbolos é denominada mensagem M-**ária**.

Mensagens analógicas são caracterizadas por dados cujos valores variam em um intervalo contínuo e são definidas em um período contínuo de tempo. Por exemplo, a temperatura ou a pressão atmosférica de um certo local, ao longo do tempo, pode variar em um intervalo contínuo e pode assumir um número infinito (incontável) de valores possíveis. Uma peça musical gravada por um pianista também é um sinal analógico. De modo semelhante, a amplitude de uma particular forma de onda de voz varia em um intervalo contínuo. Em um dado intervalo de tempo, existe um número infinito de diferentes possíveis formas de onda de voz, em contraste com apenas um número finito de possíveis mensagens digitais.

1.2.1 Imunidade de Sinais Digitais ao Ruído

Não é segredo, mesmo para um observador casual, que, toda vez que alguém examina os mais recentes produtos de comunicação eletrônica, uma "tecnologia digital" mais nova e melhor substitui a velha tecnologia analógica. Na última década, os telefones celulares completaram a transformação da primeira geração analógica AMPS na atual segunda geração (por exemplo, GSM, CDMA) e sua descendente digital, a terceira geração (por exemplo, WCDMA). Mais visível nas residências, a tecnologia de vídeo digital (DVD) tornou o sistema analógico de VHS praticamente obsoleto. A televisão digital dá prosseguimento ao ataque digital à tecnologia de vídeo analógico, removendo o último remanescente da televisão em cores. É razoável perguntar: por que as tecnologias digitais são melhores? A resposta está associada a aspectos econômicos e à qualidade. A motivação econômica é a facilidade de adoção dos versáteis, poderosos e baratos microprocessadores digitais de alta velocidade. Ainda mais importante, no que diz respeito ao nível de qualidade, uma característica proeminente das comunicações digitais é a maior imunidade de sinais digitais ao ruído e a interferências.

As mensagens digitais são transmitidas como um conjunto finito de formas de ondas elétricas. Em outras palavras, uma mensagem digital é gerada a partir de um alfabeto finito, e cada caractere no alfabeto pode ser representado por uma forma de onda ou por uma combinação sequencial de formas de onda. Por exemplo, no envio de uma mensagem em código Morse, um traço pode ser transmitido por um pulso elétrico de amplitude $A/2$ e o ponto, por um pulso de amplitude negativa $-A/2$ (Fig. 13a). Em um caso M-ário, M pulsos (ou formas de onda) elétricos são usados; cada um dos M pulsos representa um dos M símbolos possíveis. Após a transmissão, o receptor deve extrair a mensagem do sinal distorcido e ruidoso na saída do canal. É mais fácil, em geral, extrair mensagens de sinais digitais que de sinais analógicos, pois a decisão digital deve pertencer a um alfabeto de tamanho finito. Consideremos um caso binário: dois símbolos são codificados como pulsos retangulares de amplitudes $A/2$ e $-A/2$. A única decisão no receptor consiste em escolher entre dois possíveis pulsos recebidos; detalhes da forma do pulso não são relevantes. Um alfabeto finito resulta em imunidade ao ruído e a interferências. A decisão de um receptor pode ser feita com razoável certeza mesmo se os pulsos forem afetados por distorção modesta e ruído (Fig. 1.3). A mensagem digital na Fig. 1.3a é distorcida pelo canal, como ilustrado na Fig. 1.3b. Contudo, caso a distorção não seja muito grande, os dados podem ser recuperados sem erro, pois basta uma simples decisão binária: o pulso recebido é positivo ou negativo? A Fig. 1.3c mostra os mesmos dados com distorção do canal e ruído. Os dados podem

*Aqui nos referimos à informação contida na fala e não a detalhes da mesma, como pronúncia das palavras, variações de inflexão, tom e ênfase. O sinal de voz de um microfone contém todos esses detalhes e é, portanto, um sinal analógico; seu conteúdo de informação é mais de mil vezes maior que a informação acessível de uma versão escrita da mesma mensagem.

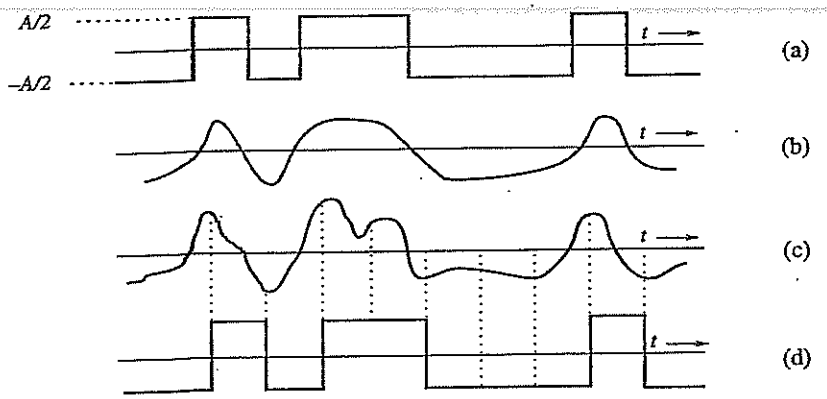


Figura 1.3 (a) Sinal transmitido. (b) Sinal distorcido recebido (sem ruído). (c) Sinal distorcido recebido (com ruído). (d) Sinal regenerado (com atraso).

ser recuperados corretamente desde que a distorção e o ruído estejam dentro de certos limites. Em contraste, em uma mensagem analógica, a própria forma de onda transporta a informação desejada, e mesmo uma pequena distorção ou interferência será percebida no sinal recebido. Fica claro que um sistema de comunicação digital é mais robusto que um sistema de comunicação analógico, pois suporta melhor ruído e distorção (desde que estes se mantenham dentro de um limite).

1.2.2 Viabilidade de Repetidores Regenerativos sem Distorção

Uma das principais razões para a qualidade superior de sistemas digitais em relação a sistemas analógicos é a viabilidade de implementação, nos primeiros, de repetidores e nós de rede **regenerativos**. Em um sistema digital, estações repetidoras são posicionadas ao longo do percurso de comunicação a distancias curtas o bastante para assegurar que ruído e distorção permaneçam dentro de um limite. Isso permite a detecção de pulsos com grande precisão. Em cada estação repetidora ou nó da rede, os pulsos que chegam são detectados, de modo que pulsos novos e "limpos" são retransmitidos à próxima estação repetidora ou nó. Esse processo evita o acúmulo de ruído e distorção ao longo do percurso, pois limpa os pulsos a intervalos regulares entre repetidores. Desta forma, mensagens podem ser transmitidas por distâncias mais longas com maior precisão. A aplicação de regeneração sem distorção, por meio de repetidores, é bastante disseminada em sistemas de comunicação de longas distâncias e em nós de redes grandes (e, possivelmente, heterogêneas).

Em sistemas analógicos, sinais e ruído em uma mesma largura de banda não podem ser separados. Nesses casos, regeneradores consistem basicamente em filtros e amplificadores e não são "regenerativos". Portanto, é impossível evitar acúmulo de ruído e distorção na banda passante ao longo do percurso. Em consequência, a distorção e a interferência de ruído podem se acumular por todo o percurso de transmissão à medida que o sinal transpõe a rede. Para piorar o problema, o sinal é continuamente atenuado no percurso da transmissão. Assim, com o aumento da distância, o sinal se torna mais fraco, enquanto a distorção e o ruído se acumulam mais. Por fim, o sinal, destruído por distorção e ruído, torna-se irreconhecível. A amplificação é de pouca valia, pois amplifica igualmente o sinal e o ruído. Em consequência, a distância que uma mensagem analógica pode percorrer para ainda ser recebida com sucesso é limitada pela potência do primeiro transmissor. Apesar dessas limitações, a comunicação analógica foi, no passado, amplamente utilizada com sucesso para comunicações de curtas e médias distâncias. Hoje, com o advento da comunicação por fibras ópticas e a dramática redução de custos alcançada na fabricação de circuitos digitais de alta velocidade e de dispositivos digitais de armazenamento de dados, praticamente todos os novos sistemas de comunicação instalados são digitais. Contudo, alguns antigos sistemas de comunicação analógicos ainda estão em uso, incluindo os empregados na difusão de rádio AM e FM.

1.2.3 Conversão Analógico-Digital (A/D)

Apesar da diferença entre sinais analógicos e digitais, há uma base comum entre os dois: a conversão de mensagens analógicas em sinais digitais (conversão A/D). O conversor analógico-digital (A/D), um dispositivo essencial em eletrônica, permite que sistemas de comunicação digital transportem sinais analógicos, como áudio e vídeo. Em geral, sinais analógicos são contínuos no tempo e em um intervalo de valores; ou seja, têm valores em cada instante de tempo e seus valores podem ter qualquer amplitude no dado intervalo. Sinais digitais, por sua vez, existem apenas em momentos discretos no tempo e podem assumir somente um número finito de valores. A conversão A/D jamais é 100% precisa. Contudo, como a percepção humana não exige precisão infinita, a conversão A/D pode, de modo efetivo, capturar a informação necessária de uma fonte analógica para a transmissão do sinal digital.

A conversão A/D se dá em dois passos: um sinal em tempo contínuo é *amostrado* para produzir um sinal em tempo discreto, cujas amplitudes contínuas são, então, *quantizadas* em níveis discretos de sinal. Primeiro, o espectro de um sinal indica as magnitudes relativas das diversas componentes de frequência. O **teorema da amostragem** (Capítulo 6) afirma que, se a mais alta frequência no espectro do sinal for B (em hertz), o sinal pode ser reconstruído a partir das amostras discretas, tomadas a uma taxa uniforme não menor que $2B$ amostras por segundo. Isso significa que, para preservar a informação de um sinal em tempo contínuo, basta transmitir apenas suas amostras (Fig. 1.4). Entretanto, os valores das amostras ainda não estão na forma digital, pois assumem valores em um intervalo contínuo. Então, o segundo passo da **quantização** é efetuado. Por meio da quantização, cada amostra é aproximada, ou "arredondada", ao nível de quantização mais próximo, como ilustrado na Fig. 1.4. Como a precisão da percepção humana é limitada, a quantização feita com granularidade suficiente não compromete a qualidade do sinal. Se as amplitudes do sinal de mensagem $m(t)$ tiverem valores no intervalo $(-m_p, m_p)$, o quantizador parte esse intervalo em L subintervalos. Cada amostra de amplitude é aproximada ao valor médio do subintervalo que a contém e passa a ser representada por um dos L números. A informação é, então, digitalizada dessa forma. Após os dois passos de amostragem e quantização, a conversão analógico-digital (A/D) se completa.

O sinal quantizado é uma aproximação do sinal original. A precisão do sinal quantizado pode ser aumentada a qualquer grau desejado por meio do aumento do número de níveis L. Para a inteligibilidade de sinais de voz, por exemplo, é suficiente tomar $L = 8$ ou 16. Para uso comercial, $L = 32$ é um valor mínimo; para comunicação telefônica, $L = 128$ e 256 são valores típicos.

Um típico sinal binário distorcido com ruído adquirido ao longo do canal é ilustrado na Fig. 1.3. Se A for suficientemente grande em comparação às amplitudes de ruído, o receptor ainda pode distinguir corretamente entre os dois pulsos. A amplitude dos pulsos é, em geral, 5 a 10 vezes a amplitude rms do ruído. Com um valor tão alto para a relação sinal-ruído (SNR — *signal-to-noise ratio*), a probabilidade de erro no receptor é menor que 10^{-6}; ou seja, em média, o receptor cometerá menos que um erro a cada milhão de pulsos. Portanto, o efeito aleatório de ruído e distorção do canal é praticamente eliminado. Quando sinais analógicos são transmitidos na forma digital, algum erro ou incerteza no sinal recebido pode ser causado pela quantização, além do ruído e das interferências do canal.

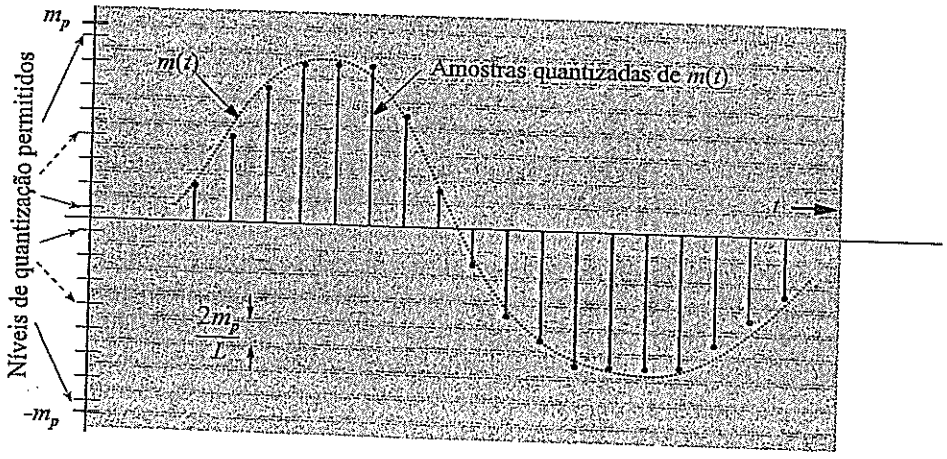


Figura 1.4 Conversão analógica-digital de um sinal.

A incerteza ou o erro causado pela quantização pode ser reduzido a qualquer valor desejado através do aumento de L. Além disso, o uso de repetidores regenerativos permite a transmissão de sinais ao longo de distâncias muito maiores do que seria possível com sinais analógicos. Como veremos mais adiante no texto, o preço de todos esses benefícios da comunicação digital é pago em termos de maiores complexidade de processamento e largura de banda de transmissão.

1.2.4 Modulação por Codificação de Pulsos – Uma Representação Digital

Uma vez completada a conversão A/D, a mensagem analógica original é representada por uma sequência de amostras, cada uma assumindo um dos L níveis de quantização preestabelecidos. A transmissão dessa sequência quantizada é tarefa de sistemas de comunicação digital. Formas de onda de sinal devem, então, ser usadas para representar a sequência de amostras quantizadas no processo de transmissão. Do mesmo modo, um dispositivo digital de armazenamento também precisa representar as amostras como formas de onda de sinal. A *modulação por codificação de pulsos* (PCM — *pulse-coded modulation*)* é um mecanismo muito simples e comumente empregado para este propósito.

Primeiro, um *bit* de informação se refere a um *dígito binário***, **1** ou **0**. A ideia por trás de PCM é representar cada amostra quantizada por uma combinação ordenada de dois pulsos básicos: $p_1(t)$ para representar o bit **1** e $p_0(t)$ para representar o bit **0**. Como cada um dos possíveis L valores de amostra pode ser escrito como uma sequência de bits de comprimento $\log_2 L$, cada amostra pode, assim, ser mapeada em uma sequência curta de pulsos que representa a sequência binária de bits. Por exemplo, se $L = 16$, cada nível de quantização pode ser descrito de forma única por 4 bits. Se empregarmos dois pulsos básicos, $p_1(t) = A/2$ e $p_0(t) = -A/2$. Uma sequência de quatro desses pulsos resulta em $2 \times 2 \times 2 \times 2 = 16$ padrões distintos, como ilustrado na Fig. 1.5. Podemos alocar um padrão a cada um dos 16 valores quantizados a serem transmitidos. Cada amostra quantizada é, então, codificada em uma sequência de quatro pulsos binários. Esse é o princípio da transmissão PCM, em que a sinalização é efetuada por meio de apenas dois pulsos (ou símbolos) básicos. O caso binário é de grande importância prática, devido à sua simplicidade e facilidade de detecção. A maior parte da comunicação digital da atualidade é binária.***

Embora PCM tenha sido inventada por P. M. Rainey em 1926 e redescoberta por A. H. Reeves em 1939, apenas na década de 1960 o grupo Bell System**** instalou o primeiro enlace de comunicação a utilizar PCM para transmissão de voz. O custo e o tamanho de circuitos a válvulas foram o principal impedimento para o emprego de PCM, até a descoberta de dispositivos semicondutores. PCM se tornou prática com o advento do transistor.

Dessa discussão sobre PCM, chegamos a uma conclusão interessante (e, até certo ponto, nada óbvia): qualquer comunicação possível pode ser transmitida por meio de um mínimo de dois símbolos. Assim, com apenas uma sequência adequada de piscadas de olho, podemos transmitir qualquer mensagem, seja uma conversa, um livro, um filme ou uma ópera. Todos os possíveis detalhes (como os variados tons de cores de objetos, timbres de voz etc.) reproduzíveis em uma tela de cinema ou na televisão colorida de alta definição podem ser transmitidos, sem perda de precisão, por meras piscadas de um olho.*****

---

*É comum, também, o termo *modulação por código de pulsos*. (N.T.)

**Na língua inglesa, a palavra "bit" significa uma pequena parte ou pedaço de algo; no caso, um pedaço de informação. Bit também é interpretado como uma contração do termo em inglês para dígito binário, *binary digit*. (N.T.)

***Existe um caso intermediário em que usamos quatro pulsos básicos (pulsos quaternários) de amplitudes $\pm A/2$ e $\pm 3A/2$. Uma sequência de dois pulsos quaternários podem formar $4 \times 4 = 16$ níveis distintos de valores.

****Grupo de companhias que explorou o serviço telefônico nos Estados Unidos de 1877 a 1984, quando foi desmembrado por lei federal. (N.T.)

*****É óbvio que, para transmitir a informação contida em um filme ou programa de televisão em tempo real, as piscadas de olho deveriam ser efetuadas em alta velocidade, inatingível para humanos. Por exemplo, um sinal de TV digital de alta definição (HDTV — *High-Definition TeleVision*) é representado por 19 milhões de bits (piscadas) por segundo.

Figura 1.5
Exemplo de codificação PCM.

Dígito	Equivalente binário	Forma de onda da codificação por pulsos
0	0000	
1	0001	
2	0010	
3	0011	
4	0100	
5	0101	
6	0110	
7	0111	
8	1000	
9	1001	
10	1010	
11	1011	
12	1100	
13	1101	
14	1110	
15	1111	

1.3 EFEITO, RELAÇÃO SINAL-RUÍDO E CAPACIDADE DO CANAL

No projeto de sistemas de comunicação, são essenciais o entendimento e a análise de fatores importantes como características do sinal e do canal, amplitude relativa de ruído, número máximo de bits por segundo que pode ser transmitido por um canal e, principalmente, a qualidade do sinal.

1.3.1 Largura de Banda e Potência de Sinal

Em um dado sistema de comunicação (digital), os parâmetros e as limitações físicas fundamentais que controlam a taxa de transmissão e qualidade do canal são a largura de banda B e a potência de sinal P_S. Em capítulos posteriores, esses parâmetros serão discutidos em detalhes tanto qualitativa como quantitativamente. Por ora, nosso objetivo é discuti-los apenas de modo qualitativo.

A **largura de banda** de um canal é o intervalo de frequências que ele é capaz de transmitir com razoável fidelidade. Por exemplo, se um canal transmite com razoável fidelidade um sinal cujas componentes de frequência variam de 0 Hz (dc) a um máximo de 5.000 Hz (5 kHz), a largura de banda B do canal é 5 kHz. Da mesma forma, cada sinal tem uma largura de banda que mede a separação máxima entre suas componentes de frequência.

Quanto mais rápidas as variações de um sinal, maior sua frequência máxima e maior sua largura de banda. Sinais ricos em conteúdo que variam rapidamente (como os associados a cenas de batalhas em um vídeo) têm maior largura de banda do que sinais maçantes que variam lentamente (como os de uma novela diurna ou de um vídeo de um animal dormindo). Um sinal pode ser transmitido com sucesso através de um canal se a largura de banda do canal for maior que a do sinal.

Para entendermos o papel de B, consideremos a possibilidade de um aumento na velocidade de transmissão por meio da compressão do sinal no tempo. A compressão do sinal no tempo por um fator

de 2 permite que ele seja transmitido na metade do tempo, de modo que a velocidade (taxa) de transmissão dobra. Contudo, a compressão do sinal no tempo por um fator de 2 faz com que o sinal "se agite" duas vezes mais rápido, o que também dobra as frequências de seus componentes. Muitas pessoas já experimentaram esse efeito tocando fitas de áudio no dobro da velocidade normal, fazendo com que as vozes das pessoas soassem como a fala esganiçada de personagens de desenho animado. Para transmitir o sinal comprimido sem distorção, a largura de banda do canal também deve ser dobrada. Portanto, a taxa de transmissão de informação que um canal pode acomodar com sucesso é diretamente proporcional a B. De modo mais genérico, se um canal de largura de banda B pode transmitir N pulsos por segundo, para transmitir KN pulsos por segundo com a mesma tecnologia é necessário um canal com largura de banda KB. Em outras palavras, o número de pulsos por segundo que pode ser transmitido através de um canal é diretamente proporcional à sua largura de banda B.

A **potência de sinal** P_S desempenha duplo papel na transmissão. Primeiro, P_S está relacionada à qualidade de transmissão. O aumento de P_S reforça o pulso de sinal e diminui os efeitos do ruído e interferência do canal. Na verdade, a qualidade de sistemas de comunicação analógicos e digitais varia com a relação sinal-ruído (SNR). Em qualquer situação, um valor mínimo da SNR no receptor é necessário para comunicação com qualidade. Portanto, um grande valor da potência de sinal P_S permite que o sistema mantenha uma SNR mínima ao longo de uma distância maior, garantindo comunicação com qualidade com um alcance maior.

O segundo papel da potência de sinal é menos óbvio, embora igualmente importante. Do ponto de vista da teoria da informação, a largura de banda B e a potência de sinal P_S são, até certo ponto, intercambiáveis; ou seja, para manter determinada taxa e precisão de transmissão de informação, podemos negociar entre os valores de P_S e B. Assim, podemos reduzir B se estivermos dispostos a aumentar P_S ou podemos diminuir P_S se uma maior largura de banda B estiver disponível. A prova rigorosa dessa afirmação será apresentada no Capítulo 13.

Em resumo, os dois principais recursos para comunicação são a largura de banda e a potência transmitida. Em um dado canal de comunicação, um recurso pode ser mais valioso que o outro e o esquema de comunicação deve ser projetado levando isso em consideração. Um típico canal telefônico, por exemplo, tem largura de banda limitada (3 kHz), mas maior liberdade em relação à potência. Em naves espaciais, por sua vez, a largura de banda disponível é enorme, mas a potência é severamente limitada. Portanto, as soluções de comunicação para esses dois casos são radicalmente distintas.

1.3.2 Capacidade do Canal e Taxa de Dados

A largura de banda do canal limita as larguras de banda de sinais que podem ser transmitidos com sucesso, enquanto a relação sinal-ruído (SNR) no receptor determina a qualidade da recuperação dos sinais transmitidos. Uma maior SNR significa que o pulso de sinal transmitido pode usar mais níveis de sinais, o que permite o transporte de um número maior de bits em cada transmissão de pulso. Uma maior largura de banda B também significa que podemos transmitir mais pulsos (variação mais rápida) ao longo do canal. Em consequência, SNR e largura de banda B afetam a "vazão" do canal. A máxima vazão que pode ser transportada de modo confiável por um canal é denominada capacidade do canal.

Um dos tipos de canais mais comumente encontrados é conhecido como canal com ruído gaussiano branco aditivo (AWGN — *Additive White Gaussian Noise*). O modelo de canal AWGN assume que a única distorção presente é ruído gaussiano branco aditivo com largura de banda finita B. Esse modelo ideal acomoda aplicações com canal sem distorção e provê um limite superior de desempenho para canais genéricos sujeitos a distorções. A capacidade do canal AWGN com largura de banda limitada foi estabelecida, de forma brilhante, na equação de Shannon,

$$C = B \log_2(1 + \text{SNR}) \quad \text{bit/s} \tag{1.1}$$

Nessa equação, a capacidade do canal (C) representa o limite superior da taxa de transmissão de informação por segundo. Em outras palavras, C é o máximo número de bits que podem ser transmitidos por segundo com uma probabilidade de erro arbitrariamente próxima de zero; ou seja, a transmissão é

tão confiável quanto desejarmos. Contudo, a capacidade apenas ressalta essa *possibilidade*, pois não especifica como pode ser implementada. É importante ressaltar que é impossível transmitir a uma taxa superior a essa sem incorrer em erros. A equação de Shannon deixa bem clara a limitação imposta à taxa de comunicação por B e pela SNR. Se não houvesse ruído no canal (o que corresponderia a SNR = ∞), a capacidade C seria ∞ e a taxa de comunicação poderia ser arbitrariamente alta. Poderíamos, então, transmitir qualquer quantidade de informação em um canal sem ruído. Essa afirmação pode ser comprovada com facilidade. Se o ruído fosse zero, não haveria incerteza na amplitude do pulso recebido e o receptor seria capaz de detectar qualquer amplitude de pulso sem erro. A menor separação entre amplitudes de pulsos pode ser tão pequena quanto desejarmos e, para um dado pulso qualquer, teríamos disponível um número infinito de níveis próximos. Poderíamos alocar um nível a cada mensagem possível. Como um número infinito de níveis estaria disponível, seria possível alocar um nível a cada mensagem concebível. A implementação de um código desse tipo pode não ser prática, mas isso não é relevante nesta argumentação. O importante, aqui, é que, se o ruído fosse zero, a comunicação deixaria de ser um problema, ao menos em teoria. A implementação de um esquema como esse seria difícil, pois requereria a geração e detecção de pulsos com amplitudes precisas. As dificuldades práticas decorrentes imporiam um limite sobre a taxa de comunicação. Vale a pena ressaltar que o resultado de Shannon, que representa um limite superior para a taxa de comunicação em um canal, seria alcançável somente com um sistema de complexidade monstruosa e impraticável, e o atraso temporal na recepção tenderia ao infinito. Sistemas práticos operam a taxas inferiores à de Shannon.

Em resumo, a equação de capacidade de Shannon demonstra, de modo qualitativo, os papéis básicos desempenhados por B e SNR na limitação do desempenho de um sistema de comunicações. Esses dois parâmetros representam, então, a principal limitação sobre a taxa de comunicação. A possibilidade da troca de valores entre esses dois parâmetros básicos também é demonstrada pela equação de Shannon.

Como exemplo prático da troca de valores entre SNR e a largura de banda B, consideremos o cenário em que um homem de voz suave fala um pouco rápido demais e não conseguimos entender completamente o que diz. Isso significa que, como ouvintes, nossa largura de banda B é muito pequena e, em consequência, a capacidade C não é alta o bastante para acomodar as frases faladas rapidamente. Contudo, se o homem puder falar mais alto (aumentando a potência e, portanto, a SNR), é provável que o entendamos melhor sem nenhuma outra alteração. Tal exemplo ilustra o conceito de permuta de recursos entre SNR e B. Vale notar que, no entanto, essa relação não é linear. Dobrar o volume do orador permite que ele fale um pouco mais rápido, mas não duas vezes mais rápido. Esse efeito de permuta desigual é totalmente capturado na equação de Shannon [Eq. (1.1)], na qual a duplicação da SNR não pode sempre compensar a perda de 50% em B.

1.4 MODULAÇÃO E DETECÇÃO

Sinais analógicos gerados por fontes de mensagens ou sinais digitais gerados por meio da conversão A/D de sinais analógicos são, em geral, referidos como sinais em banda base, pois têm natureza passa-baixos. Sinais em banda base podem ser transmitidos diretamente por meio de um canal apropriado (por exemplo, telefone, fax). Contudo, dependendo das características do canal e do sinal no domínio da frequência, sinais em banda base produzidos por diversas fontes de informação nem sempre são adequados à transmissão direta através de um dado canal. Quando as bandas de frequências do sinal e do canal não coincidem exatamente, os canais não podem ser deslocados (em frequência). Portanto, as mensagens devem ser deslocadas para as bandas de frequências adequadas aos canais. Sinais de mensagem devem, em consequência, ser modificados para que a transmissão se torne possível. Nesse processo de conversão, conhecido como **modulação**, o sinal em banda base é usado para modificar (isto é, modular) algum parâmetro de uma *portadora* de sinal de radiofrequência (RF).

Uma **portadora** é uma senoide de alta frequência. Através da modulação, um dos parâmetros da portadora senoidal — como amplitude, frequência ou fase — é variado proporcionalmente ao sinal em banda base $m(t)$. Dessa forma, temos modulação em amplitude (AM — *amplitude modulation*),

modulação em frequência (FM — *frequency modulation*) e modulação em fase (PM — *phase modulation*). A Fig. 1.6 ilustra um sinal em banda base $m(t)$ e as correspondentes formas de onda AM e FM. Na modulação AM, a amplitude da portadora varia em proporção a $m(t)$; na modulação FM, a frequência da portadora varia em proporção a $m(t)$. Para reconstruir o sinal em banda base no receptor, o sinal modulado deve passar por um processo reverso denominado **demodulação**.

Como mencionado, a modulação é utilizada para permitir a transmissão. Algumas das principais razões para a modulação são discutidas a seguir.

1.4.1 Facilidade de Radiação/Transmissão

Para radiação eficiente de energia eletromagnética, a antena radiante deve ter dimensões que sejam da ordem de uma fração ou mais do comprimento de onda do sinal de excitação. Para muitos sinais em banda base, os comprimentos de onda são demasiadamente grandes para dimensões razoáveis de antenas. Por exemplo, a potência em um sinal de voz é concentrada em frequências no intervalo de 100 a 3000 MHz. Os correspondentes comprimentos de onda variam de 100 a 3000 km. Esses longos comprimentos de onda requerem antenas cujas dimensões são demasiadamente grandes para ser construídas. Com a modulação de uma portadora, transladamos o espectro do sinal para as vizinhanças da frequência da portadora, que corresponde a um comprimento de onda muito menor. Por exemplo, uma portadora de 10 MHz tem um comprimento de onda de apenas 30 m e sua transmissão pode ser feita com uma antena com dimensões da ordem de 3 m. Nesse sentido, a modulação corresponde a fazer o sinal em banda base pegar uma carona com uma senoide de alta frequência (portadora). A portadora e o sinal em banda base também podem ser comparados a uma pedra e a um pedaço de papel. Se jogarmos o papel no ar, ele não irá muito longe. Contudo, se embrulharmos uma pedra (portadora) com ele, poderá ser lançado a uma distância muito maior.

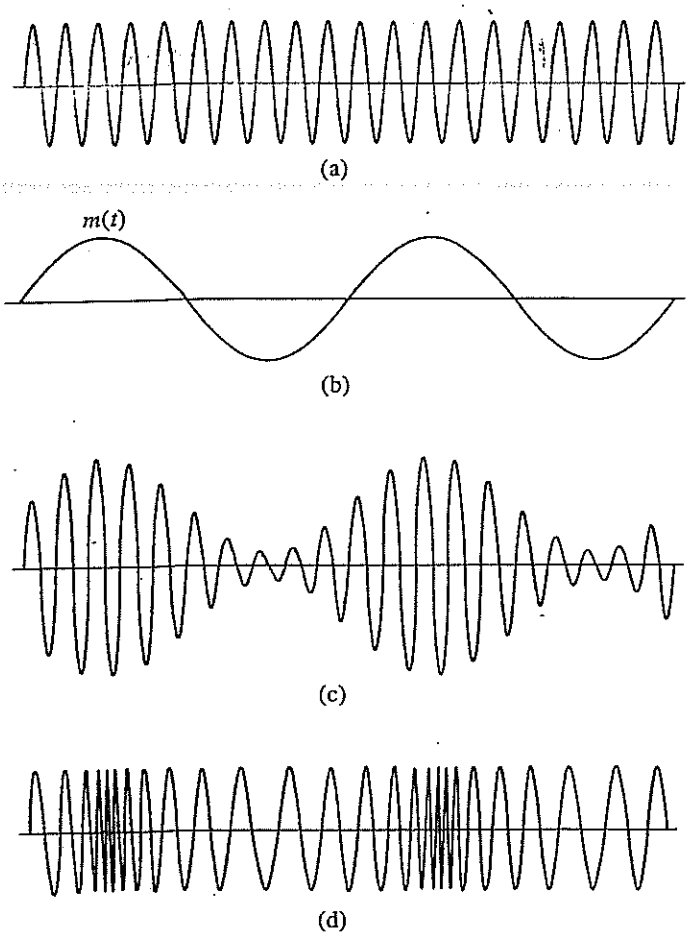


Figura 1.6 Modulação: (a) portadora; (b) sinal modulante (em banda base); (c) onda modulada em amplitude; (d) onda modulada em frequência.

1.4.2 Transmissão Simultânea de Múltiplos Sinais — Multiplexação

A modulação também permite que múltiplos sinais sejam transmitidos ao mesmo tempo em uma mesma área geográfica sem interferência mútua direta. Um exemplo simples disso são as diversas estações de televisão transportadas por um mesmo cabo (ou pelo ar) até os receptores de TV nas casas das pessoas. Sem modulação, múltiplos sinais de vídeo interferirão uns com os outros, pois todos os sinais de vídeo em banda base têm a mesma largura de banda. Portanto, sem modulação, a TV a cabo ou pelo ar seria limitada a uma estação em uma dada região — o que seria um protocolo com grandes perdas, pois a largura de banda do canal é muitas vezes maior que a do sinal.

Uma forma de resolver esse problema é o uso de modulação. Podemos usar as várias estações de TV para modular portadoras de distintas frequências; dessa forma, cada sinal é transladado a uma faixa de frequências diferentes. Se as diversas portadoras forem suficientemente espaçadas em frequência, os espectros dos sinais modulados (conhecidos como canais de TV) não se sobreporão e, portanto, não haverá interferência entre os mesmos. No receptor (aparelho de TV), um filtro passa-faixa sintonizável poderá selecionar a desejada estação ou canal de TV. Esse método de transmissão de vários sinais simultaneamente, em faixas de frequências que não se sobrepõem, é conhecido como **multiplexação por divisão em frequência (FDM — *frequency divison multiplexing*)**. Um esquema semelhante também é empregado na difusão de rádio AM e FM. A largura de banda do canal é compartilhada por diversos sinais sem nenhuma sobreposição.

Outro método de multiplexar vários sinais é conhecido como **multiplexação por divisão no tempo (TDM — *time division multiplexing*)**. Esse método é adequado quando um sinal tem a forma de um trem de pulsos (como em PCM). Quando os pulsos têm as larguras reduzidas, os espaços vazios entre pulsos de um sinal são preenchidos por pulsos de outros sinais. Assim, entrelaçando os trens de pulsos dos vários sinais em uma ordem específica, o tempo de transmissão é compartilhado por um certo número de sinais. No receptor, os trens de pulsos correspondentes aos diversos sinais são separados.

1.4.3 Demodulação

Uma vez que múltiplos sinais tenham chegado ao receptor, o sinal desejado deve ser detectado e sua forma original em banda base, recuperada. Notemos que, por conta da FDM, o primeiro estágio de um demodulador, em geral, requer um filtro passa-faixa, de modo que o receptor possa selecionar o sinal modulado em uma banda de frequências predeterminada especificada pela estação transmissora ou canal. Depois que um dado sinal modulado tenha sido isolado, o demodulador deve converter a variação de amplitude, frequência ou fase da portadora de volta ao sinal de tensão em banda base.

Para os três esquemas básicos de modulação, AM, FM e PM, os correspondentes demoduladores devem ser projetados de modo que a tensão de saída do detector varie proporcionalmente à amplitude, frequência ou fase, respectivamente, do sinal modulado de entrada. Após a implementação de circuitos com tais características de resposta, o demodulador pode converter os sinais modulados (de RF) de volta aos sinais em banda base que representam a mensagem original, seja ela de áudio, vídeo ou dados.

1.5 CODIFICAÇÕES DIGITAIS DE FONTE E PARA CORREÇÃO DE ERROS

Como mencionado, a relação sinal-ruído (SNR) e a largura da banda são dois fatores que determinam o desempenho de uma dada comunicação. Diferentemente do que se passa em sistemas de comunicação analógica, sistemas digitais, em geral, adotam medidas agressivas para reduzir a taxa de dados da fonte e combater o ruído do canal. Em particular, a *codificação da fonte* é aplicada para gerar o menor número possível de bits para uma mensagem sem sacrificar a precisão da detecção. O combate aos erros advindos de ruído e interferências, por sua vez, requer a introdução sistemática de *redundância*

no transmissor, da qual o receptor pode se valer para corrigir erros causados por distorção do canal e ruído. Esse processo é conhecido como codificação para correção de erros no transmissor e decodificação para correção de erros no receptor.

Codificação de fonte e codificação para correção de erros são dois estágios sucessivos em um sistema de comunicação digital e atuam em uma batalha sem fim. Por um lado, o trabalho da codificação de fonte consiste em remover o máximo possível de redundância de uma mensagem, de modo a encurtar a sequência de mensagens digitais a ser transmitida. O objetivo da codificação de fonte é utilizar a menor largura de banda possível sem considerar ruído e interferência do canal. Por outro lado, a codificação para correção de erros introduz redundância de forma intencional e inteligente, para que, se erros ocorrerem na detecção, a redundância possa ajudar a corrigir os mais frequentes.

Aleatoriedade, Redundância e Codificação de Fonte

Para entender a codificação de fonte, é importante que, primeiro, discutamos o papel da *aleatoriedade* em comunicações. Como observado, o ruído do canal é um importante fator que limita o desempenho da comunicação, pois é aleatório e não pode ser removido por predição. A aleatoriedade também está intimamente associada aos sinais desejados em comunicações. Na verdade, ela é a essência das comunicações; significa imprevisibilidade ou incerteza de uma mensagem de fonte. Se uma fonte não tivesse imprevisibilidade, como aquele amigo que sempre quer repetir a mesma história de "como fui abduzido por um alienígena", a mensagem seria previamente conhecida e não conteria nenhuma informação. Do mesmo modo, se uma pessoa pisca, transmite alguma informação em um dado contexto. Contudo, se a pessoa piscar continuamente com a regularidade de um relógio, as piscadas não transmitirão nenhuma informação. Em resumo, um sinal previsível não é aleatório e é completamente redundante. Portanto, uma mensagem conterá informação somente se for imprevisível. Maior previsibilidade significa maior redundância e, em consequência, menos informação. Em contraste, sinais aleatórios têm mais imprevisibilidade, ou menos previsibilidade, e contêm mais informação.

A codificação de fonte reduz a redundância com base na previsibilidade da fonte da mensagem. O objetivo da codificação de fonte é representar o sinal da fonte por meio de códigos que sejam tão curtos quanto possível. Códigos mais curtos são mais eficientes, pois exigem menor tempo de transmissão a uma dada taxa de dados. Portanto, no processo de codificação e transmissão da parte imprevisível e aleatória do sinal, a codificação de fonte deve remover a redundância do sinal. As mensagens mais previsíveis contêm mais redundância e requerem códigos mais curtos; mensagens menos previsíveis contêm mais informação e devem ser codificadas com códigos mais longos. Com a representação de mensagens mais previsíveis por meio de códigos mais curtos e de mensagens menos previsíveis com códigos mais longos, obtemos uma codificação de fonte mais eficiente. Consideremos o código Morse, por exemplo; nesse código, diferentes combinações de traços e pontos (palavras de código) são associadas às letras do alfabeto. Para minimizar o tempo de transmissão, palavras de código mais curtas são usadas para representar as letras que ocorrem com mais frequência (mais previsíveis, como *e*, *t* e *a*); palavras de código mais longas são usadas para representar as letras que ocorrem raramente (menos previsíveis, como *x*, *q* e *z*). Dessa forma, em média, uma mensagem em português tende a seguir uma distribuição conhecida de letras, resultando em sequências de código mais curtas, que podem ser transmitidas rapidamente. Isso explica por que o código Morse é uma boa codificação de fonte.

Mostraremos no Capítulo 13 que, para sinais digitais, o tempo total de transmissão é minimizado se uma mensagem (ou símbolo) de probabilidade P for representada por uma palavra de código cujo comprimento é proporcional a $\log(1/P)$. Isso é conhecido como codificação (de fonte) pela entropia.

Codificação para Correção de Erros

A codificação para correção de erros desempenha um papel importante nas comunicações. Enquanto a codificação de fonte remove redundâncias, os códigos corretores de erros adicionam redundância. A introdução sistemática de redundância viabiliza a comunicação confiável.[4] Devido à redundância, se alguns bits forem corrompidos por ruído ou interferências, outros bits a eles relacionados podem aju-

dar a recuperá-los, permitindo que a mensagem seja decodificada com precisão, apesar dos erros no sinal recebido. Todas as linguagens são redundantes. Por exemplo, o inglês tem uma redundância de cerca de 50%; isto é, em média, podemos descartar metade das letras ou palavras do inglês sem que o significado de uma dada mensagem seja perdido. Isso também significa que, em qualquer mensagem em inglês, o orador ou escritor tem livre escolha de metade das letras ou palavras, em média. A outra metade é determinada pela estrutura estatística do idioma. Se todas as redundâncias do inglês forem removidas, a transmissão de um telegrama ou uma conversa telefônica poderia ser feita em cerca da metade do tempo. Contudo, caso ocorresse um erro no receptor, seria muito difícil entender o significado da mensagem recebida. A redundância em uma mensagem desempenha, portanto, um papel útil no combate aos ruídos e interferências do canal.

Pode parecer paradoxal que, na codificação de fonte, removamos redundância e, na subsequente codificação para correção de erros, adicionemos mais redundância. Para explicar por que isso faz sentido, consideremos a remoção total das redundâncias do inglês através da codificação de fonte. Isso encurtaria a mensagem em 50% (para economizar largura de banda). Entretanto, para correção de erros, podemos recuperar alguma redundância sistemática; porém, essa redundância bem projetada tem apenas a metade do comprimento da que foi removida pela codificação de fonte, mas provê a mesma proteção contra erros. Portanto, fica claro que uma boa combinação de codificação de fonte e codificação para correção de erros pode remover redundância ineficiente sem sacrificar a correção de erros. Na verdade, um problema muito comum nesse campo é a eterna busca da *codificação conjunta fonte-canal* que permita máxima remoção de redundância sem afetar a correção de erros.

Para comprovarmos como redundância viabiliza a correção de erros, consideremos um exemplo: para a transmissão de amostras com $L = 16$ níveis de quantização, podemos utilizar um grupo de quatro pulsos binários, como mostrado na Fig. 1.5. Nesse esquema de codificação, não há redundância. Caso ocorra um erro na recepção de até mesmo apenas um pulso, o receptor produzirá um valor errado. Aqui, podemos usar redundância para eliminar o efeito de possíveis erros causados por ruído ou imperfeições do canal. Por exemplo, se adicionarmos a cada palavra de código um pulso com polaridade tal que o número de pulsos positivos seja par, teremos um código capaz de detectar um erro em qualquer posição. Assim, à palavra de código **0001**, adicionamos um quinto pulso, de polaridade positiva, e obtemos a nova palavra de código **00011**. Assim, o número de pulsos positivos é 2 (par). Se ocorrer um erro em alguma posição, essa paridade será violada. O receptor saberá que ocorreu um erro e poderá solicitar retransmissão da mensagem. Esse é um esquema de codificação muito simples e pode detectar apenas a ocorrência de um erro; não é capaz de localizar ou corrigir o erro. Além disso, esse esquema também não é capaz de detectar um número par de erros. A introdução de mais redundância permite não apenas detectar erros, mas também corrigi-los. Por exemplo, para $L = 16$ pode ser mostrado que a adição adequada de três pulsos permite a detecção e a correção de um erro que ocorra em qualquer localização. Códigos corretores de erros serão discutidos em detalhes no Capítulo 14.

1.6 BREVE REVISÃO HISTÓRICA DE TELECOMUNICAÇÕES MODERNAS

As telecomunicações (literalmente: comunicações a distância) sempre foram fundamentais para a sociedade humana. Mesmo na antiguidade, governos e unidades militares dependiam fortemente das telecomunicações para reunir informação e emitir ordens. O primeiro tipo de telecomunicação se valeu de mensageiros a pé ou a cavalo; mas a necessidade de transportar uma mensagem curta por uma longa distância (como avisar uma cidade da aproximação de invasores) levou ao emprego de sinais de fogo ou fumaça. O uso de espelhos para refletir a luz do sol (heliógrafos) foi outra forma eficaz de telecomunicação. O primeiro uso registrado dessa técnica se deu na Grécia antiga. Espelhos de sinalização também são mencionados no relato feito por Marco Polo de sua viagem ao Extremo Oriente.[1] Essas antigas tecnologias de comunicação *visual* são digitais, o que não deixa de ser surpreendente. Fogo e fumaça, em diferentes configurações, formariam distintas palavras de código. Para esse tipo

de comunicação, pessoal especializado era posicionado em colinas ou montanhas nas proximidades de cidades gregas, formando uma cadeia de repetidores regenerativos. Na verdade, plataformas de sinalização com fogo ou fumaça ainda encontram-se espalhadas pela Grande Muralha da China. O mais interessante é que refletores ou lentes, equivalentes aos amplificadores e antenas que utilizamos hoje, também eram empregados para guiar a luz a distâncias maiores.

É natural que a montagem desses primitivos sistemas de comunicação *visual* fosse tediosa e que eles pudessem transmitir apenas alguns bits de informação por hora. Um sistema de comunicação visual muito mais eficiente foi desenvolvido há pouco mais de dois séculos. Em 1793, o francês Claude Chappe inventou o conceito de "telegrafia semafórica" e o explorou em uma série de experimentos. Esse sistema consistia em uma sequência de dispositivos sinalizadores, denominados semáforos, montados em torres e, geralmente, espaçados por uma distância de 10 km. (Um semáforo tinha a aparência de uma grande figura humana com bandeiras de sinalização nas duas mãos.) O operador de um semáforo de recepção transcreveria a mensagem visualmente, em geral com o auxílio de um telescópio, e a retransmitiria de sua torre para a próxima, e assim por diante. Esse telégrafo visual se tornou o principal sistema de comunicação na França e se espalhou por outros países, incluindo os Estados Unidos. O telégrafo semafórico acabou sendo substituído pelo telégrafo elétrico. Hoje em dia, apenas algumas ruas e pontos de referência remanescentes, identificados com a denominação "Colina do Telégrafo", nos lembram do papel desse sistema na história. Ainda no século XX, a comunicação visual (via lâmpadas de Aldis, bandeiras de sinalização ou heliógrafos) permaneceu uma importante parte da comunicação marítima.

Esses antigos sistemas de telecomunicações eram sistemas ópticos baseados em receptores visuais. Em consequência, podiam cobrir apenas distâncias em visada direta e dependiam de operadores humanos para decodificação dos sinais. Um evento importante que mudou a história das telecomunicações ocorreu em 1820, quando o dinamarquês Hans Christian Oersted descobriu a interação entre eletricidade e magnetismo.[2] **Michael Faraday** fez a descoberta crucial seguinte, que **mudou a história da eletricidade e das telecomunicações**, ao concluir que a corrente elétrica pode ser induzida em um condutor por um campo magnético variante no tempo, o que tornou possível a geração de eletricidade pelo movimento de campo magnético. Além disso, a transmissão de sinais elétricos também se tornou possível, pois a variação de um campo eletromagnético induz a alteração de corrente em um circuito distante. O aspecto notável da descoberta da indução de corrente por Faraday é ter estabelecido as bases para telecomunicação sem fio por distâncias sem visada direta e, o que é mais importante, ter mostrado como gerar eletricidade como uma fonte de energia para alimentar tais sistemas. O telégrafo elétrico foi inventado logo depois, dando origem à era dos sistemas de telecomunicação modernos.

Os sistemas de telecomunicações modernos percorreram um longo caminho desde seu surgimento. Como seria difícil detalhar todos os eventos históricos que marcaram o desenvolvimento recente das telecomunicações, usaremos a Tabela 1.1 para resumir alguns dos mais notáveis eventos na evolução dos sistemas de telecomunicações. Como nosso foco é a telecomunicação elétrica, não comentaremos a igualmente longa história das comunicações (por fibras) ópticas.

É digno de nota o fato de que todos os sistemas de telecomunicações primitivos eram sistemas digitais baseados em símbolos. Apenas após a invenção do sistema telefônico por Alexander Graham Bell é que **sinais vivos** analógicos foram transmitidos. Sinais vivos podem ser instantaneamente ouvidos ou vistos pelos usuários receptores. A invenção de Bell, que inicia uma nova era (comunicação analógica), é um importante marco na história das telecomunicações. A Fig. 1.7 mostra uma cópia de uma ilustração da patente revolucionária do telefone de Bell (de 1876). Historiadores da ciência, com frequência, aclamam tal invenção como a *mais valiosa* patente já concedida na história.

A invenção de sistemas telefônicos também marca o início da era da comunicação analógica e da transmissão de sinal vivo. Em uma linha separada e igualmente excitante, a comunicação sem fio teve início em 1887, quando Heinrich Hertz demonstrou, pela primeira vez, um modo para detectar a presença de ondas eletromagnéticas. O cientista francês Edouard Branly, o físico inglês Oliver Lodge e o inventor russo Alexander Popov deram importantes contribuições ao desenvolvimento do receptor de rádio. O gênio croata Nikola Tesla também fez grandes contribuições a essa área. Valendo-se de

Tabela 1.1
Importantes Eventos dos Dois Últimos Séculos das Telecomunicações*

Ano	Principais Eventos
1820	**Primeiro experimento de magnetismo causado por corrente elétrica (por Hans C. Oersted)**
1831	**Descoberta da indução de corrente por radiação eletromagnética (por Michael Faraday)**
1830-32	Nascimento do telégrafo (creditado a Joseph Henry e Pavel Schilling)
1837	Invenção do código Morse por Samuel F. B. Morse
1864	Teoria de ondas eletromagnéticas desenvolvida por James C. Maxwell
1866	Operação do primeiro cabo telegráfico transatlântico
1876	**Invenção do telefone por Alexander G. Bell**
1878	Primeira central telefônica em New Haven, Connecticut, EUA
1887	**Detecção de ondas eletromagnéticas por Heinrich Hertz**
1896	**Telegrafia sem fio (radiotelegrafia) patenteada por Guglielmo Marconi**
1901	Primeira transmissão radiotelegráfica transatlântica por Marconi
1906	**Primeira difusão de rádio com modulação em amplitude (por Reginald A. Fessenden)**
1907	Serviço regular de radiotelegrafia transatlântica
1915	Primeiro serviço telefônico transcontinental
1920	Primeiras estações comerciais de rádio AM
1921	Rádio móvel adotado pelo Departamento de Polícia de Detroit
1925	Demonstração do primeiro sistema de televisão (por Charles F. Jenkins)
1928	Primeira estação de televisão, W3XK, nos Estados Unidos
1935	**Primeira demonstração de rádio FM (por Edwin H. Armstrong)**
1941	Padrão NTSC de televisão em preto e branco
	Primeiro serviço comercial de rádio FM
1947	**Conceito celular proposto, pela primeira vez, por Bell Labs**
1948	**Primeiro artigo importante sobre teoria da informação publicado por Claude E. Shannon**
	Invenção do transistor, por William Shockley, Walter Brattain e John Bardeen
1949	Construção do código de Golay para correção de 3 (ou menos) erros de bits
1950	Construção dos códigos de Hamming para simples correções de erros
1953	Padrão NTSC para televisão em cores
1958	**Conceito de circuito integrado proposto por Jack Kilby (Texas Instruments)**
1960	**Construção dos poderosos códigos de Reed-Solomon para correção de erros**
1962	Desenvolvimento dos primeiros modems telefônicos para computadores: Bell Dataphone 103A (300 bits/s)
1962	Códigos corretores de erros por verificação de paridade de baixa densidade propostos por Robert G. Gallager
1968-9	Primeiros codificadores com correção de erros a bordo de sondas espaciais da NASA (Pioneer IX e Mariner VI)
1971	**Primeira rede de computadores sem fio: AlohaNet**
1973	Primeira demonstração de telefone celular portátil feita pela Motorola à Comissão Federal de Comunicação dos Estados Unidos
1978	Primeiro teste de telefonia celular móvel, pela AT&T
1984	Primeiro serviço AMPS (analógico) de telefone celular portátil, pela Motorola
1989	Desenvolvimento de modems DSL para conexões de alta velocidade com computadores
1991	Lançamento do primeiro serviço celular (digital) GSM (Finlândia)
	Desenvolvimento da primeira rede de área local (LAN) sem fio (AT&T-NCR)
1993	Estabelecimento do padrão digital ATSC
1993	Códigos-turbo propostos por Berrou, Glavieux e Thitimajshima
1996	Lançamento do primeiro serviço celular CDMA (IS-95)
	Primeira difusão de HDTV
1997	Padrão IEEE 80 2.11(b) para rede LAN sem fio
1998	Exploração comercial de ADSL em larga escala
1999	Padrão IEEE 80 2.11a para rede LAN sem fio
2000	Lançamento do primeiro serviço celular 3D
2003	Padrão IEEE 80 2.11g para rede LAN sem fio

*A menos que especificado de modo diferente, os eventos listados nesta tabela ocorreram nos Estados Unidos. (N.T.)

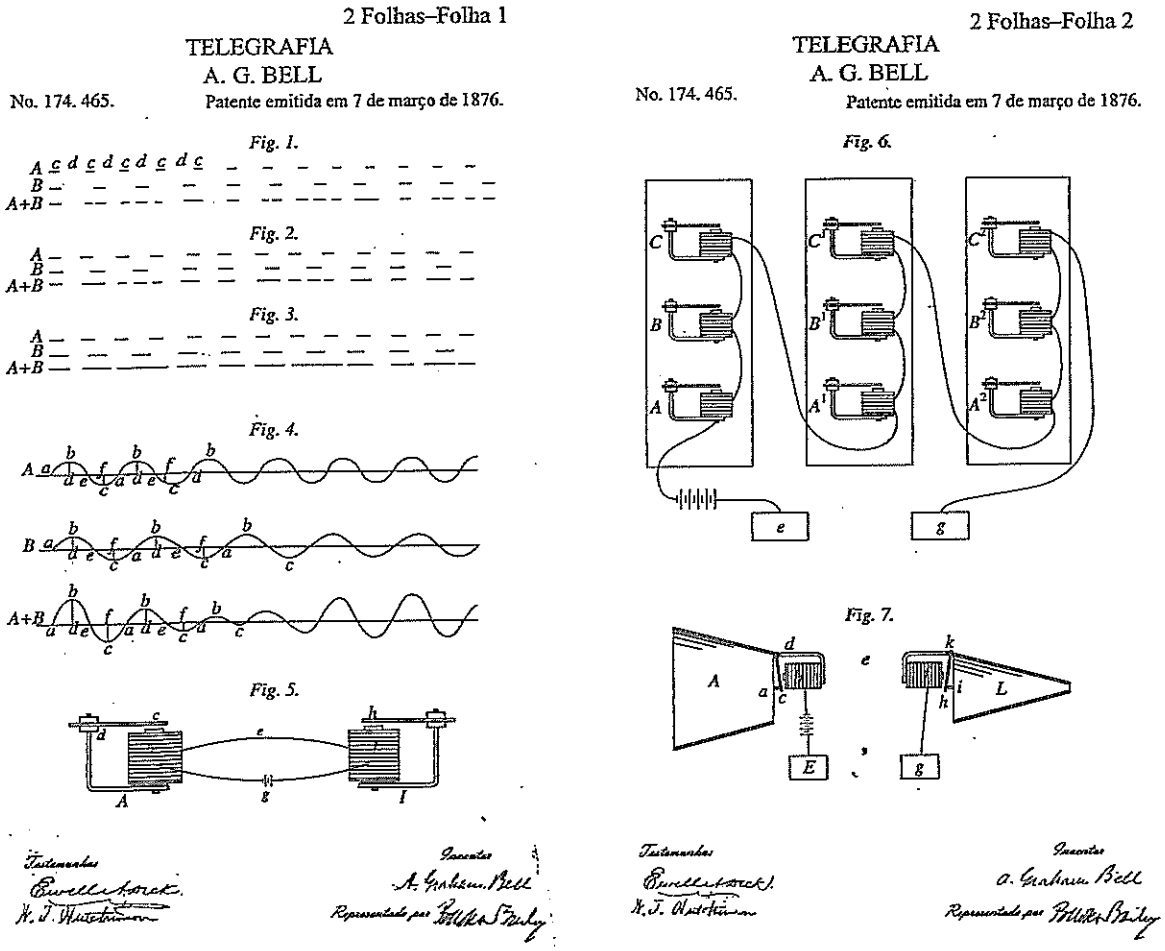


Figura 1.7 Ilustração da patente nº 174.465 obtida por Bell em 07 de março de 1876. (Dos arquivos do U.S. Patent and Trademark Office.)

experimentos e invenções anteriores, o cientista e inventor italiano Guglielmo Marconi desenvolveu, em 1895, um sistema de telégrafo sem fio pelo qual foi um dos agraciados com o Prêmio Nobel de Física de 1909. A telegrafia sem fio de Marconi foi um evento histórico das comunicações comerciais sem fio. Em pouco tempo, o casamento das invenções de Bell e Marconi permitiu que sinais analógicos de áudio fossem transmitidos sem fio, graças à tecnologia de modulação em amplitude (AM). A transmissão de música com qualidade via rádio FM foi demonstrada pela primeira vez pelo major Edwin H. Armstrong, inventor americano, em 1935, em uma reunião do IEEE (Institute of Electrical and Electronic Engineers) na cidade de Nova York, no Edifício Empire State.

Um ano histórico para as comunicações e para a eletrônica foi o de 1948: o ano que testemunhou o renascimento das comunicações digitais e a invenção do transistor semicondutor. O renascimento das comunicações digitais decorreu da criatividade e genialidade de Claude E. Shannon, conhecido como o pai da moderna teoria da comunicação e informação. Em dois artigos seminais publicados em 1948, ele estabeleceu, pela primeira vez, o fundamental conceito de capacidade do canal e sua relação com a taxa de transmissão de informação. Ao determinar a capacidade do canal de diversos modelos importantes, Shannon[3] provou que, desde que a informação fosse transmitida a uma taxa inferior à capacidade do canal, a comunicação sem erro era possível. No caso de canais ruidosos, Shannon mostrou a existência de bons códigos que tornariam a probabilidade de erros de transmissão arbitrariamente pequenos. Esse teorema relativo a canais ruidosos deu origem ao moderno campo de códigos corretores de erros. Por coincidência, a invenção do transistor (por Bill Shockley, Walter Brattain e John Bardeen) no mesmo ano preparou o caminho para o projeto e implementação de circuitos mais compactos, de maior potência e menos ruidosos que permitissem a exploração prática dos teoremas de Shannon. A sonda espacial Mariner IX, lançada em março de 1971 para exploração de Marte, foi

a primeira missão da NASA oficialmente equipada com códigos corretores de erros, que permitiram a transmissão confiável de fotografias de Marte.

Hoje, vivemos uma era de comunicação multimídia e digital, marcada pela ampla utilização de redes de computadores e telefones celulares. O primeiro modem telefônico para conexão de um computador doméstico a um *mainframe* foi desenvolvido pela AT&T Bell Labs em 1962. O dispositivo fazia uso de um acoplador acústico como interface a um aparelho telefônico comum. O acoplador acústico convertia os dados do computador local em tons audíveis e se valia do microfone do telefone para transmiti-los pelas linhas telefônicas. O acoplador recebia os dados do computador *mainframe* via receptor auricular do telefone e os convertia em bits para o computador local, a uma taxa típica abaixo de 300 bits/s. Na década de 1990, a taxa de transmissão atingiu 56 kbit/s, devido aos rápidos avanços das tecnologias de circuitos integrados (propostos por Jack Kilby em 1958) e de comunicações digitais. Nos anos 2000, foram desenvolvidos modems para redes de área local sem fio (WLAN — *wireless local area netwok*), permitindo a conexão de computadores a 11 Mbit/s. A versão comercial desses modems para WLAN, do tamanho de um cartão de crédito, foi padronizada sob a denominação IEEE 802.11b.

Os avanços tecnológicos também permitiram a reformatação de sistemas celulares. Embora o conceito celular tivesse sido desenvolvido pela Bell Labs em 1947, sistemas comerciais surgiram apenas em 1983. Os telefones "móveis" da década de 1980 eram grandes e caros, usados principalmente por empresas. O primeiro telefone celular, desenvolvido pela Motorola em 1983 e denominado DynaTAC 8000X, pesava cerca de 800 g, ganhou o apelido de "tijolo" e custava quase quatro mil dólares. Esses telefones analógicos eram, basicamente, rádios duplex FM para transmissão de voz apenas. Hoje em dia, um telefone celular é, de fato, um dispositivo digital multimídia e multitarefa, não apenas útil para comunicação de voz, mas também para enviar e receber mensagens de correio eletrônico, acessar páginas da internet e exibir vídeos. Os dispositivos celulares atuais são muito pequenos e pesam poucos gramas. Agora, ao contrário do passado, telefones celulares são para as massas. Na verdade, na Europa há mais telefones celulares que pessoas. Na África, 13% da população adulta possui telefones celulares.*

Ao longo da história, o progresso da civilização humana tem sido inseparável dos avanços tecnológicos em telecomunicações, que tiveram um papel fundamental em praticamente todos os eventos importantes. Não é exagero afirmar que as telecomunicações ajudaram a moldar o mundo em que vivemos hoje e continuarão a definir nosso futuro. Os autores esperam, portanto, que este texto estimule o interesse de numerosos estudantes das tecnologias de telecomunicações. Com a discussão dos princípios essenciais dos modernos sistemas analógicos e digitais de comunicação, os autores esperam prover uma base sólida para o treinamento de futuras gerações de cientistas e engenheiros de comunicações.

REFERÊNCIAS

1. M. G. Murray, "Aimable Air/Sea Rescue Signal Mirrors," *The Bent of Tau Beta Pi*, pp. 29–32, Fall 2004.
2. B. Bunch and A. Hellemans, Eds., *The History of Science and Technology: A Browser's Guide to the Great Discoveries, Inventions, and the People Who Made Them from the Dawn of Time to Today*, Houghton Mifflin, Boston, 2004.
3. C. E. Shannon, "A Mathematical Theory of Communications," *Bell Syst. Tech. J.* part I: pp. 379–423; part II: 623–656, July 1948.
4. S. Lin and D. J. Costello Jr., *Error Control Coding*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2004.

*Em outubro de 2010, segundo a Agência Nacional de Telecomunicações — ANATEL, o número de telefones celulares no Brasil era 194.439.250. Em novembro de 2010, a população do país, segundo o Censo de 2010, realizado pelo Instituto Brasileiro de Geografia e Estatística — IBGE, era de 190.732.694 pessoas. (N.T.)

$x^{10} = x$

$x \log 10 = x$

Lio 24/4/13

2 SINAIS E ESPAÇO DE SINAIS

1 Fu2 pag. → termo. 1 ma. p/ ler no porto

Neste capítulo, discutiremos certos conceitos básicos relativos a sinais, que são processados por sistemas. Iniciaremos com uma explicação dos termos *sinal* e *sistema*.

Sinais

Um sinal, como implica o termo, é um conjunto de informação ou dados. Como exemplo, podemos citar o sinal de telefone ou o de televisão, os dados de vendas mensais de uma empresa, os preços de ações no fechamento do pregão da bolsa de valores (no caso do Brasil, a média do índice Bovespa). Em todos esses casos, os sinais são funções da variável independente *tempo*. No entanto, nem sempre é assim. Quando uma carga elétrica é distribuída em uma superfície, por exemplo, o sinal é a densidade de carga, uma função do *espaço*, e não do tempo. Neste livro, trataremos quase exclusivamente de sinais que são funções do tempo. A discussão, entretanto, se aplica também a outras variáveis independentes.

Sistemas

Sinais podem ser processados por **sistemas**, que podem modificá-los ou deles extrair informação adicional. Por exemplo, um lançador de míssil antiaéreo pode necessitar conhecer a localização futura de um alvo móvel hostil, que é rastreado pelo radar. Como o sinal de radar fornece a localização passada e a velocidade do alvo, com o processamento adequado do sinal do radar (a entrada) é possível estimar a posição futura do alvo. Assim, um sistema é uma entidade que *processa* um conjunto de sinais (**entradas**) para produzir outro conjunto de sinais (**saídas**). Um sistema pode ser constituído de componentes físicos, como em sistemas elétricos, mecânicos ou hidráulicos (implementação em hardware), ou pode ser um algoritmo que calcula uma saída a partir de um sinal de entrada (implementação em software).

2.1 TAMANHO DE UM SINAL

Energia de um Sinal

O tamanho de qualquer entidade é uma grandeza que indica sua intensidade. De maneira geral, um sinal varia com o tempo. Para estabelecer uma grandeza padrão que meça a intensidade de um sinal, normalmente vemos um sinal $g(t)$ como uma tensão através de um resistor de um ohm. Definimos a **energia do sinal** E_g do sinal $g(t)$ como a energia que a tensão $g(t)$ dissipa no resistor. De modo mais formal, definimos E_g (para um sinal de valores reais) como

$$E_g = \int_{-\infty}^{\infty} g^2(t)\, dt \tag{2.1}$$

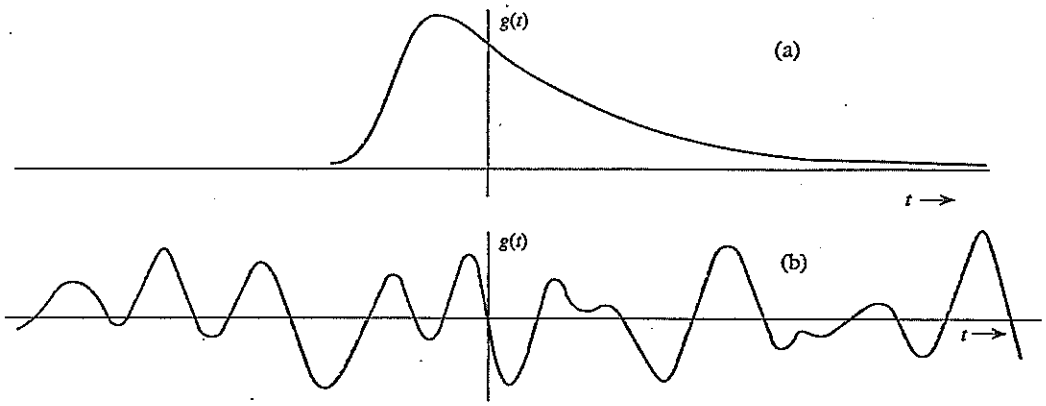


Figura 2.1 Exemplos de sinais. (a) Sinal com energia finita. (b) Sinal com potência finita.

Essa definição pode ser generalizada para um sinal $g(t)$ de valores complexos como

$$E_g = \int_{-\infty}^{\infty} |g(t)|^2 \, dt \tag{2.2}$$

Potência de um Sinal

Para ser uma medida do tamanho do sinal que faça sentido, a energia do sinal deve ser finita. Uma condição necessária para que a energia seja finita é que a amplitude do sinal tenda a zero à medida que $|t|$ tenda ao infinito (Fig. 2.1a). Caso contrário, a integral na Eq. (2.1) não convergirá.

Se a amplitude de $g(t)$ não tender a zero à medida que $|t|$ tender ao infinito (Fig. 2.1b), a energia do sinal será infinita. Em casos como esse, uma medida mais razoável do tamanho do sinal é a média temporal da energia (caso exista), que é a potência média P_g, definida (para um sinal de valores reais) por

$$P_g = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} g^2(t) \, dt \tag{2.3}$$

Podemos generalizar esta definição para um sinal $g(t)$ de valores complexos como

$$P_g = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} |g(t)|^2 \, dt \tag{2.4}$$

Observemos que a potência do sinal P_g é a média temporal (valor médio) da amplitude quadrática do sinal, ou seja, o valor **quadrático médio** de $g(t)$. Na verdade, a raiz quadrada de P_g é o valor **rms** (*Root Mean Square*), ou valor eficaz, de $g(t)$.

O valor médio de uma grandeza, calculado em um longo intervalo de tempo que tende ao infinito, existe se a grandeza for periódica ou tiver regularidade estatística. Caso essa condição não seja satisfeita, pode não existir um valor médio. Por exemplo, um sinal rampa $g(t) = t$ aumenta indefinidamente quando $|t| \to \infty$, de modo que nem energia e nem potência existirão para esse sinal.

Unidades de Energia e de Potência de um Sinal

As unidades padrão de energia e de potência de um sinal são, respectivamente, joule e watt. Contudo, na prática, é comum o uso de escalas logarítmicas para descrever a potência de um sinal. Esta notação evita a dificuldade de trabalhar com casas decimais quando a potência do sinal é muito alta ou muito baixa. Por convenção, dizemos que um sinal com potência média de P watts tem potência de

$$[10 \cdot \log_{10} P] \text{ dBw} \qquad \text{ou} \qquad [30 + 10 \cdot \log_{10} P] \text{ dBm}$$

Por exemplo, –30 dBm representa uma potência de sinal de 10^{-6} W na escala decimal padrão.

a potência do cosseno = $\frac{A^2}{2}$

O somatório de cossenos dá um sinal periódico

Exemplo 2.1 Determinemos medidas razoáveis dos sinais na Fig. 2.2.

O sinal na Fig. 2.2a tende a 0 à medida que $|t| \to \infty$. Portanto, uma medida razoável para este sinal é a energia E_g, dada por

$$E_g = \int_{-\infty}^{\infty} g^2(t)\,dt = \int_{-1}^{0} (2)^2\,dt + \int_{0}^{\infty} 4e^{-t}\,dt = 4 + 4 = 8$$

O sinal na Fig. 2.2b não tende a 0 à medida que $|t| \to \infty$. Entretanto, este sinal é periódico; logo, sua potência existe e podemos calculá-la usando a Eq. (2.3). Para sinais periódicos, este processo pode ser simplificado se observamos que um sinal periódico se repete regularmente a cada período (2 segundos, neste caso). Assim, calcular a média de $g^2(t)$ em um intervalo de tempo infinitamente longo é equivalente a calculá-la em um período (2 segundos, neste caso), ou seja

$$P_g = \frac{1}{2}\int_{-1}^{1} g^2(t)\,dt = \frac{1}{2}\int_{-1}^{1} t^2\,dt = \frac{1}{3}$$

Período = Potência

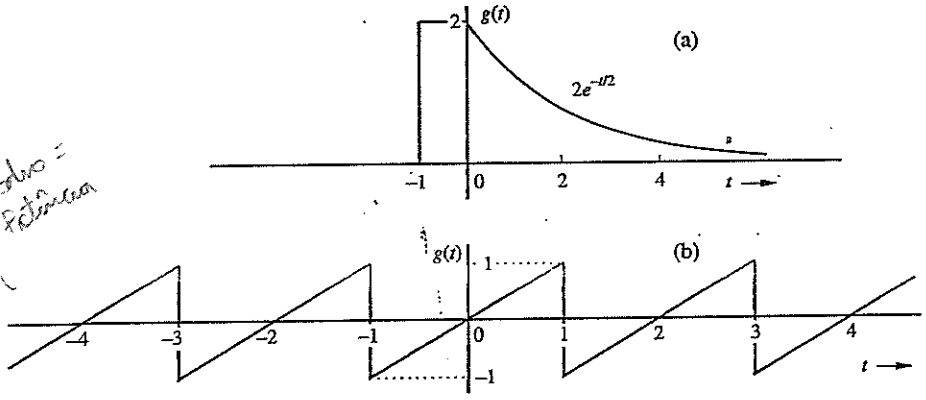


Figura 2.2 Sinais para o Exemplo 2.1.

Recordemos que a potência de um sinal é o quadrado de seu valor rms. Portanto, o valor rms deste sinal é $1/\sqrt{3}$.

2.2 CLASSIFICAÇÃO DE SINAIS

Existem várias classes de sinais. Aqui, consideraremos apenas os seguintes pares de classes, adequadas ao escopo deste livro:

1. Sinais em tempo contínuo e sinais em tempo discreto
2. Sinais analógicos e sinais digitais
3. Sinais periódicos e sinais aperiódicos
4. Sinais de energia e sinais de potência
5. Sinais determinísticos e sinais probabilísticos

2.2.1 Sinais em Tempo Contínuo e Sinais em Tempo Discreto

Um sinal que é especificado em todos os valores do tempo t (Fig. 2.3a) é um **sinal em tempo contínuo**; um sinal que é especificado apenas em pontos discretos $t = nT$ (Fig. 2.3b) é um **sinal em tem-**

Ex 2.2 → 3ª Edição → procurar na internet e baixar, pois irá ser requerido toda hora!

Ok! Tem um exemplo parecido do livro - sinais e sistemas

a) Det P_g e $g(t)_{rms}$

$g(t) = C\cos(\omega_0 t + \theta)$

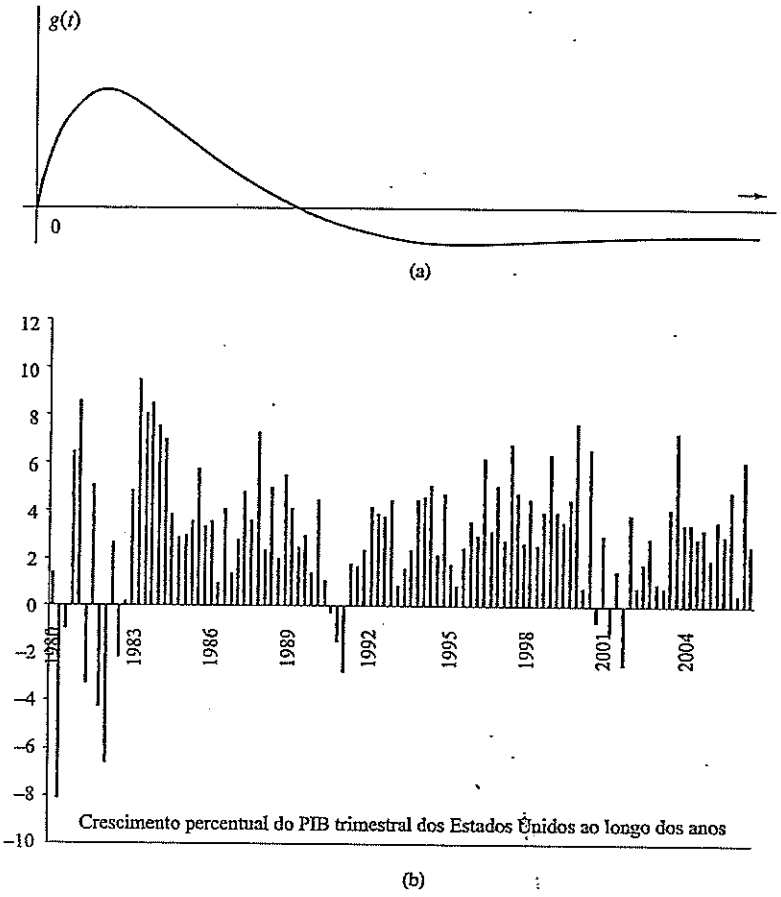


Figura 2.3 (a) Sinal em tempo contínuo. (b) Sinal em tempo discreto

po discreto. Gravações de áudio ou de vídeo são sinais em tempo contínuo, o produto interno bruto (PIB) trimestral, venda mensal de uma empresa e índice médio diário da bolsa de valores são sinais em tempo discreto.

2.2.2 Sinais Analógicos e Sinais Digitais

Não devemos confundir sinais analógicos com sinais em tempo contínuo. Os dois conceitos não são iguais. Isso também se aplica aos conceitos de sinais digitais e sinais em tempo discreto. Um sinal cuja amplitude pode assumir apenas valores em um intervalo contínuo de tempo é um **sinal analógico**, o que significa que a amplitude de um sinal analógico pode assumir um número infinito (incontável) de valores. Um **sinal digital**, por sua vez, é aquele cuja amplitude pode assumir somente um número finito de valores. Sinais associados com um computador digital são digitais porque assumem apenas dois valores (sinais binários). Para que um sinal se qualifique como digital, o número de valores não precisa ser limitado a dois: pode ser um número finito qualquer. Um sinal digital cuja amplitude pode assumir M valores é um sinal **M-ário**, do qual o binário ($M = 2$) é um caso especial. Os termos "tempo contínuo" e "tempo discreto" descrevem a natureza do sinal ao longo do eixo do tempo (eixo horizontal). Os termos "analógico" e "digital" descrevem a natureza do eixo de amplitude do sinal (eixo vertical). A Fig. 2.4 mostra exemplos de sinais de vários tipos. Fica claro que analógico não é necessariamente em tempo contínuo, assim como digital não tem que ser em tempo discreto. A Fig. 2.4c mostra um exemplo de um sinal analógico em tempo discreto. Um sinal analógico pode ser convertido em um sinal digital (via conversão analógico-digital, A/D) por meio de quantização (arredondamento), como explicado no Capítulo 6.

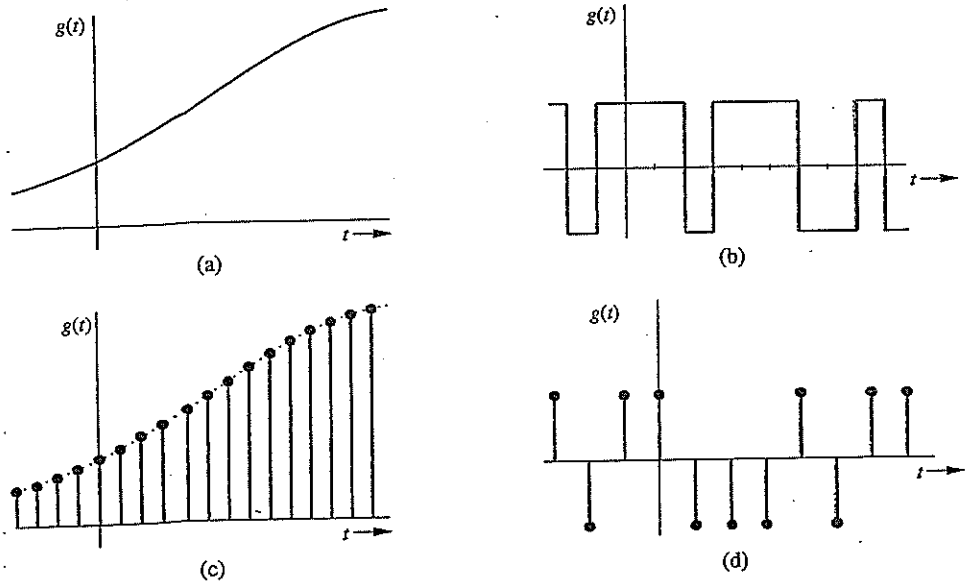


Figura 2.4 Exemplos de sinais: (a) Sinal analógico em tempo contínuo, (b) Sinal digital em tempo contínuo, (c) Sinal analógico em tempo discreto, (d) Sinal digital em tempo discreto.

2.2.3 Sinais Periódicos e Sinais Aperiódicos

Um sinal $g(t)$ é dito **periódico** se existir uma constante positiva T_0 tal que

$$g(t) = g(t + T_0) \quad \text{para todo } t \tag{2.5}$$

O **menor** valor de T_0 que satisfaz a condição de periodicidade da Eq. (2.5) é o **período** de $g(t)$. O sinal na Fig. 2.2b é um sinal periódico com período 2. Naturalmente, um sinal é **aperiódico** se não for periódico. O sinal na Fig. 2.2a é aperiódico.

Por definição, um sinal periódico $g(t)$ não se altera quando sofre deslocamento ou translação temporal igual a um período. Isso significa que um sinal periódico deve ter início em $t = -\infty$, pois, se tivesse início em algum instante de tempo finito, digamos, $t = 0$, o sinal deslocado $g(t + T_0)$ teria início em $t = -T_0$ e, portanto, $g(t + T_0)$ não seria o mesmo que $g(t)$. Em consequência, um **sinal periódico, por definição, deve ter início em $t = -\infty$ e continuar indefinidamente**, como ilustrado na Fig. 2.5. Vale observar que um sinal periódico não se altera quando é deslocado no tempo por um múltiplo de T_0. Dessa forma, $g(t)$ pode ser considerado um sinal periódico com período mT_0, sendo m um inteiro qualquer. No entanto, pela definição, o período é o menor intervalo que satisfaz a condição de periodicidade da Eq. (2.5). Ou seja, o período é T_0.

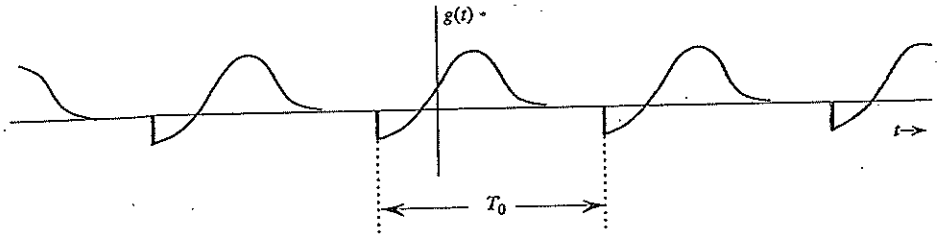


Figura 2.5 Sinal periódico de período T_0.

2.2.4 Sinais de Energia e Sinais de Potência

Um sinal com energia finita é um **sinal de energia**; um sinal com potência finita é um **sinal de potência**. Em outras palavras, um sinal $g(t)$ é um sinal de energia se

$$\int_{-\infty}^{\infty} |g(t)|^2 \, dt < \infty \tag{2.6}$$

De modo similar, um sinal com potência (valor quadrático médio) finita(o) e não nula(o) é um sinal de potência. Em outras palavras, um sinal é de potência se

$$0 < \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} |g(t)|^2 \, dt < \infty \tag{2.7}$$

Os sinais nas Figs. 2.2a e 2.2b são exemplos de sinais de energia e de potência, respectivamente. Observemos que potência é o valor temporal de energia. Como o valor é tomado em um intervalo de tempo infinitamente longo, um sinal com energia finita tem potência nula, e um sinal com potência finita tem energia infinita. Portanto, um sinal não pode, ao mesmo tempo, ser de energia e de potência. Se for um, não poderá ser o outro. Contudo, alguns sinais com potência infinita não são nem de energia nem de potência. O sinal rampa é um exemplo.

Comentários

Todo sinal observado na vida real é um sinal de energia. Um sinal de potência, por sua vez, deve ter uma duração infinita. Caso contrário, sua potência, que é o valor temporal de energia (calculado em um intervalo de tempo infinitamente longo) não tenderá a um limite (não nulo). É óbvio que, na prática, é impossível gerar um verdadeiro sinal de potência, pois este sinal deveria ter duração e energia infinitas.

Por fim, devido à repetição periódica, sinais periódicos para os quais a área sob a curva de $|g(t)|^2$ ao longo de um período é finita são sinais de potência; no entanto, nem todos os sinais de potência são periódicos.

2.2.5 Sinais Determinísticos e Sinais Aleatórios

Um sinal cuja descrição física seja completamente conhecida, de forma matemática ou gráfica, é um **sinal determinístico**. Um sinal que seja conhecido apenas em termos de uma descrição probabilística, como valor médio, valor quadrático médio ou distribuições, e não de sua completa descrição matemática ou gráfica, é um **sinal aleatório**. A maioria dos sinais de ruído encontrados na prática é de sinais aleatórios. Todos os sinais de mensagens são sinais aleatórios, pois, como mostraremos posteriormente, para que um sinal transporte informação, deve haver alguma incerteza (aleatoriedade) em relação ao mesmo. O tratamento de sinais aleatórios será discutido em capítulos ulteriores.

2.3 SINAL IMPULSO UNITÁRIO

A função impulso unitário $\delta(t)$ é uma das mais importantes funções no estudo de sinais e sistemas. Suas definição e aplicação resultam em conveniências não permitidas em matemática pura.

A função impulso unitário $\delta(t)$ foi definida por P. A. M. Dirac (e, por isso, também é conhecida como função "delta de Dirac") como

$$\delta(t) = 0, \quad t \neq 0 \tag{2.8}$$

$$\int_{-\infty}^{\infty} \delta(t) \, dt = 1 \tag{2.9}$$

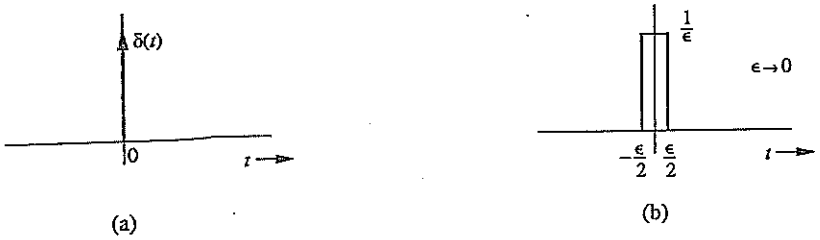


Figura 2.6
Um impulso unitário e sua aproximação.

Podemos visualizar um impulso como um pulso retangular alto e estreito de área unitária, como mostrado na Fig. 2.6. A largura desse pulso retangular é um valor muito pequeno ϵ; no limite $\epsilon \to 0$, sua altura tem um valor muito grande $1/\epsilon$. O impulso unitário pode, portanto, ser considerado um pulso retangular cuja largura se torna infinitamente pequena, cuja altura se torna infinitamente grande e uma área total que permanece constante e igual à unidade.* Assim, $\delta(t) = 0$ para todos os valores de t, exceto $t = 0$, onde, formalmente, a função não é definida. Por essa razão, um impulso unitário é representado de forma gráfica por uma seta, como na Fig. 2.6a.

Multiplicação de uma Função por um Impulso

Consideremos, agora, o que ocorre quando multiplicamos o impulso unitário $\delta(t)$ por uma função $\phi(t)$, contínua em $t = 0$. Como o impulso existe apenas em $t = 0$ e o valor de $\phi(t)$ em $t = 0$ é $\phi(0)$, obtemos

$$\phi(t)\delta(t) = \phi(0)\delta(t) \tag{2.10a}$$

De modo similar, se $\phi(t)$ for multiplicada por um impulso $\delta(t - T)$ (um impulso localizado em $t = T$), obtemos

$$\phi(t)\delta(t - T) = \phi(T)\delta(t - T) \tag{2.10b}$$

desde que $\phi(t)$ seja definida em $t = T$.

Propriedade de Amostragem da Função Impulso Unitário

Da Eq. (2.10), temos que

$$\int_{-\infty}^{\infty} \phi(t)\delta(t - T)\,dt = \phi(T)\int_{-\infty}^{\infty} \delta(t - T)\,dt = \phi(T) \tag{2.11a}$$

desde que $\phi(t)$ seja contínua em $t = T$. Esse resultado significa que *a área sob a curva do produto de uma função por um impulso $\delta(t)$ é igual ao valor da função no instante de tempo em que o impulso está localizado*. Essa importante e útil propriedade é conhecida como **propriedade de amostragem** (ou peneiramento) do impulso unitário.

Dependendo do valor de T e dos limites de integração, a função impulso pode ou não estar no intervalo de integração. Assim, temos

$$\int_a^b \phi(t)\delta(t - T)\,dt = \phi(T)\int_a^b \delta(t - T)\,dt = \begin{cases} \phi(T) & a \le T < b \\ 0 & T < a \le b, \text{ ou } T \ge b > a \end{cases} \tag{2.11b}$$

*A função impulso também pode ser aproximada por outros pulsos, como um triângulo positivo, pulso exponencial ou pulso gaussiano.

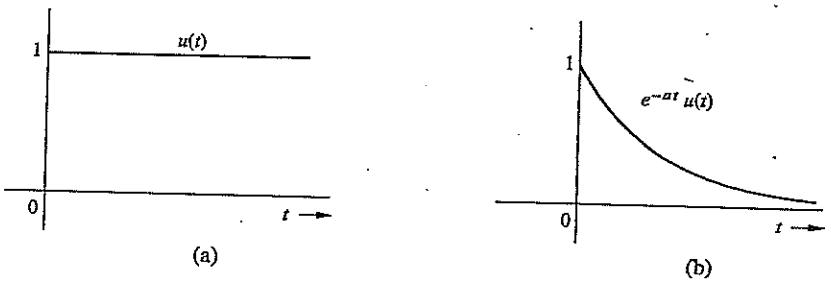


Figura 2.7 (a) Função degrau unitário $u(t)$. (b) Exponencial causal $e^{-at}u(t)$.

Função Degrau Unitário $u(t)$

Outra função útil e familiar é a função **degrau unitário** $u(t)$, encontrada com frequência na análise de circuitos e definida na Fig. 2.7a como:

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{2.12}$$

Se desejarmos que um sinal tenha início em $t = 0$ (de modo que tenha valor nulo para $t < 0$), basta que o multipliquemos por $u(t)$. Um sinal que começa após $t = 0$ é denominado **sinal causal**. Em outras palavras, $g(t)$ é um sinal causal se

$$g(t) = 0 \qquad t < 0$$

O sinal e^{-at} representa uma exponencial que tem início em $t = -\infty$. Se desejarmos que este sinal comece em $t = 0$ (forma causal), podemos descrevê-lo como $e^{-at}u(t)$ (Fig. 2.7b). Da Fig. 2.6b, observamos que, de $-\infty$ a t, a área sob a curva da forma aproximada de $\delta(t)$ é zero se $t < 0$ e igual à unidade se $t \geq 0$. Portanto, podemos escrever

$$\int_{-\infty}^{t} \delta(\tau)\, d\tau = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$$
$$= u(t) \tag{2.13a}$$

Deste resultado, obtemos

$$\frac{du}{dt} = \delta(t) \tag{2.13b}$$

2.4 ANALOGIA ENTRE SINAIS E VETORES

Existe uma forte conexão entre sinais e vetores. Sinais definidos em apenas um número finito (digamos, N) de instantes de tempo podem ser escritos como vetores (de dimensão N). Assim, consideremos um sinal $g(t)$ definido em um intervalo fechado $[a, b]$. Tomemos N pontos uniformemente espaçados no intervalo $[a, b]$, tais que

$$t_1 = a, \quad t_2 = a + \epsilon, \quad t_3 = a + 2\epsilon, \quad t_N = a + (N-1)\epsilon = b, \qquad \epsilon = \frac{b-a}{N-1}$$

Podemos, então, escrever um sinal vetorial $\mathbf{g}$ como um vetor de dimensão N

$$\mathbf{g} = [\ g(t_1) \quad g(t_2) \quad \cdots \quad g(t_N)\ ]$$

À medida que o número N de instantes de tempo aumenta, a dimensão do sinal vetorial amostrado $\mathbf{g}$ aumentará. Se $N \to \infty$, os valores do sinal formarão um vetor $\mathbf{g}$ de dimensão infinita. Como, neste

caso, $\epsilon \to 0$, o sinal vetorial g se transformará no sinal em tempo contínuo $g(t)$, definido no intervalo $[a, b]$. Em outras palavras,

$$\lim_{N\to\infty} \mathbf{g} = g(t) \quad t \in [a, b]$$

Essa relação mostra claramente que sinais em tempo contínuo são generalizações diretas de vetores de dimensão finita. Em consequência, definições e operações básicas em um espaço vetorial também podem ser aplicadas a sinais em tempo contínuo. A seguir, exploraremos esta conexão entre o espaço de vetores de dimensão finita e o espaço de sinais em tempo contínuo.

Denotaremos todos os vetores por letras em negrito. Por exemplo, $\mathbf{x}$ é um certo vetor com magnitude ou comprimento $||\mathbf{x}||$. Um vetor tem magnitude e direção. Em um espaço vetorial, podemos definir o produto interno (ou escalar) de dois vetores de valores reais $\mathbf{g}$ e $\mathbf{x}$ como

$$< \mathbf{g}, \mathbf{x} >= ||\mathbf{g}|| \cdot ||\mathbf{x}|| \cos\theta \tag{2.14}$$

em que θ é o ângulo entre os vetores $\mathbf{g}$ e $\mathbf{x}$. Com essa definição, podemos expressar $||\mathbf{x}||$, o comprimento (ou norma) do vetor $\mathbf{x}$ como

$$||\mathbf{x}||^2 =< \mathbf{x}, \mathbf{x} > \tag{2.15}$$

Isso define um espaço vetorial normado.

2.4.1 Componente de um Vetor na Direção de Outro Vetor

Consideremos dois vetores $\mathbf{g}$ e $\mathbf{x}$, como mostrado na Fig. 2.8. Seja $c\mathbf{x}$ a componente de $\mathbf{g}$ na direção de $\mathbf{x}$ (ou ao longo de $\mathbf{x}$). Do ponto de vista geométrico, a componente de $\mathbf{g}$ ao longo de $\mathbf{x}$ é a projeção de $\mathbf{g}$ sobre $\mathbf{x}$, obtida traçando, a partir da extremidade de $\mathbf{g}$, uma linha reta perpendicular ao vetor $\mathbf{x}$, como ilustrado na Fig. 2.8. Qual é o significado matemático da componente de um vetor na direção de outro vetor? Como visto na Fig. 2.8, o vetor $\mathbf{g}$ pode ser expresso em termos do vetor $\mathbf{x}$ como

$$\mathbf{g} = c\mathbf{x} + \mathbf{e} \tag{2.16}$$

Contudo, isso não descreve uma única forma de decompor $\mathbf{g}$ em termos de $\mathbf{x}$ e de $\mathbf{e}$. A Fig. 2.9 mostra duas das outras infinitas possibilidades. Das Fig. 2.9a e b, temos

$$\mathbf{g} = c_1\mathbf{x} + \mathbf{e}_1 = c_2\mathbf{x} + \mathbf{e}_2 \tag{2.17}$$

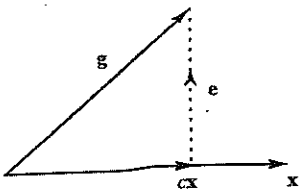
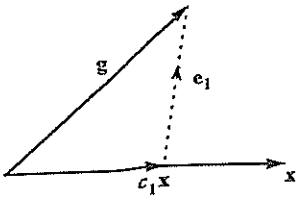
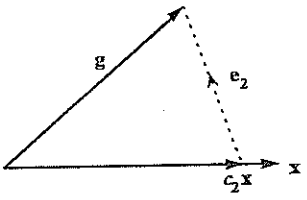


Figura 2.8 Componente (projeção) de um vetor ao longo de outro vetor.



(a)



(b)

Figura 2.9 Aproximações de um vetor em termos de outro vetor.

Cabe a seguinte pergunta: qual é a "melhor" decomposição? O conceito de ótimo depende de nosso objetivo ao decompor **g** em duas componentes.

Em cada uma das três representações anteriores, **g** é dado em termos de **x** e de outro vetor, denominado **vetor de erro**. Se nosso objetivo for aproximar **g** por $c\mathbf{x}$ (Fig. 2.8),

$$\mathbf{g} \simeq \hat{\mathbf{g}} = c\mathbf{x} \tag{2.18}$$

o erro nessa aproximação é o vetor (diferença) $\mathbf{e} = \mathbf{g} - c\mathbf{x}$. De modo similar, os erros nas aproximações nas Figs. 2.9a e b são $\mathbf{e}_1$ e $\mathbf{e}_2$, respectivamente. A aproximação na Fig. 2.8 é única, pois seu vetor de erro é o menor (menor magnitude ou norma). Podemos, agora, definir matematicamente a componente (ou projeção) de um vetor **g** na direção de um vetor **x** como $c\mathbf{x}$, onde c é escolhido para minimizar a magnitude do vetor de erro $\mathbf{e} = \mathbf{g} - c\mathbf{x}$.

Do ponto de vista geométrico, a magnitude da componente de **g** ao longo de **x** é $||\mathbf{g}||\cos\theta$, que também é igual a $c||\mathbf{x}||$. Portanto,

$$c||\mathbf{x}|| = ||\mathbf{g}||\cos\theta$$

Com base na definição do produto interno de dois vetores, multiplicando os dois lados por $||\mathbf{x}||$, obtemos

$$c||\mathbf{x}||^2 = ||\mathbf{g}||\,||\mathbf{x}||\cos\theta = <\mathbf{g}, \mathbf{x}>$$

e

$$c = \frac{<\mathbf{g}, \mathbf{x}>}{<\mathbf{x}, \mathbf{x}>} = \frac{1}{||\mathbf{x}||^2} <\mathbf{g}, \mathbf{x}> \tag{2.19}$$

Da Fig. 2.8, fica aparente que, quando **g** e **x** são perpendiculares ou ortogonais, **g** tem componente nula na direção de **x**; neste caso, $c = 0$. Tendo em vista a Eq. (2.19), podemos definir **g** e **x** como **ortogonais** se o produto interno dos dois vetores for zero, ou seja; se

$$<\mathbf{g}, \mathbf{x}> = 0 \tag{2.20}$$

2.4.2 Decomposição de um Sinal e Componentes de um Sinal

Os conceitos de componentes e ortogonalidade de vetores podem ser diretamente estendidos a sinais em tempo contínuo. Consideremos o problema de aproximar um sinal de valores reais $g(t)$ em termos de outro sinal de valores reais $x(t)$, em um intervalo $[t_1, t_2]$:

$$g(t) \simeq cx(t) \qquad t_1 \le t \le t_2 \tag{2.21}$$

O erro $e(t)$ nesta aproximação é

$$e(t) = \begin{cases} g(t) - cx(t) & t_1 \le t \le t_2 \\ 0 & \text{para qualquer outro valor de } t \end{cases} \tag{2.22}$$

Para a "melhor aproximação", precisamos minimizar o sinal de erro, ou seja, minimizar a norma do mesmo. A mínima norma de um sinal corresponde à mínima energia E_e no intervalo $[t_1, t_2]$, dada por

$$\begin{aligned} E_e &= \int_{t_1}^{t_2} e^2(t)\,dt \\ &= \int_{t_1}^{t_2} [g(t) - cx(t)]^2\,dt \end{aligned}$$

Observemos que o lado direito é uma integral definida cuja variável de integração é t. Portanto, E_e é uma função do parâmetro c (não de t), e E_e será mínima para alguma escolha do valor de c. Uma condição necessária para a minimização de E_e é

$$\frac{dE_e}{dc} = 0 \tag{2.23}$$

ou

$$\frac{d}{dc}\left[\int_{t_1}^{t_2} [g(t) - cx(t)]^2\, dt\right] = 0$$

Expandindo o termo quadrático no integrando, obtemos

$$\frac{d}{dc}\left[\int_{t_1}^{t_2} g^2(t)\, dt\right] - \frac{d}{dc}\left[2c\int_{t_1}^{t_2} g(t)x(t)\, dt\right] + \frac{d}{dc}\left[c^2\int_{t_1}^{t_2} x^2(t)\, dt\right] = 0$$

portanto,

$$-2\int_{t_1}^{t_2} g(t)x(t)\, dt + 2c\int_{t_1}^{t_2} x^2(t)\, dt = 0$$

e

$$c = \frac{\int_{t_1}^{t_2} g(t)x(t)\, dt}{\int_{t_1}^{t_2} x^2(t)\, dt} = \frac{1}{E_x}\int_{t_1}^{t_2} g(t)x(t)\, dt \tag{2.24}$$

Em resumo, se um sinal $g(t)$ for aproximado por outro sinal $x(t)$ na forma

$$g(t) \simeq cx(t)$$

o valor ótimo de c que minimiza a energia do sinal de erro nesta aproximação é dado pela Eq. (2.24).

Seguindo o que é feito com vetores, dizemos que o sinal $g(t)$ tem uma componente $cx(t)$, com c dado pela Eq. (2.24). Como no caso de um espaço vetorial, $cx(t)$ é a projeção de $g(t)$ sobre $x(t)$. Mantendo a terminologia de espaço vetorial, se a componente de um sinal $g(t)$ da forma $cx(t)$ for zero (ou seja, se $c = 0$), dizemos que os sinais $g(t)$ e $x(t)$ são ortogonais no intervalo $[t_1, t_2]$. Em outras palavras, para sinais de valores reais, dois sinais $g(t)$ e $x(t)$ são ortogonais quando a contribuição de um sinal ao outro for zero ($c = 0$). Assim, $x(t)$ e $g(t)$ são ortogonais se e somente se

$$\int_{t_1}^{t_2} g(t)x(t)\, dt = 0 \tag{2.25}$$

Com base nas ilustrações de vetores na Fig. 2.9, podemos dizer que dois sinais são ortogonais se e somente se o produto interno dos dois for zero. Isto indica que a integral da Eq. (2.25) guarda uma relação muito próxima com o conceito de produto interno de vetores.

A definição padrão de produto interno de dois vetores $\mathbf{g}$ e $\mathbf{x}$ de dimensão N

$$< \mathbf{g}, \mathbf{x} >= \sum_{i=1}^{N} g_i x_i$$

tem forma quase idêntica à integral na Eq. (2.25). Portanto, definimos o produto interno de dois sinais (de valores reais) $g(t)$ e $x(t)$, ambos especificados no intervalo de tempo $[t_1, t_2]$, como

$$< g(t),\ x(t) >= \int_{t_1}^{t_2} g(t)x(t)\, dt \tag{2.26}$$

Recordemos da álgebra linear que o quadrado do comprimento de um vetor, $||\mathbf{x}||^2$, é igual a $< \mathbf{x}, \mathbf{x} >$. Com esse conceito em mente e prosseguindo com a analogia com análise vetorial, definimos a norma de um sinal $g(t)$ como

$$||g(t)|| = \sqrt{< g(t),\ g(t) >} \tag{2.27}$$

que é a raiz quadrada da energia do sinal no dado intervalo de tempo. Fica claro que a norma de um sinal é análoga ao comprimento de um vetor de dimensão finita. De modo mais geral, sinais podem não ser meramente definidos em um segmento contínuo de tempo $[t_1, t_2]$.*

Exemplo 2.2 Para o sinal quadrado $g(t)$ mostrado na Fig. 2.10, determinemos a componente em $g(t)$ da forma de sen t. Em outras palavras, aproximemos $g(t)$ em termos de sen t.

$$g(t) \simeq c \operatorname{sen} t \qquad 0 \le t \le 2\pi$$

de modo que a energia do sinal seja mínima.

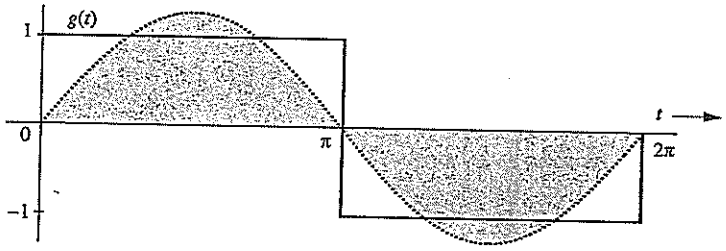


Figura 2.10 Aproximação do sinal quadrado em termos de uma única senoide.

Neste caso,

$$x(t) = \operatorname{sen} t \qquad \text{e} \qquad E_x = \int_0^{2\pi} \operatorname{sen}^2(t)\, dt = \pi$$

Da Eq. (2.24), obtemos

$$c = \frac{1}{\pi}\int_0^{2\pi} g(t) \operatorname{sen} t\, dt = \frac{1}{\pi}\left[\int_0^{\pi} \operatorname{sen} t\, dt + \int_{\pi}^{2\pi} (-\operatorname{sen} t)\, dt\right] = \frac{4}{\pi} \tag{2.29}$$

Portanto,

$$g(t) \simeq \frac{4}{\pi} \operatorname{sen} t \tag{2.30}$$

representa a melhor aproximação de $g(t)$ pela função sen t, que minimiza a energia do sinal de erro. Essa componente senoidal de $g(t)$ é mostrada hachurada na Fig. 2.10. Como em um espaço vetorial, dizemos que a função quadrada $g(t)$ ilustrada na Fig. 2.10 tem uma componente de sinal sen t com magnitude $4/\pi$.

2.4.3 Espaço Complexo de Sinais e Ortogonalidade

Até aqui, restringimos a atenção a funções de valores reais de t. Para estender os resultados a funções de valores complexos de t, consideremos, mais uma vez, o problema de aproximar a função $g(t)$ por uma função $x(t)$, em um intervalo $(t_1 \le t \le t_2)$

$$g(t) \simeq cx(t) \tag{2.31}$$

*De fato, o espaço de sinais em consideração pode consistir em um conjunto de intervalos de tempo representado simplesmente por Θ. No caso desses espaços de sinais mais genéricos, o produto interno é definido como uma integral em todo o domínio temporal Θ. Para sinais de valores complexos, o produto interno é modificado para

$$< g(t), x(t) >= \int_{\Theta} g(t)x^*(t)\, dt \tag{2.28}$$

Dada a definição de produto interno, a norma do sinal $\|g(t)\| = \sqrt{< g(t), g(t) >}$ e o espaço de sinais podem ser definidos para qualquer sinal no domínio do tempo.

em que $g(t)$ e $x(t)$ são funções de valores complexos de t. Em geral, tanto o coeficiente c como o sinal de erro

$$e(t) = g(t) - cx(t) \tag{2.32}$$

são complexos. Recordemos que a energia E_x do sinal complexo $x(t)$ no intervalo $[t_1, t_2]$ é

$$E_x = \int_{t_1}^{t_2} |x(t)|^2 \, dt$$

Para a melhor aproximação, devemos escolher o valor de c que minimiza E_e, a energia do sinal de erro $e(t)$, dada por

$$E_e = \int_{t_1}^{t_2} |g(t) - cx(t)|^2 \, dt \tag{2.33}$$

Recordemos, também, que

$$|u + v|^2 = (u + v)(u^* + v^*) = |u|^2 + |v|^2 + u^*v + uv^* \tag{2.34}$$

Usando esse resultado, podemos, após alguns cálculos, expressar a integral E_e na Eq. (2.33) como

$$E_e = \int_{t_1}^{t_2} |g(t)|^2 \, dt - \left| \frac{1}{\sqrt{E_x}} \int_{t_1}^{t_2} g(t)x^*(t) \, dt \right|^2 + \left| c\sqrt{E_x} - \frac{1}{\sqrt{E_x}} \int_{t_1}^{t_2} g(t)x^*(t) \, dt \right|^2$$

Já que os primeiros dois termos no lado direito independem de c, fica claro que E_e é minimizada com a escolha do valor de c que anula o terceiro termo. Isso leva ao coeficiente ótimo

$$c = \frac{1}{E_x} \int_{t_1}^{t_2} g(t)x^*(t) \, dt \tag{2.35}$$

Com esse resultado, devemos redefinir ortogonalidade para o caso complexo, o que é feito da seguinte forma: funções (sinais) de valores complexos $x_1(t)$ e $x_2(t)$ são ortogonais em um intervalo ($t_1 \leq t \leq t_2$) se

$$\int_{t_1}^{t_2} x_1(t)x_2^*(t) \, dt = 0 \qquad \text{ou} \qquad \int_{t_1}^{t_2} x_1^*(t)x_2(t) \, dt = 0 \tag{2.36}$$

Na verdade, basta qualquer uma das igualdades. Essa é uma definição geral de ortogonalidade, que se reduz à Eq. (2.25) quando as funções têm valores reais.

Do mesmo modo, a definição de produto interno para sinais de valores complexos em um domínio temporal Θ pode ser modificada como:

$$< g(t), \, x(t) >= \int_{\{t:t\in\Theta\}} g(t)x^*(t) \, dt \tag{2.37}$$

Em consequência, a norma de um sinal $g(t)$ fica dada por

$$||g(t)|| = \left[\int_{\{t:t\in\Theta\}} |g(t)|^2 \, dt \right]^{1/2} \tag{2.38}$$

2.4.4 Energia da Soma de Sinais Ortogonais

Sabemos que o quadrado do comprimento geométrico (ou magnitude) da soma de dois vetores ortogonais é igual à soma dos quadrados das magnitudes dos vetores. Portanto, se os vetores $\mathbf{x}$ e $\mathbf{y}$ forem ortogonais e se $\mathbf{z} = \mathbf{x} + \mathbf{y}$, então,

$$||\mathbf{z}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2$$

Sejam os sinais $g(t)$ e $x(t)$, pelo. Retirando o componente de $g(t)$ contido (ou na forma da) normal $x(t)$

Um resultado semelhante se aplica a sinais. A energia da soma de dois sinais ortogonais é igual à soma das energias dos dois sinais. Assim, se os sinais $x(t)$ e $y(t)$ forem ortogonais em um intervalo $[t_1, t_2]$ e se $z(t) = x(t) + y(t)$, então

$$E_z = E_x + E_y \tag{2.39}$$

A seguir, provaremos esse resultado para sinais de valores complexos, dos quais aqueles com valores reais são um caso especial. Da Eq. (2.34), temos

$$\begin{aligned}\int_{t_1}^{t_2} |x(t) + y(t)|^2\, dt &= \int_{t_1}^{t_2} |x(t)|^2\, dt + \int_{t_1}^{t_2} |y(t)|^2\, dt + \int_{t_1}^{t_2} x(t)y^*(t)\, dt + \int_{t_1}^{t_2} x^*(t)y(t)\, dt \\ &= \int_{t_1}^{t_2} |x(t)|^2\, dt + \int_{t_1}^{t_2} |y(t)|^2\, dt \end{aligned} \tag{2.40}$$

A última igualdade resulta da ortogonalidade dos dois sinais, que anula as duas integrais dos produtos cruzados $x(t)y^*(t)$ e $x^*(t)y(t)$. Tal resultado pode ser estendido à soma de um número qualquer de sinais mutuamente ortogonais.

? Comparação de sinais

2.5 CORRELAÇÃO DE SINAIS

Com as definições do produto interno e da norma de sinais, preparamos o caminho para a comparação de sinais. Mais uma vez, podemos tirar proveito da analogia com os familiares espaços vetoriais. Dois vetores **g** e **x** são similares se **g** tiver uma grande componente ao longo de **x**. Em outras palavras, se c na Eq. (2.19) for grande, os vetores **g** e **x** serão similares. Podemos considerar c como uma medida quantitativa da similaridade entre **g** e **x**. Contudo, essa medida seria defeituosa, pois varia com as normas (ou comprimentos) de **g** e **x**. Para ser adequada, a medida da similaridade entre **g** e **x** deveria independer dos comprimentos de **g** e **x**. Se, por exemplo, dobrarmos o comprimento de **g**, a medida da similaridade entre **g** e **x** não deveria se alterar. A Eq. (2.19) nos mostra que, se dobrarmos **g**, o valor de c também dobra (se dobrarmos **x**, o valor de c será dividido por dois). Fica claro que a medida de similaridade com base na correlação de sinais é falha. A similaridade entre dois vetores é indicada pelo ângulo θ entre eles. Quanto menor o valor de θ, maior a similaridade e vice-versa. A similaridade pode, então, ser adequadamente medida por cos θ. Maior o valor de cos θ, maior a similaridade entre os dois vetores. Assim, uma medida apropriada seria $\rho = \cos\theta$, dada por

$$\rho = \cos\theta = \frac{<\mathbf{g}, \mathbf{x}>}{||\mathbf{g}||\,||\mathbf{x}||} \tag{2.41}$$

Podemos verificar prontamente que esta medida independe dos comprimentos de **g** e **x**. Essa medida de similaridade, ρ, é conhecida como **coeficiente de correlação**. Vale observar que

$$-1 \le \rho \le 1 \tag{2.42}$$

A magnitude de ρ, portanto, jamais é maior que a unidade. Se os dois vetores estiverem alinhados, a similaridade é máxima ($\rho = 1$). Dois vetores alinhados em direções opostas têm máxima dissimilaridade ($\rho = -1$). Se os dois vetores forem ortogonais, a similaridade será zero.

Usamos os mesmos argumentos na definição de um índice de similaridade (coeficiente de correlação) para sinais. Por conveniência, consideraremos sinais em todo o eixo do tempo, de $-\infty$ a ∞. Para definirmos um índice de similaridade que independa das energias (tamanhos) de $g(t)$ e $x(t)$, devemos

normalizar c; para isso, normalizamos os dois sinais, de modo que tenham energias unitárias. Assim, um adequado índice de similaridade ρ, análogo à Eq. (2.41), é dada por

$$\rho = \frac{1}{\sqrt{E_g E_x}} \int_{-\infty}^{\infty} g(t)x(t)\,dt \tag{2.43}$$

Vale ressaltar que a multiplicação de $g(t)$ ou de $x(t)$ por qualquer constante não altera o valor desse índice. Por conseguinte, o mesmo independe dos tamanhos (energias) de $g(t)$ e $x(t)$. Fazendo uso da desigualdade de Cauchy-Schwarz (provada no Apêndice B),* podemos mostrar que a magnitude ρ jamais é maior que a unidade:

$$-1 \le \rho \le 1 \tag{2.44}$$

2.5.1 Funções de Correlação

É conveniente que examinemos a aplicação de correlação à detecção de sinais em um equipamento de radar, onde um pulso é transmitido para detectar um alvo suspeito. A detecção da presença ou ausência do pulso refletido confirma a presença ou ausência do alvo. A medida do atraso temporal entre os pulsos transmitido e recebido (refletido) nos permite determinar a distância ao alvo. Sejam os pulsos transmitido e recebido denotados por $g(t)$ e $z(t)$, respectivamente. Se usássemos a Eq. (2.43) diretamente para medir o coeficiente de correlação ρ, obteríamos

$$\rho = \frac{1}{\sqrt{E_g E_z}} \int_{-\infty}^{\infty} z(t)g^*(t)\,dt = 0 \tag{2.45}$$

A correlação é zero porque os pulsos são separados (não se sobrepõem no tempo). A integral na Eq. (2.45) será igual a zero mesmo que os pulsos sejam idênticos, mas guardem um deslocamento temporal. Para evitar essa dificuldade, comparamos o pulso recebido $z(t)$ com o pulso transmitido $g(t)$ deslocado de τ. Se, para algum valor de τ, houver uma forte correlação, não apenas detectaremos a presença do pulso, mas também determinaremos o deslocamento temporal de $z(t)$ em relação a $g(t)$. Por isso, em vez da integral no lado direito, usamos a integral modificada $\psi_{gz}(\tau)$, denominada função de **correlação cruzada** dos dois sinais de valores complexos $g(t)$ e $z(t)$ e definida como

$$\psi_{gz}(\tau) \equiv \int_{-\infty}^{\infty} z(t)g^*(t-\tau)\,dt = \int_{-\infty}^{\infty} z(t+\tau)g^*(t)\,dt \tag{2.46}$$

$\psi_{gz}(\tau)$ é, portanto, uma indicação da similaridade (correlação) entre $g(t)$ e $z(t)$ adiantado (deslocado para a esquerda) de τ segundos.

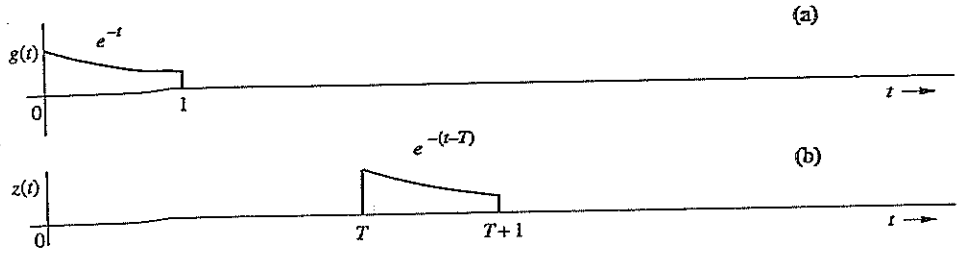


Figura 2.11 Explicação gráfica da função de autocorrelação.

*Segundo a desigualdade de Cauchy-Schwarz, para dois sinais de energias reais $g(t)$ e $x(t)$, $\left(\int_{-\infty}^{\infty} g(t)x(t)\,dt\right)^2 \le E_g E_x$, sendo que a igualdade ocorre se e somente se $x(t) = Kg(t)$, em que K é uma constante arbitrária. Existe uma desigualdade equivalente para sinais de valores complexos.

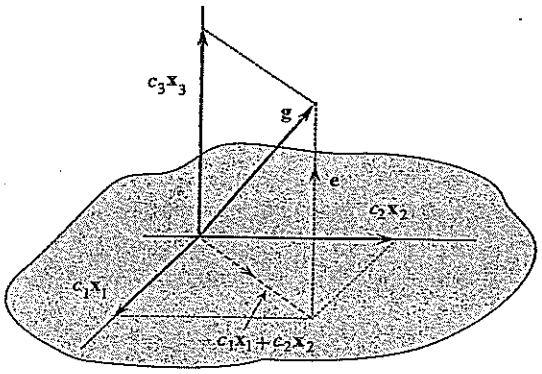


Figura 2.12 Representação de um vetor no espaço tridimensional.

2.5.2 Função de Autocorrelação

A correlação de um sinal como ele próprio, como ilustrado na Fig. 2.11, é denominada **autocorrelação**. A função de autocorrelação $\psi_g(\tau)$ de um sinal de valores reais $g(t)$ é definida como

$$\psi_g(\tau) \equiv \int_{-\infty}^{\infty} g(t)g(t+\tau)\,dt \tag{2.47}$$

A função de autocorrelação mede a similaridade do sinal $g(t)$ com sua versão deslocada no tempo. No Capítulo 3, mostraremos que a função de autocorrelação fornece valiosas informações espectrais a respeito do sinal.

2.6 CONJUNTO ORTOGONAL DE SINAIS

Nesta seção, mostraremos uma forma de representar um sinal como a soma de um conjunto ortogonal de sinais. Na verdade, os sinais nesse conjunto ortogonal formam uma base para o espaço de sinais em questão. Aqui, mais uma vez, podemos nos beneficiar do entendimento de um problema similar envolvendo vetores. Sabemos que um vetor pode ser representado como a soma de vetores ortogonais, que formam o sistema de coordenadas de um espaço vetorial. Com sinais, o problema é análogo e os resultados são similares aos do problema com vetores. Por isso, façamos uma revisão da representação de vetores.

2.6.1 Espaço Vetorial Ortogonal

Consideremos um espaço vetorial cartesiano multidimensional, representado na Fig. 2.12 por três vetores ortogonais $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$, para o caso especial de um espaço tridimensional. Primeiro, procuramos aproximar um vetor tridimensional $\mathbf{g}$ em termos de dois vetores ortogonais $\mathbf{x}_1$ e $\mathbf{x}_2$:

$$\mathbf{g} \simeq c_1\mathbf{x}_1 + c_2\mathbf{x}_2$$

O erro $\mathbf{e}$ nesta aproximação é

$$\mathbf{e} = \mathbf{g} - (c_1\mathbf{x}_1 + c_2\mathbf{x}_2)$$

ou, o que é equivalente,

$$\mathbf{g} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \mathbf{e}$$

De acordo com a argumentação geométrica anterior, a Fig. 2.12 deixa claro que o comprimento do vetor de erro $\mathbf{e}$ é mínimo quando o mesmo for perpendicular ao plano $(\mathbf{x}_1, \mathbf{x}_2)$ e quando $c_1\mathbf{x}_1$ e $c_2\mathbf{x}_2$ forem as projeções (componentes) de $\mathbf{g}$ sobre $\mathbf{x}_1$ e $\mathbf{x}_2$, respectivamente. Portanto, as constantes c_1 e c_2 são dadas pela fórmula na Eq. (2.19).

Agora, determinemos a melhor aproximação de g em termos dos três vetores mutuamente ortogonais $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$:

$$\mathbf{g} \simeq c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 \tag{2.48}$$

A Fig. 2.12 mostra que existe uma única escolha para os valores de c_1, c_2 e c_3 para a qual a Eq. (2.48) deixa de ser uma aproximação e se torna uma igualdade:

$$\mathbf{g} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3$$

Neste caso, $c_1\mathbf{x}_1$, $c_2\mathbf{x}_2$ e $c_3\mathbf{x}_3$ são as projeções (componentes) de g sobre $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$, respectivamente. Vale notar que, agora, o erro e da aproximação é zero quando g é aproximado em termos de três vetores mutuamente ortogonais $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$. Isso ocorre porque g é um vetor tridimensional e os vetores $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$ representam um *conjunto completo* de vetores ortogonais no espaço tridimensional. Completeza, aqui, significa que é impossível, nesse espaço, encontrar qualquer outro vetor $\mathbf{x}_4$ que seja ortogonal a todos os três vetores $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$. Nesse espaço, qualquer vetor pode, portanto, ser representado (com erro zero) em termos desses três vetores. Vetores como esses são conhecidos como vetores de base e o conjunto dos vetores, como uma **base ortogonal completa** desse espaço vetorial. Se um conjunto de vetores $\{\mathbf{x}_i\}$ não for completo, o erro da aproximação, na maioria dos casos, não será zero. Por exemplo, no espaço tridimensional que acabamos de considerar, em geral, não é possível representar um vetor g em termos de apenas dois vetores da base sem erro.

A escolha dos vetores de base não é única. Na verdade, cada conjunto de vetores de base corresponde a uma escolha particular de sistema de coordenadas. Assim, um vetor tridimensional g pode ser representado de diferentes formas, dependendo do sistema de coordenadas utilizado.

Em resumo, se um conjunto de vetores mutuamente ortogonais $\{\mathbf{x}_i\}$, ou seja, vetores tais que

$$< \mathbf{x}_m, \mathbf{x}_n > = \begin{cases} 0 & m \neq n \\ |\mathbf{x}_m|^2 & m = n \end{cases}$$

e se esse conjunto for completo, um vetor g nesse espaço pode ser expresso como

$$\mathbf{g} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 \tag{2.49}$$

onde as constantes c_i são dadas por

$$c_i = \frac{< \mathbf{g}, \mathbf{x}_i >}{< \mathbf{x}_i, \mathbf{x}_i >} \tag{2.50a}$$

$$= \frac{1}{||\mathbf{x}_i||^2} < \mathbf{g}, \mathbf{x}_i > \qquad i = 1, 2, 3 \tag{2.50b}$$

2.6.2 Espaço Ortogonal de Sinais

Dando prosseguimento ao problema de aproximação de sinais, usaremos conceitos e resultados da aproximação de vetores. Como antes, definimos ortogonalidade de um conjunto de sinais $x_1(t)$, $x_2(t)$,..., $x_N(t)$ em um domínio temporal Θ (pode ser um intervalo $[t_1, t_2]$) como

$$\int_{t\in\Theta} x_m(t)x_n^*(t)\,dt = \begin{cases} 0 & m \neq n \\ E_n & m = n \end{cases} \tag{2.51}$$

Se todos os sinais tiverem energia unitária $E_n = 1$, dizemos que o conjunto é *normalizado* e o denominamos **conjunto ortonormal**. Um conjunto ortogonal sempre pode ser normalizado, bastando dividir

$x_n(t)$ por $\sqrt{E_n}$, para todo n. Consideremos, agora, o problema de aproximar um sinal $g(t)$ em Θ por N sinais mutuamente ortogonais $x_1(t), x_2(t), \ldots, x_N(t)$:

$$g(t) \simeq c_1x_1(t) + c_2x_2(t) + \cdots + c_Nx_N(t) \tag{2.52a}$$

$$= \sum_{n=1}^{N} c_n x_n(t) \qquad t \in \Theta \tag{2.52b}$$

Pode ser mostrado que E_e, a energia do sinal de erro $e(t)$ nessa aproximação, é minimizada quando escolhemos

$$c_n = \frac{\int_{t\in\Theta} g(t)x_n^*(t)\,dt}{\int_{t\in\Theta} |x_n(t)|^2\,dt}$$

$$= \frac{1}{E_n}\int_{\Theta} g(t)x_n^*(t)\,dt \qquad n = 1, 2, \ldots, N \tag{2.53}$$

Além disso, se o conjunto ortogonal for **completo**, a energia do erro $E_e \to 0$ e a aproximação em (2.52) deixa de ser uma aproximação, tornando-se uma igualdade. Mais precisamente, definamos o erro da aproximação de N termos por

$$e_N(t) = g(t) - c_1x_1(t) + c_2x_2(t) + \cdots + c_Nx_N(t) = g(t) - \sum_{n=1}^{N} c_n x_n(t) \qquad t \in \Theta \tag{2.54}$$

Se a base ortogonal for **completa**, a energia do sinal de erro converge a zero, ou seja

$$\lim_{N\to\infty} \int_{t\in\Theta} |e_N(t)|^2 dt = 0 \tag{2.55}$$

Em termos estritamente matemáticos, no entanto, um sinal pode não convergir a zero mesmo que sua energia convirja a zero. Isso ocorre porque um sinal pode ser não nulo em alguns pontos isolados.* Entretanto, na prática, sinais são contínuos para todo t e a igualdade (2.55) afirma que o sinal de erro tem energia zero quando $N \to \infty$. Assim, para $N \to \infty$, a igualdade (2.55) pode ser escrita como

$$g(t) = c_1x_1(t) + c_2x_2(t) + \cdots + c_nx_n(t) + \cdots$$

$$= \sum_{n=1}^{\infty} c_n x_n(t) \qquad t \in \Theta \tag{2.56}$$

em que os coeficientes c_n são dados pela Eq. (2.53). Como a energia do sinal de erro se aproxima de zero, a energia de $g(t)$ fica igual à soma das energias de suas componentes ortogonais.

A série no lado direito da Eq. (2.53) é denominada **série de Fourier generalizada** de $g(t)$ em termos do conjunto $\{x_n(t)\}$. Quando o conjunto $\{x_n(t)\}$ é tal que a energia do erro $E_N \to 0$ à medida que $N \to \infty$, para qualquer membro de alguma classe particular de sinais, dizemos que o conjunto $\{x_n(t)\}$ é completo em $\{t: \Theta\}$ para aquela classe de $g(t)$, e o conjunto $\{x_n(t)\}$ é denominado um conjunto de **funções de base** ou **sinais de base**. Em particular, a classe de sinais de energia (finita) em Θ é denotado por $L^2\{\Theta\}$. De aqui em diante, consideraremos apenas a classe de sinais de energia, a menos que especifiquemos de modo diferente.

2.6.3 Teorema de Parseval

Recordemos que a energia da soma de sinais ortogonais é igual à soma das energias dos mesmos. Portanto, a energia do lado direito da Eq. (2.56) é a soma das energias das componentes ortogonais.

*Conhecidos como conjunto de medida zero.

A energia de uma componente $c_n x_n(t)$ é $c_n^2 E_n$. Igualando as energias nos dois lados da Eq. (2.56), obtemos

$$\begin{aligned} E_g &= c_1^2 E_1 + c_2^2 E_2 + c_3^2 E_3 + \cdots \\ &= \sum_n c_n^2 E_n \end{aligned} \tag{2.57}$$

Esse importante resultado é conhecido como **teorema de Parseval**. Recordemos que a energia de um sinal (área sob a curva do quadrado do valor do sinal) é análoga ao quadrado do comprimento de um vetor, segundo a analogia vetor-sinal. Em um espaço vetorial, sabemos que o quadrado do comprimento de um vetor é igual à soma dos quadrados dos comprimentos de suas componentes ortogonais. O teorema de Parseval [Eq. (2.57)] é a afirmação desse fato para o caso de sinais.

2.7 SÉRIE DE FOURIER EXPONENCIAL

Observamos anteriormente que a representação ortogonal de sinais NÃO é única. Embora a tradicional série de Fourier trigonométrica permita uma boa representação de todos os sinais periódicos, nesta seção apresentamos uma representação ortogonal de sinais periódicos **equivalente**, mas que tem forma mais simples.

Primeiro, observemos que o conjunto de exponenciais $e^{jn\omega_0 t}$ ($n = 0, \pm 1, \pm 2, \ldots$) é ortogonal em um intervalo de duração $T_0 = 2\pi/\omega_0$, ou seja,

$$\int_{T_0} e^{jm\omega_0 t}(e^{jn\omega_0 t})^* \, dt = \int_{T_0} e^{j(m-n)\omega_0 t} \, dt = \begin{cases} 0 & m \neq n \\ T_0 & m = n \end{cases} \tag{2.58}$$

Além disso, esse conjunto é completo.[1,2] Das Eqs. (2.53) e (2.56), um sinal $g(t)$ pode ser expresso em um intervalo de duração T_0 segundos por uma série de Fourier exponencial

$$\begin{aligned} g(t) &\stackrel{\text{ñ}}{=} \sum_{n=-\infty}^{\infty} D_n e^{jn\omega_0 t} \\ &= \sum_{n=-\infty}^{\infty} D_n e^{jn2\pi f_0 t} \end{aligned} \tag{2.59}$$

em que [ver Eq. (2.53)]

$$D_n = \frac{1}{T_0} \int_{T_0} g(t) e^{-jn2\pi f_0 t} \, dt \tag{2.60}$$

A série de Fourier exponencial na Eq. (2.59) consiste em componentes da forma $e^{jn2\pi f_0 t}$, com n variando de $-\infty$ a ∞, e é periódica, com período T_0.

Exemplo 2.3 Determinemos a série de Fourier exponencial para o sinal na Fig. 2.13b.

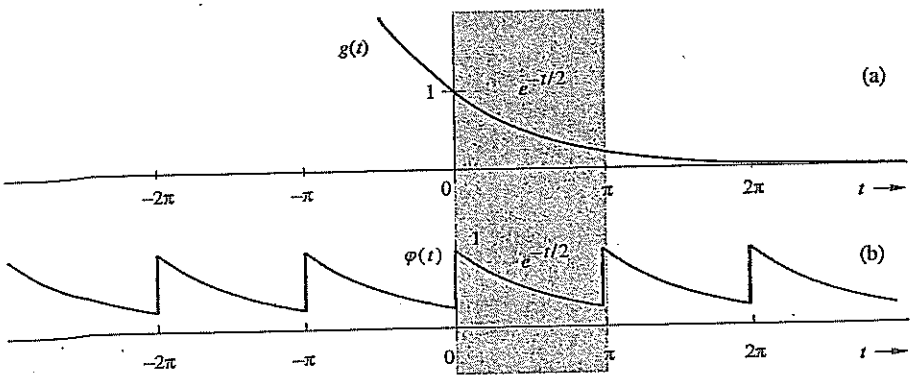


Figura 2.13 Um sinal periódico.

Neste caso, $T_0 = \pi$, $2\pi f_0 = 2\pi/T_0 = 2$ e

$$\varphi(t) = \sum_{n=-\infty}^{\infty} D_n e^{j2nt}$$

em que

$$\begin{aligned} D_n &= \frac{1}{T_0}\int_{T_0} \varphi(t) e^{-j2nt}\, dt \\ &= \frac{1}{\pi}\int_0^{\pi} e^{-t/2} e^{-j2nt}\, dt \\ &= \frac{1}{\pi}\int_0^{\pi} e^{-(\frac{1}{2}+j2n)t}\, dt \\ &= \frac{-1}{\pi\left(\frac{1}{2}+j2n\right)} e^{-(\frac{1}{2}+j2n)t}\Big|_0^{\pi} \\ &= \frac{0{,}504}{1+j4n} \end{aligned} \tag{2.61}$$

e

$$\varphi(t) = 0{,}504 \sum_{n=-\infty}^{\infty} \frac{1}{1+j4n} e^{j2nt} \tag{2.62a}$$

$$\begin{aligned} &= 0{,}504 \left[1 + \frac{1}{1+j4} e^{j2t} + \frac{1}{1+j8} e^{j4t} + \frac{1}{1+j12} e^{j6t} + \cdots \right. \\ &\quad \left. + \frac{1}{1-j4} e^{-j2t} + \frac{1}{1-j8} e^{-j4t} + \frac{1}{1-j12} e^{-j6t} + \cdots \right] \\ &\quad \vdots \end{aligned} \tag{2.62b}$$

Vale observar que os coeficientes D_n são complexos. Além disso, D_n e D_{-n} são complexos conjugados, como esperado.

Espectro de Fourier Exponencial

No espectro exponencial, representamos graficamente os valores dos coeficientes D_n em função de ω. Contudo, como em geral D_n é complexo, necessitamos de dois gráficos: um para a parte real e outro para a parte imaginária de D_n, ou um para a amplitude (magnitude) e outro para a fase de D_n. Preferimos a última opção, devido à conexão com amplitudes e fases das componentes correspondentes da série de Fourier trigonométrica. Portanto, desenhamos gráficos de $|D_n|$ em função de ω e de $\angle D_n$ em função de ω. Isso requer que os coeficientes D_n sejam expressos na forma polar $|D_n| e^{j\angle D_n}$.

Para um sinal periódico, os coeficientes D_n e D_{-n} são complexos conjugados,

$$|D_n| = |D_{-n}| \tag{2.63a}$$

$$\angle D_n = \theta_n \qquad \text{e} \qquad \angle D_{-n} = -\theta_n \tag{2.63b}$$

Logo,

$$D_n = |D_n| e^{j\theta_n} \qquad \text{e} \qquad D_{-n} = |D_n| e^{-j\theta_n} \tag{2.64}$$

Notemos que $|D_n|$ representa a amplitude (magnitude) e $\angle D_n$, a fase de cada componente. Pela Eq. (2.63), quando o sinal $g(t)$ tiver valores reais, o espectro de amplitude ($|D_n|$ *versus* f) será uma função par da frequência e o espectro de fase (ou angular, $\angle D_n$ *versus* f), uma função *ímpar* de f.

Para a série no Exemplo 2.3,

$$D_0 = 0{,}504$$

$$D_1 = \frac{0{,}504}{1+j4} = 0{,}122e^{-j75{,}96^\circ} \Longrightarrow |D_1| = 0{,}122,\ \angle D_1 = -75{,}96^\circ$$

$$D_{-1} = \frac{0{,}504}{1-j4} = 0{,}122e^{j75{,}96^\circ} \Longrightarrow |D_{-1}| = 0{,}122,\ \angle D_{-1} = 75{,}96^\circ$$

e

$$D_2 = \frac{0{,}504}{1+j8} = 0{,}0625e^{-j82{,}87^\circ} \Longrightarrow |D_2| = 0{,}0625,\ \angle D_2 = -82{,}87^\circ$$

$$D_{-2} = \frac{0{,}504}{1-j8} = 0{,}0625e^{j82{,}87^\circ} \Longrightarrow |D_{-2}| = 0{,}0625,\ \angle D_{-2} = 82{,}87^\circ$$

e assim por diante. Vale notar que D_n e D_{-n} são complexos conjugados, como esperado [ver Eq. (2.63b)].

A Fig. 2.14 mostra os espectros de frequência (amplitude e fase) da série de Fourier exponencial para o sinal periódico $\varphi(t)$ na Fig. 2.13b.

Podemos observar algumas características interessantes nesses espectros. Primeira, o espectro existe para valores positivos e negativos de f (frequência). Segunda, o espectro de amplitude é uma função par de f e o espectro de fase, uma função *ímpar* de f. As Eqs. (2.63) mostram as características de simetria da amplitude e da fase de D_n.

O que Significa Frequência Negativa?

A existência do espectro em frequências negativas é, de alguma forma, algo intrigante para algumas pessoas, pois a frequência (número de repetições por segundo) é, por definição, uma grandeza positiva. Como devemos interpretar uma frequência negativa f_0? Para expressar uma senoide de uma frequência negativa $-f_0$, podemos subtrair $\omega_0 = 2\pi f_0$ e usar uma identidade trigonométrica, como

$$\cos(-\omega_0 t + \theta) = \cos(\omega_0 t - \theta)$$

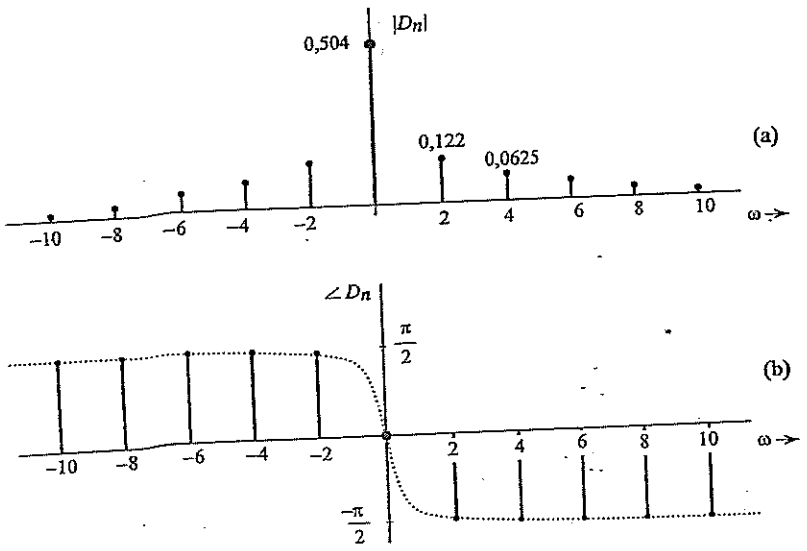


Figura 2.14 Espectro de Fourier exponencial para o sinal na Fig. 2.13a.

Isso mostra claramente que a frequência angular de uma senoide $\cos(-\omega_0 t + \theta)$ é $|\omega_0|$, que é uma grandeza positiva. A afirmação, baseada no bom senso, de que uma frequência deve ser positiva advém da tradicional noção de que frequência está associada a uma senoide de valores reais (como seno ou cosseno). Na verdade, o conceito de frequência associado a senoides de valores reais descreve somente a taxa de variação senoidal, sem ater-se ao sentido da variação. Isso ocorre porque sinais senoidais de valores reais NÃO contêm informação a respeito do sentido de sua variação.

O conceito de frequência negativa faz sentido **apenas** quando consideramos senoides de valores complexos, para as quais a taxa e o *sentido* da variação têm significado. Devemos observar que

$$e^{\pm j\omega_0 t} = \cos \omega_0 t \pm j \operatorname{sen} \omega_0 t$$

Essa relação deixa claro que uma frequência ω positiva ou negativa leva a variações periódicas com a mesma taxa. Contudo, os resultantes sinais de valores complexos NÃO são iguais. Dado que $|e^{\pm j\omega_0 t}| = 1$, tanto $e^{+j\omega_0 t}$ como $e^{-j\omega_0 t}$ são variáveis complexas de comprimento unitário, que podem ser representadas no plano complexo. Na Fig. 2.15, mostramos as duas senoides exponenciais como grandezas complexas de comprimento unitário que variam no tempo t. A taxa de rotação para as duas exponenciais $e^{\pm j\omega_0 t}$ é $|\omega_0|$. Fica claro que, para frequências positivas, a senoide exponencial gira no sentido anti-horário ou trigonométrico; para frequências negativas, a senoide exponencial gira no sentido horário. Isso ilustra o real significado de frequências negativas.

Existe uma boa analogia entre frequência positiva/negativa e velocidade positiva/negativa. Assim como pessoas relutam em usar velocidade *negativa* para descrever o movimento de um objeto, também relutam em aceitar o conceito de frequência "negativa". Contudo, uma vez que entendamos que velocidade negativa simplesmente se refere tanto ao sentido negativo do deslocamento como à real velocidade de um objeto em movimento, velocidade negativa passa a fazer sentido. De modo similar, frequência negativa NÃO descreve a taxa de variação periódica de um seno ou cosseno, mas, sim, o sentido e a taxa de rotação de uma exponencial senoidal de comprimento unitário.

Outra maneira de ver essa situação é dizer que os *espectros exponenciais são uma representação gráfica dos coeficientes D_n como uma função de f. A existência do espectro em $f = -nf_0$ simplesmente indica que uma componente exponencial $e^{-jn2\pi f_0 t}$ existe na série.* Sabemos, pela identidade de Euler

$$\cos(\omega t + \theta) = \frac{e^{j\theta}}{2} \exp(j\omega t) + \frac{e^{-j\theta}}{2} \exp(-j\omega t)$$

que uma senoide de frequência $n\omega_0$ pode ser expressa em termos de um par de exponenciais $e^{jn\omega_0 t}$ e $e^{-jn\omega_0 t}$. O fato de seno e cosseno consistirem em componentes senoidais exponenciais de frequências positiva e negativa indica que NÃO é possível descrever o *sentido* de suas variações periódicas. Na verdade, as funções seno e cosseno da frequência ω_0 consistem em duas senoides exponenciais de mesma amplitude e frequências $\pm\,\omega_0$. Em consequência, a frequência de um seno ou cosseno é o valor absoluto das duas componentes de frequência e denota apenas a taxa das variações senoidais.

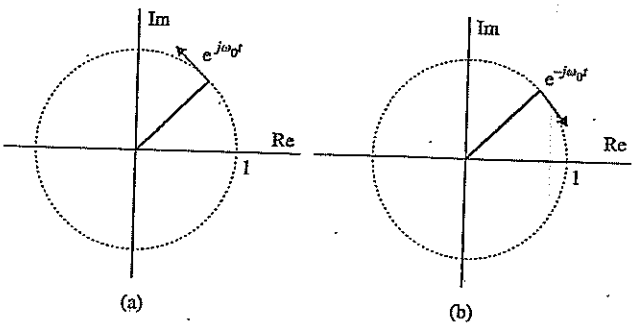


Figura 2.15 Variável complexa de comprimento unitário e frequência positiva (rotação no sentido anti-horário) comparada com variável complexa de comprimento unitário e frequência negativa (rotação no sentido horário).

Exemplo 2.4 Determinemos a série de Fourier exponencial da onda quadrada periódica mostrada na Fig. 2.16.

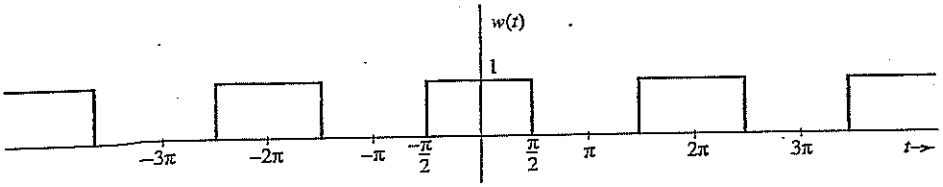


Figura 2.16
Sinal de pulso quadrado periódico.

$$w(t) = \sum_{n=-\infty}^{\infty} D_n e^{jn2\pi f_0 t}$$

onde

$$D_0 = \frac{1}{T_0}\int_{T_0} w(t)\,dt = \frac{1}{2}$$

$$\begin{aligned} D_n &= \frac{1}{T_0}\int_{T_0} w(t)e^{-jn2\pi f_0 t}\,dt, \qquad n \neq 0 \\ &= \frac{1}{T_0}\int_{-T_0/4}^{T_0/4} e^{-jn2\pi f_0 t}\,dt \\ &= \frac{1}{-jn2\pi f_0 T_0}\left[e^{-jn2\pi f_0 T_0/4} - e^{jn2\pi f_0 T_0/4}\right] \\ &= \frac{2}{n2\pi f_0 T_0}\,\text{sen}\left(\frac{n2\pi f_0 T_0}{4}\right) = \frac{1}{n\pi}\,\text{sen}\left(\frac{n\pi}{2}\right) \end{aligned}$$

Neste caso, D_n é real. Consequentemente, não precisamos do gráfico de fase ou ângulo e, em vez de representarmos o espectro de amplitude ($|D_n|$ *versus f*), podemos traçar apenas o gráfico de D_n *versus f*, como na Fig. 2.17.

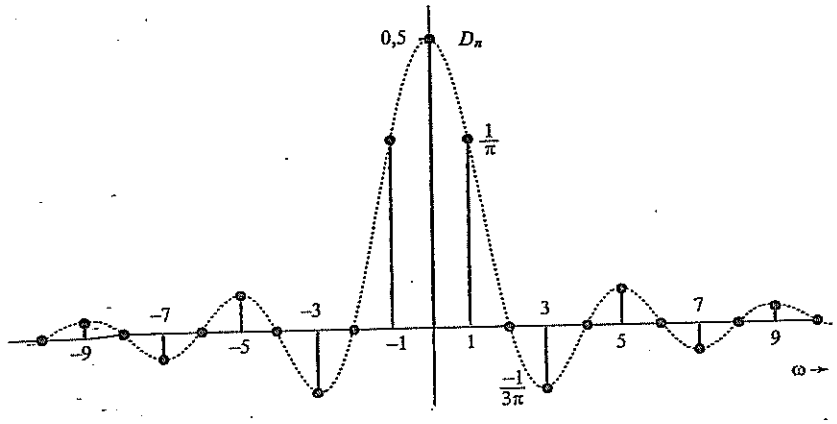


Figura 2.17
Espectro de Fourier exponencial do sinal de pulso quadrado periódico.

Exemplo 2.5 Determinemos a série de Fourier exponencial para o trem de pulsos $\delta_{T_0}(t)$ mostrado na Fig. 2.18a, e tracemos o gráfico dos correspondentes espectros.

A série de Fourier exponencial é dada por

$$\delta_{T_0}(t) = \sum_{n=-\infty}^{\infty} D_n e^{jn2\pi f_0 t} \qquad f_0 = \frac{1}{T_0} \tag{2.65}$$

com

$$D_n = \frac{1}{T_0}\int_{T_0} \delta_{T_0}(t) e^{-jn2\pi f_0 t}\, dt$$

Escolhendo o intervalo de integração $\left(\frac{-T_0}{2}, \frac{T_0}{2}\right)$ e observando que, neste intervalo, $\delta_{T_0}(t) = \delta(t)$, obtemos

$$D_n = \frac{1}{T_0}\int_{-T_0/2}^{T_0/2} \delta(t) e^{-jn2\pi f_0 t}\, dt$$

Nessa integral, o impulso está localizado em $t = 0$. Da propriedade de amostragem da função impulso, a integral no lado direito é o valor de $e^{-jn2\pi f_0 t}$ em $t = 0$ (localização do impulso). Logo,

$$D_n = \frac{1}{T_0} \tag{2.66}$$

e

$$\delta_{T_0}(t) = \frac{1}{T_0}\sum_{n=-\infty}^{\infty} e^{jn2\pi f_0 t} \qquad f_0 = \frac{1}{T_0} \tag{2.67}$$

As Eqs. (2.67) mostram que o espectro exponencial é uniforme ($D_n = 1/T_0$) para todas as frequências, como ilustrado na Fig. 2.18b. Como o espectro é real, basta que tracemos o gráfico da amplitude. Todas as fases são iguais a zero.

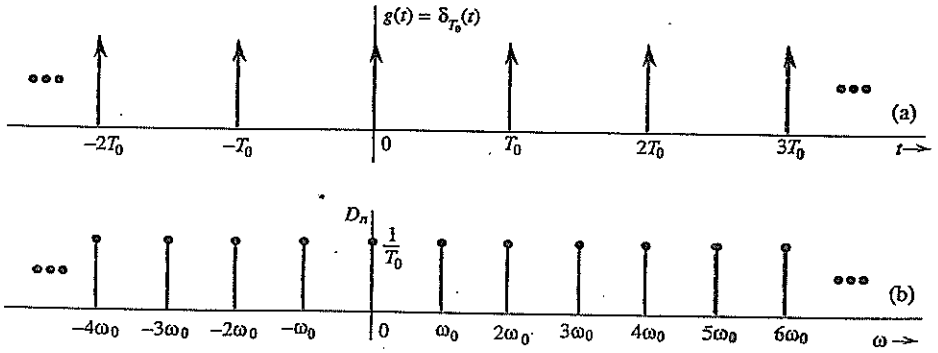


Figura 2.18 Trem de impulsos e correspondente espectro de Fourier exponencial.

Teorema de Parseval na Série de Fourier

Um sinal periódico $g(t)$ é um sinal de potência, e cada termo em sua série de Fourier também é um sinal de potência. A potência P_g de $g(t)$ é igual à potência de sua série de Fourier. Como a série de Fourier consiste em termos que são mutuamente ortogonais em um período, a potência da série de Fourier é igual à soma das potências das componentes de Fourier. Esse resultado advém do teorema de Parseval.

Portanto, para a série de Fourier exponencial

$$g(t) = D_0 + \sum_{n=-\infty,\, n\neq 0}^{\infty} D_n e^{jn\omega_0 t}$$

a potência é dada por (veja o Exercício 2.1-7)

$$P_g = \sum_{n=-\infty}^{\infty} |D_n|^2 \tag{2.68a}$$

Para um sinal $g(t)$ de valores reais, $|D_{-n}| = |D_n|$. Logo,

$$P_g = D_0{}^2 + 2\sum_{n=1}^{\infty} |D_n|^2 \tag{2.68b}$$

Comentário: O teorema de Parseval admite diferentes formas, como as das Eq. (2.57) e Eq. (2.68a). No próximo capítulo, apresentaremos mais uma forma para o caso de sinais periódicos. Embora essas várias representações do teorema pareçam distintas, todas afirmam o mesmo princípio: o quadrado do comprimento de um vetor é igual à soma dos quadrados dos comprimentos de suas componentes ortogonais. A primeira forma [Eq. (2.57)] se aplica a sinais de energia; a segunda [Eq. (2.68a)], a sinais periódicos representados pela série de Fourier exponencial.

Alguns Exemplos Adicionais de Conjuntos Ortogonais de Sinais

A representação de um sinal pela série de Fourier mostra que sinais são, em todos os sentidos, vetores. Assim como há diversas maneiras de representar um vetor como a soma de componentes, dependendo da escolha do sistema de coordenadas, também existem numerosas formas de representar um sinal como a soma de componentes. Assim como um sistema de coordenadas vetoriais é formado por um conjunto de vetores mutuamente ortogonais (retangulares, cilíndricos, esféricos etc.), um sistema de coordenadas de sinais também é formado por um conjunto de sinais mutuamente ortogonais, os sinais de base. Existe um grande número de conjuntos ortogonais de sinais que podem ser usados como sinais de base para séries de Fourier generalizadas. Alguns conjuntos de sinais bem conhecidos são os das funções trigonométricas (senoidais), funções exponenciais, funções de Walsh, funções de Bessel, polinômios de Legendre, funções de Laguerre, polinômios de Jacobi, polinômios de Hermite e polinômios de Chebyshev. Neste livro, as funções de maior interesse são as exponenciais, discutidas no próximo capítulo.

2.8 EXERCÍCIOS COM O MATLAB

Nesta seção, apresentamos alguns exercícios básicos com o MATLAB para ilustrar o processo de geração de sinais, operações com sinais e análise de sinais por série de Fourier.

Sinais Básicos e Representação Gráfica de Sinais

Funções básicas podem ser definidas por meio de arquivos m de MATLAB. A seguir, apresentamos três programas MATLAB para implementar três funções básicas, a partir de um vetor `t` variante no tempo:

. `ustep.m` implementa a função degrau unitário $u(t)$
. `rect.m` implementa a função retangular padrão $rect(t)$
. `triangl.m` implementa a função triangular padrão $\Delta(t)$

```
% (nome do arquivo:   ustep.m)
%  A função degrau unitário é uma função do tempo 't'.
%  Uso  y = ustep(t)
%
%  ustep(t) = 0     se t < 0
%  ustep(t) = 1,    se t >= 1
%
%  t - deve ser uma variável real, podendo ser um vetor ou uma matriz
%
function y=ustep(t)
   y = (t>=0) ;
end
```

```
% (nome do arquivo: rect.m)
% A função rectangular é uma função do tempo 't'.
%
% Uso   y = rect(t)
% t - deve ser uma variável real, podendo ser um vetor ou uma matriz
%
% rect(t) = 1,       se |t| < 0.5
% rect(t) = 0,       se |t| > 0.5
%
function y=rect(t)
  y =(sign(t+0.5)-sign(t-0.5) >0);
end
```

```
% (nome do arquivo: triangl.m)
% A função triangular é uma função do tempo 't'.
%
% triangl(t) = 1-|t|, se  |t| < 1
% triangl(t) = 0, se      |t| > 1
%
% Uso   y = triangl(t)
% t - deve ser uma variável real, podendo ser um vetor ou uma matriz
%
function y=triangl(t)
  y = (1-abs(t)).*(t>=-1).*(t<1);
end
```

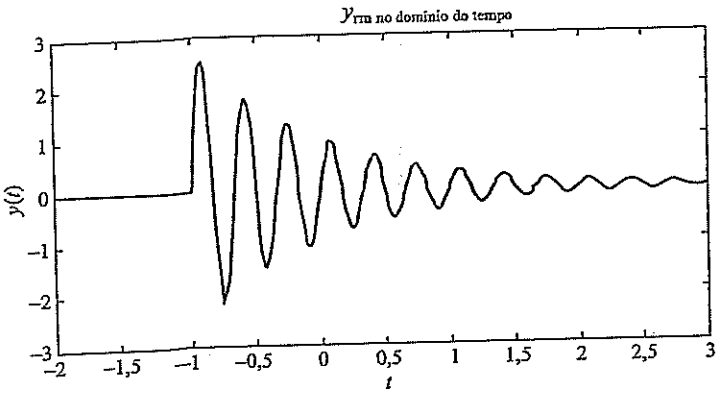


Figura 2.19 Gráfico de um sinal.

A seguir, mostramos um exemplo de como usar o MATLAB para gerar o gráfico de um sinal. Para isso, fornecemos o arquivo `siggraf.m`. Neste exemplo, construímos o sinal

$$y(t) = \exp(-t)\operatorname{sen}(6\pi t)u(t+1)$$

e traçamos o gráfico de sua variação com o tempo. A Fig. 2.19 mostra a curva resultante.

```
% (nome do arquivo: siggraf.m)
% Para traçar o gráfico de um sinal, o primeiro passo é especificar
% os eixos x e y
% Podemos, primeiro, especificar o comprimento do eixo x
  t=[-2:0.01:3];     % "t" varia de -2 a 3, em incrementos de 0,01
% A seguir, calculamos valores do sinal no intervalo de "t"
  especificado para o gráfico
  y=exp(-t).*sin(10*pi*t).*ustep(t+1);
  figure(1); fig1=plot(t,y);    % gráfico de t versus y na figura 1
  set(fig1,'Linewidth',2);      % escolhemos uma linha mais larga
  xlabel('\it t');      % usamos 't' itálico para o título do eixo x
  ylabel('\{\bf y\}(\{\it t\})'); % usamos 'y' em negrito para o título
                                    do eixo y
  title('\{\bf y\}\_\{\rm no domínio do tempo\}'); % podemos usar
                                                     subscritos
```

Sinais Periódicos e Sinais de Potência

Para gerar sinais periódicos, primeiro calculamos valores do sinal em um período e, depois, repetimos esses valores múltiplas vezes.

Com o programa de MATLAB `PfuncEx.m`, fornecido a seguir, geramos um sinal periódico e observamos seu comportamento ao longo de $2M$ períodos. Neste exemplo, o período é $T = 6$. O programa também calcula a potência média do sinal, armazenada na variável `y_power`, e a energia do sinal em um período, armazenada na variável `y_energyT`.

```
% (nome do arquivo: PfuncEx.m)
% Este exemplo gera um sinal periódico, traça o gráfico de sua variação
  com o tempo
```

```
% e calcula a potência média do sinal, na variável y_power, e a energia
% do sinal em um período T, na variável: y_energyT
   echo off;clear;clf;
% Para gerar um sinal periódico g_T(t),
% podemos, primeiro, calcular valores do sinal em um período 'T'
   Dt=0.002; % Intervalo de tempo (para amostra do sinal)
   T=6;      % período=T
   M=3;      % Para gerar 2M períodos do sinal
   t=[0:Dt:T-Dt]; %"t" varia em um período, [0, T], em incrementos de Dt
% A seguir, o sinal é calculado em um período "T"
   y=exp(-abs(t)/2).*sin(2*pi*t).*(ustep(t)-ustep(t-4));
% Agora, múltiplos períodos podem ser gerados.
   time=[];
   y_periodic=[];
for i=-M:M-1,
   time=[time i*T+t];
   y_periodic=[y_periodic y];
end
   figure(1); fy=plot(time,y_periodic);
   set(fy,'Linewidth',2);xlabel('{\it t}');
   echo on
%  Cálculo da potência média
   y_power=sum(y_periodic*y_periodic')*Dt/(max(time)-min(time))
%  Cálculo da energia em um período T
   y_energyT=sum(y.*conj(y))*Dt
```

O programa gera um sinal periódico, como mostrado na Fig. 2.20, e produz as respostas numéricas:

```
y_power =
       0.0813

y_energyT =
       0.4878
```

Correlação de Sinais

Um programa de MATLAB pode implementar diretamente o conceito de correlação de sinais introduzido na Seção 2.5. No próximo exemplo, fornecemos um programa, `sign_cor.m`, que calcula os coeficientes de correlação entre o sinal $x(t)$ e sinais $g_1(t), g_2(t),..., g_5(t)$. O programa, primeiro, gera a Fig. 2.21, que ilustra os seis sinais no domínio do tempo.

```
% (nome do arquivo: sign_cor.m)
clear
% Para gerar os seis sinais x(t), g_1(t), ... g_5(t);
% deste exemplo
% podemos primeiro, gerar o sinal g(t) em um periodo 'T'
```

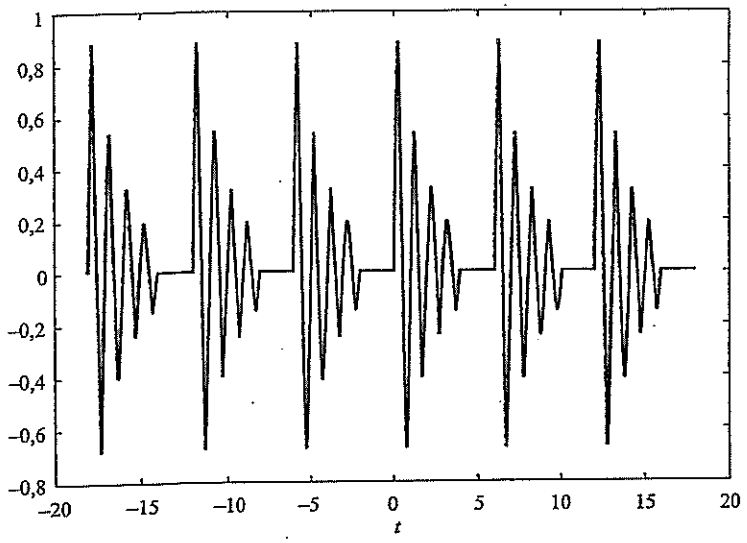


Figura 2.20 Geração de um sinal periódico.

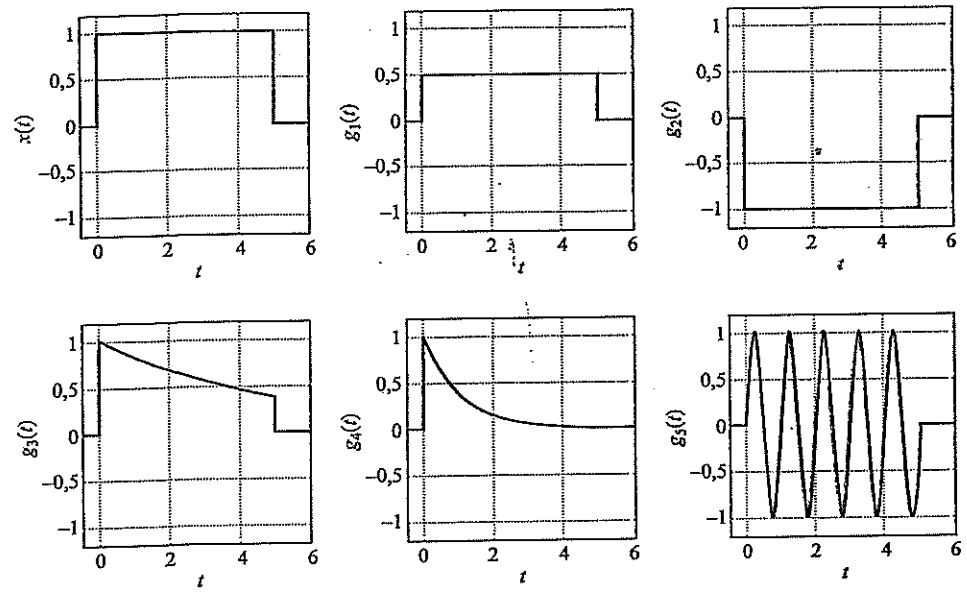


Figura 2.21 Seis sinais simples.

```
Dt=0.01; % incremento temporal Dt
T=6.0;   % duração temporal = T
t=[-1:Dt:T]; %"t" varia em [-1, T] em incrementos Dt
% A seguir, o sinal é calculado no interval de variação de "t" para
gerar o gráfico
x=ustep(t)-ustep(t-5);
g1=0.5*(ustep(t)-ustep(t-5));
g2=-(ustep(t)-ustep(t-5));
g3=exp(-t/5).*(ustep(t)-ustep(t-5));
g4=exp(-t).*(ustep(t)-ustep(t-5));
g5=sin(2*pi*t).*(ustep(t)-ustep(t-5));
subplot(231); sig1=plot(t,x,'k');
```

```
xlabel('\it t'); ylabel('{\it x}({\it t})'); % Título do eixo x
set(sig1,'Linewidth',2); % especificação da largura da linha
axis([-.5 6 -1.2 1.2]); grid  % especificação dos intervalos nos eixos
x (de -0,5 a 6) e y (de -1,2 a 1,2)
subplot(232); sig2=plot(t,g1,'k');
xlabel('\it t'); ylabel('{\it g}_1({\it t})');
set(sig2,'Linewidth',2);
axis([-.5 6 -1.2 1.2]); grid
subplot(233); sig3=plot(t,g2,'k');
xlabel('\it t'); ylabel('{\it g}_2({\it t})');
set(sig3,'Linewidth',2);
axis([-.5 6 -1.2 1.2]); grid
subplot(234); sig4=plot(t,g3,'k');
xlabel('\it t'); ylabel('{\it g}_3({\it t})');
set(sig4,'Linewidth',2);
axis([-.5 6 -1.2 1.2]); grid
subplot(235); sig5=plot(t,g4,'k');
xlabel('\it t'); ylabel('{\it g}_4({\it t})');
set(sig5,'Linewidth',2);grid
axis([-.5 6 -1.2 1.2]);
subplot(236); sig6=plot(t,g5,'k');
xlabel('\it t'); ylabel('{\it g}_5({\it t})');
set(sig6,'Linewidth',2);grid
axis([-.5 6 -1.2 1.2]);

% Cálculo da energia de cada sinal
E0=sum(x.*conj(x))*Dt;
E1=sum(g1.*conj(g1))*Dt;
E2=sum(g2.*conj(g2))*Dt;
E3=sum(g3.*conj(g3))*Dt;
E4=sum(g4.*conj(g4))*Dt;
E5=sum(g5.*conj(g5))*Dt;

c0=sum(x.*conj(x))*Dt/(sqrt(E0*E0))
c1=sum(x.*conj(g1))*Dt/(sqrt(E0*E1))
c2=sum(x.*conj(g2))*Dt/(sqrt(E0*E2))
c3=sum(x.*conj(g3))*Dt/(sqrt(E0*E3))
c4=sum(x.*conj(g4))*Dt/(sqrt(E0*E4))
c5=sum(x.*conj(g5))*Dt/(sqrt(E0*E5))
```

Os seis coeficientes de correlação obtidos com o programa são:

```
c0 =
    1
c1 =
    1
c2 =
   -1
c3 =
```

```
    0.9614
c4 =
    0.6282
c5 =
    8.6748e-17
```

Cálculo Numérico de Coeficientes D_n

Existem numerosas maneiras de calcular valores dos coeficientes D_n de uma série de Fourier. A seguir, usaremos o MATLAB para mostrar como usar integração numérica no cálculo de séries de Fourier.

Para efetuar a integração numérica direta da Eq. (2.60), o primeiro passo consiste em definir a expressão simbólica do sinal $g(t)$ em análise. Neste exemplo, usamos a função triangular $\Delta(t)$.

```
% (funct_tri.m)
% Função triangular padrão, com base de -1 a 1
     function y = funct_tri(t)
% Uso y = func_tri(t)
%     t = variável de entrada i
y=((t>-1)-(t>1)).*(1-abs(t));
```

Uma vez que o arquivo `funct_tri.m` define a função $y = g(t)$, podemos efetuar diretamente a integração numérica da Eq. (2.60), considerando um número finito de coeficientes da série de Fourier $\{D_n, n = -N,..., -1, 0, 1,..., N\}$. A seguir, fornecemos um programa em MATLAB denominado `FSexample.m`, que permite o cálculo da série de Fourier de $\Delta(t/2)$, com período $[a, b]$ ($a = -2$, $b = 2$). Neste exemplo, $N = 11$. A execução desse pequeno programa em MATLAB gerará a Fig. 2.22, que mostra a variação da amplitude e da fase de D_n.

```
% (nome do arquivo: FSexp_a.m)
% Este exemplo mostra como calcular, numericamente
% Coeficientes Dn da série de Fourier exponencial
% diretamente.
% O usuário deve definir uma função simbólica
% g(t). Neste exemplo, g(t)=funct_tri(t).
echo off; clear; clf;
  j=sqrt(-1); % Define j para algebra complexa*
  b=2; a=-2;  % Definição de um período do sinal
  tol=1.e-5;  % Especificação da tolerância para erro de integração
  T=b-a;      % comprimento de um período
  N=11;       % Número de componentes da série de Fourier
              % em cada lado da frequência zero
  Fi=[-N:N]*2*pi/T; % Especificação do intervalo de frequência
% D_0 é calculado e armazenado em D(N+1);
  Func= @(t) funct_tri(t/2);
  D(N+1)=1/T*quad(Func,a,b,tol); % a função quad,m de MATLAB usada
                                 para cálculo da integral
for i=1:N
```

*Na verdade, isso não é necessário. Em MATLAB, as letras i e j são definidas internamente como representações da unidade imaginária, embora possam também ser usadas para representar entidades genéricas. (N.T.)

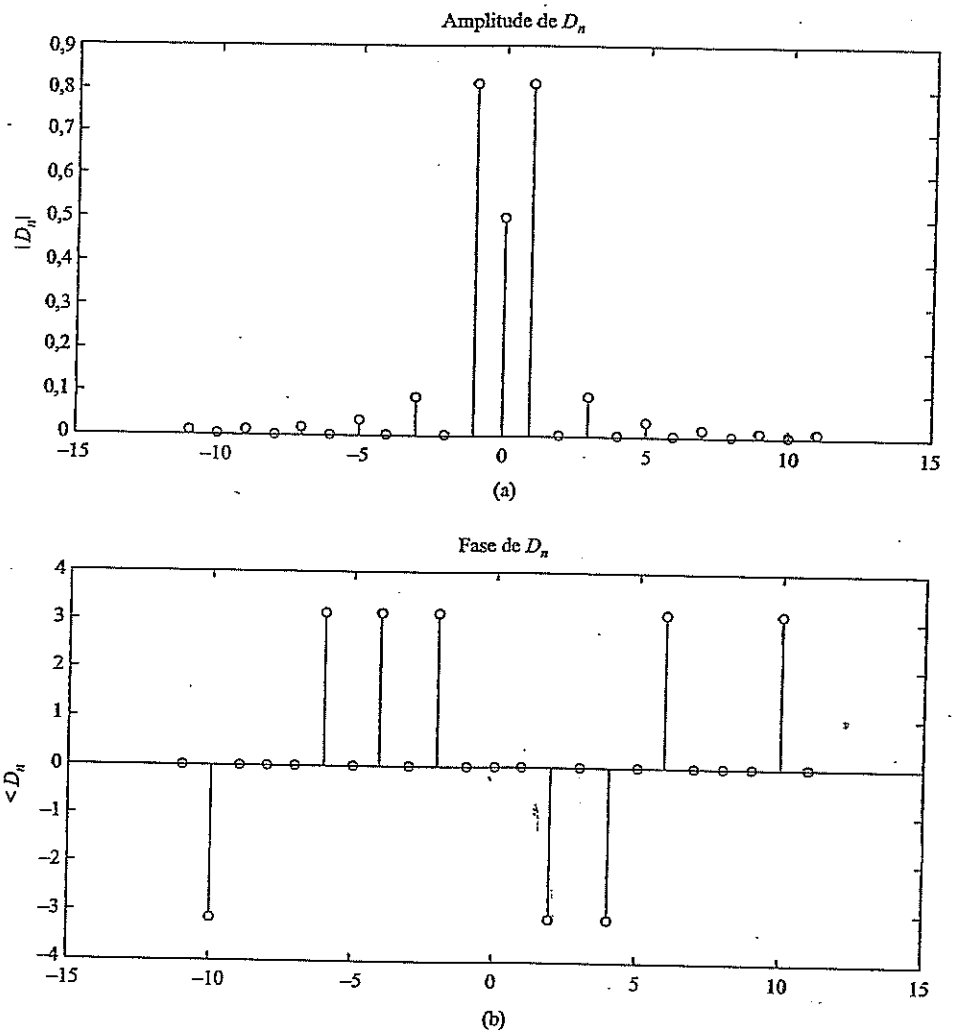


Figura 2.22 Coeficientes da série de Fourier exponencial da repetição de $\Delta(t/2)$ com período $T = 4$.

```
% Dn é calculado, para n=1,...,N, e armazenado em D(N+2), ... D(2N+1)
     Func= @(t) exp(-j*2*pi*t*i/T).*funct_tri(t/2);
     D(i+N+1)=quad(Func,a,b,tol);
% Dn é calculado, para n=-N,...,-1 e armazenado em D(1), ... D(N)
     Func= @(t) exp(j*2*pi*t*(N+1-i)/T).*func_tri(t/2);
     D(i)= quad(Func,a,b,tol);
end
figure(1);
subplot(211);s1=stem([-N:N],abs(D));
set(s1,'Linewidth',2); ylabel('|{\it D}_{\it n}|');
title('Amplitude de {\it D}_{\it n}')
subplot(212);s2=stem([-N:N],angle(D));
set(s2,'Linewidth',2); ylabel('<{\it D}_{\it n}');
title('Fase de {\it D}_{\it n}');
```

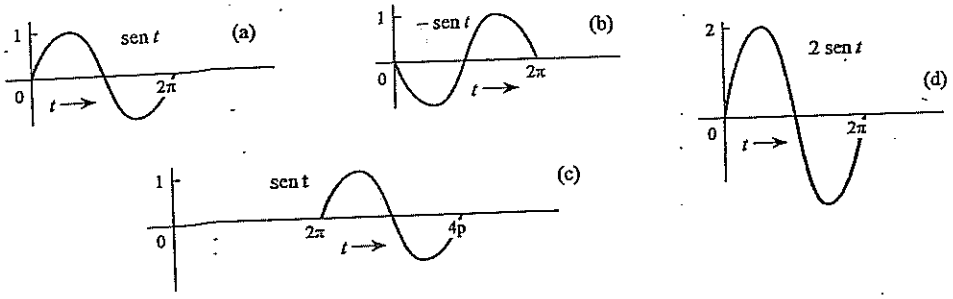
REFERÊNCIAS

1. P. L. Walker, *The Theory of Fourier Series and Integrals*, Wiley-Interscience, New York, 1986.
2. R. V. Churchill, and J. W. Brown, *Fourier Series and Boundary Value Problems*, 3rd ed., McGraw-Hill, New York, 1978.

EXERCÍCIOS

2.1-1 Determine a energia de cada um dos sinais representados na Fig. E.2.1-1. Comente os efeitos de mudança de sinal, translação temporal e multiplicação do sinal por dois sobre a energia. Qual é o efeito sobre a energia se o sinal for multiplicado por k?

Figura E.2.1-1



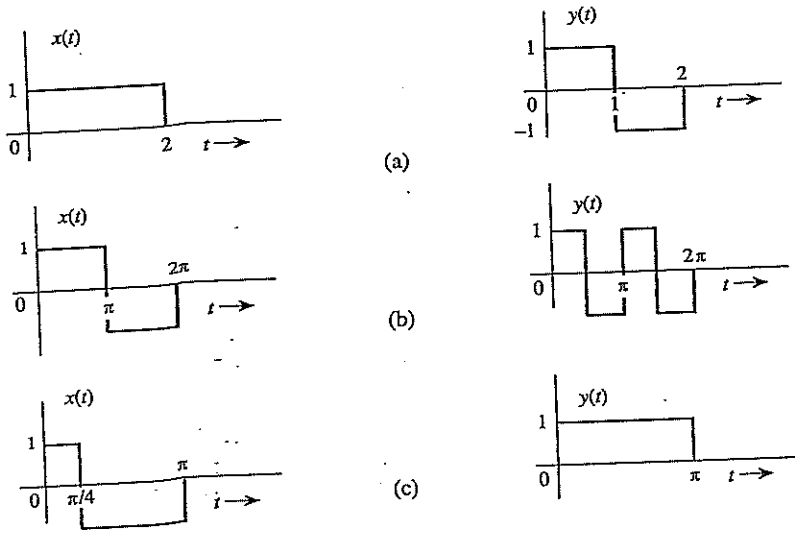
2.1-2 (a) Determine E_x e E_y, energias dos sinais $x(t)$ e $y(t)$ mostrados na Fig. E.2.1-2a. Esboce os gráficos dos sinais $x(t) + y(t)$ e $x(t) - y(t)$, e mostre que a energia de qualquer um desses sinais é igual a $E_x + E_y$. Repita o procedimento para o par de sinais na Fig. E.2.1-2b.

(b) Agora, repita o procedimento para o par de sinais na Fig. E.2.1-2c. Nesse caso, as energias dos sinais $x(t) + y(t)$ e $x(t) - y(t)$ são iguais?

2.1-3 Determine a energia de uma senoide $C\cos(\omega_0 t + \theta)$.

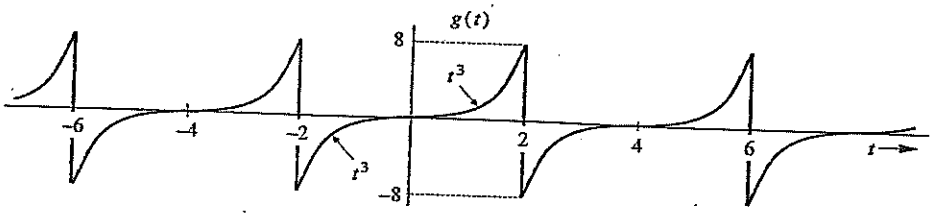
2.1-4 Mostre que, se $\omega_1 = \omega_2$, a potência de $g(t) = C_1 \cos(\omega_1 t + \theta_1) + C_2 \cos(\omega_2 t + \theta_2)$ é $[C_1^2 + C_2^2 + 2C_1C_2 \cos(\theta 1 - \theta 2)]/2$, que não é igual a $\left(C_1^2 + C_2^2\right)/2$.

Figura E.2.1-2



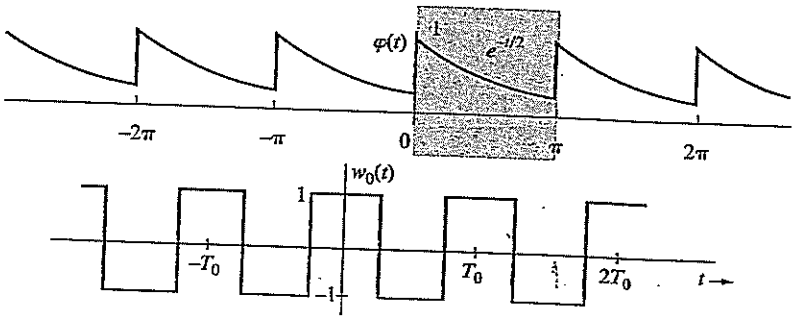
2.1-5 Determine a potência do sinal periódico $g(t)$ mostrado na Fig. E.2.1-5. Determine, também, a potência e valor rms de **(a)** $-g(t)$, **(b)** $2g(t)$ e **(c)** $cg(t)$. Comente os resultados.

Figura E.2.1-5



2.1-6 Determine a potência e valor rms dos sinais na **(a)** Fig. E.2.1-6a, **(b)** Fig. 2.16, **(c)** Fig. E.2.1-6b, **(d)** Fig. E.2.7-4a, **(e)** Fig. E.2.7-4c.

Figura E.2.1-6



2.1-7 Mostre que a potência de um sinal $g(t)$ dado por

$$g(t) = \sum_{k=m}^{n} D_k e^{j\omega_k t} \qquad \omega_i \neq \omega_k \text{ para todo } i \neq k$$

é (teorema de Parseval)

$$P_g = \sum_{k=m}^{n} |D_k|^2$$

2.1-8 Determine a potência e o valor rms dos seguintes sinais:

(a) $10\cos\left(100t + \frac{\pi}{3}\right)$

(b) $10\cos\left(100t + \frac{\pi}{3}\right) + 16\,\text{sen}\left(150t + \frac{\pi}{5}\right)$

(c) $(10 + 2\,\text{sen}\,3t)\cos 10t$

(d) $10\cos 5t \cos 10t$

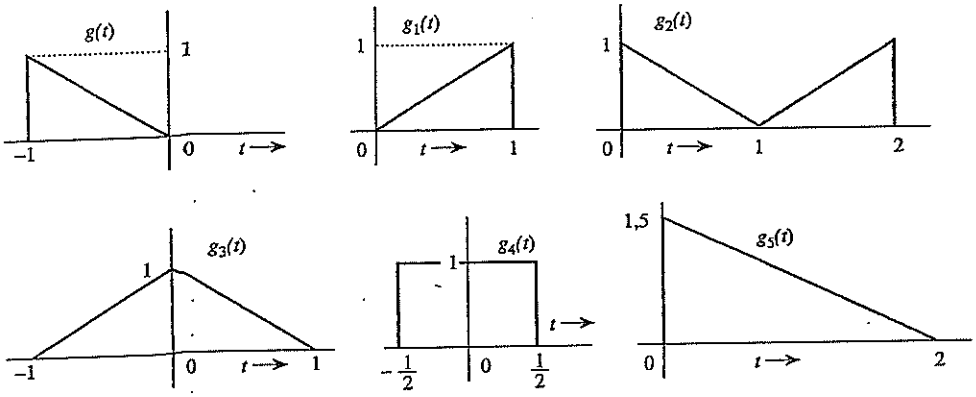
(e) $10\,\text{sen}\,5t \cos 10t$

(f) $e^{j\alpha t}\cos\omega_0 t$

2.2-1 Mostre que uma exponencial e^{-at} com início em $-\infty$ não é um sinal de energia e nem um sinal de potência, para qualquer valor real de a. A seguir, para a imaginário, mostre que é um sinal de potência, com potência $P_g = 1$, independentemente do valor de a.

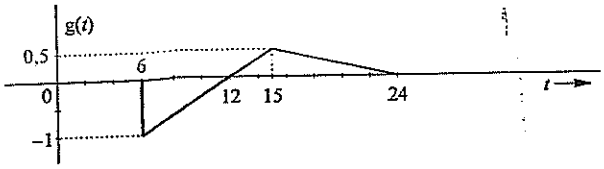
2.3-1 Na Fig. E.2.3-1, o sinal $g_1(t) = g(-t)$. Expresse os sinais $g_2(t)$, $g_3(t)$, $g_4(t)$ e $g_5(t)$ em termos dos sinais $g(t)$, $g_1(t)$ e de suas versões nas quais o eixo do tempo é deslocado, dilatado e invertido. Por exemplo, $g_2(t) = g(t-T) + g_1(t-T)$, para algum valor adequado de T. Do mesmo modo, $g_3(t)$ e $g_4(t)$ podem ser expressos como $g(t-T) + g(t-T)$, para algum valor adequado de T. Por sua vez, $g_5(t)$ pode ser expresso em termos de $g(t)$ com o eixo do tempo deslocado, dilatado e, por fim, $g(t)$ multiplicado por uma constante.

Figura E.2.3-1



2.3-2 Para o sinal $g(t)$ mostrado na Fig. E.2.3-2, esboce gráficos dos seguintes sinais: **(a)** $g(-t)$, **(b)** $g(t+6)$, **(c)** $g(3t)$, **(d)** $g(6-t)$.

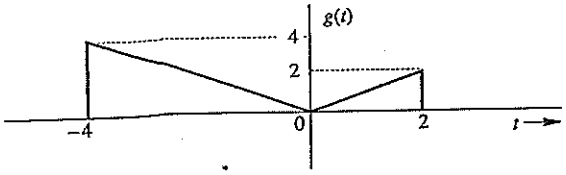
Figura E.2.3-2



2.3-3 Para o sinal $g(t)$ mostrado na Fig. E.2.3-3, esboce gráficos dos seguintes sinais: **(a)** $g(t-4)$, **(b)** $g(t/1{,}5)$, **(c)** $g(2t-4)$, **(d)** $g(2-t)$.

Sugestão: Lembre-se de que a substituição de t por $t-T$ atrasa o sinal em T. Portanto, $g(2t-4)$ é $g(2t)$ com t substituído por $t-2$. Da mesma forma, $g(2-t)$ é $g(-t)$ com t substituído por $t-2$.

Figura E.2.3-3



2.3-4 Para um sinal de energia $g(t)$ com energia E_g, mostre que a energia de qualquer um dos sinais $-g(t)$, $g(-t)$ e $g(t-T)$ é E_g. Mostre, também, que a energia de $g(at)$ e a de $g(at-b)$ são iguais a E_g/a. Isso demonstra que a inversão e a translação temporais não afetam a energia de sinais. Contudo, a compressão temporal de um sinal por um fator a (t substituído por at) divide a energia pelo mesmo fator. Qual é o efeito sobre a energia quando o sinal é **(a)** dilatado temporalmente por um fator a (t substituído por at, $a > 1$) e **(b)** multiplicado por uma constante a?

2.3-5 Simplifique as seguintes expressões:

(a) $\left(\dfrac{\tan t}{2t^2+1}\right)\delta(t)$

(b) $\left(\dfrac{j\omega-3}{\omega^2+9}\right)\delta(\omega)$

(c) $\left[e^{-t}\cos(3t-\pi/3)\right]\delta(t+\pi)$

(d) $\left(\dfrac{\operatorname{sen}\pi(t+2)}{t^2-4}\right)\delta(t-1)$

(e) $\left(\dfrac{\cos(\pi t)}{t+2}\right)\delta(2t+3)$

(f) $\left(\dfrac{\operatorname{sen}k\omega}{\omega}\right)\delta(\omega)$

Sugestão: Use a Eq. (2.10b). Para a parte (f), use a regra de L'Hôpital.

2.3-6 Calcule as seguintes integrais:

(a) $\int_{-\infty}^{\infty} g(\tau)\delta(t-\tau)\,d\tau$

(b) $\int_{-\infty}^{\infty} \delta(\tau)g(t-\tau)\,d\tau$

(c) $\int_{-\infty}^{\infty} \delta(t)e^{-j\omega t}\,dt$

(d) $\int_{-\infty}^{1} \delta(t-2)\operatorname{sen}\pi t\,dt$

(e) $\int_{-2}^{\infty} \delta(3+t)e^{-t}\,dt$

(f) $\int_{-2}^{2} (t^3+4)\delta(1-t)\,dt$

(g) $\int_{-\infty}^{\infty} g(2-t)\delta(3-t)\,dt$

(h) $\int_{-\infty}^{\infty} e^{(x-1)}\cos\frac{\pi}{2}(x-5)\delta(2x-3)\,dx$

Sugestão: $\delta(x)$ está localizado em $x=0$. Por exemplo, $\delta(1-t)$ está localizado em $1-t=0$, ou seja, em $t=1$, e assim por diante.

2.3-7 Prove que

$$\delta(at)=\frac{1}{|a|}\delta(t)$$

Depois, mostre que

$$\delta(\omega)=\frac{1}{2\pi}\delta(f) \qquad \text{onde} \qquad \omega=2\pi f$$

Sugestão: mostre que

$$\int_{-\infty}^{\infty}\phi(t)\delta(at)\,dt=\frac{1}{|a|}\phi(0)$$

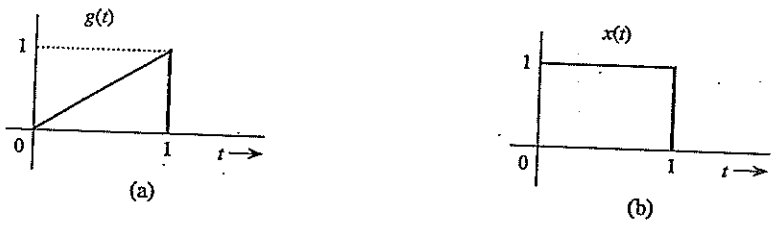
2.4-1 Deduza a Eq. (2.19) de modo alternativo; para isto, observe que $\mathbf{e}=(\mathbf{g}-c\mathbf{x})$ e

$$|\mathbf{e}|^2=(\mathbf{g}-c\mathbf{x})\cdot(\mathbf{g}-c\mathbf{x})=|\mathbf{g}|^2+c^2|\mathbf{x}|^2-2c\mathbf{g}\cdot\mathbf{x}$$

Para minimizar $|\mathbf{e}|^2$, iguale a zero sua derivada em relação a c.

2.4-2 Para os sinais $g(t)$ e $x(t)$ mostrados na Fig. E.2.4-2, determine a componente da forma de $x(t)$ contida em $g(t)$. Em outras palavras, determine o valor ótimo de c na aproximação $g(t)\approx cx(t)$, de modo que a energia do sinal de erro seja mínima. Qual é o valor da energia do resultante sinal de erro?

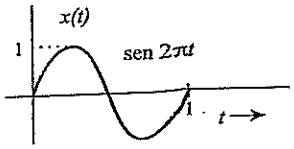
Figura E.2.4-2



2.4-3 Para os sinais $g(t)$ e $x(t)$ mostrados na Fig. E.2.4-2, determine a componente da forma de $g(t)$ contida em $x(t)$. Em outras palavras, determine o valor ótimo de c na aproximação $x(t) \approx cg(t)$, de modo que a energia do sinal de erro seja mínima. Qual é o valor da energia do resultante sinal de erro?

2.4-4 Refaça o Exercício 2.4-2 para o caso em que $x(t)$ é o pulso senoidal na Fig. E.2.2-4

Figura E.2.4-4



2.4-5 As energias de dois sinais $x(t)$ e $y(t)$ são E_x e E_y, respectivamente.

(a) Se $x(t)$ e $y(t)$ forem ortogonais, mostre que a energia do sinal $x(t) + y(t)$ é igual à energia do sinal $x(t) - y(t)$ e dada por $Ex + Ey$.

(b) Se $x(t)$ e $y(t)$ forem ortogonais, determine as energias dos sinais $c_1x(t) + c_2y(t)$ e $c_1x(t) - c_2y(t)$.

(c) A energia cruzada de dois sinais de energia $x(t)$ e $y(t)$, E_{xy}, é definida como

$$E_{xy} = \int_{-\infty}^{\infty} x(t)y^*(t)\,dt$$

Se $z(t) = x(t) \pm y(t)$, mostra que

$$E_z = E_x + E_y \pm (E_{xy} + E_{yx})$$

2.4-6 Sejam $x_1(t)$ e $x_2(t)$ dois sinais de energia unitária e ortogonais no intervalo de $t = t_1$ a t_2. Como os sinais $x_1(t)$ e $x_2(t)$ têm energia unitária e são ortogonais, podem ser representados como dois vetores ortogonais e de comprimentos unitários $(\mathbf{x}_1, \mathbf{x}_2)$. Considere um sinal $g(t)$, dado por

$$g(t) = c_1x_1(t) + c_2x_2(t) \qquad t_1 \le t \le t_2$$

Este sinal pode ser representado como um vetor $\mathbf{g}$ por meio de um ponto (c_1, c_2) no plano $x_1 - x_2$.

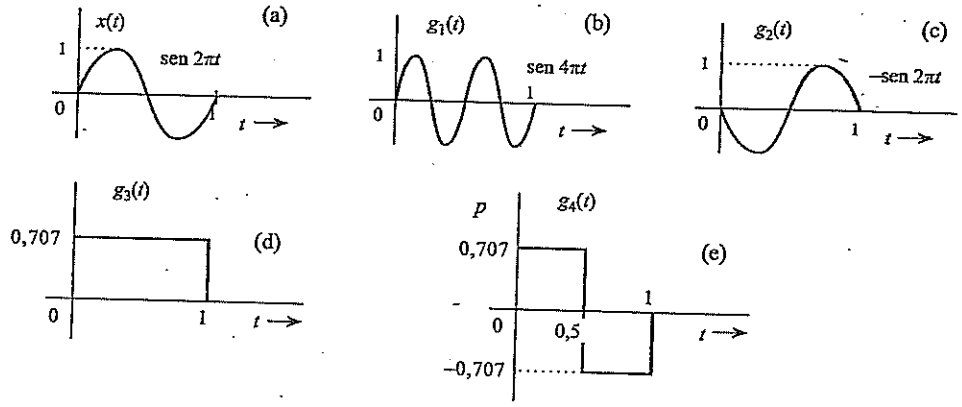
(a) Determine, neste espaço bidimensional, a representação vetorial dos seguintes seis sinais:

(i) $g_1(t) = 2x_1(t) - x_2(t)$
(ii) $g_2(t) = -x_1(t) + 2x_2(t)$
(iii) $g_3(t) = -x_2(t)$
(iv) $g_4(t) = x_1(t) + 2x_2(t)$
(v) $g_5(t) = 2x_1(t) + x_2(t)$
(vi) $g_6(t) = 3x_1(t)$

(b) Entre os seis vetores do item (a), identifique pares de vetores mutuamente ortogonais. Verifique que os pares de sinais correspondentes a estes vetores também são ortogonais.

2.5-1 Determine os coeficientes de correlação c_n do sinal $x(t)$ e de cada um dos quatro pulsos $g_1(t)$, $g_2(t)$, $g_3(t)$ e $g_4(t)$ mostrados na Fig. E.2.5-1. Para garantir máxima margem contra ruído ao longo da rota de transmissão, que par de pulsos você selecionaria para uma comunicação binária?

Figura E.2.5-1



2.7-1 (a) Esboce o gráfico do sinal $g(t) = t^2$ e determine a série de Fourier exponencial para representar $g(t)$ no intervalo $(-1, 1)$. Esboce o gráfico da série de Fourier $\phi(t)$ para todos os valores de t.

(b) Comprove o teorema de Parseval [Eq. (2.68a)] para este caso, dado que

$$\sum_{n=1}^{\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$$

2.7-2 (a) Esboce o gráfico do sinal $g(t) = t$ e determine a série de Fourier exponencial para representar $g(t)$ no intervalo $(-\pi, \pi)$. Esboce o gráfico da série de Fourier $\phi(t)$ para todos os valores de t.

(b) Comprove o teorema de Parseval [Eq. (2.68a)] para este caso, dado que

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

2.7-3 Se um sinal periódico satisfizer certas condições de simetria, a evolução dos coeficientes da série de Fourier pode ser um pouco simplificada.

(a) Mostre que, se $g(t) = g(-t)$ (simetria par), os coeficientes da série de Fourier exponencial são reais.

(b) Mostre que, se $g(t) = -g(-t)$ (simetria ímpar), os coeficientes da série de Fourier exponencial são imaginários.

(c) Mostre que, em cada caso, os coeficientes de Fourier podem ser calculados por meio da integração de sinais periódicos ao longo de apenas meio período. Isso ocorre porque toda a informação de um período está implícita em um meio período, devido à simetria.

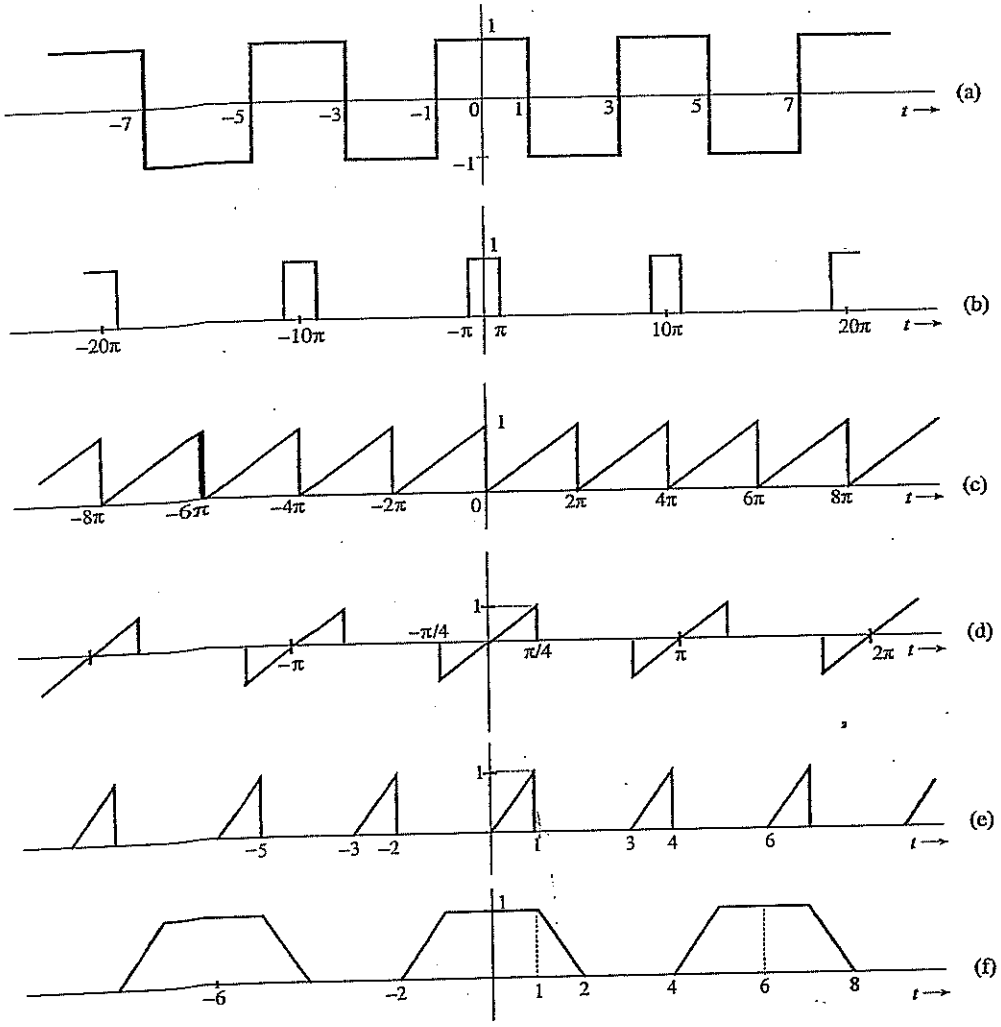
Sugestão: Se $g_e(t)$ e $g_o(t)$ forem funções par e ímpar, respectivamente, de t, então (assumindo que não há impulso ou suas derivadas na origem),

$$\int_{-a}^{a} g_e(t)\,dt = \int_{0}^{2a} g_e(t)\,dt \quad \text{e} \quad \int_{-a}^{a} g_o(t)\,dt = 0$$

Além disso, o produto de uma função par por uma função ímpar é uma função ímpar; o produto de duas funções pares é uma função par; o produto de duas funções ímpares é uma função ímpar.

2.7-4 Para cada sinal periódico mostrado na Fig. E.2.7-4, determine a série de Fourier exponencial e esboce os gráficos dos espectros de amplitude e de fase. Relate qualquer propriedade de simetria.

Figura E.2.7-4



2.7-5 (a) Mostre que uma função arbitrária $g(t)$ pode ser expressa como uma soma de uma função par $g_e(t)$ e de uma função ímpar $g_o(t)$:

$$g(t) = g_e(t) + g_o(t)$$

$$\textit{Sugestão:} \quad g(t) = \underbrace{\frac{1}{2}\Big[g(t) + g(-t)\Big]}_{g_e(t)} + \underbrace{\frac{1}{2}\Big[g(t) - g(-t)\Big]}_{g_o(t)}$$

(b) Determine as componentes pares e ímpares das seguintes funções: **(i)** $u(t)$; **(ii)** $e^{-\omega(t)}u(t)$; **(iii)** e^{jt}.

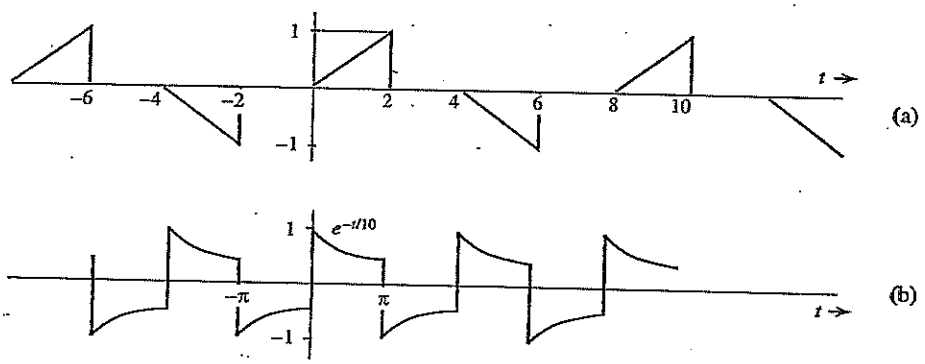
2.7-6 (a) Se, em um período, as duas metades de sinal periódico tiverem a mesma forma e fases opostas, o sinal periódico tem **simetria de meia onda**. Se um sinal periódico $g(t)$, de período T_0, satisfizer a condição de simetria de meia onda, então,

$$g\left(t - \frac{T_0}{2}\right) = -g(t)$$

Neste caso, mostre que todos os harmônicos (coeficientes) de ordem par são nulos.

(b) Use este resultado para determinar a série de Fourier para os sinais periódicos na Fig. E.2.7-6.

Figura E.2.7-6



2.8-1 Um sinal periódico $g(t)$ é expresso pela seguinte série de Fourier:

$$g(t) = 3\,\text{sen}\,t + \cos\left(3t - \frac{2\pi}{3}\right) + 2\cos\left(8t + \frac{\pi}{3}\right)$$

(a) Aplique as identidades de Euler diretamente ao sinal $g(t)$ e escreva sua série de Fourier exponencial.

(b) Aplique as identidades de Euler diretamente ao sinal $g(t)$ e esboce gráficos dos espectros da correspondente série de Fourier exponencial.

3 ANÁLISE E TRANSMISSÃO DE SINAIS

Engenheiros eletricistas, de modo instintivo, quando pensam sobre sinais e sistema o fazem em termos de espectros de frequência e de respostas em frequência, respectivamente. Até mesmo adolescentes sabem que sinais de áudio têm largura de banda de 20 kHz e que alto-falantes de boa qualidade devem responder até 20 kHz. Isso significa pensar no domínio da frequência. No capítulo anterior, discutimos representações espectrais de sinais periódicos (série de Fourier). No presente capítulo, estenderemos esta representação espectral a sinais aperiódicos.

3.1 REPRESENTAÇÃO DE SINAIS APERIÓDICOS ATRAVÉS DA INTEGRAL DE FOURIER

Aplicaremos um processo de limite para mostrar que um sinal aperiódico pode ser expresso como uma soma contínua (integral) de exponenciais eternas. Para representar um sinal aperiódico $g(t)$, como o mostrado na Fig. 3.1a, em termos de exponenciais infinitas no tempo, construamos um sinal periódico $g_{T_0}(t)$ com a repetição do sinal $g(t)$ a cada T_0 segundos, como indicado na Fig. 3.1b. O período T_0 deve ser suficientemente longo para evitar sobreposição de pulsos repetidos. O sinal periódico $g_{T_0}(t)$ pode ser representado por uma série de Fourier exponencial. Se fizermos $T_0 \to \infty$, os pulsos no sinal periódico se repetirão após um intervalo infinito, de modo que

$$\lim_{T_0 \to \infty} g_{T_0}(t) = g(t)$$

Portanto, no limite $T_0 \to \infty$, a série de Fourier que representa $g_{T_0}(t)$ também representa $g(t)$. A série de Fourier exponencial para $g_{T_0}(t)$ é dada por

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} D_n e^{jn\omega_0 t} \tag{3.1}$$

com

$$D_n = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} g_{T_0}(t) e^{-jn\omega_0 t}\, dt \tag{3.2a}$$

e

$$\omega_0 = \frac{2\pi}{T_0} = 2\pi f_0 \tag{3.2b}$$

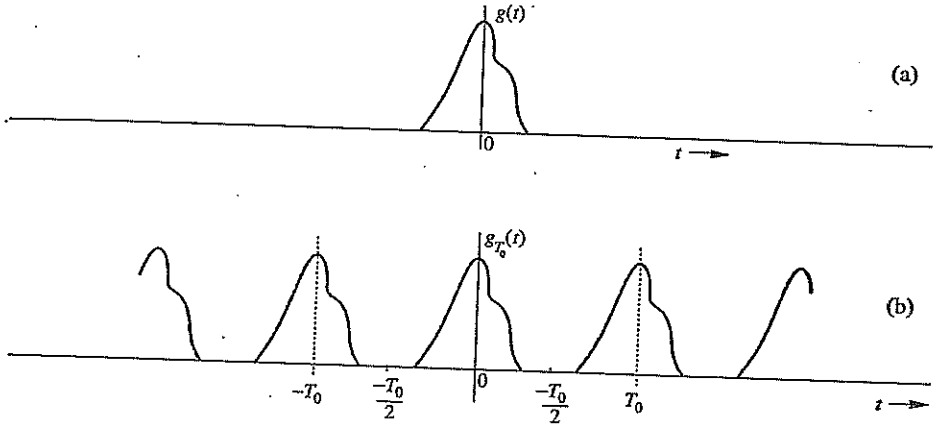


Figura 3.1 Construção de um sinal periódico através da repetição periódica de $g(t)$.

Vale observar que integrar $g_{T_0}(t)$ no intervalo $(-T_0/2, T_0/2)$ é o mesmo que integrar $g(t)$ em $(-\infty, \infty)$. Portanto, a Eq. (3.2a) pode ser expressa como

$$D_n = \frac{1}{T_0}\int_{-\infty}^{\infty} g(t)e^{-jn\omega_0 t}\,dt = \frac{1}{T_0}\int_{-\infty}^{\infty} g(t)e^{-j2\pi n f_0 t}\,dt \tag{3.2c}$$

É interessante notar como a natureza do espectro é alterada à medida que T_0 aumenta. Para entender esse fascinante comportamento, definamos $G(f)$, uma função contínua de ω, como

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-j\omega t}\,dt \tag{3.3}$$

$$= \int_{-\infty}^{\infty} g(t)e^{-j2\pi f t}\,dt \tag{3.4}$$

em que $\omega = 2\pi f$. Um exame das Eqs. (3.2c) e (3.3) revela que

$$D_n = \frac{1}{T_0}G(nf_0) \tag{3.5}$$

E também que os coeficientes de Fourier D_n são iguais a $(1/T_0)$ vezes as amostras de $G(f)$ uniformemente espaçadas a intervalos de f_0 Hertz, como mostrado na Fig. 3.2a.*

Portanto, $(1/T_0)G(f)$ é o envelope dos coeficientes D_n. Agora, para tomar o limite $T_0 \to \infty$, dobremos o valor de T_0 sucessivamente. Dobrar T_0 significa dividir por dois o valor da frequência fundamental f_0, ou seja, dobrar o número de componentes (amostras) no espectro. No entanto, dobrar o valor de T_0 implica dividir a amplitude do envelope $(1/T_0)G(f)$ por dois, como ilustrado na Fig. 3.2b. Se continuarmos com esse processo de dobrar o valor de T_0 sucessivamente, o espectro se torna cada vez mais denso e sua magnitude, menor. Vale observar que, no entanto, a forma do espectro permanece a mesma [proporcional a $G(f)$ da Eq. (3.3)]. No limite, à medida que $T_0 \to \infty$, $f_0 \to 0$ e $D_n \to 0$. Isso significa que o espectro se torna tão denso que as componentes espectrais são espaçadas por um intervalo nulo (infinitesimal). Ao mesmo tempo, a amplitude de cada componente é zero (infinitesimal). Temos *nada de tudo, embora tenhamos algo*! Isto soa como *Alice no País das Maravilhas* e, como veremos, é característico de um fenômeno muito familiar.**

*Para simplificar, consideramos, na Fig. 3.2, que D_n e, portanto, $G(f)$ são reais. Contudo, o argumento também se aplica a D_n [ou $G(f)$] complexo.

**Considere isso uma prova irrefutável da máxima que diz que ter 0% de tudo é melhor do que ter 100% de nada!

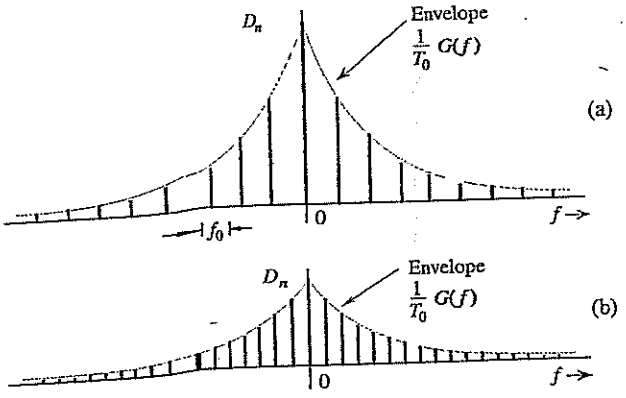


Figura 3.2 Mudança no espectro de Fourier à medida que o período T_0 na Fig. 3.1 é dobrado.

A substituição da Eq. (3.5) na Eq. (3.1) leva a

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} \frac{G(nf_0)}{T_0} e^{jn2\pi f_0 t} \tag{3.6}$$

À medida que $T_0 \to \infty, f_0 = 1/T_0$ se torna infinitesimal ($f_0 \to 0$). Em consequência, podemos substituir f_0 por uma notação mais apropriada, Δf. Em termos desta nova notação, a Eq. (3.2b) passa a

$$\Delta f = \frac{1}{T_0}$$

e a Eq. (3.6),

$$g_{T_0}(t) = \sum_{n=-\infty}^{\infty} [G(n\Delta f)\Delta f]\, e^{(j2\pi n\Delta f)t} \tag{3.7a}$$

A Eq. (3.7a) mostra que $g_{T_0}(t)$ pode ser expressa como uma soma de exponenciais infinitas no tempo de frequências $0, \pm\Delta f, \pm 2\Delta f, \pm 3\Delta f, \ldots$ (a série de Fourier). A amplitude da componente de frequência $n\Delta f$ é $[G(n\Delta f)\Delta f]$. No limite, quando $T_0 \to \infty$, $\Delta f \to 0$ e $g_{T_0}(t) \to g(t)$. Portanto,

$$g(t) = \lim_{T_0 \to \infty} g_{T_0}(t) = \lim_{\Delta f \to 0} \sum_{n=-\infty}^{\infty} G(n\Delta f) e^{(j2\pi n\Delta f)t} \Delta f \tag{3.7b}$$

A soma no lado direito da Eq. (3.7b) pode ser vista como a área sob a curva da função $G(f)e^{j2\pi ft}$, como mostrado na Fig. 3.3. Logo,

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df \tag{3.8}$$

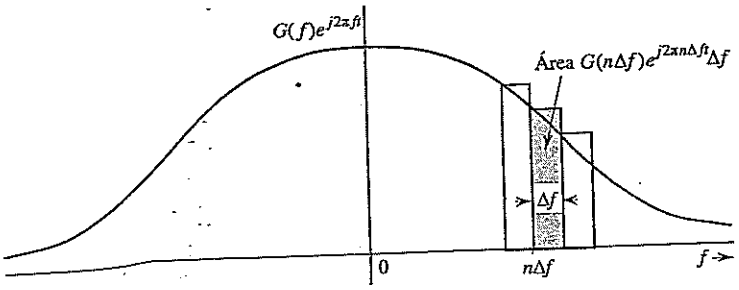


Figura 3.3 No limite $T_0 \to \infty$, a série de Fourier se torna a integral de Fourier.

A integral no lado direito da Eq. (3.8) é denominada **integral de Fourier.** Conseguimos, então, representar um sinal aperiódico $g(t)$ por uma integral de Fourier* (em vez de uma série de Fourier). Essa integral é basicamente (no limite) uma série de Fourier com frequência fundamental $\Delta f \to 0$, como visto na Eq. (3.7b). A amplitude da exponencial $e^{j2\pi n\Delta f t}$ é $G(n\Delta f)\Delta f$. Assim, a função $G(f)$ dada pela Eq. (3.3) atua como uma função espectral.

Denominamos $G(f)$ a transformada de Fourier **direta** de $g(t)$, e $g(t)$ a transformada de Fourier **inversa** de $G(f)$. A mesma informação é transmitida quando dizemos que $g(t)$ e $G(f)$ formam um par de transformadas de Fourier. Simbolicamente, isso é representado como

$$G(f) = \mathcal{F}[g(t)] \qquad \text{e} \qquad g(t) = \mathcal{F}^{-1}[G(f)]$$

ou

$$g(t) \Longleftrightarrow G(f)$$

Recapitulando,

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-j\omega t}\,dt \tag{3.9a}$$

e

$$g(t) = \int_{-\infty}^{\infty} G(f)e^{j\omega t}\,df \tag{3.9b}$$

em que $\omega = 2\pi f$.

É interessante termos em mente que a integral de Fourier na Eq. (3.9b) tem a natureza de uma série de Fourier com frequência fundamental Δf que tende a zero [Eq. (3.7b)]. Portanto, a maioria das considerações sobre a série de Fourier, assim como suas propriedades, se aplica também à transformada de Fourier. Podemos desenhar o gráfico do espectro $G(f)$ em função de f. Como $G(f)$ tem valores complexos, temos espectros de amplitude e de ângulo (fase):

$$G(f) = |G(f)|e^{j\theta_g(f)}$$

em que $|G(f)|$ é a amplitude e $\theta_g(f)$, o ângulo (ou fase) de $G(f)$. Da Eq. (3.9a),

$$G(-f) = \int_{-\infty}^{\infty} g(t)e^{j2\pi ft}\,dt$$

f versus ω

Na representação de sinais no domínio da frequência, usamos as notações equivalentes de frequência angular ω e frequência f indiscriminadamente. Não existe qualquer diferença conceitual entre o emprego de frequência angular ω (cuja unidade é radiano por segundo) e frequência (cuja unidade é hertz, Hz). Dada a relação direta entre as duas frequências, podemos substituir $\omega = 2\pi f$ em $G(f)$ e obter a expressão da transformada de Fourier no domínio ω:

$$\mathcal{F}[g(t)] = \int_{-\infty}^{\infty} g(t)e^{-j\omega t}dt \tag{3.10}$$

Devido ao fator adicional 2π na variável ω usada na Eq. (3.10), a expressão da transformada inversa em função de ω requer uma divisão por 2π. Portanto, a notação em termos de f é, em geral, preferida na prática para a expressão de transformadas de Fourier. Por isto, neste livro, denotaremos, na maioria dos casos, a transformada de Fourier de sinais como funções $G(f)$. A notação de frequência angular ω, por sua vez, também pode oferecer alguma conveniência na representação de senoides. Assim, em capítulos posteriores, sempre que for *conveniente e não gerar confusão* usaremos as duas notações equivalentes de modo indiscriminado.

*Isso não deve ser considerado uma prova rigorosa da Eq. (3.8). A situação não é tão simples como fizemos parecer.[1]

Propriedade de Simetria Conjugada

A partir da Eq. (3.9a), concluímos que, se $g(t)$ for uma função de t de valores reais, $G(f)$ e $G(-f)$ são complexos conjugados, ou seja,*

$$G(-f) = G^*(f) \tag{3.11}$$

Logo,

$$|G(-f)| = |G(f)| \tag{3.12a}$$

$$\theta_g(-f) = -\theta_g(f) \tag{3.12b}$$

Assim, para $g(t)$ de valores reais, o espectro de amplitude $|G(f)|$ é uma função par e o espectro de fase $\theta_g(f)$, uma função ímpar de f. Esta propriedade (**propriedade de simetria conjugada**) é válida apenas para funções $g(t)$ de valores reais. Esse resultado não deve causar surpresa, pois foi obtido no Capítulo 2 para o espectro de Fourier de um sinal periódico. *A transformada $G(f)$ é a especificação de $g(t)$ no domínio da frequência.*

Exemplo 3.1 Determinemos a transformada de Fourier de $e^{-at}u(t)$.

Por definição [Eq. (3.9a)],

$$G(f) = \int_{-\infty}^{\infty} e^{-at}u(t)e^{-j2\pi ft}\,dt = \int_0^{\infty} e^{-(a+j2\pi f)t}\,dt = \frac{-1}{a+j2\pi f}e^{-(a+j2\pi f)t}\Big|_0^{\infty}$$

Mas $|e^{-j2\pi ft}| = 1$. Logo, quando $t \to \infty$, $e^{-(a+j2\pi f)t} = e^{-at}e^{-j2\pi ft} = 0$ se $a > 0$. Portanto,

$$G(f) = \frac{1}{a+j\omega} \qquad a > 0 \tag{3.13a}$$

em que $\omega = 2\pi f$. Expressando $a + j\omega$ na forma polar como $\sqrt{a^2+\omega^2}\,e^{j\tan^{-1}(\frac{\omega}{a})}$, obtemos

$$G(f) = \frac{1}{\sqrt{a^2+(2\pi f)^2}}e^{-j\tan^{-1}(\frac{2\pi f}{a})} \tag{3.13b}$$

Logo,

$$|G(f)| = \frac{1}{\sqrt{a^2+(2\pi f)^2}} \qquad \text{e} \qquad \theta_g(f) = -\tan^{-1}\left(\frac{2\pi f}{a}\right)$$

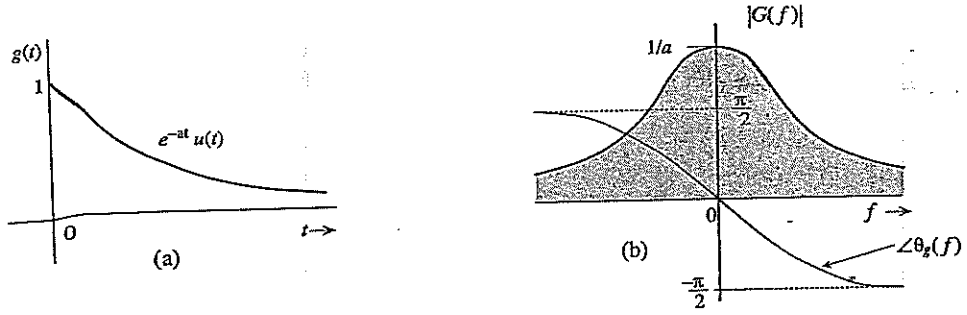


Figura 3.4 $e^{-at}u(t)$ e correspondente espectro de Fourier.

Os espectros de amplitude, $|G(f)|$, e de fase, $\theta_g(f)$, de $e^{-at}u(t)$ são mostrados na Fig. 3.4b. Observemos que $|G(f)|$ é uma função par de f e $\theta_g(f)$, uma função ímpar de f, como esperado.

Simetria hermitiana é o termo usado para descrever funções complexas que satisfazem a Eq. (3.11).

Existência da Transformada de Fourier

No Exemplo 3.1, observamos que, quando $a < 0$, a integral de Fourier para $e^{-at}u(t)$ não converge. Em consequência, a transformada de Fourier de $e^{-at}u(t)$ não existe para $a < 0$ (exponencial crescente). Fica claro que nem todos os sinais podem ser transformados pela integral de Fourier. A existência da transformada de Fourier é assegurada para qualquer $g(t)$ que satisfaça as condições de Dirichlet, sendo a primeira delas*

$$\int_{-\infty}^{\infty} |g(t)|\, dt < \infty \tag{3.14}$$

Para mostrar isso, recordemos que $|e^{-j2\pi ft}| = 1$. Assim, a partir da Eq. (3.9a), obtemos

$$|G(f)| \leq \int_{-\infty}^{\infty} |g(t)|\, dt$$

Isto mostra que a existência da transformada de Fourier fica assegurada se a condição (3.14) for satisfeita. Caso contrário, não há garantia. Vimos no Exemplo 3.1 que a transformada de Fourier de um sinal exponencial crescente (que viola essa condição) não existe. Embora tal condição seja suficiente, não é necessária para a existência da transformada de Fourier de um sinal. Por exemplo, o sinal (sen at)/t viola a condição (3.14), mas tem uma transformada de Fourier. Qualquer sinal que possa ser gerado na prática satisfaz as condições de Dirichlet e, portanto, tem uma transformada de Fourier. Dessa forma, a existência física de um sinal é uma condição suficiente para a existência de sua transformada de Fourier.

Linearidade da Transformada de Fourier (Teorema da Superposição)

A transformada de Fourier é linear, ou seja, se

$$g_1(t) \Longleftrightarrow G_1(f) \quad \text{e} \quad g_2(t) \Longleftrightarrow G_2(f)$$

então, para quaisquer constantes a_1 e a_2, temos

$$a_1 g_1(t) + a_2 g_2(t) \Longleftrightarrow a_1 G_1(f) + a_2 G_2(f) \tag{3.15}$$

A prova é simples e resulta diretamente da Eq. (3.9a). Esse teorema simplesmente afirma que combinações lineares de sinais no domínio do tempo correspondem, no domínio da frequência, a combinações lineares de suas transformadas de Fourier. Esse resultado pode ser estendido a um número finito arbitrário de termos:

$$\sum_k a_k g_k(t) \Longleftrightarrow \sum_k a_k G_k(f)$$

para quaisquer constantes $\{a_k\}$ e sinais $\{g_k(t)\}$.

Interpretação Física da Transformada de Fourier

Para o entendimento de aspectos da transformada de Fourier, devemos lembrar que a representação de Fourier é uma forma de expressar um sinal em termos de senoides infinitas no tempo ou exponenciais. O espectro de Fourier de um sinal indica as amplitudes e fases relativas das senoides necessárias à síntese do sinal. O espectro de Fourier de um sinal periódico tem amplitudes finitas e existe em frequências discretas (f e seus múltiplos). É fácil visualizar um espectro desse tipo; mas, no caso de sinais aperiódicos, a visualização do espectro não é simples, pois o mesmo é contínuo e existe em todas as frequências. O conceito de espectro contínuo pode ser interpretado por meio da consideração de um fenômeno análogo e mais tangível. Um exemplo familiar de uma distribuição contínua é o carregamento de uma viga. Consideremos uma viga submetida à carga de pesos $D_1, D_2, D_3, \ldots, D_n$ loca-

*As condições de Dirichlet restantes são: em qualquer intervalo finito, $g(t)$ pode ter apenas um número finito de máximos e mínimos e um número finito de descontinuidades finitas. Quando essas condições são satisfeitas, a integral de Fourier no lado direito da Eq. (3.9b) converge para $g(t)$ em todos os pontos em que $g(t)$ é contínua; em pontos em que $g(t)$ é descontínua, a integral converge para o valor médio dos limites de $g(t)$ à direita e à esquerda do ponto de descontinuidade.

lizados em pontos uniformemente espaçados $x_1, x_2, ..., x_n$, como mostrado na Fig. 3.5a. A carga total W_T sobre a viga é dada pela soma das cargas localizadas nestes n pontos:

$$W_T = \sum_{i=1}^{n} D_i$$

Consideremos o caso de uma viga submetida a uma carga contínua, como ilustrado na Fig. 3.5b. Nesse caso, embora pareça existir uma carga em cada ponto, a carga em cada ponto é zero. Isso não significa que não exista carga sobre a viga. Nesse caso, uma medida adequada do carregamento não é a carga em cada ponto, mas a densidade de carga por unidade de comprimento no ponto. Seja $G(x)$ a densidade de carga por unidade de comprimento da viga. Isso significa que a carga sobre um comprimento de viga Δx ($\Delta x \to 0$), em algum ponto x, é $G(x)\Delta x$. Para determinar a carga total sobre a viga, dividimos a viga em segmentos de comprimento Δx ($\Delta x \to 0$). A carga sobre o n-ésimo segmento de comprimento Δx é $[G(n\Delta x)]\Delta x$. A carga total W_T é, então, dada por

$$W_T = \lim_{\Delta x \to 0} \sum_{x_1}^{x_n} G(n\Delta x)\,\Delta x$$

$$= \int_{x_1}^{x_n} G(x)\,dx$$

No caso de carregamento discreto (Fig. 3.5a), a carga total existe apenas nos n pontos discretos. Nos outros pontos, não há carga. No caso de carregamento contínuo, no entanto, a carga existe em cada ponto, mas, em qualquer ponto específico x, a carga é zero. Contudo, a carga em um pequeno comprimento de viga Δx é $[G(n\Delta x)]\Delta x$ (Fig. 3.5b). Dessa forma, embora a carga em um ponto x seja zero, a carga relativa neste ponto é $G(x)$.

O espectro de um sinal é completamente análogo ao carregamento de uma viga. Quando o sinal $g(t)$ é periódico, o espectro é discreto e $g(t)$ pode ser expresso como a soma de exponenciais discretas e de amplitudes finitas:

$$g(t) = \sum_{n} D_n e^{j2\pi n f_0 t}$$

No caso de um sinal aperiódico, o espectro se torna contínuo; ou seja, o espectro existe para todos os valores de f, mas a amplitude de cada componente espectral é zero. Aqui, a medida adequada não é a amplitude de uma componente em uma dada frequência, mas a densidade espectral por unidade de largura de banda. Da Eq. (3.7b), fica claro que $g(t)$ é sintetizada por meio da soma de exponenciais da forma $e^{j2\pi n\Delta f t}$, onde a contribuição de qualquer uma das componentes exponenciais é zero. Contudo, a contribuição de exponenciais em uma largura de banda infinitesimal Δf localizada em $f = n\Delta f$ é $G(n\Delta f)\Delta f$, e a adição de todas estas componentes produz a forma integral de $g(t)$:

$$g(t) = \lim_{\Delta f \to 0} \sum_{n=-\infty}^{\infty} G(n\Delta f) e^{(jn2\pi f)t}\,\Delta f = \int_{-\infty}^{\infty} G(f) e^{j2\pi f t}\,df$$

A contribuição das componentes na largura de banda df é $G(f)df$, em que df é a largura de banda em hertz. Fica claro que $G(f)$ é a **densidade espectral** por unidade de largura de banda (em hertz).

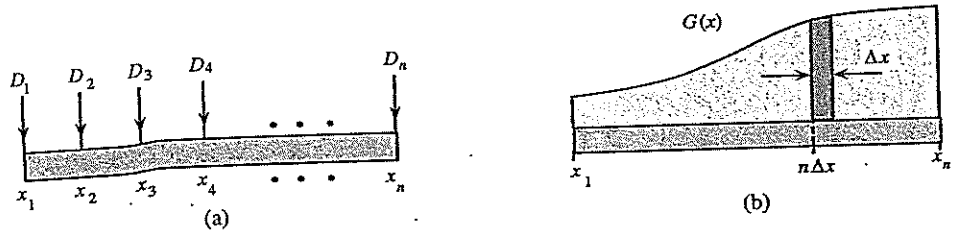


Figura 3.5 Analogia com a transformada de Fourier.

Isso também significa que, mesmo que a amplitude de uma componente qualquer seja zero, a amplitude relativa de uma componente de frequência f é $G(f)$. Embora $G(f)$ seja uma densidade espectral, na prática é costume denominá-la **espectro** de $g(t)$, em vez de densidade espectral de $g(t)$. Seguindo essa convenção, chamaremos $G(f)$ de espectro de Fourier (ou transformada de Fourier) de $g(t)$.

3.2 TRANSFORMADAS DE ALGUMAS FUNÇÕES ÚTEIS

Por conveniência, a seguir, introduziremos uma notação compacta para algumas funções úteis, como as funções retangular, triangular e de interpolação.

Função Retangular Unitária

Usamos a notação $\Pi(x)$ para representar um pulso retangular de altura e largura unitárias, centrado na origem, como ilustrado na Fig. 3.6a:

$$\Pi(x) = \begin{cases} 1 & |x| \leq \frac{1}{2} \\ 0{,}5 & |x| = \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases} \tag{3.16}$$

Vale notar que o pulso retangular na Fig. 3.6b é o pulso retangular unitário $\Pi(x)$ expandido por um fator τ e, portanto, pode ser expresso como $\Pi(x/\tau)$. Observemos, ainda, que o denominador τ em $\Pi(x/\tau)$ indica a largura do pulso.

Função Triangular Unitária

Usamos a notação $\Delta(x)$ para representar um pulso triangular de altura e largura unitárias, centrado na origem, como ilustrado na Fig. 3.7a:

$$\Delta(x) = \begin{cases} 1 - 2|t| & |x| < \frac{1}{2} \\ 0 & |x| > \frac{1}{2} \end{cases} \tag{3.17}$$

Vale notar que o pulso na Fig. 3.7b é o pulso $\Delta(x/\tau)$. Observemos que, como no caso do pulso retangular, o denominador τ em $\Delta(x/\tau)$ indica a largura do pulso.

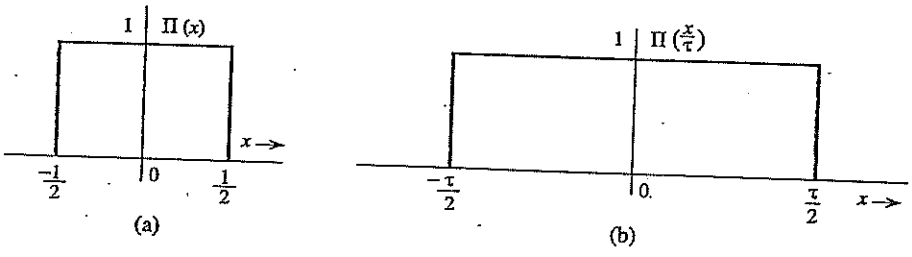


Figura 3.6 Pulso retangular.

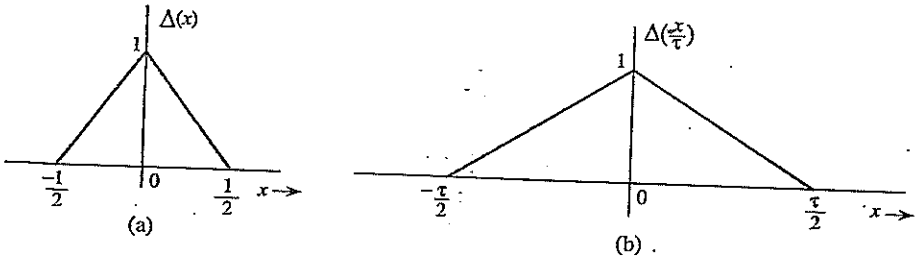


Figura 3.7 Pulso triangular.

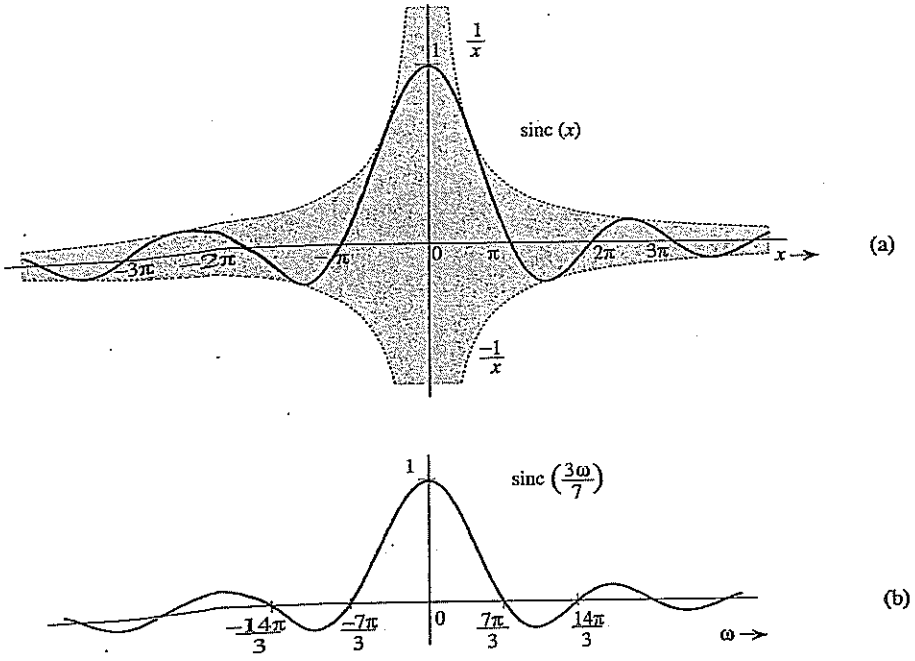


Figura 3.8
Pulso sinc.

Função Sinc sinc(x)

A função sen x/x é a função "seno dividido pelo argumento", denotada por sinc (x).*

Essa função tem um papel importante no processamento de sinais. Definimos

$$\text{sinc}\,(x) = \frac{\text{sen}\,x}{x} \tag{3.18}$$

Uma inspeção da Eq. (3.18) mostra que

1. sinc (x) é uma função par de x.
2. sinc (x) = 0 quando sen $x = 0$, exceto em $x = 0$, em que a função é indeterminada. Isso significa que sinc (x) = 0 para $t = \pm\,\pi, \pm\,2\pi, \pm\,3\pi, \ldots$
3. Usando a regra de L'Hôpital, determinamos sinc (0) = 1.
4. sinc (x) é o produto de um sinal oscilatório sen x (de período 2π) por uma função monótona decrescente $1/x$. Portanto, sinc (x) exibe oscilações senoidais de período 2π, cuja amplitude decresce continuamente na forma $1/x$.
5. Resumindo, sinc (x) é uma função oscilatória par, de amplitude decrescente. A função tem um pico unitário em $x = 0$ e cruza o eixo x em múltiplos inteiros de π.

A Fig. 3.8a mostra um gráfico de sinc (x). Observemos que sinc (x) = 0 para valores de x múltiplos inteiros positivos e negativos de π. A Fig. 3.8b mostra o gráfico de sinc ($3\omega/7$). O argumento $3\omega/7 = \pi$ quando $\omega = 7\pi/3$ ou $f = 7/6$. Portanto, o primeiro zero dessa função ocorre em $\omega = 7\pi/3$ ($f = 7/6$).

*Na literatura, sinc (x) também é denotada por Sa (x). Alguns autores definem sinc (x) como

$$\text{sinc}\,(x) = \frac{\text{sen}\,\pi x}{\pi x}$$

Exemplo 3.2 Determinemos a transformada de Fourier de $g(t) = \Pi(t/\tau)$ (Fig. 3.9a).

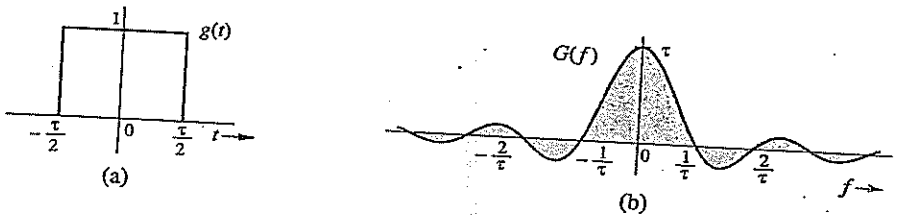


Figura 3.9 Pulso retangular e correspondente espectro de Fourier.

Temos

$$G(f) = \int_{-\infty}^{\infty} \Pi\left(\frac{t}{\tau}\right) e^{-j2\pi ft}\, dt$$

Como $\Pi(t/\tau) = 1$ para $|t| < \tau/2$ e é zero para $|t| > \tau/2$,

$$\begin{aligned} G(f) &= \int_{-\tau/2}^{\tau/2} e^{-j2\pi ft}\, dt \\ &= \frac{1}{j2\pi f}(e^{-j\pi f\tau} - e^{j\pi f\tau}) = \frac{2\,\mathrm{sen}(\pi f\tau)}{2\pi f} \\ &= \tau\frac{\mathrm{sen}(\pi f\tau)}{(\pi f\tau)} = \tau \ \mathrm{sinc}\ (\pi f\tau). \end{aligned}$$

Logo,

$$\Pi\left(\frac{t}{\tau}\right) \Longleftrightarrow \tau\ \mathrm{sinc}\left(\frac{\omega\tau}{2}\right) = \tau\ \mathrm{sinc}\ (\pi f\tau) \tag{3.19}$$

Recordemos que sinc $(x) = 0$ quando $x = \pm n\pi$. Logo, sinc $(\omega\tau/2) = 0$ quando $\omega\tau/2 = \pm n\pi$; ou seja, quando $f = \pm n/\tau$ $(n = 1, 2, 3,...)$, como ilustrado na Fig. 3.9b. Observemos que, neste caso, $G(f)$ é real; portanto, a informação espectral está contida em apenas um gráfico de $G(f)$, mostrado na Fig. 3.9b.

Exemplo 3.3 Determinemos a transformada do sinal impulso unitário $\delta(t)$.

Usando a propriedade de amostragem da função impulso, obtemos

$$\mathcal{F}[\delta(t)] = \int_{-\infty}^{\infty} \delta(t) e^{-j2\pi ft}\, dt = e^{-j2\pi f\cdot 0} = 1 \tag{3.20a}$$

ou

$$\delta(t) \Longleftrightarrow 1 \tag{3.20b}$$

A Fig. 3.10 mostra $\delta(t)$ e seu espectro.

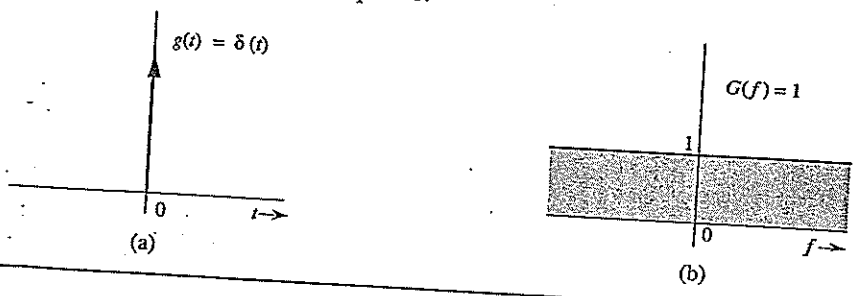


Figura 3.10 Impulso unitário e correspondente espectro de Fourier.

Exemplo 3.4 Determinemos a transformada de Fourier inversa de $\delta(2\pi f) = \frac{1}{2\pi}\cdot\delta(f)$.

A partir da Eq. (3.9b) e da propriedade de amostragem da função impulso,

$$\mathcal{F}^{-1}[\delta(2\pi f)] = \int_{-\infty}^{\infty} \delta(2\pi f)e^{j2\pi ft}\,df = \frac{1}{2\pi}\int_{-\infty}^{\infty} \delta(2\pi f)e^{j2\pi ft}\,d(2\pi f)$$
$$= \frac{1}{2\pi}\cdot e^{-j2\pi f\cdot 0} = \frac{1}{2\pi}$$

Portanto,

$$\frac{1}{2\pi} \Longleftrightarrow \delta(2\pi f) \tag{3.21a}$$

ou

$$1 \Longleftrightarrow \delta(f) \tag{3.21b}$$

Isso mostra que o espectro de um sinal constante $g(t) = 1$ é um impulso $\delta(f) = 2\pi\,\delta(2\pi f)$, como ilustrado na Fig. 3.11.

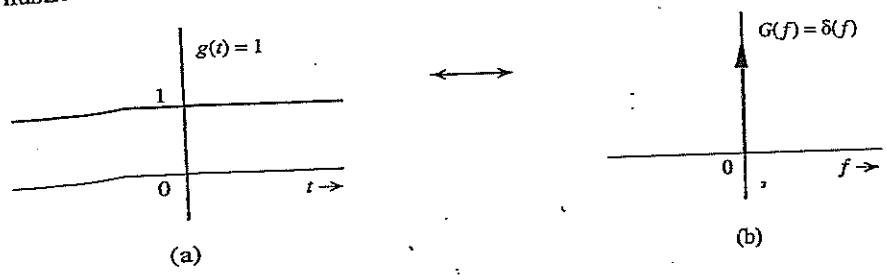


Figura 3.11 Sinal constante (dc) e correspondente espectro de Fourier.

Esse resultado [Eq. (3.21b)] também poderia ter sido obtido por uma análise qualitativa. Recordemos que a transformada de Fourier de $g(t)$ é uma representação espectral de $g(t)$ em termos de componentes exponenciais infinitas da forma $e^{j2\pi ft}$. Para representar um sinal constante $g(t) = 1$, precisamos apenas de uma exponencial infinita $e^{j2\pi ft}$ com $f = 0$. Isso resulta em um espectro em uma única frequência $f = 0$. Também poderíamos dizer que $g(t) = 1$ é um sinal dc que tem uma única componente de frequência em $f = 0$ (dc).

Se um impulso em $f = 0$ é o espectro de um sinal dc, o que representará um impulso em $f = f_0$? Responderemos a esta pergunta no próximo exemplo.

Exemplo 3.5 Determinemos a transformada de Fourier inversa de $\delta(f - f_0)$.

Usando a propriedade de amostragem da função impulso, obtemos

$$\mathcal{F}^{-1}[\delta(f - f_0)] = \int_{-\infty}^{\infty} \delta(f - f_0)e^{j2\pi ft}\,df = e^{j2\pi f_0 t}$$

Portanto,

$$e^{j2\pi f_0 t} \Longleftrightarrow \delta(f - f_0) \tag{3.22a}$$

Esse resultado mostra que o espectro de uma exponencial eterna $e^{j2\pi f_0 t}$ é um único impulso em $f = f_0$. Também chegaríamos a essa conclusão por uma análise qualitativa. Para representar a exponencial infinita $e^{j2\pi f_0 t}$, precisamos de uma única exponencial infinita $e^{j2\pi ft}$ com $\omega = 2\pi f_0$. Portanto, o espectro consiste em uma única componente na frequência $f = f_0$.

Da Eq. (3.22a), temos

$$e^{-j2\pi f_0 t} \Longleftrightarrow \delta(f + f_0) \tag{3.22b}$$

Exemplo 3.6 Determinemos a transformada de Fourier da senoide eterna $\cos 2\pi f_0 t$.

Recordemos a fórmula de Euler

$$\cos 2\pi f_0 t = \frac{1}{2}(e^{j2\pi f_0 t} + e^{-j2\pi f_0 t})$$

Somando as Eqs. (3.22a) e (3.22b) e usando a fórmula anterior, obtemos

$$\cos 2\pi f_0 t \Longleftrightarrow \frac{1}{2}[\delta(f + f_0) + \delta(f - f_0)] \tag{3.23}$$

O espectro de $\cos 2\pi f_0 t$ consiste em dois impulsos, em f_0 e $-f_0$, no domínio da frequência, ou dois impulsos em $\pm\,\omega_0 = \pm 2\pi f_0$, no domínio ω, como ilustrado na Fig. 3.12. Esse resultado também poderia ter sido obtido por análise qualitativa. Uma senoide infinita $\cos \omega_0 t$ pode ser sintetizada por duas exponenciais infinitas, $e^{j\omega_0 t}$ e $e^{-j\omega_0 t}$. Portanto, o espectro consiste em apenas duas componentes de frequência, em ω_0 e $-\omega_0$.

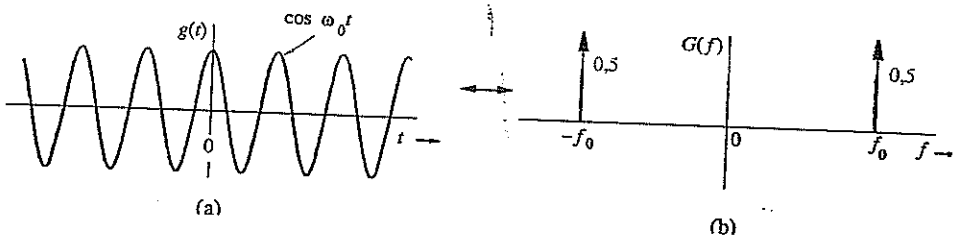


Figura 3.12 Sinal cosseno e correspondentes espectro de Fourier.

Exemplo 3.7 Determinemos a transformada de Fourier da função sgn(t) (pronunciada signum de t), mostrada na Fig. 3.13. O valor da função é +1 ou −1, dependendo se t é positivo ou negativo:

$$\operatorname{sgn}(t) = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{cases} \tag{3.24}$$

Não podemos usar integração para determinar a transformada de sgn(t) diretamente, pois sgn(t) viola a condição de Dirichlet [ver Eq. (3.14) e a nota de rodapé associada]. Para sermos específicos, sgn(t) não é absolutamente integrável. Contudo, a transformada pode ser obtida considerando sgn(t) como a soma de duas exponenciais, como mostrado na Fig. 3.13, no limite quando $a \to 0$:

$$\operatorname{sgn} t = \lim_{a \to 0} \left[e^{-at}u(t) - e^{at}u(-t)\right]$$

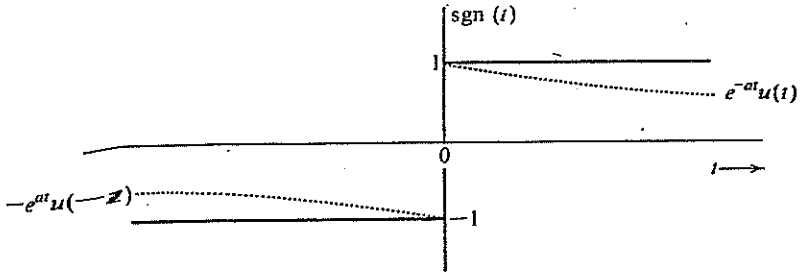


Figura 3.13
Função signum.

Portanto,

$$\mathcal{F}[\text{sgn}(t)] = \lim_{a\to 0}\left\{\mathcal{F}[e^{-at}u(t)] - \mathcal{F}[e^{at}u(-t)]\right\}$$

$$= \lim_{a\to 0}\left(\frac{1}{a+j2\pi f} - \frac{1}{a-j2\pi f}\right) \quad \text{(ver pares 1 e 2 na Tabela 3.1)}$$

$$= \lim_{a\to 0}\left(\frac{-j4\pi f}{a^2+4\pi^2 f^2}\right) = \frac{1}{j\pi f} \tag{3.25}$$

3.3 ALGUMAS PROPRIEDADES DA TRANSFORMADA DE FOURIER

Agora, estudaremos algumas propriedades importantes da transformada de Fourier, suas implicações e aplicações. Antes de embarcarmos neste estudo, devemos ressaltar um aspecto essencial da transformada de Fourier: a **dualidade tempo-frequência**.

3.3.1 Dualidade Tempo-Frequência

As Eqs. (3.9) mostram um fato interessante: a notável similaridade entre as operações de transformação direta e inversa. Essas operações, necessárias para passar de $g(t)$ a $G(f)$ e de $G(f)$ a $g(t)$, são ilustradas graficamente na Fig. 3.14. A única pequena diferença entre essas duas operações reside nos sinais opostos usados nos argumentos das exponenciais.

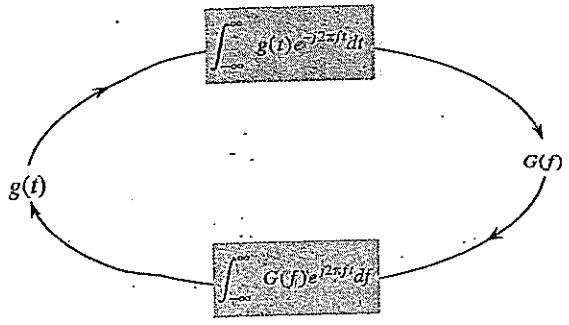


Figura 3.14
Quase simetria entre transformadas de Fourier direta e inversa.

Tabela 3.1
Pequena Tabela de Transformadas de Fourier

	$g(t)$	$G(f)$	
1	$e^{-at}u(t)$	$\dfrac{1}{a+j2\pi f}$	$a>0$
2	$e^{at}u(-t)$	$\dfrac{1}{a-j2\pi f}$	$a>0$
3	$e^{-a\|t\|}$	$\dfrac{2a}{a^2+(2\pi f)^2}$	$a>0$
4	$te^{-at}u(t)$	$\dfrac{1}{(a+j2\pi f)^2}$	$a>0$
5	$t^n e^{-at}u(t)$	$\dfrac{n!}{(a+j2\pi f)^{n+1}}$	$a>0$
6	$\delta(t)$	1	
7	1	$\delta(f)$	
8	$e^{j2\pi f_0 t}$	$\delta(f-f_0)$	
9	$\cos 2\pi f_0 t$	$0{,}5\,[\delta(f+f_0)+\delta(f-f_0)]$	
10	$\operatorname{sen} 2\pi f_0 t$	$j0{,}5\,[\delta(f+f_0)-\delta(f-f_0)]$	
11	$u(t)$	$\frac{1}{2}\delta(f)+\dfrac{1}{j2\pi f}$	
12	$\operatorname{sgn} t$	$\dfrac{2}{j2\pi f}$	
13	$\cos 2\pi f_0 t\, u(t)$	$\dfrac{1}{4}[\delta(f-f_0)+\delta(f+f_0)]+\dfrac{j2\pi f}{(2\pi f_0)^2-(2\pi f)^2}$	
14	$\operatorname{sen} 2\pi f_0 t\, u(t)$	$\dfrac{1}{4j}[\delta(f-f_0)-\delta(f+f_0)]+\dfrac{2\pi f_0}{(2\pi f_0)^2-(2\pi f)^2}$	
15	$e^{-at}\operatorname{sen} 2\pi f_0 t\, u(t)$	$\dfrac{2\pi f_0}{(a+j2\pi f)^2+4\pi^2 f_0^2}$	$a>0$
16	$e^{-at}\cos 2\pi f_0 t\, u(t)$	$\dfrac{a+j2\pi f}{(a+j2\pi f)^2+4\pi^2 f_0^2}$	$a>0$
17	$\Pi\left(\dfrac{t}{\tau}\right)$	$\tau\operatorname{sinc}(\pi f\tau)$	
18	$2B\operatorname{sinc}(2\pi Bt)$	$\Pi\left(\dfrac{f}{2B}\right)$	
19	$\Delta\left(\dfrac{t}{\tau}\right)$	$\dfrac{\tau}{2}\operatorname{sinc}^2\left(\dfrac{\pi f\tau}{2}\right)$	
20	$B\operatorname{sinc}^2(\pi Bt)$	$\Delta\left(\dfrac{f}{2B}\right)$	
21	$\sum_{n=-\infty}^{\infty}\delta(t-nT)$	$f_0\sum_{n=-\infty}^{\infty}\delta(f-nf_0)$	$f_0=\frac{1}{T}$
22	$e^{-t^2/2\sigma^2}$	$\sigma\sqrt{2\pi}e^{-2(\sigma\pi f)^2}$	

Essa similaridade tem amplas consequências no estudo de transformadas de Fourier e é a base da chamada dualidade entre tempo e frequência. *O princípio da dualidade pode ser comparado a uma fotografia e seu negativo. Uma fotografia pode ser obtida de seu negativo e, por meio de procedimento semelhante, o negativo pode ser obtido da fotografia.* Para qualquer resultado ou relação envolvendo $g(t)$ e $G(f)$, existe um resultado ou relação dual, obtido(a) por meio de troca dos papéis de $g(t)$ e $G(f)$ no resultado/relação original (juntamente com algumas pequenas modificações, devido ao fator 2π e a uma mudança de sinal). Por exemplo, a propriedade de translação temporal, que será provada posteriormente, afirma que, se $g(t) \Longleftrightarrow G(f)$, então

$$g(t-t_0) \Longleftrightarrow G(f)e^{-j2\pi f t_0}$$

O dual dessa propriedade (propriedade de translação em frequência) afirma que

$$g(t)e^{j2\pi f_0 t} \Longleftrightarrow G(f - f_0)$$

Observemos a troca de papéis entre tempo e frequência nestas duas equações (com a pequena diferença devido à mudança de sinal no argumento da exponencial). A utilidade desse princípio reside no fato de que, *sempre que deduzimos algum resultado, podemos ter certeza de que existe um dual do mesmo*. Essa informação propicia um maior entendimento de muitas propriedades ou resultados inesperados em processamento de sinais.

As propriedades da transformada de Fourier são úteis não apenas no cálculo de transformadas direta e inversa de diversas funções, mas também na obtenção de resultados valiosos em processamento de sinais. Nesta discussão, o leitor não deve deixar de observar a sempre presente dualidade entre tempo e frequência. Iniciamos com a propriedade de dualidade, que é uma das consequências do princípio de dualidade.

3.3.2 Propriedade de Dualidade

A propriedade de dualidade afirma que

$$g(t) \Longleftrightarrow G(f)$$

e

$$G(t) \Longleftrightarrow g(-f) \tag{3.26}$$

Pela propriedade de dualidade, se a transformada de Fourier de $g(t)$ for $G(f)$, a transformada de Fourier de $G(t)$, com f substituído por t, é $g(-f)$, que é o sinal original no domínio do tempo, com t substituído por $-f$.

Prova: A partir da Eq. (3.9b),

$$g(t) = \int_{-\infty}^{\infty} G(x)e^{j2\pi xt}\,dx$$

Logo,

$$g(-t) = \int_{-\infty}^{\infty} G(x)e^{-j2\pi xt}\,dx$$

A substituição de t por f resulta na Eq. (3.26). ■

Exemplo 3.8 Neste exemplo, aplicaremos a propriedade de dualidade [Eq. (3.26)] ao par de funções na Fig. 3.15a.

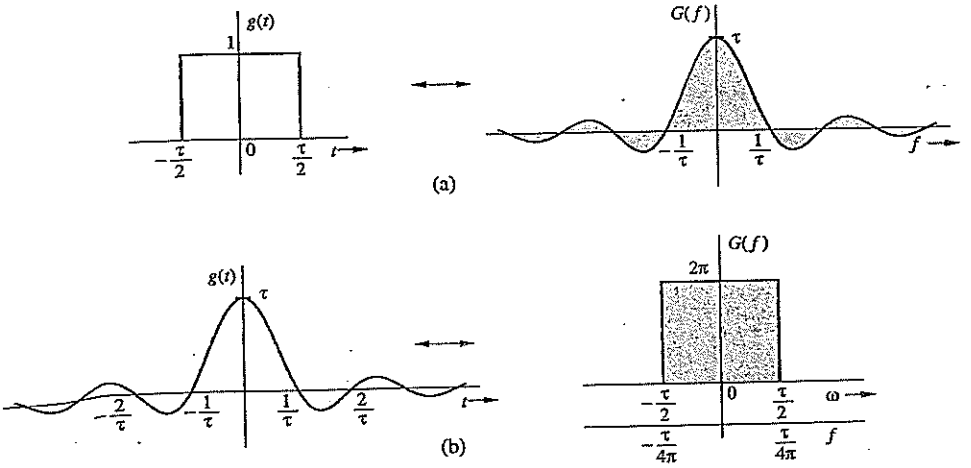


Figura 3.15 Propriedade de dualidade da transformada de Fourier.

Da Eq. (3.19), temos

$$\Pi\left(\frac{t}{\tau}\right) \Longleftrightarrow \tau \operatorname{sinc}(\pi f \tau) \tag{3.27a}$$

$$\underbrace{\Pi\left(\frac{t}{\alpha}\right)}_{g(t)} \Longleftrightarrow \underbrace{\alpha \operatorname{sinc}(\pi f \alpha)}_{G(f)} \tag{3.27b}$$

$G(t)$ tem a mesma forma de $G(f)$, apenas com f substituído por t; $g(-f)$ é $g(t)$ com t substituído por $-f$. Portanto, a propriedade de dualidade (3.26) fornece

$$\underbrace{\alpha \operatorname{sinc}(\pi \alpha t)}_{G(t)} \Longleftrightarrow \underbrace{\Pi\left(-\frac{f}{\alpha}\right)}_{g(-f)} = \Pi\left(\frac{f}{\alpha}\right) \tag{3.28a}$$

Substituindo $\tau = 2\pi\alpha$, obtemos

$$\tau \operatorname{sinc}\left(\frac{\alpha t}{2}\right) \Longleftrightarrow 2\pi\, \Pi\left(\frac{2\pi f}{\tau}\right) \tag{3.28b}$$

Na Eq. (3.8), usamos o fato de que $\Pi(-t) = \Pi(t)$, pois $\Pi(t)$ é uma função par de t. A Fig. 3.15b mostra gráficos deste par de funções. Observemos a troca de papéis entre t e $2\pi f$ (com o pequeno ajuste do fator 2π). Esse resultado aparece como o par 18 na Tabela 3.1 (com $\tau/2 = W$).

Como um exercício interessante, propomos a aplicação da propriedade de dualidade para gerar o dual de cada par na Tabela 3.1.

3.3.3 Propriedade de Dilatação no Tempo

Se

$$g(t) \Longleftrightarrow G(f)$$

então, para uma constante real positiva a,

$$g(at) \Longleftrightarrow \frac{1}{|a|} G\left(\frac{f}{a}\right) \tag{3.29}$$

Prova: Para uma constante real positiva a,

$$\mathcal{F}[g(at)] = \int_{-\infty}^{\infty} g(at) e^{-j2\pi f t}\, dt = \frac{1}{a} \int_{-\infty}^{\infty} g(x) e^{(-j2\pi f/a)x}\, dx = \frac{1}{a} G\left(\frac{f}{a}\right)$$

Do mesmo modo, pode ser mostrado que, se $a < 0$,

$$g(at) \Longleftrightarrow \frac{-1}{a} G\left(\frac{f}{a}\right)$$

Disto resulta a Eq. (3.29). ■

Importância da Propriedade de Dilatação no Tempo

A função $g(at)$ representa a função $g(t)$ comprimida no tempo por um fator a ($|a| > 1$). Da mesma forma, uma função $G(f/a)$ representa a função $G(f)$ expandida em frequência pelo mesmo fator a. A

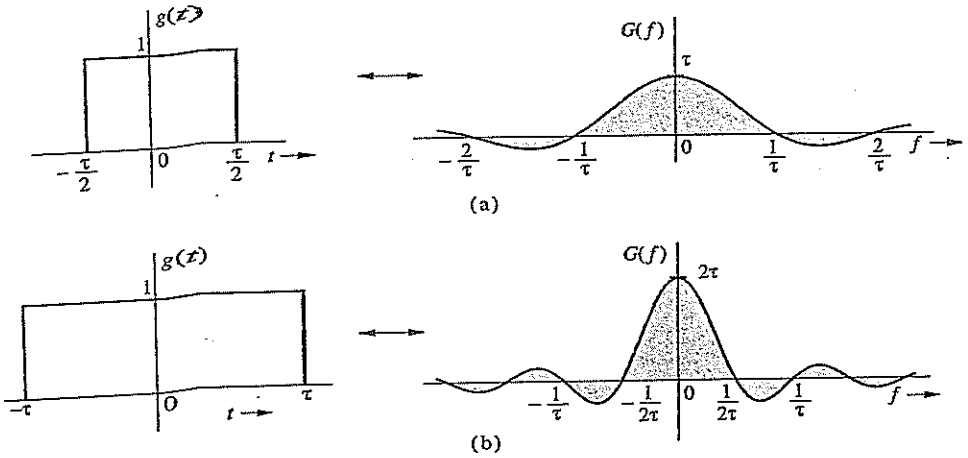


Figura 3.16 Propriedade de dilatação da transformada de Fourier.

propriedade de dilatação temporal afirma que a compressão temporal de um sinal resulta em sua expansão espectral, e que a expansão temporal de um sinal resulta em sua compressão espectral. De modo intuitivo, podemos perceber que a compressão temporal de um sinal por um fator a significa que o sinal varia **mais rapidamente**, sendo a taxa de variação aumentada pelo mesmo fator. Para sintetizar esse sinal, as frequências de suas componentes senoidais devem ser aumentadas pelo fator a, implicando que o espectro de frequências do sinal é expandido pelo fator a. Outrossim, a expansão temporal de um sinal o faz variar mais lentamente; ou seja, as frequências de suas componentes são reduzidas, implicando a compressão do espectro de frequências. Por exemplo, o sinal $\cos 4\pi f_0 t$ é o sinal $\cos 2\pi f_0 t$ é comprimido no tempo por um fator 2. Fica claro que o espectro do primeiro (impulsos em $\pm 2f_0$) é uma versão expandida do espectro do último (impulsos em $\pm f_0$). O efeito da dilatação temporal é ilustrado na Fig. 3.16.

Reciprocidade entre Duração de um Sinal e Sua Largura de Banda

A propriedade de dilatação temporal implica que, se a largura de $g(t)$ aumentar, seu espectro se torna mais estreito, e vice-versa. Dobrar a duração de um sinal significa dividir sua largura espectral por dois, e vice-versa. Isso sugere que a largura de banda de um sinal é inversamente proporcional à duração ou largura temporal (em segundos). Já comprovamos tal fato para o pulso retangular, quando determinamos que a largura de banda de um pulso de duração τ segundos era $1/\tau$ Hz. Uma discussão mais aprofundada sobre este interessante tema pode ser encontrada na literatura.[2]

Exemplo 3.9 Mostremos que

$$g(-t) \Longleftrightarrow G(-f) \tag{3.30}$$

A seguir, usemos esse resultado e o fato de que $e^{-at}u(t) \Longleftrightarrow 1/(a + j2\pi f)$ para calcular as transformadas de Fourier de $e^{-at}u(-t)$ e $e^{-a|t|}$.

A Eq. (3.30) resulta da Eq. (3.29) com $a = -1$. Aplicação da Eq. (3.30) ao par 1 da Tabela 3.1 fornece

$$e^{at}u(-t) \Longleftrightarrow \frac{1}{a - j2\pi f}$$

e

$$e^{-a|t|} = e^{-at}u(t) + e^{at}u(-t)$$

Logo,

$$e^{-a|t|} \Longleftrightarrow \frac{1}{a+j2\pi f} + \frac{1}{a-j2\pi f} = \frac{2a}{a^2+(2\pi f)^2} \tag{3.31}$$

Gráficos do sinal $e^{-a|t|}$ e de seu espectro são mostrados na Fig. 3.17.

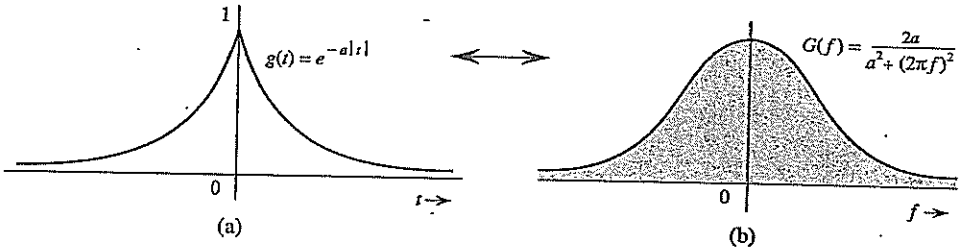


Figura 3.17 $e^{-a|t|}$ e correspondente espectro de Fourier.

3.3.4 Propriedade de Translação no Tempo

Se

$$g(t) \Longleftrightarrow G(f)$$

então

$$g(t-t_0) \Longleftrightarrow G(f)e^{-j2\pi f t_0} \tag{3.32a}$$

Prova: Por definição,

$$\mathcal{F}[g(t-t_0)] = \int_{-\infty}^{\infty} g(t-t_0)e^{-j2\pi f t}\,dt$$

Fazendo $t - t_0 = x$, obtemos

$$\begin{aligned}\mathcal{F}[g(t-t_0)] &= \int_{-\infty}^{\infty} g(x)e^{-j2\pi f(x+t_0)}\,dx \\ &= e^{-j2\pi f t_0}\int_{-\infty}^{\infty} g(x)e^{-j2\pi f x}\,dx = G(f)e^{-j2\pi f t_0}\end{aligned} \tag{3.32b}$$

Esse resultado mostra que *atrasar um sinal em t_0 segundos não altera seu espectro de amplitude. O espectro de fase, contudo, sofre uma alteração de $-2\pi f t_0$.*

Interpretação Física da Fase Linear

O atraso temporal de um sinal provoca um deslocamento de fase linear no espectro do mesmo. Esse resultado também pode ser obtido por meio de uma análise heurística. Imaginemos que $g(t)$ seja sintetizada por suas componentes de Fourier, que são senoides de certas amplitudes e fases. O sinal atrasado $g(t-t_0)$ pode ser sintetizado pelas mesmas componentes senoidais, cada uma atrasada em t_0 segundos. As amplitudes das componentes não são alteradas; portanto, o espectro de amplitude de $g(t-t_0)$ é idêntico ao de $g(t)$. Contudo, o atraso temporal t_0 aplicado a cada senoide altera a fase de cada componente. Uma senoide $\cos 2\pi f t$ atrasada de t_0 é dada por

$$\cos 2\pi f(t-t_0) = \cos(2\pi f t - 2\pi f t_0)$$

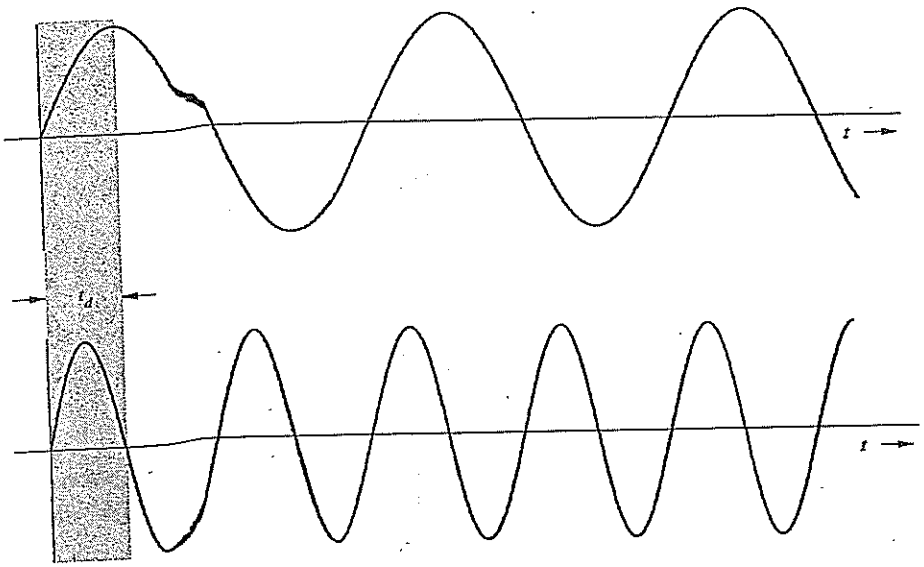


Figura 3.18 Interpretação física da propriedade de translação no tempo.

Isso significa que um atraso temporal t_0 em uma senoide de frequência f se manifesta como um atraso de fase $2\pi f t_0$, que é uma função linear de f. Isso implica que, para um dado atraso temporal, as componentes de frequências mais elevadas sofrerão deslocamentos de fase proporcionalmente mais altos. Esse efeito é ilustrado na Fig. 3.18 com duas senoides, sendo a frequência da senoide na parte inferior da figura o dobro da frequência da senoide na parte superior. O mesmo atraso temporal t_0 resulta em um deslocamento de fase de $\pi/2$ na senoide de cima e um deslocamento de fase de π na outra senoide. Isso comprova que, *para obter o mesmo atraso temporal, senoides de frequências mais elevadas devem sofrer deslocamentos de fase proporcionalmente mais altos.*

Exemplo 3.10 Determinemos a transformada de Fourier de $e^{-a|t-t_0|}$.

Essa função, cujo gráfico é mostrado na Fig. 3.19a, é uma versão de $e^{-a|t|}$ (cujo gráfico é mostrado na Fig. 3.17a) deslocada no tempo. Das Eqs. (3.31) e (3.32), temos

$$e^{-a|t-t_0|} \Longleftrightarrow \frac{2a}{a^2+(2\pi f)^2} e^{-j2\pi f t_0} \tag{3.33}$$

O espectro de $e^{-a|t-t_0|}$ (Fig. 3.19b) é igual ao de $e^{-a|t|}$ (Fig. 3.17b), a menos do deslocamento de fase de $-2\pi f t_0$.

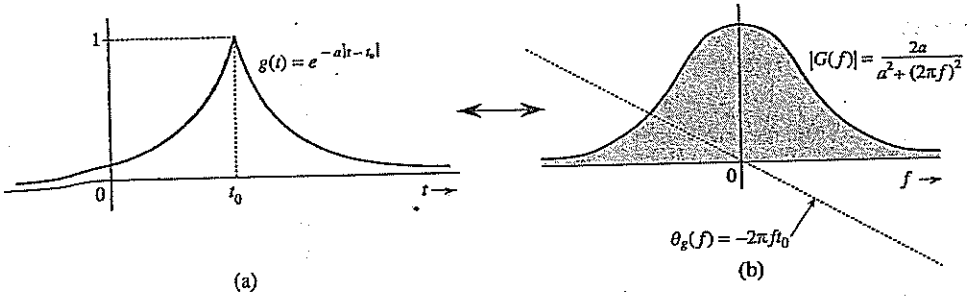


Figura 3.19 Efeito do deslocamento temporal no espectro de Fourier de um sinal.

Vale observar que o atraso temporal t_0 provoca um deslocamento de fase **linear** $-2\pi f t_0$ no espectro. Esse exemplo demonstra, de maneira clara, o efeito de um deslocamento temporal.

3.3.5 Propriedade de Translação na Frequência

Se

$$g(t) \Longleftrightarrow G(f)$$

então,

$$g(t)e^{j2\pi f_0 t} \Longleftrightarrow G(f - f_0) \tag{3.34}$$

Essa propriedade também é conhecida como propriedade de modulação.

Prova: Por definição,

$$\mathcal{F}[g(t)e^{j2\pi f_0 t}] = \int_{-\infty}^{\infty} g(t)e^{j2\pi f_0 t}e^{-j2\pi f t}\,dt = \int_{-\infty}^{\infty} g(t)e^{-j(2\pi f - 2\pi f_0)t}\,dt = G(f - f_0)$$

■

Essa propriedade afirma que a multiplicação de um sinal por um fator $e^{j2\pi f_0 t}$ desloca o espectro do sinal por $f = f_0$. Vale notar a dualidade entre as propriedades de translação no tempo e na frequência.

Substituindo f_0 por $-f_0$ na Eq. (3.34), obtemos

$$g(t)e^{-j2\pi f_0 t} \Longleftrightarrow G(f + f_0) \tag{3.35}$$

Como $e^{j2\pi f_0 t}$ não é uma função real que possa ser gerada, o deslocamento em frequência é obtido, na prática, pela multiplicação de $g(t)$ por uma senoide. Isso pode ser visto de

$$g(t)\cos 2\pi f_0 t = \frac{1}{2}\left[g(t)e^{j2\pi f_0 t} + g(t)e^{-j2\pi f_0 t}\right]$$

Das Eqs. (3.34) e (3.35), temos

$$g(t)\cos 2\pi f_0 t \Longleftrightarrow \frac{1}{2}[G(f - f_0) + G(f + f_0)] \tag{3.36}$$

Esse resultado mostra que a multiplicação de um sinal $g(t)$ por uma senoide de frequência f_0 desloca o espectro $G(f)$ por $\pm f_0$. Multiplicar $g(t)$ por uma senoide $\cos 2\pi f_0 t$ corresponde a modular a amplitude da senoide. Esse tipo de modulação é conhecido como **modulação em amplitude**. A senoide $\cos 2\pi f_0 t$ é denominada **portadora**; o sinal $g(t)$ é o **sinal modulante** e o sinal $g(t)\cos 2\pi f_0 t$, o **sinal modulado**. Modulação e demodulação serão discutidas nos Capítulos 4 e 5.

Para esboçar o gráfico do sinal $g(t)\cos 2\pi f_0 t$, observemos que

$$g(t)\cos 2\pi f_0 t = \begin{cases} g(t) & \text{quando } \cos 2\pi f_0 t = 1 \\ -g(t) & \text{quando } \cos 2\pi f_0 t = -1 \end{cases}$$

Portanto, $g(t)\cos 2\pi f_0 t$ toca $g(t)$ quando a senoide $\cos 2\pi f_0 t$ passa por seus picos positivos, e toca $-g(t)$ quando $\cos 2\pi f_0 t$ passa por seus picos negativos. Isso significa que $g(t)$ e $-g(t)$ atuam como envelopes para o sinal $g(t)\cos 2\pi f_0 t$ (ver Fig. 3.20c). O sinal $-g(t)$ é o reflexo de $g(t)$ em relação ao eixo horizontal. A Fig. 3.20 mostra gráficos dos sinais $g(t)$, $g(t)\cos 2\pi f_0 t$ e dos respectivos espectros.

Translação do Espectro de Fase de um Sinal Modulado

Podemos transladar a fase de cada componente espectral de um sinal modulado por um valor constante θ_0 simplesmente usando uma portadora $\cos(2\pi f_0 t + \theta_0)$ no lugar de $\cos(2\pi f_0 t)$. Se um sinal $g(t)$ for multiplicado por $\cos(2\pi f_0 t + \theta_0)$, podemos fazer uso de um argumento similar ao empregado na dedução da Eq. (3.36) e mostrar que

$$g(t)\cos(2\pi f_0 t + \theta_0) \Longleftrightarrow \frac{1}{2}\left[G(f - f_0)\,e^{j\theta_0} + G(f + f_0)\,e^{-j\theta_0}\right] \tag{3.37}$$

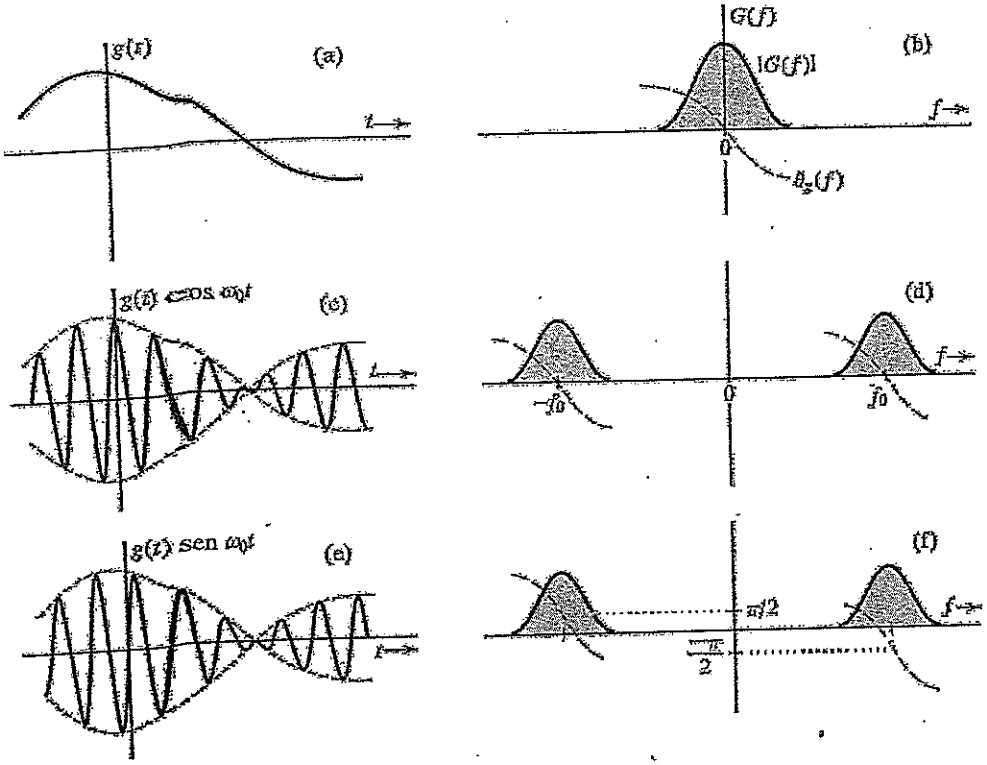


Figura 3.20 Modulação em amplitude de um sinal causa deslocamento espectral.

Para um caso especial em que $\theta_0 = -\pi/2$, a Eq. (3.37) se torna

$$g(t)\,\text{sen}\,2\pi f_0 t \Longleftrightarrow \frac{1}{2}\left[G(f-f_0)\,e^{-j\pi/2} + G(f+f_0)\,e^{j\pi/2}\right] \tag{3.38}$$

Vale notar que sen $2\pi f_0 t$ é cos $2\pi f_0 t$ com um atraso de fase de $\pi/2$. Portanto, deslocar a fase da portadora por $\pi/2$ implica deslocar a fase de cada componente espectral por $\pi/2$. As Fig. 3.20e e 3.20f mostram gráficos do sinal $g(t)$ sen $2\pi f_0 t$ e de seu espectro.

Modulação é uma aplicação comum que desloca o espectro de sinais. Em particular, se diversos sinais de mensagem, todos ocupando a mesma faixa de frequências, forem transmitidos simultaneamente ao longo de um meio de transmissão comum, interferirão mutuamente; será impossível separá-los ou recuperá-los no receptor. Por exemplo, se todas as estações de rádio decidissem transmitir sinais de áudio simultaneamente, os receptores não seriam capazes de separá-los. Esse problema é resolvido com o uso da modulação, e a cada estação de rádio é alocada uma frequência portadora distinta. Cada estação transmite um sinal modulado e, dessa forma, desloca o espectro do sinal à faixa de frequências a ela alocada, que não é ocupada por qualquer outra estação. Um receptor de rádio pode sintonizar uma estação qualquer selecionando a faixa de frequências da estação desejada. O receptor deve, então, demodular o sinal recebido (desfazer o efeito da modulação). Demodulação, portanto, consiste em outro deslocamento espectral, necessário para transladar o sinal à sua faixa original de frequências.

Sinais Passa-Faixa

A Fig. 3.20(d)(f) mostra que, se $g_c(t)$ e $g_s(t)$ forem sinais passa-baixos, cada um com largura de banda B Hz ou $2\pi B$ rad/s, os sinais $g_c(t)$ cos $2\pi f_0 t$ e $g_s(t)$ sen $2\pi f_0 t$ serão sinais passa-faixa que ocupam a mesma faixa de frequências, cada um com largura de banda $2B$ Hz. Portanto, uma combinação linear desses dois sinais também será um sinal passa-faixa que ocupará a mesma faixa de frequências

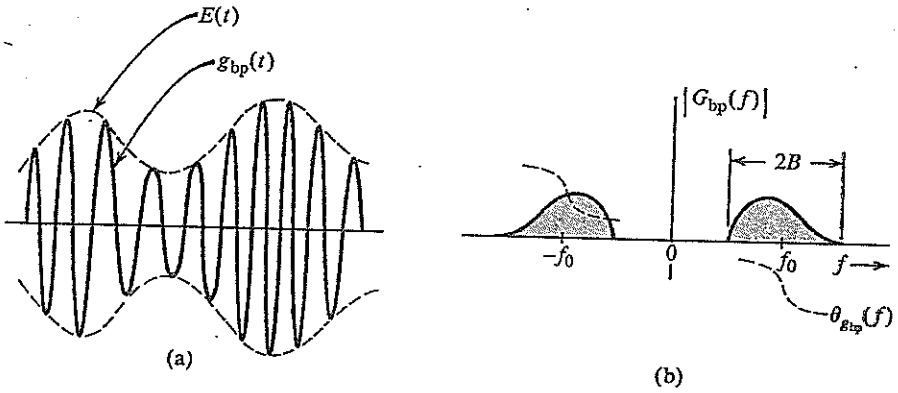


Figura 3.21 Sinal passa-faixa e correspondente espectro.

de cada um deles e terá a mesma largura de banda ($2B$ Hz). Assim, em geral, um sinal passa-faixa $g_{bp}(t)$ pode ser expresso como*

$$g_{bp}(t) = g_c(t)\cos 2\pi f_0 t + g_s(t)\,\text{sen}\, 2\pi f_0 t \tag{3.39}$$

O espectro de $g_{bp}(t)$ é centrado em $\pm f_0$ e tem largura de banda $2B$, como ilustrado na Fig. 3.21. Embora os espectros de magnitude de $g_c(t)\cos 2\pi f_0 t$ e de $g_s(t)\,\text{sen}\, 2\pi f_0 t$ sejam simétricos em relação a $\pm f_0$, o espectro de magnitude da soma desses sinais, $g_{bp}(t)$, não é necessariamente simétrico em relação a $\pm f_0$. Isto se deve ao fato de que as fases distintas dos dois sinais não permitem que suas amplitudes se somem diretamente, pois

$$a_1 e^{j\varphi_1} + a_2 e^{j\varphi_2} \neq (a_1 + a_2)e^{j(\varphi_1+\varphi_2)}$$

Um sinal passa-faixa típico $g_{bp}(t)$ e seu espectro são ilustrados na Fig. 3.21. Podemos usar uma identidade trigonométrica bem conhecida para expressar a Eq. (3.39) como

$$g_{bp}(t) = E(t)\cos[2\pi f_0 t + \psi(t)] \tag{3.40}$$

em que

$$E(t) = +\sqrt{g_c^2(t) + g_s^2(t)} \tag{3.41a}$$

$$\psi(t) = -\tan^{-1}\left[\frac{g_s(t)}{g_c(t)}\right] \tag{3.41b}$$

Como $g_c(t)$ e $g_s(t)$ são sinais passa-baixos, $E(t)$ e $\psi(t)$ também são sinais passa-baixos. Como $E(t)$ é não negativo [Eq. (3.41a)], a Eq. (3.40) indica que $E(t)$ é um envelope de variação lenta e que $\psi(t)$ é uma fase de variação lenta do sinal passa-faixa $g_{bp}(t)$, como ilustrado na Fig. 3.21. Assim, o sinal passa-faixa $g_{bp}(t)$ será uma senoide cuja amplitude varia lentamente. Devido à fase variante no tempo $\psi(t)$, a frequência da senoide também variará lentamente** com o tempo, em torno da frequência central f_0.

*Ver a Seção 9.9 para uma prova rigorosa dessa afirmação.

**Para um envelope bem definido, é necessário $B \ll f_0$; caso contrário, as variações de $E(t)$ serão da mesma ordem que às da portadora e será difícil separar o envelope da portadora.

Exemplo 3.11 Determinemos a transformada de Fourier de um sinal periódico $g(t)$ de período T_0 e, a seguir, determinemos a transformada de Fourier do trem periódico de impulsos $\delta_{T_0}(t)$ mostrado na Fig. 3.22a.

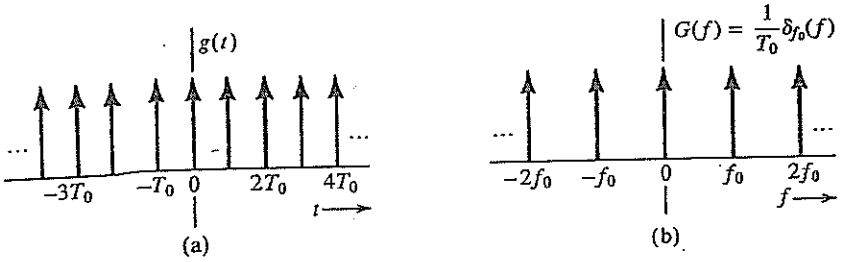


Figura 3.22 Trem de impulsos e correspondente espectro.

Um sinal periódico $g(t)$ pode ser expresso por uma série de Fourier exponencial, da seguinte forma

$$g(t) = \sum_{n=-\infty}^{\infty} D_n e^{jn2\pi f_0 t} \qquad f_0 = \frac{1}{T_0}$$

Logo,

$$g(t) \Longleftrightarrow \sum_{n=-\infty}^{\infty} \mathcal{F}[D_n e^{jn2\pi f_0 t}]$$

Da Eq. (3.22a), temos

$$g(t) \Longleftrightarrow \sum_{n=-\infty}^{\infty} D_n \delta(f - nf_0) \tag{3.42}$$

A Eq. (2.67) mostra que o trem de impulso $\delta_{T_0}(t)$ pode ser expresso por uma série de Fourier como

$$\delta_{T_0}(t) = \frac{1}{T_0} \sum_{n=-\infty}^{\infty} e^{jn2\pi f_0 t} \qquad f_0 = \frac{1}{T_0}$$

Aqui, $D_n = 1/T_0$. Portanto, da Eq. (3.42), temos

$$\begin{aligned} \delta_{T_0}(t) &\Longleftrightarrow \frac{1}{T_0} \sum_{n=-\infty}^{\infty} \delta(f - nf_0) \\ &= \frac{1}{T_0} \delta_{f_0}(f) \qquad f_0 = \frac{1}{T_0} \end{aligned} \tag{3.43}$$

Logo, o espectro de um trem de impulsos também é um trem de impulsos (no domínio da frequência), como ilustrado na Fig. 3.23b.

3.3.6 Teorema da Convolução

A convolução de duas funções $g(t)$ e $w(t)$, denotada por $g(t) * w(t)$, é definida pela integral

$$g(t) * w(t) = \int_{-\infty}^{\infty} g(\tau) w(t - \tau)\, d\tau$$

A propriedade de convolução no domínio do tempo e sua dual, a propriedade de convolução no domínio da frequência, afirmam que, se

$$g_1(t) \Longleftrightarrow G_1(f) \qquad \text{e} \qquad g_2(t) \Longleftrightarrow G_2(f)$$

então **(convolução no domínio do tempo)**

$$g_1(t) * g_2(t) \Longleftrightarrow G_1(f)G_2(f) \tag{3.44}$$

e **(convolução no domínio da frequência)**

$$g_1(t)g_2(t) \Longleftrightarrow G_1(f) * G_2(f) \tag{3.45}$$

Essas duas relações do teorema da convolução afirmam que a convolução de dois sinais no domínio do tempo se torna uma multiplicação no domínio da frequência, enquanto a multiplicação de dois sinais no domínio do tempo se torna a convolução no domínio da frequência.
Prova: Por definição

$$\mathcal{F}[g_1(t) * g_2(t)] = \int_{-\infty}^{\infty} e^{-j2\pi ft}\left[\int_{-\infty}^{\infty} g_1(\tau)g_2(t-\tau)d\tau\right]dt$$

$$= \int_{-\infty}^{\infty} g_1(\tau)\left[\int_{-\infty}^{\infty} e^{-j2\pi ft}g_2(t-\tau)dt\right]d\tau$$

A integral interna é a transformada de Fourier de $g_2(t-\tau)$, dada por $G_2(f)e^{-j2\pi f\tau}$ [propriedade de translação no tempo, Eq. (3.32a)]. Logo,

$$\mathcal{F}[g_1(t) * g_2(t)] = \int_{-\infty}^{\infty} g_1(\tau)e^{-j2\pi f\tau}G_2(f)\,d\tau$$

$$= G_2(f)\int_{-\infty}^{\infty} g_1(\tau)e^{-j2\pi f\tau}\,d\tau = G_1(f)G_2(f)$$

■

A propriedade de convolução no domínio da frequência [Eq. (3.45)] pode ser provada exatamente da mesma forma, simplesmente trocando os papéis de $g(t)$ e $G(f)$.

Largura de Banda do Produto de Dois Sinais

Se $g_1(t)$ e $g_2(t)$ tiverem larguras de banda B_1 e B_2 Hz, respectivamente, a largura de banda de $g_1(t)\,g_2(t)$ é $B_1 + B_2$ Hz. Esse resultado advém da aplicação da propriedade de largura de banda da convolução[3] à Eq. (3.45). Essa propriedade afirma que a largura de banda de $x * y$ é a soma das larguras de banda de x e de y. Logo, se a largura de banda de $g(t)$ for B Hz, a largura de banda de $g^2(t)$ será $2B$ Hz, e a largura de banda de $g^n(t)$, nB Hz.*

*A propriedade de largura de banda da convolução não se aplica a alguns casos patológicos. A propriedade falha quando a convolução de duas funções é zero em um intervalo, mesmo que as duas funções sejam não nulas [por exemplo, sen $2\pi f_0 t\ u(t) * u(t)$]. Tecnicamente, a propriedade é válida até mesmo nesse caso desde que, no cálculo da largura de banda da função que resulta da convolução, levemos em consideração o intervalo em que a convolução é zero.

Exemplo 3.12 Usando a propriedade de convolução no domínio do tempo, mostremos que, se

$$g(t) \Longleftrightarrow G(f)$$

então

$$\int_{-\infty}^{t} g(\tau)d\tau \Longleftrightarrow \frac{G(f)}{j2\pi f} + \frac{1}{2}G(0)\delta(f) \tag{3.46}$$

Como

$$u(t-\tau) = \begin{cases} 1 & \tau \leq t \\ 0 & \tau > t \end{cases}$$

temos

$$g(t) * u(t) = \int_{-\infty}^{\infty} g(\tau)u(t-\tau)\,d\tau = \int_{-\infty}^{t} g(\tau)\,d\tau$$

Agora, da propriedade de convolução no domínio do tempo [Eq. (3.44)], obtemos

$$\begin{aligned} g(t) * u(t) &\Longleftrightarrow G(f)U(f) \\ &= G(f)\left[\frac{1}{j2\pi f} + \frac{1}{2}\delta(f)\right] \\ &= \frac{G(f)}{j2\pi f} + \frac{1}{2}G(0)\delta(f) \end{aligned}$$

Na dedução desse último resultado, usamos o par 11 da Tabela 3.1 e a Eq. (2.10a).

3.3.7 Diferenciação e Integração no Domínio do Tempo

Se

$$g(t) \Longleftrightarrow G(f),$$

então (**diferenciação no domínio do tempo**)*

$$\frac{dg(t)}{dt} \Longleftrightarrow j2\pi f G(f) \tag{3.47}$$

e (**integração no domínio do tempo**)

$$\int_{-\infty}^{t} g(\tau)d\tau \Longleftrightarrow \frac{G(f)}{j2\pi f} + \frac{1}{2}G(0)\delta(f) \tag{3.48}$$

Prova: Diferenciando os dois lados da Eq. (3.9b), obtemos

$$\frac{dg(t)}{dt} = \int_{-\infty}^{\infty} j2\pi f G(f) e^{j2\pi ft}\,df$$

Isso mostra que

$$\frac{dg(t)}{dt} \Longleftrightarrow j2\pi f G(f)$$

*Válida apenas se a transformada de $dg(t)/dt$ existir.

Sucessivas aplicações desta propriedade levam a

$$\frac{d^n g(t)}{dt^n} \Longleftrightarrow (j2\pi f)^n G(f) \tag{3.49}$$

A propriedade de integração no domínio do tempo [Eq. (3.48)] já foi provada no Exemplo 3.12. ■

Exemplo 3.13 Usemos a propriedade de diferenciação no domínio do tempo para calcular a transformada de Fourier do pulso triangular $\Delta(t/\tau)$ mostrado na Fig. 3.23a.

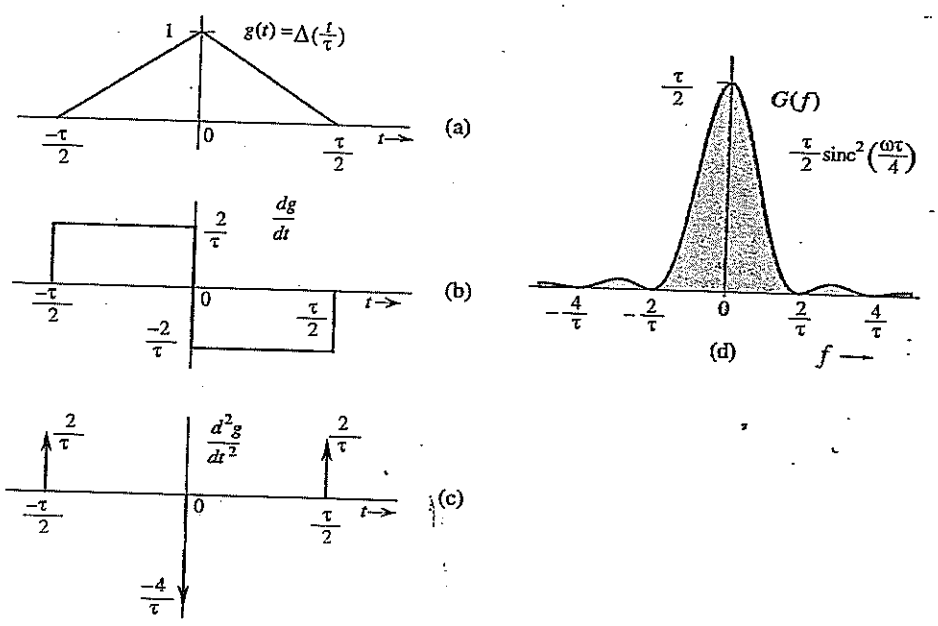


Figura 3.23 Uso da propriedade de diferenciação no domínio do tempo para calcular a transformada de Fourier de um sinal linear por partes.

Para calcular a transformada de Fourier deste pulso, o diferenciamos em relação ao tempo sucessivas vezes, como indicado na Fig. 3.23b e c. A derivada de segunda ordem consiste em uma sequência de impulsos (Fig. 3.23c). Recordemos que a derivada de um sinal em uma descontinuidade do tipo degrau é um impulso de amplitude igual à altura do degrau. A função $dg(t)/dt$ tem um degrau ou salto positivo $2/\tau$ em $t = \pm\,\tau/2$, e um degrau negativo $4/\tau$ em $t = 0$. Portanto,

$$\frac{d^2g(t)}{dt^2} = \frac{2}{\tau}\left[\delta\left(t+\frac{\tau}{2}\right) - 2\delta(t) + \delta\left(t-\frac{\tau}{2}\right)\right] \tag{3.50}$$

Da propriedade de diferenciação no domínio do tempo [Eq. (3.49)], temos

$$\frac{d^2g}{dt^2} \Longleftrightarrow (j2\pi f)^2 G(f) = -(2\pi f)^2 G(f) \tag{3.51a}$$

Adicionalmente, a propriedade de translação no tempo [Eqs. (3.32)] fornece

$$\delta(t - t_0) \Longleftrightarrow e^{-j2\pi f t_0} \tag{3.51b}$$

Tomando a transformada de Fourier da Eq. (3.50) e usando os resultados da Eq. (3.51), obtemos

$$(j2\pi f)^2 G(f) = \frac{2}{\tau}\left(e^{j\pi f\tau} - 2 + e^{-j\pi f\tau}\right) = \frac{4}{\tau}(\cos \pi f\tau - 1) = -\frac{8}{\tau}\text{sen}^2\left(\frac{\pi f\tau}{2}\right)$$

e

$$G(f) = \frac{8}{(2\pi f)^2 \tau} \operatorname{sen}^2\left(\frac{\pi f \tau}{2}\right) = \frac{\tau}{2}\left[\frac{\operatorname{sen}(\pi f \tau/2)}{\pi f \tau/2}\right]^2 = \frac{\tau}{2}\operatorname{sinc}^2\left(\frac{\pi f \tau}{2}\right) \tag{3.52}$$

O espectro $G(f)$ é ilustrado na Fig. 3.23d. Esse procedimento de cálculo da transformada de Fourier pode ser aplicado a qualquer função $g(t)$ que consista em segmentos de reta, com $g(t) \to 0$ quando $|t| \to \infty$. A derivada de segunda ordem de um sinal desse tipo é uma sequência de impulsos, cujas transformadas de Fourier podem ser determinadas por inspeção. Este exemplo sugere um método numérico para o cálculo da transformada de Fourier de um sinal arbitrário $g(t)$, em que o sinal é aproximado por segmentos de reta.

Para facilitar a consulta, diversas propriedades importantes da transformada de Fourier são resumidas na Tabela 3.2.

Tabela 3.2
Propriedades de Operações com Transformada de Fourier

Operação	$g(t)$	$G(f)$
Superposição	$g_1(t) + g_2(t)$	$G_1(f) + G_2(f)$
Multiplicação por escalar	$kg(t)$	$kG(f)$
Dualidade	$G(t)$	$g(-f)$
Dilatação no tempo	$g(at)$	$\frac{1}{\lvert a\rvert}G\left(\frac{f}{a}\right)$
Translação no tempo	$g(t - t_0)$	$G(f)e^{-j2\pi f t_0}$
Translação na frequência	$g(t)e^{j2\pi f_0 t}$	$G(f - f_0)$
Convolução no domínio do tempo	$g_1(t) * g_2(t)$	$G_1(f)G_2(f)$
Convolução no domínio da frequência	$g_1(t)g_2(t)$	$G_1(f) * G_2(f)$
Diferenciação no domínio do tempo	$\frac{d^n g(t)}{dt^n}$	$(j2\pi f)^n G(f)$
Integração no domínio do tempo	$\int_{-\infty}^{t} g(x)\,dx$	$\frac{G(f)}{j2\pi f} + \frac{1}{2}G(0)\delta(f)$

3.4 TRANSMISSÃO DE SINAL EM UM SISTEMA LINEAR

Um sistema linear invariante no tempo (LIT) em tempo contínuo pode ser igualmente bem caracterizado no domínio do tempo ou no domínio da frequência. O modelo de sistema LIT ilustrado na Fig. 3.24 pode, muitas vezes, ser usado para caracterizar canais de comunicação. Em sistemas de comunicação e no processamento de sinais, o interesse reside apenas em sistemas lineares estáveis com entrada limitada e saída limitada (sistemas BIBO — *bounded-input-bounded-output systems*). Uma discussão detalhada sobre estabilidade de sistemas pode ser encontrada em um livro de Lathi.[3]

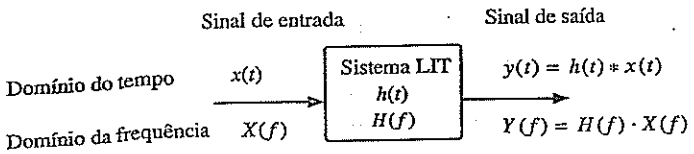


Figura 3.24 Transmissão de um sinal através de um sistema linear invariante no tempo.

Um sistema LIT estável pode ser caracterizado no domínio do tempo por sua resposta ao impulso $h(t)$, que é a resposta do sistema a um impulso unitário de entrada:

$$y(t) = h(t) \quad \text{quando} \quad x(t) = \delta(t)$$

A resposta do sistema a um sinal de entrada limitado $x(t)$ é obtida da relação de convolução

$$y(t) = h(t) * x(t) \tag{3.53}$$

No domínio da frequência, a relação entre os sinais de entrada e de saída é obtida da transformada de Fourier dos dois lados da Eq. (3.53). Sejam,

$$x(t) \Longleftrightarrow X(f)$$
$$y(t) \Longleftrightarrow Y(f)$$
$$h(t) \Longleftrightarrow H(f)$$

Aplicando o teorema da convolução, a Eq. (3.53) passa a

$$Y(f) = H(f) \cdot X(f) \tag{3.54}$$

Em geral, $H(f)$, a transformada de Fourier da resposta ao impulso $h(t)$, recebe a denominação **função de transferência** ou **resposta em frequência** do sistema LIT. É comum que $H(f)$ seja uma função de valores complexos, podendo ser escrita como

$$H(f) = |H(f)|e^{j\theta_h(f)}$$

em que $|H(f)|$ é a resposta de amplitude e $\theta_h(f)$, a resposta de fase do sistema LIT.

3.4.1 Distorção do Sinal Durante a Transmissão

A transmissão de um sinal de entrada $x(t)$ através de um sistema o transforma no sinal de saída $y(t)$. A Eq. (3.54) ilustra a natureza dessa transformação ou modificação. Aqui, $X(f)$ e $Y(f)$ são os espectros de entrada e de saída, respectivamente. Portanto, $H(f)$ é a resposta espectral do sistema. O espectro de saída é obtido multiplicando o espectro de entrada pela resposta espectral do sistema. A Eq. (3.54) ressalta a formatação espectral (ou modificação) do sinal de entrada pelo sistema. A Eq. (3.54) pode ser expressa na forma polar como

$$|Y(f)|e^{j\theta_y(f)} = |X(f)||H(f)|e^{j[\theta_x(f)+\theta_h(f)]}$$

Assim, obtemos as relações de amplitude e de fase

$$|Y(f)| = |X(f)|\,|H(f)| \tag{3.55a}$$
$$\theta_y(f) = \theta_x(f) + \theta_h(f) \tag{3.55b}$$

Durante a transmissão, o espectro de amplitude do sinal de entrada, $|X(f)|$, é alterado para $|X(f)| \times |H(f)|$. Do mesmo modo, o espectro de fase do sinal de entrada, $\theta_x(f)$ é alterado para $\theta_x(f) + \theta_h(f)$.

Uma componente espectral do sinal de entrada de frequência f é modificada em amplitude por um fator $|H(f)|$ e deslocada em fase por um ângulo $\theta_h(f)$. $|H(f)|$ é a resposta de amplitude e $\theta_h(f)$, a resposta de fase do sistema. Os gráficos de $|H(f)|$ e $\theta_h(f)$ em função de f mostram como o sistema modifica as amplitudes e fases das diversas entradas senoidais. Por isso, $H(f)$ é denominada **resposta em frequência** do sistema. Durante a transmissão através do sistema, algumas componentes de frequência podem ter a amplitude amplificada, enquanto outras podem ter a amplitude atenuada. As fases relativas das diversas componentes também são alteradas. Em geral, a forma de onda de saída será diferente da de entrada.

3.4.2 Transmissão sem Distorção

Em diversas aplicações, como amplificação de sinal ou transmissão de sinal de mensagem através de um canal de comunicação, é necessário que a forma de onda de saída seja uma réplica da forma de onda de entrada. Nesses casos, é necessário minimizar a distorção causada pelo amplificador ou canal de comunicação. Portanto, a determinação das características de um sistema que permita a passagem de um sinal sem distorção (**transmissão sem distorção**) é de interesse prático.

A transmissão é sem distorção se a entrada e a saída tiverem formas de onda idênticas, a menos de uma constante multiplicativa. Uma saída atrasada que mantenha a forma de onda de entrada também é considerada sem distorção. Assim, na transmissão sem distorção, a entrada $x(t)$ e a saída $y(t)$ satisfazem a condição

$$y(t) = k \cdot x(t - t_d) \tag{3.56}$$

A transformada de Fourier desta equação fornece

$$Y(f) = kX(f)e^{-j2\pi f t_d}$$

Contudo, como

$$Y(f) = X(f)H(f)$$

temos,

$$H(f) = k\,e^{-j2\pi f t_d}$$

Esta é a função de transferência necessária para transmissão sem distorção. Desta equação, obtemos

$$|H(f)| = k \tag{3.57a}$$

$$\theta_h(f) = -2\pi f t_d \tag{3.57b}$$

Isto mostra que, para transmissão sem distorção, a resposta de amplitude $|H(f)|$ deve ser constante e a resposta de fase $\theta_h(f)$, uma função linear de f que passa pela origem em $f = 0$, como ilustrado na Fig. 3.25. A inclinação de $\theta_h(f)$ em relação à frequência angular $\omega = 2\pi f$ é $-t_d$, onde t_d é o atraso da saída em relação à entrada.*

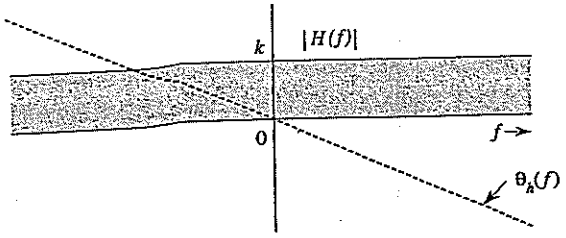


Figura 3.25 Resposta em frequência de um sistema linear invariante no tempo para transmissão sem distorção.

Sistemas Passa-Tudo *versus* Sistemas sem Distorção

Na análise de circuitos e no projeto de filtros, algumas vezes o maior interesse reside no ganho de um sistema. Um sistema passa-tudo tem ganho constante para todas as frequências [ou seja, $|H(f)| = k$], sem a exigência de fase linear. Notemos que, pela Eq. (3.57), um sistema sem distorção é um sistema passa-tudo, embora a recíproca não seja verdadeira. Como é muito comum que os principiantes se confundam com a diferença entre sistemas sem distorção e sistemas passa-tudo, este é o melhor momento para esclarecer os conceitos.

*Adicionalmente, é necessário que $\theta_h(0)$ seja 0 (como na Fig. 3.25) ou um valor constante $n\pi$ (n inteiro), ou seja, $\theta_h(f) = n\pi - 2\pi f t_d$. O efeito da adição da fase $n\pi$ é, no máximo, uma mudança de sinal da função.

Para comprovar como um sistema passa-tudo pode introduzir distorção, consideremos um exemplo ilustrativo. Imaginemos que queiramos transmitir um sinal musical gravado de um dueto violino-violoncelo. O violino contribui com a parte de frequências altas desse sinal musical, enquanto o violoncelo contribui com a parte de frequências baixas. Quando esse sinal musical é transmitido através de um dado sistema *passa-tudo*, as duas partes recebem o mesmo ganho. Contudo, suponhamos que esse sistema passa-tudo cause um atraso *extra* de um segundo no conteúdo de alta frequência da música (a parte do violino). Em consequência, na recepção, a audiência ouvirá um sinal "musical" que está totalmente fora de sincronismo, embora *todas as componentes do sinal tenham recebido o mesmo ganho e todas estejam presentes*. A diferença no atraso de transmissão das distintas componentes de frequência advém da fase não linear de $H(f)$ no filtro passa-tudo.

Para sermos mais precisos, o ganho da função de transferência $|H(f)|$ determina o ganho de cada componente de frequência de entrada, e $\angle H(f)$ determina o atraso de cada componente. Imaginemos que a entrada do sistema $x(t)$ consista em múltiplas senoides (suas componentes espectrais). Para que o sinal de saída $y(t)$ seja sem distorção, o mesmo deve ser igual ao sinal de entrada multiplicado pelo ganho k e atrasado por t_d. Para sintetizar um sinal desse tipo, $y(t)$ precisa ter exatamente as mesmas componentes de $x(t)$, sendo cada uma multiplicada por k e atrasada por t_d. Isso significa que a função de transferência do sistema, $H(f)$, deve ser tal que cada componente senoidal receba o mesmo ganho (ou perda) k e sofra o mesmo atraso temporal t_d segundos. A primeira condição requer que

$$|H(f)| = k$$

Vimos anteriormente (Seção 3.3) que, para que todas as componentes de frequências sofram o mesmo atraso temporal t_d, é necessário um atraso de fase linear $2\pi f t_d$ (Fig. 3.18) que passe pela origem

$$\theta_h(f) = -2\pi f t_d$$

Na prática, muitos sistemas têm característica de fase que pode apenas ser aproximada como linear. Um método conveniente para verificar a linearidade de fase consiste em traçar o gráfico da inclinação de $\angle H(f)$ em função da frequência. A inclinação pode ser uma função de f, no caso geral, dada por

$$t_d(f) = -\frac{1}{2\pi} \cdot \frac{d\theta_h(f)}{df} \tag{3.58}$$

Se a inclinação de θ_h for constante (ou seja, se θ_h for linear em relação a f), todas as componentes de frequência sofrerão o mesmo atraso temporal t_d. Se a inclinação não for constante, o atraso temporal t_d variará com a frequência. Isso significa que diferentes componentes de frequência sofrerão diferentes atrasos temporais e, em consequência, a forma de onda de saída não será uma réplica da forma de onda de entrada (como no exemplo do dueto de violino-violoncelo). Para que a transmissão de um sinal seja sem distorção, $t_d(f)$ deve ser uma constante t_d em toda a faixa de frequências de interesse.*

Existe, então, uma clara distinção entre sistemas passa-tudo e sistemas sem distorção. Um erro comum consiste em pensar que a planura da resposta de amplitude $|H(f)|$ baste para garantir a qualidade

*A Fig. 3.25 mostra que, para transmissão sem distorção, a resposta de fase deve não apenas ser linear, como também deve passar pela origem. Essa última exigência pode ser um pouco relaxada no caso de sinais passa-faixa. A fase na origem pode ser uma constante qualquer [$\theta_h(f) = \theta_0 - 2\pi f t_d$ ou $\theta_h(0) = \theta_0$]. A razão para isso reside na Eq. (3.37), que mostra que a adição de uma fase constante θ_0 ao espectro de um sinal passa-faixa equivale a deslocar a fase da portadora por θ_0. O sinal modulante (envelope) não é afetado. O envelope de saída é igual ao de entrada, atrasado por

$$t_g = -\frac{1}{2\pi}\frac{d\theta_h(f)}{df}$$

Esse atraso t_d é chamado de **atraso de grupo** ou **atraso de envelope**; a portadora de saída é igual à de entrada, atrasada por

$$t_p = -\frac{\theta_h(f)}{2\pi f}$$

Esse atraso é denominado **atraso de fase**, sendo f_0 a frequência central da banda passante.

do sinal. Um sistema que tenha uma resposta de amplitude plana pode distorcer um sinal a ponto de torná-lo irreconhecível, caso a resposta de fase não seja linear (t_d não constante).

Natureza da Distorção em Sinais de Áudio e de Vídeo

De modo geral, um ouvido humano é capaz de perceber prontamente uma distorção de amplitude, embora seja insensível a distorção de fase. Para que a distorção de fase se torne perceptível, a variação no atraso temporal (variação da inclinação de θ_h) deve ser comparável à duração do sinal (ou a duração fisicamente perceptível, se o sinal, propriamente dito, seja longo). No caso de sinais de áudio, cada sílaba falada pode ser considerada um sinal individual. A duração média de uma sílaba falada é da ordem de 0,01 a 0,1 segundo. Embora sistemas de áudio possam ter resposta de fase não linear, é possível que nenhuma distorção perceptível resulte, pois, em sistemas de áudio usados na prática, a variação máxima na inclinação de θ_h é apenas uma pequena fração de um milissegundo. Essa é a justificativa da afirmação "o ouvido humano é relativamente insensível à distorção de fase".[4] Um resultado disso é o fato de os fabricantes de equipamentos de áudio disponibilizarem apenas $|H(f)|$, a característica de resposta de amplitude de seus sistemas.

No caso de sinais de vídeo, a situação é exatamente oposta. O olho humano é sensível à distorção de fase e relativamente insensível à distorção de amplitude. Em sinais de televisão, a distorção de amplitude se manifesta como uma destruição parcial dos valores relativos de meios-tons da imagem resultante, o que não é prontamente percebido pelo olho humano. A distorção de fase (resposta de fase não linear), por sua vez, causa atrasos temporais diferentes aos distintos elementos da imagem. Isso resulta em uma imagem embaralhada, o que é imediatamente percebido pelo olho humano. A distorção de fase também é importante em sistemas de comunicação digital, pois a característica de fase não linear de um canal causa dispersão (alargamento temporal) dos pulsos, o que resulta em interferência entre pulsos vizinhos. Essa interferência pode provocar erros na amplitude dos pulsos no receptor: um 1 binário pode ser interpretado como 0 e vice-versa.

3.5 FILTROS IDEAIS *VERSUS* FILTROS PRÁTICOS

Filtros ideais permitem a transmissão sem distorção de uma determinada faixa de frequências e suprimem todas as outras frequências. O filtro passa-baixos ideal (Fig. 3.26), por exemplo, permite que todas as componentes de frequência abaixo de $f = B$ Hz passem sem distorção e suprime todas as componentes acima de $f = B$. A Fig. 3.27 mostra características de filtros passa-altos e passa-faixa ideais.

O filtro passa-baixos ideal na Fig. 3.26a tem inclinação de fase linear $-t_d$, que resulta em um atraso temporal de t_d segundos para todas as componentes de frequências abaixo de B Hz. Portanto, se o sinal de entrada $g(t)$ for limitado em banda a B Hz, a saída $y(t)$ será igual a $g(t)$ atrasado por t_d, ou seja,

$$y(t) = g(t - t_d)$$

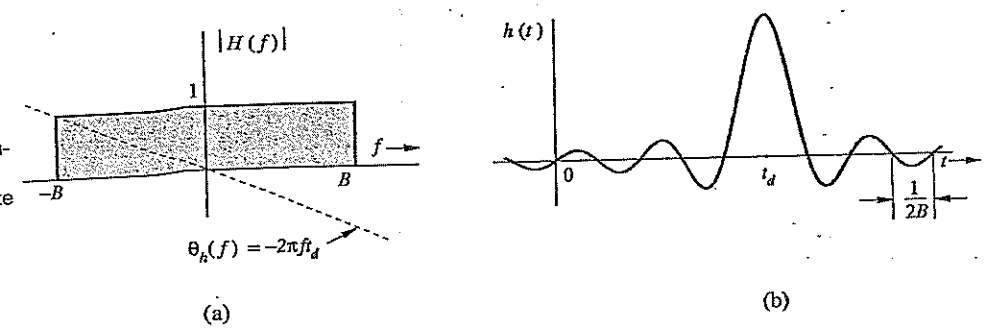


Figura 3.26 Resposta em frequência de um filtro passa-baixos ideal e correspondente resposta ao impulso.

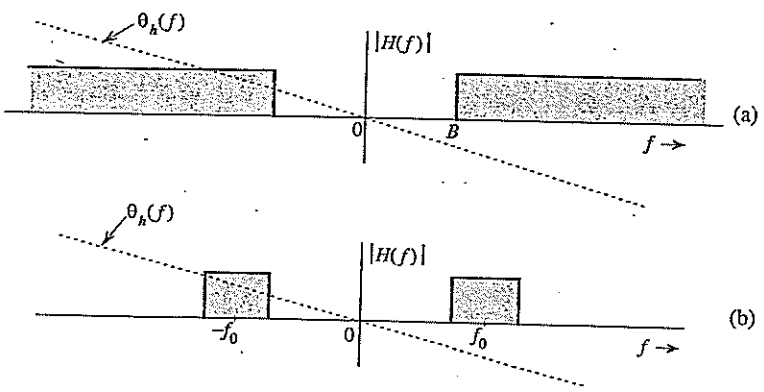


Figura 3.27 Respostas em frequência de filtros passa-altos e passa-faixa ideais.

O sinal $g(t)$ é transmitido por esse sistema sem distorção, mas com atraso temporal t_d. Para este filtro, $|H(f)| = \Pi(f/2B)$ e $\theta_h(f) = -2\pi f t_d$, de modo que

$$H(f) = \Pi\left(\frac{f}{2B}\right) e^{-j2\pi f t_d} \tag{3.59a}$$

A resposta ao impulso unitário $h(t)$ deste filtro pode ser obtida do par 18 na Tabela 3.1, e da propriedade de translação no tempo:

$$\begin{aligned} h(t) &= \mathcal{F}^{-1}\left[\Pi\left(\frac{f}{2B}\right) e^{-j2\pi f t_d}\right] \\ &= 2B \operatorname{sinc}[2\pi B(t - t_d)] \end{aligned} \tag{3.59b}$$

Recordemos que $h(t)$ é a resposta do sistema ao impulso unitário $\delta(t)$ aplicado em $t = 0$. A Fig. 3.26b ilustra um fato curioso: a resposta $h(t)$ tem início antes mesmo da aplicação da entrada (em $t = 0$). É óbvio que o filtro é não causal e, portanto, não realizável; ou seja, um sistema como este é fisicamente impossível, pois nenhum sistema razoável é capaz de responder a uma entrada **antes** que a mesma seja aplicada ao sistema. De modo similar, podemos mostrar que outros filtros ideais (como os filtros passa-altos e passa-faixa mostrados na Fig. 3.27) também não são fisicamente realizáveis.

Para um sistema fisicamente realizável, $h(t)$ deve ser causal, ou seja,

$$h(t) = 0 \quad \text{para } t < 0$$

No domínio da frequência, esta condição é equivalente ao **critério de Paley-Winer**, segundo o qual a condição necessária e suficiente para que $|H(f)|$ seja a resposta de amplitude de um sistema realizável (ou causal) é*

$$\int_{-\infty}^{\infty} \frac{|\ln |H(f)||}{1 + (2\pi f)^2} df < \infty \tag{3.60}$$

Se $H(f)$ não satisfizer esta condição, não será realizável. Vale notar que, se $|H(f)| = 0$ em uma banda finita qualquer, $|\ln|H(f)|| = \infty$ nesta banda, violando a condição (3.60). Se, no entanto, $H(f) = 0$ em uma única frequência (ou em um conjunto de frequências discretas), a integral na Eq. (3.60) ainda pode ser finita, embora o integrando seja infinito. Portanto, para um sistema fisicamente realizável, $H(f)$ pode ser zero em algumas frequências, mas não pode ser zero em uma faixa finita de frequências. Segundo esse critério, as características de filtros ideais (Figs. 3.26 e 3.27) são claramente não realizáveis.

*Assumimos que $|H(f)|$ seja quadrático-integrável, ou seja, assumimos que

$$\int_{-\infty}^{\infty} |H(f)|^2 df$$

seja finita

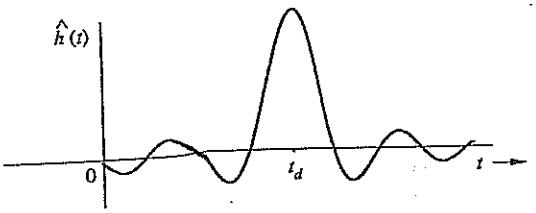


Figura 3.28 Realização aproximada de uma característica de filtro passa-baixos ideal com truncamento da resposta ao impulso.

A resposta ao impulso $h(t)$ na Fig. 3.26 não é realizável. Uma abordagem prática ao projeto de filtros consiste em cortar a cauda de $h(t)$ para $t < 0$. A resultante resposta ao impulso, $\widehat{h}(t)$ é causal e dada por

$$\widehat{h}(t) = h(t)u(t)$$

por ser causal (Fig. 3.28), $\widehat{h}(t)$ é fisicamente realizável. Se t_d for suficientemente grande, $\widehat{h}(t)$ será uma boa aproximação de $h(t)$ e o filtro correspondente, $\widehat{H}(f)$ será uma boa aproximação de um filtro ideal. Esta realização de um filtro ideal é possível devido ao maior valor do atraso temporal t_d. Isso significa que o preço de uma boa aproximação física é um maior atraso na saída, o que, com frequência, é verdade para sistemas não causais. Teoricamente, a realização da característica ideal requer um atraso $t_d = \infty$. Contudo, um exame da Fig. 3.27b mostra que um atraso t_d de três ou quatro vezes π/W fará com que $\widehat{h}(t)$ seja uma boa versão aproximada de $h(t - t_d)$. Por exemplo, filtros de áudio devem passar frequências de até 20 kHz (a maior frequência que o ouvido humano pode distinguir). Neste caso, um atraso t_d da ordem de 10^{-4} segundo (0,1 ms) seria uma escolha razoável. No entanto, a operação de truncamento [corte da cauda de $h(t)$ para torná-la causal] cria alguns problemas inesperados de espalhamento e vazamento espectrais, que podem ser parcialmente corrigidos com o emprego de uma função de trancamento gradual para que a cauda de $h(t)$ seja cortada de forma gradual (e não de forma abrupta).[5]

Na prática, podemos realizar uma variedade de características de filtros para aproximar características ideais. Filtros práticos (ou realizáveis) têm resposta de amplitude $|H(f)|$ com características graduais, sem descontinuidades do tipo degrau. Por exemplo, filtros de Butterworth ou de Chebychev são largamente utilizados em várias aplicações, incluindo circuitos práticos de comunicação.

Sinais analógicos também podem ser processados de forma digital (conversão A/D). Isto envolve amostragem, quantização e codificação. O resultante sinal digital pode ser processado por um pequeno computador digital especialmente projetado para converter a sequência de entrada na sequência de saída desejada. A sequência de saída é convertida de volta ao desejado sinal analógico. Um algoritmo especial para o computador digital de processamento pode ser usado para implementar uma dada operação com sinais (por exemplo, filtragem passa-baixos, passa-faixa ou passa-altos). O tema de filtragem digital foge um pouco do escopo deste livre. Existem livros excelentes sobre esse assunto.[3]

3.6 DISTORÇÃO DE SINAL EM UM CANAL DE COMUNICAÇÃO

Um sinal transmitido ao longo de um canal é distorcido por várias imperfeições do canal. A seguir, estudaremos a natureza da distorção de um sinal.

3.6.1 Distorção Linear

Primeiro, consideremos canais lineares invariantes no tempo. Em um canal desse tipo, a distorção do sinal pode ser causada por distorção de magnitude ou de fase, ou ambas, devido a características não ideais do canal. Podemos determinar os efeitos que tais **não idealidades** terão sobre um pulso

$g(t)$ transmitido ao longo do canal. Admitamos que o pulso exista em um intervalo (a, b) e seja zero fora dele. As componentes do espectro de Fourier do pulso terão um equilíbrio perfeito e delicado de magnitudes e fases, de modo que a soma das mesmas forma precisamente o pulso $g(t)$ no intervalo (a, b) e é zero fora dele. A transmissão de $g(t)$ através de um canal ideal que satisfaz as condições de transmissão sem distorção não perturba esse equilíbrio, pois um canal sem distorção multiplica todas as componentes pelo mesmo fator e as atrasa pelo mesmo intervalo de tempo. Contudo, se a resposta de amplitude do canal for não ideal [ou seja, se $|H(f)|$ não for constante], esse delicado equilíbrio será quebrado e a soma de todas as componentes deixará de ser zero fora do intervalo (a, b). Em resumo, o pulso se espalhará (ver Exemplo 3.14). O mesmo ocorre se a característica de fase do canal não for ideal, ou seja, se $\theta_h(f) \neq -2\pi f t_d$. Assim, alargamento, ou **dispersão**, do pulso ocorrerá se a resposta de amplitude ou a resposta de fase não for ideal, ou se ambas não forem ideais.

Distorção de canal linear (alargamento temporal) é particularmente danosa a sistemas de comunicação digital, pois introduz o que é conhecido como interferência intersimbólica (ISI — *InterSymbol Interference*). Em outras palavras, ao ser transmitido através de um canal dispersivo, um símbolo digital tende a se alargar além do intervalo de tempo a ele reservado. Portanto, símbolos adjacentes interferirão uns com os outros, aumentando a probabilidade de erro de detecção no receptor.

Exemplo 3.14 A função de transferência de um filtro passa-baixos $H(f)$ (Fig. 3.19a) é dada por

$$H(f) = \begin{cases} (1 + k\cos 2\pi f T)e^{-j2\pi f t_d} & |f| < B \\ 0 & |f| > B \end{cases} \tag{3.61}$$

Um pulso $g(t)$ limitado em banda a B Hz (Fig. 3.29b) é aplicado à entrada desse filtro. Determinemos a saída $y(t)$.

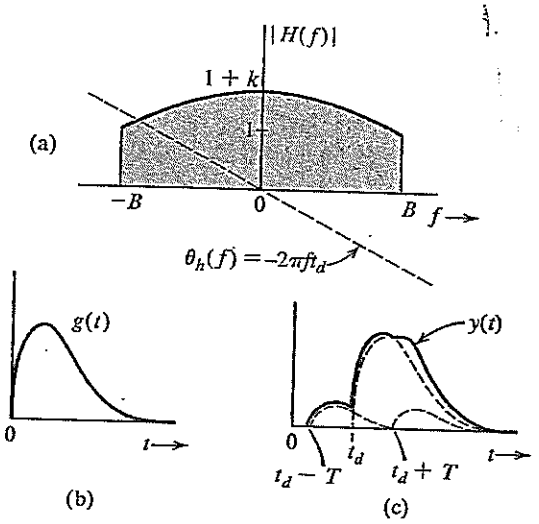


Figura 3.29 Pulso dispersado ao passar por um sistema que não é sem distorção.

A característica de fase desse filtro é ideal, mas a de amplitude é não ideal. Como $g(t) \Longleftrightarrow G(f)$, $y(t) \Longleftrightarrow Y(f)$, e

$$\begin{aligned} Y(f) &= G(f)H(f) \\ &= G(f) \cdot \Pi\left(\frac{f}{2B}\right)(1 + k\cos 2\pi f T)e^{-j2\pi f t_d} \\ &= G(f)e^{-j2\pi f t_d} + k\,[G(f)\cos 2\pi f T]\; e^{-j2\pi f t_d} \end{aligned} \tag{3.62}$$

Vale notar que, na dedução da Eq. (3.62), como $g(t)$ é limitado em banda a B Hz, temos $G(f) \times \Pi\left(\frac{f}{2B}\right) = G(f)$. Fazendo uso da propriedade de translação no tempo e da Eq. (3.32a), obtemos

$$y(t) = g(t - t_d) + \frac{k}{2}[g(t - t_d - T) + g(t - t_d + T)] \tag{3.63}$$

A saída é, na verdade, $g(t) + (k/2)[g(t - T) + g(t + T)]$ atrasado de t_d. Ou seja, a saída consiste em $g(t)$ e seus ecos deslocados por $\pm t_d$. A dispersão do pulso causada pelos ecos fica evidente na Fig. 3.29c. Se $H(f)$ representar resposta ideal de amplitude e resposta não ideal de fase, o efeito será similar (ver Exercício 3.6-1).

3.6.2 Distorção Causada por Não Linearidades do Canal

Até aqui, consideremos o canal como linear. Essa aproximação é válida apenas para pequenos sinais. Para sinais de grande amplitude, não linearidades não podem ser desprezadas. Uma discussão detalhada de sistemas não lineares está além do escopo deste livro. Consideraremos um caso simples de um canal não ideal sem memória, onde a entrada g e a saída y se relacionam por alguma equação não linear (sem memória),

$$y = f(g)$$

O lado direito dessa equação pode ser expandido em uma série de Maclaurin, como

$$y(t) = a_0 + a_1 g(t) + a_2 g^2(t) + a_3 g^3(t) + \cdots + a_k g^k(t) + \cdots$$

Recordemos o resultado na Seção 3.3.6 (convolução), segundo o qual, se a largura de banda de $g(t)$ for B hz, então a largura de banda de $g^k(t)$ será kB Hz. Portanto, a largura de banda de $y(t)$ é **maior que** kB Hz. Em consequência, o espectro de saída será mais largo que o de entrada, e o sinal de saída conterá componentes de frequências não contidas no sinal de entrada. Em comunicação por difusão, é necessário amplificar sinais a níveis de potência muito elevados, o que requer amplificadores de alta eficiência (de classe C). Infelizmente, esses amplificadores são não lineares e, ao serem usados na amplificação de sinais, causam distorção. Esse é um dos sérios problemas de sinais AM. No entanto, sinais FM não são afetados por distorção não linear, como mostrado no Capítulo 5. Se um sinal for transmitido através de um canal não linear, a não linearidade não apenas distorce o sinal, mas também causa interferência em outros sinais no canal, devido a sua dispersão (alargamento) espectral.

No caso de sistemas de comunicação digital, o efeito da distorção não linear difere do efeito de alargamento temporal causado por distorção linear. A distorção linear provoca interferência entre sinais em um mesmo canal, enquanto a dispersão espectral, devido à distorção não linear, causa interferência entre sinais que usam diferentes canais de frequência.

Exemplo 3.15 A relação entre a entrada $x(t)$ e a saída $y(t)$ de um certo canal não linear é dada por

$$y(t) = x(t) + 0{,}000158x^2(t)$$

Determinemos o sinal de saída $y(t)$ e seu espectro $Y(f)$, considerando um sinal de entrada $x(t) = 2000\,\text{sinc}(2000\pi t)$. Comprovemos que a largura de banda do sinal de saída é o dobro da largura de banda do sinal de entrada. Esse é o resultado de usar o quadrado de um sinal. Será possível recuperar o sinal $x(t)$ (sem distorção) da saída $y(t)$?

Como

$$x(t) = 2000\,\text{sinc}\,(2000\pi t) \qquad \Longleftrightarrow \qquad X(f) = \Pi\left(\frac{f}{2000}\right)$$

Temos,

$$y(t) = x(t) + 0{,}000158x^2(t) = 2000\,\text{sinc}\,(2000\pi t) + 0{,}316 \cdot 2000\,\text{sinc}^2\,(2000\pi t)$$

$$\Longleftrightarrow$$

$$Y(f) = \Pi\left(\frac{f}{2000}\right) + 0{,}316\,\Delta\left(\frac{f}{4000}\right)$$

Vale notar que $0{,}316 \times 2000\ \text{sinc}^2\,(2000\pi t)$ é o termo indesejado (distorção) no sinal recebido. A Fig. 3.30a mostra o espectro do sinal de entrada (desejado) $X(f)$; a Fig. 3.30b mostra o espectro do termo indesejado (distorção); a Fig. 3.30c mostra o espectro recebido $Y(f)$. Façamos as seguintes observações:

1. A largura de banda do sinal recebido $y(t)$ é o dobro da largura de banda do sinal de entrada $x(t)$ (pois o sinal foi elevado ao quadrado).
2. O sinal recebido contém o sinal de entrada $x(t)$ e o sinal indesejado $632\ \text{sinc}^2\,(2000\pi t)$. Os espectros desses dois sinais são ilustrados na Fig. 3.30a e b. A Fig. 3.30c mostra $Y(f)$, o espectro do sinal recebido. Os espectros do sinal desejado e do sinal de distorção se sobrepõem, o que impede que o sinal $x(t)$ seja recuperado do sinal $y(t)$ sem alguma distorção.
3. Podemos reduzir a distorção aplicando o sinal recebido a um filtro passa-baixos com largura de banda de 1000 Hz. O espectro da saída desse filtro é ilustrado na Fig. 3.30d. A saída do filtro é o sinal de entrada desejado $x(t)$ com alguma distorção residual.

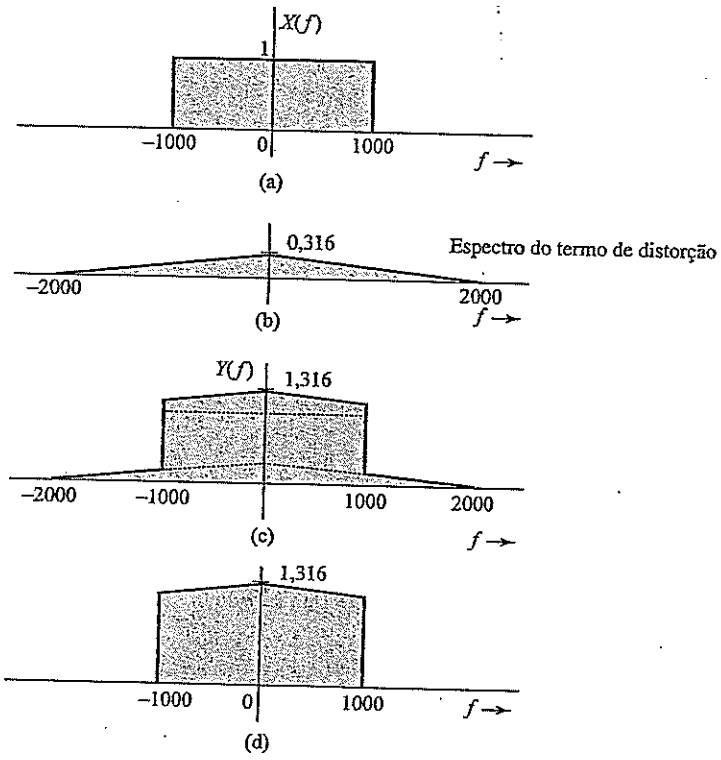


Figura 3.30 Distorção de sinal causada por operação não linear. (a) Espectro do sinal desejado (de entrada). (b) Espectro do sinal indesejado (distorção) no sinal recebido. (c) Espectro do sinal recebido. (d) Espectro do sinal recebido após filtragem passa-baixos.

4. Haverá um problema adicional de interferência com outros sinais caso o sinal de entrada $x(t)$ seja multiplexado por divisão em frequência juntamente com outros sinais nesse mesmo canal. Isso significa que vários sinais que ocupam faixas de frequências que não se sobrepõem são transmitidos simultaneamente no mesmo canal. Se o espectro $X(f)$ se alargar além de sua banda original de 1000 Hz, interferirá com o sinal que ocupa a faixa de frequências entre 1000 e 2000 Hz. Assim, em adição à distorção de $x(t)$, haverá interferência com a banda adjacente.
5. Se $x(t)$ fosse um sinal digital consistindo em um trem de pulsos, cada pulso seria distorcido, mas não haveria interferência entre pulsos adjacentes. Mesmo com pulsos distorcidos, os dados podem ser recebidos sem perda, pois comunicação digital é capaz de suportar considerável distorção de pulsos sem perda de informação. Portanto, se esse canal fosse usado para transmitir um sinal multiplexado por divisão no tempo e consistindo em dois trens de pulsos entrelaçados, os dados nos dois trens seriam recuperados no receptor.

3.6.3 Distorção Causada por Efeitos de Multipercurso

Uma transmissão em multipercurso ocorre quando um sinal transmitido chega ao receptor por dois ou mais percursos, com atrasos distintos. Por exemplo, se um sinal for transmitido através de um cabo que apresenta irregularidades (descasamento) de impedâncias ao longo do percurso, o sinal chegará ao receptor na forma de uma onda direta mais diversas reflexões com atrasos variáveis. No caso de enlaces de rádio, o sinal pode ser recebido de um percurso direto entre as antenas transmissora e receptora e, também, de reflexões por outros objetos, como montanhas e edifícios. Em enlaces de rádio de longa distância que utilizam a ionosfera, efeitos similares ocorrem por conta de percursos de um e de múltiplos saltos. Em cada caso, o canal de transmissão pode ser representado como vários canais em paralelo, cada um com diferentes valores de atenuação relativa e de atraso temporal. Consideremos o caso de apenas dois percursos: um com ganho unitário e atraso t_d, e outro com ganho α e atraso $t_d + \Delta t$, como ilustrado na Fig. 3.31a. As funções de transferências dos dois percursos são dadas $e^{-j2\pi f t_d}$ e $\alpha e^{-j2\pi f(t_d+\Delta t)}$, respectivamente. A função de transferência global de um canal como esse é $H(f)$, dada por

$$\begin{aligned} H(f) &= e^{-j2\pi f t_d} + \alpha e^{-j2\pi f(t_d+\Delta t)} \\ &= e^{-j2\pi f t_d}(1 + \alpha e^{-j2\pi f \Delta t}) \\ &= e^{-j2\pi f t_d}(1 + \alpha \cos 2\pi f \Delta t - j\alpha \operatorname{sen} 2\pi f \Delta t) \end{aligned} \tag{3.64a}$$

$$= \underbrace{\sqrt{1 + \alpha^2 + 2\alpha \cos 2\pi f \Delta t}}_{|H(f)|} \exp\left[-j\underbrace{\left(2\pi f t_d + \tan^{-1}\frac{\alpha \operatorname{sen} 2\pi f \Delta t}{1 + \alpha \cos 2\pi f \Delta t}\right)}_{\theta_h(f)}\right] \tag{3.64b}$$

As características de amplitude e de fase de $H(f)$ são periódicas em f, com período $1/\Delta t$ (Fig. 3.31b). O canal de multipercurso pode, portanto, exibir não idealidades nas características de amplitude e de fase e pode causar distorção linear (dispersão de pulsos), como discutido anteriormente.

Se, por exemplo, os ganhos dos dois percursos forem muito próximos, ou seja, se $\alpha \approx 1$, os sinais recebidos nos dois percursos podem ter fases opostas (defasadas por π radianos) em certas frequências. Isso significa que, em frequências nas quais os dois percursos resultam em fases opostas, os sinais dos dois percursos quase se cancelarão mutuamente. A Eq. (3.64b) mostra que, em frequências tais que $f = n/(2\Delta t)$ (n *ímpar*), $\cos 2\pi f \Delta t = -1$, e $|H(f)| \approx 0$ quando $\alpha \approx 1$. Essas são as frequências de nulo devido a multipercurso. Nas frequências $f = n/(2\Delta t)$ (n par), os dois sinais interferirão de forma construtiva, reforçando o ganho. Tais canais são responsáveis pelo **desvanecimento seletivo em fre-**

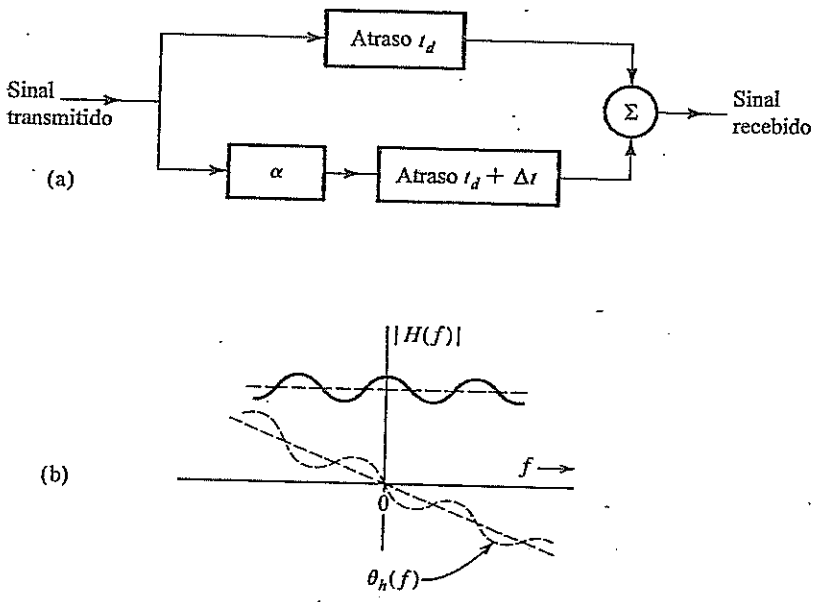


Figura 3.31 Transmissão em multipercurso.

quência de sinais transmitidos. Esse tipo de distorção pode ser parcialmente corrigido com o emprego de um equalizador com linha de retardo gradual, como mostrado no Exercício 3.6-2. Tais equalizadores são úteis em várias aplicações em comunicação; o projeto dos mesmos é abordado nos Capítulos 7 e 12.

3.6.4 Canais com Desvanecimento

Até aqui, presumimos que as características do canal eram constantes no tempo. Na prática, encontramos canais cujas características de transmissão variam com o tempo. Por exemplo, canais baseados no espalhamento troposférico e canais que, para alcançar comunicação de longa distância, utilizam a ionosfera para a reflexão de rádio. As variações temporais das propriedades do canal têm origem em alterações semiperiódicas e aleatórias das características de propagação do meio. As propriedades de reflexão da ionosfera, por exemplo, estão relacionadas às condições meteorológicas, que sofrem mudanças sazonais, diárias ou, até mesmo, horárias, assim como acontece com o clima. Períodos de tempestades repentinas também ocorrem. Portanto, a função de transferência efetiva do canal varia de modo semiperiódico e aleatório, causando atenuação randômica do sinal. Esse fenômeno é conhecido como **desvanecimento**. Efeitos de desvanecimento lento podem ser reduzidos com o emprego de **controle automático de ganho (AGC** – *Automatic Gain Control*).*

O desvanecimento poder exibir grande dependência da frequência, de modo que diferentes componentes de frequência são afetadas de modo não uniforme. Este tipo de desvanecimento, conhecido como desvanecimento seletivo em frequência, pode causar sérios problemas às comunicações. Propagação em multipercurso pode originar desvanecimento seletivo em frequência.

3.7 ENERGIA E DENSIDADE ESPECTRAL DE ENERGIA DE SINAIS

A energia E_g de um sinal $g(t)$ é definida como a área sob a curva de $|g(t)|^2$. A energia também pode ser calculada da transformada de Fourier $G(f)$, através do teorema de Parseval.

*O AGC também elimina variações lentas do sinal original.

3.7.1 Teorema de Parseval

A energia de um sinal pode ser relacionada ao espectro do sinal $G(f)$, por substituição da Eq. (3.9b) na Eq. (2.2):

$$E_g = \int_{-\infty}^{\infty} g(t)g^*(t)\,dt = \int_{-\infty}^{\infty} g(t)\left[\int_{-\infty}^{\infty} G^*(f)e^{-j2\pi ft}\,df\right]dt$$

Nesta equação, usamos o fato de que $g^*(t)$, por ser o complexo conjugado de $g(t)$, pode ser expresso como o conjugado do lado direito da Eq. (3.9b). Agora, trocando a ordem de integração, obtemos,

$$\begin{aligned} E_g &= \int_{-\infty}^{\infty} G^*(f)\left[\int_{-\infty}^{\infty} g(t)e^{-j2\pi ft}\,dt\right]df \\ &= \int_{-\infty}^{\infty} G(f)G^*(f)\,df \\ &= \int_{-\infty}^{\infty} |G(f)|^2\,df \end{aligned} \tag{3.65}$$

Este último resultado é o conhecido teorema de Parseval. Um resultado similar foi obtido, na Eq. (2.68), para um sinal periódico e sua série de Fourier. O teorema de Parseval permite que determinemos a energia de um sinal tanto de sua especificação no domínio do tempo, $g(t)$, como de sua especificação no domínio da frequência, $G(f)$.

Exemplo 3.16 Comprovemos o teorema de Parseval para o sinal $g(t) = e^{-at}u(t)$ $(a > 0)$.

Temos

$$E_g = \int_{-\infty}^{\infty} g^2(t)\,dt = \int_0^{\infty} e^{-2at}\,dt = \frac{1}{2a} \tag{3.66}$$

Agora, determinemos E_g do espectro do sinal $G(f)$, dado por

$$G(f) = \frac{1}{j2\pi f + a}$$

e, da Eq. (3.65)

$$E_g = \int_{-\infty}^{\infty} |G(f)|^2\,df = \int_{-\infty}^{\infty} \frac{1}{(2\pi f)^2 + a^2}\,df = \frac{1}{2\pi a}\tan^{-1}\frac{2\pi f}{a}\bigg|_{-\infty}^{\infty} = \frac{1}{2a}$$

o que comprova o teorema de Parseval.

3.7.2 Densidade Espectral de Energia (ESD)

Uma interpretação da Eq. (3.65) nos diz que a energia de um sinal $g(t)$ é o resultado das energias contribuídas por todas as componentes espectrais do sinal $g(t)$. A contribuição de uma componente espectral de frequência f é proporcional a $|G(f)|^2$. Para detalhar um pouco mais, consideremos um sinal $g(t)$ aplicado à entrada de um filtro passa-faixa ideal, cuja função de transferência $H(f)$ é mostrada na Fig. 3.32a. Esse filtro suprime todas as frequências, exceto aquelas em uma estreita faixa Δf $(\Delta f \to 0)$ cen-

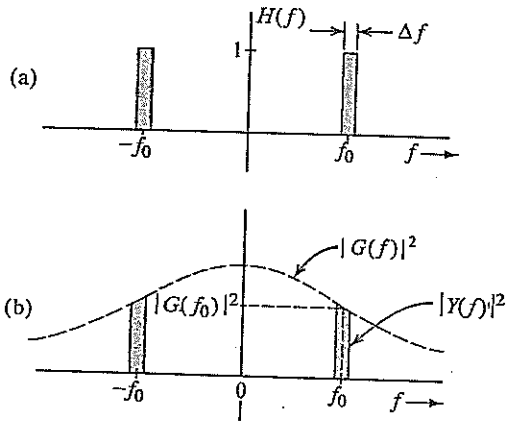


Figura 3.32 Interpretação da densidade espectral de energia de um sinal.

trada na frequência f_0 (Fig. 3.32b). Seja $y(t)$ a saída do filtro; então sua transformada de Fourier $Y(f)$ é $G(f)H(f)$, e a energia E_y de $y(t)$ é dada por

$$E_y = \int_{-\infty}^{\infty} |G(f)H(f)|^2 \, df \tag{3.67}$$

Como $H(f) = 1$ na banda passante Δf e é zero em todas as outras frequências, a integral no lado direito é a soma das duas áreas hachuradas na Fig. 3.32b; logo (para $\Delta f \to 0$)

$$E_y = 2\,|G(f_0)|^2 \, df$$

Portanto, $2|G(f)|^2 df$ é a energia contribuída pelas componentes espectrais contidas nas duas bandas estreitas, cada uma com largura Δf Hz, centradas em $\pm f_0$. Podemos, então, interpretar $|G(f)|^2$ como a energia por unidade de largura de banda (em hertz) das componentes espectrais de $g(t)$ centradas na frequência f. Em outras palavras, $|G(f)|^2$ é a densidade espectral de energia (por unidade de largura de banda em hertz) de $g(t)$. Na verdade, como as componentes de frequências positivas e negativas se combinam para formar as componentes na banda Δf, a energia contribuída por unidade de largura de banda é $2|G(f)|^2$. Contudo, por conveniência, consideramos frequências positivas e negativas como independentes. Assim a **densidade espectral de energia (ESD** — *Energy Spectral Density*) $\Psi_g(f)$ é definida como

$$\Psi_g(f) = |G(f)|^2 \tag{3.68}$$

e a Eq. (3.65) pode ser expressa como

$$E_g = \int_{-\infty}^{\infty} \Psi_g(f) \, df \tag{3.69a}$$

Dos resultados do Exemplo 3.16, a ESD do sinal $g(t) = e^{-at}u(t)$ é

$$\Psi_g(f) = |G(f)|^2 = \frac{1}{(2\pi f)^2 + a^2} \tag{3.69b}$$

3.7.3 Largura de Banda Essencial de um Sinal

O espectro da maioria dos sinais se estende ao infinito. Contudo, como a energia de um sinal prático é finita, o espectro do sinal deve tender a 0 quando $f \to \infty$. A maior parte da energia do sinal está contida em uma certa banda de B Hz, e a energia contribuída pelas componentes de frequência maiores que

B Hz é desprezível. Podemos, então, suprimir o espectro do sinal que se estende além de B Hz, com pequeno efeito na forma e na energia do sinal. A largura de banda B é denominada **largura de banda essencial** do sinal. O critério para a determinação de B depende da tolerância ao erro em uma particular aplicação. Podemos, por exemplo, selecionar B de modo que a largura de banda contenha 95% da energia do sinal.* O nível de energia pode ser maior ou menor que 95%, dependendo da precisão necessária. Podemos fazer uso de um critério como este para determinar a largura de banda essencial de um sinal. A supressão de todas as componentes espectrais de $g(t)$ fora da largura de banda essencial resulta em um sinal $\hat{g}(t)$, que é uma boa aproximação de $g(t)$.** Caso usemos o critério de 95% para a largura de banda essencial, a energia do erro (diferença) $g(t) - \hat{g}(t)$ será 5% de E_g. O exemplo a seguir ilustra o procedimento para estimar a largura de banda.

Exemplo 3.17 Estimemos a largura de banda essencial W (em rad/s) do sinal $e^{-at}u(t)$ para que a banda essencial contenha 95% da energia do sinal.

Neste caso,

$$G(f) = \frac{1}{j2\pi f + a}$$

e a ESD é

$$|G(f)^2| = \frac{1}{(2\pi f)^2 + a^2}$$

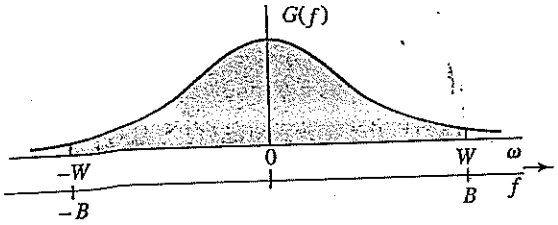


Figura 3.33 Estimativa da largura de banda essencial de um sinal.

A ESD é ilustrada na Fig. 3.33. A energia do sinal E_g é a área sob a curva da ESD, e já foi calculada como $1/2a$. Seja W rad/s a largura de banda essencial, que contém 95% da energia total do sinal E_g. Isso significa que $1/2\pi$ vezes a área hachurada na Fig. 3.33 é $0{,}95/2a$, ou seja,

$$\frac{0{,}95}{2a} = \int_{-W/2\pi}^{W/2\pi} \frac{df}{(2\pi f)^2 + a^2}$$

$$= \frac{1}{2\pi a} \tan^{-1} \frac{2\pi f}{a}\bigg|_{-W/2\pi}^{W/2\pi} = \frac{1}{\pi a} \tan^{-1} \frac{W}{a}$$

ou

$$\frac{0{,}95\pi}{2} = \tan^{-1} \frac{W}{a} \Longrightarrow W = 12{,}7\,a \text{ rad/s}$$

*No caso de um sinal passa-baixos, a largura de banda essencial também pode ser definida como a frequência na qual o valor da amplitude do espectro é uma pequena fração (digamos, 5–10%) do valor de pico. No Exemplo 3.16, o pico de $|G(f)|$ é $1/a$ e ocorre em $f = 0$.

**Na prática, o truncamento é efetuado de forma gradual, com o uso de janelas graduais, de modo a evitar vazamento espectral excessivo devido a um truncamento abrupto.[5]

O valor da largura de banda essencial em hertz é

$$B = \frac{W}{2\pi} = 2{,}02\,a \quad \text{Hz}$$

Isso significa que, na banda de 0 (dc) a $12{,}7 \times a$ rad/s ($2{,}02 \times a$ Hz), as componentes espectrais de $g(t)$ contribuem com 95% da energia total do sinal; todas as componentes espectrais restantes (na banda de $2{,}02 \times a$ Hz a ∞) contribuem com apenas 5% da energia do sinal.*

Exemplo 3.18 Estimemos a largura de banda essencial de um pulso retangular $g(t) = \Pi(t/T)$ (Fig. 3.34a), para que a largura de banda essencial contenha pelo menos 90% da energia do pulso.

Para este pulso, a energia E_g é

$$E_g = \int_{-\infty}^{\infty} g^2(t)\,dt = \int_{-T/2}^{T/2} dt = T$$

E, como

$$\Pi\left(\frac{t}{T}\right) \Longleftrightarrow T\ \text{sinc}\,(\pi f T)$$

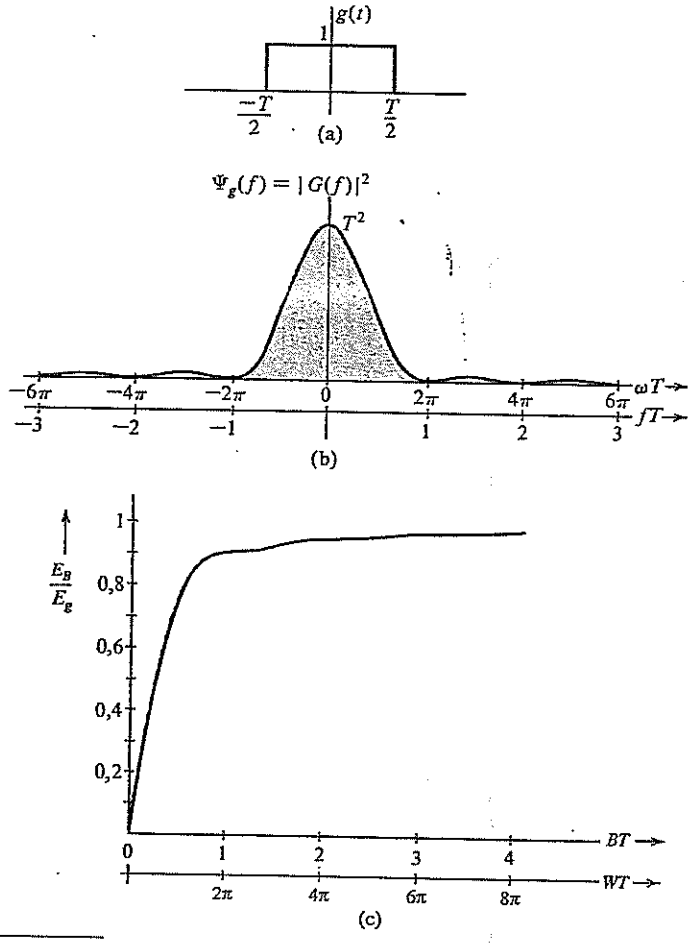


Figura 3.34 (a) Função retangular EX-FGN/FGC, (b) correspondente densidade espectral de energia e (c) fração da energia na banda B (H_2).

*Vale notar que, embora a ESD exista em toda a banda de $-\infty$ a ∞, o espectro trigonométrico existe apenas na banda de 0 a ∞. O intervalo espectral $-\infty$ a ∞ se aplica ao espectro exponencial. Contudo, na prática, sempre que nos referimos a uma largura de banda, o fazemos no sentido trigonométrico. Portanto, a banda essencial é de 0 a B Hz (ou W rad/s) e não de $-B$ a B.

a ESD desse pulso é dada por

$$\Psi_g(f) = |G(f)|^2 = T^2 \operatorname{sinc}^2(\pi f T)$$

A ESD é ilustrada na Fig. 3.34b como função de ωT e de fT, sendo f a frequência em hertz. A energia E_B na banda de 0 a B Hz é dada por

$$E_B = \int_{-B}^{B} T^2 \operatorname{sinc}^2(\pi f T)\, df$$

Fazendo $2\pi f T = x$ na integral, de modo que $df = dx/(2\pi T)$, obtemos

$$E_B = \frac{T}{\pi} \int_0^{2\pi BT} \operatorname{sinc}^2\left(\frac{x}{2}\right) dx$$

E, como $E_g = T$, temos

$$\frac{E_B}{E_g} = \frac{1}{\pi} \int_0^{2\pi BT} \operatorname{sinc}^2\left(\frac{x}{2}\right) dx$$

A integral no lado direito é calculada numericamente; um gráfico de E_B/E_g em função de BT é mostrado na Fig. 3.34c. Notemos que 90,28% da energia total do pulso $g(t)$ estão contidos na banda $B = 1/T$ Hz. Portanto, segundo o critério de 90%, a largura de banda de um pulso retangular de largura T segundos é $1/T$ Hz.

3.7.4 Energia de Sinais Modulados

Vimos que a modulação desloca o espectro de sinal $G(f)$ para a esquerda e a direita de f_0. A seguir, mostraremos que algo semelhante ocorre com a ESD do sinal modulado.

Seja $g(t)$ um sinal em banda base, limitado a B Hz. O sinal modulado em amplitude $\varphi(t)$ é dado por

$$\varphi(t) = g(t) \cos 2\pi f_0 t$$

O espectro (transformada de Fourier) de $\varphi(t)$ é calculado como

$$\Phi(f) = \frac{1}{2}[G(f + f_0) + G(f - f_0)]$$

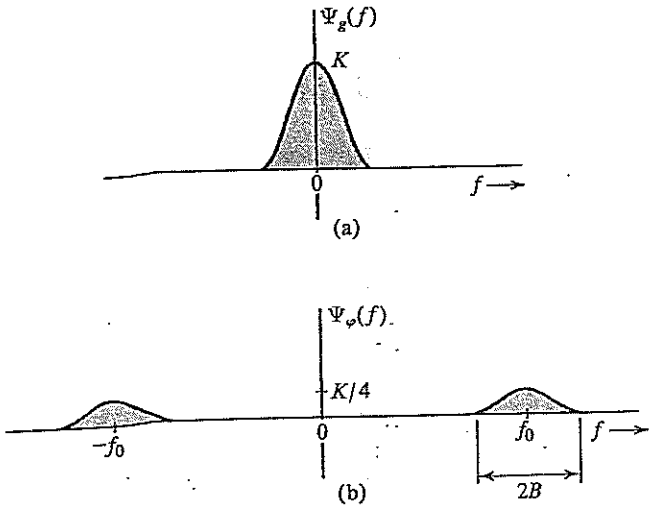


Figura 3.35 Densidades espectrais de energia de sinais modulante e modulado.

A ESD do sinal modulado $\varphi(t)$ é $|\Phi(f)|^2$, ou seja,

$$\Psi_\varphi(f) = \frac{1}{4}|G(f+f_0) + G(f-f_0)|^2$$

Se $f_0 \geq B$, $G(f+f_0)$ e $G(f-f_0)$ não se sobrepõem (ver Fig. 3.35), e

$$\begin{aligned}\Psi_\varphi(f) &= \frac{1}{4}\left[|G(f+f_0)|^2 + |G(f-f_0)|^2\right] \\ &= \frac{1}{4}\Psi_g(f+f_0) + \frac{1}{4}\Psi_g(f-f_0)\end{aligned} \tag{3.70}$$

As ESDs de $g(t)$ e do sinal modulado $\varphi(t)$ são mostradas na Fig. 3.35. Fica claro que a modulação desloca a ESD de $g(t)$ de $\pm f_0$. Observemos que a área sob $\Psi_\varphi(f)$ é a metade da área sob $\Psi_g(f)$. Como a energia de um sinal é proporcional à área sob a curva da ESD, concluímos que a energia de $\varphi(t)$ é a metade da energia de $g(t)$, ou seja,

$$E_\varphi = \frac{1}{2}E_g \qquad f_0 \geq B \tag{3.71}$$

Pode parecer surpreendente que um sinal $\varphi(t)$, que aparenta ser tão energético em comparação com $g(t)$, tenha apenas a metade da energia de $g(t)$. As aparências enganam, como sempre. A energia de um sinal é proporcional ao quadrado de sua amplitude; amplitudes mais altas contribuem mais para a energia. O sinal $g(t)$ mantém altos níveis de amplitude na maior parte do tempo. O sinal modulado $\varphi(t)$, por sua vez, devido ao fator $\cos 2\pi f_0 t$, cai ao nível de amplitude zero muitas vezes, o que reduz sua energia.

3.7.5 Função de Autocorrelação Temporal e Densidade Espectral de Energia

No Capítulo 2, mostramos que uma boa medida de comparação de dois sinais $g(t)$ e $z(t)$ é a função de correlação cruzada $\psi_{gz}(\tau)$, definida na Eq. (2.46). Também definimos a correlação de um sinal $g(t)$ com ele próprio [função de autocorrelação $\psi_g(\tau)$] na Eq. (2.47). Para um sinal $g(t)$ de valores reais, a função de autocorrelação $\psi_g(\tau)$ é dada por*

$$\psi_g(\tau) = \int_{-\infty}^{\infty} g(t)g(t+\tau)\,dt \tag{3.72a}$$

Fazendo $x = t + \tau$ na Eq. (3.72a), obtemos

$$\psi_g(\tau) = \int_{-\infty}^{\infty} g(x)g(x-\tau)\,dx$$

Nesta equação, x é uma simples variável de integração e pode ser substituída por t. Assim,

$$\psi_g(\tau) = \int_{-\infty}^{\infty} g(t)g(t\pm\tau)\,dt \tag{3.72b}$$

Isso mostra que, para um sinal $g(t)$ de valores reais, a função de autocorrelação é uma função par de τ, ou seja,

$$\psi_g(\tau) = \psi_g(-\tau) \tag{3.72c}$$

*Para um sinal de valores complexos, definimos

$$\psi_g(\tau) = \int_{-\infty}^{\infty} g(t)g^*(t-\tau)\,dt = \int_{-\infty}^{\infty} g^*(t)g(t+\tau)\,dt$$

Existe, na verdade, uma relação muito importante entre a função de autocorrelação de um sinal e sua ESD. Especificamente, a função de autocorrelação de um sinal $g(t)$ e sua ESD $\Psi_g(f)$ formam um par de transformadas de Fourier:

$$\psi_g(\tau) \Longleftrightarrow \Psi_g(f) \tag{3.73a}$$

Logo,

$$\Psi_g(f) = \mathcal{F}\left\{\psi_g(\tau)\right\} = \int_{-\infty}^{\infty} \psi_g(\tau) e^{-j2\pi f\tau}\, d\tau \tag{3.73b}$$

$$\psi_g(\tau) = \mathcal{F}^{-1}\left\{\Psi_g(f)\right\} = \int_{-\infty}^{\infty} \Psi_g(f) e^{-j2\pi f\tau}\, df \tag{3.73c}$$

Notemos que a transformada de Fourier da Eq. (3.73a) é calculada em relação a τ, e não em relação a t.

A seguir, mostraremos que a ESD $\Psi_g(f) = |G(f)|^2$ é a transformada de Fourier da função de autocorrelação $\psi_g(\tau)$. Embora esse resultado seja mostrado aqui para sinais de valores reais, também é válido para sinais de valores complexos. Devemos notar que a função de autocorrelação é uma função de τ e não de t. Portanto, sua transformada de Fourier é $\int \psi_g(\tau) e^{-j2\pi f\tau} d\tau$. Assim,

$$\begin{aligned}\mathcal{F}[\psi_g(\tau)] &= \int_{-\infty}^{\infty} e^{-j2\pi f\tau} \left[\int_{-\infty}^{\infty} g(t) g(t+\tau)\, dt\right] d\tau \\ &= \int_{-\infty}^{\infty} g(t) \left[\int_{-\infty}^{\infty} g(\tau + t) e^{-j2\pi f\tau} d\tau\right] dt\end{aligned}$$

A integral interna é a transformada de Fourier de $g(\tau + t)$, que é $g(\tau)$ deslocado para a esquerda por t. Portanto, a transformada é dada por $G(f)e^{j2\pi ft}$, segundo a propriedade de translação no tempo, Eq. (3.32a). Portanto,

$$\mathcal{F}[\psi_g(\tau)] = G(f) \int_{-\infty}^{\infty} g(t) e^{j2\pi ft} dt = G(f)G(-f) = |G(f)|^2$$

Isto completa a prova de que

$$\psi_g(\tau) \Longleftrightarrow \Psi_g(f) = |G(f)|^2 \tag{3.74}$$

Uma análise cuidadosa da operação de correlação revela uma relação próxima com convolução. De fato, a função de autocorrelação $\psi_g(\tau)$ é a convolução de $g(\tau)$ com $g(-\tau)$, pois

$$g(\tau) * g(-\tau) = \int_{-\infty}^{\infty} g(x) g[-(\tau - x)]\, dx = \int_{-\infty}^{\infty} g(x) g(x - \tau)\, dx = \psi_g(\tau)$$

Aplicação da propriedade de convolução no domínio do tempo [Eq. (3.44)] a esse resultado produz a Eq. (3.74).

ESD da Entrada e da Saída

Sejam $x(t)$ e $y(t)$, respectivamente, o sinal de entrada e o correspondente sinal de saída de um sistema linear invariante no tempo (LIT); logo,

$$Y(f) = H(f)X(f)$$

Portanto,

$$|Y(f)|^2 = |H(f)|^2 |X(f)|^2$$

Isso mostra que

$$\Psi_y(f) = |H(f)|^2 \Psi_x(f) \tag{3.75}$$

Assim, a ESD do sinal de saída é $|H(f)|^2$ vezes a ESD do sinal de entrada.

3.8 POTÊNCIA E DENSIDADE ESPECTRAL DE POTÊNCIA DE SINAIS

Para um sinal de potência, uma medida razoável de seu tamanho é a potência [definida na Eq. (2.4)], tomada como o valor médio da energia do sinal em um intervalo de tempo infinito. A potência P_g de um sinal de valores reais $g(t)$ é dada por

$$P_g = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} g^2(t)\, dt \tag{3.76}$$

A potência do sinal e os conceitos a ela associados podem ser entendidos com facilidade se definirmos um sinal truncado $g_T(t)$ como

$$g_T(t) = \begin{cases} g(t) & |t| \le T/2 \\ 0 & |t| > T/2 \end{cases}$$

O sinal truncado é ilustrado na Fig. 3.36. A integral no lado direito da Eq. (3.76) corresponde a E_{gT}, energia do sinal truncado $g_T(t)$. Assim,

$$P_g = \lim_{T\to\infty} \frac{E_{gT}}{T} \tag{3.77}$$

Esta equação descreve a relação entre potência e energia de sinais não periódicos. O entendimento desta relação auxiliará a compreensão de todos os conceitos associados a potência e energia, assim como o relacionamento entre eles. Como a potência de um sinal é o valor médio de sua energia, todos os conceitos e resultados relativos à energia de um sinal também se aplicam à potência do sinal, bastando que modifiquemos os conceitos de forma adequada, tomando seus valores médios.

3.8.1 Densidade Espectral de Potência (PSD)

Caso um sinal $g(t)$ seja um sinal de potência, sua potência é finita, e o sinal truncado $g_T(t)$ é um sinal de energia desde que T permaneça finito. Se $g_T(t) \Longleftrightarrow G_T(f)$, do teorema de Parseval, temos

$$E_{gT} = \int_{-\infty}^{\infty} g_T^2(t)\, dt = \int_{-\infty}^{\infty} |G_T(f)|^2\, df$$

Portanto, P_g, a potência de $g(t)$, fica dada por

$$P_g = \lim_{T\to\infty} \frac{E_{gT}}{T} = \lim_{T\to\infty} \frac{1}{T} \left[\int_{-\infty}^{\infty} |G_T(f)|^2\, df \right] \tag{3.78}$$

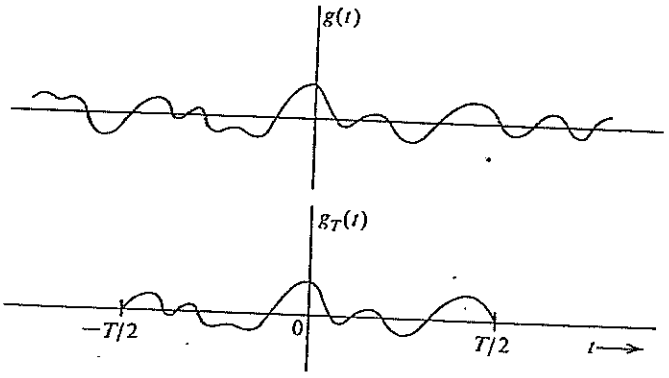


Figura 3.36 Processo de limite para o cálculo da PSD.

À medida que T aumenta, a duração de $g_T(t)$ aumenta e sua energia E_{gT} aumenta proporcionalmente. Isso significa que $|G_T(f)|^2$ também aumenta com T e, quando $T \to \infty$, $|G_T(f)|^2$ tende a ∞. Contudo, $|G_T(f)|^2$ deve tender a ∞ com a mesma taxa que T, pois, para um sinal de potência, o lado direito da Eq. (3.78) deve convergir. Essa convergência permite que troquemos a ordem do processo de limite e da integração na Eq. (3.78) e obtenhamos

$$P_g = \int_{-\infty}^{\infty} \lim_{T\to\infty} \frac{|G_T(f)|^2}{T}\, df \tag{3.79}$$

Definimos a **densidade espectral de potência (PSD** — *power spectral density*) $S_g(\omega)$ como

$$S_g(f) = \lim_{T\to\infty} \frac{|G_T(f)|^2}{T} \tag{3.80}$$

Logo,*

$$P_g = \int_{-\infty}^{\infty} S_g(f)\, df \tag{3.81a}$$

$$= 2\int_{0}^{\infty} S_g(f)\, df \tag{3.81b}$$

Esse resultado reproduz o obtido [Eq. (3.69a)] para sinais de energia. A potência é a área sob a curva da PSD. Vale notar que a PSD é o valor médio temporal da ESD de $g_T(t)$ [Eq. (3.80)].

A PSD, assim como a ESD, é uma função par de f, com valores reais e positivos. Se $g(t)$ for um sinal de tensão, a unidade da PSD será volt quadrado por hertz.

3.8.2 Função de Autocorrelação Temporal de Sinais de Potência

A função de autocorrelação (temporal) $\mathcal{R}_g(\tau)$ de um sinal de potência de valores reais $g(t)$ é definida como**

$$\mathcal{R}_g(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g(t)g(t-\tau)\, dt \tag{3.82a}$$

Podemos lançar mão do mesmo argumento que usamos no caso de sinais de energia [Eqs. (3.72b) e (3.72c)] para mostrar que $\mathcal{R}_g(\tau)$ é uma função par de τ. Isso significa que, para $g(t)$ de valores reais,

$$\mathcal{R}_g(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g(t)g(t+\tau)\, dt \tag{3.82b}$$

e

$$\mathcal{R}_g(\tau) = \mathcal{R}_g(-\tau) \tag{3.83}$$

Para sinais de energia, a ESD $\Psi_g(f)$ é a transformada de Fourier da função de autocorrelação $\psi_g(\tau)$. Um resultado similar se aplica a sinais de potência. Agora, mostraremos que, para um sinal de potência, a PSD $S_g(f)$ é a transformada de Fourier da função de autocorrelação $\mathcal{R}_g(\tau)$. A partir da Eq. (3.82b) e da Fig. 3.36,

$$\mathcal{R}_g(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-\infty}^{\infty} g_T(t)g_T(t+\tau)\, dt = \lim_{T\to\infty} \frac{\psi_{gT}(\tau)}{T} \tag{3.84}$$

---

*Devemos ser cautelosos no uso de uma expressão unilateral, como $P_g = 2\int_0^\infty S_g(f)\, df$, quando $S_g(f)$ contém um impulso na origem (uma componente dc). O impulso não deve ser multiplicado pelo fator 2.

**Para um sinal de valores complexos, definimos

$$\mathcal{R}_g(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g(t)g^*(t-\tau)\, dt = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g^*(t)g(t+\tau)\, dt$$

Recordemos, do teorema de Wiener-Khintchiné, que $\psi_{gT}(\tau) \Longleftrightarrow |G_T(f)|^2$. Logo, a transformada de Fourier da equação anterior nos fornece

$$\mathcal{R}_g(\tau) \Longleftrightarrow \lim_{T\to\infty} \frac{|G_T(f)|^2}{T} = S_g(f) \tag{3.85}$$

Embora tenhamos provado esses resultados para um sinal $g(t)$ de valores reais, as Eqs. (3.80), (3.81a), (3.81b) e (3.85) são igualmente válidas para sinais $g(t)$ de valores complexos.

O conceito e relações associados à potência de sinal reproduzem os associados à energia de sinal. Isso fica claro na Tabela 3.3.

Tabela 3.3

$E_g = \int_{-\infty}^{\infty} g^2(t)\,dt$	$P_g = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g^2(t)\,dt = \lim_{T\to\infty} \frac{E_{gT}}{T}$
$\psi_g(\tau) = \int_{-\infty}^{\infty} g(t)g(t+\tau)\,dt$	$\mathcal{R}_g(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g(t)g(t+\tau)\,dt = \lim_{T\to\infty} \frac{\psi_{gT}(\tau)}{T}$
$\Psi_g(f) = \lvert G(f)\rvert^2$	$S_g(f) = \lim_{T\to\infty} \frac{\lvert G_T(f)\rvert^2}{T} = \lim_{T\to\infty} \frac{\Psi_{gT}(f)}{T}$
$\psi_g(\tau) \Longleftrightarrow \Psi_g(f)$	$\mathcal{R}_g(\tau) \Longleftrightarrow S_g(f)$
$E_g = \int_{-\infty}^{\infty} \Psi_g(f)\,df$	$P_g = \int_{-\infty}^{\infty} S_g(f)\,df$

A Potência de um Sinal É Seu Valor Quadrático Médio

Um exame da Eq. (3.76) revela que a potência do sinal é o valor médio temporal de seu valor quadrático. Em outras palavras, P_g é o valor quadrático médio de $g(t)$. Contudo, devemos recordar que este é um valor médio temporal, e não estatístico (a ser discutido em capítulos posteriores). Valores médios estatísticos são denotados por uma barra. Assim, o valor quadrático médio (estatístico) de uma variável x é denotado por $\overline{x^2}$. Usaremos uma barra ondulada para denotar o valor médio temporal e distingui-lo do valor médio estatístico. Assim, o valor quadrático médio temporal de $g(t)$ será denotado por $\widetilde{g^2(t)}$. Valores médios temporais são, em geral, representados como $\langle g^2(t)\rangle$. No entanto, usaremos a notação da barra ondulada, pois é a associação de valor médio com barra ondulada é mais simples que com colchetes. Com essa notação, vemos que

$$P_g = \widetilde{g^2(t)} = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g^2(t)\,dt \tag{3.86a}$$

O valor rms de um sinal é a raiz quadrada de seu valor quadrático médio; logo

$$[g(t)]_{\text{rms}} = \sqrt{P_g} \tag{3.86b}$$

Da Eq. (3.82), fica claro que, para um sinal $g(t)$ de valores reais, a função de autocorrelação temporal $\mathcal{R}_g(\tau)$ é o valor médio temporal de $g(t)g(t \pm \tau)$. Assim,

$$\mathcal{R}_g(\tau) = \widetilde{g(t)g(t \pm \tau)} \tag{3.87}$$

Esta discussão também explica por que dizemos "autocorrelação temporal" em vez de "autocorrelação". O objetivo é distinguir, de forma clara, essa função de autocorrelação (um valor médio temporal) da função de autocorrelação estatística (um valor médio estatístico) a ser introduzida no Capítulo 9, no contexto de teoria da probabilidade e processos aleatórios.

Interpretação da Densidade Espectral de Potência

Como a PSD é o valor médio temporal da ESD de $g(t)$, podemos seguir a mesma argumentação usada na interpretação da ESD. Podemos mostrar prontamente que a PSD $S_g(f)$ representa a potência por unidade de largura de banda (em hertz) das componentes espectrais na frequência f. A contribuição das componentes de frequência na banda de f_1 a f_2 à potência é dada por

$$\Delta P_g = 2\int_{f_1}^{f_2} S_g(f)\,df \tag{3.88}$$

Método de Autocorrelação: Uma Ferramenta Poderosa

Para um sinal $g(t)$, a ESD, que é igual a $|G(f)|^2$, também pode ser obtida da transformada de Fourier da correspondente função de autocorrelação. Se a transformada de Fourier de um sinal basta para o cálculo da ESD, por que complicar desnecessariamente nossa vida com a função de autocorrelação? A razão para seguir esse caminho alternativo é estabelecer a base para tratar de sinais de potência e sinais aleatórios. A transformada de Fourier de um sinal de potência, em geral, não existe. Além disso, o luxo de poder calcular a transformada de Fourier só é possível para sinais determinísticos, que podem ser descritos como funções do tempo. Sinais de mensagem aleatórios que ocorrem em problemas de comunicação (por exemplo, uma sequência aleatória de pulsos binários) não podem ser descritos como funções do tempo, o que torna impossível o cálculo de suas transformadas de Fourier. Contudo, a função de autocorrelação de sinais deste tipo pode ser calculada a partir da correspondente informação estatística. Isso permite que determinemos a PSD (informação espectral) desses sinais. De fato, podemos considerar a abordagem da autocorrelação como a generalização de técnicas de Fourier a sinais de potência e sinais aleatórios. O exemplo a seguir, de um trem aleatório de pulsos binários, ilustra, de forma dramática, o poder dessa técnica.

Exemplo 3.19 A Fig. 3.37a mostra um trem aleatório de pulsos binários $g(t)$. A largura do pulso é $T_b/2$ e um dígito binário é transmitido a cada T_b segundos. Um **1** binário é transmitido por um pulso positivo, e um **0** binário, por um pulso negativo. Os dois símbolos têm a mesma probabilidade de ocorrência, que é aleatória. Determinemos a função de autocorrelação, a PSD e a largura de banda essencial desse sinal.

Não podemos descrever esse sinal como uma função do tempo, pois sua forma de onda, por ser aleatória, não é conhecida. Podemos, no entanto, conhecer seu comportamento por meio de valores médios (informação estatística). A função de autocorrelação, por ser um parâmetro de valor médio (temporal) do sinal, é determinada da informação estatística (média) fornecida. Temos [Eq. (3.82a)]

$$\mathcal{R}_g(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} g(t)g(t-\tau)\,dt$$

Na Fig. 3.37b, as linhas sólidas representam $g(t)$ e as tracejadas, $g(t-\tau)$, que é $g(t)$ atrasado por τ. Para obter o valor integral no lado direito da equação anterior, multiplicamos $g(t)$ por $g(t-\tau)$, calculamos a área sob a curva do produto $g(t)g(t-\tau)$ e dividimos o resultado pela largura T do intervalo. Admitamos que há N bits (pulsos) neste intervalo T, de modo que $T = NT_b$ e $N \to \infty$ quando $T \to \infty$. Assim,

$$\mathcal{R}_g(\tau) = \lim_{N\to\infty} \frac{1}{NT_b}\int_{-NT_b/2}^{NT_b/2} g(t)g(t-\tau)\,dt$$

Consideremos, primeiro, o caso $\tau < T_b/2$, em que todos os pulsos de $g(t)$ e de $g(t-\tau)$ se sobrepõem (região hachurada). A área sob a curva do produto $g(t)g(t-\tau)$ é, para cada pulso,

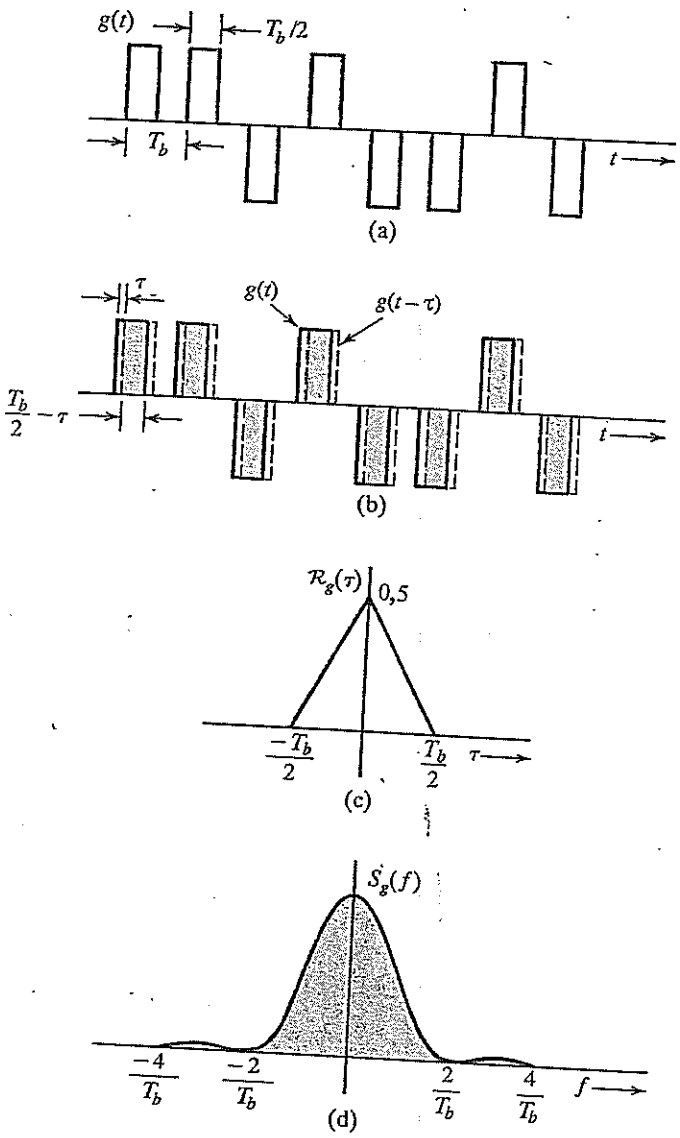


Figura 3.37 Função de autocorrelação e função de densidade espectral de potência de um trem aleatório de pulsos binários.

$T_b/2 - \tau$. Como há N pulsos no intervalo em que o valor médio é calculado, a área total sob $g(t)$ $g(t-\tau)$ é $N(T_b/2 - \tau)$, e

$$\mathcal{R}_g(\tau) = \lim_{N\to\infty} \frac{1}{NT_b}\left[N\left(\frac{T_b}{2} - \tau\right)\right]$$
$$= \frac{1}{2}\left(1 - \frac{2\tau}{T_b}\right) \qquad \tau < \frac{T_b}{2}$$

Como $\mathcal{R}_g(\tau)$ é uma função par de τ,

$$\mathcal{R}_g(\tau) = \frac{1}{2}\left(1 - \frac{2|\tau|}{T_b}\right) \qquad |\tau| < \frac{T_b}{2} \tag{3.89a}$$

como mostrado na Fig. 3.37c.

À medida que aumentamos o valor de τ além de $T_b/2$, pulsos adjacentes passarão a se sobrepor. Dois pulsos que se sobrepõem têm igual probabilidade de serem de mesma polaridade ou de polaridades opostas. Portanto, no intervalo de sobreposição, seu produto tem igual probabilidade de ser 1 ou -1. Na média, metade dos produtos de pulsos será 1 (combinações positivo-positivo ou negativo-negativo de pulsos), e metade será -1 (combinações positivo-negativo ou negativo-positivo de pulsos). Em consequência, a área sob a curva de $g(t)g(t-\tau)$ será zero, se a média for calculada em um intervalo de tempo de duração infinita ($T \to \infty$), e

$$\mathcal{R}_g(\tau) = 0 \qquad |\tau| > \frac{T_b}{2} \tag{3.89b}$$

As duas partes da Eq. (3.89) mostram que, neste caso, a função de autocorrelação é a função triangular $\frac{1}{2}\,\Delta(t/T_b)$, obtida no Exemplo 3.13 (ou na Tabela 3.1, par 19), pois

$$S_g(f) = \frac{T_b}{4}\,\mathrm{sinc}^2\left(\frac{\pi f T_b}{2}\right) \tag{3.90}$$

A PSD é o quadrado da função sinc, como mostrado na Fig. 3.37d. Do resultado no Exemplo 3.18, concluímos que 90,28% da área deste espectro estão contidos na banda de 0 a $4\pi/T_b$ rad/s ou de 0 a $2/T_b$ Hz. Assim, a largura de banda essencial pode ser tomada como $2/T_b$ Hz (presumindo um critério de 90%). Este exemplo ilustra, de forma dramática, como a função de autocorrelação pode ser usada para obter informação espectral de um sinal (aleatório) quando métodos convencionais de obtenção da transformada de Fourier não forem aplicáveis.

3.8.3 Densidades Espectrais de Potência de Entrada e de Saída

Como a PSD é o valor médio temporal da ESD, a relação entre as PSDs dos sinais de entrada e de saída de um sistema linear invariante no tempo (LIT) é similar à existente entre as correspondentes ESDs. Seguindo os argumentos usados no caso da ESD [Eq. (3.75)], podemos mostrar que, se $g(t)$ e $y(t)$ forem os sinais de entrada e de saída, respectivamente, de um sistema LIT com função de transferência $H(f)$, então,

$$S_y(f) = |H(f)|^2 S_g(f) \tag{3.91}$$

Exemplo 3.20 Um sinal de ruído $n_i(t)$, com PSD $S_{ni}(f) = K$ é aplicado à entrada de um diferenciador ideal (Fig. 3.38a). Determinemos a PSD e a potência do sinal de ruído de saída, $n_o(t)$.

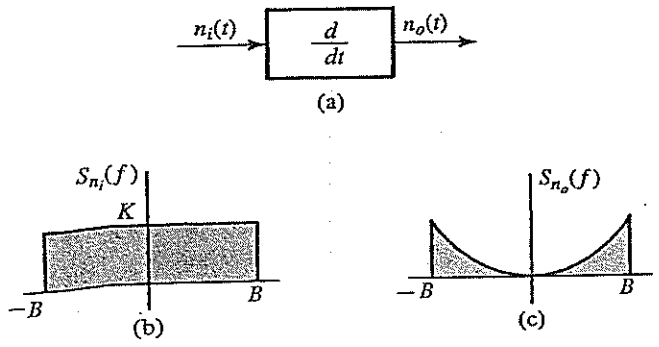


Figura 3.38 Densidades espectrais dos sinais de entrada e de saída de um diferenciador ideal.

A função de transferência de um diferenciador ideal é $H(f) = j2\pi f$. Se o ruído na saída do diferenciador for $n_o(t)$, da Eq. (3.91), temos

$$S_{n_o}(f) = |H(f)|^2 S_{n_i}(f) = |j2\pi f|^2 K$$

A PSD do sinal de saída $S_{no}(f)$ é parabólica, como indicado na Fig. 3.38c. A potência de ruído de saída N_o é a área sob a curva da PSD de saída. Logo,

$$N_o = \int_{-B}^{B} K(2\pi f)^2\, df = 2K \int_0^B (2\pi f)^2\, df = \frac{8\pi^2 B^3 K}{3}$$

3.8.4 PSD de Sinais Modulados

Seguindo o mesmo argumento empregado na dedução das Eqs. (3.70) e (3.71) para sinais de energia, podemos, no caso de sinais de potência, obter resultados similares se calcularmos valores médios. Podemos mostrar que, para um sinal de potência $g(t)$, se

$$\varphi(t) = g(t) \cos 2\pi f_0 t$$

a PSD $S_\varphi(f)$ de um sinal modulado $\varphi(t)$ é dada por

$$S_\varphi(f) = \frac{1}{4}\left[S_g(f + f_0) + S_g(f - f_0)\right] \tag{3.92}$$

Detalhes da dedução desse resultado são dados na Seção 7.8. Assim, a modulação desloca a PSD de $g(t)$ por $\pm f_0$. A potência de $\varphi(t)$ é a metade da potência de $g(t)$, ou seja,

$$P_\varphi = \frac{1}{2} P_g \qquad f_0 \geq B \tag{3.93}$$

3.9 CÁLCULO NUMÉRICO DA TRANSFORMADA DE FOURIER: A DFT

Para efetuar o cálculo numérico de $G(f)$, a transformada de Fourier de $g(t)$, devemos utilizar amostras de $g(t)$. Além disso, conseguimos calcular $G(f)$ apenas em número discreto de frequências. Assim, podemos calcular somente amostras de $G(f)$. Por isso, a seguir, estabeleceremos algumas relações entre amostras de $g(t)$ e de $G(f)$.

Em cálculos numéricos, os dados devem ser finitos. Isso significa que o número de amostras de $g(t)$ e de $G(f)$ deve ser finito. Em outras palavras, devemos lidar com sinais limitados no tempo. Se um sinal não for limitado no tempo, é necessário truncá-lo para que tenha duração finita. O mesmo se aplica a $G(f)$. Consideremos, inicialmente, um sinal $g(t)$ de duração τ segundo e início em $t = 0$, como mostrado na Fig. 3.39a. Contudo, por motivos que se tornarão claros mais adiante, consideremos que a duração de $g(t)$ seja T_0, com $T_0 \geq \tau$, de modo que $g(t) = 0$ no intervalo $\tau < t \leq T_0$, como mostrado na Fig. 3.39a. Obviamente, isso em nada afeta o cálculo de $G(f)$. Tomemos amostras de $g(t)$ a intervalos uniformes de T_s segundos, o que produz um total de N_0 amostras, com

$$N_0 = \frac{T_0}{T_s} \tag{3.94}$$

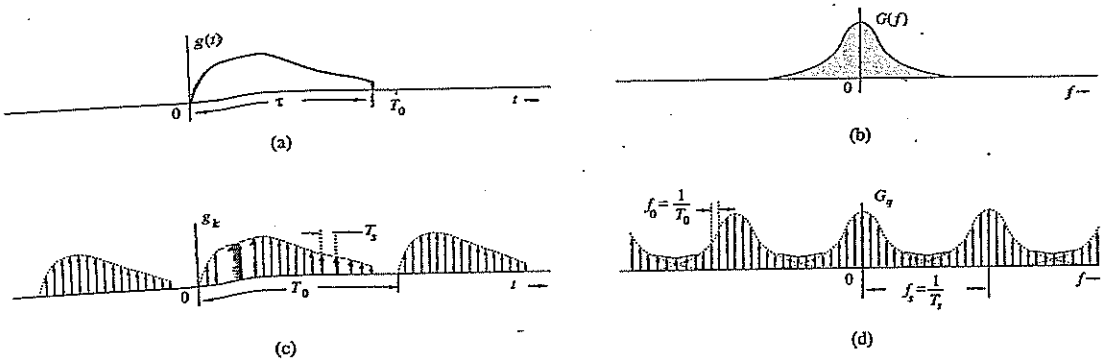


Figura 3.39 Relação entre amostras de $g(t)$ e de $G(f)$.

Mas,*

$$G(f) = \int_0^{T_0} g(t)e^{-j2\pi ft}\,dt$$

$$= \lim_{T_s \to 0} \sum_{k=0}^{N_0-1} g(kT_s)e^{-j2\pi fkT_s}\,T_s \tag{3.95}$$

Tomemos amostras de $G(f)$ a intervalos uniformes de largura f_0. Seja G_q a q-ésima amostra, ou seja, $G_q = G(qf_0)$; com isso, a partir da Eq. (3.95), temos

$$G_q = \sum_{k=0}^{N_0-1} T_s g(kT_s)e^{-jq2\pi f_0 T_s k}$$

$$= \sum_{k=0}^{N_0-1} g_k e^{-jq\Omega_0 k} \tag{3.96}$$

em que

$$g_k = T_s g(kT_s), \qquad G_q = G(qf_0), \qquad \Omega_0 = 2\pi f_0 T_s \tag{3.97}$$

Portanto, a Eq. (3.96) relaciona as mostras de $g(t)$ às amostras de $G(f)$. Nesta dedução, supomos que $T_s \to 0$. Na prática, não é possível fazer $T_s \to 0$, o que aumentaria a quantidade de dados demasiadamente. Devemos nos esforçar para fazer T_s tão pequeno quanto possível, do ponto de vista prático. Isso resultará em algum erro computacional.

Cabe uma interessante observação à Eq. (3.96). As amostras de G_q são periódicas, com período $2\pi/\Omega_0$. Isto resulta da Eq. (3.96), que mostra que $G_{(q+2\pi/\Omega_0)} = G_q$. Assim, apenas $2\pi/\Omega_0$ amostras G_q podem ser independentes. A Eq. (3.96) mostra que G_q é determinada por N_0 valores independentes de g_k. Portanto, para que exista uma única inversa para a Eq. (3.96), deve haver somente N_0 amostras independentes de G_q. Isso significa que

$$N_0 = \frac{2\pi}{\Omega_0} = \frac{2\pi}{2\pi f_0 T_s} = \frac{2\pi N_0}{2\pi f_0 T_0} \tag{3.98}$$

Em outras palavras, temos

$$2\pi f_0 = \frac{2\pi}{T_0} \qquad \text{e} \qquad f_0 = \frac{1}{T_0} \tag{3.99}$$

Portanto, o intervalo de amostragem espectral f_0 Hz pode ser ajustado por meio de uma escolha adequada de T_0: maior T_0, menor f_0. Com isto, fica clara a motivação para escolher $T_0 \geq \tau$. Quando T_0

*O limite superior do somatório na Eq. (3.95) é $N_0 - 1$ (não N_0), pois o último termo do mesmo tem início em $(N_0 - 1)T_s$ e cobre a área sob a curva até $N_0 T_s = T_0$.

é maior que τ, temos várias amostras g_k com valor zero no intervalo entre τ e T_0. Assim, ao aumentar o número de amostras g_k com valor zero, reduzimos f_0 [amostras de $G(f)$ mais próximas umas das outras], produzindo mais detalhes de $G(f)$. Esse processo de redução de f_0 com a inclusão de amostras g_k com valor zero é conhecido como **preenchimento com zero**. Adicionalmente, para um dado intervalo de amostragem T_0, um maior valor de T_0 implica maior N_0. Ou seja, com a escolha de um valor suficientemente grande para N_0, podemos obter amostras de $G(f)$ tão próximas quanto possível.

Para determinar a relação inversa, multiplicamos os dois lados da Eq. (3.96) por $e^{jm\Omega_0 q}$ e efetuamos a soma em q:

$$\sum_{q=0}^{N_0-1} G_q e^{jm\Omega_0 q} = \sum_{q=0}^{N_0-1} \left[\sum_{k=0}^{N_0-1} g_k e^{-jq\Omega_0 k} \right] e^{jm\Omega_0 q}$$

Trocando a ordem das somas no lado direito, obtemos

$$\sum_{q=0}^{N_0-1} G_q e^{jm\Omega_0 q} = \sum_{k=0}^{N_0-1} g_k \left[\sum_{q=0}^{N_0-1} e^{j(m-k)\Omega_0 q} \right] \tag{3.100}$$

Para calcular o valor da soma interna no lado direito, mostraremos que

$$\sum_{k=0}^{N_0-1} e^{jn\Omega_0 k} = \begin{cases} N_0 & n = 0, \pm N_0, \pm 2N_0, \ldots \\ 0 & \text{para todos os outros valores de } n \end{cases} \tag{3.101}$$

Para mostrar isto, recordemos que $\Omega_0 N_0 = 2\pi$ e $e^{jn\Omega_0 k} = 1$ para $n = 0, \pm N_0, \pm 2N_0, \ldots$, de modo que

$$\sum_{k=0}^{N_0-1} e^{jn\Omega_0 k} = \sum_{k=0}^{N_0-1} 1 = N_0 \qquad n = 0, \pm N_0, \pm 2N_0, \ldots$$

Para calcular a soma para outros valores de n, notemos que a soma no lado esquerdo da Eq. (3.101) é uma série geométrica de razão $\alpha = e^{jn\Omega_0}$. Portanto, a soma parcial dos primeiros N_0 termos é

$$\sum_{k=0}^{N_0-1} e^{jn\Omega_0 k} = \frac{e^{jn\Omega_0 N_0} - 1}{e^{jn\Omega_0} - 1} = 0,$$

com

$$e^{jn\Omega_0 N_0} = e^{j2\pi n} = 1$$

Isso prova a Eq. (3.101).

Assim, a soma interna no lado direito da Eq. (3.100) é igual a zero para $k \neq m$, e igual a N_0 para $k = m$. Portanto, a soma externa terá somente um termo não zero quando $k = m$, e este termo será $N_0 g_k = N_0 g_m$. Logo,

$$g_m = \frac{1}{N_0} \sum_{q=0}^{N_0-1} G_q e^{jm\Omega_0 q} \qquad \Omega_0 = \frac{2\pi}{N_0} \tag{3.102}$$

A Eq. (3.102) revela um fato interessante: que $g_{(m+N_0)} = g_m$. Isso significa que a sequência g_k também é periódica, com período N_0 (correspondendo à duração temporal $N_0 T_s = T_0$ segundos). Adicionalmente, G_q também é periódica, com período N_0, o que corresponde ao intervalo de frequência $N_0 f_0 = (T_0/T_s)(T_0) = 1/T_s = f_s$ hertz. Notemos que $1/T_s$ é o número de amostras de $g(t)$ por segundo. Assim, $1/T_s = f_s$ é a frequência de amostragem (em hertz) de $g(t)$. Em outras palavras, G_q é periódica,

com período N_0 e se repete a cada f_s Hz. Resumamos os resultados obtidos até aqui. Provamos o par de transformadas de Fourier discretas (DFT – *discrete Fourier transform*).

$$G_q = \sum_{k=0}^{N_0-1} g_k e^{-jq\Omega_0 k} \tag{3.103a}$$

$$g_k = \frac{1}{N_0} \sum_{q=0}^{N_0-1} G_q e^{jk\Omega_0 q} \tag{3.103b}$$

com

$$\begin{aligned} g_k &= T_s g(kT_s) & G_q &= G(qf_0) \\ 2\pi f_0 &= \frac{2\pi}{T_0} & 2\pi f_s &= \frac{2\pi}{T_s} \\ N_0 &= \frac{T_0}{T_s} = \frac{f_s}{f_0} & \Omega_0 &= 2\pi f_0 T_s = \frac{2\pi}{N_0} \end{aligned} \tag{3.104}$$

As duas sequências g_k e G_q são periódicas, com período N_0. Isso implica que g_k se repete com um período de T_0 segundos e G_q, com um período de $f_s = 1/T_s$ Hz (a frequência de amostragem). O intervalo de amostragem de g_k é de T_s segundos e o intervalo de amostragem de G_q é de $f_0 = 1/T_0$ Hz, como indicado na Fig. 3.39c e d. Por conveniência, usamos a variável de frequência f (em hertz) em vez de ω (em radianos por segundo).

Presumimos que $g(t)$ é limitado a τ segundos. Isso torna $G(f)$ não limitado em frequência.* Em consequência, a repetição do espectro G_q, como mostrado na Fig. 3.39d, causará sobreposição de componentes espectrais, resultando em erro. A natureza deste erro, conhecido como **erro de mascaramento** (*aliasing error*), é explicado em detalhes no Capítulo 6. O espectro de G_q se repete a cada f_s Hz. O erro de mascaramento é reduzido com o aumento de f_s, a frequência de repetição (ver Fig. 3.39d). Em resumo, o cálculo de G_q por DFT tem erro de mascaramento quando $g(t)$ é limitado no tempo. Este erro pode ser feito tão pequeno quanto desejado, com o aumento da frequência de amostragem $f_s = 1/T_s$ (ou com a redução do intervalo de amostragem T_s). O erro de mascaramento é um resultado direto da não satisfação, na Eq. (3.95), da exigência $T_s \to 0$.

Quando $g(t)$ não é limitado no tempo, precisamos truncá-lo para torná-lo limitado no tempo. Isso provocará erros adicionais em G_q. Esse erro pode ser reduzido tanto quanto desejado, por meio do aumento adequado do intervalo de truncamento T_0.**

No cálculo da transformada de Fourier inversa [por meio da DFT inversa na Eq. (3.103b)], encontramos problemas semelhantes. Se $G(f)$ for limitado em frequência, $g(t)$ não será limitado no tempo e haverá sobreposição das repetições periódicas das amostras g_k (mascaramento no domínio do tempo). Podemos reduzir o erro de mascaramento se aumentarmos T_0, o período de g_k (em segundos). Isso equivale a reduzir o intervalo $f_0 = 1/T_0$ entre as frequências em que amostras de $G(f)$ são colhidas. Caso $G(f)$ não seja limitado em frequência, será necessário truncá-lo. Isto causará erro adicional no cálculo de g_k. Com o aumento da largura de banda de truncamento, podemos reduzir esse erro. Na prática, funções (janelas) graduais são usadas para o truncamento[5], para reduzir a severidade de alguns problemas provocados por funções de truncamento abrupto (também conhecidas como janela retangular).

*Podemos mostrar que um sinal não pode ser simultaneamente limitado no tempo e na frequência. Se for um, não pode ser o outro, e vice-versa.[3]

** As relações de DFT também representam um par de transformadas e são exatas. Se, no entanto, associarmos g_k e G_q às amostras de um sinal $g(t)$ e de sua transformada de Fourier $G(f)$, respectivamente, as relações de DFT se tornam aproximações, devido aos efeitos de mascaramento e truncamento.

Como G_q é periódica, com período N_0, precisamos determinar os valores de G_q em qualquer período. É comum, para a determinação de G_q, considerar o período no intervalo $(0, N_0 - 1)$, em vez de no intervalo $(-N_0/2, N_0/2 - 1)$. Observação idêntica se aplica a g_k.

Escolha dos Valores de T_s, T_0 e N_0

No cálculo de DFT, primeiro, precisamos selecionar valores adequados para N_0, T_s e T_0. Para isso, devemos antes especificar o valor de B, a largura de banda essencial de $g(t)$. Da Fig. 3.39d, fica claro que a sobreposição espectral (mascaramento) ocorre na frequência $f_s/2$ Hz. Essa sobreposição espectral também pode ser interpretada como se o espectro além de $f_s/2$ sofresse uma dobra em $f_s/2$. Por conseguinte, essa frequência também recebe a denominação **frequência de dobramento**. Se a frequência de dobramento for escolhida de modo que o espectro $G(f)$ além da mesma seja desprezível, o mascaramento (ou sobreposição espectral) não será relevante. Portanto, a frequência de dobramento deve ser pelo menos igual à maior frequência significativa, ou seja, a frequência além da qual $G(f)$ é desprezível. Chamaremos essa frequência **largura de banda essencial** B (em hertz). Caso $g(t)$ seja limitado em frequência, sua largura de banda será igual à largura de banda essencial. Assim,

$$\frac{f_s}{2} \geq B \quad \text{Hz} \tag{3.105a}$$

Dado que o intervalo de amostragem $T_s = 1/f_s$ [Eq. (3.104)], temos

$$T_s \leq \frac{1}{2B} \tag{3.105b}$$

Uma vez determinado o valor de B, podemos escolher o valor de T_s segundo a Eq. (3.105b). Como

$$f_0 = \frac{1}{T_0} \tag{3.106}$$

em que f_0 é a **resolução de frequência** [separação entre amostras de $G(f)$], caso f_0 seja dado, podemos selecionar o valor de T_0 segundo a Eq. (3.106). Conhecidos os valores de T_0 e T_s, determinamos o de N_0 de

$$N_0 = \frac{T_0}{T_s} \tag{3.107}$$

Em geral, se o sinal for limitado no tempo, $G(f)$ não será limitado em frequência e haverá mascaramento no cálculo de G_q. Para reduzir o efeito de mascaramento, precisamos aumentar a frequência de dobramento, ou seja, devemos reduzir o valor de T_s (intervalo de amostragem) tanto quanto praticamente possível. Se o sinal for limitado em frequência, $g(t)$ não será limitado no tempo, de modo que haverá mascaramento (sobreposição) no cálculo de g_k. Para reduzir este mascaramento, precisamos aumentar o valor de T_0, o período de g_k. Isso implica a redução do intervalo f_0 (em hertz) de amostragem em frequência. Em qualquer dos casos (redução de T_s, no caso de sinal limitado no tempo, ou aumento de T_0, no caso de sinal limitado em frequência), para maior precisão, precisamos aumentar o número de amostras N_0, pois $N_0 = T_0/T_s$. Existem, ainda, sinais que não são limitados nem no tempo nem em frequência. Para estes sinais, devemos reduzir T_s e aumentar T_0.

Pontos de Descontinuidade

Caso $g(t)$ tenha, em um ponto de amostragem, uma descontinuidade do tipo degrau, o valor da amostra deve ser tomado como a média dos valores nos dois lados da descontinuidade, pois a representação de Fourier em um ponto de descontinuidade converge para o valor médio.

Uso de Algoritmo de FFT no Cálculo de DFT

O número de contas necessário para o cálculo de uma DFT foi drasticamente reduzido em um algoritmo desenvolvido por Tukey e Cooley, em 1965.[6] Esse algoritmo, conhecido como **transformada**

de Fourier rápida* (FFT—*fast Fourier transform*), reduz o número de contas de algo da ordem de N_0^2 para $N_0 \log N_0$. Para calcular o valor de uma amostra G_r pela Eq. (3.103a), precisamos de N_0 multiplicações complexas e $N_0 - 1$ adições complexas. Para calcular N_0 valores de G_r ($r = 0, 1, ..., N_0 - 1$), precisamos de um total de N_0^2 multiplicações complexas e $N_0(N_0 - 1)$ adições complexas. Para grandes valores de N_0, isto pode exigir um tempo proibitivamente grande, mesmo com o uso de computador de alta velocidade. O algoritmo de FFT é um salva-vidas em aplicações de processamento de sinais. O algoritmo de FFT fica simplificado se escolhermos N_0 como uma potência de 2, embora isto não seja, em geral, necessário. Detalhes da FFT podem ser encontrados em qualquer livro sobre processamento de sinais (por exemplo, Ref. 3).

3.10 EXERCÍCIOS COM O MATLAB

Cálculo de Transformadas de Fourier

Nesta seção de exercícios baseados em computador, consideremos dois exemplos para ilustrar o uso de DFT no cálculo da transformada de Fourier. Usaremos MATLAB para calcular a DFT com o algoritmo de FFT. No primeiro exemplo, o sinal é $g(t) = e^{-2t}u(t)$, com início em $t = 0$, e no segundo, $g(t) = \Pi(t)$, com início em $t = -½$.

EXEMPLO COMPUTACIONAL C3.1

Empreguemos a DFT (implementada pelo algoritmo de FFT) para calcular a transformada de Fourier de $e^{-2t}u(t)$ e, a seguir, tracemos o gráfico do resultante espectro de Fourier.

Primeiro, devemos determinar T_s e T_0. A transformada de Fourier de $e^{-2t}u(t)$ é $1/(2\pi f + 2)$. Esse sinal passa-faixa não é limitado em frequência. Tomemos sua largura de banda essencial como a frequência em que $|G(f)|$ se torna igual à 1% do valor de pico, que ocorre em $f = 0$. Observemos que

$$|G(f)| = \frac{1}{\sqrt{(2\pi f)^2 + 4}} \approx \frac{1}{2\pi f} \qquad 2\pi f \gg 2$$

O pico de $|G(f)|$ ocorre em $f = 0$, em que $|G(0)| = 0{,}5$. Portanto, a largura de banda essencial B corresponde a $f = B$, com

$$|G(f)| \approx \frac{1}{2\pi B} = 0{,}5 \times 0{,}01 \quad \Rightarrow B = \frac{100}{\pi} \text{ Hz}$$

e, da Eq. (3.105b),

$$T_s \leq \frac{1}{2B} = 0{,}005\pi = 0{,}0157$$

Arredondemos esse valor para $T_s = 0{,}015625$ segundo, de modo que tenhamos 64 amostras por segundo. Agora, devemos determinar T_0. O sinal não é limitado no tempo. Precisamos truncá-lo em T_0, tal que $g(T_0) << 1$. Escolhamos $T_0 = 4$ (oito constantes de tempo do sinal), o que resulta em $N_0 = T_0/T_s = 256$, que é uma potência de 2. Vale ressaltar que há muita flexibilidade na determinação de T_s e T_0, dependendo da precisão desejada e da capacidade computacional disponível. Poderíamos ter escolhido $T_0 = 8$ e $T_s = 1/32$, o que também resultaria em $N_0 = 256$, mas implicaria um erro de mascaramento ligeiramente maior.

*No Brasil, é igualmente empregada a denominação "transformada rápida de Fourier". (N.T.)

Como o sinal tem uma descontinuidade do tipo degrau em $t = 0$, o valor da primeira amostra (em $t = 0$) é 0,5, média dos valores nos dois lados da descontinuidade. O programa de MATLAB que implementa a DFT com o algoritmo de FFT é o seguinte:

```
Ts=1/64; T0=4; N0=T0/Ts;
t=0:Ts:Ts*(N0-1);t=t';
g=Ts*exp(-2*t);
g(1)=Ts*0.5;
G=fft(g);
$[Gp,Gm]$=cart2pol($real(G),imag(G)$);
k=0:N0-1; k=k';
w=2*pi*k/T0;
subplot(211),stem(w(1:32),Gm(1:32));
subplot(212),stem(w(1:32),Gp(1:32))
```

Como G_q tem período N_0, $G_q = G_{(q+256)}$, de modo que $G_{256} = G_0$. Portanto, basta traçar o gráfico de G_q no intervalo $q = 0$ a $q = 255$ (e não 256). Além disso, devido à periodicidade, $G_{-q} = G_{(-q+256)}$, ou seja, os valores de G_q no intervalo $q = -127$ a $q = -1$ são idênticos aos valores de G_q no intervalo $q = 129$ a $q = 255$. Logo, $G_{-127} = G_{129}$, $G_{-126} = G_{130}, \ldots, G_{-1} = G_{255}$. Adicionalmente, devido à propriedade de simetria conjugada da transformada de Fourier, $G_{-q} = G_q^*$; assim, $G_{129} = G_{127}^*$, $G_{130} = G_{126}^*, \ldots, G_{255} = G_1^*$. Consequentemente, para sinais de valores reais, não é necessário marcar no gráfico os valores de G_q com q maior que $N_0/2$ (128, neste caso), pois são os complexos conjugados dos valores de G_q com $q = 0$ a 128.

O gráfico do espectro de Fourier na Fig. 3.40 mostra amplitude e fase das amostras de $G(f)$ tomadas em intervalos de $1/T_0 = 1/4$ Hz, ou $\omega_0 = 1{,}5708$ rad/s. Na Fig. 3.40, mostramos apenas os primeiros 28 pontos (em vez dos 128 pontos), para evitar o acúmulo excessivo de dados no gráfico.

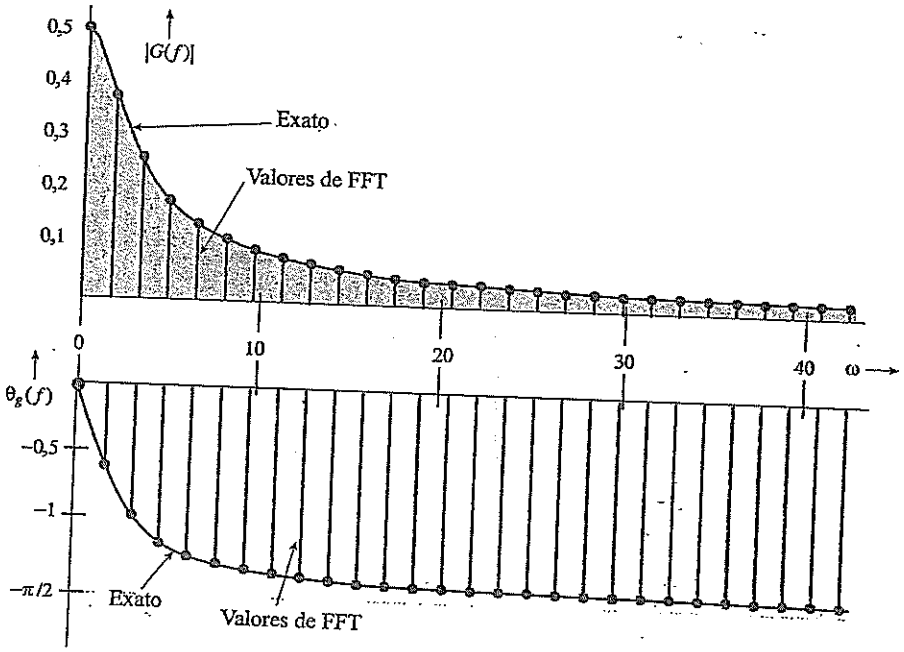


Figura 3.40 Transformada de Fourier discreta de um sinal exponencial $e^{-2t}u(t)$. O eixo horizontal é ω (em radianos por segundo).

Neste exemplo, dispúnhamos da expressão analítica de $G(f)$, o que nos permitiu fazer escolhas INTELIGENTES para B (ou frequência de amostragem f_s). Na prática, em geral, não conhecemos $G(f)$. Na verdade, isso é exatamente o que desejamos calcular. Nesses casos, para determinar B ou f_s, devemos lançar mão de evidências circunstanciais. Devemos, sucessivamente, reduzir o valor de T_s e calcular a transformada até que o resultado satisfaça o desejado número de algarismos significativos.

A seguir, calcularemos a transformada de Fourier de $g(t) = 8\,\Pi(t)$.

EXEMPLO COMPUTACIONAL C3.2

Empreguemos a DFT (implementada pelo algoritmo de FFT) para calcular a transformada de Fourier de $8\,\Pi(t)$ e tracemos o gráfico do resultante espectro de Fourier.

Essa função retangular e sua transformada de Fourier são mostradas na Fig. 3.41a e b. Para determinar o valor do intervalo de amostragem T_s, devemos, primeiro, definir a largura de banda essencial B. Da Fig. 3.41b, vemos que $G(f)$ decai lentamente com f. Consequentemente, a largura de banda essencial B é bastante grande. Por exemplo, em $B = 15{,}5$ Hz (97,39 rad/s), $G(f) = -0{,}1643$, o que corresponde a cerca de 2% do valor de pico, $G(0)$. Poderíamos, então, tomar a largura de banda essencial como 16 Hz. No entanto, deliberadamente, tomaremos $B =$ 4 Hz, por dois motivos: (1) mostrar o efeito de mascaramento e (2) o uso de $B > 4$ implicaria enorme número de amostras, que não poderiam ser mostradas de forma adequada em uma página de livro sem perda de detalhes fundamentais. Portanto, aceitaremos a aproximação para que possamos esclarecer conceitos de DFT por meio de gráficos.

A escolha $B = 4$ resulta em um intervalo de amostragem $T_s = 1/2B = 1/8$ segundos. Examinando novamente o espectro na Fig. 3.41b, vemos que a escolha da resolução de frequência $f_0 = 1/4$ Hz é razoável, e corresponde a quatro amostras em cada lóbulo de $G(f)$. Neste caso, $T_0 = 1/f_0 = 4$ segundos, e $N_0 = T_0/T_s = 32$. A duração de $g(t)$ é de apenas 1 segundo. Devemos repetir $g(t)$ a cada 4 segundos, como indicado na Fig. 3.41c, e tomar amostras a cada 0,125 segundo. Isso nos dará 32 amostras ($N_0 = 32$). Também temos

$$g_k = T_s g(kT)$$
$$= \frac{1}{8} g(kT)$$

Como $g(t) = 8\,\Pi(t)$, os valores de g_k são 1, 0 ou 0,5 (nos pontos de descontinuidade), como mostrado na Fig. 3.41c; nessa figura, por conveniência, g_k é mostrado como função de t e de k.

Na dedução da DFT, supomos que $g(t)$ tem início em $t = 0$ (Fig. 3.39a) e tomamos N_0 amostras no intervalo $(0, T_0)$. No caso em consideração, contudo, $g(t)$ tem início em $t = -½$. Essa dificuldade é facilmente resolvida quando observamos que a DFT obtida por este procedimento é, na verdade, a DFT de g_k repetido a cada T_0 segundos. Da Fig. 3.41c, fica claro que a repetição periódica do segmento de g_k no intervalo de -2 a 2 segundos é equivalente à repetição do segmento de g_k no intervalo de 0 a 4 segundos. Portanto, a DFT das amostras colhidas entre -2 e 2 segundos é igual à DFT das amostras colhidas entre 0 e 4 segundos. Assim, independentemente do instante em que $g(t)$ tem início, sempre podemos tomar as amostras de $g(t)$ e repetilas periodicamente no intervalo de 0 a T_0. No presente exemplo, os valores das 32 amostras são

$$g_k = \begin{cases} 1 & 0 \le k \le 3 \quad \text{e} \quad 29 \le k \le 31 \\ 0 & 5 \le k \le 27 \\ 0{,}5 & k = 4.28 \end{cases}$$

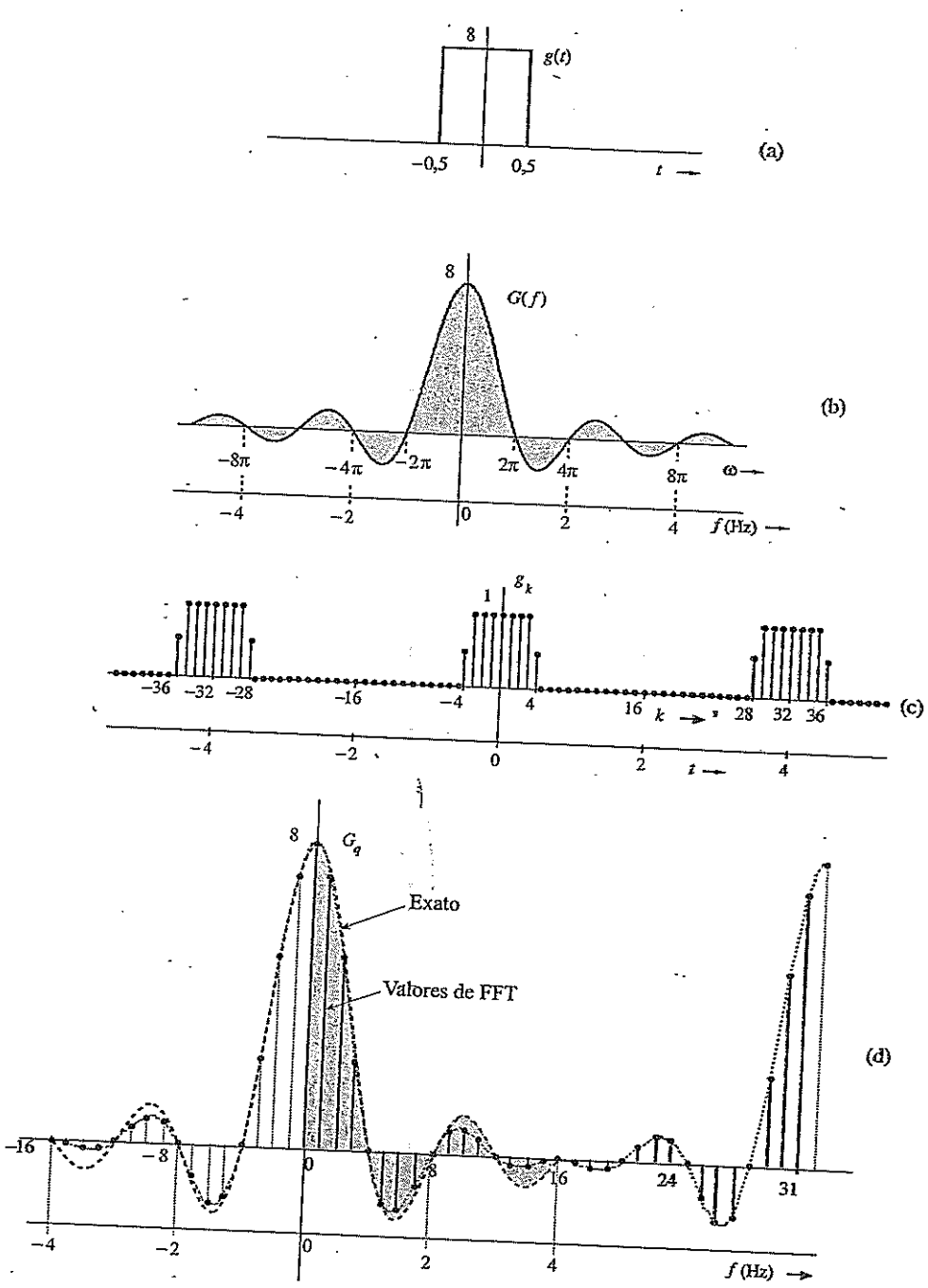


Figura 3.41 Transformada de Fourier discreta de um pulso retangular.

Vale ressaltar que a última amostra é tomada em $t = 31/8$ e não em $t = 4$, pois a repetição do sinal reinicia em $t = 4$, de modo que a amostra em $t = 4$ é igual à amostra em $t = 0$. Com $N_0 = 32$, $\Omega_0 = 2\pi/32 = \pi/16$. Logo, [ver a Eq. (3.103a)],

$$G_q = \sum_{k=0}^{31} g_k e^{-jq\frac{\pi}{16}k}$$

O programa MATLAB que usa o algoritmo de FFT para calcular a DFT é dado a seguir. Primeiro, escrevemos um programa MATLAB para gerar 32 amostras de g_k e, então, calculamos a DFT.

```
% (c32.m)
B=4;     f0=1/4;
Ts=1/(2*B); T0=1/f0;
N0=T0/Ts;
k=0:N0; k=k';
for m=1:length(k)
$  $ if k(m)$>$=0 & k(m)$<$=3, gk(m)=1; end
$  $ if k(m)==4 & k(m)==28 gk(m)=0.5; end
$  $ if k(m)$>$=5 & k(m)$<$=27, gk(m)=0; end
$  $ if k(m)$>$=29 & k(m)$<$=31, gk(m)=1; end
end
gk=gk';
Gr=fft(gk);
subplot(211), stem(k,gk)
subplot(212), stem(k,Gr)
```

A Fig. 3.41d mostra o gráfico de G_q.

As amostras G_q são espaçadas de $f_0 = 1/T_0$ Hz. Neste exemplo, $T_0 = 4$ segundos, de modo que a resolução de frequência f_0 é ¼ Hz, como desejado. A frequência de dobramento $f_s/2 = B = 4$ Hz corresponde a $q = N_0/2 = 16$. Como G_q tem período N_0 ($N_0 = 32$), os valores de G_q para q entre -16 e -1 são iguais àqueles para q entre 16 e 31. A DFT nos fornece amostras do espectro $G(f)$.

Para facilitar a comparação, a Fig. 3.41d também mostra a curva hachurada 8 sinc(πf), que é a transformada de Fourier de 8 $\Pi(t)$. Os valores de G_q calculados pela DFT exibem erro de mascaramento, o que fica claro quando comparamos os dois gráficos. O erro em G_2 é da ordem de apenas 1,3%. No entanto, o erro de mascaramento aumenta rapidamente com r. Por exemplo, o erro em G_6 é de cerca de 12%, e o erro em G_{10}, 33%. O erro em G_{14} é de assustadores 72%. O erro percentual aumenta de forma muito rápida nas proximidades da frequência de dobramento ($r = 16$), pois $g(t)$ tem uma descontinuidade degrau, o que faz com que $G(f)$ decaia muito lentamente, como $1/f$. Assim, nas proximidades da frequência de dobramento, a cauda invertida (devido ao mascaramento) é quase igual a $G(f)$. Além disso, os valores extremos são a diferença entre os valores exato e da parte que sofreu dobra (quase iguais aos exatos). Consequentemente, o erro percentual nas proximidades da frequência de dobramento ($r = 16$, neste exemplo) é muito alto, embora o erro absoluto seja muito pequeno. Fica claro que, para sinais com descontinuidades do tipo degrau, o erro de mascaramento nas proximidades da frequência de dobramento sempre será grande (em termos percentuais), qualquer que seja o valor escolhido para N_0. Para garantir erro de mascaramento desprezível para qualquer valor de q, devemos assegurar que $N_0 >> q$. Essa observação se aplica a todos os sinais com descontinuidade do tipo degrau.

Filtragem

Quando pensamos em filtragem, em geral, o fazemos em termos de uma solução orientada a hardware (ou seja, montagem de um circuito com componentes *RLC* e amplificadores operacionais). Contudo, a filtragem também admite uma solução orientada a software [algoritmo computacional que fornece a saída filtrada $y(t)$, para uma dada entrada $g(t)$]. Isso pode ser implementado de modo conveniente via DFT. Seja $g(t)$ o sinal a ser filtrado; então, os valores G_q, DFT de g_k, são calculados. O espectro G_q é formatado (filtrado) como desejado através da multiplicação de G_q por H_q, em que H_q são as amostras da função de transferência do filtro, $H(f)$ [$H_q = H(qf_0)$]. Por fim, calculamos a DFT inversa (ou IDFT) de G_qH_q e obtemos a saída filtrada y_k [$y_k = T_s y(kT)$]. O próximo exemplo ilustra este procedimento.

EXEMPLO COMPUTACIONAL C3.3

O sinal $g(t)$ na Fig. 3.42a é aplicado a um filtro passa-baixos ideal, cuja função de transferência $H(f)$ é mostrada na Fig. 3.42b. Usemos a DFT para calcular a saída do filtro.

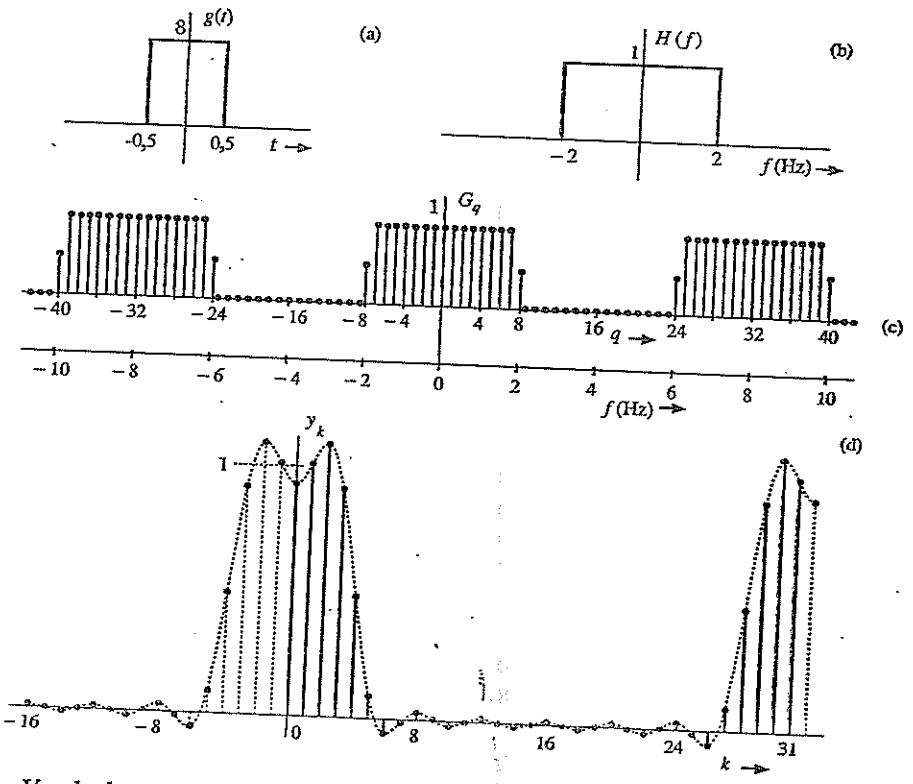


Figura 3.42 Filtragem de $g(t)$ por $H(f)$.

Já calculamos a DFT de $g(t)$ com 32 amostras (Fig. 3.41d). Agora, devemos multiplicar G_q por H_q. Para calcular H_q, recordemos que, na determinação da DFT de $g(t)$ com 32 amostras, usamos $f_0 = 0{,}25$ Hz. Como G_q tem período $N_0 = 32$, H_q deve ter o mesmo período e, portanto, amostras espaçadas de 0,25 Hz. Isso significa que H_q deve se repetir a cada 8 Hz ou 16π rad/s (ver Fig. 3.42c). Assim, as 32 amostras de H_q são produzidas, no intervalo $0 \le f \le 8$, como

$$H_q = \begin{cases} 1 & 0 \le q \le 7 \quad \text{e} \quad 25 \le q \le 31 \\ 0 & 9 \le q \le 23 \\ 0{,}5 & q = 8.24 \end{cases}$$

Multiplicamos G_q por H_q e calculamos a DFT inversa. O resultante sinal de saída é mostrado na Fig. 3.42d. A Tabela 3.4 lista valores de g_k, G_q, H_q, Y_q e y_k.

No Exemplo C.32, já calculamos a DFT de $g(t)$ com 32 amostras (G_q). O programa MATLAB do Exemplo C3.2 pode ser armazenado como um arquivo.m (por exemplo, "c32.m"). Podemos importar G_q no ambiente MATLAB via comando "c32". A seguir, geramos 32 amostras de H_q, multiplicamos G_q por H_q e, para obter y_k, calculamos a DFT inversa. Também podemos obter y_k calculando a convolução de g_k e h_k.

```
c32;
q=0:32; q=q';
for m=1:length(q)
     if q(m)$>$=0 & q(m)$<$=7, Hq(m)=1; end
     if q(m)$>$=25 & q(m)$<$=31, Hq(m)=1; end
     if q(m)$>$=9 & q(m)$<$=23, Hq(m)=0; end
     if q(m)==8 & q(m)==24, Hq(m)=0.5; end
```

```
end
Hq=Hq';
Yq=Gq.*Hq;
yk=ifft(Yq);
clf,stem(k,yk)
```

Tabela 3.4

No.	gk	G_q	H_q	G_qH_q	y_k
0	1	8,000	1	8,000	0,9285
1	1	7,179	1	7,179	1,009
2	1	5,027	1	5,027	1,090
3	1	2,331	1	2,331	0,9123
4	1	0,000	1	0,000	0,4847
5	0,5	−1,323	1	−1,323	0,08884
6	0	−1,497	1	−1,497	−0,05698
7	0	−0,8616	1	−0,8616	−0,01383
8	0	0,000	0,5	0,000	0,02933
9	0	0,5803	0	0,000	0,004837
10	0	0,6682	0	0,000	−0,01966
11	0	0,3778	0	0,000	−0,002156
12	0	0,000	0	0,000	0,01534
13	0	−0,2145	0	0,000	0,0009828
14	0	−0,1989	0	0,000	−0,01338
15	0	−0,06964	0	0,000	−0,0002876
16	0	0,000	0	0,000	0,01280
17	0	−0,06964	0	0,000	−0,0002876
18	0	−0,1989	0	0,000	−0,01338
19	0	−0,2145	0	0,000	0,0009828
20	0	0,000	0	0,000	0,01534
21	0	0,3778	0	0,000	−0,002156
22	0	0,6682	0	0,000	−0,01966
23	0	0,5803	0	0,000	0,004837
24	0	0,000	0,5	0,000	0,03933
25	0	−0,8616	1	−0,8616	−0,01383
26	0	−1,497	1	−1,497	−0,05698
27	0	−1,323	1	−1,323	0,08884
28	0,5	0,000	1	0,000	0,4847
29	1	2,331	1	2,331	0,9123
30	1	5,027	1	5,027	1,090
31	1	7,179	1	7,179	1,009

REFERÊNCIAS

1. R. V. Churchill and J. W. Brown, *Fourier Series and Boundary Value Problems*, 3rd ed., McGraw-Hill, New York, 1978.
2. R. N. Bracewell, *Fourier Transform and Its Applications*, rev. 2nd ed., McGraw-Hill, New York, 1986.
3. B. P. Lathi, *Signal Processing and Linear Systems*, Oxford University Press, 2000.
4. E. A. Guillemin, *Theory of Linear Physical Systems*, Wiley, New York, 1963.
5. F. J. Harris, "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," *Proc. IEEE*, vol. 66, pp. 51–83, Jan. 1978.
6. J. W. Tukey and J. Cooley, "An Algorithm for the Machine Calculation of Complex Fourier Series," *Mathematics of Computation*, Vol. 19, pp. 297–301, April 1965.

EXERCÍCIOS

3.1-1 Mostre que a transformada de Fourier de $g(t)$ pode ser expressa como

$$G(f) = \int_{-\infty}^{\infty} g(t)\cos 2\pi ft\, dt - j\int_{-\infty}^{\infty} g(t)\operatorname{sen} 2\pi ft\, dt$$

A seguir, mostre que, caso $g(t)$ seja uma função par de t,

$$G(f) = 2\int_{0}^{\infty} g(t)\cos 2\pi ft\, dt$$

e, caso $g(t)$ seja uma função ímpar de t,

$$G(f) = -2j\int_{0}^{\infty} g(t)\operatorname{sen} 2\pi ft\, dt$$

Agora, mostre que:

Se $g(t)$ for:	*Então $G(f)$ é:*
uma função real e par de t	uma função real e par de f
uma função real e ímpar de t	uma função imaginária e ímpar de f
uma função imaginária e par de t	uma função imaginária e par de f
uma função complexa e par de t	uma função complexa e par de f
uma função real e ímpar de t	uma função real e ímpar de f

3.1-2 **(a)** Mostre que, para uma função $g(t)$ de valores reais, a transformada inversa, Eq. (3.9b), pode ser expressa como

$$g(t) = 2\int_{0}^{\infty} |G(f)|\cos[2\pi ft + \theta_g(2\pi f)]\, df$$

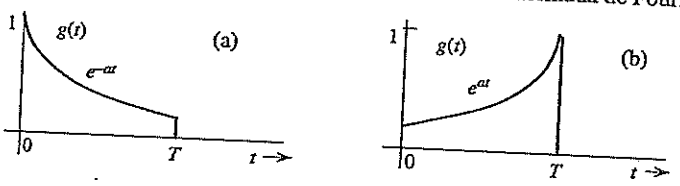
Essa é a forma trigonométrica da transformada de Fourier (inversa).

(b) Expresse a integral de Fourier (transformada de Fourier inversa) para $g(t) = e^{-at}u(t)$ na forma trigonométrica dada na parte (a).

3.1-3 Se $g(t) \Longleftrightarrow G(f)$, mostre que $g^*(t) \Longleftrightarrow G^*(-f)$.

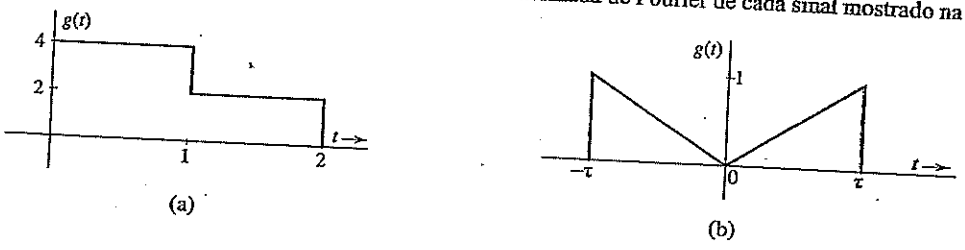
3.1-4 A partir da definição (3.9a), determine a transformada de Fourier de cada sinal mostrado na Fig. E.3.1-4.

Figura E.3.1-4



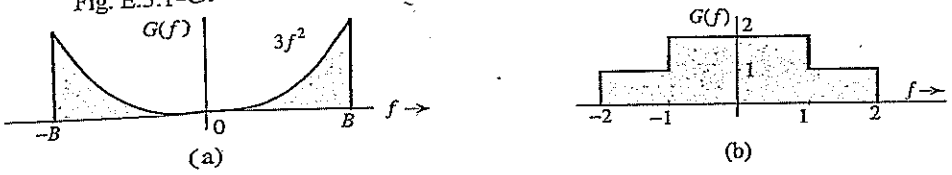
3.1-5 A partir da definição (3.9a), determine a transformada de Fourier de cada sinal mostrado na Fig. E.3.1-5.

Figura E.3.1-5



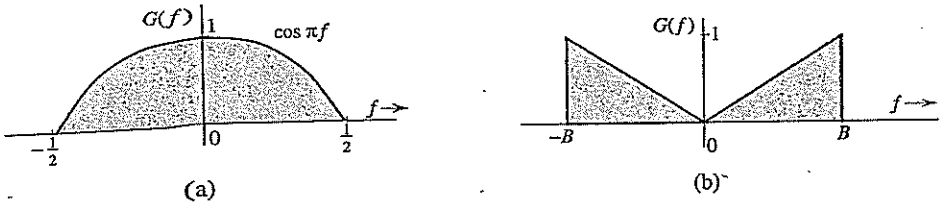
3.1-6 A partir da definição (3.9b), determine a transformada de Fourier inversa de cada espectro mostrado na Fig. E.3.1-6.

Figura E.3.1-6



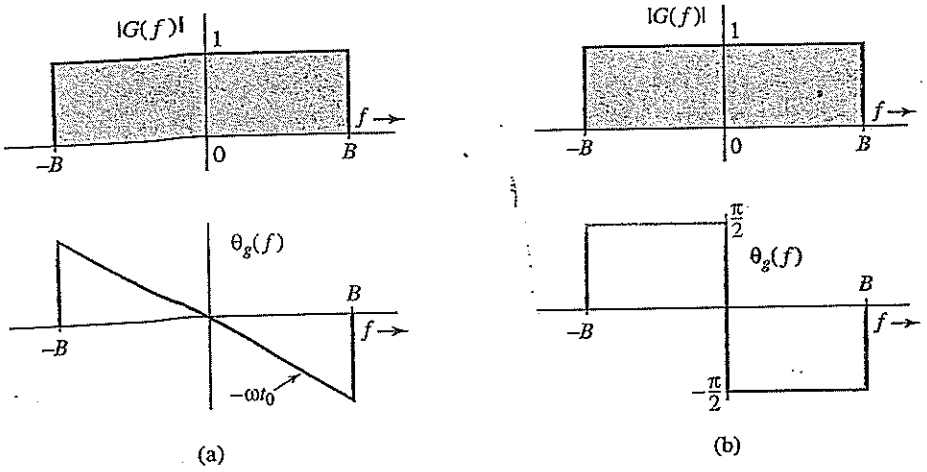
3.1-7 A partir da definição (3.9b), determine a transformada de Fourier inversa de cada espectro mostrado na Fig. E.3.1-7.

Figura E.3.1-7



3.1-8 Mostre que, apesar da semelhança, os dois sinais nas partes **(a)** e **(b)** da Fig. E.3.1-8 são totalmente distintos no domínio do tempo.

Figura E.3.1-8



Sugestão: $G(f) = |G(f)|e^{j\theta_g(f)}$. Na parte **(a)**, $G(f) = 1 \times e^{-j2\pi f t_0}$, $|f| \le B$, enquanto na parte **(b)**,

$$G(f) = \begin{cases} 1e^{-j\pi/2} = -j & 0 < f \le B \\ 1e^{j\pi/2} = j & 0 > f \ge -B \end{cases}$$

3.2-1 Esboce gráficos das seguintes funções: **(a)** $\Pi(t/2)$; **(b)** $\Delta(3\omega/100)$; **(c)** $\Pi(t - 10/8)$; **(d)** sinc $(\pi\omega/5)$; **(e)** sinc $[(\omega - 10\pi)/5]$; **(f)** sinc $(t/5)\,\Pi(t/10\pi)$.

Sugestão: $g(\frac{x-a}{b})$ é $g(\frac{x}{b})$ deslocada para a direita por a.

3.2-2 Da definição (3.9a), mostre que a transformada de Fourier de $\Pi\,(t-5)$ é sinc $(\pi f)e^{-j10\pi f}$.

3.2-3 Da definição (3.9b), mostre que a transformada de Fourier inversa de $\Pi\,[(2\pi f - 10)/2\pi]$ é sinc $(\pi t)e^{j10t}$.

3.2-4 Use os pares 7 e 12 na Tabela 3.1 e mostre que $u(t) \Longleftrightarrow 0{,}5\,\delta(f) + 1/j2\pi f$.

Sugestão: Some 1 a sgn (t) e verifique o sinal resultante.

3.2-5 Mostre que $\cos(2\pi f_0 t + \theta) \Longleftrightarrow \frac{1}{2}[\delta(f+f_0)e^{-j\theta} + \delta(f-f_0)e^{j\theta}]$.

Sugestão: Use a fórmula de Euler para expressar $\cos(2\pi f_0 t + \theta)$ em termos de exponenciais.

3.3-1 Aplique a propriedade de dualidade a pares adequados na Tabela 3.1 e mostre que:

(a) $0{,}5[\delta(t) + (j/\pi t)] \Longleftrightarrow u(f)$

(b) $\delta(t+T) + \delta(t-T) \Longleftrightarrow 2\cos 2\pi fT$

(c) $\delta(t+T) - \delta(t-T) \Longleftrightarrow 2j \operatorname{sen} 2\pi fT$

Sugestão: $g(-t) \Longleftrightarrow G(-f)$ e $\delta(t) = \delta(-t)$.

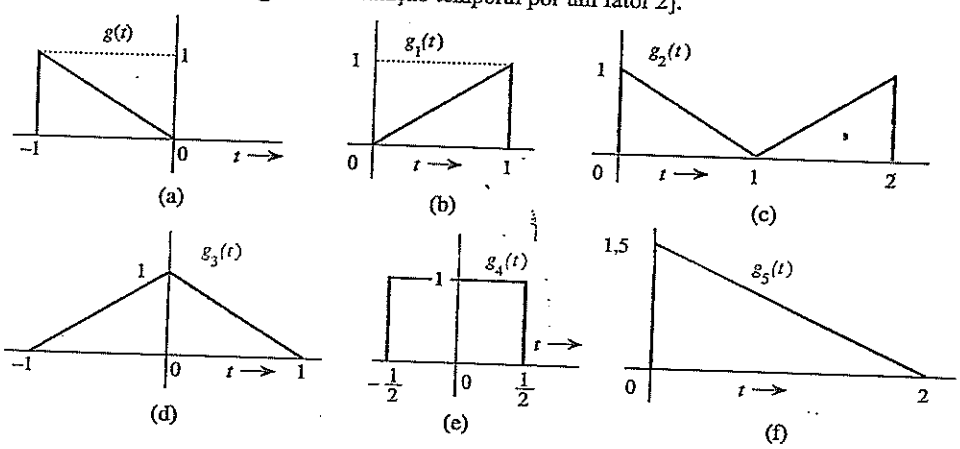
3.3-2 A transformada de Fourier do pulso triangular $g(t)$ na Fig. E.3.3-2a é dada por

$$G(f) = \frac{1}{(2\pi f)^2}(e^{j2\pi f} - j2\pi f e^{j2\pi f} - 1)$$

Use essa informação e as propriedades de translação e dilatação no domínio do tempo para determinar a transformada de Fourier de cada sinal mostrado na Fig. E.3.3-2b–f.

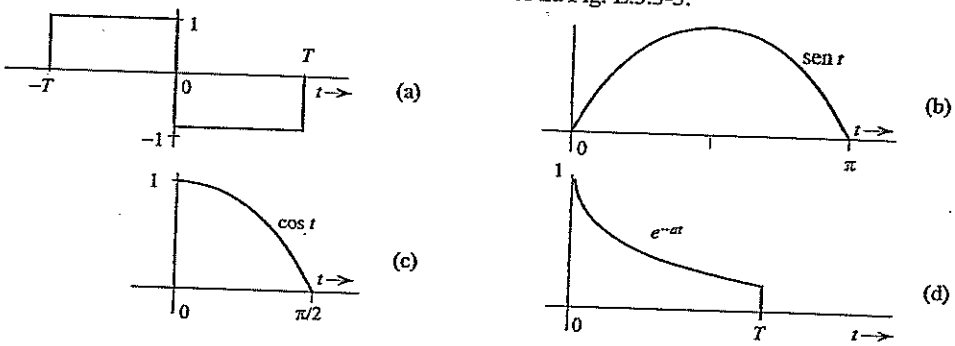
Sugestão: O reflexo de $g(t)$ em relação ao eixo vertical resulta no pulso $g_1(t)$ na Fig. E.3.3-2b; em consequência, $g_1(t) = g(-t)$. O pulso na Fig. E.3.3-2c pode ser expresso como $g(t-T) + g_1(t-T)$ [soma de $g(t)$ e $g_1(t)$ atrasados de T]. Os dois pulsos na Fig. 3.3-2d-e podem ser expressos como $g(t-T) + g_1(t+T)$) [soma de $g(t)$ atrasado de T e $g_1(t)$ adiantado de T], para alguma escolha adequada de T. O pulso na Fig. 3.3-2f pode ser obtido pela dilatação temporal de $g(t)$ por um fator 2, seguida de um atraso de 2 segundos [ou atraso de 1 segundo e dilatação temporal por um fator 2].

Figura E.3.3-2



3.3-3 Usando apenas a propriedade de translação no domínio do tempo e a Tabela 3.1, determine a transformada de Fourier de cada um dos sinais mostrados na Fig. E.3.3-3.

Figura E.3.3-3



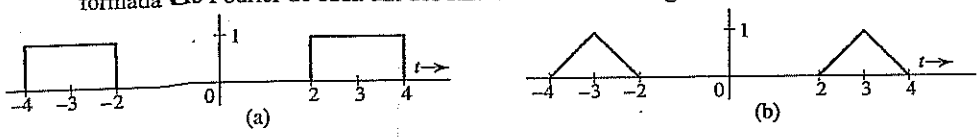
Sugestão: O sinal na Fig. E.3.3-3a é a soma de dois pulsos retangulares deslocados. O sinal na Fig. E.3.3-3b é $\operatorname{sen} t\,[u(t) - u(t-\pi)] = \operatorname{sen} t\, u(t) - \operatorname{sen} t\, u(t-\pi) = \operatorname{sen} t\, u(t) + \operatorname{sen}(t-\pi)\, u(t-\pi)$. Comprove que a soma dessas duas senoides, de fato, resulta no pulso na Fig. E.3.3-3b. De modo similar, o sinal na Fig. E.3.3-3c pode ser expresso como $\cos t\, u(t) + \operatorname{sen}(t - \pi/2)\, u(t - \pi/2)]$ (comprove isso traçando os gráficos dos sinais). O sinal na Fig. 3.3-3d é $e^{-at}[u(t) - u(t-T)] = e^{-at}u(t) - e^{-aT}e^{-a(t-T)}u(t-T)$.

3.3-4 Use a propriedade de translação no domínio do tempo e mostre que, se $g(t) \Longleftrightarrow G(f)$, então

$$g(t+T)+g(t-T) \Longleftrightarrow 2G(f)\cos 2\pi fT$$

Isso é o dual da Eq. (3.36). Com base neste resultado e nos pares 17 e 19 na Tabela 3.1, determine a transformada de Fourier de cada um dos sinais mostrados na Fig. E.3.3-4.

Figura 3.3-4



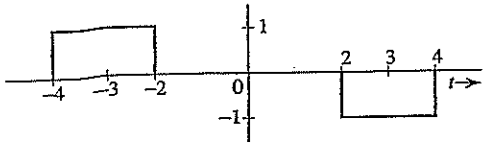
3.3-5 Comprove os seguintes resultados:

$$g(t)\,\mathrm{sen}\,2\pi f_0 t \Longleftrightarrow \frac{1}{2j}[G(f-f_0)-G(f+f_0)]$$

$$\frac{1}{2j}[g(t+T)-g(t-T)] \Longleftrightarrow G(f)\,\mathrm{sen}\,2\pi fT$$

Com base nesse resultado e na Tabela 3.1, determine a transformada de Fourier do sinal na Fig. E.3.3-5.

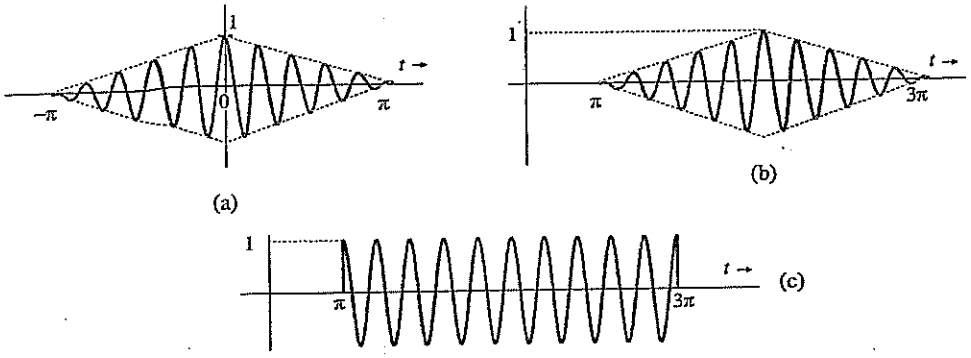
Figura 3.3-5



3.3-6 Os sinais na Fig. E.3.3-6 são modulados com uma portadora cos $10t$. Utilizando propriedades adequadas da transformada de Fourier e a Tabela 3.1, determine as transformadas de Fourier desses sinais. Esboce os gráficos dos espectros de amplitude e de fase para os sinais nas Fig. E.3.3-6a e b.

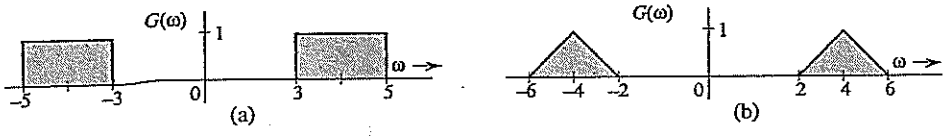
Sugestão: Essas funções podem ser expressas na forma $g(t)\cos 2\pi f_0 t$.

Figura 3.3-6



3.3-7 Utilizando a propriedade de translação no domínio da frequência e a Tabela 3.1, determine a transformada de Fourier inversa do espectro mostrado na Fig. E.3.3-7. Observe que, nesse exercício, a transformada de Fourier é dada no domínio ω.

Figura 3.3-7



3.3-8 Um sinal $g(t)$ é limitado em frequência a B Hz. Mostre que o sinal $g^n(t)$ é limitado em frequência a nB Hz.

Sugestão: $g^2(t) \Longleftrightarrow [G(f) * G(f)]$, e assim por diante. Use a propriedade de largura de banda da convolução.

3.3-9 Determine a transformada de Fourier do sinal na Fig. E.3.3-3a empregando três métodos diferentes:

(a) Por integração direta, utilizando a definição (3.9a).

(b) Empregando apenas o par 17 na Tabela 3.1 e a propriedade de translação no domínio do tempo.

(c) Usando as propriedades de diferenciação e translação no domínio do tempo, assim como o par $\delta(t) \Longleftrightarrow 1$.

Sugestão: $1 - \cos 2x = 2 \operatorname{sen}^2 x$.

3.3-10 O processo de recuperação de um sinal $g(t)$ de um sinal modulado $g(t) \cos 2\pi f_0 t$ é denominado **demodulação**. Mostre que o sinal $g(t) \cos 2\pi f_0 t$ pode ser demodulado multiplicando-o por $2 \cos 2\pi f_0 t$ e aplicando o produto a um filtro passa-baixos de largura de banda B Hz [a largura de banda de $g(t)$]. Considere $B < f_0$. *Sugestão*: $2 \cos^2 2\pi f_0 t = 1 + \cos 4\pi f_0 t$. Observe que o espectro de $g(t) \cos 4\pi f_0 t$ é centrado em $2f_0$ e será suprimido pelo filtro passa-baixos de largura de banda B Hz.

3.4-1 Os sinais $g_1(t) = 10^4\, \Pi(10^4 t)$ e $g_2(t) = \delta(t)$ são aplicados às entradas de filtros passa-baixos ideais $H_1(f) = \Pi(f/20.000)$ e $H_2(f) = \Pi(f/10.000)$ (Fig. E.3.4-1). As saídas $y_1(t)$ e $y_2(t)$ desses filtros são multiplicadas para produzir o sinal $y(t) = y_1(t)\, y_2(t)$.

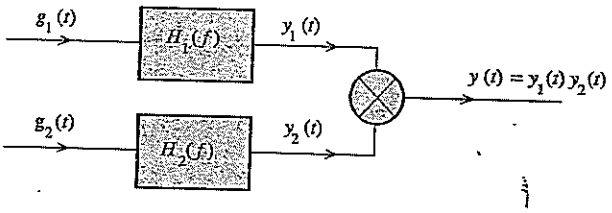
(a) Trace os gráficos de $G_1(f)$ e $G_2(f)$.

(b) Trace os gráficos de $H_1(f)$ e $H_2(f)$.

(c) Trace os gráficos de $Y_1(f)$ e $Y_2(f)$.

(d) Determine as larguras de banda de $y_1(t)$, $y_2(t)$ e de $y(t)$.

Figura E.3.4-1



3.5-1 Dentre os sistemas com as respostas ao impulso listadas a seguir, qual é causal?

(a) $h(t) = e^{-at} u(t), \quad a > 0$

(b) $h(t) = e^{-a|t|}, \quad a > 0$

(c) $h(t) = e^{-a(t-t_0)} u(t - t_0), \quad a > 0$

(d) $h(t) = \operatorname{sinc}(at), \quad a > 0$

(e) $h(t) = \operatorname{sinc}[a(t - t_0)], \quad a > 0$.

3.5-2 Considere um filtro com a seguinte função de transferência:

$$H(f) = e^{-k(2\pi k f)^2 - j2\pi f t_0}$$

Usando os critérios dos domínios do tempo [$h(t)$ não causal] e da frequência (critério de Paley-Wiener), mostre que este filtro é fisicamente irrealizável. É possível, com a escolha de um valor suficientemente grande para t_0, tornar este filtro aproximadamente realizável? Usando um critério próprio (razoável) de realização aproximada, determine t_0.

Sugestão: Use o par 22 na Tabela 3.1.

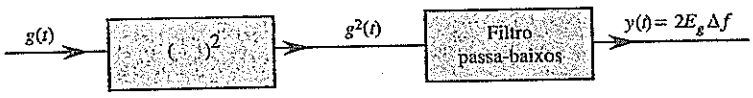
3.5-3 Mostre que um filtro com a função de transferência

$$H(f) = \frac{2(10^5)}{(2\pi f)^2 + 10^{10}} e^{-j2\pi f t_0}$$

é irrealizável. É possível, com a escolha de um valor suficientemente grande para t_0, tornar esse filtro aproximadamente realizável? Usando um critério próprio (razoável) de realização aproximada, determine t_0.

Sugestão: Mostre que a resposta ao impulso é não causal.

Figura E.3.7-6



3.8-1 Mostre que a função de autocorrelação de $g(t) = C\cos(2\pi f_0 t + \theta_0)$ é dada por $\mathcal{R}_g(\tau) = (C^2/2)\cos 2\pi f_0 \tau$ e que a correspondente PSD é $S_g(f) = (C^2/4)[\delta(f-f_0) + \delta(f+f_0)]$. A seguir, mostre que, para um sinal $y(t)$ dado por

$$y(t) = C_0 + \sum_{n=1}^{\infty} C_n \cos(n2\pi f_0 t + \theta_n)$$

a função de autocorrelação e a PSD são dadas por

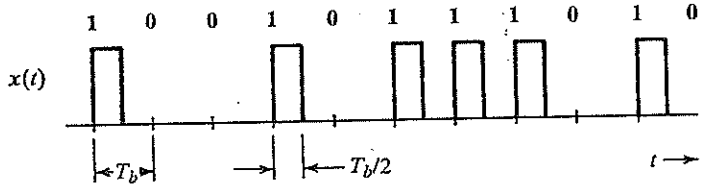
$$\mathcal{R}_y(\tau) = C_0{}^2 + \frac{1}{2}\sum_{n=1}^{\infty} C_n{}^2 \cos n2\pi f_0 \tau$$

$$S_y(f) = C_0{}^2\delta(f) + \frac{1}{4}\sum_{n=1}^{\infty} C_n{}^2[\delta(f - nf_0) + \delta(f + nf_0)]$$

Sugestão: Mostre que, se $g(t) = g_1(t) + g_2(t)$, então $\mathcal{R}_g(\tau) = \mathcal{R}_{g_1}(\tau) + \mathcal{R}_{g_2}(\tau) + \mathcal{R}_{g_1 g_2}(\tau) + \mathcal{R}_{g_2 g_1}(\tau)$, em que $\mathcal{R}_{g_1 g_2}(\tau) = \lim_{T\to\infty}(1/T)\int_{-T/2}^{T/2} g_1(t)g_2(t+\tau)\,dt$. Caso $g_1(t)$ e $g_2(t)$ representem quaisquer dois dos infinitos termos em $y(t)$, mostre que $\mathcal{R}_{g_1 g_2}(\tau) = \mathcal{R}_{g_2 g_1}(\tau) = 0$. Para isso, use o fato de que, considerando um intervalo de tempo muito grande, a área sob qualquer senoide é, no máximo, igual à área correspondente a um meio ciclo da senoide.

3.8-2 O sinal binário aleatório $x(t)$ mostrado na Fig. E.3.8-2 transmite um dígito a cada T_b segundos. Um **1** binário é transmitido por um pulso $p(t)$ de largura $T_b/2$ e amplitude A; um **0** binário é transmitido pela ausência de pulso. Os dígitos **1** e **0** têm igual probabilidade de ocorrência, que é aleatória. Determine a função de autocorrelação $\mathcal{R}_x(\tau)$ e a PSD $S_x(f)$.

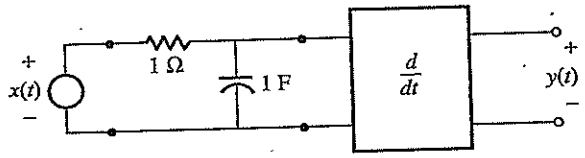
Figura E.3.8-2



3.8-3 Determine o valor quadrático médio (ou potência) da tensão de saída $y(t)$ do circuito RC mostrado na Fig. E.3.5-4, com $RC = 2\pi$, considerando que a PSD da tensão de entrada $S_x(f)$ seja dada por **(a)** K; **(b)** $\Pi(\pi f)$; **(c)** $[\delta(f+1) + \delta(f-1)]$. Em cada caso, calcule a potência (valor quadrático médio) do sinal de entrada $x(t)$.

3.8-4 Determine o valor quadrático médio (ou potência) da tensão de saída $y(t)$ do sistema mostrado na Fig. E.3.8-4, considerando que a PSD da tensão de entrada $S_x(f)$ seja dada por $\Pi(\pi f)$. Calcule a potência (valor quadrático médio) do sinal de entrada $x(t)$.

Figura 3.8-4



4 MODULAÇÕES E DEMODULAÇÕES EM AMPLITUDE

Modulação, em geral, se refere a um processo que desloca o sinal de mensagem a uma banda específica de frequências, ditada pelo canal físico (por exemplo, um modem telefônico para a banda de sinal de voz). A modulação provê diversas vantagens, como mencionado no Capítulo 1, entre as quais se incluem a facilidade de transmissão de RF e a multiplexação por divisão em frequência. As modulações podem ser analógicas ou digitais. Embora sistemas de comunicação tradicionais, como rádios AM/FM e sinais de televisão NTSC, sejam baseados em modulações analógicas, sistemas mais recentes, como os de telefonia celular 2G e 3G, HDTV e DSL, são todos digitais.

Neste capítulo e no próximo, focaremos as clássicas modulações analógicas: modulação em amplitude e modulação em ângulo. Antes de iniciarmos a discussão de modulações analógicas, é importante que façamos a distinção entre sistemas de comunicação que não usam modulação (**sistemas de comunicação em banda base**) e sistemas que utilizam modulação (**sistemas de comunicação por portadora**).

4.1 COMUNICAÇÃO EM BANDA BASE *VERSUS* COMUNICAÇÃO POR PORTADORA

O termo **banda base** designa a banda de frequências do sinal de mensagem original proveniente da fonte ou transdutor de entrada (Fig. 1.2). Em telefonia, a banda base é a faixa de frequências de áudio (banda de sinais de voz), de 0 a 3,5 kHz. No caso de sinal de televisão NTSC, a banda base de vídeo é a faixa de frequências de 0 a 4,3 MHz. Para dados digitais ou modulação por codificação de pulsos (PCM) que usam sinalização bipolar à taxa de R_b pulsos por segundo, a banda base é, aproximadamente, a faixa de frequências de 0 a R_b Hz.

Comunicação em Banda Base

Na comunicação em banda base, sinais de mensagem são transmitidos diretamente sem qualquer modificação. Dado que, em sua maioria, sinais em banda base, como sinais de áudio e de vídeo, têm um conteúdo significativo em frequências baixas, não podem ser transmitidos de modo eficiente através de enlaces de rádio (enlaces sem fio). Consequentemente, canais de usuário dedicados, como pares trançados e cabos coaxiais, são alocados a cada usuário para comunicações de longas distâncias. Uma vez que sinais em banda base têm bandas que se sobrepõem, haveria forte interferência entre sinais que compartilhassem um mesmo canal. Por esse motivo, na comunicação em banda base, uma boa parte do espectro não é utilizada. Com a modulação de vários sinais em banda base e o resultante deslocamento de seus espectros a bandas de frequências que não se sobrepõem, o que constitui a multiplexação por divisão em frequência (FDM – *frequency division multiplexing*), muitos usuários

4 MODULAÇÕES E DEMODULAÇÕES EM AMPLITUDE

Modulação, em geral, se refere a um processo que desloca o sinal de mensagem a uma banda específica de frequências, ditada pelo canal físico (por exemplo, um modem telefônico para a banda de sinal de voz). A modulação provê diversas vantagens, como mencionado no Capítulo 1, entre as quais se incluem a facilidade de transmissão de RF e a multiplexação por divisão em frequência. As modulações podem ser analógicas ou digitais. Embora sistemas de comunicação tradicionais, como rádios AM/FM e sinais de televisão NTSC, sejam baseados em modulações analógicas, sistemas mais recentes, como os de telefonia celular 2G e 3G, HDTV e DSL, são todos digitais.

Neste capítulo e no próximo, focaremos as clássicas modulações analógicas: modulação em amplitude e modulação em ângulo. Antes de iniciarmos a discussão de modulações analógicas, é importante que façamos a distinção entre sistemas de comunicação que não usam modulação (**sistemas de comunicação em banda base**) e sistemas que utilizam modulação (**sistemas de comunicação por portadora**).

4.1 COMUNICAÇÃO EM BANDA BASE *VERSUS* COMUNICAÇÃO POR PORTADORA

O termo **banda base** designa a banda de frequências do sinal de mensagem original proveniente da fonte ou transdutor de entrada (Fig. 1.2). Em telefonia, a banda base é a faixa de frequências de áudio (banda de sinais de voz), de 0 a 3,5 kHz. No caso de sinal de televisão NTSC, a banda base de vídeo é a faixa de frequências de 0 a 4,3 MHz. Para dados digitais ou modulação por codificação de pulsos (PCM) que usam sinalização bipolar à taxa de R_b pulsos por segundo, a banda base é, aproximadamente, a faixa de frequências de 0 a R_b Hz.

Comunicação em Banda Base

Na comunicação em banda base, sinais de mensagem são transmitidos diretamente sem qualquer modificação. Dado que, em sua maioria, sinais em banda base, como sinais de áudio e de vídeo, têm um conteúdo significativo em frequências baixas, não podem ser transmitidos de modo eficiente através de enlaces de rádio (enlaces sem fio). Consequentemente, canais de usuário dedicados, como pares trançados e cabos coaxiais, são alocados a cada usuário para comunicações de longas distâncias. Uma vez que sinais em banda base têm bandas que se sobrepõem, haveria forte interferência entre sinais que compartilhassem um mesmo canal. Por esse motivo, na comunicação em banda base, uma boa parte do espectro não é utilizada. Com a modulação de vários sinais em banda base e o resultante deslocamento de seus espectros a bandas de frequências que não se sobrepõem, o que constitui a multiplexação por divisão em frequência (FDM – *frequency division multiplexing*), muitos usuários

podem compartilhar um canal, utilizando a maior parte da largura de banda disponível. Comunicação de longa distância através de um enlace de rádio também requer modulação para deslocar o espectro do sinal a frequências mais altas, para permitir radiação mais eficiente de potência, com a utilização de antenas de dimensões razoáveis. Outro uso da modulação é a troca da largura de banda de transmissão por melhor desempenho diante de interferências.

Modulações com Portadora

A comunicação que faz uso de modulação para deslocar o espectro de frequência de um sinal é conhecida como **comunicação por portadora**. Em termos de modulação analógica, um dos parâmetros básicos (amplitude, frequência ou fase) de uma **portadora senoidal** de alta frequência f_c Hz (ou $\omega_c = 2\pi f_c$ rad/s) é variado linearmente com o sinal em banda base $m(t)$. Isso resulta em modulação em amplitude (AM – *amplitude modulation*), modulação em frequência (FM – *frequency modulation*) ou modulação em fase (PM – *phase modulation*), respectivamente. A modulação em amplitude é uma operação linear, enquanto os outros dois tipos de modulação com portadora são similares e não lineares, geralmente designadas pelo termo comum **modulação em ângulo**.

Aqui, cabe um comentário a respeito de sinais modulados por pulsos [modulação em amplitude de pulso (PAM – *pulse amplitude modulation*), modulação em largura de pulso (PWM – *pulse width modulation*), modulação em posição de pulso (PPM – *pulse position modulation*), modulação por codificação de pulsos (PCM – *pulse code modulation*) e modulação delta (DM – *delta modulation*)]. Apesar do termo *modulação*, são sinais digitais em banda base. Nesses casos, "modulação" é usada não para designar deslocamento de frequência ou de banda, mas, sim, os esquemas de codificação de pulsos digitais usados na representação dos sinais analógicos originais. Em outras palavras, o sinal de mensagem analógico modula parâmetros de um trem de pulsos digitais. Esses sinais ainda podem modular uma portadora, de modo que seus espectros sejam deslocados.

Modulações em Amplitude e Modulações em Ângulo

Denotemos por $m(t)$ o sinal de mensagem de fonte a ser transmitido pelo emissor aos receptores; a transformada de Fourier do sinal é denotada por $M(f)$. Para deslocar a resposta de frequência de $m(t)$ à uma nova faixa de frequências centrada em f_c Hz, notamos, primeiro, que a transformada de Fourier já revelou uma propriedade muito forte, conhecida como propriedade de *translação na frequência*, que permite alcançar este objetivo. Em outras palavras, tudo o que precisamos fazer é multiplicar $m(t)$ por uma frequência senoidal f_c, de modo que

$$s_1(t) = m(t) \cos 2\pi f_c t$$

Com isso, alcançamos imediatamente o objetivo básico da modulação e movemos o conteúdo de frequências do sinal para que fique centrado em $\pm f_c$:

$$S_1(f) = \frac{1}{2}M(f - f_c) + \frac{1}{2}M(f + f_c)$$

Esta simples multiplicação, na verdade, permite que variações na amplitude da senoide $s_1(t)$ sejam proporcionais ao sinal de mensagem. O método é, de fato, uma forma valiosa de modulação, conhecida como modulação em amplitude.

De modo mais geral, consideremos um sinal senoidal

$$s(t) = A(t) \cos [\omega_c t + \phi(t)]$$

Há três variáveis em uma senoide: amplitude, frequência (instantânea) e fase. O sinal de mensagem pode ser usado para modular qualquer um desses três parâmetros e permitir que $s(t)$ transporte a informação do transmissor ao receptor:

Amplitude $A(t)$ varia linearmente com $m(t)$	$\Longleftrightarrow$	modulação em amplitude
Frequência varia linearmente com $m(t)$	$\Longleftrightarrow$	modulação em frequência
Fase $\phi(t)$ varia linearmente com $m(t)$	$\Longleftrightarrow$	modulação em fase

Essas operações são conhecidas, respectivamente, como modulação em amplitude, modulação em frequência e modulação em fase. Neste capítulo, descreveremos várias formas de modulação em amplitude em sistemas de comunicação usados na prática. Modulações em amplitude são lineares e sua análise nos domínios do tempo e da frequência, mais simples. No Capítulo 5, discutiremos as modulações em ângulo não lineares separadamente.

Uso de f ou ω

No Capítulo 3, observamos a equivalência das respostas de frequência denotadas pela frequência f e pela frequência angular ω. Cada uma dessas duas notações tem suas próprias vantagens e desvantagens. Após o estudo dos exemplos e exercícios do Capítulo 3, o leitor deve se sentir familiarizado e confortável com o uso das duas notações. Assim, de aqui em diante, usaremos ambas, selecionando uma ou outra com base na simplicidade de notação ou representação ou gráfica.

4.2 MODULAÇÃO EM AMPLITUDE COM BANDA LATERAL DUPLA

Modulação em amplitude é caracterizada por uma amplitude $A(t)$ **portadora** de informação, que é uma função linear do sinal de banda base (mensagem) $m(t)$. Ao mesmo tempo, a frequência ω_c e fase θ_c da portadora permanecem constantes. Podemos considerar $\theta_c = 0$ sem perda de generalidade. Se a amplitude A da portadora for diretamente proporcional ao sinal modulante $m(t)$, o *sinal modulado* será $m(t) \cos \omega_c t$ (Fig. 4.1). Como vimos anteriormente, [Eq. (3.36)], esse tipo de modulação simplesmente desloca o espectro de $m(t)$ para a frequência da portadora (Fig. 4.1a). Portanto, se

$$m(t) \Longleftrightarrow M(f)$$

então

$$m(t) \cos 2\pi f_c t \Longleftrightarrow \tfrac{1}{2}[M(f+f_c) + M(f-f_c)] \tag{4.1}$$

Recordemos que $M(f-f_c)$ é $M(f)$ deslocada para a direita por f_c, e $M(f+f_c)$ é $M(f)$ deslocada para a esquerda por f_c. Dessa forma, o processo de modulação desloca o espectro do sinal modulante para a esquerda e para a direita por f_c. Devemos notar também que, se a largura de banda de $m(t)$ for B Hz, o sinal modulado terá largura de banda $2B$ Hz, como indicado na Fig. 4.1.c. Observemos ainda que o espectro do sinal modulado centrado em $\pm f_c$ (ou $\pm\omega_c$ rad/s) consiste em duas partes: uma no lado externo de $\pm f_c$, conhecida como *banda lateral superior* (*USB – upper sideband*) e uma no lado interno de $\pm f_c$, conhecida como *banda lateral inferior* (*LSB – lower sideband*). Na Fig. 4.1c, vemos ainda que, a menos que o sinal de mensagem $M(f)$ tenha um impulso na frequência zero, nesse esquema o sinal modulado não contém uma componente discreta da frequência da portadora f_c. Em outras palavras, o processo de modulação não introduz uma senoide em f_c. Por esse motivo, é referido como **modulação em banda lateral dupla com portadora suprimida** (DSB-SC – *double-sideband supressed carrier*).*

A relação entre B e f_c é relevante. A Fig. 4.1c mostra que $f_c \geq B$, o que evita a sobreposição dos espectros modulados centrados em f_c e em $-f_c$. Se $f_c < B$, as duas cópias do espectro da mensagem se sobreporiam e a informação de $m(t)$ seria perdida durante a modulação, o que impossibilitaria a recuperação de $m(t)$ do sinal modulado $m(t) \cos \omega_c t$.

Fatores práticos podem impor restrições adicionais a f_c. Por exemplo, em aplicações de difusão (*broadcasting*), a antena transmissora pode irradiar sem distorção apenas uma estreita largura de ban-

*O termo **portadora suprimida** não significa, necessariamente, a ausência de espectro na frequência da portadora f_c, mas que não existe uma componente discreta da frequência da portadora. Isso implica que o espectro da modulação DSB-SC não tem impulsos em $\pm f_c$; o que também implica que o sinal modulado $m(t) \cos 2\pi f t$ não contém um termo da forma $k \cos 2\pi f t$ [supondo que $m(t)$ tenha valor médio nulo].

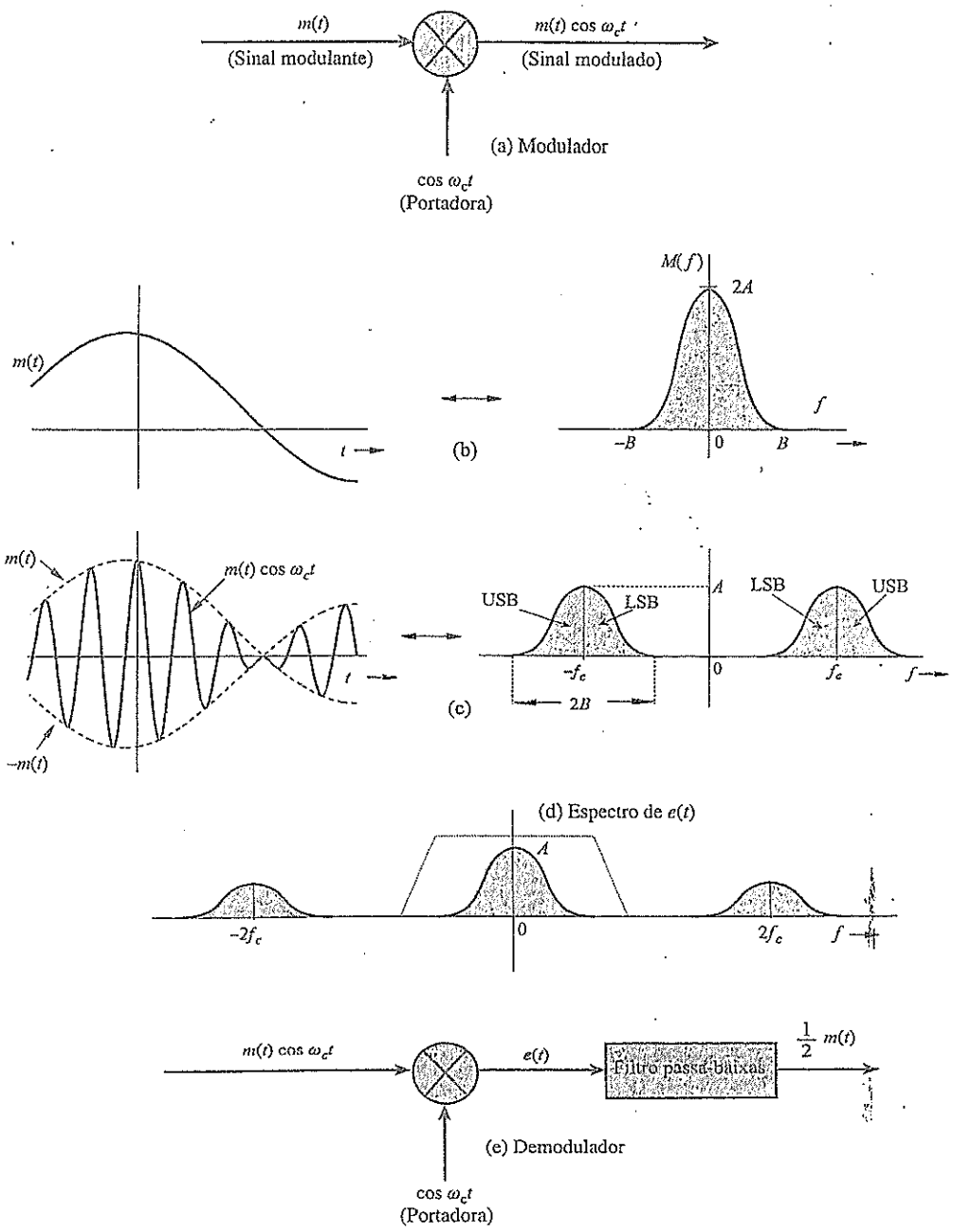


Figura 4.1 Modulação DSB-SC e demodulação.

da. Isso significa que, para evitar distorção causada pela antena transmissora, devemos ter $f_c/B >> 1$. Na radiodifusão AM, por exemplo, com $B = 5$ kHz e frequências portadoras entre 550 e 1600 kHz, a razão f_c/B tem valor entre 100 e 300.

Demodulação

A modulação DSB-SC traslada ou desloca o espectro de frequência para a esquerda e para a direita por f_c (ou seja, para $+f_c$ e $-f_c$), como mostra a Eq. (4.1). Para recuperar o sinal original $m(t)$ do sinal modulado, é necessário devolver o espectro à posição original. O processo de recuperação do sinal original do sinal modulado (translação do espectro de volta à sua posição original) é referido como **demodulação**. Observemos que, na Fig. 4.1c, se o espectro do sinal modulado for deslocado para a esquerda e para a direita por f_c (e multiplicado por meio), obteremos o espectro mostrado na Fig. 4.1d,

que contém o espectro de banda base desejado mais espectros indesejáveis em $\pm 2f_c$. Esses últimos podem ser suprimidos com o uso de um filtro passa-baixas. Assim, a demodulação, quase idêntica à modulação, consiste em multiplicar o sinal modulado recebido $m(t)\cos\omega_c t$ por uma portadora $\cos\omega_c t$ e aplicar o resultado a um filtro passa-baixas, como indicado na Fig. 4.1e. Podemos chegar à mesma conclusão diretamente no domínio do tempo; para isso, observemos que o sinal $e(t)$ na Fig. 4.1e é

$$\begin{aligned} e(t) &= m(t)\cos^2\omega_c t \\ &= \frac{1}{2}[m(t) + m(t)\cos 2\omega_c t] \end{aligned} \tag{4.2a}$$

Portanto, a transformada de Fourier do sinal $e(t)$ é

$$E(f) = \frac{1}{2}M(f) + \frac{1}{4}[M(f+2f_c) + M(f-2f_c)] \tag{4.2b}$$

Esta análise mostra que o sinal $e(t)$ consiste em duas componentes $(1/2)m(t)$ e $(1/2)m(t)\cos 2\omega_c t$, juntamente com seus espectros que não se sobrepõem, como ilustrado na Fig. 4.1d. O espectro da segunda componente, que é um sinal modulado com portadora $2f_c$, é centrado em $\pm 2f_c$. Logo, essa componente é suprimida pelo filtro passa-baixas na Fig. 4.1e. A componente desejada $(1/2)M(f)$, por ser um espectro de frequências baixas (centrado em $f = 0$), passa pelo filtro e resulta na saída $(1/2)m(t)$. Uma forma possível de característica de filtro passa-baixas é mostrada (em linha tracejada) na Fig. 4.1d. O filtro permite a demodulação sem distorção do sinal de mensagem $m(t)$ do sinal DSB-SC. Para nos livrarmos da inconveniente fração 1/2 na saída, podemos usar uma portadora $2\cos\omega_c t$ em vez de $\cos\omega_c t$. Na verdade, mais adiante, usaremos muito essa estratégia, que não afeta as conclusões gerais.

Esse método de recuperação do sinal em banda base é denominado *detecção síncrona* ou *detecção coerente*, em que usamos uma portadora com exatamente a mesma frequência (e fase) da portadora usada na modulação. Assim, para demodulação, torna-se necessário gerar uma portadora local no receptor, com coerência (sincronismo) de frequência e fase com a portadora usada na modulação.

Exemplo 4.1 Para o sinal em banda base

$$m(t) = \cos\omega_m t = \cos 2\pi f_m t,$$

determinemos o sinal DSB-SC e esbocemos seu espectro. Identifiquemos a USB e a LSB. Verifiquemos que o sinal modulado DSB-SC pode ser demodulado pelo demodulador ilustrado na Fig. 4.1e.

O caso neste exemplo é referido como *modulação por tom*, pois o sinal modulante é uma senoide pura, ou tom puro, $\cos\omega_m t$. Para deixar claros os conceitos básicos da modulação DSB-SC, resolveremos esse problema nos domínios da frequência e do tempo. Na abordagem do domínio da frequência, trabalhamos com espectros dos sinais. O espectro do sinal em banda base $m(t) = \cos\omega_m t$ é dado por

$$\begin{aligned} M(f) &= \frac{1}{2}[\delta(f - f_m) + \delta(f + f_m)] \\ &= \pi[\delta(\omega - \omega_m) + \delta(\omega + \omega_m)] \end{aligned}$$

O espectro da mensagem consiste em dois impulsos localizados em $\pm f_m$, como ilustrado na Fig. 4.2a. O espectro (modulado) DSB-SC, como visto na Eq. (4.1), é o espectro da banda base na Fig. 4.2a deslocado para a direita e para a esquerda por f_c (e multiplicado por meio), como indicado na Fig. 4.2b. Este espectro consiste em impulsos nas frequências $\pm(f_c - f_m)$ e $\pm(f_c + f_m)$. O espectro no lado externo de f_c é a USB, e o espectro no lado interno de f_c, a LSB. Observemos que o espectro DSB-SC não tem uma componente na frequência da portadora f_c. Por isso é denominado *portadora suprimida*.

Na abordagem do domínio do tempo, trabalhamos diretamente com sinais no domínio do tempo. Para o sinal em banda base $m(t) = \cos \omega_m t$, o sinal DSB-SB $\varphi_{\text{DSB-SC}}(t)$ é

$$\begin{aligned}\varphi_{\text{DSB-SC}}(t) &= m(t) \cos \omega_c t \\ &= \cos \omega_m t \cos \omega_c t \\ &= \frac{1}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]\end{aligned}$$

Isso mostra que, quando o sinal em banda base (mensagem) é uma pura senoide de frequência f_m, o sinal modulado consiste em duas senoides: uma componente de frequência $f_c + f_m$ (a USB) e uma componente de frequência $f_c - f_m$ (a LSB). A Fig. 4.2b mostra o espectro de $\varphi_{\text{DSB-SC}}(t)$. Portanto, cada componente de frequência f_m no sinal modulante origina duas componentes de frequências $f_c + f_m$ e $f_c - f_m$ no sinal modulado. Notemos o fato curioso de que não há componente da frequência portadora f_c no lado direito da equação anterior. Como mencionado, é por isso que a modulação é denominada modulação em banda lateral dupla com portadora suprimida (DSB-SC).

Agora, comprovemos que o sinal modulado $\varphi_{\text{DSB-SC}}(t) = \cos \omega_m t \cos \omega_c t$, quando aplicado à entrada do demodulador na Fig. 4.1e, leva a uma saída proporcional ao desejado sinal em banda base $\cos \omega_m t$. O sinal $e(t)$ na Fig. 4.1e é dado por

$$\begin{aligned}e(t) &= \cos \omega_m t \cos^2 \omega_c t \\ &= \frac{1}{2} \cos \omega_m t \,(1 + \cos 2\omega_c t)\end{aligned}$$

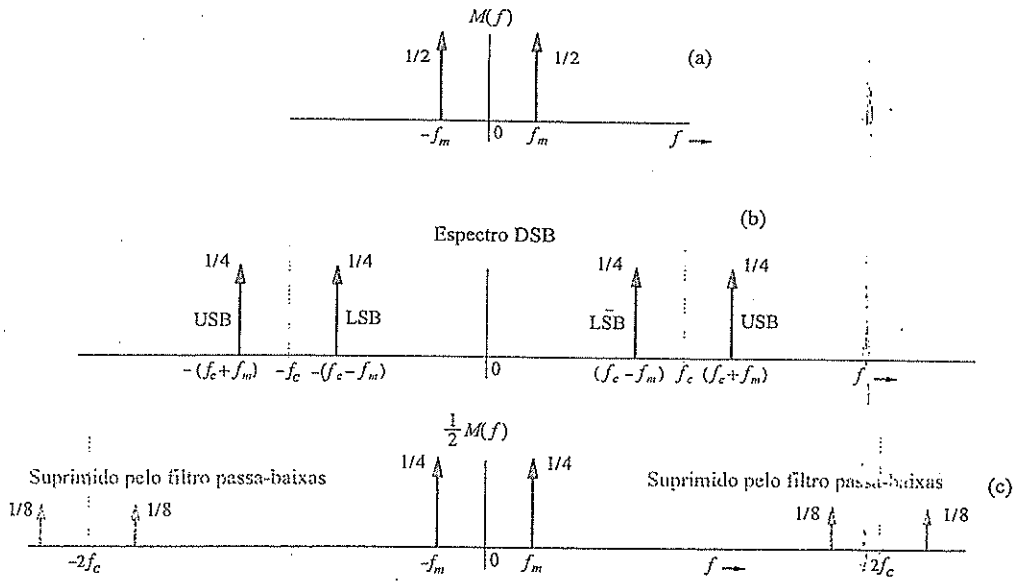


Figura 4.2 Exemplo de modulação DSB-SC.

O espectro do termo $\cos \omega_m t \cos 2\omega_c t$ é centrado em $2\omega_c$ e será suprimido pelo filtro passa-baixas, produzindo a saída $\frac{1}{2}\cos \omega_m t$. Esse resultado também pode ser obtido no domínio da frequência. A demodulação faz com que o espectro na Fig. 4.2b seja deslocado para a esquerda e para a direita por ω_c (e multiplicado por meio). Isso resulta no espectro mostrado na Fig. 4.2c. O filtro passa-baixas suprime os espectros centrados em $\pm 2\omega_c$, produzindo o espectro $\frac{1}{2}M(f)$.

Moduladores

Moduladores podem ser construídos de diferentes maneiras. A seguir, discutiremos algumas importantes categorias de moduladores.

Moduladores Multiplicadores: Neste caso, a modulação é feita diretamente com a multiplicação de $m(t)$ por $\cos \omega_c t$, com o uso de um multiplicador analógico, cuja saída é proporcional ao produto dos dois sinais de entrada. Um modulador desse tipo pode, tipicamente, ser obtido a partir de um amplificador de ganho variável, em que o parâmetro de ganho (como o β de um transistor) é controlado por um dos sinais – digamos, $m(t)$. Quando o sinal $\cos \omega_c t$ é aplicado à entrada desse multiplicador, a saída é proporcional a $m(t) \cos \omega_c t$.

No passado, a multiplicação de dois sinais em uma faixa dinâmica razoável era um desafio para projetistas de circuitos. No entanto, com o avanço contínuo da tecnologia de semicondutores, a multiplicação de sinais deixou de ser um grande problema. Contudo, apresentaremos várias configurações clássicas de moduladores que evitam o uso de multiplicadores. O estudo desses moduladores provê um inestimável entendimento do conceito e uma excelente oportunidade para aprender algumas novas técnicas de análise de sinais.

Moduladores Não Lineares: A modulação também pode ser obtida com o uso de dispositivos não lineares, como diodo ou transistor semicondutor. A Fig. 4.3 mostra um esquema possível, que faz uso de dois elementos não lineares idênticos (as caixas rotuladas NL).

Admitamos que a característica entrada-saída de qualquer dos elementos não lineares seja aproximada por uma série de potências

$$y(t) = ax(t) + bx^2(t) \tag{4.3}$$

em que $x(t)$ e $y(t)$ são a entrada e a saída, respectivamente, do elemento não linear. A saída $z(t)$ do somador na Fig. 4.3 é dada por

$$z(t) = y_1(t) - y_2(t) = [ax_1(t) + bx_1{}^2(t)] - [ax_2(t) + bx_2{}^2(t)]$$

A substituição das duas entradas $x_1(t) = \cos \omega_c t + m(t)$ e $x_2(t) = \cos \omega_c t - m(t)$ nessa equação resulta em

$$z(t) = 2a \cdot m(t) + 4b \cdot m(t) \cos \omega_c t$$

O espectro de $m(t)$ é centrado na origem, enquanto o espectro de $m(t) \cos \omega_c t$ é centrado em $\pm\omega_c$. Em consequência, quando $z(t)$ é aplicado a um filtro passa-faixa centrado em ω_c, o sinal $am(t)$ é suprimido e o desejado sinal modulado $4bm(t) \cos \omega_c t$ pode passar pelo sistema sem distorção.

Há duas entradas nesse circuito: $m(t)$ e $\cos \omega_c t$. A saída do último somador, $z(t)$, não contém uma das entradas, o sinal da portadora $\cos \omega_c t$. Em consequência, o sinal da portadora não aparece na entrada do filtro passa-faixa final. O circuito atua como uma ponte balanceada para uma das entradas (a portadora). Circuitos que têm essa característica são denominados *circuitos balanceados*. O modulador não linear na Fig. 4.3 é um exemplo da classe de moduladores conhecida como *moduladores balanceados*. Esse circuito é balanceado em relação a apenas uma das entradas (a portadora); a outra entrada, $m(t)$, ainda aparece no filtro passa-faixa final, que deve rejeitá-la. Por tal razão, o circuito é referido como *modulador simplesmente balanceado*. Um circuito balanceado em relação às duas entradas é denominado *modulado duplamente balanceado*, do qual o modulador em anel (Fig. 4.6) é um exemplo.

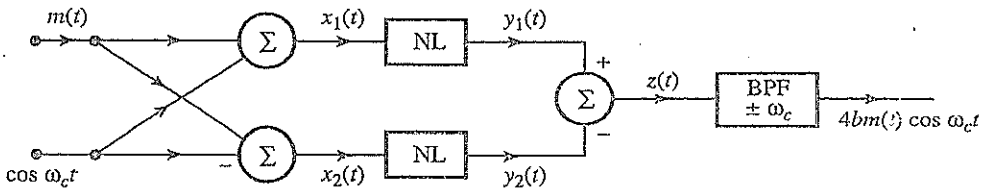


Figura 4.3 Modulador DSB-SC não linear.

Moduladores Chaveados: A operação de multiplicação requerida para modulação pode ser substituída por uma operação de chaveamento mais simples, se nos dermos conta de que um sinal modulado pode ser obtido multiplicando $m(t)$ não apenas por uma senoide pura, mas também por qualquer sinal periódico $\phi(t)$ da frequência angular fundamental ω_c. Um sinal periódico desse tipo pode ser expresso por uma série de Fourier trigonométrica como

$$\phi(t) = \sum_{n=0}^{\infty} C_n \cos(n\omega_c t + \theta_n) \tag{4.4a}$$

Logo,

$$m(t)\phi(t) = \sum_{n=0}^{\infty} C_n\, m(t) \cos(n\omega_c t + \theta_n) \tag{4.4b}$$

Isso mostra que o espectro do produto $m(t)\phi(t)$ é o espectro $M(\omega)$ deslocado para $\pm\omega_c$, $\pm 2\omega_c$, ..., $\pm n\omega_c$, ... Caso esse sinal seja aplicado a um filtro passa-faixa de largura de banda $2B$ Hz e centrado em ω_c, obteremos o desejado sinal modulado $c_1 m(t) \cos(\omega_c t + \theta_1)$.*

O trem de pulsos quadrados $w(t)$ na Fig. 4.4b é um sinal periódico, cuja série de Fourier foi calculada anteriormente como (reescrevendo os resultados do Exemplo 2.4):

$$w(t) = \frac{1}{2} + \frac{2}{\pi}\left(\cos\,\omega_c t - \frac{1}{3}\cos\,3\omega_c t + \frac{1}{5}\cos\,5\omega_c t - \cdots\right) \tag{4.5}$$

O sinal $m(t)w(t)$ é dado por

$$m(t)w(t) = \frac{1}{2}m(t) + \frac{2}{\pi}\left[m(t)\cos\,\omega_c t - \frac{1}{3}m(t)\cos\,3\omega_c t + \frac{1}{5}m(t)\cos\,5\omega_c t - \cdots\right] \tag{4.6}$$

O sinal $m(t)w(t)$ consiste não apenas na componente $m(t)$, mas também em um número infinito de sinais modulados com frequências portadoras ω_c, $3\omega_c$, $5\omega_c$, ... Portanto, o espectro de $m(t)w(t)$ consiste em múltiplas cópias do espectro $M(f)$ da mensagem, deslocadas a 0, $\pm f_c$, $\pm 3f_c$, $\pm 5f_c$,... (com pesos relativos decrescentes), como mostrado na Fig. 4.4c.

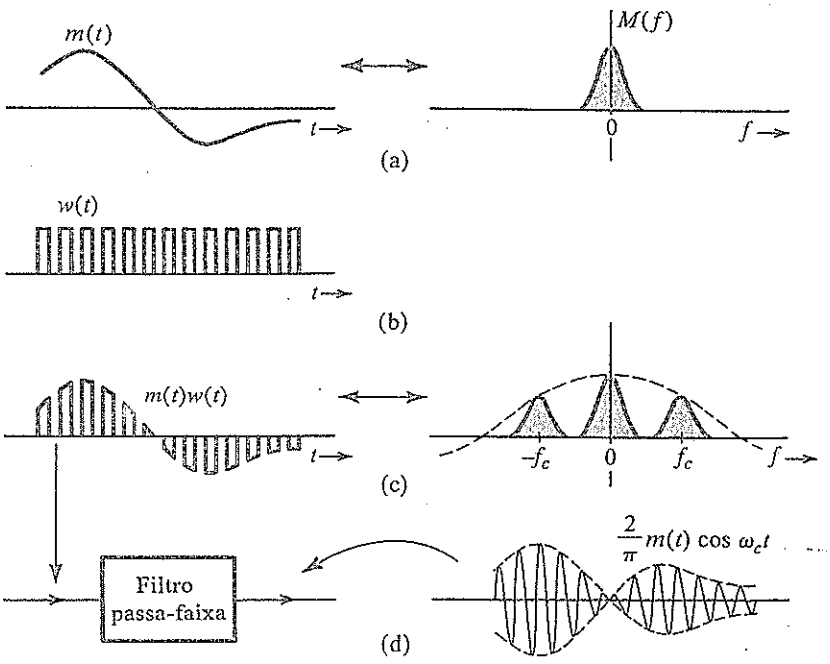


Figura 4.4 Modulador chaveado para DSB-SC.

*A fase θ_1 não é relevante.

Para modulação, nosso interesse reside na extração apenas da componente modulada $m(t)\cos\omega_c t$. Para separar essa componente da multidão, aplicamos o sinal $m(t)w(t)$ a um filtro passa-faixa de largura de banda $2B$ Hz (ou $4\pi B$ rad/s), centrado em $\pm f_c$. Desde que a frequência da portadora $f_c \geq 2B$ (ou $\omega_c \geq 4\pi B$), esse procedimento suprimirá as componentes espectrais não centradas em $\pm f_c$, produzindo o desejado sinal modulado $(2/\pi)m(t)\cos\omega_c t$ (Fig. 4.4d).

Vemos, agora, o resultado deste método. A multiplicação de um sinal por um trem de pulsos quadrados é, *na verdade*, uma operação de chaveamento, na qual o sinal $m(t)$ é ligado e desligado periodicamente; isso pode ser feito por meio de simples elementos comutadores controlados por $w(t)$. A Fig. 4.5a ilustra um desses elementos comutadores, o **modulador em ponte de diodos**, alimentado por uma senoide $A\cos\omega_c t$ para produzir a ação de chaveamento. Os diodos D_1, D_2 e D_3, D_4 são pares casados. Quando o sinal $\cos\omega_c t$ tem uma polaridade que torne o terminal c positivo em relação a d, todos os diodos conduzem. Como os diodos D_1 e D_2 são casados, os terminais a e b têm o mesmo potencial e são, efetivamente, curto-circuitados. Durante o próximo meio ciclo, o terminal d é positivo em relação a c e os quatro diodos não conduzem, fazendo com que os terminais a e b fiquem em circuito aberto. A ponte de diodos na Fig. 4.5a, portanto, funciona como uma chave eletrônica, cujos terminais a e b abrem e fecham periodicamente, com a frequência da portadora f_c, quando uma senoide $A\cos\omega_c t$ é aplicada aos terminais c e d. Para obter o sinal $m(t)w(t)$, podemos posicionar essa chave eletrônica (terminais a e b) em série (Fig. 4.5b) ou em paralelo com o sinal $m(t)$, como na Fig. 4.5c. Esses moduladores são conhecidos, respectivamente, como **modulador em ponte de diodos série** e **modulador em ponte de diodos paralela**. O chaveamento de $m(t)$ se repete a cada meio ciclo da portadora, resultando no sinal chaveado $m(t)w(t)$ que, quando aplicado a um filtro passa-faixa, produz o desejado sinal modulado $(2/\pi)m(t)\cos\omega_c t$.

Outro modulador chaveado, conhecido como **modulador em anel**, é mostrado na Fig. 4.6a. Durante o meio ciclo positivo da portadora, os diodos D_1 e D_3 conduzem, enquanto os diodos D_2 e D_4 estão abertos. Assim, o terminal a é conectado a c e o terminal b, a d. Durante o meio ciclo negativo da portadora, os diodos D_1 e D_3 estão abertos, enquanto os diodos D_2 e D_4 conduzem, de modo que o terminal a é conectado a d e o terminal b, a c. Em consequência, a saída é proporcional a $m(t)$ durante o meio ciclo positivo e a $-m(t)$ durante o meio ciclo negativo. Efetivamente, $m(t)$ é multiplicado por um trem de pulsos quadrados $w_0(t)$, como mostrado na Fig. 4.6b. A série de Fourier para $w_0(t)$ pode ser obtida com o uso do sinal $w(t)$ da Eq. (4.5), resultando em $w_0(t) = 2w(t) - 1$. Portanto, podemos usar a série de Fourier de $w(t)$ [Eq. (4.5)] para determinar a série de Fourier de $w_0(t)$ como

$$w_0(t) = \frac{4}{\pi}\left(\cos\omega_c t - \frac{1}{3}\cos 3\omega_c t + \frac{1}{5}\cos 5\omega_c t - \cdots\right) \tag{4.7a}$$

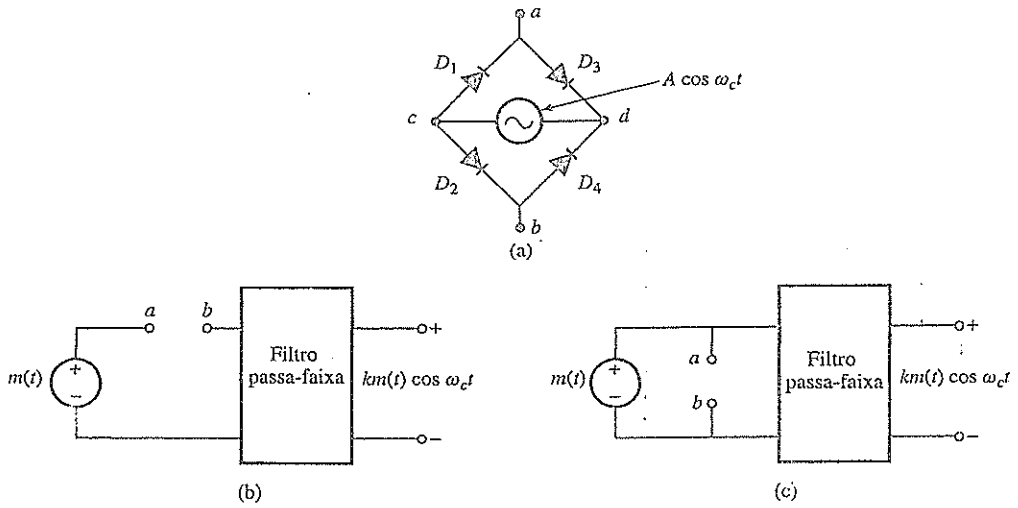
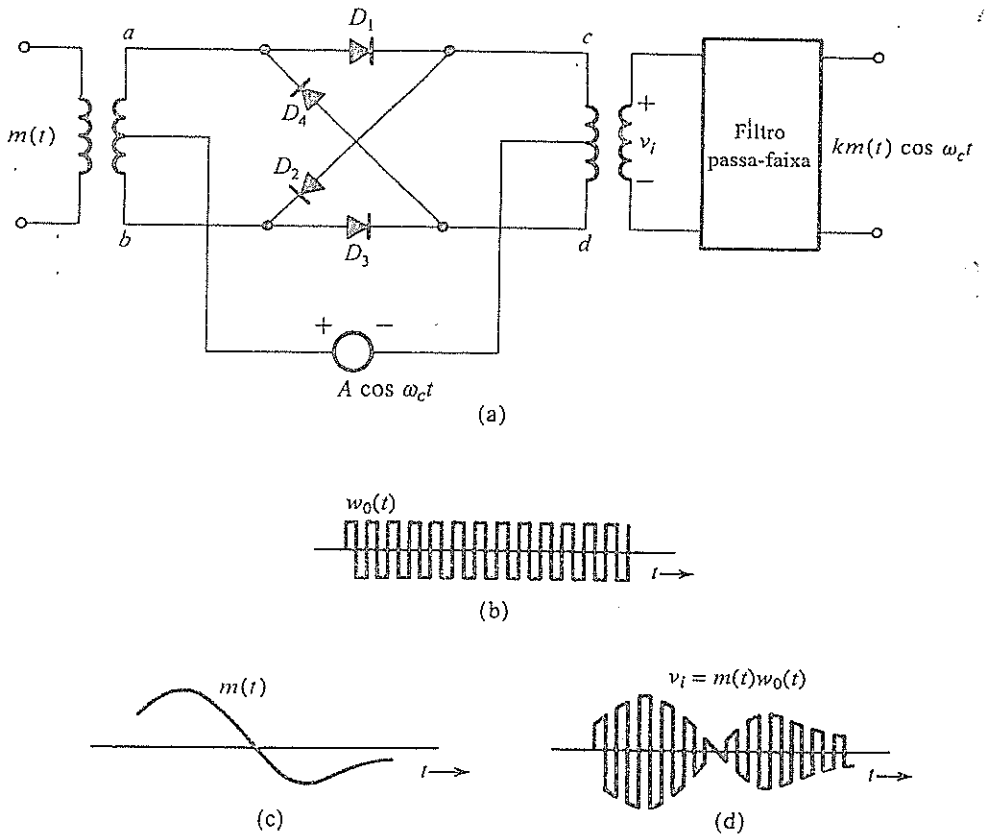


Figura 4.5 (a) Chave eletrônica em ponte de diodos, (b) Modulador em ponte de diodos série, (c) Modulador em ponte de diodos paralela.

Figura 4.6
Modulador em anel.



Assim, temos

$$v_i(t) = m(t)w_0(t) = \frac{4}{\pi}\left[m(t)\cos\omega_c t - \frac{1}{3}m(t)\cos 3\omega_c t + \frac{1}{5}m(t)\cos 5\omega_c t - \cdots\right] \tag{4.7b}$$

O sinal $m(t)w_0(t)$ é representado na Fig. 4.6d. Quando essa forma de onda é aplicada a um filtro passa-faixa centrado em ω_c (Fig. 4.6a), a saída do filtro será o desejado sinal $(4/\pi)m(t)\cos\omega_c t$.

Esse circuito tem duas entradas: $m(t)$ e $\cos\omega_c t$. Nenhum desses sinais aparece na entrada do filtro passa-faixa final. Assim, esse circuito é um exemplo de um **modulador duplamente balanceado**.

Exemplo 4.2 Misturador ou Conversor de Frequências

Analisemos um misturador de frequências, ou conversor de frequências, usado para alterar a frequência portadora de um sinal modulado $m(t)\cos\omega_c t$ de ω_c para outra frequência ω_I.

Isso pode ser feito com a multiplicação de $m(t)\cos\omega_c t$ por $2\cos\omega_{mix}t$, em que $\omega_{mix} = \omega_c + \omega_I$ ou $\omega_c - \omega_I$ e a aplicação do produto a um filtro passa-faixa, como indicado na Fig. 4.7a.
O produto $x(t)$ é

$$\begin{aligned} x(t) &= 2m(t)\cos\omega_c t\cos\omega_{mix}t \\ &= m(t)[\cos(\omega_c - \omega_{mix})t + \cos(\omega_c + \omega_{mix})t] \end{aligned}$$

Se escolhermos $\omega_{mix} = \omega_c - \omega_I$, temos

$$x(t) = m(t)[\cos\omega_I t + \cos(2\omega_c + \omega_I)t]$$

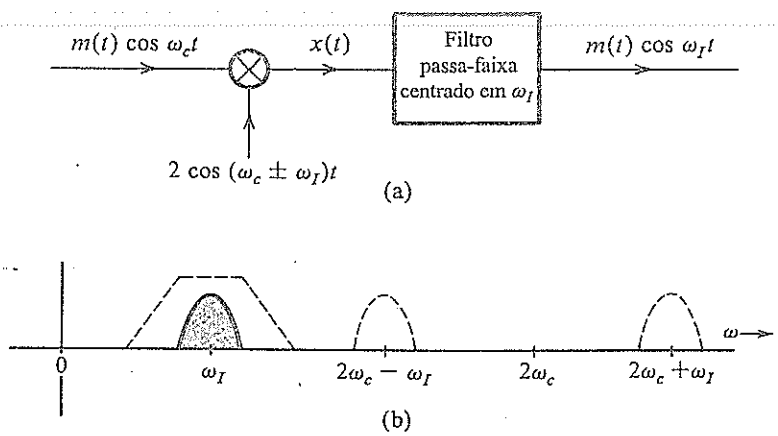


Figura 4.7 Misturador ou conversor de frequências.

Se escolhermos $\omega_{\text{mix}} = \omega_c + \omega_I$, obtemos

$$x(t) = m(t)[\cos\omega_I t + \cos(2\omega_c + \omega_I)t]$$

Em qualquer dos casos, desde que $\omega_c - \omega_I \geq 2\pi B$ e $\omega_I \geq 2\pi B$, os vários espectros na Fig. 4.7b não se sobreporão. Em consequência, um filtro passa-faixa na saída, centrado em ω_I, deixará passar o termo $m(t)\cos\omega_I t$ e suprimirá o outro termo, produzindo a saída $m(t)\cos\omega_I t$. Dessa forma, a frequência portadora foi trasladada de ω_c para ω_I.

A operação de mistura/conversão de frequências (também conhecida como heterodinagem) consiste, basicamente, na translação do espectro por ω_{mix}. Isso é equivalente à operação de modulação, com uma frequência portadora de modulação (a frequência ω_{mix} do oscilador local) que difere da frequência de entrada por ω_I. Qualquer um dos moduladores discutidos anteriormente pode ser usado para a mistura de frequências. Quando selecionamos a frequência portadora local como $\omega_{\text{mix}} = \omega_c + \omega_I$, a operação é denominada **conversão ascendente** (*upconversion*); quando selecionamos a frequência portadora local como $\omega_{\text{mix}} = \omega_c - \omega_I$, a operação é denominada **conversão descendente** (*downconversion*).

Demodulação de Sinais DSB-SC

Como discutido anteriormente, a demodulação de um sinal DSB-SC envolve, essencialmente, a multiplicação pelo sinal da portadora e é idêntica à modulação (Fig. 4.1). No receptor, multiplicamos o sinal de entrada por uma portadora local de frequência e fase em sincronismo com a portadora recebida. O produto é, então, aplicado a um filtro passa-baixas. A **única diferença** entre modulador e demodulador se refere ao sinal de entrada e ao filtro de saída. No modulador, a mensagem $m(t)$ é a entrada e a saída do multiplicador é aplicada a um filtro passa-faixa centrado em ω_c; no demodulador, a entrada é o sinal DSB-SC e a saída do multiplicador é aplicada a um filtro passa-baixas. Portanto, todos os moduladores discutidos anteriormente, sem multiplicadores, podem ser usados como demoduladores, desde que os filtros passa-faixa de saída sejam substituídos por filtros passa-baixas de largura de banda B.

Para demodulação, o receptor deve gerar uma portadora em sincronismo de fase e de frequência com a portadora de entrada. Esses demoduladores são também denominados demoduladores **síncronos** ou **coerentes** (ou, ainda, **homódinos**).

Exemplo 4.3 Analisemos o demodulador chaveado que utiliza a chave eletrônica (ponte de diodos) na Fig. 4.5a como comutador (em série ou em paralelo).

O sinal de entrada é $m(t)\cos\omega_c t$. A portadora provoca o chaveamento periódico do sinal de entrada. Portanto, a saída é $m(t)\cos\omega_c t \times w(t)$. Fazendo uso da identidade trigonométrica $\cos x \cos y = 0{,}5[\cos(x+y) + \cos(x-y)]$, obtemos

$$\begin{aligned} m(t)\cos\omega_c t \times w(t) &= m(t)\cos\omega_c t\left[\frac{1}{2} + \frac{2}{\pi}\left(\cos\omega_c t - \frac{1}{3}\cos 3\omega_c t + \cdots\right)\right] \\ &= \frac{2}{\pi}m(t)\cos^2\omega_c t + \text{termos da forma } m(t)\cos n\omega_c t \\ &= \frac{1}{\pi}m(t) + \frac{1}{\pi}m(t)\cos 2\omega_c t + \text{termos da forma } m(t)\cos n\omega_c t \end{aligned}$$

Os termos da forma $m(t)\cos n\omega_c t$ têm espectros centrados em $\pm n\omega_c$ e são suprimidos pelo filtro passa-baixas, que produz a saída $(1/\pi)\,m(t)$. Deixamos como exercício para o leitor a comprovação de que, ao funcionar como demodulador, a saída do circuito em anel na Fig. 4.6a (com um filtro passa-baixas na saída) é $(2/\pi)\,m(t)$ (duas vezes a saída do demodulador chaveado desse exemplo).

4.3 MODULAÇÃO EM AMPLITUDE (AM)

Na seção anterior, iniciamos a discussão da modulação em amplitude com a modulação em amplitude DSB-SC, pois essa pode ser facilmente entendida e analisada tanto no domínio do tempo como no domínio da frequência. Contudo, nem sempre a simplicidade analítica equivale à simplicidade de implementação prática. A demodulação (coerente) de um sinal DSB-SC requer que o receptor possua um sinal de portadora em sincronismo com a portadora recebida. Na prática, essa exigência não é atendida com facilidade. Como o sinal modulado pode ter viajado centenas de quilômetros e pode ter sofrido algum deslocamento de frequência desconhecido, o sinal recebido tem, na verdade, a forma

$$r(t) = A_c m(t-t_0)\cos[(\omega_c + \Delta\omega)(t-t_0)] = A_c m(t-t_0)\cos[(\omega_c+\Delta\omega)t - \theta_d)]$$

em que $\Delta\omega$ representa o efeito de deslocamento Doppler e

$$\theta_d = (\omega_c + \Delta\omega)t_d$$

resulta de um atraso temporal t_0 desconhecido. Para demodulação coerente, o receptor deve ser suficientemente sofisticado para gerar um oscilador local $\cos[(\omega_c+\Delta\omega)t - \theta_d]$ a partir apenas do sinal recebido $r(t)$. Um receptor desse tipo pode ser de difícil implementação e ter custo elevado. Em sistemas de difusão, o alto custo deve ser evitado, pois podem existir numerosos receptores para cada transmissor.

Como alternativa ao demodulador coerente, o transmissor deve enviar uma portadora $A\cos\omega_c t$ [juntamente com o sinal modulado $m(t)\cos\omega_c t$], de modo que não seja necessária a geração de uma portadora no receptor. Nesse caso, o transmissor deve transmitir a uma potência de saída muito mais alta, o que aumenta os custos da alternativa. Em comunicação ponto a ponto, em que há um transmissor para cada receptor, uma maior complexidade no receptor pode ser justificável, desde que seu custo seja compensado por um transmissor mais barato. Em sistemas de difusão, em que há um grande número de receptores para cada transmissor, é mais econômico ter um transmissor caro de alta potência e receptores simples e baratos, pois qualquer redução de custo de receptores é multiplicada pelo número de unidades receptoras. Por isso, sistemas de difusão tendem a privilegiar a transferência de custos dos (numerosos) receptores para os (poucos) transmissores.

A segunda opção – transmissão de uma portadora juntamente com o sinal modulado – é a escolha óbvia no caso de sistemas de difusão, tendo em vista os benefícios de custos. Isso nos leva aos chamados sistemas AM (de modulação em amplitude), em que o sinal transmitido $\varphi_{AM}(t)$ é dado por

$$\varphi_{AM}(t) = A\cos\omega_c t + m(t)\cos\omega_c t \tag{4.8a}$$

$$= [A + m(t)]\cos\omega_c t \tag{4.8b}$$

O espectro de $\varphi_{AM}(t)$ é basicamente o mesmo que o de $\varphi_{DSB\text{-}SC}(t) = m(t)\cos\omega_c t$, exceto pelos dois impulsos adicionais em $\pm f_c$,

$$\varphi_{AM}(t) \iff \frac{1}{2}[M(f+f_c)+M(f-f_c)]+\frac{A}{2}[\delta(f+f_c)+\delta(f-f_c)] \tag{4.8c}$$

Uma comparação entre $\varphi_{AM}(t)$ e $\varphi_{DSB\text{-}SC}(t) = m(t)\cos\omega_c t$ deixa claro que o sinal AM é idêntico ao sinal DSB-SC, com $A + m(t)$ como o sinal modulante [no lugar de $m(t)$]. O valor de A é sempre tomado como positivo. Portanto, para esboçar a forma de onda de $\varphi_{AM}(t)$, desenhamos o envelope ou envoltória $|A + m(t)|$ e sua imagem $-|A + m(t)|$, e preenchemos o espaço entre as duas curvas com a senoide na frequência da portadora f_c. O valor de A afeta o envelope temporal do sinal modulado.

Os dois casos são considerados na Fig. 4.8. No primeiro caso, A tem valor suficientemente grande para que $A + m(t) \geq 0$ seja sempre não negativo. No segundo, o valor de A não é bastante para satisfazer esta condição. No primeiro caso, o envelope tem a mesma forma que $m(t)$ (deslocada de um valor dc constante A). No segundo, a formas do envelope e de $m(t)$ são diferentes, pois a parte negativa de $A + m(t)$ é retificada. Isso significa que podemos detectar o desejado sinal $m(t)$ se detectarmos o envelope no primeiro caso, em que $A + m(t) > 0$. Essa detecção não é possível no segundo caso. Veremos que a detecção de envelope é uma operação extremamente simples e barata, que não requer a geração de uma portadora local para demodulação. No entanto, como visto anteriormente, o envelope AM contém informação sobre $m(t)$ apenas se o sinal AM $[A + m(t)]\cos\omega_c t$ satisfizer a condição $A + m(t) > 0$ para todo t.

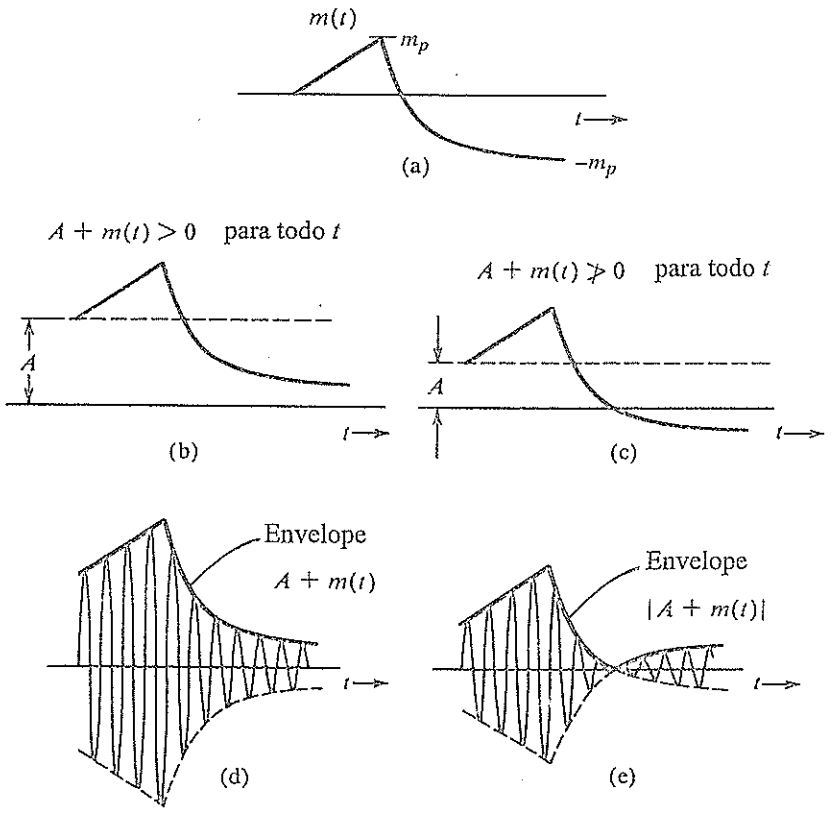


Figura 4.8 Sinal AM e seu envelope.

Agora, sejamos mais específicos a respeito da definição de "envelope". Consideremos um sinal $E(t)\cos\omega_c t$. Caso $E(t)$ varie lentamente em comparação à portadora senoidal $\cos\omega_c t$, o envelope de $E(t)\cos\omega_c t$ é $|E(t)|$. Isso significa [ver Eq. (4.8b)] que, se e somente se $A + m(t) \geq 0$ para todo t, o envelope de $\varphi_{AM}(t)$ é

$$|A + m(t)| = A + m(t)$$

Em outras palavras, para que a detecção de envelope detecte $m(t)$ corretamente, duas condições devem ser satisfeitas:

(a) $f_c >>$ largura de banda de $m(t)$
(b) $A + m(t) \geq 0$

Essa conclusão pode ser prontamente comprovada a partir da Fig. 4.8d e e. Na Fig. 4.8d, em que $A + m(t) \geq 0$, $A + m(t)$ é, de fato, o envelope e $m(t)$ pode ser recuperado desse envelope. Na Fig. 4.8e, em que $A + m(t)$ não é sempre positivo, o envelope $|A + m(t)|$ é retificado de $A + m(t)$ e $m(t)$ não pode ser recuperado desse envelope. Em consequência, na Fig. 4.8d, a demodulação de $\varphi_{AM}(t)$ corresponde à simples detecção de envelope. Assim, a **condição para detecção de envelope** de um sinal AM é

$$A + m(t) \geq 0 \qquad \text{para todo } t \tag{4.9a}$$

Se $m(t) \geq 0$ para todo t, $A = 0$ satisfaz a condição (4.9a). Nesse caso, não há necessidade de adicionar qualquer portadora, pois o envelope do sinal DSB-SC $m(t) \cos \omega_c t$ é $m(t)$, de modo que esse sinal DSB-SC pode ser detectado via detecção de envelope. Na discussão a seguir, presumiremos que $m(t) \ngeq 0$ para todo t, ou seja, $m(t)$ pode ser negativo para alguns valores de t.

Sinais de Mensagem $m(t)$ com Deslocamento Nulo: Sejam $\pm m_p$ os valores máximo e mínimo de $m(t)$, respectivamente (Fig. 4.8). Isto significa que $m(t) \geq -m_p$. Logo, a condição para detecção de envelope (4.9a) é equivalente a

$$A \geq -m_{\min} \tag{4.9b}$$

Dessa forma, a mínima amplitude de portadora necessária para viabilizar a detecção de envelope é m_p. Isso fica bem claro da Fig. 4.8. Definimos o índice de modulação μ como

$$\mu = \frac{m_p}{A} \tag{4.10a}$$

Para que a detecção de envelope seja sem distorção, a condição é $A \geq m_p$. Portanto,

$$0 \leq \mu \leq 1 \tag{4.10b}$$

é a condição necessária para demodulação sem distorção de sinais AM com detector de envelope.

Quando $A < m_p$, a Eq. (4.10a) mostra que $\mu > 1$ (sobremodulação). Nesse caso, a opção de detecção de envelope deixa de ser viável. Precisamos, então, lançar mão da demodulação síncrona. Vale notar que a demodulação síncrona pode ser usada para qualquer valor de μ, pois o demodulador recuperará o sinal $A + m(t)$. Basta um bloco dc adicional para remover a tensão DC A. O detector de envelope, consideravelmente mais simples e barato que o detector síncrono, pode ser usado apenas para $\mu \leq 1$.

Sinais de Mensagem m(t) com Deslocamento Não Nulo: Em raras ocasiões, o sinal de mensagem $m(t)$ terá um deslocamento não nulo, de modo que seus valores máximo, $m_{\max}$, e mínimo, $m_{\min}$, não sejam simétricos:

$$m_{\min} \neq -m_{\max}$$

Nesse caso, podemos observar que qualquer deslocamento do envelope não altera a forma da saída do detector de envelope. Na verdade, um deslocamento constante não carrega qualquer informação.

Assim, a detecção de envelope ainda seria sem distorção se

$$0 \leq \mu \leq 1 \tag{4.11a}$$

com a definição do índice de modulação modificada para

$$\mu = \frac{m_{\max} - m_{\min}}{2A + m_{\max} + m_{\min}} \tag{4.11b}$$

Exemplo 4.4 Esbocemos a forma de onda de $\varphi_{AM}(t)$ para índices de modulação $\mu = 0{,}5$ e $\mu = 1$, com $m(t) = b \cos \omega_m t$. Esse caso é referido como **modulação por tom**, pois o sinal modulante é uma senoide pura (ou tom puro).

Aqui, $m_{max} = b$ e $m_{min} = -b$. Logo, segundo a Eq. (4.10a), o índice de modulação é

$$\mu = \frac{b - (-b)}{2A + b + (-b)} = \frac{b}{A}$$

Com isso, $b = \mu A$ e

$$m(t) = b \cos \omega_m t = \mu A \cos \omega_m t$$

Portanto,

$$\varphi_{AM}(t) = [A + m(t)] \cos \omega_c t = A[1 + \mu \cos \omega_m t] \cos \omega_c t$$

A Fig. 4.9 mostra os sinais modulados correspondentes a $\mu = 0{,}5$ e $\mu = 1$, respectivamente.

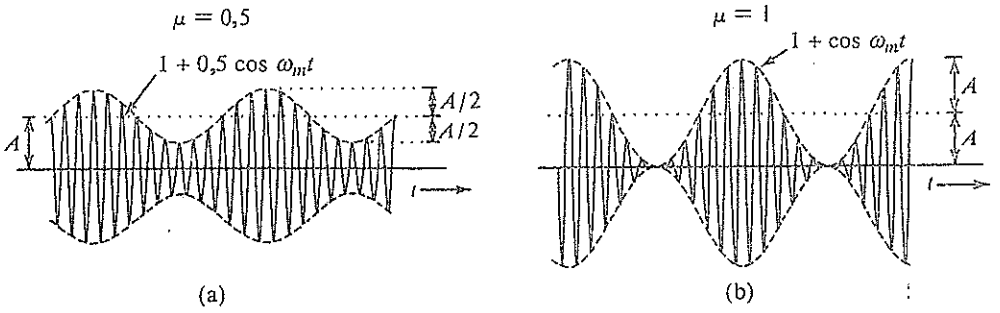


Figura 4.9 Sinal AM modulado por tom. (a) $\mu = 0{,}5$ e (b) $\mu = 1$.

Potências da Banda Lateral e da Portadora

Em AM, a vantagem da detecção de envelope tem um preço, pois o termo da portadora não transporta informação, de modo que, desse ponto de vista, potência da portadora é desperdiçada:

$$\varphi_{AM}(t) = \underbrace{A \cos \omega_c t}_{\text{portadora}} + \underbrace{m(t) \cos \omega_c t}_{\text{banda lateral}}$$

A potência da portadora P_c é o valor quadrático médio de $A \cos \omega_c t$, que é $A^2/2$. A potência da banda lateral P_s é a potência de $m(t) \cos \omega_c t$, que é $0{,}5\ \overline{m^2(t)}$ [Eq. (3.93)]. Logo,

$$P_c = \frac{A^2}{2} \qquad \text{e} \qquad P_s = \frac{1}{2}\overline{m^2(t)}$$

A informação útil da mensagem reside na potência da banda lateral, enquanto a potência da portadora é usada por conveniência na modulação e na demodulação. A potência total é a soma das potências da portadora (desperdiçada) e da banda lateral (útil). A eficiência de potência, η, é definida como

$$\eta = \frac{\text{potência útil}}{\text{potência total}} = \frac{P_s}{P_c + P_s} = \frac{\overline{m^2(t)}}{A^2 + \overline{m^2(t)}} 100\%$$

Para o caso especial de modulação por tom

$$m(t) = \mu A \cos \omega_m t \qquad \text{e} \qquad \overline{m^2(t)} = \frac{(\mu A)^2}{2}$$

Logo,

$$\eta = \frac{\mu^2}{2+\mu^2}100\%$$

com a condição $0 \leq \mu \leq 1$. Podemos observar que η cresce monotonamente com μ e que η_{max} ocorre em $\mu = 1$, de modo que

$$\eta_{max} = 33\%$$

Portanto, para modulação por tom, na melhor condição ($\mu = 1$), apenas um terço da potência transmitida é usada para transportar mensagens. No caso de sinais práticos, a eficiência é ainda pior, da ordem de 25% ou menos, em comparação com sinais DSB-SC. A melhor condição implica $\mu = 1$. Menores valores de μ degradam a eficiência ainda mais. Por essa razão, compressão de volume e limitação de pico são comumente empregadas em AM para assegurar que modulação completa ($\mu = 1$) é mantida na maior parte do tempo.

Exemplo 4.5 Determinemos η e a porcentagem da potência total transportada pelas bandas laterais da onda AM com modulação por tom, quando $\mu = 0{,}5$ e $\mu = 0{,}3$.

Para $\mu = 0{,}5$,

$$\eta = \frac{\mu^2}{2+\mu^2}100\% = \frac{(0{,}5)^2}{2+(0{,}5)^2}100\% = 11{,}11\%$$

Logo, apenas 11% da potência total reside nas bandas laterais. Para $\mu = 0{,}3$,

$$\eta = \frac{(0{,}3)^2}{2+(0{,}3)^2}100\% = 4{,}3\%$$

Ou seja, apenas 4,3% da potência está nas bandas laterais que contêm o sinal de mensagem.

Geração de Sinais AM

Em princípio, a geração de sinais AM é idêntica à de sinais DSB-SC discutidos na Seção 4.2, exceto pela componente adicional de portadora $A \cos \omega_c t$, que deve ser adicionada ao sinal DSB-SC.

Demodulação de Sinais AM

Como no caso de sinais DSB-SC, o sinal AM pode ser demodulado de forma coerente por uma portadora gerada localmente. Contudo, a demodulação AM coerente ou síncrona viola o propósito de AM, pois não tira vantagem da componente adicional de portadora $A \cos \omega_c t$. Como vimos, no caso $\mu \leq 1$, o envelope do sinal AM segue o sinal de mensagem $m(t)$. A seguir, consideremos dois métodos incoerentes de demodulação AM sob a condição $0 < \mu \leq 1$: detecção por retificação e detecção de envelope.

Detector Retificador: Se um sinal AM for aplicado a um circuito consistindo em um diodo e um resistor (Fig. 4.10), a parte negativa da onda AM será removida. A saída através do resistor é uma versão retificada de meia onda do sinal AM. Visualmente, o diodo funciona como um tesoura que corta os meios ciclos negativos da senoide modulada. Em essência, na saída do retificador, o sinal AM é multiplicado por $w(t)$. Logo, a saída retificada de meia onda $v_R(t)$ é

$$v_R(t) = \{[A + m(t)] \cos \omega_c t\} w(t) \tag{4.12}$$

$$= [A + m(t)] \cos \omega_c t \left[\frac{1}{2} + \frac{2}{\pi}\left(\cos \omega_c t - \frac{1}{3}\cos 3\omega_c t + \frac{1}{5}\cos 5\omega_c t - \cdots\right)\right] \tag{4.13}$$

$$= \frac{1}{\pi}[A + m(t)] + \text{ outros termos de frequências mais altas} \tag{4.14}$$

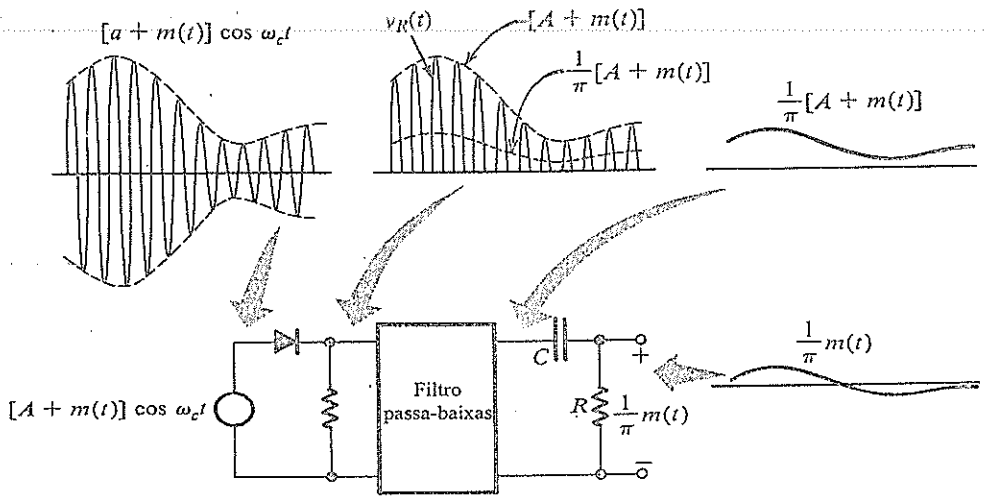


Figura 4.10 Detector retificador de AM.

Quando $v_R(t)$ é aplicado a um filtro passa-baixas com frequência de corte B Hz, a saída é $[A + m(t)]/\pi$ e todos os outros termos em v_R de frequências maiores que B Hz são suprimidos. O termo dc A/π pode ser bloqueado por um capacitor (Fig. 4.10), resultando na saída desejada $m(t)/\pi$. A saída pode ser dobrada com o uso de um retificador de onda completa.

É interessante observar que, por causa da multiplicação por $w(t)$, a detecção por retificação é, para todos os efeitos, uma detecção síncrona realizada sem o uso de um oscilador local. Em AM, a alta componente da portadora assegura que as passagens por zero sejam periódicas e, portanto, que informação a respeito da frequência e da fase da portadora no transmissor estejam embutidas no próprio sinal AM.

Detector de Envelope: A saída de um detector de envelope segue o envelope do sinal modulado. O circuito simples ilustrado na Fig. 4.11a funciona como um detector de envelope. No ciclo positivo do sinal de entrada, a entrada cresce e pode ultrapassar a tensão carregada no capacitor $v_C(t)$. Ligando o diodo e permitindo que o capacitor C se carregue até a tensão de pico do ciclo do sinal de entrada. À medida que o sinal de entrada cai abaixo do valor de pico, seu valor rapidamente fica abaixo da tensão no capacitor (que é muito próxima do valor de pico), causando a abertura do diodo. Agora, o capacitor se descarrega através do resistor R a uma taxa mais baixa (com constante de tempo RC). Durante o próximo ciclo positivo, o mesmo drama se repete. Quando o sinal de entrada ultrapassa a tensão no capacitor, o diodo volta a conduzir. Novamente, o capacitor se carrega até o valor de pico deste (novo) ciclo. O capacitor se descarrega lentamente durante o ciclo negativo.

Durante cada ciclo positivo, o capacitor se carrega até a tensão de pico do sinal de entrada e, então, se descarrega lentamente até o próximo ciclo positivo, como ilustrado na Fig. 4.11b. Assim, a tensão de saída, $v_C(t)$, segue o (crescente) envelope do sinal AM de entrada. Igualmente importante, o lento descarregamento do capacitor através do resistor R permite que a tensão no capacitor siga o envelope decrescente. O descarregamento do capacitor entre picos positivos origina um sinal oscilatório de frequência ω_c na saída. Essa oscilação (*ripple*) pode ser reduzida com a escolha de uma maior constante de tempo RC, de modo que o capacitor se descarregue muito pouco entre picos positivos ($RC >> 1/\omega_c$). Contudo, a escolha de um valor muito grande para RC impossibilitaria que a tensão no capacitor seguisse um envelope de declínio rápido (ver Fig. 4.11b). Uma vez que a máxima taxa de declínio do envelope AM é determinada pela largura de banda B do sinal de mensagem $m(t)$, o critério de projeto para RC deve ser

$$1/\omega_c \ll RC < 1/(2\pi B) \qquad \text{ou} \qquad 2\pi B < \frac{1}{RC} \ll \omega_c$$

A saída do detector de envelope é $v_C(t) = A + m(t)$, com um *ripple* de frequência ω_c. O termo dc A pode ser bloqueado por um capacitor ou um simples filtro passa-altas RC. O *ripple* pode ser ainda mais reduzido com o uso de um outro filtro RC (passa-baixas).

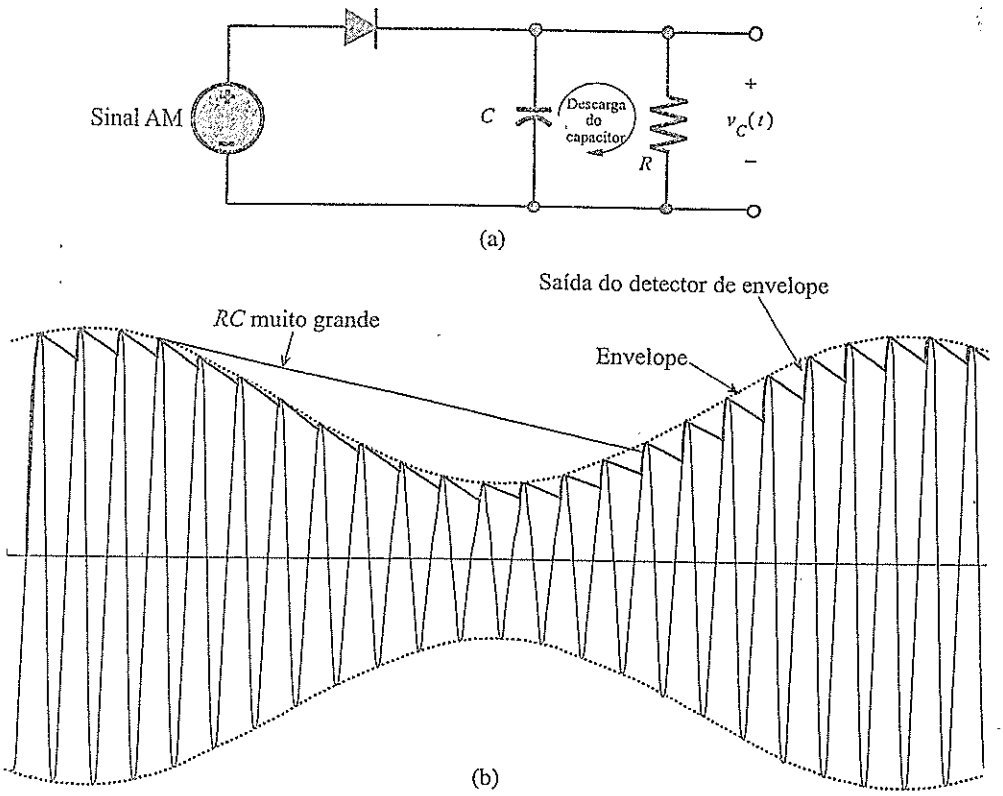


Figura 4.11
Detector de envelope para AM.

4.4 MODULAÇÕES EM AMPLITUDE COM EFICIÊNCIA DE LARGURA DE BANDA

Como ilustrado na Fig. 4.12, o espectro DSB (tanto para portadora suprimida como para AM) tem duas bandas laterais: uma banda lateral superior (USB) e uma banda lateral inferior (LSB), cada uma contendo a informação completa do sinal em banda base $m(t)$. Em consequência, para um sinal em banda base $m(t)$ com largura de banda B Hz, modulações DSB requerem, para transmissão, o dobro da largura de banda de radiofrequência. Para melhorar a eficiência espectral da modulação em amplitude, existem dois esquemas básicos que ou tiram proveito da redundância espectral de 100% ou a removem:

- Modulação em banda lateral simples (SSB – *single-sideband*), que remove a LSB ou a USB e usa apenas uma largura de banda de B Hz para um sinal de mensagem $m(t)$;
- modulação em amplitude em quadratura (QAM – *quadrature amplitude modulation*), que usa a redundância espectral para enviar duas mensagens na mesma largura de banda de $2B$ Hz.

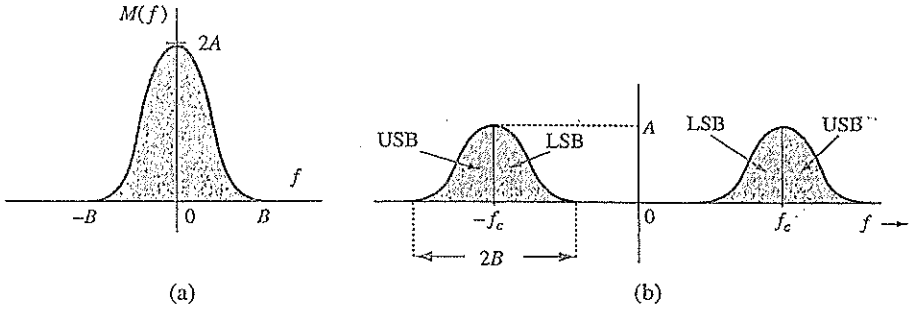


Figura 4.12
(a) Espectro original da mensagem.
(b) Redundância de largura de banda na modulação DSB.

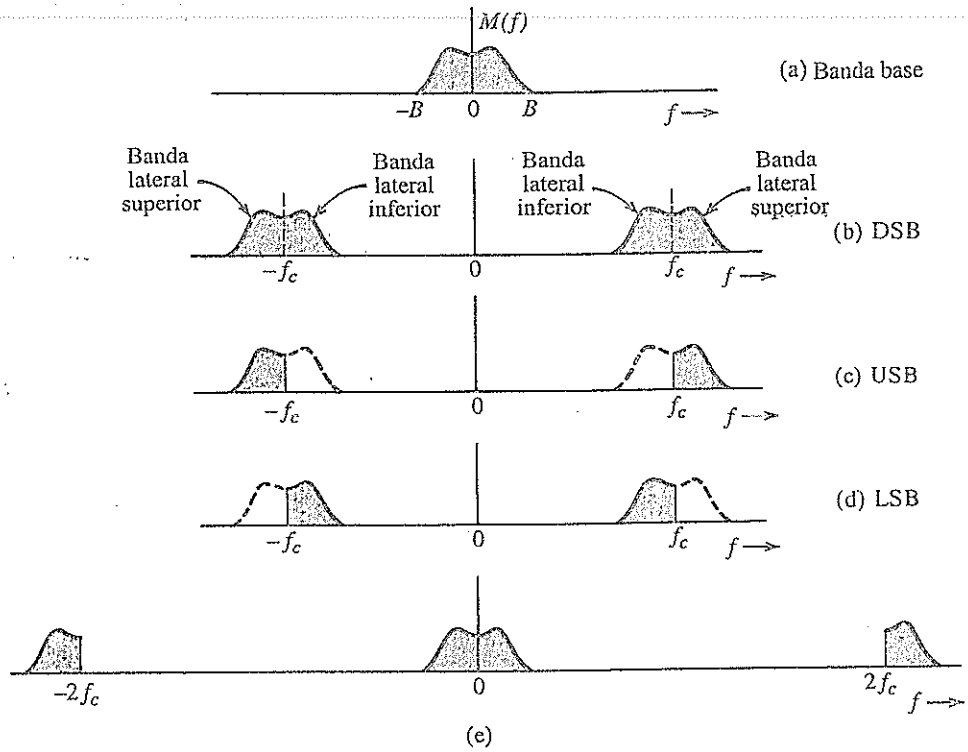


Figura 4.13 Espectros SSB, obtidos com supressão de uma das bandas laterais de DSB.

Modulação em Amplitude: Banda Lateral Simples (SSB)

Como ilustrado na Fig. 4.13, a LSB ou a USB pode ser suprimida do sinal DSB via filtragem passa-faixa. Esquemas em que apenas uma das bandas laterais é transmitida são conhecidos como de **transmissão em banda lateral simples (SSB)** e requerem apenas a metade da largura de banda do sinal DSB.

Um sinal SSB pode ser demodulado de modo coerente (síncrono), da mesma forma que um sinal DSB-SC. Por exemplo, a multiplicação de um sinal USB (Fig. 4.13c) por $\cos \omega_c t$ desloca o espectro do mesmo para a esquerda e para a direita por ω_c, resultando no espectro ilustrado na Fig. 4.13e. Uma filtragem passa-baixas desse sinal produz o desejado sinal em banda base. O mesmo se aplica a sinais LSB. Como a demodulação de sinais SSB é idêntica à de sinais DSB, os transmissores podem utilizar apenas a metade da largura de banda de sinais DSB-SC, sem qualquer custo adicional nos receptores. Como uma componente de portadora não acompanha o sinal modulado SSB, as saídas desse tipo de moduladores são referidas como sinais com portadora suprimida (SSB-SC).

Transformada de Hilbert

A seguir, apresentaremos uma nova ferramenta, que usaremos mais adiante, conhecida como **transformada de Hilbert**. Usamos $x_h(t)$ e $\mathcal{H}\{x(t)\}$ para denotar a transformada de Hilbert do sinal $x(t)$:

$$x_h(t) = \mathcal{H}\{x(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(\alpha)}{t-\alpha}\, d\alpha \tag{4.15}$$

Vale observar que o lado direito da Eq. (4.15) tem a forma de uma convolução

$$x(t) * \frac{1}{\pi t}$$

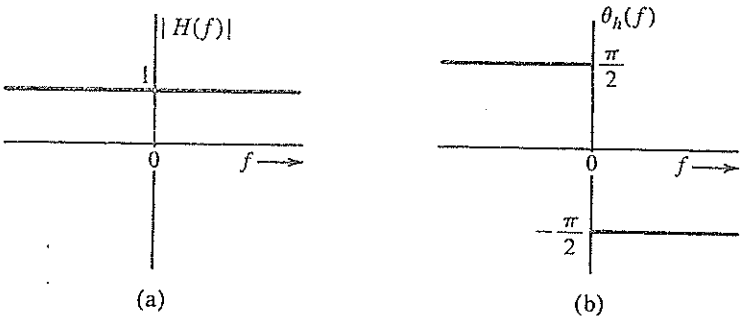


Figura 4.14 Função de transferência de um deslocador de fase ideal de $\pi/2$ (transformador de Hilbert).

A aplicação da propriedade de dualidade do par 12 da Tabela 3.1 resulta em $1/\pi t \Longleftrightarrow -j\,\mathrm{sgn}\,(f)$. Aplicação da propriedade da convolução temporal à convolução da Eq. (4.15) leva a

$$X_h(f) = -jX(f)\,\mathrm{sgn}\,(f) \tag{4.16}$$

Da Eq. (4.16), temos que, se $m(t)$ passar por uma função de transferência $H(f) = -j\,\mathrm{sgn}\,(f)$, a saída será $m_h(t)$, a transformada de Hilbert de $m(t)$. Como

$$H(f) = -j\,\mathrm{sgn}\,(f) \tag{4.17}$$

$$= \begin{cases} -j = 1 \cdot e^{-j\pi/2} & f > 0 \\ j = 1 \cdot e^{j\pi/2} & f < 0 \end{cases} \tag{4.18}$$

temos que $|H(f)| = 1$ e $\theta_h(f) = -\pi/2$ para $f > 0$, e $\pi/2$ para $f < 0$, como mostrado na Fig. 4.14. Dessa forma, se alterarmos a fase de cada componente de $m(t)$ de $\pi/2$ (sem alterar sua amplitude), o sinal resultante será $m_h(t)$, a transformada de Hilbert de $m(t)$. Portanto, um transformador de Hilbert é um deslocador de fase ideal que translada a fase de cada componente espectral de $-\pi/2$.

Representação de Sinais SSB no Domínio do Tempo

Uma vez que os blocos fundamentais de um sinal SSB são as bandas laterais, busquemos, primeiro, uma expressão para cada banda lateral no domínio do tempo.

A Fig. 4.15a mostra o espectro de mensagem $M(f)$. A Fig. 4.15b mostra sua metade direita, $M_+(f)$, e a Fig. 4.15c, a metade esquerda, $M_-(f)$. Das Fig. 4.15b e c, observamos que

$$M_+(f) = M(f) \cdot u(f) = M(f)\frac{1}{2}[1 + \mathrm{sgn}(f)] = \frac{1}{2}[M(f) + jM_h(f)] \tag{4.19a}$$

$$M_-(f) = M(f)u(-f) = M(f)\frac{1}{2}[1 - \mathrm{sgn}(f)] = \frac{1}{2}[M(f) - jM_h(f)] \tag{4.19b}$$

Agora, podemos expressar o sinal SSB em termos de $m(t)$ e $m_h(t)$. A partir da Fig. 4.15d, fica claro que o espectro USB, $\Phi_{\mathrm{USB}}(f)$, pode ser expresso como

$$\begin{aligned}\Phi_{\mathrm{USB}}(f) &= M_+(f - f_c) + M_-(f + f_c) \\ &= \frac{1}{2}[M(f - f_c) + M(f + f_c)] - \frac{1}{2j}[M(f - f_c) - M(f + f_c)]\end{aligned}$$

Da propriedade de translação na frequência, a transformada inversa dessa equação é dada por

$$\varphi_{\mathrm{USB}}(t) = m(t)\cos\,\omega_c t - m_h(t)\,\mathrm{sen}\,\omega_c t \tag{4.20a}$$

Da mesma forma, podemos mostrar que

$$\varphi_{\mathrm{LSB}}(t) = m(t)\cos\,\omega_c t + m_h(t)\,\mathrm{sen}\,\omega_c t \tag{4.20b}$$

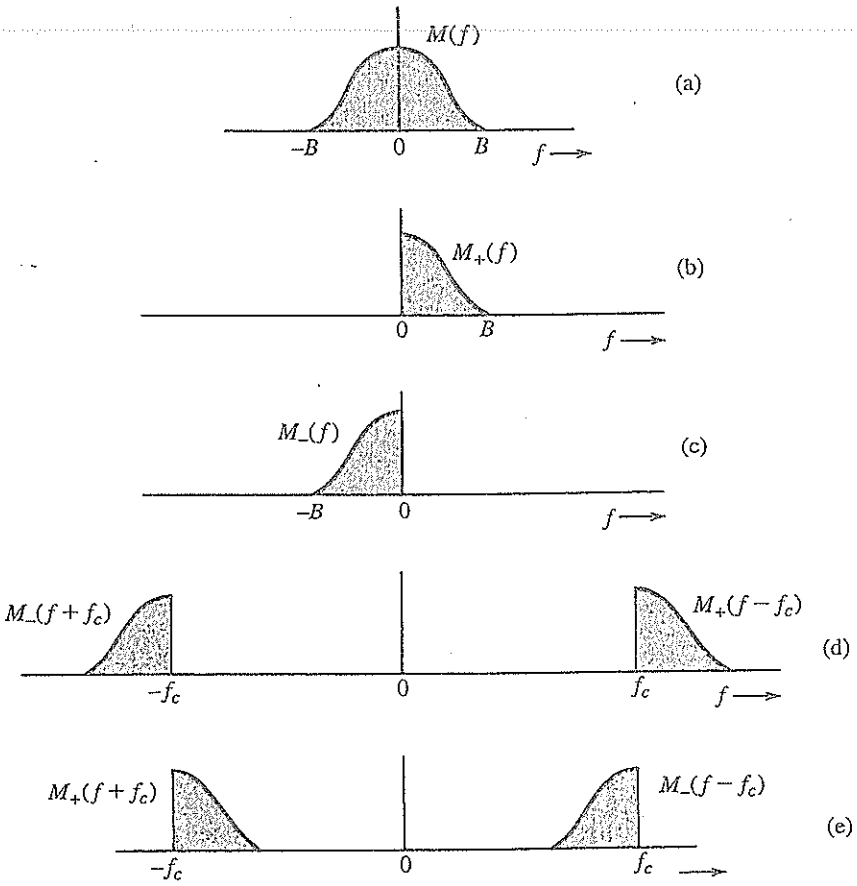


Figura 4.15 Expressão de espectros SSB em termos de $M_+(f)$ e $M_-(f)$.

Logo, um sinal SSB genérico $\varphi_{SSB}(t)$ pode ser expresso como

$$\varphi_{SSB}(t) = m(t)\cos\omega_c t \mp m_h(t)\,\text{sen}\,\omega_c t \tag{4.20c}$$

em que o sinal menos se aplica a USB e o sinal mais, a LSB.

Dada a expressão de sinais SSB-SC no domínio do tempo, podemos agora confirmar analiticamente (e não graficamente) que sinais SSB-SC podem ser demodulados coerentemente:

$$\begin{aligned}\varphi_{SSB}(t)\cos\omega_c t &= [m(t)\cos\omega_c t \mp m_h(t)\,\text{sen}\,\omega_c t]\,2\cos\omega_c t\\ &= m(t)[1+\cos 2\omega_c t] \mp m_h(t)\,\text{sen}\,2\omega_c t\\ &= m(t) + \underbrace{[m(t)\cos 2\omega_c t \mp m_h(t)\,\text{sen}\,2\omega_c t]}_{\text{Sinal SSB-SC com portadora } 2\omega_c}\end{aligned}$$

Logo, o produto $\varphi_{SSB}(t) \times 2\cos\omega_c t$ produz o sinal em banda base e outro sinal SSB com portadora $2\omega_c$. O espectro na Fig. 4.13e mostra precisamente esse resultado. Um filtro passa-baixas suprimirá os termos SSB indesejados, fornecendo o desejado sinal em banda base $m(t)$. Assim, o demodulador é idêntico ao demodulador síncrono usado para sinais DSB-SC. Portanto, qualquer um dos demoduladores DSB-SC discutidos na Seção 4.2 pode ser usado para demodular um sinal SSB-SC.

Exemplo 4.6 Modulação por Tom: SSB

Determinemos $\varphi_{SSB}(t)$ para o caso simples de uma modulação por tom, ou seja, em que o sinal modulante é uma senoide $m(t) = \cos\omega_m t$. Determinemos, também, a demodulação coerente desse sinal SSB.

Recordemos que a transformada de Hilbert atrasa a fase de cada componente espectral de $\pi/2$. No caso considerado, há apenas uma componente espectral, de frequência ω_m. O atraso da fase de $m(t)$ de $\pi/2$ resulta em

$$m_h(t) = \cos\left(\omega_m t - \frac{\pi}{2}\right) = \operatorname{sen} \omega_m t$$

Logo, da Eq. (4.20c),

$$\begin{aligned}\varphi_{\text{SSB}}(t) &= \cos\ \omega_m t \cos\ \omega_c t \mp \operatorname{sen} \omega_m t \operatorname{sen} \omega_c t \\ &= \cos\ (\omega_c \pm \omega_m)t\end{aligned}$$

Portanto,

$$\varphi_{\text{USB}}(t) = \cos\ (\omega_c + \omega_m)t \qquad \text{e} \qquad \varphi_{\text{LSB}}(t) = \cos\ (\omega_c - \omega_m)t$$

Para comprovar esses resultados, consideremos o espectro de $m(t)$ (Fig. 4.16a) e seus espectros DSB-SC (Fig. 4.16b), USB (Fig. 4.16c) e LSB (Fig. 4.16d). Fica evidente que os espectros nas Fig. 4.16c e d correspondem, de fato, aos espectros de $\varphi_{\text{USB}}(t)$ e $\varphi_{\text{LSB}}(t)$ deduzidos anteriormente.

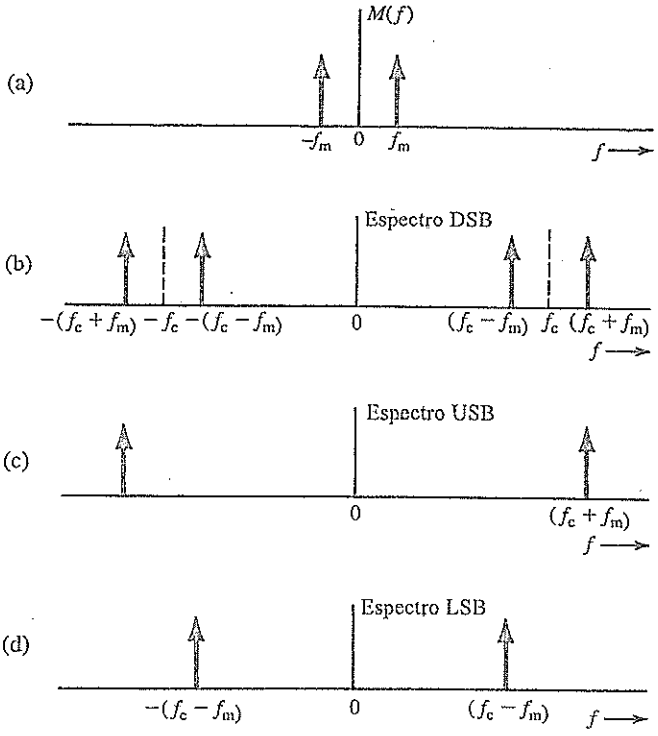


Figura 4.16 Espectros SSB para modulação por tom.

Por fim, a demodulação coerente da modulação SSB por tom é obtida de

$$\begin{aligned}\varphi_{\text{SSB}}(t) 2\cos\ \omega_c t &= 2\cos\ (\omega_c \pm \omega_m)t \cos\ \omega_c t \\ &= \cos\ \omega_m t + \cos\ (\omega_c + \omega_m)t\end{aligned}$$

que pode ser aplicado a um filtro passa-baixas para recuperar o tom de mensagem $\cos\ \omega_m t$.

Sistemas de Modulação SSB

Três métodos são comumente empregados para gerar sinais SSB: deslocamento de fase, filtragem seletiva e o método de Weaver.[1] Nenhum desses métodos de modulação é preciso e todos, em geral, requerem que o espectro do sinal em banda base tenha baixa potência nas proximidades da origem.

O **método de deslocamento de fase** usa diretamente a Eq. (4.20) como base. A Fig. 4.17 mostra sua implementação. A caixa rotulada "$-\pi/2$" é um deslocador de fase, que atrasa a fase de cada componente espectral positiva de $\pi/2$. Portanto, é um transformador de Hilbert. Vale notar que um deslocador de fase de Hilbert ideal é irrealizável, pois o deslocador de fase de Hilbert requer uma mudança brusca de fase de π na frequência zero. Quando a mensagem $m(t)$ tem um nulo dc e pouco conteúdo de baixa frequência, a aproximação prática desse deslocador de fase ideal praticamente não afeta a precisão da modulação SSB.

No **método de filtragem seletiva**, o mais usado na geração de sinais SSB, um sinal DSB-SC é passado por um filtro de corte abrupto para eliminar a banda lateral indesejada. Para a obtenção de USB, o filtro deve deixar passar sem atenuação todas as componentes acima da frequência f_c e deve suprimir completamente todas as componentes abaixo de f_c. Uma operação como essa exige um filtro ideal, que é irrealizável. Contudo, uma boa aproximação é possível caso exista alguma separação entre a banda passante e a banda de rejeição. Felizmente, o sinal de voz satisfaz esta condição, pois seu espectro exibe pouco conteúdo de potência na origem (Fig. 4.18a). Adicionalmente, testes mostraram que, para sinais de voz, componentes de frequências abaixo de 300 Hz não são importantes. Em outras palavras, podemos suprimir todas as componentes abaixo de 300 Hz (e acima de 3500 Hz) sem afetar a inteligibilidade de modo considerável. Assim, a filtragem da banda lateral indesejada se torna relativamente fácil no caso de sinais de voz, pois nos resta uma região de transição de 600 Hz em torno da frequência de corte f_c. Para minimizar interferência em canais adjacentes, a banda lateral indesejada deve ser atenuada em pelo menos 40 dB.

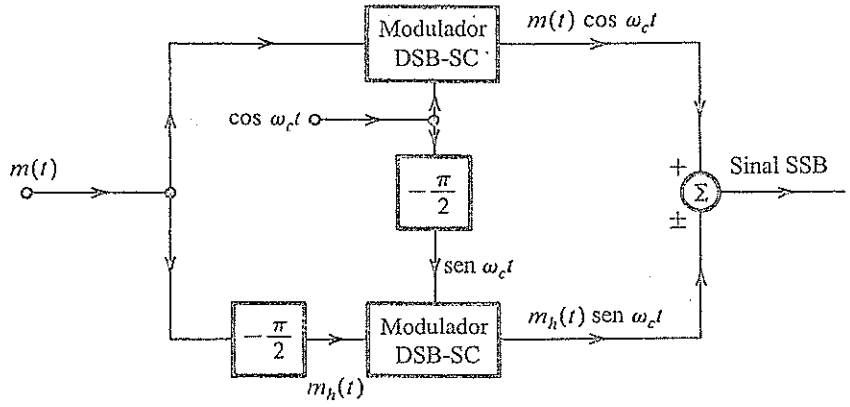


Figura 4.17 Geração de SSB com o método de deslocamento de fase.

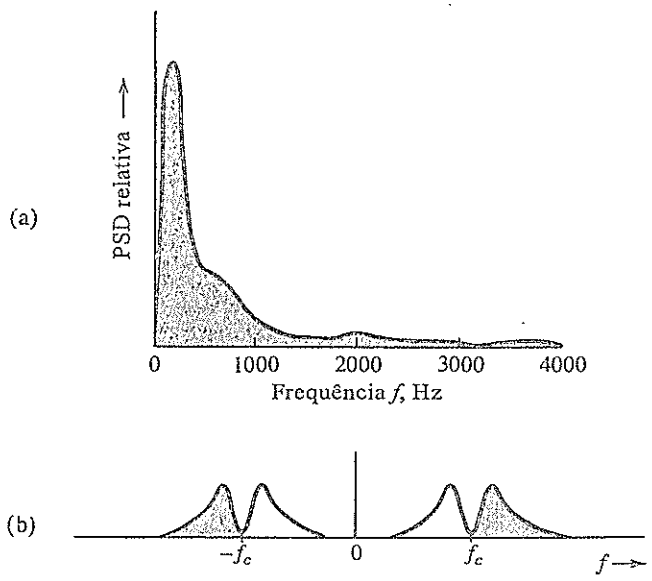


Figura 4.18 (a) Espectro de potência relativa de um sinal de voz e (b) correspondente espectro USB.

Para frequência portadora f_c muito alta, a razão entre a separação de banda (600 Hz) e a frequência portadora pode ser demasiadamente pequena e, em consequência, uma transição de 40 dB em amplitude pode ser difícil de alcançar. Nesse caso, um terceiro método, conhecido como **método de Weaver**,[1] faz uso de dois estágios de modulação em amplitude SSB. Primeiro, a modulação é efetuada com uma frequência portadora baixa (f_{c_1}). O resultante sinal SSB efetivamente alarga a separação para $2f_{c_1}$ (ver espectro hachurado na Fig. 4.18b). A seguir, esse sinal é tratado como um novo sinal em banda base, o que permite alcançar a modulação SSB em uma frequência portadora mais elevada.

Detecção de Sinais SSB com Portadora (SSB+C)

Agora, consideremos sinais SSB com uma portadora adicional (SSB+C). Tais sinais podem ser expressos como

$$\varphi_{\text{SSB+C}} = A\cos\omega_c t + [m(t)\cos\omega_c t + m_h(t)\,\text{sen}\,\omega_c t]$$

e $m(t)$ pode ser recuperado por detecção síncrona [multiplicando $\varphi_{\text{SSB+C}}$ por $\cos\omega_c t$] se a componente de portadora $A\cos\omega_c t$ puder ser extraída (via filtragem de banda estreita) de $\varphi_{\text{SSB+C}}$. Alternativamente, se a amplitude A da portadora for suficientemente grande, $m(t)$ também pode ser (aproximadamente) recuperado de $\varphi_{\text{SSB+C}}$ por detecção de envelope ou com retificação. Para vermos isso, reescrevemos $\varphi_{\text{SSB+C}}$ como

$$\begin{aligned}\varphi_{\text{SSB+C}} &= [A + m(t)]\cos\omega_c t + m_h(t)\,\text{sen}\,\omega_c t \\ &= E(t)\cos(\omega_c t + \theta)\end{aligned} \tag{4.21}$$

em que $E(t)$, o envelope de $\varphi_{\text{SSB+C}}$, é dado por [ver Eq. (3.41a)]

$$\begin{aligned}E(t) &= \{[A + m(t)]^2 + m_h^2(t)\}^{1/2} \\ &= A\left[1 + \frac{2m(t)}{A} + \frac{m^2(t)}{A^2} + \frac{m_h^2(t)}{A^2}\right]^{1/2}\end{aligned}$$

Se $A >> |m(t)|$, em geral* $A >> |m_h(t)|$, e os termos $m^2(t)/A^2$ e $m_h^2(t)/A^2$ podem ser ignorados. Assim,

$$E(t) \simeq A\left[1 + \frac{2m(t)}{A}\right]^{1/2}$$

Usando expansão em série de Taylor e descartando termos de ordens superiores [pois $m(t)/A << 1$], obtemos

$$\begin{aligned}E(t) &\simeq A\left[1 + \frac{m(t)}{A}\right] \\ &= A + m(t)\end{aligned}$$

Fica evidente que, para uma portadora intensa, o sinal SSB+C pode ser demodulado por um detector de envelope.

Em AM, a detecção de envelope exige a condição $A \geq |m(t)|$, enquanto em SSB+C, a condição é $A >> |m(t)|$. Portanto, no caso SSB, a necessária amplitude da portadora é muito maior que em AM e, consequentemente, a eficiência de SSB+C é pateticamente baixa.

Modulação em Amplitude em Quadratura (QAM)

Dada a dificuldade em gerar sinais SSB-SC com precisão, a modulação em amplitude em quadratura (QAM) representa uma alternativa atraente a SSB-SC. QAM pode ser gerada com exatidão, sem a exigência de filtros passa-faixa com corte abrupto. O funcionamento de QAM é baseado na trans

*Isto pode não ser verdade para todo t, mas é verdade para a maioria dos valores de t.

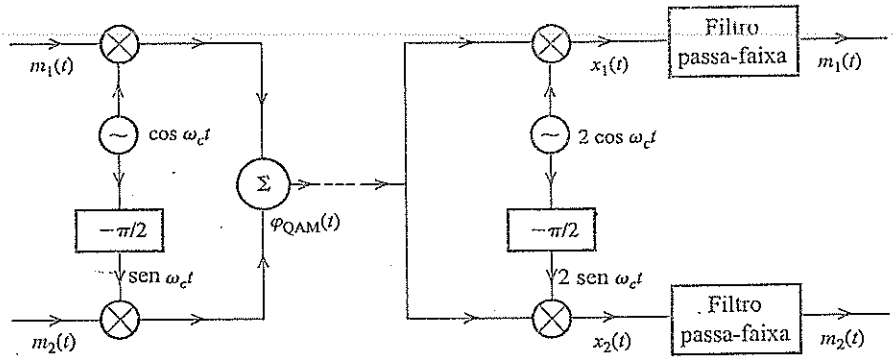


Figura 4.19 Multiplexação em amplitude em quadratura.

missão de dois sinais DSB com portadoras de mesma frequência, mas em quadratura de fase, como ilustrado na Fig. 4.19. Esse esquema é referido como **modulação em amplitude em quadratura (QAM) ou multiplexação em quadratura**.

Como mostrado na Fig. 4.19, as caixas rotuladas $-\pi/2$ são deslocadores de fase que atrasam a fase de uma entrada senoidal de $-\pi/2$ rad. Se os dois sinais de mensagens em banda base a serem transmitidos forem $m_1(t)$ e $m_2(t)$, o correspondente sinal QAM $\varphi_{QAM}(t)$, a soma de dois sinais modulados DSB, é

$$\varphi_{QAM}(t) = m_1(t) \cos \omega_c t + m_2(t) \operatorname{sen} \omega_c t$$

Os dois sinais modulados ocupam a mesma banda. No entanto, os dois sinais em banda base podem ser separados no receptor por detecção síncrona, caso as duas portadoras locais sejam usadas em quadratura de fase, como mostrado na Fig. 4.19. Para comprovarmos isso, consideremos a saída do multiplicador $x_1(t)$ no braço superior do receptor (Fig. 4.19):

$$\begin{aligned} x_1(t) &= 2\varphi_{QAM}(t) \cos \omega_c t = 2[m_1(t) \cos \omega_c t + m_2(t) \operatorname{sen} \omega_c t] \cos \omega_c t \\ &= m_1(t) + m_1(t) \cos 2\omega_c t + m_2(t) \operatorname{sen} 2\omega_c t \end{aligned} \tag{4.22a}$$

Os dois últimos termos são sinais passa-faixa centrados em $2\omega_c$. Na verdade, formam um sinal QAM com frequência portadora $2\omega_c$. Esses termos são suprimidos por um filtro passa-faixa, resultando na desejada saída demodulada $m_1(t)$. Do mesmo modo, podemos mostrar que a saída do braço inferior do receptor é $m_2(t)$.

$$\begin{aligned} x_2(t) &= 2\varphi_{QAM}(t) \operatorname{sen} \omega_c t = 2[m_1(t) \cos \omega_c t + m_2(t) \operatorname{sen} \omega_c t] \operatorname{sen} \omega_c t \\ &= m_2(t) - m_2(t) \cos 2\omega_c t + m_1(t) \operatorname{sen} 2\omega_c t \end{aligned} \tag{4.22b}$$

Assim, dois sinais em banda base, cada um com largura de banda de B Hz, podem ser transmitidos simultaneamente em uma largura de banda $2B$ com o uso de transmissão DSB e multiplexação em quadratura. O canal superior também é referido como canal **em fase** (I) e o canal inferior, como canal em **quadratura** (Q). Os dois sinais $m_1(t)$ e $m_2(t)$ podem ser demodulados separadamente.

No entanto, vale notar que a demodulação QAM deve ser totalmente síncrona. Um erro na fase ou na frequência da portadora no demodulador resultará em perda e interferência entre os dois canais. Para mostrarmos isso, consideremos que a portadora no demodulador seja $2 \cos (\omega_c t + \theta)$. Nesse caso,

$$\begin{aligned} x_1(t) &= 2[m_1(t) \cos \omega_c t + m_2(t) \operatorname{sen} \omega_c t] \cos (\omega_c t + \theta) \\ &= m_1(t) \cos \theta - m_2(t) \operatorname{sen} \theta + m_1(t) \cos (2\omega_c t + \theta) + m_2(t) \operatorname{sen} (2\omega_c t + \theta) \end{aligned}$$

Os dois filtros passa-baixas suprimem os dois sinais modulados por portadora de frequência angular $2\omega_c$, resultando na primeira saída do demodulador

$$m_1(t) \cos \theta - m_2(t) \operatorname{sen} \theta$$

Assim, além do desejado sinal $m_1(t)$, também temos o sinal recebido $m_2(t)$ no braço superior do receptor. Um fenômeno semelhante ocorre no braço inferior. Essa **interferência cocanal** é indesejável. Dificuldades semelhantes surgem quando há erro na frequência local (ver Exercício 4.4-1). Adicionalmente, atenuação desigual na USB e na LSB durante a transmissão leva a interferência cruzada ou cocanal.

Multiplexação em quadratura é usada em televisão analógica em cores para multiplexar os chamados sinais de crominância, que transportam informação sobre cores. Nesse caso, a sincronização é alcançada com a inserção periódica de uma breve explosão de sinal da portadora (referida como **explosão de cor** no sinal transmitido). A transmissão de televisão digital por satélite também faz uso de QAM.

Em termos de exigências de largura de banda, a SSB é semelhante à QAM, mas menos exigente em termos de frequência e fase da portadora e da necessidade de um meio de transmissão sem distorção. Contudo, a geração de SSB se torna difícil quando o sinal em banda base $m(t)$ tem significante conteúdo espectral próximo a dc.

4.5 MODULAÇÕES EM AMPLITUDE: BANDA LATERAL VESTIGIAL (VSB)

Como discutido anteriormente, a geração de sinais SSB exatos é difícil. Em geral, isso requer que o sinal de mensagem $m(t)$ tenha um nulo em torno de dc. Um deslocador de fase, necessário para o método de deslocamento de fase, é irrealizável, ou pode ser realizado apenas de forma aproximada. A geração de sinais DSB é muito mais simples, mas requer o dobro da largura de banda de sinal. A modulação em **banda lateral vestigial (VSB** – *vestigial sideband*), também referida como sistema de banda lateral assimétrica, representa um equilíbrio entre DSB e SSB. A VSB herda as vantagens de DSB e de SSB, mas evita suas desvantagens, a um pequeno custo. A geração de sinais VSB é relativamente simples e, ao mesmo tempo, a largura de banda dos mesmos é apenas um pouco maior (tipicamente, 25%) que a de sinais SSB.

Em VSB, em vez de rejeitar uma banda lateral completamente (como em SSB), um corte gradual de uma banda lateral, como mostrado na Fig. 4.20d, é aceitável. O sinal em banda base pode ser recuperado exatamente por um detector síncrono em conjunto com um filtro equalizador apropriado, $H_o(f)$, na saída do receptor (Fig. 4.21). Caso uma portadora mais intensa seja transmitida juntamente com o sinal VSB, o sinal em banda base pode ser recuperado por um detector de envelope (ou detector retificador).

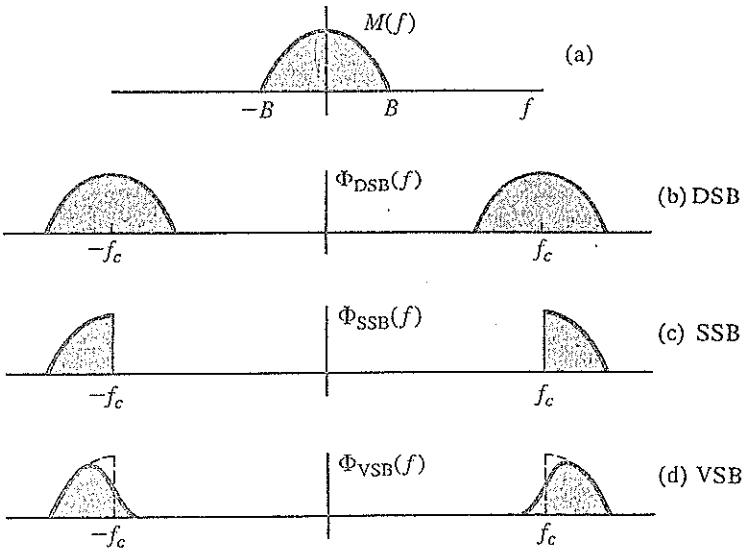


Figura 4.20 Espectros do sinal modulante e dos correspondentes sinais DSB, SSB e VSB.

Se o filtro formatador vestigial que produz VSB a partir de DSB for $H_i(f)$ (Fig. 4.21), o espectro do resultante sinal VSB é

$$\Phi_{VSB}(f) = [M(f+f_c) + M(f-f_c)]H_i(f) \tag{4.23}$$

Este filtro formatador VSB $H_i(f)$ permite a transmissão de uma banda lateral e suprime a outra, não completamente, mas de forma gradual. Isso facilita a realização do filtro, mas resulta em uma largura de banda ligeiramente maior do que a de SSB (em que a outra banda lateral é totalmente suprimida). A largura de banda do sinal VSB é tipicamente 25 a 33% maior que a de sinais SSB.

Desejamos que $m(t)$ seja recuperável de $\varphi_{VSB}(t)$ com demodulação síncrona no receptor. Para isso, o sinal VSB recebido, $\varphi_{VSB}(t)$, é multiplicado por 2 cos $\omega_c t$. O produto $e(t)$ é dado por

$$e(t) = 2\varphi_{VSB}(t)\cos\omega_c t \Longleftrightarrow [\Phi_{VSB}(f+f_c) + \Phi_{VSB}(f-f_c)]$$

O sinal $e(t)$ é, ainda, aplicado a um filtro passa-baixas equalizador, com função de transferência $H_o(f)$. A saída do filtro equalizador deve ser $m(t)$. O espectro do sinal de saída é dado por

$$M(f) = [\Phi_{VSB}(f+f_c) + \Phi_{VSB}(f-f_c)]H_o(f)$$

A substituição da Eq. (4.23) na equação anterior e a eliminação do espectro em $\pm 4f_c$ [suprimido pelo filtro passa-baixas $H_o(f)$] resultam em

$$M(f) = M(f)[H_i(f+f_c) + H_i(f-f_c)]H_o(f) \tag{4.24}$$

Logo,

$$H_o(f) = \frac{1}{H_i(f+f_c) + H_i(f-f_c)} \qquad |f| \le B \tag{4.25}$$

Como $H_i(f)$ é um filtro passa-faixa, os termos $H_i(f \pm f_c)$ contêm componentes passa-baixas.

Filtro VSB Complementar e Detecção de Envelope de Sinais VSB+C

Como um caso especial de um filtro no modulador VSB, podemos escolher $H_i(f)$ tal que

$$H_i(f+f_c) + H_i(f-f_c) = 1 \qquad |f| \le B \tag{4.26}$$

O filtro de saída consiste apenas em um filtro passa-baixas com função de transferência

$$H_o(f) = 1 \qquad |f| \le B$$

O sinal resultante VSB mais portadora (VSB+C) pode ser demodulado por detecção de envelope. Podemos comprovar esse método de demodulação usando exatamente o mesmo argumento empregado no caso de sinais SSB+C. Em particular, devido à Eq. (4.26), podemos definir um novo filtro passa-baixas

$$F(f) = j[1 - 2H_i(f-f_c)] = -j[1 - 2H_i(f+f_c)] \qquad |f| \le B$$

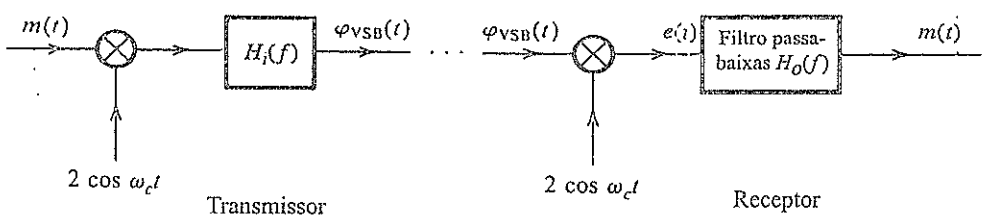


Figura 4.21 Modulador e demodulador VSB.

Definindo um novo sinal passa-baixas (complexo) como

$$m_v(t) \Longleftrightarrow M_v(f) = F(f)M(f)$$

podemos reescrever o sinal VSB como

$$\Phi_{\text{VSB}}(f) = \frac{M(f-f_c)+M(f+f_c)}{2} + \frac{M_v(f-f_c)-M_v(f+f_c)}{2j} \tag{4.27a}$$

$$\Longleftrightarrow$$

$$\varphi_{\text{VSB}}(t) = m(t)\cos 2\pi f_c t + m_v(t)\,\text{sen}\,2\pi f_c t \tag{4.27b}$$

Fica claro que os sinais modulados SSB e VSB têm a mesma forma, com $m_h(t)$ de SSB substituído por um sinal passa-baixas $m_v(t)$ em VSB. Aplicando a mesma análise feita para a detecção de envelope de SSB+C, a adição de uma portadora intensa a $\varphi_{\text{VSB}}(t)$ permitirá a detecção de envelope de VSB+C.

Mostramos que, para detecção de envelope, SSB+C requer uma portadora muito mais intensa que DSB+C (AM). Como VSB+C é um caso intermediário, a portadora adicional exigida em VSB é mais intensa que em AM, mas menos intensa que em SSB+C.

Exemplo 4.7 A frequência portadora de um certo sinal VSB é $f_c = 20$ kHz, e a largura de banda do sinal em banda base é 6 kHz. O filtro formatador VSB na saída, $H_i(f)$, que corta a gradualmente a banda lateral inferior ao longo de 2 kHz, é mostrado na Fig. 4.22a. Determinemos o filtro de saída, $H_o(f)$, necessário à recepção sem distorção.

A Fig. 4.22b mostra os segmentos passa-baixas de $H_i(f+f_c)+H_i(f-f_c)$. Desse espectro, nos interessa apenas a porção na banda base (a restante porção indesejada é suprimida pelo filtro de saída). Na banda de 0 a 2 kHz, a amplitude desse espectro é 0,5; de 2 a 6 kHz, a amplitude é 1, como mostrado na Fig. 4.22b. A Fig. 4.22c mostra o desejado filtro de saída $H_o(f)$, que é o recíproco do espectro na Fig. 4.22b [Eq. (4.25)].

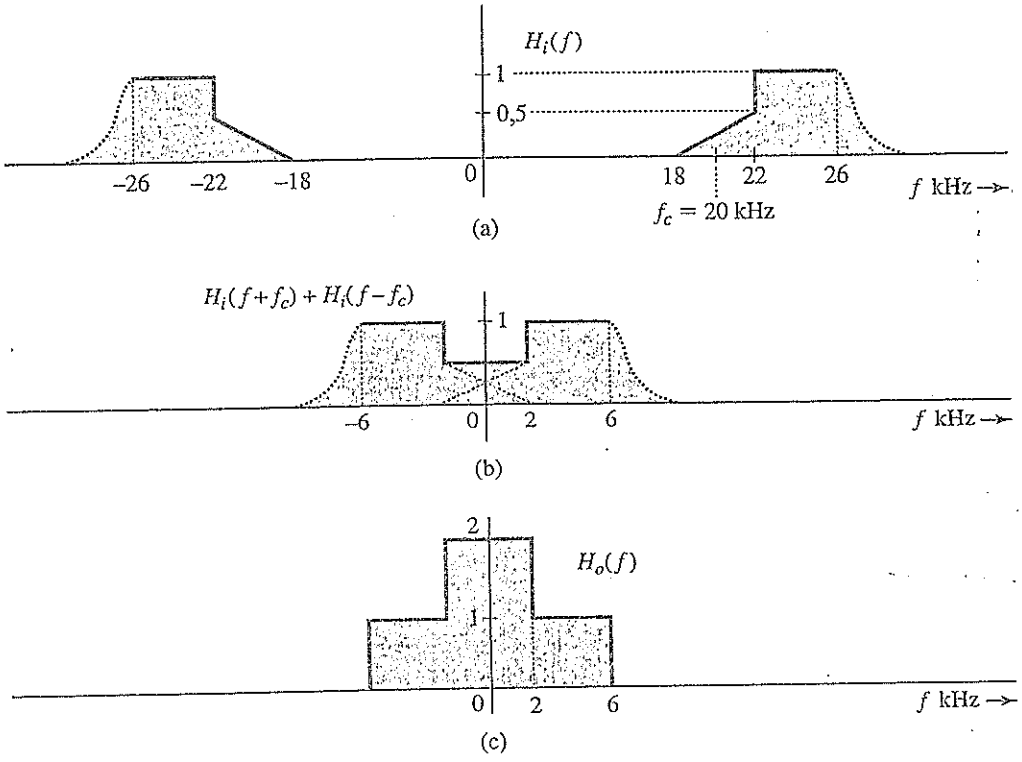


Figura 4.22 Modulador VSB e filtros do receptor.

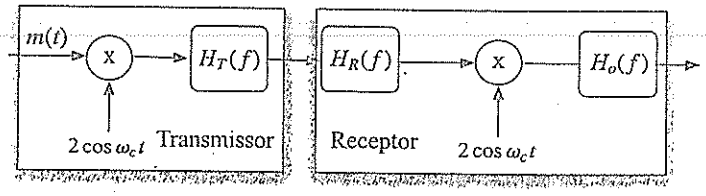


Figura 4.23 Filtro do transmissor, $H_T(f)$, filtro do *front-end* do receptor, $H_R(f)$, e filtro passa-baixas de saída do receptor, $H_o(f)$, para sistemas de televisão VSB.

Uso de VSB na Difusão de Televisão

VSB é um interessante equilíbrio entre SSB e DSB, o que a torna muito atraente para sistemas de difusão de televisão. O sinal de vídeo de televisão em banda base ocupa uma enorme largura de banda de 4,5 MHz, de modo que um sinal DSB requer uma largura de banda de 9 MHz. Para economizar largura de banda, poderia ser desejável o uso de SSB. Infelizmente, isso criaria diversos problemas. Primeiro, o sinal de vídeo em banda base tem potência razoável na região de baixas frequências, o que dificultaria a supressão completa de uma banda lateral. Segundo, para um receptor de difusão, um detector de envelope é mais conveniente do que um detector síncrono, pois reduz o custo do receptor. Vimos, anteriormente, que SSB+C tem baixa eficiência de potência. Adicionalmente, o uso de SSB aumentaria o custo do receptor.

A formatação espectral de sinais VSB de televisão é ilustrada na Fig. 4.23. O espectro vestigial é controlado por dois filtros: o filtro de RF do transmissor, $H_T(f)$, e o filtro de RF do receptor, $H_R(f)$. No conjunto, temos

$$H_i(f) = H_T(f)H_R(f)$$

Portanto, o projeto do filtro de saída do receptor, $H_o(f)$, segue a Eq. (4.25).

O espectro DSB de um sinal de televisão é mostrado na Fig. 4.24a. O filtro de formatação vestigial, $H_i(f)$, corta gradualmente o espectro da banda lateral inferior, de 0,75 MHz a 1,25 MHz abaixo da frequência portadora f_c, como ilustrado na Fig. 4.24b. O filtro de saída do receptor, $H_o(f)$, é projetado segundo a Eq. (4.25). A resultante largura espectral VSB é de 6 MHz. Esse valor deve ser comparado com os 9 MHz da largura de banda DSB e os 4,5 MHz da largura de banda SSB.

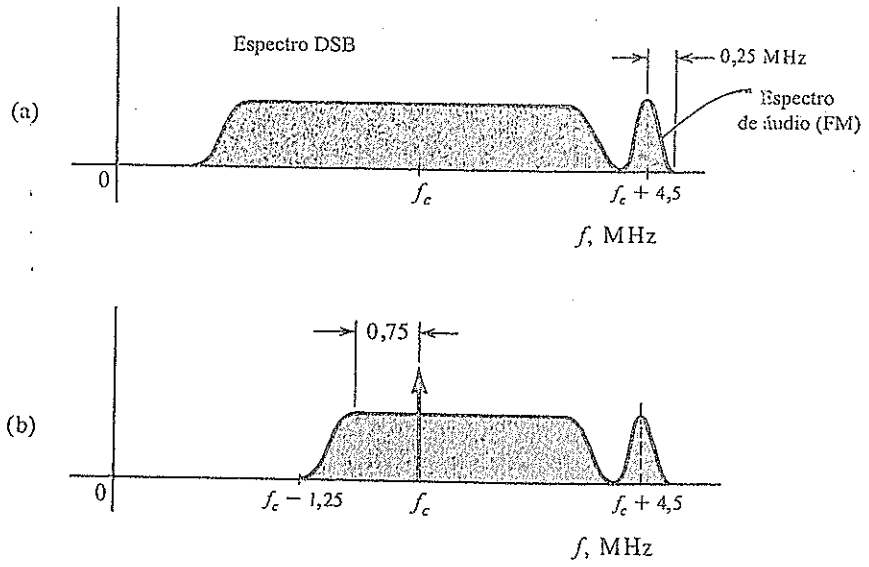


Figura 4.24 Espectro de sinal de televisão: (a) sinal DSB, (b) sinal transmitido.

4.6 SINCRONIZAÇÃO DA PORTADORA LOCAL

Em um sistema modulado em amplitude com portadora suprimida (DSB-SC, SSB-SC e VSB-SC), o receptor coerente deve gerar uma portadora local em sincronia (de frequência e fase) com a portadora recebida. Como discutido anteriormente, qualquer discrepância de frequência ou fase da portadora local causa distorção na saída do detector.

Consideremos um caso SSB-SC em que o sinal recebido é

$$m(t)\cos\,[(\omega_c+\Delta\omega)t+\delta]-m_h(t)\,\mathrm{sen}\,[(\omega_c+\Delta\omega)t+\delta]$$

devido ao atraso de propagação e ao deslocamento de frequência Doppler. A portadora local continua sendo $2\cos\omega_c t$. O produto do sinal recebido pela portadora local é $e(t)$, dado por

$$\begin{aligned} e(t) &= 2\cos\,\omega_c t\,[m(t)\cos(\omega_c t+\Delta\omega t+\delta)-m_h(t)\,\mathrm{sen}(\omega_c t+\Delta\omega t+\delta)] \\ &= m(t)\cos(\Delta\omega t+\delta)-m_h(t)\,\mathrm{sen}(\Delta\omega t+\delta) \\ &+ \underbrace{m(t)\cos[(2\omega_c+\Delta\omega)t+\delta]-m_h(t)\,\mathrm{sen}[(2\omega_c+\Delta\omega)t+\delta]}_{\text{sinal SSB-SC passa-faixa em torno de } 2\omega_c+\Delta\omega} \end{aligned} \tag{4.28}$$

A componente passa-faixa é suprimida pelo filtro passa-baixas do receptor, deixando a saída $e_o(t)$ como

$$e_o(t)=m(t)\cos(\Delta\omega t+\delta)-m_h(t)\,\mathrm{sen}(\Delta\omega t+\delta) \tag{4.29}$$

Se $\Delta\omega$ e δ forem ambos nulos (nenhum erro de frequência nem de fase), então

$$e_o(t)=m(t)$$

como esperado.

Na prática, se a onda de rádio viajar uma distância de d metros à velocidade da luz c, o atraso de fase será

$$\delta=-(\omega_c+\Delta\omega)d/c$$

que pode ser qualquer valor no intervalo $[-\pi,+\pi]$. Dois osciladores com, inicialmente, a mesma frequência também podem se afastar. Se o receptor ou o transmissor estiver viajando a uma velocidade v_e, o máximo deslocamento de frequência Doppler será

$$\Delta f_{\max}=\frac{v_e}{c}f_c$$

A velocidade v_e depende do veículo em questão (nave espacial, avião, carro etc.). Por exemplo, para uma velocidade v_e de 108 km/h e uma frequência portadora de 100 MHz, o máximo deslocamento de frequência Doppler será de 10 Hz. Um deslocamento fixo $\Delta\omega$ de cada componente de frequência destrói a relação harmônica entre as mesmas. Para $\Delta f=10$ Hz, as componentes de frequência em 1000 e 2000 Hz serão deslocadas para 1010 e 2010 Hz, respectivamente. Isso afeta a relação harmônica entre elas e, em consequência, a qualidade de sinais que não são de áudio.

É interessante observar que sinais de áudio são altamente redundantes e, a menos que Δf seja muito grande, uma alteração como essa não destrói a inteligibilidade da saída. Para sinais de áudio, $\Delta f<30$ Hz não afeta a qualidade dos mesmos de modo significativo. $\Delta f>30$ Hz resultará em uma qualidade de som semelhante à voz do Pato Donald. Contudo, a inteligibilidade não é completamente perdida.

Em geral, há duas maneiras de recuperar uma portadora que chega ao receptor. Uma consiste em o transmissor transmitir um sinal piloto (senoidal) que pode ser a portadora exata ou diretamente a ela relacionado (por exemplo, um piloto na metade da frequência portadora). O piloto é separado no receptor por um filtro de largura de banda muito estreita e sintonizado na frequência do piloto. O sinal piloto é amplificado e usado para sincronizar o oscilador local. Em outro método, no qual um pi-

loto não é transmitido, o receptor emprega um dispositivo não linear para processar o sinal recebido e gerar uma componente separada da portadora, que pode ser extraída por meio de filtros passa-faixa estreitos. Fica claro que filtros passa-faixa eficientes e de pequena largura de banda são muito importante para os dois métodos. Além disso, o filtro passa-faixa também deve ter a capacidade de ajustar, de modo adaptativo, sua frequência central para combater deslocamentos de frequência ou deslocamentos Doppler significativos. Afora algumas configurações típicas de filtros passa-faixa, a *malha de captura de fase* (PLL – *phase locked loop*), que tem papel crucial na aquisição de portadora de vários esquemas de modulação, pode ser vista como um desses filtros passa-faixa adaptativos. Os princípios de PLL serão discutidos mais adiante neste capítulo.

4.7 MULTIPLEXAÇÃO POR DIVISÃO EM FREQUÊNCIA (FDM)

A multiplexação de sinais permite a transmissão de vários sinais em um mesmo canal. No Capítulo 6, discutiremos da multiplexação por divisão no tempo (TDM – *time division multiplexing*), na qual vários canais compartilham um mesmo canal no tempo. Em FDM (*frequency division multiplexing*), diversos sinais compartilham a banda de um canal. Cada sinal é modulado por uma frequência portadora diferente. Essas portadoras, referidas como **subportadoras**, são adequadamente separadas para evitar sobreposição (interferência) dos espectros dos vários sinais modulados. Cada sinal pode usar um tipo distinto de modulação (por exemplo, DSB-SC, AM, SSB-SC, VSB-SC ou, até mesmo, modulação em frequência ou modulação em fase). O espectro do sinal modulado pode ser separado por uma estreita banda de guarda para evitar interferência e facilitar a separação de sinais no receptor.

Quando todos os espectros dos sinais modulados são somados, obtemos um sinal composto que pode ser considerado como um sinal em banda base, que, para transmissão, pode modular uma portadora de radiofrequência (RF).

No receptor, o sinal de entrada é, primeiro, demodulado pela portadora de RF, para recuperar o sinal em banda base composto; a seguir, esse é aplicado a filtros passa-faixa para separar todos os sinais modulados. Cada sinal modulado é, então, demodulado individualmente por uma subportadora apropriada para fornecer os desejados sinais em banda base originais.

Um exemplo simples de FDM é o sistema de telefonia analógica de longa distância. Há dois tipos de sistemas de portadoras de telefonia de longa distância: o sistema legado de hierarquia analógica de portadora L e o sistema norte-americano de hierarquia digital de portadora T (ou o sistema europeu de portadora E).[3] Ambos foram padronizados pela União Internacional de Telecomunicações (*International Telecommunications Union*), conhecida (antes de 1992) como CCITT (*Comité Consultatif International Téléphonique et Télégraphique*). Discutiremos, primeiro, a hierarquia de telefonia analógica que utiliza FDM e modulação SSB; a discussão da hierarquia digital será feita posteriormente (no Capítulo 6).

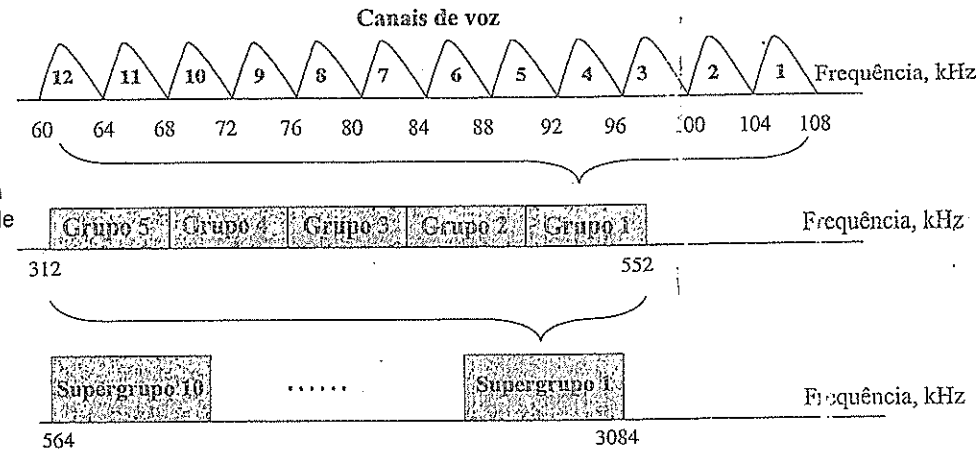


Figura 4.25 Sistema de multiplexação por divisão em frequência para telefonia analógica de longa distância na hierarquia de portadora L.

Na hierarquia analógica de portadora L,[4] cada canal de voz é modulado com uso de SSB+C. Doze canais de voz formam um **grupo** básico de canais que ocupa a banda de 60 a 108 kHz. Como mostrado na Fig. 4.25, cada canal de usuário utiliza a LSB, e a multiplexação por divisão em frequência (FDM) é alcançada mantendo uma separação de 4 kHz entre portadoras de canal.

Acima na hierarquia,[5] cinco grupos formam um **supergrupo**, via FDM. A multiplexação de 10 supergrupos gera um **grupo mestre**; a multiplexação de seis supergrupos forma um **grupo jumbo**, que consiste em 3600 canais de voz em uma banda de frequência de 16,984 MHz no sistema L4. A cada nível da hierarquia a partir do supergrupo, espaçamentos adicionais de frequência são previstos para redução de interferência e para a inserção de frequências-piloto. O sinal multiplexado pode ser alimentado à entrada de banda base de um canal de rádio de micro-ondas ou diretamente a um sistema de transmissão coaxial.

4.8 MALHA DE CAPTURA DE FASE (*PHASE LOCKED LOOP* — PLL) E ALGUMAS APLICAÇÕES

Malha de Captura de Fase (PLL)*

A **malha de captura de fase (PLL)** é um dispositivo muito importante, geralmente usado para rastrear a fase e a frequência de uma componente de portadora de um sinal recebido. Portanto, é um dispositivo útil para a demodulação síncrona de sinais AM com portadora suprimida ou com uma portadora de baixa intensidade (piloto). A PLL também pode ser usada para a demodulação de sinais modulados em ângulo, especialmente sob condições de baixa relação sinal-ruído (SNR — *signal-to-noise ratio*). A PLL também encontra importantes aplicações em vários sistemas de recuperação de relógio, incluindo recuperação de tempo em receptores digitais. Por essas razões, a PLL tem um papel essencial em praticamente todos os modernos sistemas de comunicação digitais e analógicos.

Uma PLL tem três componentes básicos:

1. Um oscilador controlado por tensão (VCO — *voltage-controlled oscillator*)
2. Um multiplicador, que funciona como um detector de fase (PD — *phase detector*) ou comparador de fase.
3. Um filtro de malha, $H(s)$

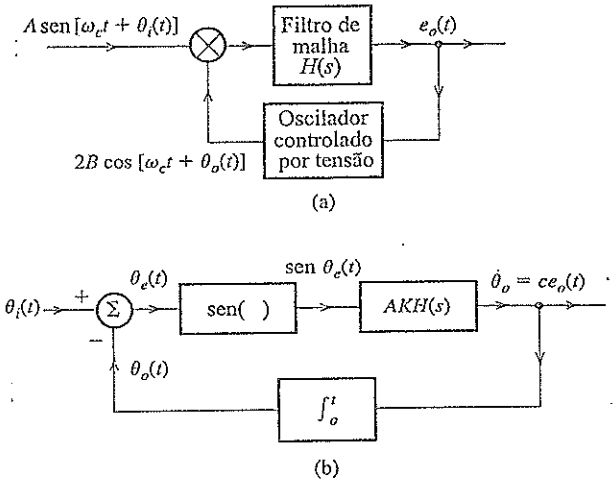


Figura 4.26 Malha de captura de fase e seu circuito equivalente.

*O termo "malha de sincronismo de fase" é igualmente empregado para PLL. (N.T.)

Funcionamento Básico da PLL

O funcionamento da PLL é similar ao de um sistema de realimentação (Fig. 4.26a). Em um típico sistema de realimentação, o sinal de realimentação tende a seguir o sinal de entrada. Caso o sinal de realimentação não seja igual ao de entrada, a diferença entre os dois (conhecida como erro) alterará o sinal de realimentação até que este se aproxime do sinal de entrada. O princípio de funcionamento de uma PLL é semelhante, exceto que a grandeza realimentada e comparada não é a amplitude, mas a fase. O VCO ajusta sua própria frequência até que suas frequência e fase acompanhem a frequência e a fase do sinal de entrada. Nesse ponto, os dois sinais estão em sincronismo (exceto se por uma possível diferença de fase constante).

O **oscilador controlado por tensão (VCO)** é um oscilador cuja frequência pode ser controlada linearmente por uma tensão de entrada. Se a tensão de entrada em um VCO for $e_o(t)$, sua saída será uma senoide com frequência instantânea dada por

$$\omega(t) = \omega_c + ce_o(t) \tag{4.30}$$

em que c é uma constante do VCO e ω_c, a **frequência livre** do VCO [quando $e_o(t) = 0$]. A saída do multiplicador sofre uma filtragem passa-baixas pelo filtro de malha e, então, é aplicada à entrada do VCO. Essa tensão altera a frequência do oscilador e mantém a malha **travada** (*locked*), forçando a saída do VCO a rastrear a fase (e, consequentemente, a frequência) da senoide de entrada.

Se a saída do VCO for $B \cos [\omega_c t + \theta_o(t)]$, sua frequência instantânea será $\omega_c + \dot{\theta}_o(t)$. Portanto,

$$\dot{\theta}_o(t) = ce_o(t) \tag{4.31}$$

Notemos que c e B são parâmetros constantes da PLL.

Admitamos que o sinal que chega (sinal de entrada da PLL) seja A sen $[\omega_c t + \theta_i(t)]$. Caso o sinal que chega seja A sen $[\omega_o t + \psi(t)]$, ainda pode ser expresso como A sen $[\omega_c t + \theta_i(t)]$, em que $\theta_i(t) = (\omega_o - \omega_c)t + \psi(t)$. Logo, a análise a seguir é geral e não é restrita à condição de que a frequência do sinal que chega seja igual à do sinal livre do VCO.

A saída do multiplicador é

$$AB \operatorname{sen}(\omega_c t + \theta_i) \cos(\omega_c t + \theta_o) = \frac{AB}{2}[\operatorname{sen}(\theta_i - \theta_o) + \operatorname{sen}(2\omega_c t + \theta_i + \theta_o)]$$

O termo da soma de frequências é suprimido pelo filtro de malha. Assim, a saída efetiva do filtro de malha é $\frac{1}{2}AB$ sen $[\theta_i(t) - \theta_o(t)]$. Seja $h(t)$ a resposta do filtro de malha ao impulso unitário; portanto,

$$\begin{aligned} e_o(t) &= h(t) * \frac{1}{2}AB \operatorname{sen}[\theta_i(t) - \theta_o(t)] \\ &= \frac{1}{2}AB \int_0^t h(t-x) \operatorname{sen}[\theta_i(x) - \theta_o(x)]\, dx. \end{aligned} \tag{4.32}$$

Substituindo a Eq. (4.32) na Eq. (4.31) e tomando $K = \frac{1}{2}cB$, obtemos

$$\dot{\theta}_o(t) = AK \int_0^t h(t-x) \operatorname{sen} \theta_e(x)\, dx \tag{4.33}$$

em que $\theta_e(t)$ é o erro de fase, definido como

$$\theta_e(t) = \theta_i(t) - \theta_o(t)$$

Essas equações [juntamente com a Eq. (4.31)] sugerem prontamente um modelo para a PLL, como ilustrado na Fig. 4.26b.

O projeto da PLL requer seleção criteriosa do filtro de malha $H(s)$ e do ganho da malha AK. Diferentes filtros de malha podem permitir que a PLL capture e rastreie sinais de entrada com distintos tipos de variação de frequência. O ganho da malha, por sua vez, pode afetar a faixa de variações de frequência que podem ser rastreadas.

Análise de PLL sob Pequeno Erro

Na análise de PLL sob pequeno erro, sen $\theta_e \simeq \theta_e$, de modo que o diagrama de blocos na Fig. 4.26b se reduz ao sistema linear (invariante no tempo) mostrado na Fig. 4.27a. Uma simples análise de realimentação fornece

$$\frac{\Theta_o(s)}{\Theta_i(s)} = \frac{AKH(s)/s}{1+[AKH(s)/s]} = \frac{AKH(s)}{s+AKH(s)} \tag{4.34}$$

Portanto, a PLL atua como um filtro com função de transferência $AKH(s)/[s + AKH(s)]$, como indicado na Fig. 4.27b. O erro $\Theta_e(s)$ é dado por

$$\begin{aligned}\Theta_e(s) = \Theta_i(s) - \Theta_o(s) &= \left[1 - \frac{\Theta_o(s)}{\Theta_i(s)}\right]\Theta_i(s) \\ &= \frac{s}{s+AKH(s)}\Theta_i(s)\end{aligned} \tag{4.35}$$

Uma das importantes aplicações de PLL é a aquisição de frequência e de fase para efeitos de sincronização. Seja A sen $(\omega_0 t + \varphi_0)$ o sinal que chega. Desejamos gerar um sinal local de frequência ω_0 e fase* φ_0. Supondo que a frequência quiescente do VCO seja ω_c, o sinal que chega pode ser expresso como A sen $[\omega_c t + \theta_i(t)]$, em que

$$\theta_i(t) = (\omega_0 - \omega_c)t + \varphi_0$$

e

$$\Theta_i(s) = \frac{\omega_0 - \omega_c}{s^2} + \frac{\varphi_0}{s}$$

Consideremos o caso especial de $H(s) = 1$. A substituição dessa equação na Eq. (4.35) resulta em

$$\begin{aligned}\Theta_e(s) &= \frac{s}{s+AK}\left[\frac{\omega_0-\omega_c}{s^2} + \frac{\varphi_0}{s}\right] \\ &= \frac{(\omega_o-\omega_c)/AK}{s} - \frac{(\omega_0-\omega_c)/AK}{s+AK} + \frac{\varphi_0}{s+AK}\end{aligned}$$

Logo,

$$\theta_e(t) = \frac{(\omega_0-\omega_c)}{AK}\left(1-e^{-AKt}\right) + \varphi_0 e^{-AKt} \tag{4.36a}$$

Observemos que

$$\lim_{t\to\infty}\theta_e(t) = \frac{\omega_0-\omega_c}{AK} \tag{4.36b}$$

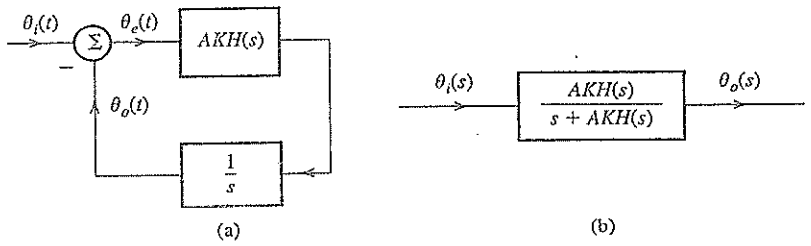


Figura 4.27 Circuitos equivalentes de uma PLL linearizada.

*Com uma diferença $\pi/2$.

Portanto, depois da morte do transiente (em cerca de $4/AK$ segundos), o erro de fase mantém um valor constante $(\omega_0 - \omega_c)/AK$. Isso significa que a frequência da PLL, por fim, se iguala à frequência recebida ω_0. Existe, no entanto, um erro de fase constante. A saída da PLL é

$$B \cos\left[\omega_0 t + \varphi_0 - \frac{\omega_0 - \omega_c}{AK}\right]$$

Para uma PLL de segunda ordem que usa

$$H(s) = \frac{s+a}{s} \tag{4.37a}$$

$$\Theta_e(s) = \frac{s}{s + AKH(s)}\Theta_i(s) \tag{4.37b}$$

$$= \frac{s^2}{s^2 + AK(s+a)}\left[\frac{\omega_0 - \omega_c}{s^2} + \frac{\varphi_0}{s}\right]$$

o teorema do valor final fornece diretamente[6]

$$\lim_{t\to\infty} \theta_e(t) = \lim_{s\to 0} s\Theta_e(s) = 0 \tag{4.38}$$

Nesse caso, a PLL adquire, por fim, tanto a frequência como a fase do sinal recebido.

Podemos usar a análise sob pequeno erro para mostrar que uma malha de primeira ordem não é capaz de rastrear um sinal recebido cuja frequência instantânea varia linearmente com o tempo. No entanto, com o emprego de uma malha de segunda ordem [Eq. (4.37)], um sinal desse tipo pode ser rastreado, a menos de uma fase constante (erro de fase constante); com uma malha de terceira ordem, o sinal pode ser rastreado com erro de fase nulo.[7]

Vale ressaltar que a análise anterior leva em consideração um modelo linear, válido apenas quando $\theta_e(t) << \pi/2$. Isso significa que as frequências ω_0 e ω_c devem ser muito próximas para que essa análise seja válida. No caso geral, devemos usar o modelo não linear na Fig. 4.26b. Para esse tipo de análise, sugerimos uma consulta a Viterbi,[7] Gardner[8] ou Lindsey.[9]

Análise de Malha de Primeira Ordem

Aqui, usaremos o modelo não linear na Fig. 4.26b, considerando o caso simples $H(s) = 1$, para o qual $h(t) = \delta(t)$,* e a Eq. (4.33) fornece

$$\dot{\theta}_o(t) = AK \operatorname{sen} \theta_e(t)$$

Como $\theta_e = \theta_i - \theta_o$

$$\dot{\theta}_e = \dot{\theta}_i - AK \operatorname{sen} \theta_e(t). \tag{4.39}$$

Consideremos, agora, o problema da aquisição de frequência e de fase. Suponhamos que o sinal que chega seja A sen $(\omega_0 t + \varphi_0)$ e que o VCO tenha uma frequência quiescente ω_c. Logo,

$$\theta_i(t) = (\omega_0 - \omega_c)t + \varphi_0$$

e

$$\theta_e = (\omega_0 - \omega_c) - AK \operatorname{sen} \theta_e(t) \tag{4.40}$$

*Na verdade, $h(t) = 2B$ sinc $(2\pi Bt)$, em que B é a largura de banda do filtro de malha. Esse é um filtro passa-baixos estreito, que suprime o sinal de alta frequência centrado em $2\omega_c$. Isto leva a $H(s) = 1$ em uma estreita banda passa-baixos de B Hz.

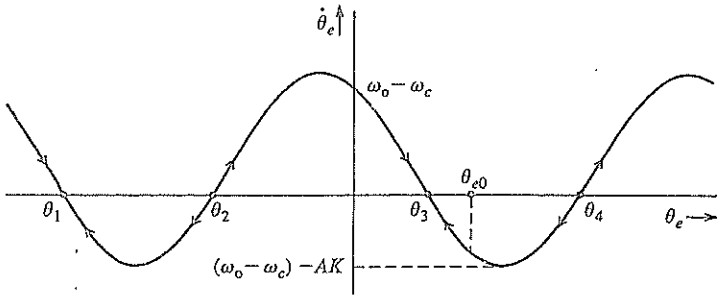


Figura 4.28 Trajetória de uma PLL de primeira ordem.

Para uma melhor compreensão do comportamento da PLL, usemos a Eq. (4.40) para esboçar o gráfico de $\dot{\theta}_e$ em função de θ_e. A Eq. (4.40) mostra que $\dot{\theta}_e$ é uma senoide deslocada verticalmente, como mostrado na Fig. 4.28. Para satisfazer a Eq. (4.40), a operação da malha deve permanecer na trajetória da senoide, como indicado na Fig. 4.28. Quando $\dot{\theta}_e = 0$, o sistema está em equilíbrio, pois, nesses pontos, θ_e deixa de variar no tempo. Assim, $\theta_e = \theta_1$, θ_2, θ_3 e θ_4 são pontos de equilíbrio.

Para um erro de fase inicial $\theta_e(0) = \theta_{e0}$ (Fig. 4.28), então $\dot{\theta}_e$ correspondente a esse valor de θ_e é negativo. Portanto, o erro de fase, inicialmente, diminuirá ao longo da trajetória da senoide, até o valor θ_3, em que é alcançado equilíbrio. Assim, em estado estacionário, o erro de fase é uma constante θ_3. Isso significa que a malha está travada em frequência; ou seja, a frequência do VCO passa a ser ω_0, mas há um erro de fase θ_3. Notemos, contudo, que, se $|\omega_0 - \omega_c| > AK$, não há pontos de equilíbrio na Fig. 4.28, a malha jamais alcança a condição de travamento e θ_e se desloca ao longo da trajetória indefinidamente. Portanto, essa malha simples pode alcançar travamento de fase desde que a diferença entre a frequência de entrada ω_0 e a frequência quiescente ω_c do VCO não seja superior a AK.

Na Fig. 4.28, existem vários pontos de equilíbrio. Metade desses pontos, no entanto, são pontos de equilíbrio instável, o que implica que uma pequena perturbação no estado do sistema afastará mais o ponto de operação destes pontos de equilíbrio. Os pontos θ_1 e θ_3 são estáveis, pois qualquer pequena perturbação ao estado do sistema tenderá a trazê-lo de volta a esses pontos. Consideremos, por exemplo, o ponto θ_3. Caso o estado seja perturbado ao longo da trajetória em direção à direita, $\dot{\theta}_e$ será negativo, o que tende a reduzir θ_e e trazê-lo de volta a θ_3. Caso o ponto de operação seja perturbado a partir de θ_3 para a esquerda, $\dot{\theta}_e$ será positivo, e θ_e tenderá a crescer e o ponto de operação retornará a θ_3. Contudo, se o ponto de operação for perturbado em θ_2 em direção à direita, $\dot{\theta}_e$ será positivo, e θ_e aumentará até alcançar θ_3. Se, em θ_2, o ponto de operação for perturbado para a esquerda, $\dot{\theta}_e$ será negativo e θ_e aumentará até alcançar θ_1. Portanto, θ_2 é um ponto de equilíbrio instável. Qualquer perturbação, como ruído, deslocará o ponto de operação para θ_1 ou θ_3. Do mesmo modo, podemos mostrar que θ_4 é um ponto instável e que θ_1 é um ponto de equilíbrio estável.

O ponto de equilíbrio θ_3 ocorre onde $\dot{\theta}_e = 0$. Logo, da Eq. (4.40),

$$\theta_3 = \text{sen}^{-1} \frac{\omega_0 - \omega_c}{AK}$$

Se $\theta_e << \pi/2$, então

$$\theta_3 \simeq \frac{\omega_0 - \omega_c}{AK}$$

o que concorda com o resultado obtido anteriormente na análise sob pequeno erro [Eq. (4.36b)].

A malha de primeira ordem está sujeita a um erro de fase constante. Além disso, pode alcançar travamento em frequência somente se a diferença entre a frequência que chega e a frequência quiescente do VCO não ultrapassar AK rad/s. Malhas de ordens superiores superam essas desvantagens, mas criam um novo problema de estabilidade. Uma análise mais detalhada pode ser encontrada em Gardener.[8]

Generalização de Comportamentos de PLL

Para generalização, suponhamos que a malha esteja *travada*, ou seja, que as frequências das senoides de entrada e de saída sejam iguais. Nesse caso, dizemos que os dois sinais são mutuamente **coeren-**

tes em fase ou estão em **travamento de fase**. O VCO rastreia a frequência e a fase do sinal recebido. Uma PLL é capaz de rastrear a frequência de entrada somente em uma faixa finita de deslocamentos de frequência. Essa faixa é denominada faixa de **retenção** (*hold-in range*) ou de **travamento** (*lock range*). Caso, inicialmente, as frequências de entrada e de saída não sejam suficientemente próximas, a malha pode não alcançar o travamento. A faixa de frequências na qual a entrada provocará o travamento da malha é denominada faixa de **aquisição** (*pull-in range*) ou de **captura** (*capture range*). Caso a frequência de entrada varie muito rapidamente, a malha pode não alcançar travamento.

Se a senoide de entrada for ruidosa, a PLL não apenas é capaz de seguir a senoide, mas também de limpá-la A PLL ainda pode ser usada como um demodulador de sinais modulados em frequência (FM) e como sintetizador de frequência, como mostraremos no próximo capítulo. Multiplicadores e divisores de frequência também podem ser construídos com PLL. Por ser um circuito integrado relativamente barato, a PLL se tornou um dos circuitos de comunicação mais utilizados.

Em naves espaciais, devido ao deslocamento Doppler e à deriva do oscilador, a frequência do sinal recebido contém muita incerteza. O deslocamento Doppler da portadora pode chegar a ±75 kHz, enquanto a banda do sinal modulado desejado pode ser de apenas 10 Hz. A recepção de um sinal desse tipo com receptores comuns exigiria um filtro de largura de banda de 150 kHz, quando o sinal desejado tem uma largura de banda de apenas 10 Hz. Isso causaria um aumento indesejável no ruído do receptor (por um fator de 15.000), pois a potência de ruído é proporcional à largura de banda. Em casos como esse, a PLL se mostra conveniente, pois é capaz de seguir continuamente a frequência recebida, de modo que a necessária largura de banda do filtro passa a ser de apenas 10 Hz.

Aquisição de Portadora em DSB-SC

A seguir, discutiremos dois métodos de regeneração de portadora com o uso de PLL no receptor em DSB-SC: método quadrático e malha Costas.

Método Quadrático:

Este esquema é delineado na Fig. 4.29. O sinal recebido é elevado ao quadrado e aplicado a um filtro passa-faixa estreito (de alto Q) sintonizado em $2\omega_c$. A saída desse filtro é a senoide $k \cos 2\omega_c t$, juntamente com algum sinal residual indesejável. Esse sinal é aplicado a uma PLL, para a obtenção de uma senoide mais limpa no dobro da frequência portadora; a senoide é, então, passada por um divisor de frequências 2:1, resultando em uma portadora local em sincronismo de fase e frequência com a portadora recebida. A análise é simples. O sinal quadrático $x(t)$ é

$$x(t) = [m(t) \cos \omega_c t]^2 = \tfrac{1}{2}m^2(t) + \tfrac{1}{2}m^2(t) \cos 2\omega_c t$$

Como $m^2(t)$ é um sinal não negativo, tem valor médio não nulo [ao contrário de $m(t)$, que, em geral, tem valor médio nulo]. Suponhamos que o valor médio, componente dc de $m^2(t)/2$, seja k. Com isso, podemos expressar $m^2(t)/2$ como

$$\frac{1}{2}m^2(t) = k + \phi(t)$$

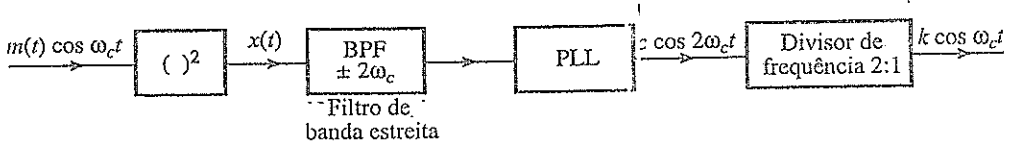


Figura 4.29 Uso do quadrado do sinal para gerar uma portadora para demodulação coerente.

em que $\phi(t)$ é um sinal em banda base de média zero [$m^2(t)$ menos sua componente dc]. Assim,

$$x(t) = \frac{1}{2}m^2(t) + \frac{1}{2}m^2(t)\cos 2\omega_c t$$
$$= \frac{1}{2}m^2(t) + k\cos 2\omega_c t + \phi(t)\cos 2\omega_c t$$

O filtro passa-faixa é um filtro de banda estreita (de alto Q) sintonizado em $2\omega_c$. Esse filtro suprime totalmente o sinal $m^2(t)$, cujo espectro é centrado em $\omega = 0$. O filtro também elimina a maior parte do sinal $\phi(t)\cos 2\omega_c t$. Isso ocorre porque o sinal, embora tenha espectro centrado em $2\omega_c$, tem potência zero (infinitesimal) em $2\omega_c$, pois $\phi(t)$ tem um valor dc zero. Além disso, essa componente é distribuída ao longo da banda de $4B$ Hz centrada em $2\omega_c$. Portanto, muito pouco desse sinal passa pelo filtro de banda estreita.* Em contraste, o espectro de $k\cos 2\omega_c t$ consiste em impulsos localizados em $\pm 2\omega_c$. Logo, toda sua potência é centrada em $2\omega_c$ e passará pelo filtro. Portanto, a saída do filtro é $k\cos 2\omega_c t$ mais um pequeno resíduo indesejado de $\phi(t)\cos 2\omega_c t$. Esse resíduo pode ser eliminado com o uso de uma PLL que siga $k\cos 2\omega_c t$. A saída da PLL, depois da passagem por um divisor de frequências 2:1, fornece a portadora desejada. Aqui, cabe uma ressalva. Como o sinal recebido é perdido ao ser elevado ao quadrado, temos uma ambiguidade de polaridade (ou ambiguidade de fase π) na portadora gerada. No caso de sinais analógicos, isso é irrelevante. Contudo, no caso de sinais digitais em banda base, a polaridade da portadora é essencial e esse método não pode, portanto, ser usado diretamente.

Malha de Costas: Outro esquema para a geração de uma portadora local, proposto por Costas,[10] é mostrado na Fig. 4.30. O sinal recebido é $m(t)\cos(\omega_c t + \theta_i)$. No receptor, um VCO gera a portadora $\cos(\omega_c t + \theta_o)$. O erro de fase é $\theta_e = \theta_i - \theta_o$. Vários sinais são indicados na Fig. 4.30. Os dois filtros passa-baixas suprimem termos de altas frequências e fornecem $m(t)\cos\theta_e$ e $m(t)\,\text{sen}\,\theta_e$. Essas saídas são, então, multiplicadas, resultando em $m^2(t)\,\text{sen}\,2\theta_e$. A aplicação desse sinal a um filtro passa-baixas de banda estreita produz a saída $R\,\text{sen}\,2\theta_e$, em que R é a componente dc de $m^2(t)/2$. O sinal $R\,\text{sen}\,2\theta_e$ é aplicado à entrada de um VCO com frequência quiescente ω_c. A entrada $R\,\text{sen}\,2\theta_e$ aumenta a frequência de saída, que, por sua vez, reduz θ_e. Esse mecanismo foi discutido em detalhes no contexto da Fig. 4.26.

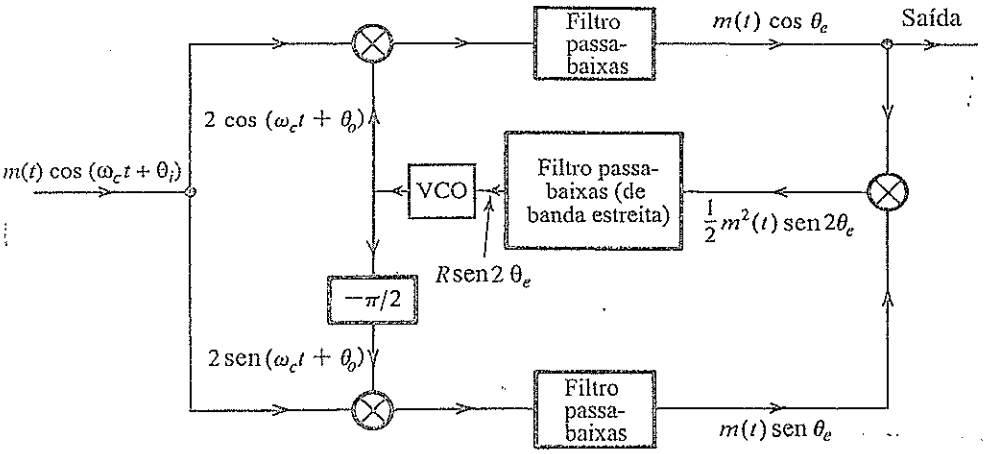


Figura 4.30 Malha de aquisição de fase de Costas para a geração de uma portadora para demodulação coerente.

*Isso também explica por que não podemos extrair a portadora diretamente de $m(t)\cos\omega_c t$ passando por um filtro de banda estreita centrado em ω_c. A razão é que a potência de $m(t)\cos\omega_c t$ em ω_c é zero, pois $m(t)$ não tem componente dc [o valor médio de $m(t)$ é zero].

Aquisição de Portadora em SSB-SC

Para sincronização em um receptor SSB-SC, podemos usar osciladores a cristal altamente estáveis, com cristais cortados para a mesma frequência no transmissor e no receptor. Em frequências muito altas, nas quais cristais de quartzo podem não ter desempenho adequado, uma portadora piloto pode ser transmitida. Esses são os mesmos métodos empregados para DSB-SC. Contudo, nem o método quadrático nem a malha de Costas, utilizados em DSB-SC, podem ser usados em SSB-SC. Para comprovar isto, escrevemos o sinal SSB-SC como

$$\begin{aligned}\varphi_{\text{SSB}}(t) &= m(t)\cos\,\omega_c t \mp m_h(t)\,\text{sen}\,\omega_c t\\ &= E(t)\cos\,[\omega_c t + \theta(t)]\end{aligned}$$

em que

$$E(t) = \sqrt{m^2(t) + m_h^2(t)}$$

$$\theta(t) = \tan^{-1}\left[\frac{\pm m_h(t)}{m(t)}\right]$$

Elevando este sinal ao quadrado, obtemos

$$\begin{aligned}\varphi_{\text{SSB}}^2(t) &= E^2(t)\cos^2\,[\omega_c t + \theta(t)]\\ &= \frac{E^2(t)}{2}\{1 + \cos\,[2\omega_c t + 2\theta(t)]\}\end{aligned}$$

O sinal $E^2(t)$ é eliminado por um filtro passa-faixa. Lamentavelmente, o sinal remanescente não é uma senoide pura de frequência $2\omega_c$ (como no caso DSB). Nada podemos fazer para remover a fase variante no tempo $2\theta(t)$ dessa senoide. Assim, para SSB, a técnica quadrática não funciona. O mesmo argumento pode ser usado para mostrar que a malha de Costas também não funciona. Essas conclusões se aplicam, ainda, a sinais VSB.

4.9 EXERCÍCIOS COM O MATLAB

Nesta seção, apresentamos exercícios baseados em MATLAB para reforçar alguns dos conceitos básicos de modulação analógica cobertos nas seções anteriores. Usaremos exemplos que podem ilustrar a modulação e a demodulação de sinais DSB-SC, AM, SSB-SC e QAM.

Modulação e Demodulação DSB-SC

O primeiro programa MATLAB, `triplesinc.m`, gera um sinal que é (quase) estritamente limitado em banda e consiste em três versões atrasadas de um mesmo sinal:

$$m_2(t) = 2\,\text{sinc}\,(2t/T_a) + \text{sinc}\,(2t/T_a + 1) + \text{sinc}\,(2t/T_a - 1)$$

```
% (triplesinc.m)
%   Sinal em banda base para AM
%   Uso: m=triplesinc(t,Ta)
    function m=triplesinc(t,Ta)
%   t é a duração do sinal
%   Ta é o parâmetro, igual ao dobro do atraso
%
    sig_1=sinc(2*t/Ta);
    sig_2=sinc(2*t/Ta-1);
    sig_3=sinc(2*t/Ta+1);
    m=2*sig_1+sig_2+sig_3;
end
```

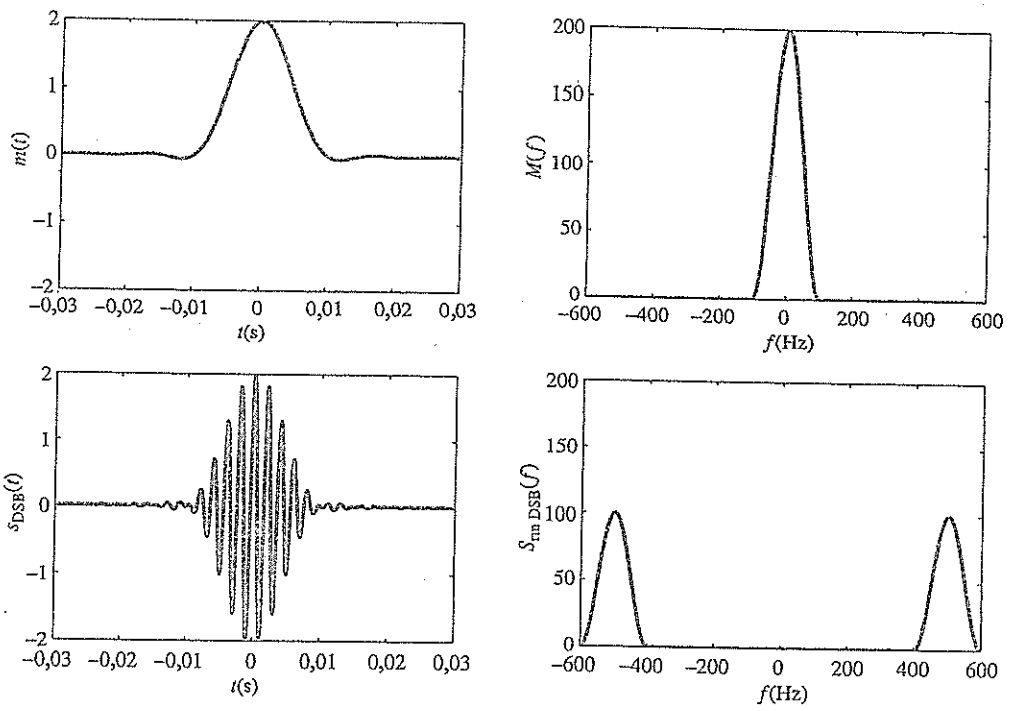


Figura 4.31 Sinais do exemplo, nos domínios do tempo e da frequência durante a modulação DSB.

O sinal DSB-SC pode ser gerado com o arquivo MATLAB `ExampleDSB.m` que gera um sinal DSB-SC para $t = \in (-0{,}04, 0{,}04)$. A frequência portadora é 300 Hz. O sinal de mensagem original e o sinal DSB-SC, nos domínios do tempo e da frequência, são ilustrados na Fig. 4.31.

```
% (ExampleDSB.m)
% Este programa usa triplesinc.m para ilustrar a modulação
% e a demodulação DSB

ts=1.e-4

t=-0.04:ts:0.04;
Ta=0.01;
m_sig=triplesinc(t,Ta);
Lfft=length(t); Lfft=2^ceil(log2(Lfft));
M_fre=fftshift(fft(m_sig,Lfft));
freqm=(-Lfft/2:Lfft/2-1)/(Lfft*ts);

s_dsb=m_sig.*cos(2*pi*500*t);
Lfft=length(t); Lfft=2^ceil(log2(Lfft)+1);
S_dsb=fftshift(fft(s_dsb,Lfft));
freqs=(-Lfft/2:Lfft/2-1)/(Lfft*ts);

Trange=[-0.03 0.03 -2 2]
figure(1)
subplot(221);td1=plot(t,m_sig);
axis(Trange); set(td1,'Linewidth',2);
```

```
xlabel('{\it t} (sec)'); ylabel('{\it m}({\it t})')
subplot(223);td2=plot(t,s_dsb);
axis(Trange); set(td2,'Linewidth',2);
xlabel('{\it t} (sec)'); ylabel('{\it s}_{\rm DSB}({\it t})')

Frange=[-600 600 0 200]
subplot(222);fd1=plot(freqm,abs(M_fre));
axis(Frange); set(fd1,'Linewidth',2);
xlabel('{\it f} (Hz)'); ylabel('{\it M}({\it f})')
subplot(224);fd2=plot(freqs,abs(S_dsb));
axis(Frange); set(fd2,'Linewidth',2);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{rm DSB}({\it f})')
```

O primeiro exemplo de modulação, `ExampleDSBdemfilt.m`, é baseado em um sinal de mensagem estritamente passa-baixas $m_0(t)$. A seguir, geramos um sinal de mensagem diferente, que não é estritamente limitado em banda. Na verdade, o novo sinal de mensagem consiste em dois triângulos:

$$m_1(t) = \Delta\left(\frac{t+0{,}01}{0{,}01}\right) - \Delta\left(\frac{t-0{,}01}{0{,}01}\right)$$

A demodulação coerente também é implementada com um filtro passa-baixas de resposta ao impulso finita (FIR — *finite impulse response*) de ordem 40. O sinal de mensagem original $m(t)$, o sinal DSB-SC $m(t)$ dos $\omega_c t$, o sinal do demodulador $e(t) = m(t)\cos^2 \omega_c t$ e o sinal de mensagem $m_d(t)$ recuperado após filtragem passa-baixas são ilustrados na Fig. 4.32, no domínio do tempo, e na Fig. 4.33, no domínio da frequência. O filtro passa-baixas no demodulador tem largura de banda de 150 Hz. O resultado da demodulação quase não exibe distorção.

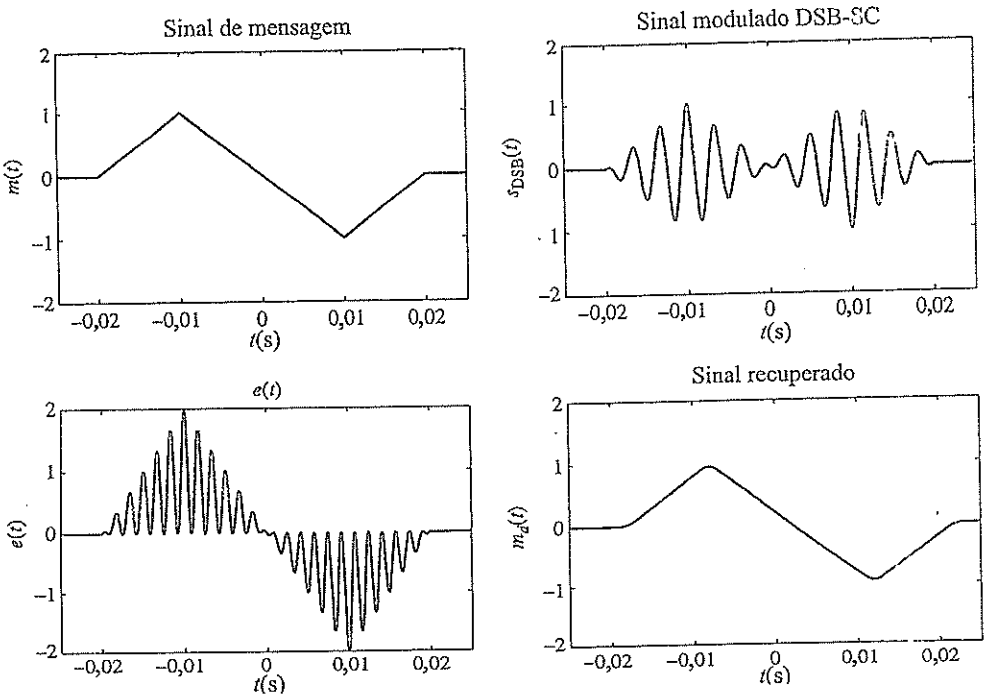


Figura 4.32 Sinais nos domínios do tempo durante a modulação e a demodulação DSB.

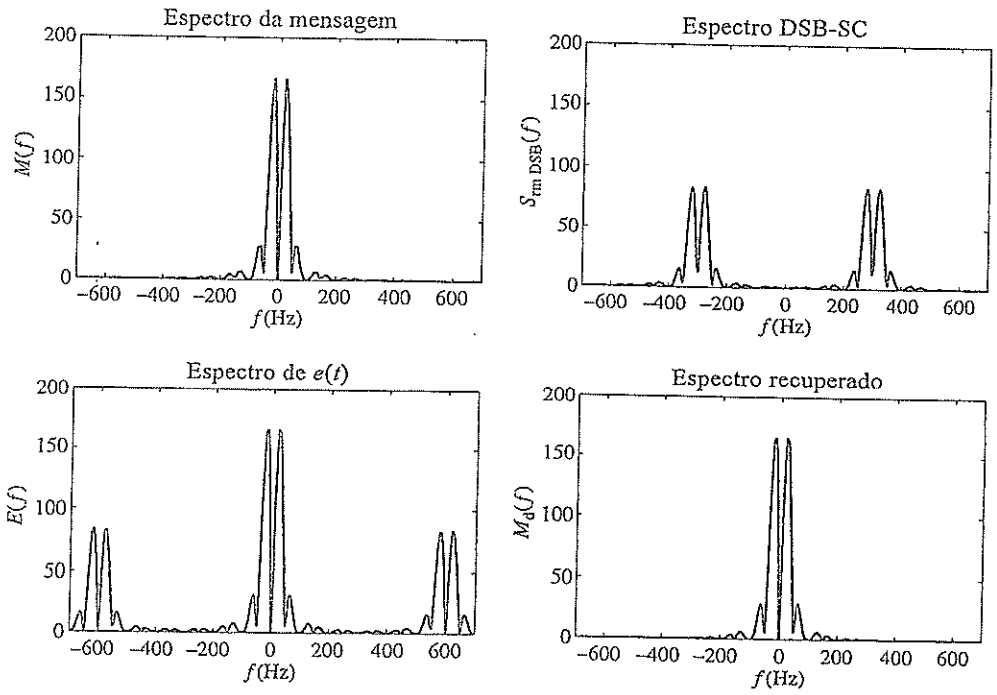


Figura 4.33 Sinais nos domínios da frequência durante a modulação e a demodulação DSB-SC.

```
% (ExampleDSBdemfilt.m)
% Este programa usa triangl.m para ilustrar modulação
% e demodulação DSB

ts=1.e-4;

t=-0.04:ts:0.04;
Ta=0.01;
m_sig=triangl((t+0.01)/0.01)-triangl((t-0.01)/0.01);
Lm_sig=length(m_sig);
Lfft=length(t);
Lfft=2^ceil(log2(Lfft));
M_fre=fftshift(fft(m_sig,Lfft));
freqm=(-Lfft/2:Lfft/2-1)/(Lfft*ts);
B_m=150; %Bandwidth of the signal is B_m Hz.
h=fir1(40,[B_m*ts]);

t=-0.04:ts:0.04;
Ta=0.01;fc=300;
s_dsb=m_sig.*cos(2*pi*fc*t);
Lfft=length(t); Lfft=2^ceil(log2(Lfft)+1);
S_dsb=fftshift(fft(s_dsb,Lfft));
freqs=(-Lfft/2:Lfft/2-1)/(Lfft*ts);

% Demodulação tem início com a multiplicação pela portadora
s_dem=s_dsb.*cos(2*pi*fc*t)*2;
S_dem=fftshift(fft(s_dem,Lfft));
```

```
% Uso de filtro passa-baixos ideal com largura de banda de 150 Hz
s_rec=filter(h,1,s_dem);
S_rec=fftshift(fft(s_rec,Lfft));

Trange=[-0.025 0.025 -2 2];
figure(1)
subplot(221);td1=plot(t,m_sig);
axis(Trange); set(td1,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}({\it t})');
title('sinal de mensagem');
subplot(222);td2=plot(t,s_dsb);
axis(Trange); set(td2,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it s}_{\rm DSB}({\it t})')
title('sinal modulado DSB-SC');
subplot(223);td3=plot(t,s_dem);
axis(Trange); set(td3,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it e}({\it t})')
title('{\it e}({\it t})');
subplot(224);td4=plot(t,s_rec);
axis(Trange); set(td4,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}_d({\it t})')
title('Sinal recuperado');

Frange=[-700 700 0 200];
figure(2)
subplot(221);fd1=plot(freqm,abs(M_fre));
axis(Frange); set(fd1,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}({\it f})');
title('espectro da mensagem');
subplot(222);fd2=plot(freqs,abs(S_dsb));
axis(Frange); set(fd2,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{rm DSB}({\it f})');
title('espectro DSB-SC');
subplot(223);fd3=plot(freqs,abs(S_dem));
axis(Frange); set(fd3,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it E}({\it f})');
title('espectro de {\it e}({\it t})');
subplot(224);fd4=plot(freqs,abs(S_rec));
axis(Frange); set(fd4,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}_d({\it f})');
title('espectro recuperado');
```

Modulação e Demodulação AM

Neste exercício, geraremos um sinal AM convencional, com índice de modulação $\mu = 1$. Usando o mesmo sinal de mensagem $m(t)$, o programa MATLAB `ExampleAMdemfilt.m` gera o sinal de mensagem, o correspondente sinal AM, o sinal retificado em demodulação não coerente e o sinal retificado após filtragem passa-baixas. O filtro passa-baixas no demodulador tem largura de banda de 150 Hz. Os sinais no domínio do tempo são ilustrados na Fig. 4.34, e os correspondentes sinais no domínio da frequência, na Fig. 4.35.

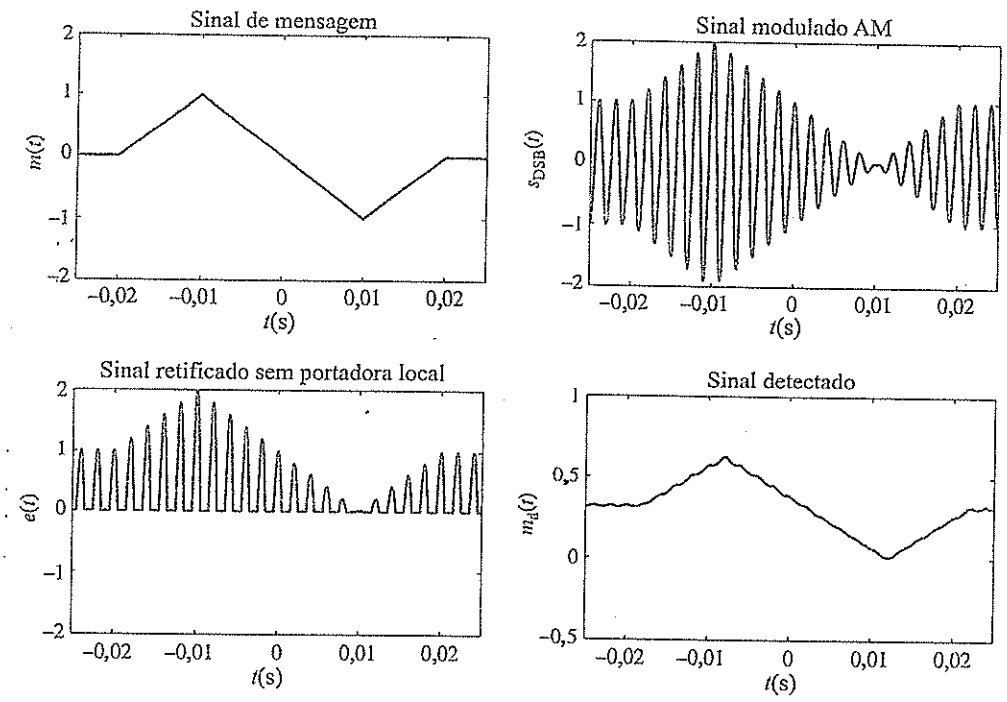


Figura 4.34 Sinais no domínio do tempo em modulação AM e demodulação não coerente.

Notemos o grande impulso no espectro do sinal AM. A limitada janela temporal significa que um impulso ideal não é possível e que apenas picos intensos, centrados na frequência portadora ±300 Hz, são visíveis. Por fim, como o sinal de mensagem não é estritamente limitado em banda, a relativamente baixa frequência portadora de 300 Hz força o filtro passa-baixas no demodulador a truncar algumas das componentes do sinal de mensagem no demodulador. Distorção é visível nas proximidades dos cantos abruptos do sinal recuperado.

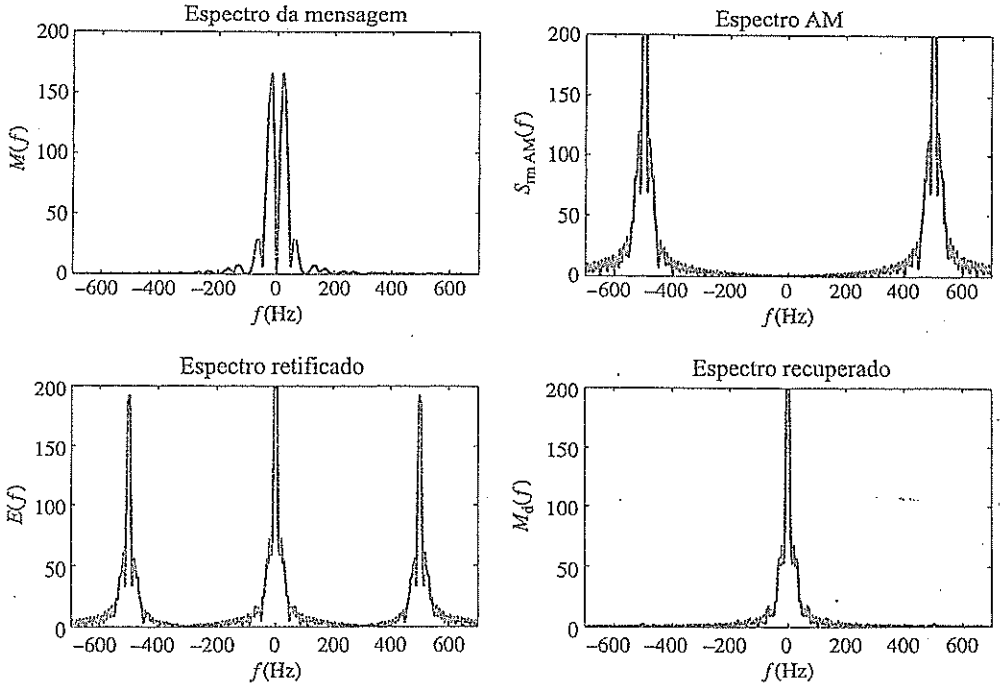


Figura 4.35 Sinais no domínio da frequência em modulação AM e demodulação não coerente.

```
% (ExampleAMdemfilt.m)
% Este programa usa triangl.m para ilustrar modulação
% e demodulação AM

ts=1.e-4;
t=-0.04:ts:0.04;
Ta=0.01; fc=500;
m_sig=triangl((t+0.01)/0.01)-triangl((t-0.01)/0.01);
Lm_sig=length(m_sig);
Lfft=length(t); Lfft=2^ceil(log2(Lfft));
M_fre=fftshift(fft(m_sig,Lfft));
freqm=(-Lfft/2:Lfft/2-1)/(Lfft*ts);
B_m=150; %Bandwidth of the signal is B_m Hz.
h=fir1(40,[B_m*ts]);

% Sinal AM gerado com a adição de uma portadora a DSB-SC
s_am=(1+m_sig).*cos(2*pi*fc*t);
Lfft=length(t); Lfft=2^ceil(log2(Lfft)+1);
S_am=fftshift(fft(s_am,Lfft));
freqs=(-Lfft/2:Lfft/2-1)/(Lfft*ts);

% Demodulação tem início com o uso de um retificador
s_dem=s_am.*(s_am>0);
S_dem=fftshift(fft(s_dem,Lfft));

% Uso de filtro passa-baixas ideal com largura de banda de 150 Hz
s_rec=filter(h,1,s_dem);
S_rec=fftshift(fft(s_rec,Lfft));

Trange=[-0.025 0.025 -2 2];
figure(1)
subplot(221);td1=plot(t,m_sig);
axis(Trange); set(td1,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}({\it t})');
title('sinal de mensagem');
subplot(222);td2=plot(t,s_am);
axis(Trange); set(td2,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it s}_{\rm DSB}({\it t})')
title('sinal modulado AM');
subplot(223);td3=plot(t,s_dem);
axis(Trange); set(td3,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it e}({\it t})')
title('sinal retificado sem portadora local');
subplot(224);td4=plot(t,s_rec);
Trangelow=[-0.025 0.025 -0.5 1];
axis(Trangelow); set(td4,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}_d({\it t})')
title('sinal detectado');

Frange=[-700 700 0 200];
figure(2)
subplot(221);fd1=plot(freqm,abs(M_fre));
axis(Frange); set(fd1,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}({\it f})');
title('espectro da mensagem');
subplot(222);fd2=plot(freqs,abs(S_am));
```

```
axis(Frange); set(fd2,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{rm AM}({\it f})');
title('espectro AM');
subplot(223);fd3=plot(freqs,abs(S_dem));
axis(Frange); set(fd3,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it E}({\it f})');
title('espectro retificado');
subplot(224);fd4=plot(freqs,abs(S_rec));
axis(Frange); set(fd4,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}_d({\it f})');
title('espectro recuperado');
```

Modulação e Demodulação SSB-SC

Para ilustrar os processos de modulação e de demodulação SSB-SC, esse exercício gera um sinal SSB-SC usando o mesmo sinal de mensagem $m_1(t)$ com triângulos duplos. A frequência portadora continua sendo 300 Hz. O programa MATLAB `ExampleSSBdemfilt.m` executa essa função. É aplicada demodulação coerente, na qual um simples filtro passa-baixas com largura de banda de 150 Hz é usado para produzir o sinal de mensagem recuperado.

Os sinais no domínio do tempo são mostrados na Fig. 4.36, enquanto os correspondentes sinais no domínio da frequência são mostrados na Fig. 4.37.

```
% (ExampleSSBdemfilt.m)
% Este programa usa triangl.m
% para ilustrar modulação SSB % e demodulação
clear;clf;

ts=1.e-4;
t=-0.04:ts:0.04;
Ta=0.01; fc=300;
m_sig=triangl((t+0.01)/0.01)-triangl((t-0.01)/0.01);
Lm_sig=length(m_sig);
Lfft=length(t); Lfft=2^ceil(log2(Lfft));
M_fre=fftshift(fft(m_sig,Lfft));
freqm=(-Lfft/2:Lfft/2-1)/(Lfft*ts);
B_m=150; % A largura de banda do sinal é B_m Hz.
h=fir1(40,[B_m*ts]);

s_dsb=m_sig.*cos(2*pi*fc*t);
Lfft=length(t); Lfft=2^ceil(log2(Lfft)+1);
S_dsb=fftshift(fft(s_dsb,Lfft));
L_lsb=floor(fc*ts*Lfft);
SSBfilt=ones(1,Lfft);
SSBfilt(Lfft/2-L_lsb+1:Lfft/2+L_lsb)=zeros(1,2*L_lsb);
S_ssb=S_dsb.*SSBfilt;
freqs=(-Lfft/2:Lfft/2-1)/(Lfft*ts);
s_ssb=real(ifft(fftshift(S_ssb)));
s_ssb=s_ssb(1:Lm_sig);

% A demodulação começa pela multiplicação com a portadora
s_dem=s_ssb.*cos(2*pi*fc*t)*2;
S_dem=fftshift(fft(s_dem,Lfft));
```

Figura 4.36 Sinais no domínio do tempo durante modulação SSB-SC e demodulação coerente.

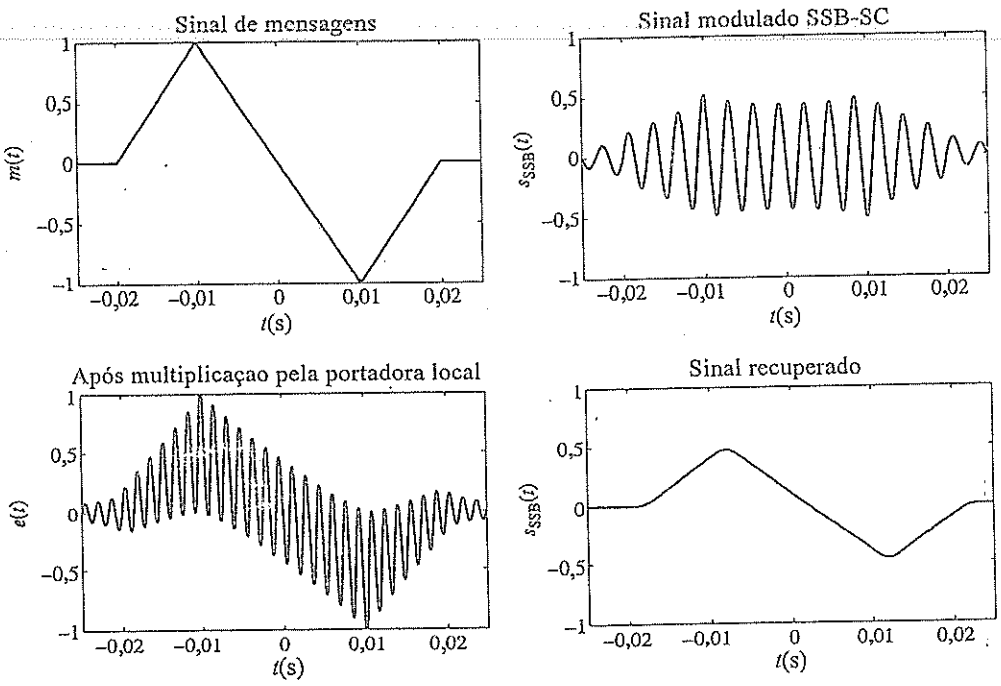
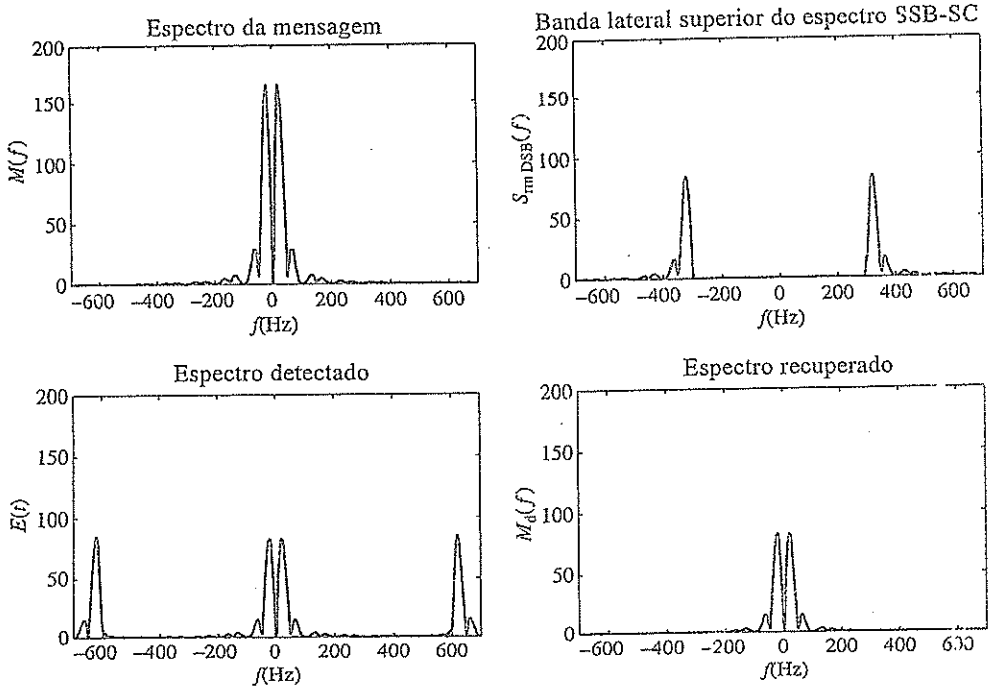


Figura 4.37 Sinais no domínio da frequência durante modulação SSB-SC e demodulação coerente.



```
% Usando um filtro PB com largura de banda 150 Hz
s_rec=filter(h,1,s_dem);
S_rec=fftshift(fft(s_rec,Lfft));

Trange=[-0.025 0.025 -1 1];
figure(1)
subplot(221);td1=plot(t,m_sig);
axis(Trange); set(td1,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}({\it t})');
title('sinal de mensagem');
subplot(222);td2=plot(t,s_ssb);
axis(Trange); set(td2,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it s}_{\rm SSB}({\it t})')
title('sinal modulado SSB-SC');
subplot(223);td3=plot(t,s_dem);
axis(Trange); set(td3,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it e}({\it t})')
title('após multiplicação pela portadora local');
subplot(224);td4=plot(t,s_rec);
axis(Trange); set(td4,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}_d({\it t})')
title('Sinal recuperado');

Frange=[-700 700 0 200];
figure(2)
subplot(221);fd1=plot(freqm,abs(M_fre));
axis(Frange); set(fd1,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}({\it f})');
title('espectro da mensagem');
subplot(222);fd2=plot(freqs,abs(S_ssb));
axis(Frange); set(fd2,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{rm DSB}({\it f})');
title('banda lateral superior do espectro SSB-SC');
subplot(223);fd3=plot(freqs,abs(S_dem));
axis(Frange); set(fd3,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it E}({\it f})');
title('espectro detectado');
subplot(224);fd4=plot(freqs,abs(S_rec));
axis(Frange); set(fd4,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}_d({\it f})');
title('espectro recuperado');
```

Modulação e Demodulação QAM

Neste exercício, aplicaremos a QAM para modular e demodular dois sinais de mensagem $m_1(t)$ e $m_2(t)$. A frequência portadora permanece em 300 Hz, mas dois sinais são modulados e detectados simultaneamente. O sinal QAM é demodulado coerentemente com a multiplicação por cos $600\pi\, t$ e sen $600\pi\, t$, para recuperar os dois sinais de mensagem. Cada produto é filtrado pelo mesmo filtro passa-baixas de ordem 40. O programa MATLAB `ExampleQAMdemfilt.m` completa essa ilustração produzindo gráficos dos sinais no domínio do tempo, durante a modulação e a demodulação do primeiro sinal $m_1(t)$ e do segundo sinal $m_2(t)$. Os resultados para $m_1(t)$ nos domínios do tempo e da frequência são mostrados nas Fig. 4.38 e Fig. 4.39, respectivamente. As Fig. 4.40 e Fig. 4.41 mostram, respectivamente, os resultados para $m_2(t)$ nos domínios do tempo e da frequência.

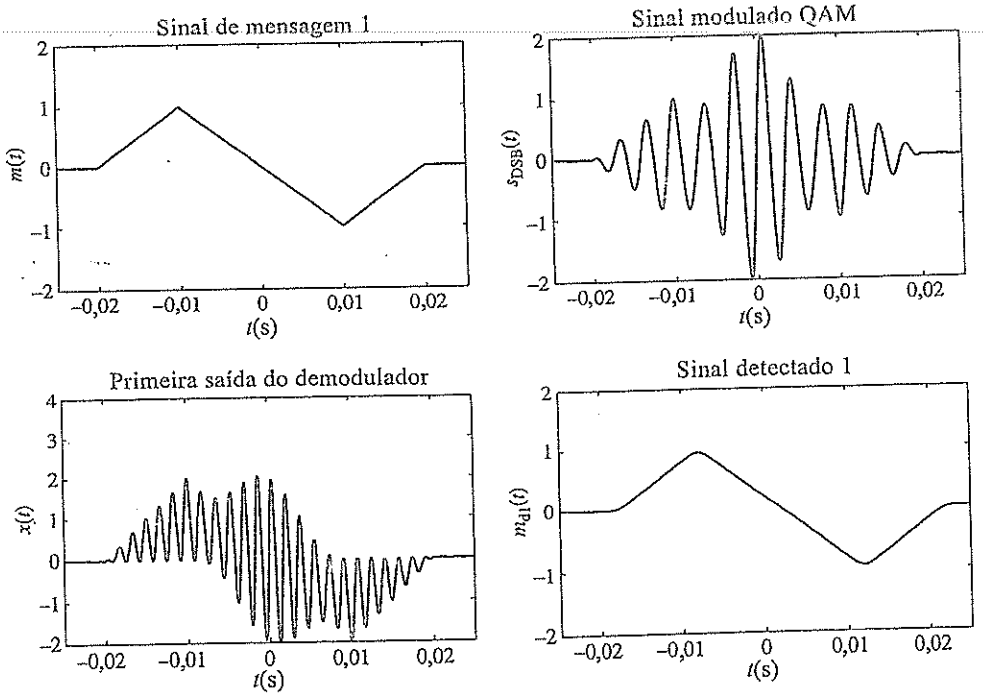


Figura 4.38 Sinais no domínio do tempo durante modulação QAM e demodulação coerente da primeira mensagem $m_2(t)$.

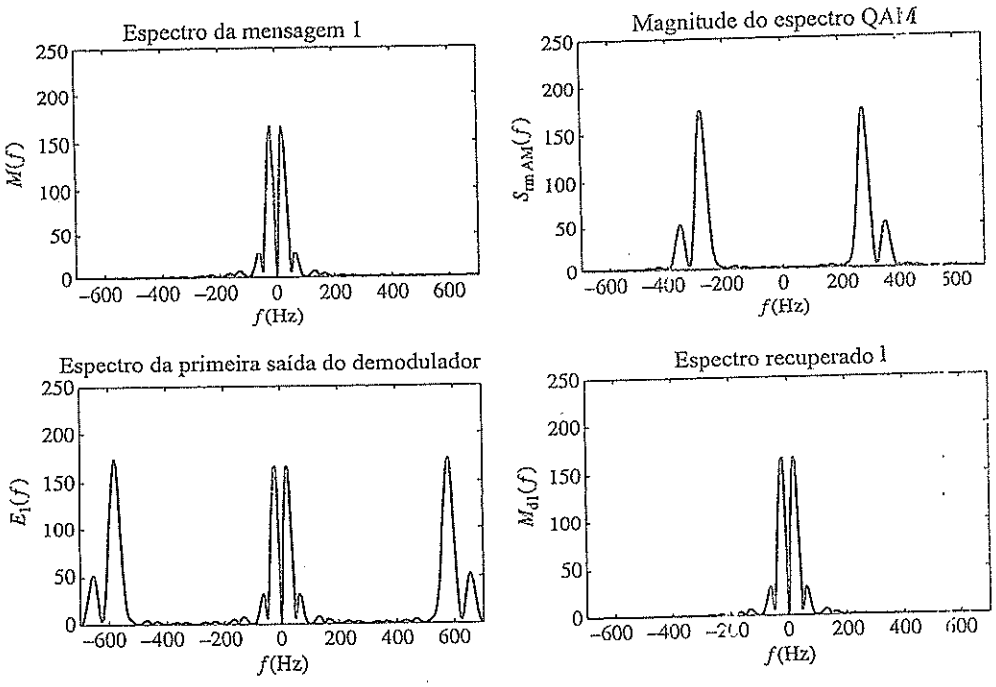


Figura 4.39 Sinais no domínio da frequência durante modulação QAM e demodulação coerente da primeira mensagem $m_2(t)$.

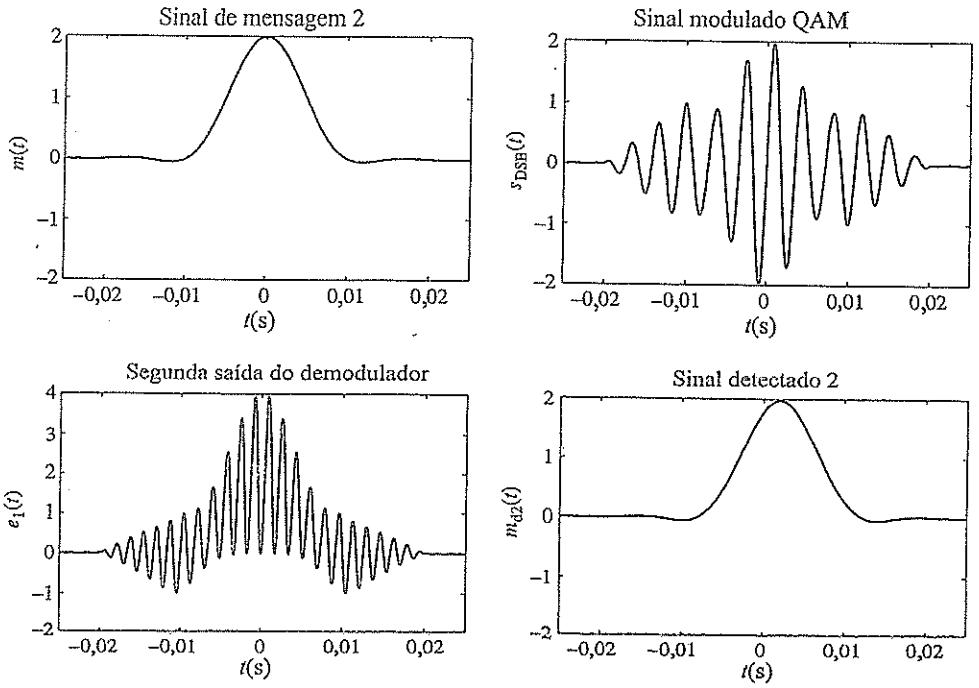


Figura 4.40 Sinais no domínio do tempo durante modulação QAM e demodulação coerente da segunda mensagem $m_2(t)$.

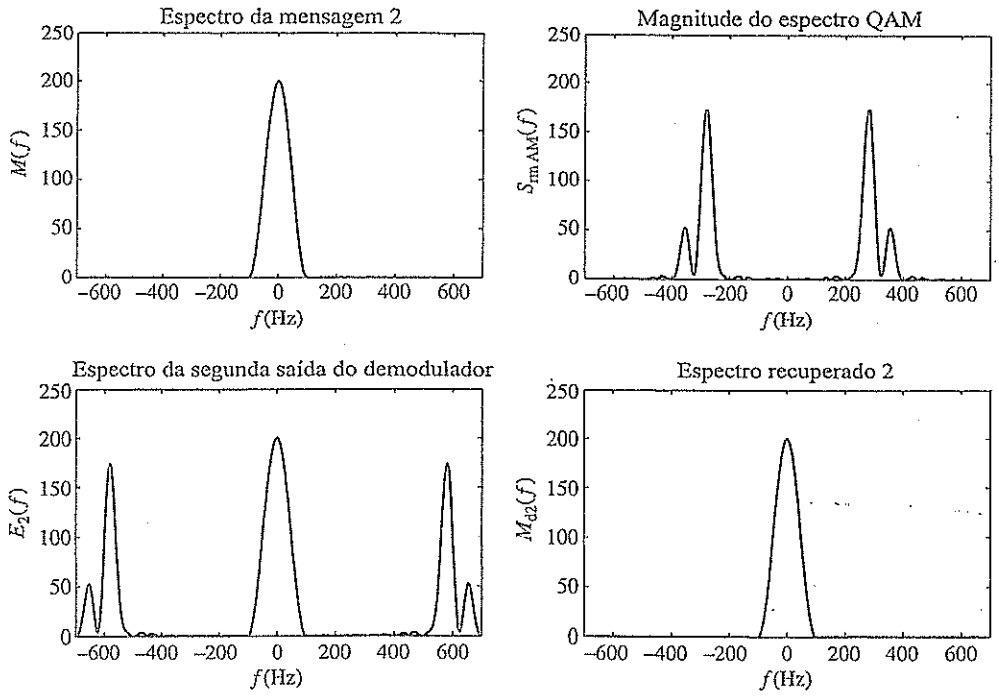


Figura 4.41 Sinais no domínio da frequência durante modulação QAM e demodulação coerente da segunda mensagem $m_2(t)$.

```
% (ExampleQAMdemfilt.m)
% Este programa usa triangl.m e triplesinc.m
% para ilustrar modulação QAM % e demodulação
% de dois sinais de mensagem
clear;clf;
ts=1.e-4;
t=-0.04:ts:0.04;
Ta=0.01; fc=300;
% Usa triangl.m e triplesinc.m para gerar
% dois sinais de mensagem de diferentes formas e espectros
m_sig1=triangl((t+0.01)/0.01)-triangl((t-0.01)/0.01);
m_sig2=triplesinc(t,Ta);
Lm_sig=length(m_sig1);
Lfft=length(t); Lfft=2^ceil(log2(Lfft));
M1_fre=fftshift(fft(m_sig1,Lfft));
M2_fre=fftshift(fft(m_sig2,Lfft));
freqm=(-Lfft/2:Lfft/2-1)/(Lfft*ts);
%
B_m=150; %A largura de banda do sinal é B_m Hz.
% Projeta um simples filtro passa-baixas com largura de banda B_m Hz.
h=fir1(40,[B_m*ts]);

% Sinal QAM gerado com a adição de uma portadora a DSB-SC
s_qam=m_sig1.*cos(2*pi*fc*t)+m_sig2.*sin(2*pi*fc*t);
Lfft=length(t); Lfft=2^ceil(log2(Lfft)+1);
S_qam=fftshift(fft(s_qam,Lfft));
freqs=(-Lfft/2:Lfft/2-1)/(Lfft*ts);

% Demodulação tem início com o uso de um retificador
s_dem1=s_qam.*cos(2*pi*fc*t)*2;
S_dem1=fftshift(fft(s_dem1,Lfft));
% Demodulação do segundo sinal
s_dem2=s_qam.*sin(2*pi*fc*t)*2;
S_dem2=fftshift(fft(s_dem2,Lfft));
%
% Uso de um filtro passa-baixas ideal com a largura de banda de 150 Hz

s_rec1=filter(h,1,s_dem1);
S_rec1=fftshift(fft(s_rec1,Lfft));
s_rec2=filter(h,1,s_dem2);
S_rec2=fftshift(fft(s_rec2,Lfft));

Trange=[-0.025 0.025 -2 2];
Trange2=[-0.025 0.025 -2 4];
figure(1)
subplot(221);td1=plot(t,m_sig1);
axis(Trange); set(td1,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}({\it t})');
title('sinal de mensagem 1');
subplot(222);td2=plot(t,s_qam);
axis(Trange); set(td2,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it s}_{\rm DSB}({\it t})')
title('sinal modulado QAM');
subplot(223);td3=plot(t,s_dem1);
axis(Trange2); set(td3,'Linewidth',1.5);
```

```
xlabel('{\it t} (sec)'); ylabel('{\it x}({\it t})')
title('primeira saída do demodulador');
subplot(224);td4=plot(t,s_rec1);
axis(Trange); set(td4,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}_{d1}({\it t})')
title('sinal detectado 1');

figure(2)
subplot(221);td5=plot(t,m_sig2);
axis(Trange); set(td5,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}({\it t})');
title('sinal de mensagem 2');
subplot(222);td6=plot(t,s_qam);
axis(Trange); set(td6,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it s}_{\rm DSB}({\it t})')
title('sinal modulado QAM');
subplot(223);td7=plot(t,s_dem2);
axis(Trange2); set(td7,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it e}_1({\it t})')
title('segunda saída do demodulador');
subplot(224);td8=plot(t,s_rec2);
axis(Trange); set(td8,'Linewidth',1.5);
xlabel('{\it t} (sec)'); ylabel('{\it m}_{d2}({\it t})')
title('sinal detectado 2');

Frange=[-700 700 0 250];
figure(3)
subplot(221);fd1=plot(freqm,abs(M1_fre));
axis(Frange); set(fd1,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}({\it f})');
title('espectro da mensagem 1');
subplot(222);fd2=plot(freqs,abs(S_qam));
axis(Frange); set(fd2,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{rm AM}({\it f})');
title('magnitude do espectro QAM');
subplot(223);fd3=plot(freqs,abs(S_dem1));
axis(Frange); set(fd3,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it E}_1({\it f})');
title('espectro da primeira saída do demodulador');
subplot(224);fd4=plot(freqs,abs(S_rec1));
axis(Frange); set(fd4,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}_{d1}({\it f})');
title('espectro recuperado 1');
figure(4)
subplot(221);fd1=plot(freqm,abs(M2_fre));
axis(Frange); set(fd1,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}({\it f})');
title('espectro da mensagem 2');
subplot(222);fd2=plot(freqs,abs(S_qam));
axis(Frange); set(fd2,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{rm AM}({\it f})');
title('magnitude do espectro QAM');
subplot(223);fd7=plot(freqs,abs(S_dem2));
axis(Frange); set(fd7,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it E}_2({\it f})');
```

```
title('espectro da segunda saída do demodulador');
subplot(224);fd8=plot(freqs,abs(S_rec2));
axis(Frange); set(fd8,'Linewidth',1.5);
xlabel('{\it f} (Hz)'); ylabel('{\it M}_{d2}({\it f})');
title('espectro recuperado 2');
```

REFERÊNCIAS

1. Single Sideband Issue, *Proc. IRE*, vol. 44, Dec. 1956.
2. D. K. Weaver Jr., "A Third Method of Generation and Detection of Single Sideband Signals," *Proc. IRE*, vol. 44, pp. 1703–1705, Dec. 1956.
3. Bell Telephone Laboratories, *Transmission Systems for Communication*, 4th ed., Murray Hill, NJ, 1970.
4. R. T. James, "AT&T Facilities and Services," *Proc. IEEE*, vol. 60, pp. 1342–1349, Nov. 1972.
5. W. L. Smith, "Frequency and Time in Communications," *Proc. IEEE*, vol. 60, pp. 589–594, May 1972.
6. B. P. Lathi, B. P., *Linear Systems and Signals*, Oxford University Press, New York, 2000.
7. A. J. Viterbi, *Principles of Coherent Communication*, McGraw-Hill, New York, 1966.
8. F. M. Gardner, *Phaselock Techniques*, 3rd ed., Wiley, Hoboken, NJ, 2005.
9. W. C. Lindsey, *Synchronization Systems in Communication and Control*, Prentice-Hall, Englewood Cliffs, NJ, 1972.
10. J. P. Costas, "Synchronous Communication," *Proc. IRE*, vol. 44, pp. 1713–1718, Dec. 1956.

EXERCÍCIOS

4.2-1 Para cada um dos sinais em banda base: (i) $m(t) = \cos 1000\pi t$; (ii) $m(t) = 2\cos 1000\pi t + \text{sen } 2000\pi t$; (iii) $m(t) = \cos 1000\pi t \cos 3000\pi t$, faça o seguinte:

(a) Esboce o espectro de $m(t)$.

(b) Esboce o espectro do sinal DSB-SC $m(t) \cos 10.000\,\pi t$.

(c) Identifique os espectros da banda lateral superior (USB) e da banda lateral inferior (LSB).

(d) Identifique as frequências na banda base e as correspondentes frequências nos espectros DSB-SC, USB e LSB. Explique, em cada caso, a natureza da translação de frequência.

4.2-2 Repita o Exercício 4.2-1 [apenas as partes **(a)**, **(b)** e **(c)**] para: **(i)** $m(t) = \text{sinc}\,(100t)$; **(ii)** $m(t) = e^{-|t|}$; **(iii)** $m(t) = e^{-|t-1|}$. Observe que $e^{-|t-1|}$ é $e^{-|t|}$ atrasado de 1 segundo. Para o último caso, é necessário considerar tanto o espectro de magnitude como o de fase.

4.2-3 Repita o Exercício 4.2-1 [apenas as partes **(a)**, **(b)** e **(c)**] para $m(t) = e^{-|t|}$, para a portadora $\cos(10.000t - \pi/4)$.

Sugestão: Use a Eq. (3.37).

4.2-4 Projete um modulador DSB-SC para gerar um sinal modulado $km(t) \cos(\omega_c t + \theta)$, em que $m(t)$ é um sinal limitado em banda a B Hz. A Fig. E.4.2-4 mostra um modulador DSB-SC disponível no estoque. O gerador de portadora disponível produz $\cos^3 \omega_c t$ em vez de $\cos \omega_c t$. É possível gerar o sinal desejado usando apenas esse equipamento? Você pode empregar qualquer tipo de filtro que achar necessário.

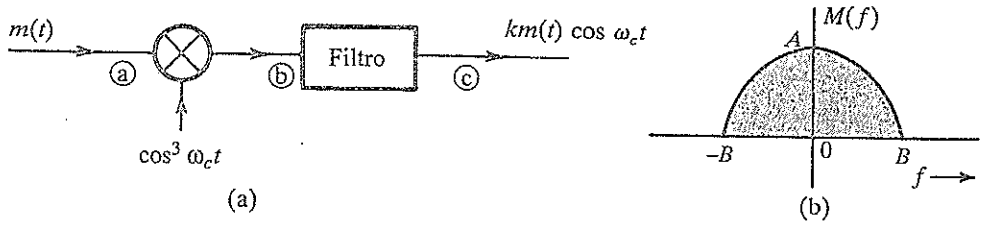
(a) Que tipo de filtro é exigido na Fig. E.4.2-4?

(b) Determine os espectros do sinal nos pontos b e c e indique as bandas de frequências ocupadas por esses espectros.

(c) Qual é o mínimo valor utilizável de ω_c?

(d) Esse esquema funcionaria se a saída do gerador de portadora fosse $\text{sen}^3\, \omega_c t$? Explique.

Figura E.4.2-4



(e) Esse esquema funcionaria se a saída do gerador de portadora fosse $\cos^n \omega_c t$, para qualquer inteiro $n \geq 2$?

4.2-5 Projete um modulador DSB-SC para gerar um sinal modulado $km(t) \cos \omega_c t + \theta)$, com uma frequência portadora $f_c = 300$ kHz ($\omega_c = 2\pi \times 300.000$). O seguinte equipamento está disponível no estoque: **(i)** gerador de sinal com frequência 100 kHz; **(ii)** modulador em anel; **(iii)** filtro passa-faixa sintonizado em 300 kHz.

(a) Mostre como gerar o sinal desejado.

(b) Se a saída do modulador for $k \times m(t) \cos \omega_c t$, determine k.

4.2-6 Moduladores e demoduladores em amplitude também podem ser construídos sem o uso de multiplicadores. Na Fig. E.4.2-6, a entrada $\phi(t) = m(t)$ e a amplitude $A >> |\phi(t)|$. Os dois diodos são idênticos, com uma resistência de r ohms no modo de condução, e resistência infinita no modo de corte. Mostre que a saída $e_0(t)$ é dada por

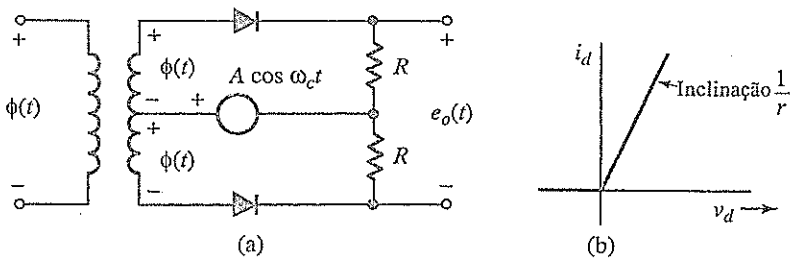
$$e_o(t) = \frac{2R}{R+r} w(t)\, m(t)$$

em que $w(t)$ é o sinal periódico chaveado mostrado na Fig. 2.20a, com período $2\pi/W_c$ segundos.

(a) Mostre que esse circuito pode ser usado como um modulador DSB-SC.

(b) Como você usaria esse circuito como demodulador síncrono para sinais DSB-SC?

Figura E.4.2-6



4.2-7 Na Fig. E.4.2-6, caso $\phi(t) = \text{sen}\,(\omega_c t + \theta)$ e a saída $e_0(t)$ seja aplicada a um filtro passa-baixas, mostre que o circuito pode ser usado como um detector de fase, ou seja, um circuito que mede a diferença de fase entre duas senoides de mesma frequência (ω_c).

Sugestão: Mostre que a saída do filtro é um sinal dc proporcional a sen θ.

4.2-8 Dois sinais $m_1(t)$ e $m_2(t)$, ambos limitados em banda a 5.000 Hz, devem ser transmitidos simultaneamente em um canal, com o esquema de multiplexação ilustrado na Fig. E.4.2-8. No ponto b, o sinal é o sinal multiplexado, que modula uma portadora de frequência 20.000 Hz. No ponto c, o sinal modulado é transmitido no canal.

(a) Esboce os espectros de sinais nos pontos a, b e c.

(b) Qual deve ser a largura de banda do canal?

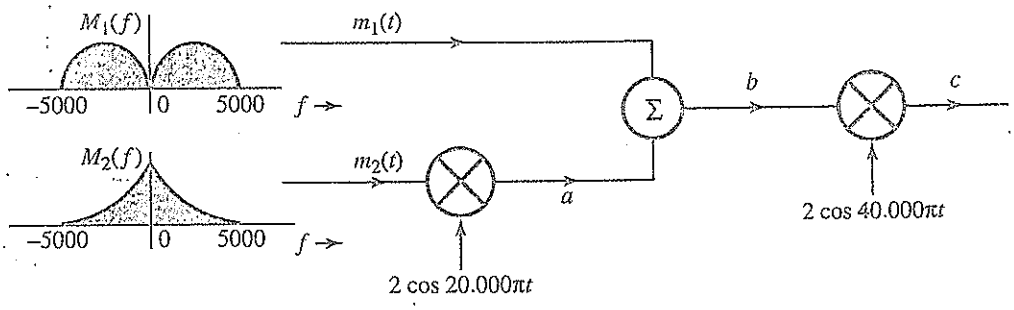
(c) Projete um receptor para recuperação dos sinais $m_1(t)$ e $m_2(t)$ a partir do sinal modulado no ponto c.

4.2-9 O sistema mostrado na Fig. E.4.2-9 é usado para embaralhar sinais de áudio. A saída $y(t)$ é a versão embaralhada da entrada $m(t)$.

(a) Determine o espectro do sinal embaralhado $y(t)$.

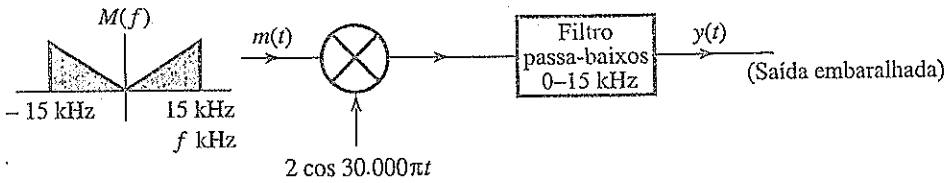
(b) Sugira um método para o desembaralhamento de $y(t)$, para recuperar $m(t)$.

Figura E.4.2-8



Uma versão ligeiramente modificada desse desembaralhador foi usada comercialmente pela primeira vez em um enlace de telefone via rádio de 25 milhas conectando Los Angeles à ilha de Santa Catalina.

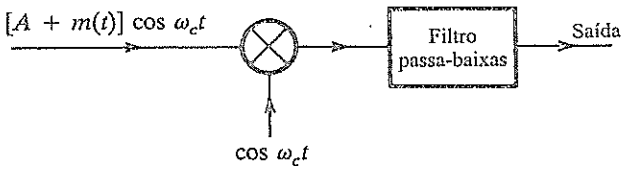
Figura E.4.2-9



4.2-10 Um sinal DSB-SC é dado por $m(t)\cos(2\pi)10^6t$. A frequência portadora desse sinal, 1 MHz, deve ser alterada para 400 kHz. O único equipamento disponível consiste em um modulador em anel, um filtro passa-faixa centrado em 400 kHz e um gerador de onda senoidal, cuja frequência pode ser variada de 150 a 210 kHz. Mostre como o sinal desejado $cm(t)\cos(2\pi \times 400 \times 10^3t)$ pode ser obtido de $m(t)\cos(2\pi)10^6t$. Determine o valor de c.

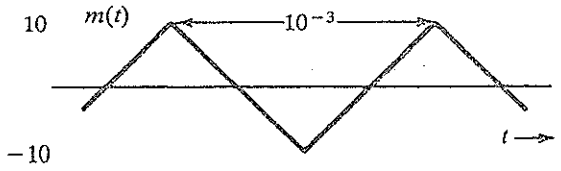
4.3-1 A Fig. E.4.3-1 mostra um esquema para demodulação coerente (síncrona). Mostre que esse esquema pode demodular o sinal AM $[A + m(t)]\cos(2\pi f_ct)$, independentemente do valor de A.

Figura E.4.3-1



4.3-2 Esboce a forma de onda do sinal AM $[A + m(t)]\cos(2\pi f_ct)$, para o sinal triangular periódico $m(t)$ mostrado na Fig. E.4.3-2, com índices de modulação **(a)** $\mu = 0{,}5$; **(b)** $\mu = 1$; **(c)** $\mu = 2$; **(d)** $\mu = \infty$. Como você interpreta o caso $\mu = \infty$?

Figura E.4.3-2



4.3-3 Para o sinal AM com $m(t)$ mostrado na Fig. E.4.3-2 e $\mu = 0{,}8$:

(a) Determine a amplitude e a potência da portadora.

(b) Determine a potência na banda lateral e a eficiência de potência η.

4.3-4 **(a)** Esboce a forma de onda do sinal DSB-SC correspondente ao sinal de mensagem $m(t) = \cos 2\pi t$.

(b) O sinal DSB-SC da parte **(a)** é aplicado à entrada de um detector de envelope. Mostre que a saída do detector de envelope não é $m(t)$, mas $|m(t)|$. Mostre que, em geral, se um sinal AM $[A + m(t)]\cos\omega_ct$ for

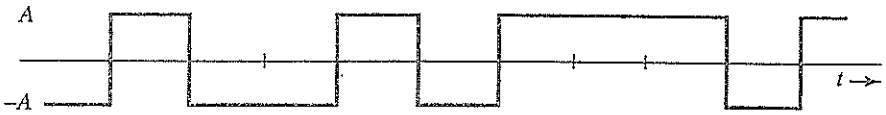
detectado por envelope, a saída é $|A + m(t)|$. Então, mostre que a condição para a recuperação de $m(t)$ com um detector de envelope é $A + m(t) > 0$ para todo t.

4.3-5 Mostre que qualquer esquema que possa ser usado para gerar sinais DSB-SC também gera sinais AM. A recíproca é verdadeira? Justifique.

4.3-6 Mostre que qualquer esquema que possa ser usado para demodular sinais DSB-SC também demodula sinais AM. A recíproca é verdadeira? Justifique.

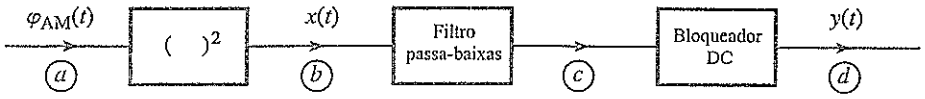
4.3-7 No texto, a eficiência de potência de AM para uma senoide $m(t)$ foi determinada. Efetue uma análise similar para o caso em que $m(t)$ é um sinal binário aleatório, como mostrado na Fig. E.4.3-7, e $\mu = 1$. Esboce a forma de onda do sinal AM com $\mu = 1$. Determine a potência na banda lateral e a potência total (potência do sinal AM), assim como a razão entre elas (eficiência de potência η).

Figura E.4.3-7



4.3-8 Na era inicial do rádio, sinais AM eram demodulados por um detector a cristal seguido por um filtro passa-baixas e um bloqueador dc, como mostrado na Fig. E.4.3-8. Considere que um detector a cristal seja basicamente um dispositivo quadrático. Determine os sinais nos pontos a, b, c e d. Identifique o termo de distorção na saída $y(t)$. Mostre que, se $A >> |m(t)|$, a distorção é pequena.

Figura E.4.3-8



4.4-1 Em um sistema QAM (Fig. 4.19), a portadora gerada localmente tem um erro de frequência $\Delta\omega$ e um erro de fase δ; ou seja, a portadora no receptor é $\cos[(\omega_c + \Delta\omega)t + \delta]$ ou $\text{sen}[(\omega_c + \Delta\omega)t + \delta]$. Mostre que a saída do braço superior do receptor é

$$m_1(t)\cos[(\Delta\omega)t + \delta] - m_2(t)\,\text{sen}[(\Delta\omega)t + \delta]$$

em vez de $m_1(t)$, e que a saída do braço inferior do receptor é

$$m_1(t)\,\text{sen}[(\Delta\omega)t + \delta] + m_2(t)\cos[(\Delta\omega)t + \delta]$$

em vez de $m_2(t)$.

4.4-2 Um sinal modulante $m(t)$ é dado por:

(a) $m(t) = \cos 100\pi t + 2\cos 300\pi t$

(b) $m(t) = \text{sen}\, 100\pi t\, \text{sen}\, 500\pi t$

Em cada caso:

(i) Esboce o espectro de $m(t)$.

(ii) Determine e esboce o espectro do sinal DSB-SC $2m(t)\cos 1000\pi t$.

(iii) Do espectro obtido em (ii), elimine o espectro LSB e obtenha o espectro USB.

(iv) Conhecendo o espectro USB em (ii), escreva a expressão $\varphi_{USB}(t)$ para o sinal USB.

(v) Repita (iii) e (iv) para obter o sinal $\varphi_{LSB}(t)$.

4.4-3 Para os sinais no Exercício 4.4-2, use a Eq. (4.20) e determine as expressões temporais $\varphi_{LSB}(t)$ e $\varphi_{USB}(t)$ com uma frequência portadora $\omega_c = 1000$ rad/s.

Sugestão: Se $m(t)$ for uma senoide, sua transformada de Hilbert $m_h(t)$ será a senoide $m(t)$ com um atraso de fase de $\pi/2$ rad.

4.4-4 Determine $\varphi_{LSB}(t)$ e $\varphi_{USB}(t)$ para o sinal modulante $m(t) = \pi B \operatorname{sinc}^2(2\pi Bt)$, com $B = 2000$ Hz e frequência portadora $f_c = 10.000$ Hz. Faça o seguinte:

(a) Esboce os espectros de $m(t)$ e do correspondente sinal DSB-SC $2m(t) \cos \omega_c t$.

(b) Para determinar o espectro LSB, elimine a USB do espectro DSB-SC obtido na parte (a).

(c) Determine o sinal $\varphi_{LSB}(t)$, que é a transformada de Fourier inversa do espectro LSB determinado na parte (b). Siga o mesmo procedimento e determine $\varphi_{USB}(t)$.

4.4-5 Se $m_h(t)$ for a transformada de Hilbert de $m(t)$, então

(a) Mostre que a transformada de Hilbert de $m_h(t)$ é $-m(t)$.

(b) Mostre que as energias de $m(t)$ e de $m_h(t)$ são idênticas.

4.4-6 Um sinal LSB é demodulado coerentemente, como mostrado na Fig. E.4.4-6. Lamentavelmente, devido ao atraso na transmissão, o sinal de portadora recebido não é $2 \cos \omega_c t$, como foi transmitido, mas $2 \cos [(\omega_c t + \Delta\omega)t + \delta]$. O oscilador local ainda é $\cos \omega_c t$. Mostre que

(a) Quando $\delta = 0$, a saída $y(t)$ é o sinal $m(t)$ com todas as componentes espectrais deslocadas de $\Delta\omega$.

Sugestão: Observe que a saída $y(t)$ é idêntica ao lado direito da Eq. (4.20a), com ω_c substituído por $\Delta\omega$.

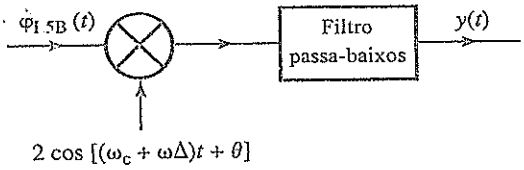
(b) Quando $\Delta\omega = 0$, a saída é o sinal $m(t)$ com as fases de todas as componentes espectrais deslocadas de δ.

Sugestão: Mostre que o espectro de saída $Y(f) = M(f)e^{j\delta}$, para $f \geq 0$, e igual a $M(f)e^{-j\delta}$, quando $f < 0$.

(c) Em cada um desses casos, explique a natureza da distorção.

Sugestão: Para a parte (a), a demodulação consiste no deslocamento de um espectro LSB para a esquerda e para a direita, por $\omega_c + \Delta\omega$ e na filtragem passa-baixas do resultado. Para a parte (b), use a expressão (4.20b) para $\varphi_{LSB}(t)$, multiplique-a pela portadora local $2 \cos (\omega_c t + \delta)$ e aplique o resultado a um filtro passa-baixas.

Figura E.4.4-6

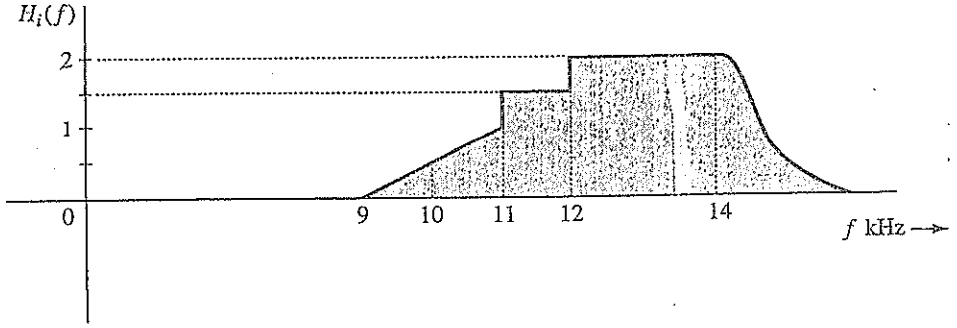


4.4-7 Um sinal USB é gerado com o método de deslocamento de fase (Fig. 4.17). Se a entrada a esse sistema for $m_h(t)$, em vez de $m(t)$, qual será a saída? Esse ainda é um sinal SSB com largura de banda igual à de $m(t)$? Esse sinal pode ser demodulado [para recuperar $m(t)$]? Em caso afirmativo, como?

4.5-1 Um filtro vestigial $H_i(f)$, mostrado no transmissor da Fig. 4.21, tem a função de transferência mostrada na Fig. E.4.5-1. A frequência portadora é $f_c = 10$ kHz e a largura de banda do sinal em banda base é de 4 kHz. Determine a correspondente função de transferência do filtro equalizador $H_o(f)$ mostrado no receptor da Fig. 4.21.

Sugestão: Use a Eq. (4.25).

Figura E.4.5-1



5 MODULAÇÕES E DEMODULAÇÕES EM ÂNGULO

Como discutido no capítulo anterior, uma modulação de portadora pode ser feita modulando a amplitude, a frequência ou a fase de uma **portadora senoidal** de frequência f_c. Nele focamos os vários sistemas de modulação linear em amplitude e as correspondentes demodulações. Agora, discutiremos modulações não lineares em frequência (FM — *frequency modulation*) e em fase (PM — *phase modulation*), também conhecidas como modulações em ângulo.

5.1 MODULAÇÃO NÃO LINEAR

No caso de sinais AM, a amplitude de uma portadora é modulada por um sinal $m(t)$ e, portanto, o conteúdo de informação em $m(t)$ está nas variações de amplitude da portadora. Como vimos, os dois outros parâmetros de uma portadora senoidal — frequência e fase — também podem ser variados proporcionalmente ao sinal de mensagem, na forma de sinais modulados em frequência e em fase, respectivamente. A seguir, descreveremos a essência da modulação em frequência (FM) e da modulação em fase (PM).

Falsa Largada

Na década de 1920, radiodifusão estava em sua infância. Contudo, havia uma busca intensa por técnicas para redução de ruído (estática). Como a potência de ruído é proporcional à largura de banda do sinal modulado (bandas laterais), esforços foram dedicados ao desenvolvimento de um esquema de modulação que reduzisse a largura de banda. Ainda mais importante, a redução da largura de banda também permitiria mais usuários; e havia rumores sobre a descoberta de um novo método que eliminava as bandas laterais (sem banda lateral, sem largura de banda!). A ideia da **modulação em frequência (FM)**, em que a frequência portadora seria variada em proporção com a mensagem $m(t)$, era intrigante. A frequência angular da portadora $\omega(t)$ seria variável no tempo, de modo que $\omega(t) = \omega_c + km(t)$, sendo k uma constante arbitrária. Se a amplitude de pico de $m(t)$ for m_p, os valores máximo e mínimo da frequência portadora serão $\omega_c + km_p$ e $\omega_c - km_p$, respectivamente. Portanto, as componentes espectrais permaneceriam nessa faixa, com uma largura de banda de $2km_p$, centrada em ω_c. O entendimento era que o controle do parâmetro constante k controlava a largura de banda do sinal modulado. Embora isso fosse verdade, havia também a esperança de que o uso de um valor arbitrariamente pequeno para k tornaria a largura de banda de informação arbitrariamente pequena. Essa possibilidade era vista como um passaporte ao paraíso das comunicações. Lamentavelmente, resultados experimentais mostraram que esse raciocínio continha sério erro. A largura de banda de FM, como foi revelado, é sempre maior que (ou, na melhor das hipóteses, igual a) a largura de banda de AM. Em alguns casos, a largura de banda de FM é muitas vezes maior que a de AM. Onde estava a falácia no raciocínio original? Logo descobriremos.

O Conceito de Frequência Instantânea

Enquanto sinais AM transportam a mensagem em suas amplitudes variantes, sinais FM podem variar a frequência instantânea proporcionalmente ao sinal modulante $m(t)$. Isso significa que a frequência portadora é alterada de modo contínuo a cada instante. À primeira vista, isso não faz muito sentido, pois, para definir uma frequência, devemos ter um sinal senoidal com a mesma frequência pelo menos ao longo de um ciclo (ou meio ciclo ou um quarto de ciclo). Esse problema nos faz recordar o primeiro encontro com **velocidade instantânea** em um curso introdutório de mecânica. Até a apresentação de derivadas, por Leibniz e Newton, estávamos acostumados a pensar em velocidade como sendo constante em um intervalo de tempo e éramos incapazes de, ao menos, imaginar que a velocidade pudesse variar a cada instante. Jamais esquecemos o encanto e o deslumbramento causados pela análise de derivadas e velocidade instantânea quando esses conceitos nos foram apresentados. Uma experiência similar aguarda o leitor no que diz respeito à **frequência instantânea**.

Consideremos um sinal senoidal generalizado $\varphi(t)$ dado por

$$\boxed{\varphi(t) = A\cos\theta(t)} \quad \text{FM e PM} \tag{5.1}$$

em que $\theta(t)$ é o **ângulo generalizado**, uma função de t. A Fig. 5.1 mostra um caso hipotético de $\theta(t)$. O ângulo generalizado para uma senoide convencional $A\cos(\omega_c t + \theta_0)$ é uma reta $\omega_c t + \theta_0$, como ilustrado na Fig. 5.1. Um caso geral hipotético de $\theta(t)$ corresponde à tangente ao ângulo $(\omega_c t + \theta_0)$ em algum instante t. O ponto crucial é que, em torno de t, em um pequeno intervalo $\Delta t \to 0$, o sinal $\varphi(t) = A\cos\theta(t)$ e a senoide $A\cos(\omega_c t + \theta_0)$ são idênticos; ou seja,

$$\varphi(t) = A\cos(\omega_c t + \theta_0) \qquad t_1 < t < t_2$$

Podemos afirmar que, nesse pequeno intervalo Δt, a frequência angular de $\varphi(t)$ é ω_c. Como $(\omega_c t + \theta_0)$ tangencia $\theta(t)$, a frequência angular de $\varphi(t)$ é a inclinação de seu ângulo $\theta(t)$ nesse pequeno intervalo. Podemos generalizar esse conceito em **cada instante** de tempo e definir a frequência instantânea ω_i em um instante qualquer de tempo t como a inclinação de $\theta(t)$ em t. Assim, para $\varphi(t)$ na Eq. (5.1), a frequência angular instantânea e o ângulo generalizado se relacionam por

$$\omega_i(t) = \frac{d\theta}{dt} \tag{5.2a}$$

$$\theta(t) = \int_{-\infty}^{t} \omega_i(\alpha)\, d\alpha \tag{5.2b}$$

FM e PM

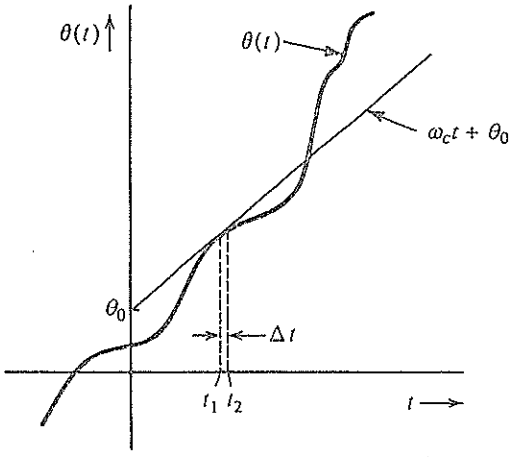


Figura 5.1 Conceito de frequência instantânea.

Agora, podemos ver a possibilidade de transmitir a informação de $m(t)$ variando o ângulo θ de uma portadora. Tais técnicas de modulação, em que o ângulo de uma portadora é variado de alguma forma por um sinal modulante $m(t)$, são conhecidas como **modulação em ângulo** ou **modulação exponencial**. Duas possibilidades simples são a **modulação em fase (PM)** e a **modulação em frequência (FM)**. Em PM, o ângulo $\theta(t)$ varia linearmente com $m(t)$:

$$\theta(t) = \omega_c t + \theta_0 + k_p m(t)$$

em que k_p é uma constante e ω_c, a frequência portadora. Supondo que $\theta_0 = 0$, sem perda de generalidade,

$$\boxed{\theta(t) = \omega_c t + k_p m(t)} \quad \text{PM} \tag{5.3a}$$

A resultante onda PM é

$$\boxed{\varphi_{\text{PM}}(t) = A\cos[\omega_c t + k_p m(t)]} \quad \text{PM} \tag{5.3b}$$

Nesse caso, a frequência angular instantânea $\omega_i(t)$ é dada por

$\Delta\omega(t) = k_p m(t)$

$$\boxed{\omega_i(t) = \frac{d\theta}{dt} = \omega_c + k_p \dot{m}(t)} \quad \text{PM} \quad \Delta f = \frac{1}{2\pi}\left[\Delta\omega(t)\Big|_{\text{MÁX}}\right] = \frac{1}{2\pi} k_p \dot{m}_p \tag{5.3c}$$

Portanto, em PM, a frequência angular instantânea ω_i varia linearmente com a derivada do sinal modulante. Se a frequência angular instantânea ω_i variar linearmente com o sinal modulante, temos FM. Assim, em FM, a frequência angular instantânea ω_i é

$$\boxed{\omega_i(t) = \omega_c + \overbrace{k_f m(t)}^{\Delta\omega(t)}} \quad \text{FM} \quad \Delta f = \frac{1}{2\pi}\left[\Delta\omega(t)\Big|_{\text{MÁX}}\right] = \tag{5.4a}$$

$\Delta f = \frac{1}{2\pi} k_f m_p$

em que k_f é uma constante. O ângulo $\theta(t)$ é, agora, dado por

$$\boxed{\begin{aligned}\theta(t) &= \int_{-\infty}^{t} [\omega_c + k_f m(\alpha)]\, d\alpha \\ &= \omega_c t + k_f \int_{-\infty}^{t} m(\alpha)\, d\alpha\end{aligned}} \quad \text{FM}$$

Aqui, assumimos que o termo constante em $\theta(t)$ é zero, sem perda de generalidade. A onda FM é

$$\boxed{\varphi_{\text{FM}}(t) = A\cos\left[\omega_c t + k_f \int_{-\infty}^{t} m(\alpha)\, d\alpha\right]} \quad \text{FM} \tag{5.5}$$

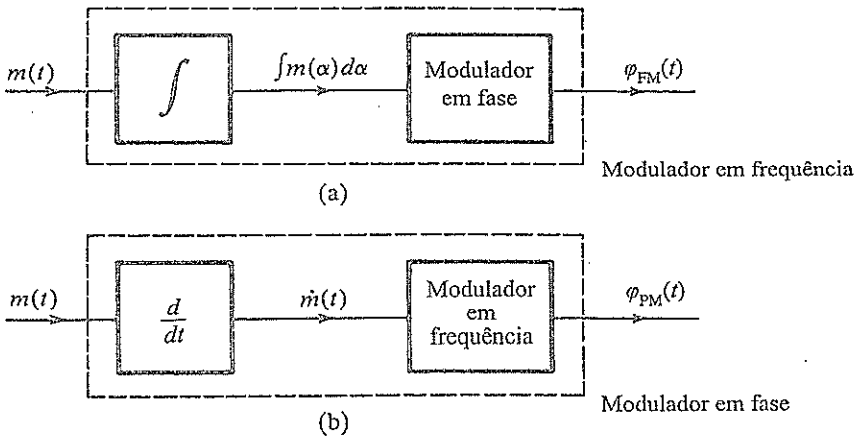


Figura 5.2 Modulações em fase e em frequência são equivalentes e intercambiáveis.

Relação entre FM e PM

Das Eqs. (5.3b) e (5.5), fica aparente que PM e FM não são apenas muito parecidas, mas também inseparáveis. A substituição de $m(t)$ na Eq. (5.3b) por $\int m(\alpha)\, d\alpha$ transforma PM em FM. Assim, um sinal que seja a onda FM correspondente a $m(t)$ também é a onda FM correspondente a $\int m(\alpha)\, d\alpha$ (Fig. 5.2a). Da mesma forma, a onda PM correspondente a $m(t)$ também é a onda FM correspondente a $\dot{m}(t)$ (Fig. 5.2b). Portanto, ao examinar um sinal modulado em ângulo $\varphi(t)$, não há como saber se o mesmo é FM ou PM. Na verdade, não faz sentido perguntar se um final modulado em ângulo é FM ou PM. Isso é análogo a perguntar a um homem casado que tem filhos se ele é um pai ou um filho. Essa discussão e a Fig. 5.2 também mostram que não é necessário discutir separadamente os métodos de geração e demodulação de cada tipo de modulação.

As Eqs. (5.3b) e (5.5) mostram que, tanto em PM como em FM, o ângulo de uma portadora varia proporcionalmente a alguma medida de $m(t)$. Em PM, o ângulo é diretamente proporcional a $m(t)$, enquanto em FM, é proporcional à integral de $m(t)$. Como mostrado na Fig. 5.2b, um modulador em frequência pode ser usado diretamente para gerar um sinal FM ou a mensagem de entrada $m(t)$ pode ser processada por um filtro (diferenciador) com função de transferência $H(s) = s$ para gerar um sinal PM. Contudo, por que limitarmos nosso estudo a esses casos? Há um número infinito de modos de processar $m(t)$ antes de FM. Caso restrinjamos a escolha a um operador linear, uma medida de $m(t)$ pode ser obtida como a saída de um sistema linear invertível (invariante no tempo), com função de transferência $H(s)$ ou resposta ao impulso $h(t)$. A portadora com modulação em ângulo generalizada $\varphi_{EM}(t)$ pode ser expressa como

$$\varphi_{EM}(t) = A\cos[\omega_c t + \psi(t)] \tag{5.6a}$$

$$= A\cos\left[\omega_c t + \int_{-\infty}^{t} m(\alpha)\, h(t-\alpha)\, d\alpha\right] \tag{5.6b}$$

Desde que $H(s)$ seja uma operação reversível (ou invertível), $m(t)$ pode ser recuperado de $\psi(t)$ passando esse sinal modulado por um sistema com função de transferência $[H(s)]^{-1}$, como mostrado na Fig. 5.3. Assim, PM e FM são apenas dois casos especiais com $h(t) = k_p\delta(t)$ e $h(t) = k_p u(t)$, respectivamente.

Isso mostra que, se analisarmos um tipo de modulação em ângulo (como FM), podemos estender os resultados prontamente a qualquer outro tipo. Historicamente, o conceito de modulação em ângulo iniciou com FM, de modo que, neste capítulo, analisaremos primeiro FM, com discussão ocasional de PM. No entanto, isso não significa que FM é superior a outros tipos de modulação em ângulo. Ao contrário, para a maioria dos sinais práticos, PM é superior a FM. Na verdade, o desempenho ótimo não é alcançado nem com PM puro, nem com FM puro, mas com algo entre esses dois tipos de modulação.

Potência de uma Onda Modulada em Ângulo

Embora as frequência e fase instantâneas de uma onda modulada em ângulo possam variar com o tempo, a amplitude A permanece constante. Portanto, a potência de uma onda modulada em ângulo (PM ou FM) sempre é $A^2/2$, independentemente do valor de k_p ou de k_f.

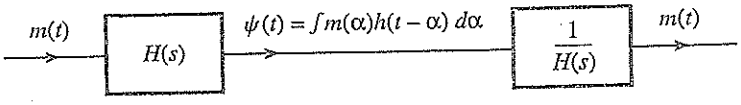


Figura 5.3 Modulação em ângulo generalizada via filtro $H(s)$ e recuperação da mensagem com a passagem do sinal modulado pelo filtro inverso $1/H(s)$.

Exemplo 5.1 Esbocemos as formas de onda FM e PM para o sinal modulante mostrado na Fig. 5.4a. As constantes k_f e k_p são $2\pi \times 10^5$ e 10π, respectivamente, e a frequência portadora f_c é de 100 MHz.

Para FM:

$$\omega_i = \omega_c + k_f m(t)$$

Dividindo tudo por 2π, temos a equação em termos da variável f (frequência em hertz). A frequência instantânea f_i é

$$\begin{aligned} f_i &= f_c + \frac{k_f}{2\pi} m(t) \\ &= 10^8 + 10^5 m(t) \\ (f_i)_{\text{mín}} &= 10^8 + 10^5\,[m(t)]_{\text{mín}} = 99{,}9 \text{ MHz} \\ (f_i)_{\text{máx}} &= 10^8 + 10^5\,[m(t)]_{\text{máx}} = 100{,}1 \text{ MHz} \end{aligned}$$

Como $m(t)$ aumenta e diminui linearmente com o tempo, a frequência instantânea aumenta linearmente de 99,9 a 100,1 MHz em um meio ciclo e cai linearmente de 100,1 a 99,1 MHz no meio ciclo restante do sinal modulante (Fig. 5.4b).

PM para $m(t)$ é FM para $\dot{m}(t)$. Isso resulta da Eq. (5.3c).

Para PM:

$$\begin{aligned} f_i &= f_c + \frac{k_p}{2\pi} \dot{m}(t) \\ &= 10^8 + 5\,\dot{m}(t) \\ (f_i)_{\text{mín}} &= 10^8 + 5\,[\dot{m}(t)]_{\text{mín}} = 10^8 - 10^5 = 99{,}9 \text{ MHz} \\ (f_i)_{\text{máx}} &= 10^8 + 5\,[\dot{m}(t)]_{\text{máx}} = 100{,}1 \text{ MHz} \end{aligned}$$

Como $\dot{m}(t)$ oscila entre os valores de −20.000 a 20.000, a frequência portadora oscila entre 99,9 e 100,1 MHz a cada meio ciclo de $\dot{m}(t)$, como mostrado na Fig. 5.4d.

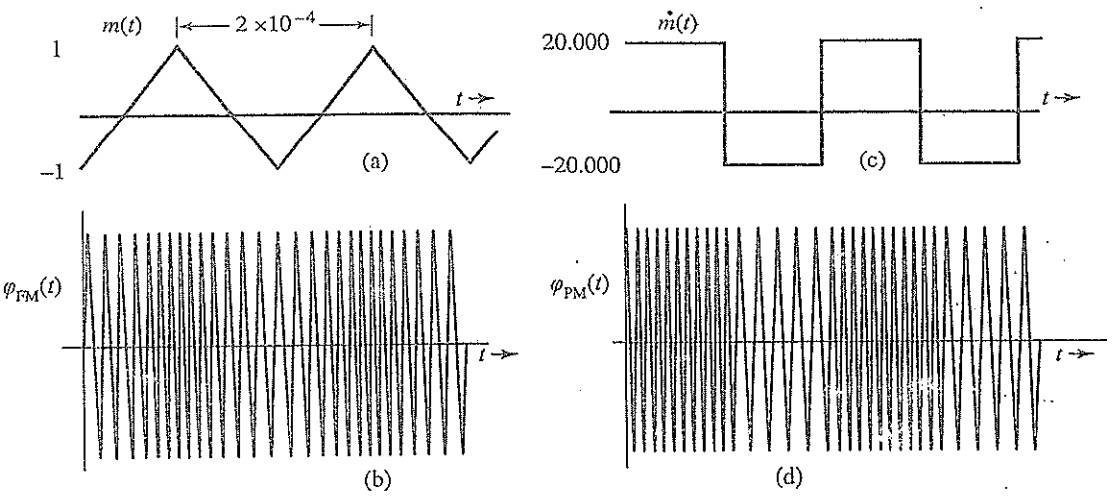


Figura 5.4 Formas de onda FM e PM.

Este método indireto para esboçar a forma de onda PM [usando $\dot{m}(t)$ para modular a portadora em frequência] funciona desde que $m(t)$ seja um sinal contínuo. Caso $m(t)$ seja descontínuo, o sinal PM sofrerá mudanças bruscas de fase e, em consequência, $\dot{m}(t)$ conterá impulsos. Esse método indireto falha em *pontos de descontinuidade*. Nesses casos, uma abordagem direta deve ser usada no ponto de descontinuidade para especificar as mudanças bruscas de fase. Isso é demonstrado no próximo exemplo.

Exemplo 5.2 Esbocemos as formas de onda FM e PM para o sinal modulante digital $m(t)$ mostrado na Fig. 5.5a. As constantes k_f e k_p valem $2\pi \times 10^5$ e $\pi/2$, respectivamente, e $f_c = 100$ MHz.

Para FM:

$$f_i = f_c + \frac{k_f}{2\pi} m(t) = 10^8 + 10^5 m(t)$$

Como $m(t)$ oscila entre -1 e 1, a frequência da onda FM oscila entre 99,9 e 100,1 MHz, como mostrado na Fig. 5.5b. O esquema de modulação de uma portadora em frequência por um sinal digital (Fig. 5.5b) é denominado **chaveamento de frequência** (FSK — *frequency shift keying*), pois os dígitos de informação são transmitidos via chaveamento de diferentes frequências (ver Seção 7.8).

Para PM:

$$f_i = f_c + \frac{k_p}{2\pi} \dot{m}(t) = 10^8 + \frac{1}{4} \dot{m}(t)$$

A derivada $\dot{m}(t)$ (Fig. 5.5c) é zero, exceto em pontos de descontinuidade de $m(t)$, em que impulsos de amplitude ± 2 estão presentes. Isso significa que a frequência do sinal PM permanece a mesma, exceto nesses instantes isolados do tempo! Não fica imediatamente claro como uma frequência instantânea pode sofrer uma alteração de tamanho infinito e retornar ao valor original em tempo zero. Consideremos a abordagem direta:

$$\begin{aligned} \varphi_{PM}(t) &= A \cos [\omega_c t + k_p m(t)] \\ &= A \cos \left[\omega_c t + \frac{\pi}{2} m(t)\right] \\ &= \begin{cases} A \operatorname{sen} \omega_c t & \text{quando } m(t) = -1 \\ -A \operatorname{sen} \omega_c t & \text{quando } m(t) = 1 \end{cases} \end{aligned}$$

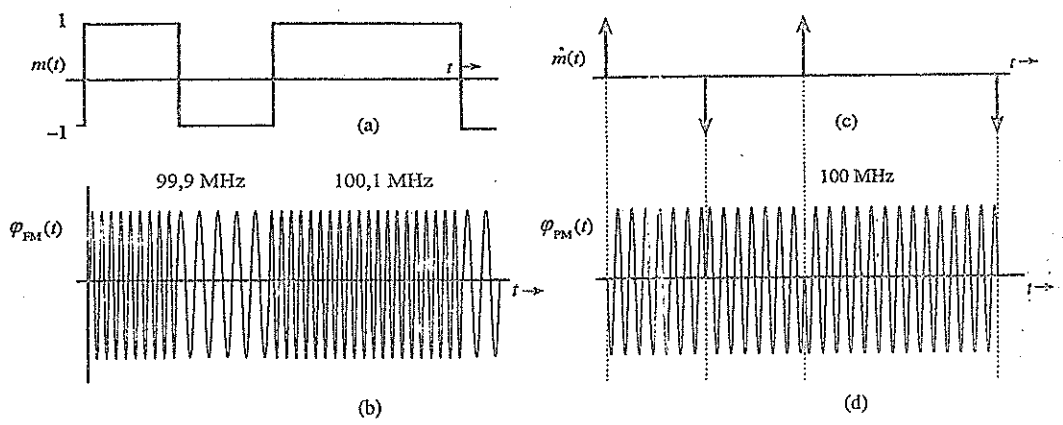


Figura 5.5 Formas de onda FM e PM.

Esta onda PM é mostrada na Fig. 5.5d. Tal esquema de modulação em fase de uma portadora por um sinal digital é denominado **chaveamento de fase** (PSK – *phase shift keying*), pois os dígitos de informação são transmitidos via chaveamento da fase da portadora. Notemos que PSK também pode ser visto como uma modulação DSB-SC por $m(t)$.

A onda PM $\varphi_{PM}(t)$, neste caso, apresenta descontinuidade de fase em instantes nos quais estão localizados impulsos de $\dot{m}(t)$. Nesses instantes, a fase da portadora sofre uma alteração de π instantaneamente. Um deslocamento de fase finito em tempo zero implica frequência instantânea infinita nesses instantes. Isso está em acordo com nossa observação a respeito de $\dot{m}(t)$.

A descontinuidade de fase de $\varphi_{PM}(t)$ no instante em que $m(t)$ é descontínuo é $k_p m_d$, em que m_d é o valor da descontinuidade de $m(t)$ no dado instante. Neste exemplo, a amplitude de $m(t)$ sofre uma descontinuidade de 2 (de -1 a 1). Logo, a descontinuidade de fase de $\varphi_{PM}(t)$ é $k_p m_d = (\pi/2) \times 2 = \pi$ rad, que confirma o resultado obtido anteriormente.

Quando $\dot{m}(t)$ é um sinal digital (como na Fig. 5.5a), $\varphi_{PM}(t)$ apresenta uma descontinuidade de fase nos pontos em que $m(t)$ tem uma descontinuidade degrau. A seguir, mostraremos que, para evitar ambiguidade na demodulação nesses casos, o desvio de fase $k_p m(t)$ deve ficar restrito ao intervalo $(-\pi, \pi)$. Por exemplo, se k_p fosse $3\pi/2$, então

$$\varphi_{PM}(t) = A \cos\left[\omega_c t + \frac{3\pi}{2} m(t)\right]$$

Neste caso, $\varphi_{PM}(t) = A$ sen $\omega_c t$ quando $m(t) = 1$ ou $-1/3$. Isso certamente causará ambiguidade no receptor quando A sen $\omega_c t$ for recebido. Especificamente, o receptor não é capaz de determinar o valor exato de $m(t)$. Tal ambiguidade jamais ocorre se $k_p m(t)$ ficar restrito ao intervalo $(-\pi, \pi)$.

O que causa essa ambiguidade? Quando $m(t)$ tem uma descontinuidade degrau, a fase de $\varphi_{PM}(t)$ muda instantaneamente. Como a fase $\varphi_0 + 2n\pi$ é indistinguível da fase φ_0, ambiguidades serão inerentes ao demodulador, a menos que as variações de fase fiquem limitadas ao intervalo $(-\pi, \pi)$. Isso significa que k_p deve ser suficientemente pequeno para restringir a mudança de fase $k_p m(t)$ ao intervalo $(-\pi, \pi)$.

Nenhuma restrição é imposta ao valor de k_p se $m(t)$ for contínuo. Nesse caso, a mudança de fase com o tempo deixa de ser instantânea e passa a ser gradual; uma fase $\varphi_0 + 2n\pi$ exibirá n ciclos adicionais da portadora, para uma fase de apenas φ_0. A onda PM pode ser detectada com um demodulador FM seguido por um integrador (ver Exercício 5.4-1). Os n ciclos adicionais serão detectados pelo demodulador FM e a subsequente integração produzirá uma fase $2n\pi$. Assim, as fases φ_0 e $\varphi_0 + 2n\pi$ podem ser detectadas sem ambiguidade. Essa conclusão pode ser comprovada a partir do Exemplo 5.1, em que a máxima mudança de fase é $\Delta\varphi = 10\pi$.

Como um sinal limitado em banda não pode ter descontinuidades do tipo degrau, também podemos dizer que, quando $m(t)$ for limitado em banda, nenhuma restrição é imposta ao valor de k_p.

5.2 LARGURA DE BANDA DE ONDAS MODULADAS EM ÂNGULO

Diferentemente de AM, a modulação em ângulo é não linear, de modo que nenhuma propriedade da transformada de Fourier pode ser aplicada diretamente para efeitos de análise de largura de banda. Para determinar a largura de banda de uma onda FM, definamos

$$a(t) = \int_{-\infty}^{t} m(\alpha)\, d\alpha \tag{5.7}$$

e

$$\hat{\varphi}_{FM}(t) = A\, e^{j[\omega_c t + k_f a(t)]} = A e^{jk_f a(t)} e^{j\omega_c t} \tag{5.8a}$$

de modo que a relação para sinais FM é

$$\varphi_{\text{FM}}(t) = \text{Re}\left[\hat{\varphi}_{\text{FM}}(t)\right] \tag{5.8b}$$

Expandindo a exponencial $e^{jk_f a(t)}$ na Eq. (5.8a) em uma série de potências, obtemos

$$\hat{\varphi}_{\text{FM}}(t) = A\left[1 + jk_f a(t) - \frac{k_f^2}{2!}a^2(t) + \cdots + j^n\frac{k_f^n}{n!}a^n(t) + \cdots\right]e^{j\omega_c t} \tag{5.9a}$$

e

$$\begin{aligned}\varphi_{\text{FM}}(t) &= \text{Re}\left[\hat{\varphi}_{\text{FM}}(t)\right] \\ &= A\left[\cos\omega_c t - k_f a(t)\,\text{sen}\,\omega_c t - \frac{k_f^2}{2!}a^2(t)\cos\omega_c t + \frac{k_f^3}{3!}a^3(t)\,\text{sen}\,\omega_c t + \cdots\right]\end{aligned} \tag{5.9b}$$

A onda modulada consiste em uma portadora não modulada mais vários termos modulados em amplitude, como $a(t)$ sen $\omega_c t$, $a^2(t)$ cos $\omega_c t$, $a^3(t)$ sen $\omega_c t$, ... O sinal $a(t)$ é uma integral de $m(t)$. Se $M(f)$ for limitado em banda a B, $A(f)$ também será limitado em banda* a B. O espectro de $a^2(t)$ é simplesmente $A(f) * A(f)$ e limitado em banda a $2B$. De modo similar, o espectro de $a^n(t)$ é limitado em banda a nB. Portanto, o espectro consiste em uma portadora não modulada mais os espectros de $a(t)$, $a^2(t)$,..., $a^n(t)$, centrados em ω_c. Fica claro que a onda modulada não é limitada em banda: tem largura de banda infinita e não guarda uma relação simples com o espectro do sinal modulante, como no caso de AM.

Embora a largura de banda de uma onda FM seja, em teoria, infinita, para sinais práticos, com $|a(t)|$ limitado, $|k_f a(t)|$ permanecerá finito. Como $n!$ aumenta muito mais rápido que $|k_f a(t)|^n$, temos

$$\frac{k_f^n a^n(t)}{n!} \simeq 0 \qquad \text{para } n \text{ grande}$$

Portanto, podemos concluir que a maior parte da potência do sinal modulado reside em uma largura de banda finita. Esse é o principal fundamento da análise de largura de banda de modulações em ângulo. Há duas possibilidades em termos de largura de banda: FM de banda estreita e FM de banda larga.

Aproximação para Modulação em Ângulo de Banda Estreita

Em contraste com AM, modulações em ângulo são não lineares. A relação não linear entre $a(t)$ e $\varphi(t)$ fica evidente dos termos que envolvem $a^n(t)$ na Eq. (5.9b). Quando k_f é muito pequeno, tal que

$$|k_f a(t)| \ll 1$$

todos os termos de ordens superiores na Eq. (5.9b) são desprezíveis, exceto os dois primeiros. Assim, temos uma boa aproximação

$$\varphi_{\text{FM}}(t) \approx A\left[\cos\omega_c t - k_f a(t)\,\text{sen}\,\omega_c t\right] \tag{5.10}$$

Essa aproximação é uma modulação linear cuja expressão é similar à de um sinal AM com sinal de mensagem $a(t)$. Como a largura de banda de $a(t)$ é B Hz, a largura de banda de $\varphi_{\text{FM}}(t)$ na Eq. (5.10) é $2B$ Hz, segundo a propriedade de translação em frequência, devido ao termo $a(t)$ sen $\omega_c t$. Por essa razão, o sinal FM correspondente ao caso $|k_f a(t)| << 1$ é denominado **FM de banda estreita** (NBFM

*Isso se deve ao fato de que integração é uma operação linear equivalente a passar o sinal por uma função de transferência $1/j2\pi f$. Assim, se $M(f)$ for limitado em banda a B, $A(f)$ também deve ser limitado em banda a B.

– *narrowband FM*). De modo similar, o sinal **PM de banda estreita** (**NBPM** – *narrowband PM*) é aproximado por

$$\varphi_{PM}(t) \approx A\left[\cos \omega_c t - k_p m(t) \operatorname{sen} \omega_c t\right] \tag{5.11}$$

NBPM também tem uma largura de banda aproximada de $2B$.

Uma comparação entre NBFM [Eq. (5.10)] e AM [Eq. (5.9a)] ressalta as semelhanças e diferenças entre os dois tipos de modulação. Os dois têm a mesma largura de banda modulada $2B$. O espectro da banda lateral apresenta, no caso de FM, um deslocamento de fase de $\pi/2$ em relação à portadora e, no caso de AM, está em fase com a portadora. Devemos, no entanto, recordar que, apesar das aparentes semelhanças, os sinais AM e FM têm formas de onda muito distintas. Em um sinal AM, a frequência de oscilação é constante e a amplitude varia no tempo; em um sinal FM, a amplitude permanece constante e a frequência varia no tempo.

Análise de FM de Banda Larga: A Falácia Revelada

Um sinal de FM faz sentido apenas caso seu desvio de frequência seja suficientemente grande. Em outras palavras, em aplicações práticas de FM, o valor escolhido para a constante k_f é grande o bastante para que a condição $|k_f a(t)| \ll 1$ não seja atendida. Nesses casos, nos referimos aos sinais FM como **FM de banda larga** (**WBFM** – *wideband FM*). Assim, na análise de largura de banda de WBFM, não podemos ignorar todos os termos de ordens superiores na Eq. (5.9b). Inicialmente, escolheremos a rota dos pioneiros que, com seu raciocínio intuitivamente simples, falharam na estimativa da largura de banda de FM. Se descobrirmos a falácia em seu raciocínio, teremos uma chance para obter uma melhor estimativa da largura de FM (de banda larga).

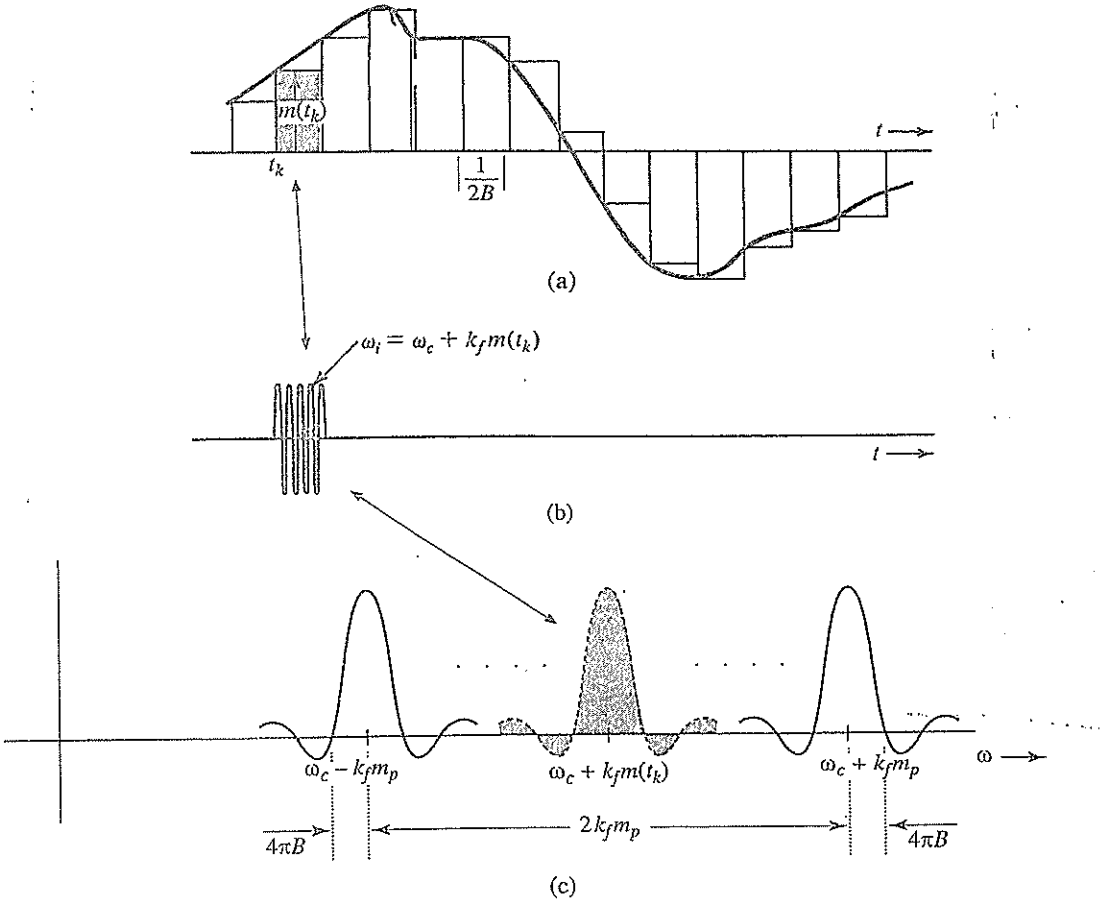


Figura 5.6 Estimativa da largura de banda de uma onda FM.

Consideremos uma mensagem $m(t)$ passa-baixos, com largura de banda de B Hz. Esse sinal pode ser bem aproximado por um sinal em escada $\hat{m}(t)$, como mostrado na Fig. 5.6a. O sinal $m(t)$ é, agora, aproximado por pulsos de amplitude constante. Por conveniência, cada um desses pulsos será denominado uma "célula". Para assegurar que $\hat{m}(t)$ contenha toda a informação de $m(t)$, a largura da célula em $\hat{m}(t)$ não deve ser maior que o intervalo de Nyquist $1/2B$ segundo, de acordo com o teorema da amostragem (Capítulo 6).

A análise do sinal FM correspondente a $\hat{m}(t)$ é relativamente mais simples, por conta dos pulsos (células) de amplitude constante e largura $T = 1/2B$ segundos. Consideremos uma célula típica, com início em $t = t_k$. Essa célula tem amplitude constante $m(t_k)$. Assim o sinal FM a ela correspondente é uma senoide de frequência $\omega_c + k_f m(t_k)$ e duração $T = 1/2B$, como mostrado na Fig. 5.6b. O sinal FM para $\hat{m}(t)$ consiste em uma sequência de pulsos senoidais como este, de frequência constante e duração $T = 1/2B$, correspondentes às várias células de $\hat{m}(t)$. O espectro FM para $\hat{m}(t)$ é a soma das transformadas de Fourier dos pulsos senoidais associados às células. A transformada de Fourier do pulso senoidal na Fig. 5.6b (que corresponde à k-ésima célula) é uma função sinc, mostrada hachurada na Fig. 5.6c, dada pela Eq. (3.27a), com $\tau = 1/2B$, e Eq. (3.26), com $f_0 = f_c + k_f m(t_k)/2\pi$:

$$\text{rect}(2Bt)\cos[\omega_c t + k_f m(t_k)t] \Longleftrightarrow \frac{1}{2}\text{sinc}\left[\frac{\omega + \omega_c + k_f m(t_k)}{4B}\right] + \frac{1}{2}\text{sinc}\left[\frac{\omega - \omega_c - k_f m(t_k)}{4B}\right]$$

Notemos que o espectro desse pulso se espalha nos dois lados da frequência central $\omega_c + k_f m(t_k)$ por $4\pi B$ como o lóbulo principal da função sinc. A Fig. 5.6c mostra os espectros de pulsos senoidais correspondentes a várias células. As amplitudes mínima e máxima das células são $-m_p$ e m_p, respectivamente. Portanto, as frequências *centrais* mínima e máxima dos pulsos senoidais curtos correspondentes ao sinal FM para todas as células são $\omega_c - k_f m_p$ e $\omega_c + k_f m_p$, respectivamente. Consideremos o lóbulo principal da função sinc dessas respostas de frequências como contribuição significativa à largura da banda FM, como mostrado na Fig. 5.6c. Assim, os valores máximo e mínimo das frequências significativas nesse espectro são $\omega_c + k_f m_p + 4\pi B$ e $\omega_c - k_f m_p - 4\pi B$, respectivamente. A largura de banda do espectro FM é, aproximadamente,

$$B_{\text{FM}} = \frac{1}{2\pi}(2k_f m_p + 8\pi B) = 2\left(\frac{k_f m_p}{2\pi} + 2B\right) \text{ Hz}$$

Agora podemos entender a falácia no raciocínio dos pioneiros. As frequências portadoras máxima e mínima são $\omega_c + k_f m_p$ e $\omega_c - k_f m_p$, respectivamente. Por isso, foi concluído que as componentes espectrais também devem estar contidas nesse intervalo, resultando em uma largura da banda de FM igual a $2k_f m_p$. A hipótese implícita é que uma senoide de frequência ω tem todo seu espectro concentrado em ω. Infelizmente, isso é verdadeiro somente no caso da senoide eterna com duração $T = \infty$ (o que transforma a função sinc em um impulso). Para uma senoide de duração finita, com duração de T segundos, o espectro se espalha como a função sinc nos dois lados de ω por, pelo menos, a largura do lóbulo principal, que é $2\pi/T$. Os pioneiros deixaram escapar esse efeito de espalhamento.

Por conveniência de notação, dado o desvio da frequência portadora (em radianos por segundos) de $\pm k_f m_p$, denotaremos o *pico do desvio de frequência* em hertz por Δf. Assim,

$$\Delta f = k_f \frac{m_{\text{máx}} - m_{\text{mín}}}{2 \cdot 2\pi} = f = k_f \frac{m_p}{2\pi}$$

A largura de banda de FM estimada (em hertz) pode, então, ser expressa como

$$B_{\text{FM}} \simeq 2(\Delta f + 2B) \tag{5.12}$$

A estimativa de largura de banda agora obtida é um pouco maior que o valor real, pois corresponde à aproximação de $m(t)$ em degraus, e não ao verdadeiro sinal $m(t)$, que é consideravelmente mais suave. Portanto, a real largura de banda de FM é menor que esse valor estimado. Com base na Fig. 5.6c, fica claro que uma melhor aproximação para a largura de banda FM fica entre

$$[2\Delta f,\ 2\Delta f + 4B]$$

Portanto, devemos reajustar nossa estimativa de largura de banda. Para efetuar essa correção, observemos que, para o caso NBFM, k_f é muito pequeno. Dessa forma, em NBFM, dado um valor fixo m_p, Δf é muito pequeno (em comparação com B). Nesse caso, ignoramos o pequeno termo Δf na Eq. (5.12), e obtemos

$$B_{\text{FM}} \approx 4B$$

Contudo, mostramos anteriormente que, no caso de banda estreita, a largura de banda FM é aproximadamente $2B$ Hz. Isso indica que uma melhor estimativa para a largura de banda é

$$B_{\text{FM}} = 2(\Delta f + B) = 2\left(\frac{k_f m_p}{2\pi} + B\right) \tag{5.13}$$

Este é exatamente o resultado obtido por Carson,[1] que investigou esse problema de forma rigorosa para modulação por tom [senoide $m(t)$]. Na literatura, esta fórmula recebe o nome de **regra de Carson**. Observemos que, para um caso verdadeiramente de banda larga, onde $\Delta f >> B$, a Eq. (5.13) pode ser aproximada por

$$B_{\text{FM}} \approx 2\Delta f \qquad \Delta f \gg B \tag{5.14}$$

Como $\Delta\omega = k_f m_p$, essa fórmula é precisamente a que os pioneiros usaram para a largura de banda FM. O único erro estava em pensar que essa fórmula seria válida para todos os casos, especialmente para o de banda estreita, em que $\Delta f << B$.

Definamos uma razão de desvio β como

$$\beta = \frac{\Delta f}{B} \tag{5.15}$$

A regra de Carson pode ser expressa em termos da razão de desvio como

$$B_{\text{FM}} = 2B(\beta + 1) \tag{5.16}$$

A razão de desvio controla o grau de modulação e, em consequência, tem um papel semelhante ao do índice de modulação em AM. De fato, para o caso especial de FM modulado por tom, a razão de desvio β é denominado **índice de modulação.**

Modulação em Fase

Todos os resultados deduzidos para FM se aplicam diretamente a PM. Assim, no caso de PM, a frequência instantânea é dada por

$$\omega_i = \omega_c + k_p \dot{m}(t)$$

Portanto, o pico de desvio de frequência Δf é

$$\Delta f = k_p \frac{[\dot{m}(t)]_{\text{máx}} - [\dot{m}(t)_{\text{mín}}]}{2 \cdot 2\pi} \tag{5.17a}$$

Supondo que

$$\dot{m}_p = [\dot{m}(t)]_{\text{máx}} = -|\dot{m}(t)_{\text{mín}}| \tag{5.17b}$$

então

$$\Delta f = k_p \frac{\dot{m}_p}{2\pi} \tag{5.17c}$$

Portanto,*

$$B_{PM} = 2(\Delta f + B) \tag{5.18a}$$

$$= 2\left(\frac{k_p \dot{m}_p}{2\pi} + B\right) \tag{5.18b}$$

Um aspecto muito interessante de FM é o fato de $\Delta\omega = k_f m_p$ depender somente do valor de pico de $m(t)$; $\Delta\omega$ é independente do espectro de $m(t)$. Em PM, por sua vez, $\Delta\omega = k_p \dot{m}_p$ depende do valor de pico de $\dot{m}(t)$. Contudo, $\dot{m}(t)$ tem forte dependência da composição espectral de $m(t)$. A presença de componentes de altas frequências em $m(t)$ implica rápidas variações temporais, resultando em um maior valor para $\dot{m}_p$. Reciprocamente, uma predominância de componentes de frequências baixas resultará em um menor valor para $\dot{m}_p$. Dessa forma, embora a largura de banda de sinais FM [Eq. (5.13)] seja praticamente independente da forma espectral de $m(t)$, a largura de banda de sinais PM [Eq. (5.18)] é fortemente afetada pela forma espectral de $m(t)$. Quando o espectro de $m(t)$ está concentrado nas frequências baixas, B_{PM} será menor que quanto o espectro de $m(t)$ estiver concentrado nas frequências altas.

Análise Espectral da Modulação em Frequência por Tom

Para uma portadora FM com um sinal de mensagem genérico $m(t)$, a análise espectral requer o uso da aproximação do sinal por degraus. Modulação por tom é um caso especial, para o qual uma análise espectral exata é possível: $m(t)$ é uma senoide. Usemos este caso especial para comprovar a aproximação de largura de banda FM. Seja

$$m(t) = \alpha \cos \omega_m t$$

Da Eq. (5.7), com a hipótese que, inicialmente, $a(-\infty) = 0$, temos

$$a(t) = \frac{\alpha}{\omega_m} \operatorname{sen} \omega_m t$$

Assim, da Eq. (5.8a), temos

$$\hat{\varphi}_{FM}(t) = Ae^{(j\omega_c t + k_f \alpha/\omega_m \operatorname{sen} \omega_m t)}$$

Além disso,

$$\Delta\omega = k_f m_p = \alpha k_f$$

e a largura de banda de $m(t)$ é $2\pi B = \omega_m$ rad/s. A razão de desvio (ou, neste caso, o índice de modulação) é

$$\beta = \frac{\Delta f}{B} = \frac{\Delta\omega}{2\pi B} = \frac{\alpha k_f}{\omega_m}$$

Logo,

$$\hat{\varphi}_{FM}(t) = Ae^{j(\omega_c t + \beta \operatorname{sen} \omega_m t)}$$

$$= Ae^{j\omega_c t}(e^{j\beta \operatorname{sen} \omega_m t}) \tag{5.19}$$

*A Eq. (5.17a) pode ser aplicada somente se $m(t)$ for uma função contínua do tempo. Caso $m(t)$ tenha descontinuidades, sua derivada não existe. Nessa situação, devemos usar a abordagem direta (discutida no Exemplo 5.2) para determinar $\varphi_{PM}(t)$ e, então, determinamos $\Delta\omega$ de $\varphi_{PM}(t)$.

Notemos que $e^{j\beta \operatorname{sen} \omega_m t}$ é um sinal periódico, com período $2\pi/\omega_m$, e pode ser expandido em uma série de Fourier exponencial como

$$e^{j\beta \operatorname{sen} \omega_m t} = \sum_{n=-\infty}^{\infty} D_n e^{j\omega_m t}$$

em que

$$D_n = \frac{\omega_m}{2\pi} \int_{-\pi/\omega_m}^{\pi/\omega_m} e^{j\beta \operatorname{sen} \omega_m t} e^{-jn\omega_m t}\, dt = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(\beta \operatorname{sen} x - nx)}\, dx$$

A integral no lado direito não pode ser calculada em uma forma fechada, e deve ser integrada por expansão em uma série infinita. Essa integral é tabelada e denotada por $J_n(\beta)$, a função de Bessel de primeira espécie e ordem n. A Fig. 5.7a mostra gráficos dessa função em função de n e para diversos valores de β. Assim,

$$e^{j\beta \operatorname{sen} \omega_m t} = \sum_{n=-\infty}^{\infty} J_n(\beta) e^{jn\omega_m t} \tag{5.20}$$

Substituindo a Eq. (5.20) na Eq. (5.19), obtemos

$$\hat{\varphi}_{\mathrm{FM}}(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) e^{j(\omega_c t + n\omega_m t)}$$

e

$$\varphi_{\mathrm{FM}}(t) = A \sum_{n=-\infty}^{\infty} J_n(\beta) \cos(\omega_c + n\omega_m)t$$

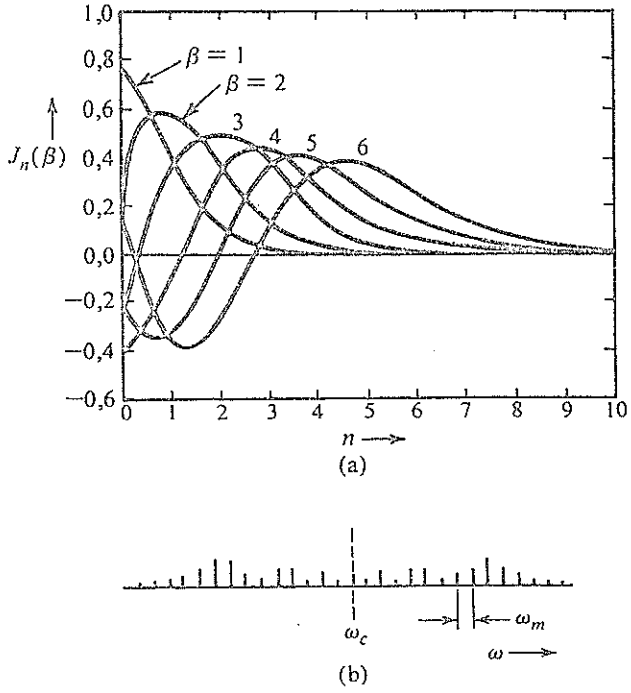


Figura 5.7 (a) Variação de $J_n(\beta)$ em função de n, para diversos valores de β, (b) Espectro de onda FM modulada por tom.

O sinal FM modulado por tom tem uma componente de portadora e um número infinito de bandas laterais de frequências $\omega_c \pm \omega_m$, $\omega_c \pm 2\omega_m$, ..., $\omega_c \pm n\omega_m$, ..., como mostrado na Fig. 5.7b. Esse resultado contrasta fortemente com o obtido para o espectro DSB-SC, que tem apenas uma banda lateral de cada lado da frequência portadora. A magnitude da n-ésima banda lateral em $\omega_c + n\omega_m$ é* $J_n(\beta)$. Dos gráficos de $J_n(\beta)$ na Fig. 5.7a, podemos verificar que, para um dado β, $J_n(\beta)$ decresce com n e que há um número finito de linhas espectrais significativas nas bandas laterais. Podemos ver na Fig. 5.7a que $J_n(\beta)$ é desprezível para $n > \beta + 1$. Portanto, o número de impulsos significativos nas bandas laterais é $\beta + 1$. A largura de banda da portadora FM é dada por

$$\begin{aligned} B_{\text{FM}} &= 2(\beta + 1)f_m \\ &= 2(\Delta f + B) \end{aligned}$$

o que está em acordo com o resultado obtido anteriormente [Eq. (5.13)]. Quando $\beta << 1$ (NBFM), existe somente uma banda lateral significativa e a largura de banda é $B_{\text{FM}} = 2f_m = 2B$. É importante notar que a análise desse caso de modulação por tom é uma comprovação, não uma prova, da fórmula de Carson.

Na literatura, modulação por tom em FM é, em geral, discutida em detalhe. Como, no entanto, modulação em ângulo é uma modulação não linear, os resultados deduzidos para modulação por tom podem ter pouca relação com situações práticas. Na verdade, na melhor das hipóteses, esses resultados não fazem sentido e, na pior, podem ser enganosos se generalizados a sinais práticos.** Como autores e professores, parece-nos que dar demasiada ênfase à modulação por tom pode ser enganador. Por isso, aqui, omitimos discussão adicional sobre o assunto.

O método para determinar o espectro de uma onda FM modulada por tom pode ser usado na determinação do espectro de uma onda FM para o caso em que $m(t)$ é um sinal periódico genérico:

$$\hat{\varphi}_{\text{FM}}(t) = Ae^{j\omega_c t}[e^{jk_f a(t)}]$$

Como $a(t)$ é um sinal periódico, $e^{jk_f a(t)}$ também é um sinal periódico, que pode ser expresso em uma série de Fourier exponencial, como na equação anterior. Feito isso, é relativamente simples escrever $\varphi_{\text{FM}}(t)$ em termos da portadora e das bandas laterais.

Exemplo 5.3 (a) Estimemos B_{FM} e B_{PM} para o sinal modulante $m(t)$ na Fig. 5.4a, para $k_f = 2\pi \times 10^5$ e $k_p = 5\pi$. Suponhamos que a largura de banda essencial do sinal periódico $m(t)$ seja igual à frequência de seu terceiro harmônico.

(b) Repitamos o exercício para o caso em que a amplitude de $m(t)$ é dobrada [$m(t)$ multiplicado por 2].

(a) A amplitude de pico de $m(t)$ é unitária. Assim, $m_p = 1$. A seguir, determinemos a largura de banda essencial B de $m(t)$. Deixamos como exercício mostrar que a série de Fourier para esse sinal periódico é dada por

$$m(t) = \sum_n C_n \cos n\omega_0 t \qquad \omega_0 = \frac{2\pi}{2 \times 10^{-4}} = 10^4\pi$$

em que

$$C_n = \begin{cases} \dfrac{8}{\pi^2 n^2} & n \text{ ímpar} \\ 0 & n \text{ par} \end{cases}$$

---

*Como $J_{-n}(\beta) = (-1)^n J_n(\beta)$, a magnitude da LSB em $\omega = \omega_c - n\omega_m$ é igual à da USB em $\omega = \omega_c + n\omega_m$.

**Por exemplo, com base na análise da modulação por tom, é comum afirmar que FM é superior a PM por um fator de 3, no que diz respeito à SNR de saída. Na verdade, isso é falso para a maioria dos sinais encontrados na prática.

Podemos observar que as amplitudes dos harmônicos decaem rapidamente com n. O terceiro harmônico é apenas 11% da fundamental e o quinto harmônico, 4%. Isso significa que as potências do terceiro e do quinto harmônicos correspondem a 1,21% e 0,16% da potência da componente fundamental, respectivamente. Assim, podemos tomar a largura de banda essencial de $m(t)$ como a frequência de seu terceiro harmônico, ou seja,

$$B = 3 \times \frac{10^4}{2} = 15\,\text{kHz}$$

Para FM

$$\Delta f = \frac{1}{2\pi} k_f m_p = \frac{1}{2\pi}(2\pi \times 10^5)(1) = 100$$

e

$$B_{\text{FM}} = 2(\Delta f + B) = 230\,\text{kHz}$$

A razão de desvio β é dado por

$$\beta = \frac{\Delta f}{B} = \frac{100}{15}$$

e

$$B_{\text{FM}} = 2B(\beta + 1) = 30\left(\frac{100}{15} + 1\right) = 230\,\text{kHz}$$

Para PM: A amplitude de pico de $\dot{m}(t)$ é 20.000 e

$$\Delta f = \frac{1}{2\pi} k_p \dot{m}_p = 50\ \text{kHz}$$

Logo,

$$B_{\text{PM}} = 2(\Delta f + B) = 130\,\text{kHz}$$

A razão de desvio β é dado por

$$\beta = \frac{\Delta f}{B} = \frac{50}{15}$$

e

$$B_{\text{PM}} = 2B(\beta + 1) = 30\left(\frac{50}{15} + 1\right) = 130\ \text{kHz}.$$

(b) Quando multiplicamos $m(t)$ por 2, dobramos seu valor de pico. Ou seja, $m_p = 2$. Contudo, a largura de banda não é alterada, de modo que $B = 15$ kHz.
Para FM:

$$\Delta f = \frac{1}{2\pi} k_f m_p = \frac{1}{2\pi}(2\pi \times 10^5)(2) = 200\,\text{kHz}$$

e

$$B_{FM} = 2(\Delta f + B) = 430\,\text{kHz}$$

A razão de desvio β é dado por

$$\beta = \frac{\Delta f}{B} = \frac{200}{15}$$

e

$$B_{FM} = 2B(\beta + 1) = 30\left(\frac{200}{15} + 1\right) = 430\,\text{kHz}$$

Para PM: Quando multiplicamos $m(t)$ por 2, dobramos sua derivada, de modo que, agora, $\dot{m}_p$ = 40.000 e

$$\Delta f = \frac{1}{2\pi} k_p \dot{m}_p = 100\,\text{kHz}$$

e

$$B_{PM} = 2(\Delta f + B) = 230\,\text{kHz}$$

A razão de desvio β é dado por

$$\beta = \frac{\Delta f}{B} = \frac{100}{15}$$

e

$$B_{PM} = 2B(\beta + 1) = 30\left(\frac{100}{15} + 1\right) = 230\,\text{kHz}$$

Observemos que dobrar a amplitude do sinal [dobrar $m(t)$] resulta, aproximadamente, no dobro do desvio de frequência Δf tanto da forma de onda FM como da forma de onda PM.

Exemplo 5.4 Repitamos o Exemplo 5.1 para o caso em que $m(t)$ é dilatado temporalmente por um fator 2; ou seja, o período de $m(t)$ passa a ser 4×10^{-4}.

Recordemos que a dilatação temporal de um sinal por um fator 2 reduz a largura espectral (largura de banda) do sinal por um fator 2. Podemos comprovar isso observando que a frequência fundamental agora é 2,5 kHz e o terceiro harmônico, 7,5 kHz. Logo, B = 7,5 kHz, que é a metade da largura de banda no exemplo anterior. Além disso, a dilatação temporal não afeta a amplitude de pico, de modo que $m_p = 1$. Contudo, $\dot{m}_p$ fica dividido por 2, ou seja, $\dot{m}_p$ = 10.000.
Para FM:

$$\Delta f = \frac{1}{2\pi} k_f m_p = 100\,\text{kHz}$$

$$B_{FM} = 2(\Delta f + B) = 2(100 + 7{,}5) = 215\,\text{kHz}$$

Para PM:

$$\Delta f = \frac{1}{2\pi} k_p \dot{m}_p = 25\,\text{kHz}$$

$$B_{PM} = 2(\Delta f + B) = 65\,\text{kHz}$$

Notemos que a dilatação temporal de $m(t)$ tem pouco efeito sobre a largura de banda FM, mas reduz a largura de banda PM à metade do valor original. Isso comprova a observação feita anteriormente, que o espectro PM depende fortemente do espectro de $m(t)$.

Exemplo 5.5 Um sinal modulado em ângulo, com portadora $\omega_c = 2\pi \times 10^5$ é descrito pela equação

$$\varphi_{EM}(t) = 10\cos(\omega_c t + 5\,\text{sen}\,3000t + 10\,\text{sen}\,2000\pi t)$$

(a) Determinemos a potência do sinal modulado.
(b) Determinemos o desvio de frequência Δf.
(c) Determinemos a razão de desvio β.
(d) Determinemos o desvio de fase $\Delta\phi$.
(e) Estimemos a largura de banda de $\varphi_{EM}(t)$.

A largura de banda do sinal é a maior frequência em $m(t)$ (ou sua derivada). Nesse caso, $B = 2000\pi/2\pi = 1000$ Hz.

(a) A amplitude da portadora é 10 e a potência,

$$P = \frac{10^2}{2} = 50$$

(b) Para determinar o desvio de frequência Δf, determinemos a frequência instantânea ω_i, que é dada por

$$\omega_i = \frac{d}{dt}\theta(t) = \omega_c + 15.000\cos 3000t + 20.000\pi\cos 2000\pi t$$

O desvio da portadora é $15.000\cos 3000t + 20.000\pi\cos 2000t$. As duas senoides se somam em fase em algum ponto, de modo que o valor máximo dessa expressão é $15.000 + 20.000\pi$. Esse é o máximo desvio da portadora $\Delta\omega$. Logo,

$$\Delta f = \frac{\Delta\omega}{2\pi} = 12.387{,}32 \text{ Hz}$$

(c) $\beta = \dfrac{\Delta f}{B} = \dfrac{12.387{,}32}{1000} = 12{,}387$

(d) O ângulo $\theta(t) = \omega t + (5\,\text{sen}\,3000t + 10\,\text{sen}\,2000\pi t)$. O desvio de fase é o valor máximo do ângulo entre parênteses, dado por $\Delta\phi = 15$ rad.

(e) $B_{EM} = 2(\Delta f + B) = 26.774{,}65$ Hz.
Vale observar a generalidade do método para estimar a largura de banda de uma onda modulada em ângulo. Não é preciso saber se é uma onda FM, PM ou de algum outro tipo de modulação em ângulo. O método é aplicável a qualquer sinal modulado em ângulo.

Uma Nota Histórica: Edwin H. Armstrong (1890-1954)

Hoje, ninguém duvida que FM tem um papel importante em radiodifusão e comunicações. No entanto, até a década de 1960, a radiodifusão FM parecia fadada ao fracasso, devido ao antieconômico uso de largura de banda.

A história de FM é repleta de estranhas ironias. O ímpeto por trás do desenvolvimento de FM residia no desejo de reduzir a largura de banda na transmissão de sinais. Um raciocínio superficial

Edwin H. Armstrong. (Reproduzido com permissão de Armstrong Family Archives.)

mostrara que seria possível reduzir a largura de banda de transmissão com o uso de FM. Contudo, resultados experimentais mostraram o contrário. Na verdade, a largura de banda de transmissão de FM era maior do que a de AM. Uma cuidadosa análise matemática efetuada por Carson mostrou que FM exigia uma largura de banda maior que AM. Lamentavelmente, Carson não reconheceu a vantagem compensadora de FM, no que diz respeito à capacidade de eliminação de ruído. Sem muita base, ele concluiu que FM introduzia distorção inerente e não exibia qualquer vantagem compensadora.[1] Em um artigo posterior, escreveu

"Na verdade, à medida que mais e mais esquemas são analisados e testados, e a natureza essencial do problema se torna mais claramente perceptível, somos forçados a concluir que a estática (ruído), assim como os pobres, sempre estará entre nós."[2] Assim, a opinião de um dos maiores matemáticos especializados na área de comunicações da época atrasou o desenvolvimento de FM em mais de uma década. A *vantagem supressora de ruído* de FM foi posteriormente provada pelo major Edwin H. Armstrong,[3] um brilhante engenheiro cujas contribuições ao campo de sistemas de rádio são comparáveis às de Hertz e Marconi. Em grande parte, o trabalho de Armstrong foi responsável pela retomada do interesse em FM.

Embora não tenha inventado o conceito, Armstrong é considerado o pai da moderna modulação FM. Nascido a 18 de dezembro de 1890, na cidade de Nova York, Edwin H. Armstrong é reconhecido como um dos pioneiros da radioeletricidade do século XX. Armstrong tem o crédito da invenção do *circuito regenerativo* (Patente americana 1.113.149, concedida em 1912 quando ele era aluno de graduação na Columbia University), do *circuito super-heteródino* (Patente americana 1.342.885, concedida em 1918 quando ele servia ao exército dos Estados Unidos, estacionado em Paris, durante a Primeira Guerra Mundial), do *circuito super-regenerativo* (Patente americana 1.424.065, concedida em 1922) e do sistema completo de radiodifusão FM (Patente americana 1.941.066, de 1933). Todas foram contribuições inovadoras ao campo do rádio. Em 1933, a revista *Fortune* declarou: a modulação em frequência de banda larga é a quarta, e talvez a maior, em uma lista de invenções de Armstrong que moldaram grande parte da moderna radiodifusão. O major Armstrong é o reconhecido inventor do circuito "de realimentação" regenerativo, que tirou a arte do rádio do estágio dos fones de ouvido baseados no detector a cristal e tornou possível a amplificação da difusão; do circuito super-heteródino, que é a base de praticamente todo o rádio moderno; e do circuito super-regenerativo, agora em largo uso em... sistemas de ondas curtas.[4]

Armstrong foi o último de uma linhagem de inventores solitários. Após receber a patente por FM em 1933, Armstrong leu seu agora famoso artigo (que, posteriormente, foi publicado nos anais do IRE[3]), acompanhado da primeira demonstração pública da radiodifusão FM no dia 5 de novembro de

1935, em uma reunião secional do Instituto de Engenheiros de Rádio (IRE – *Institute of Radio Engineers*, um predecessor do IEEE*). A dramática redução do ruído estático com FM não foi bem recebida pela indústria da radiodifusão, que viu FM como uma ameaça a seu vasto investimento comercial em rádio AM. Para estabelecer a difusão FM, Armstrong travou uma longa e cara batalha com a indústria da radiodifusão, que, encorajada pela Comissão Federal de Comunicação (FCC — *Federal Communication Commission*), lutou com unhas e dentes para resistir à modulação FM. Mesmo assim, em dezembro de 1941, 67 estações comerciais FM foram autorizadas, com cerca de meio milhão de receptores em uso, e 43 pedidos de autorização pendentes. Durante audiência da FCC em setembro de 1944, o Conselho de Planejamento Técnico de Rádio (RTPB — *Radio Technical Planning Board*) fez sua recomendação final para que 75 canais fossem alocados a FM, na faixa de 41 a 56 MHz.

Apesar da recomendação do RTPB, que representava a melhor orientação para a comunidade de engenharia de rádio, persistia intenso *lobby* para que a FCC deslocasse a banda FM, principalmente por parte daqueles que divulgavam a preocupação quanto às fortes interferências que poderiam ocorrer na banda de 40 MHz em decorrência de reflexões na ionosfera. Assim, em junho de 1945, a FCC, com base no testemunho errôneo de um especialista técnico, bruscamente deslocou a largura de banda alocada a FM da faixa de 42 a 50 MHz para 88 a 108 MHz. Esse foi um duro golpe em FM, pois tornou obsoletos mais de meio milhão de receptores e equipamentos (transmissores, antenas etc.), fabricados e vendidos pela indústria de FM às 50 estações desde 1941, para a banda de 42 a 50 MHz. Armstrong lutou contra a decisão e, posteriormente, conseguiu que o especialista técnico reconhecesse seu erro. Apesar de tudo isso, as alocações da FCC permaneceram inalteradas. Nas disputas judiciais, Armstrong gastou a considerável fortuna que havia acumulado com as invenções. Os gigantes da radiodifusão, que haviam resistido a FM com tanto empenho, mudaram de opinião e usaram suas invenções sem pagar-lhe os devidos *royalties*. Armstrong gastou muito de seu tempo em tribunais em alguns dos mais longos, notáveis e cáusticos processos judiciais de patentes da época.[5] Por fim, sem recursos, com energia exaurida e vida familiar destroçada, o desiludido Armstrong cometeu suicídio: (em 1945) pulou da janela de seu apartamento no décimo terceiro andar, em River House, Nova York.

A viúva de Armstrong prosseguiu com a batalha judicial e venceu. Na década de 1960, FM estava bem estabelecida como um sistema superior de rádio[6] e Edwin H. Armstrong, devidamente reconhecido como o inventor da modulação em frequência. Em 1955, a UIT adicionou seu nome ao rol de grandes inventores. Em 1980, Edwin H. Armstrong foi introduzido ao *U.S National Inventors Hall of Fame* [*Hall* da Fama de Inventores Americanos]; em 1983, seu retrato foi estampado em um selo do correio dos Estados Unidos.[7]

5.3 GERAÇÃO DE ONDAS FM

Existem, basicamente, duas maneiras de gerar ondas FM: **direta** e **indireta**. Primeiro, descreveremos o gerador de FM de banda estreita utilizado na **geração indireta de FM** de sinais de banda larga modulados em ângulo.

Geração de NBFM

Para NBFm e sinais NBPM, mostramos anteriormente que, como $|k_f a(t)| << 1$ e $|k_p m(t)| << 1$, respectivamente, os sinais modulados podem ser aproximados como

$$\varphi_{\text{NBFM}}(t) \simeq A[\cos \omega_c t - k_f a(t) \operatorname{sen} \omega_c t] \tag{5.21a}$$

$$\varphi_{\text{NBPM}}(t) \simeq A[\cos \omega_c t - k_p m(t) \operatorname{sen} \omega_c t] \tag{5.21b}$$

As duas aproximações são lineares e similares à expressão para a onda AM. Na verdade, as Eq. (5.21) sugerem um possível método para a geração de sinais FM e PM de banda estreita com o uso de moduladores DSB-SC. A representação em diagrama de bloco de tais sistemas é mostrada na Fig. 5.8.

**Institute of Electrical and Electronic Engineers* (Instituto de Engenheiros Elétricos e Eletrônicos). (N.T.)

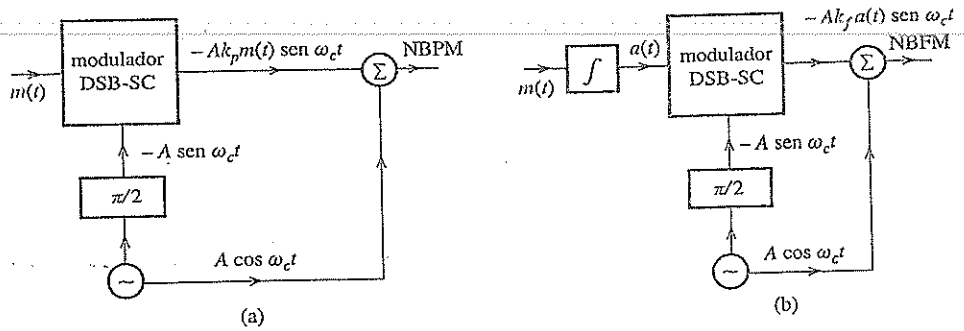


Figura 5.5 (a) Gerador PM de banda estreita, (b) Gerador de sinal FM de banda estreita.

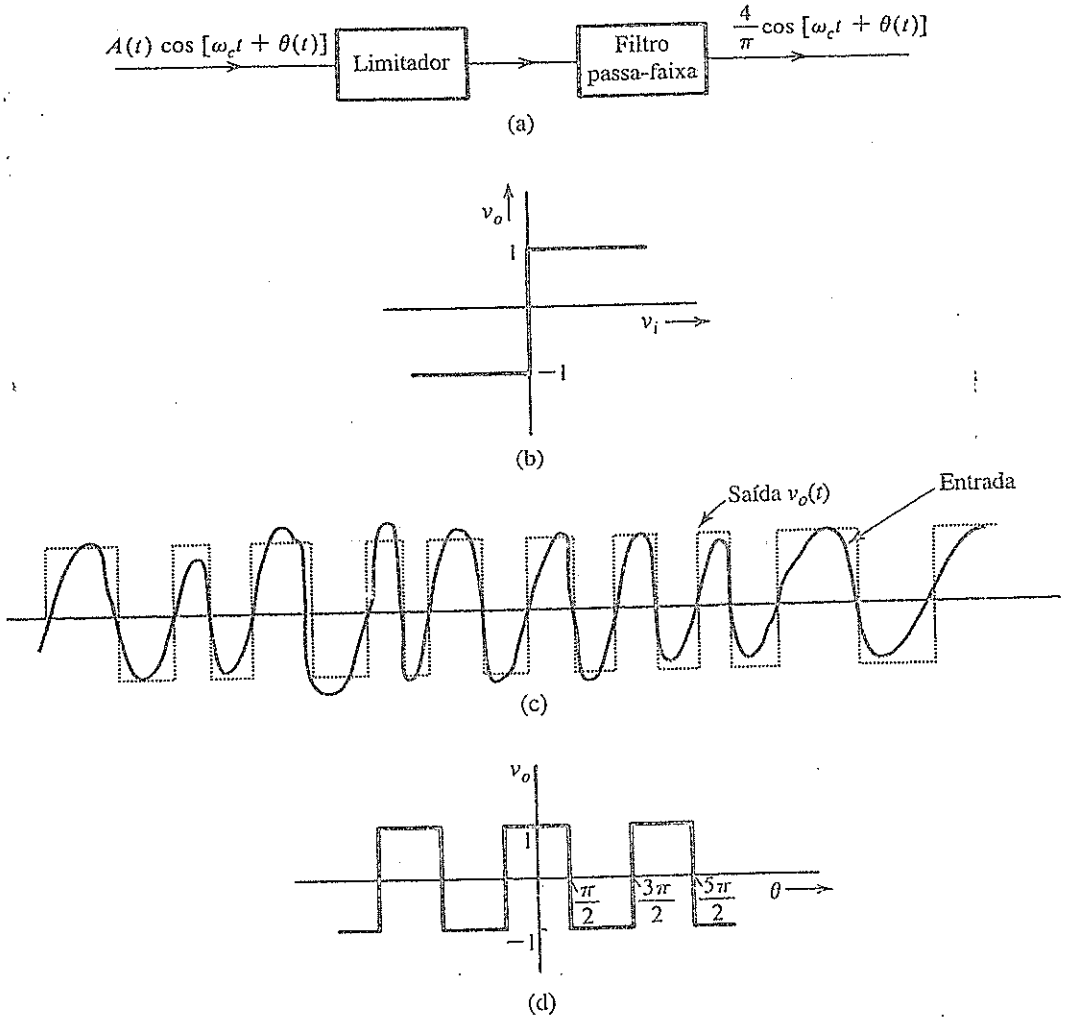


Figura 5.9 (a) Limitador e filtro passa-faixa usados para remover variações de amplitude em uma onda FM. (b) Característica entrada-saída do limitador. (c) Entrada do limitador e correspondente saída, (d) Saída do limitador em função de θ.

É importante ressaltar que o sinal NBFM gerado como na Fig. 5.8b tem alguma distorção, devido à aproximação na Eq. (5.10). A saída desse modulador NBFM também exibe algumas variações de amplitude. Um dispositivo não linear projetado para limitar a amplitude de um sinal passa-faixa pode remover a maior parte dessa distorção.

Limitador Passa-Faixa

As variações de amplitude de uma portadora modulada em ângulo podem ser eliminadas com um **limitador passa-faixa**, que consiste em um limitador seguido por um filtro passa-faixa (Fig. 5.9a). A característica entrada-saída de um limitador é mostrada na Fig. 5.9b. Observemos que, para uma entrada senoidal, a saída do limitador passa-faixa será uma onda quadrada de amplitude unitária, qualquer que seja a amplitude senoidal de entrada. Além disso, os pontos em que a senoide de entrada se anula (pontos de cruzamento do zero) são preservados na saída, pois, quando a entrada é zero, a saída também é zero (Fig. 5.9b). Assim, uma entrada senoidal modulada em ângulo $v_i(t) = A(t) \cos \theta(t)$ resulta em uma onda quadrada modulada em ângulo, de amplitude constante, $v_o(t)$, como indicado na Fig. 5.9c. Como vimos, uma operação não linear desse tipo preserva a informação da modulação em ângulo. Quando $v_o(t)$ é aplicado ao filtro passa-faixa centrado em ω_c, a saída é uma onda modulada em ângulo, de amplitude constante. Para comprovar isso, consideremos a onda de entrada modulada em ângulo

$$v_i(t) = A(t) \cos \theta(t)$$

em que

$$\theta(t) = \omega_c t + k_f \int_{-\infty}^{t} m(\alpha)\, d\alpha$$

A saída $v_o(t)$ do limitador é +1 ou −1, dependendo se $v_i(t) = A(t) \cos \theta(t)$ é positivo ou negativo (Fig. 5.9c). Como $A(t) \geq 0$, $v_o(t)$ pode ser expresso como uma função de θ:

$$v_o(\theta) = \begin{cases} +1 & \cos \theta > 0 \\ -1 & \cos \theta < 0 \end{cases}$$

Portanto, como uma função de θ, v_o é uma função periódica de onda quadrada, com período 2π (Fig. 5.9d), que pode ser expandida em uma série de Fourier (Capítulo 2)

$$v_o(\theta) = \frac{4}{\pi}\left(\cos \theta - \frac{1}{3}\cos 3\theta + \frac{1}{5}\cos 5\theta + \cdots\right)$$

Em qualquer instante de tempo t, $\theta = \omega_c t + k_f \int m(\alpha)\, d\alpha$. Logo, como função do tempo, a saída v_o é dada por

$$\begin{aligned} v_o[\theta(t)] &= v_o\left[\omega_c t + k_f \int m(\alpha)\, d\alpha\right] \\ &= \frac{4}{\pi}\left\{\cos\left[\omega_c t + k_f \int m(\alpha)\, d\alpha\right] - \frac{1}{3}\cos 3\left[\omega_c t + k_f \int m(\alpha)\, d\alpha\right]\right. \\ &\quad \left. + \frac{1}{5}\cos 5\left[\omega_c t + k_f \int m(\alpha)\, d\alpha\right] \cdots\right\} \end{aligned}$$

A saída, portanto, tem a onda original FM mais ondas FM com frequências multiplicadas por fatores 3, 5, 7,... Podemos aplicar a saída do limitador a um filtro passa-faixa com frequência central ω_c e largura da banda B_{FM}, como mostrado na Fig. 5.9a. A saída do filtro $e_o(t)$ é a desejada portadora modulada em ângulo, com amplitude constante,

$$e_o(t) = \frac{4}{\pi}\cos\left[\omega_c(t) + k_f \int m(\alpha) d\alpha\right]$$

Embora tenhamos deduzido esses resultados para FM, os mesmo também se aplicam a PM (modulação em ângulo em geral). O filtro passa-faixa não apenas mantém constante a amplitude da desejada

portadora modulada em ângulo, mas ainda suprime parcialmente o ruído de canal, quando esse é pequeno.[8]

Método Indireto de Armstrong

No método indireto de Armstrong, o NBFM é gerado como mostrado na Fig. 5.8b [ou Eq. (5.10). O sinal NBFM é, então, convertido em WBFM com o uso de **multiplicadores de frequência** adicionais.

Um multiplicador de frequência pode ser realizado por um dispositivo não linear seguido por um filtro passa-faixa. Primeiro, consideremos um dispositivo não linear cujo sinal de saída $y(t)$ correspondente a uma entrada $x(t)$ seja dado por

$$y(t) = a_2 x^2(t)$$

Caso um sinal FM passe por esse dispositivo, o sinal de saída será

$$\begin{aligned} y(t) &= a_2 \cos^2\left[\omega_c t + k_f \int m(\alpha)\, d\alpha\right] \\ &= 0.5a_2 + 0.5a_2 \cos\left[2\omega_c t + 2k_f \int m(\alpha)\, d\alpha\right] \end{aligned} \tag{5.22}$$

Portanto, um filtro passa-faixa centrado em $2\omega_c$ recuperará um sinal FM com o dobro da frequência instantânea original. Em geral, um dispositivo linear deve ter uma característica da forma

$$y(t) = a_0 + a_1 x(t) + a_2 x^2(t) + \cdots + a_n x^n(t) \tag{5.23}$$

Se $x(t) = A \cos[\omega_c t + k_f \int m(\alpha)\, d\alpha]$, com o uso de identidades trigonométricas, podemos mostrar prontamente que $y(t)$ tem a forma

$$\begin{aligned} y(t) = c_o + c_1 \cos\left[\omega_c t + k_f \int m(\alpha)\, d\alpha\right] + c_2 \cos\left[2\omega_c t + 2k_f \int m(\alpha)\, d\alpha\right] \\ + \cdots + c_n \cos\left[n\omega_c t + nk_f \int m(\alpha)\, d\alpha\right] \end{aligned} \tag{5.24}$$

Ou seja, a saída terá espectros em ω_c, $2\omega_c$,..., $n\omega_c$, com desvios de frequência Δf, $2\Delta f$,..., $n\Delta f$, respectivamente. Cada uma dessas componentes é um sinal FM separado dos outros. Assim, um filtro passa-faixa centrado em $n\omega_c$ pode recuperar um sinal FM cuja frequência instantânea é ω_c multiplicado por um fator n. Esses dispositivos, que consistem em uma não linearidade e um filtro passa-faixa, são conhecidos como **multiplicadores de frequência**. Na verdade, um multiplicador de frequência pode aumentar tanto a frequência portadora como o desvio de frequência por um fator inteiro n. Dessa forma, se desejarmos um aumento de doze vezes no desvio de frequência, podemos usar um dispositivo não linear de décima segunda ordem ou dois de segunda ordem e um de terceira ordem em cascata. A saída tem um filtro passa-faixa centrado em $12\omega_c$, para que selecione somente o termo apropriado, cujos frequência portadora e desvio de frequência Δf são 12 vezes os valores originais.

Isso forma a base do modulador indireto de Armstrong. Primeiro, o NBFM é gerado aproximadamente. A seguir, a frequência NBFM é multiplicada e suas variações de amplitude, limitadas. Em geral, desejamos aumentar Δf por um fator n muito grande. Isso também aumenta a frequência portadora por n. Um aumento tão grande na frequência portadora pode não ser necessário. Se for esse o caso, podemos aplicar mistura de frequências (ver Exemplo 4.2, Fig. 4.7) para deslocar a frequência portadora até o valor desejado.

Um diagrama simplificado de um transmissor FM comercial baseado no método de Armstrong é mostrado na Fig. 5.10. A saída final deve ter uma frequência portadora de 91,2 MHz e $\Delta f = 75$ kHz. Iniciamos com NBFM, com frequência portadora $f_{c1} = 200$ kHz, gerada por um oscilador a cristal.

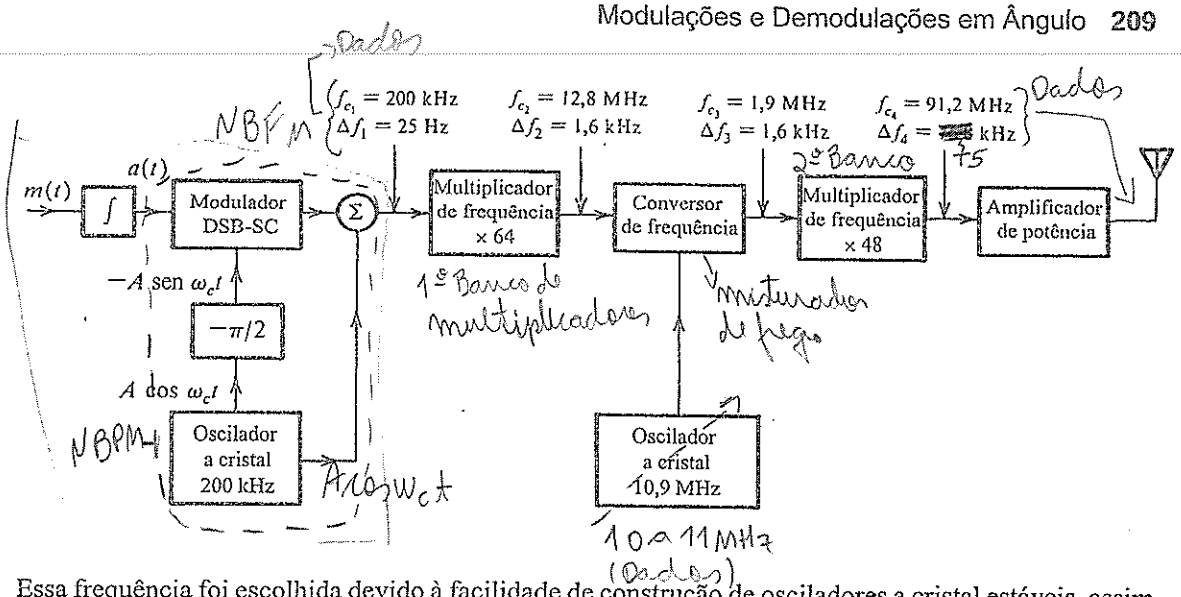


Figura 5.10 Diagrama em blocos do transmissor FM indireto de Armstrong.

Essa frequência foi escolhida devido à facilidade de construção de osciladores a cristal estáveis, assim como moduladores balanceados, para a mesma. Para manter $\beta << 1$, como necessário para NBFM, o valor do desvio Δf é escolhido como 25 Hz. Para modulação por tom, $\beta = \Delta f/f_m$. O espectro de banda base (necessário para fins de alta fidelidade) ocupa a faixa de 50 Hz a 15 kHz. A escolha de $\Delta f =$ 25 Hz é razoável, pois, para o pior caso possível ($f_m = 50$), resulta em $\beta = 0{,}5$.

Para alcançar $\Delta f = 75$ kHz, devemos multiplicar por 75.000/25 = 3.000. Isso pode ser feito com dois estágios multiplicadores, de 64 e 48, como indicado na Fig. 5.10, resultando em uma multiplicação total de 64 × 48 = 3.072, e em $\Delta f = 76{,}8$ kHz.* A multiplicação é efetuada por meio de dobradores e triplicadores de frequência em cascata, segundo a necessidade. Dessa forma, uma multiplicação por 64 pode ser obtida com seis dobradores em cascata, e uma multiplicação por 48, com quatro dobradores e um triplicador em cascata. No entanto, a multiplicação de $f_c = 200$ kHz por 3.072 resultaria em uma portadora de cerca de 600 MHz. Esse problema é resolvido por meio de uma translação ou conversão de frequência após o primeiro multiplicador (Fig. 5.10). A primeira multiplicação por 64 resulta em uma frequência portadora $f_{c2} = 200$ kHz × 64 = 12,8 MHz, e em um desvio de portadora $\Delta f_2 =$ 25 × 64 = 1,6 kHz. Agora, usamos um conversor de frequência (ou misturador), com frequência portadora de 10,9 MHz, para trasladar todo o espectro. Isso resulta em uma nova frequência portadora $f_{c3} =$ 12,8 − 10,9 = 1,9 MHz. O conversor de frequência traslada todo o espectro, sem alterar Δf. Assim, $\Delta f_3 = 1{,}6$ kHz. A multiplicação por 48 resulta em $f_{c4} = 1{,}9 \times 48 = 91{,}2$ MHz e $\Delta f_4 = 1{,}6 \times 48 = 76{,}8$ kHz.

Este esquema tem a vantagem de estabilidade de frequência, mas está sujeito a ruído inerente, causado por excessiva multiplicação e distorção nas frequências modulantes mais baixas, em que o valor de $\Delta f/f_m$ não é suficientemente pequeno.

Exemplo 5.6 Discutamos a natureza da distorção inerente ao gerador de FM indireto de Armstrong.

Dois tipos de distorção surgem neste esquema: distorção de amplitude e distorção de frequência. A onda NBFM é dada por [Eq. (5.10)]:

$$\begin{aligned}\varphi_{\text{FM}}(t) &= A[\cos\omega_c t - k_f a(t)\,\text{sen}\,\omega_c t]\\ &= AE(t)\cos[\omega_c t + \theta(t)]\end{aligned}$$

em que

$$E(t) = \sqrt{1 + k_f^2 a^2(t)} \qquad \text{e} \qquad \theta(t) = \tan^{-1}[k_f a(t)]$$

---

*Caso desejemos que Δf seja exatamente 75 kHz, em vez de 76,8 kHz, devemos reduzir o desvio Δf de banda estreita de 25 Hz para 25(75/76,8) = 24,41 Hz.

Distorção de amplitude ocorre porque a amplitude $AE(t)$ da forma de onda modulada não é constante. Esse não é um problema sério, pois variações de amplitude podem ser eliminadas por um limitador passa-faixa, como discutido anteriormente nessa seção (ver Fig. 5.9). Idealmente, $\theta(t)$ deve ser igual a $k_f a(t)$. Contudo, na equação anterior, a fase $\theta(t)$ é

$$\theta(t) = \tan^{-1}[k_f a(t)]$$

e a frequência instantânea $\omega_i(t)$,

$$\begin{aligned}\omega_i(t) = \dot{\theta}(t) &= \frac{k_f \dot{a}(t)}{1 + k_f^2 a^2(t)} \\ &= \frac{k_f m(t)}{1 + k_f^2 a^2(t)} \\ &= k_f m(t)[1 - k_f^2 a^2(t) + k_f^4 a^4(t) - \cdots]\end{aligned}$$

No caso ideal, a frequência instantânea deveria ser $k_f m(t)$. Os termos remanescentes nessa equação são a distorção.

Investiguemos o efeito dessa distorção na modulação por tom, em que $m(t) = \alpha \cos \omega_m t$, $a(t) = \alpha \operatorname{sen} \omega_m t/\omega_m$, e o índice de modulação é $\beta = \alpha k_f/\omega_m$:

$$\omega_i(t) = \beta\omega_m \cos \omega_m t(1 - \beta^2 \operatorname{sen}^2 \omega_m t + \beta^4 \operatorname{sen}^4 \omega_m t - \cdots)$$

Dessa última equação, fica evidente que esse esquema tem distorção por harmônicos impares, sendo o terceiro harmônico o termo mais importante. Desprezando os termos restantes, a última equação fica escrita como

$$\begin{aligned}\omega_i(t) &\simeq \beta\omega_m \cos \omega_m t(1 - \beta^2 \operatorname{sen}^2 \omega_m t) \\ &= \underbrace{\beta\omega_m\left(1 - \frac{\beta^2}{4}\right) \cos \omega_m t}_{\text{desejado}} + \underbrace{\frac{\beta^3\omega_m}{4} \cos 3\omega_m t}_{\text{distorção}}\end{aligned}$$

A razão entre a distorção por terceiro harmônico e o sinal desejado pode ser obtida para o gerador na Fig. 5.10. Para o estágio NBFM,

$$\beta B = \Delta f_1 = 25 \text{ Hz}$$

Logo, o pior caso ocorre na frequência modulante mais baixa. Por exemplo, se a frequência do tom for apenas 50 Hz, $\beta = 0{,}5$. Nesse caso, a distorção por terceiro harmônico é 1/15 ou 6,67%.

Geração Direta

Em um oscilador controlado por tensão (VCO), a frequência é controlada por uma tensão externa. A frequência de oscilação varia linearmente com a tensão de controle. Podemos gerar uma onda FM usando o sinal modulante como um sinal de controle. Isso fornece

$$\omega_i(t) = \omega_c + k_f m(t)$$

Podemos construir um VCO com um amplificador operacional e um comparador histerético,[9] (como um circuito disparador de Schmitt, ou Schmitt *trigger*). Outra forma de alcançar o mesmo objetivo

consiste em variar um dos parâmetros reativos (C ou L) do circuito ressonante de um oscilador. Um *diodo semicondutor com polarização reversa* atua como um capacitor cuja capacitância varia com a tensão de polarização. A capacitância desses diodos, conhecidos por diferentes nomes comerciais (por exemplo, Varicap, Varactor, Voltacap), pode ser aproximada como uma função linear da tensão de polarização $m(t)$, em um intervalo limitado. Nos osciladores Hartley ou Colpitt, por exemplo, a frequência de oscilação é dada por.

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

Se a capacitância C for variada pelo sinal modulante $m(t)$, ou seja, se

$$C = C_0 - km(t)$$

então

$$\begin{aligned}
\omega_0 &= \frac{1}{\sqrt{LC_0\left[1 - \dfrac{km(t)}{C_0}\right]}} \\
&= \frac{1}{\sqrt{LC_0}\left[1 - \dfrac{km(t)}{C_0}\right]^{1/2}} \\
&\approx \frac{1}{\sqrt{LC_0}}\left[1 + \frac{km(t)}{2C_0}\right] \qquad \frac{km(t)}{C_0} \ll 1
\end{aligned}$$

Neste cálculo, utilizamos a aproximação em série de Taylor

$$(1+x)^n \approx 1 + nx \qquad |x| \ll 1$$

com $n = 1/2$. Logo,

$$\begin{aligned}
\omega_0 &= \omega_c\left[1 + \frac{km(t)}{2C_0}\right] \quad \text{em que} \quad \omega_c = \frac{1}{\sqrt{LC_0}} \\
&= \omega_c + k_f m(t) \qquad \text{com} \quad k_f = \frac{k\omega_c}{2C_0}
\end{aligned}$$

Como $C = C_0 - km(t)$, o máximo desvio de capacitância é

$$\Delta C = km_p = \frac{2k_f C_0 m_p}{\omega_c}$$

Portanto,

$$\frac{\Delta C}{C_0} = \frac{2k_f m_p}{\omega_c} = \frac{2\Delta f}{f_c}$$

Na prática, $\Delta f/f_c$ é, em geral, pequeno; logo, ΔC é uma pequena fração de C_0, o que ajuda a limitar a distorção harmônica causada pela aproximação empregada nessa dedução.

Podemos, também, gerar FM diretamente por meio de um reator com núcleo saturável, no qual a indutância de uma bobina é variada por uma corrente através de uma segunda bobina (em volta do mesmo núcleo). Isso resulta em um indutor variável, cuja indutância é proporcional à corrente na segunda bobina.

A geração direta de FM produz, em geral, um desvio de frequência suficiente e requer pouca multiplicação de frequência. Contudo, a estabilidade de frequência desse método é pobre. Na prática, é feito uso de realimentação para estabilizar a frequência. A frequência desejada é comparada com uma frequência constante gerada por um oscilador a cristal estável. Um sinal de erro (erro na frequência) é detectado e realimentado ao oscilador para corrigir o erro.

Propriedades da Modulação em Ângulo

A FM (como modulação em ângulo, em geral) tem certas propriedades características que recomendam sua aplicação em diversos sistemas de rádio. A largura de banda de transmissão de sistemas AM não pode ser alterada. Por isso, sistemas AM não são capazes de trocar largura de banda por largura de banda de transmissão. Sistemas baseados em modulação por codificação de pulso (PCM) (Capítulo 6) têm essa capacidade, assim como sistemas de modulação em ângulo. Na modulação em ângulo, a largura de banda de transmissão pode ser ajustada pelo valor de Δf. Para sistemas de modulação em ângulo, a SNR é aproximadamente proporcional ao quadrado da largura de banda de transmissão B_T. Em PCM, a SNR varia exponencialmente com B_T e, portanto, é superior à modulação em ângulo.

Exemplo 5.7 Projetemos um modulador FM indireto para gerar um sinal FM com frequência portadora 97,3 MHz e $\Delta f = 10{,}24$ kHz. Há disponibilidade de um gerador NBFM com $f_{c_1} = 20$ kHz e $\Delta f = 5$ Hz. Somente dobradores de frequência podem ser usados como multiplicadores. Adicionalmente, para mistura de frequência, há disponibilidade de um oscilador local (LO) com frequência ajustável entre 400 e 500 kHz.

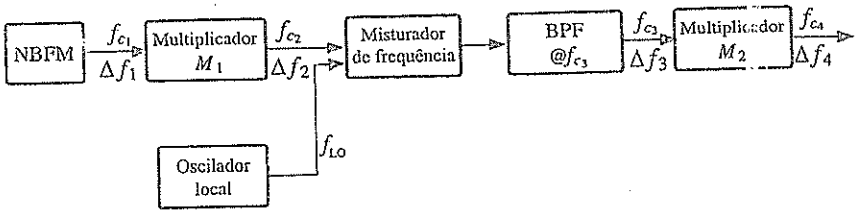


Figura 5.11 Projeto de um modulador indireto de Armstrong.

O modulador é mostrado na Fig. 5.11. Devemos determinar M_1, M_2 e f_{LO}. Primeiro, o gerador NBFM produz

$$f_{c_1} = 20.000 \qquad \text{e} \qquad \Delta f_1 = 5$$

O WBFM final deve ter

$$f_{c_4} = 97{,}3 \times 10^6 \qquad \Delta f_4 = 10.240$$

Calculemos, agora, o fator total de multiplicação de frequência:

$$M_1 \cdot M_2 = \frac{\Delta f_4}{\Delta f_1} = 2048 = 2^{11} \tag{5.25}$$

Como apenas dobradores de frequência podem ser utilizados, temos três equações:

$$M_1 = 2^{n_1}$$
$$M_2 = 2^{n_2}$$
$$n_1 + n_2 = 11$$

Temos, ainda,

$$f_{c_2} = 2^{n_1} f_{c_1} \qquad \text{e} \qquad f_{c_4} = 2^{n_2} f_{c_3}$$

Para determinar f_{LO}, há três relações possíveis:

$$f_{c_3} = f_{c_2} \pm f_{LO} \qquad \text{e} \qquad f_{c_3} = f_{LO} - f_{c_2}$$

Cada uma deve ser testada para estabelecer qual delas cai na faixa

$$400.000 \leq f_{LO} \leq 500.000$$

(a) Primeiro, testemos $f_{c3} = f_{c2} - f_{LO}$. Isso nos leva a

$$\begin{aligned} 97{,}3 \times 10^6 &= 2^{n_2}\left(2^{n_1} f_{c_1} - f_{LO}\right) \\ &= 2^{n_1+n_2} f_{c_1} - 2^{n_2} f_{LO} \\ &= 2^{11} 20 \times 10^3 - 2^{n_2} f_{LO} \end{aligned}$$

Portanto, temos

$$f_{LO} = 2^{-n_2}\left(4{,}096 \times 10^7 - 9{,}73 \times 10^7\right) < 0$$

Este valor está fora da faixa de frequência do oscilador local.

(b) A seguir, testemos $f_{c_3} = f_{c_2} + f_{LO}$, que nos leva a

$$\begin{aligned} 97{,}3 \times 10^6 &= 2^{n_2}\left(2^{n_1} f_{c_1} + f_{LO}\right) \\ &= 2^{11} 20 \times 10^3 + 2^{n_2} f_{LO} \end{aligned}$$

Assim, temos

$$f_{LO} = 2^{-n_2}\left(5{,}634 \times 10^7\right)$$

Se $n_2 = 7$, $f_{LO} = 440$ kHz, que está na faixa de valores realizáveis com o oscilador local.

(c) Se escolhermos $f_{c_3} = f_{LO} - f_{c_2}$, teremos

$$\begin{aligned} 97{,}3 \times 10^6 &= f_{LO} - 2^{n_2} 2^{n_1} f_{c_1} \\ &= 2^{n_2} f_{LO} - 2^{11}(20 \times 10^3) \end{aligned}$$

Portanto, temos

$$f_{LO} = 2^{-n_2}\left(13{,}826 \times 10^7\right)$$

Nenhum inteiro n_2 levará a um valor realizável de f_{LO}.

Assim, o projeto final é $M_1 = 16$, $M_2 = 128$ e $f_{LO} = 440$ kHz.

5.4 DEMODULAÇÃO DE SINAIS FM

A informação contida em sinais FM reside na frequência instantânea $\omega_i = \omega_c + k_f m(t)$. Portanto, um circuito seletivo em frequência, com função de transferência $|H(f)| = 2a\pi f + b$ na faixa FM, produziria uma saída proporcional à frequência instantânea (Fig. 5.12a).* Existem diversos circuitos com tais características. O mais simples é um diferenciador ideal, com função de transferência $j2\pi f$.

*Desde que as variações de ω_i sejam lentas em comparação com a constante de tempo do circuito.

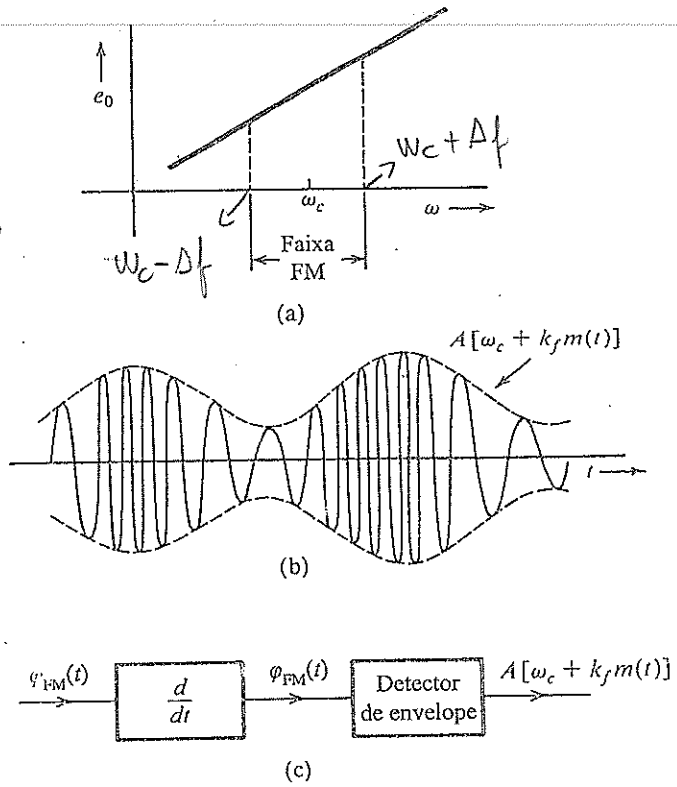


Figura 5.12 (a) Resposta do demodulador de frequência. (b) Saída de um diferenciador para a onda FM de entrada. (c) Demodulação FM por diferenciação direta.

Se aplicarmos $\varphi_{FM}(t)$ a um diferenciador ideal, a saída será

$$\begin{aligned}\dot{\varphi}_{FM}(t) &= \frac{d}{dt}\left\{A\cos\left[\omega_c t + k_f\int_{-\infty}^{t} m(\alpha)\,d\alpha\right]\right\}\\ &= A\left[\omega_c + k_f m(t)\right]\operatorname{sen}\left[\omega_c t + k_f\int_{-\infty}^{t} m(\alpha)\,d(\alpha) - \pi\right]\end{aligned} \tag{5.26}$$

Tanto a amplitude como a frequência do sinal $\dot{\varphi}_{FM}(t)$ são moduladas (Fig. 5.12); o envelope é $A[\omega_c + k_f m(t)]$. Como $\Delta\omega = k_f m_p < \omega_c$, temos $\omega_c + k_f m(t) > 0$ para todo t, e $m(t)$ pode ser obtido com a detecção de envelope de $\dot{\varphi}_{FM}(t)$ (Fig. 5.12c).

A amplitude A da portadora FM que chega ao receptor deve ser constante. Caso a amplitude A não seja constante, mas uma função do tempo, no lado direito da Eq. (5.26) haverá um termo adicional contendo dA/dt. Mesmo que este termo seja desprezível, o envelope de $\varphi_{FM}(t)$ será $A(t)[\omega_c + k_f m(t)]$ e a saída do detector de envelope será proporcional a $m(t)A(t)$, o que ainda causará distorção. Portanto, é essencial manter A constante. Diversos fatores, como ruído de canal e desvanecimento, fazem A variar. A variação em A deve ser suprimida por meio de um limitador passa-faixa (discutido na Seção 5.3) antes que o sinal seja aplicado ao detector FM.

Demoduladores de Frequência Práticos

O diferenciador representa apenas uma forma de converter a variação de frequência de sinais FM em variações de amplitude que, subsequentemente, podem ser detectadas por um detector de envelope. Podemos usar um amplificador operacional como diferenciador no receptor FM. O papel do diferenciador pode ser substituído por um sistema linear cuja resposta de frequência contenha um segmento linear com inclinação positiva. A aproximação da inclinação linear ideal na Fig. 5.12a constitui o método conhecido como **detecção de inclinação**.

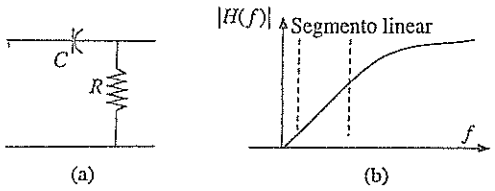


Figura 5.13
(a) Filtro passa-altos *RC*.
(b) Segmento de inclinação positiva na resposta de amplitude.

Um dispositivo simples pode ser o filtro passa-altos *RC* da Fig. 5.13. A resposta de frequência *RC* é

$$H(f) = \frac{j2\pi fRC}{1 + j2\pi fRC} \approx j2\pi fRC \quad \text{se} \quad 2\pi fRC \ll 1$$

Assim, se o parâmetro *RC* for muito pequeno, tal que seu produto pela frequência portadora $\omega_c RC \ll 1$, o filtro *RC* aproxima um diferenciador.

De modo similar, um simples circuito sintonizado *RLC* seguido por um detector de envelope também pode funcionar como detector de frequência, pois, abaixo da frequência de ressonância $\omega_o = 1/\sqrt{LC}$, sua resposta de frequência $|H(f)|$ aproxima uma inclinação linear. Assim, o projeto do receptor requer

$$\omega_c < \omega_o = \frac{1}{\sqrt{LC}}$$

Como a operação é feita com a inclinação de $|H(f)|$, esse método também é conhecido como **detecção de inclinação**. Contudo, como a inclinação de $|H(f)|$ é linear somente em uma faixa muito estreita, a saída contém considerável distorção. Essa deficiência pode ser parcialmente corrigida por um **discriminador balanceado**, formado por dois detectores de inclinação. Outro demodulador balanceado, o **detector de razão**, largamente utilizado no passado, oferece melhor proteção contra variações de amplitude que discriminadores. Durante muitos anos, detectores de razão foram o padrão em quase todos os receptores FM.[10]

Detectores de cruzamento do zero também são utilizados devido aos avanços em circuitos digitais integrados. O primeiro passo consiste em usar o amplificador limitador da Fig. 5.9a para gerar o pulso retangular de saída na Fig. 5.9c. O resultante trem de pulsos retangulares de largura variável pode, então, ser aplicado para disparar um contador digital, que é um **contador de frequência** projetado para medir a frequência instantânea a partir do número de cruzamentos do zero. A taxa de cruzamentos do zero é igual à frequência instantânea do sinal de entrada.

Demodulação FM via PLL

Consideremos uma PLL travada ao sinal de entrada sen $[\omega_c t + \theta_i(t)]$, e um sinal de erro $e_o(t)$. Quando o sinal de entrada é um sinal FM,

$$\theta_i(t) = k_f \int_{-\infty}^{t} m(\alpha)\, d\alpha + \frac{\pi}{2} \tag{5.27}$$

e,

$$\theta_o(t) = k_f \int_{-\infty}^{t} m(\alpha)\, d\alpha + 0{,}5\pi - \theta_e(t)$$

Com a PLL travada, podemos presumir um pequeno erro de frequência $\dot{\theta}_e(t) \approx 0$. Assim, o sinal de saída do filtro de malha é

$$e_o(t) = \frac{1}{c}\dot{\theta}_o(t) = \frac{1}{c}\frac{d}{dt}\left[k_f \int_{-\infty}^{t} m(\alpha)\, d\alpha + 0{,}5\pi - \theta_e(t)\right] \simeq \frac{k_f}{c} m(t) \tag{5.28}$$

Portanto, a PLL atua como um demodulador FM. Se o sinal de entrada for uma onda PM, $e_0(t) = k_p\, m(t)/c$. Neste caso, é necessário integrar $e_o(t)$ para que obtenhamos o desejado sinal $m(t)$.

Para uma análise mais precisa do comportamento da PLL como demodulador FM, consideremos o caso de um pequeno erro (modelo linear da PLL), com $H(s) = 1$. Para esse caso, a análise de realimentação da PLL sob pequeno erro do Capítulo 4 passa a

$$\Theta_o(s) = \frac{AK\,H(s)}{s + AK\,H(s)}\Theta_i(s) = \frac{AK}{s + AK}\Theta_i(s)$$

Se $E_o(s)$ e $M(s)$ forem as transformadas de Laplace de $e_o(t)$ e $m(t)$, respectivamente, das Eqs. (5.27) e (5.28), obtemos

$$\Theta_i(s) = \frac{k_f M(s)}{s} \qquad \text{e} \qquad s\Theta_o(s) = cE_o(s)$$

Logo,

$$E_o(s) = \left(\frac{k_f}{c}\right)\frac{AK}{s + AK}M(s)$$

Portanto, a saída da PLL $e_o(t)$ é uma versão distorcida de $m(t)$, e é equivalente à saída de um circuito de um polo simples (como um circuito RC simples), com função de transferência $k_f AK/c\ (s + AK)$, tendo $m(t)$ como entrada. Para reduzir a distorção, devemos escolher AK bem acima da largura de banda de $m(t)$, de modo que $e_o(t)$; $k_f m(t)/c$.

Na presença de pequeno ruído, o comportamento da PLL é comparável ao de um discriminador de frequência. A vantagem da PLL em relação ao discriminador de frequência deixa de existir somente quando o ruído é grande.

5.5 EFEITOS DE DISTORÇÃO NÃO LINEAR E INTERFERÊNCIA

Imunidade da Modulação em Ângulo a Não Linearidades

Uma propriedade muito útil da modulação em ângulo é sua amplitude constante, que a torna menos susceptível a não linearidades. Consideremos, por exemplo, um amplificador com distorção não linear de segunda ordem, cujas entrada e saída estão relacionadas por

$$y(t) = a_0 + a_1 x(t) + a_2 x^2(t) + \cdots + a_n x^n(t)$$

Fica claro que o primeiro termo é o desejado termo de amplificação do sinal, enquanto os termos restantes são as indesejadas distorções não lineares. Para o sinal modulado em ângulo

$$x(t) = A\cos[\omega_c t + \psi(t)]$$

Podemos empregar identidades trigonométricas para reescrever a saída não ideal $y(t)$ do sistema como

$$y(t) = c_o + c_1\cos[\omega_c t + \psi(t)] + c_2\cos[2\omega_c t + 2\psi(t)] + \cdots + c_n\cos[n\omega_c t + n\psi(t)]$$

Um valor suficientemente grande de ω_c torna as componentes de $y(t)$ separáveis no domínio da frequência; assim, um filtro passa-faixa centrado em ω_c e largura de banda igual a B_{FM} (ou B_{PM}) pode extrair a desejada componente de sinal FM $c_1\cos[\omega_c t + \psi(t)]$ sem distorção. Isso mostra que sinais modulados em ângulos são imunes às distorções não lineares.

Em AM, uma não linearidade similar causa não apenas modulações indesejadas, com frequências portadoras $n\omega_c$, mas também distorção do sinal desejado. Por exemplo, se um sinal DSB-SC $m(t)$ cos $\omega_c t$ passar por uma não linearidade $y(t) = a\,x(t) + b\,x^3(t)$, a saída será

$$\begin{aligned} y(t) &= a\,m(t)\cos\omega_c t + b\,m^3(t)\cos^3\omega_c t \\ &= \left[a\,m(t) + \frac{3b}{4}m^3(t)\right]\cos\omega_c t + \frac{b}{4}m^3(t)\cos 3\omega_c t \end{aligned}$$

Se este sinal for aplicado a um filtro passa-faixa, a saída será $[am(t) + (3b/4)m^3(t)]\cos\omega_c t$. Notemos a componente de distorção $(3b/4)m^3(t)$ que aparece junto com o sinal desejado $am(t)$.

Imunidade a não linearidade é a principal razão do uso de modulação em ângulo em sistemas de rádio de micro-ondas, nas quais os níveis de potência são elevados. Isso requer amplificadores de micro-ondas de classe C altamente eficientes. Adicionalmente, a amplitude constante dá a FM uma espécie de imunidade contra desvanecimento rápido. O efeito de variações de amplitude causadas por desvanecimento rápido pode ser eliminado com o uso de controle automático de ganho e limitação em banda passante. Essas vantagens tornaram a FM atraente para a tecnologia por trás da primeira geração (1G) de sistemas de telefonia celular.

As mesmas vantagens também tornam FM atraente para sistemas de rádio de micro-ondas. Nos sistemas legados de telefonia de longa distância, vários canais são multiplexados por meio de sinais SSB para formar sinais de portadora L. Os sinais multiplexados são modulados em frequência e transmitidos em sistemas de rádio de micro-ondas com muitos enlaces em tandem. Contudo, nessa aplicação, FM não é usado para reduzir efeitos de ruído, mas para tirar proveito de outras vantagens associadas à amplitude constante; assim, é usado NBFM em vez de WBFM.

Efeito de Interferência

A modulação em ângulo, quando comparada com AM, também é menos vulnerável a interferências de pequenos sinais entre canais adjacentes.

Consideremos o simples caso de interferência entre uma portadora não modulada $A\cos\omega_c t$ e outra senoide $I\cos(\omega_c+\omega)t$. O sinal recebido $r(t)$ é

$$\begin{aligned} r(t) &= A\cos\omega_c t + I\cos(\omega_c+\omega)t \\ &= (A + I\cos\omega t)\cos\omega_c t - I\,\mathrm{sen}\,\omega t\,\mathrm{sen}\,\omega_c t \\ &= E_r(t)\cos[\omega_c t + \psi_d(t)] \end{aligned}$$

em que

$$\psi_d(t) = \tan^{-1}\frac{I\,\mathrm{sen}\,\omega t}{A + I\cos\omega t}$$

Quando o sinal interferente é pequeno em comparação com a portadora ($I \ll A$),

$$\psi_d(t) \simeq \frac{I}{A}\,\mathrm{sen}\,\omega t \tag{5.29}$$

A fase de $E_r(t)\cos[\omega_c t + \psi_d(t)]$ é $\dot{\psi}_d(t)$ e a frequência instantânea, $\omega_c + \dot{\psi}_d(t)$. Se o sinal $E_r(t)\cos[\omega_c t + \psi_d(t)]$ for aplicado a um demodulador de fase ideal a saída $y_d(t)$ será $\psi_d(t)$. De modo similar, a saída $y_d(t)$ de um demodulador de frequência ideal será $\dot{\psi}_d(t)$. Logo,

$$y_d(t) = \frac{I}{A}\,\mathrm{sen}\,\omega t \qquad \text{para PM} \tag{5.30}$$

$$y_d(t) = \frac{I\omega}{A}\cos\omega t \qquad \text{para FM} \tag{5.31}$$

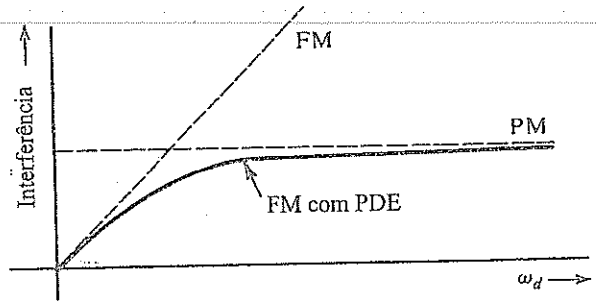


Figura 5.14 Efeito de interferência em PM, FM e FM com pré-ênfase/deênfase (PDE).

Observemos que, nos dois casos, a saída de interferência é inversamente p[illegible]orcional à amplitude da portadora A. Assim, quanto maior a amplitude da portadora A, menor o e[illegible]o da interferência. Isso difere muito do que se passa em sinais AM, nos quais a saída de interferên[illegible] independe da amplitude da portadora.* Portanto, no que diz respeito à supressão de interferência f[illegible] $(I << A)$, sistemas com modulação em ângulo são muito melhores que sistemas AM.

Devido à supressão de interferência fraca em FM, quando ouvimos r[illegible]s FM, observamos o que é conhecido como **efeito de captura**. No caso de dois transmissores em [illegible] a separação entre as frequências portadoras é menor que a faixa de áudio, em vez de interferên[illegible] observamos que a portadora mais intensa efetivamente suprime (captura) a mais fraca. Testes s[illegible]tivos mostraram que, em sinais de áudio, mesmo um baixo nível de interferência de 35 dB pod[illegible]usar efeitos indesejáveis. Portanto, em AM, o nível de interferência deve ser mantido abaixo de 3[illegible]B. No caso de FM, devido ao efeito de captura, basta que o nível de interferência fique abaixo de 6[illegible]B.

A Fig. 5.14 mostra curvas da amplitude da interferência (I/A para PM [illegible] $I\omega/A$ para FM) em função de ω na saída do receptor. Em PM, a amplitude da interferência é const[illegible] para todo ω e, no caso de FM, aumenta linearmente com ω.**

Interferência Devido a Ruído de Canal

Em um sinal modulado em ângulo, o ruído de canal atua como uma [illegible]erferência. Consideremos a forma mais comum de ruído, o ruído branco, que tem densidade esp[illegible]ctral de potência constante. Esse ruído pode ser considerado como a soma de senoides de tod[illegible]s as frequências na banda. Todas as componentes têm a mesma amplitude (devido à densidade [illegible]niforme). Isso significa que I é constante para todo ω; o espectro da amplitude da interferência [illegible]a saída do receptor; e mostrado na Fig. 5.14. O espectro da amplitude da interferência é cons[illegible]ante para PM e, no caso de FM, cresce linearmente com ω.

---

*Por exemplo, um sinal AM com uma interferência senoidal $I \cos(\omega_c + \omega)t$ é dado por

$$r(t) = [A + m(t)] \cos \omega_c t + I \cos(\omega_c + \omega)t$$
$$= [A + m(t) + I \cos \omega t] \cos \omega_c t - I \operatorname{sen} \omega t \operatorname{sen} \omega_c t$$

Para esse sinal, o envelope é

$$E(t) = \{[A + m(t) + I \cos \omega t]^2 + I^2 \operatorname{sen}^2 \omega t\}^{1/2} \approx A + m(t) + [illegible] \cos \omega t \qquad I \ll A$$

Assim, o sinal de interferência na saída do detector de envelope é $I \cos \omega t$, que independe da amplitude da portadora A. Obteríamos o mesmo resultado caso demodulação síncrona fosse utilizada. A mesma conclusão se aplica a sistemas AM-SC.

**Os resultados nas Eqs. (5.30) e (5.31) podem ser prontamente estendidos ao caso de mais de uma senoide interferente. Quando há várias senoides interferentes, o sistema se comporta linearmente, desde que as amplitudes das senoides sejam muito pequenas em comparação com a amplitude da portadora.

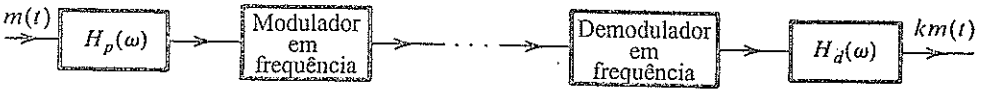


Figura 5.15 Pré-ênfase e deênfase em um sistema FM.

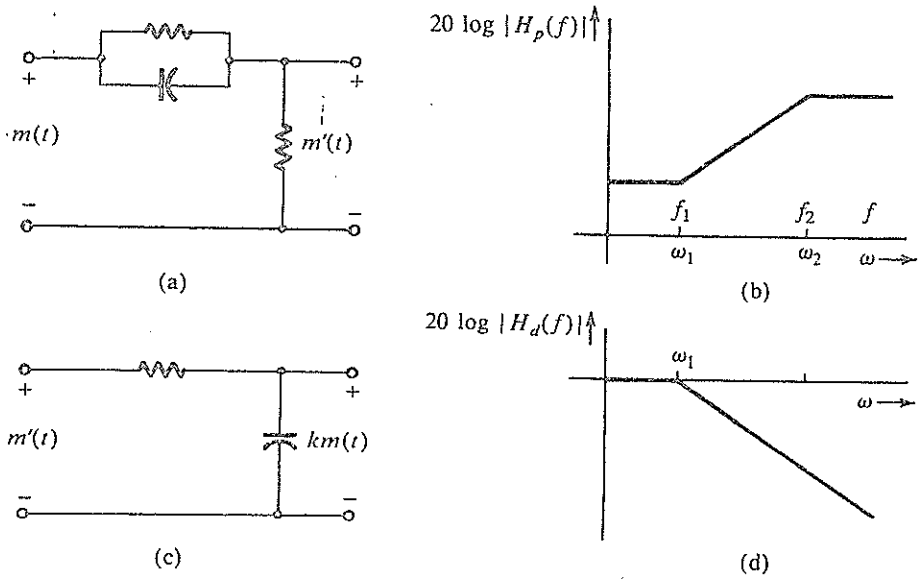


Figura 5.16 (a) Filtro de pré-ênfase e (b) sua resposta de frequência. (c) Filtro de deênfase e (d) sua resposta de frequência.

Pré-Ênfase e Deênfase na Difusão FM

A Fig. 5.14 mostra que, em FM, a interferência (ruído) aumenta linearmente com a frequência, e a potência de ruído na saída do receptor se concentra nas frequências altas. Uma análise da Fig. 4.18b mostra que a densidade espectral de potência (PSD) de um sinal de áudio $m(t)$ se concentra nas frequências abaixo de 2,1 kHz. Assim, a PSD de ruído se concentra nas frequências altas, nas quais $m(t)$ é mais fraco. Isso pode parecer uma catástrofe. Contudo, na verdade, nessa situação há uma oportunidade velada para uma grande redução do ruído. O processo, ilustrado na Fig. 5.15, funciona da seguinte maneira: no transmissor, as componentes mais fracas do sinal de áudio $m(t)$ (acima de 2,1 kHz) são amplificadas antes da modulação por um filtro de **pré-ênfase**, com função de transferência $H_p(f)$. No receptor, a saída do demodulador é aplicada a um filtro de **deênfase**, com função de transferência $H_d(f) = 1/H_p(f)$. Dessa forma, o filtro de deênfase desfaz a pré-ênfase atenuando (deenfatizando) as componentes de altas frequências (acima de 2,1 kHz) e, assim, restaura o sinal original $m(t)$. O ruído, no entanto, entra no canal e, portanto, não passa por pré-ênfase (amplificação). Todavia, o ruído passa pelo filtro de deênfase, que atenua as componentes de altas frequências, em que está concentrada a maior parte da potência de ruído (Fig. 5.14). O processo de pré-ênfase/deênfase (PDE) preserva o sinal desejado, mas reduz a potência de ruído consideravelmente.

Filtros de Pré-ênfase e de Deênfase

A Fig. 5.14 oferece uma oportunidade para pré-ênfase. Nas frequências baixas, FM é menos sujeito a interferência que PM; nas frequências altas, ocorre o oposto. Se pudermos fazer nosso sistema se comportar como FM nas frequências baixas e como PM nas frequências altas, teremos o melhor dos dois mundos. Na difusão comercial, isso é feito com um sistema (Fig. 5.15) de filtros de pré-ênfase (antes da modulação) $H_p(f)$ e deênfase (após a modulação) $H_d(f)$, mostrados na Fig. 5.16. A frequência f_1 é de 2,1 kHz e f_2, tipicamente 30 kHz ou mais (bem acima da faixa de áudio), de modo que f_2 fique

fora da faixa de interesse. Esses filtros podem ser real[illegible]os por simples circuitos RC (Fig. 5.16). A escolha de $f_1 = 2{,}1$ kHz foi, aparentemente, feita com ba[illegible]a experiência. Foi observado que essa escolha de f_1 mantinha a mesma amplitude de pico m_p cor[illegible] sem pré-ênfase.[11] Esse resultado satisfazia a condição de uma largura de banda de transmissão [illegible].

A função de transferência de pré-ênfase é

$$H_p(f) = K\frac{j2\pi f + \omega_1}{j2\pi f + \omega_2} \tag{5.32a}$$

em que K, o ganho, é tomado com o valor ω_2/ω_1. Assim,

$$H_p(f) = \left(\frac{\omega_2}{\omega_1}\right)\frac{j2\pi f + \omega_1}{j2\pi f + \omega_2} \tag{5.32b}$$

Para $2\pi f << \omega_1$,

$$H_p(f) \simeq 1 \tag{5.32c}$$

Para frequências $\omega_1 << 2\pi f << \omega_2$,

$$H_p(f) \simeq \frac{j2\pi f}{\omega_1} \tag{5.32d}$$

Portanto, nas frequências intermediárias (2,1 – 15 kHz), o circuito de pré-ênfase atua como um diferenciador, o que, efetivamente, transforma o esquema em FM nessas frequências. Isso significa que FM com PDE é FM na faixa de frequências do sinal modulante, de 0 a 2,1 kHz, e quase PM na faixa de 2,1 kHz a 15 kHz, como desejado.

O filtro de deênfase $H_d(f)$ é dado por

$$H_d(f) = \frac{\omega_1}{j2\pi f + \omega_1}$$

Notemos que, para $2\pi f << \omega_2$, $H_p(f) \simeq (j2\pi f + \omega_1)/\omega_1$. Logo, $H_p(f)\ H_d(f) \simeq 1$ na faixa de frequências de 0 a 15 kHz.

Por motivos históricos e práticos, filtros ótimos PDE não são usados na vida real. Pode ser mostrado que a PDE aumenta a SNR em 13,27 dB (uma razão de potência de 21,25).

O efeito colateral benéfico de PDE é uma melhora na característica de interferência. Como a interferência (sinais indesejados e estações vizinhas) entra após o estágio transmissor, passa apenas pela operação de deênfase e não pela de amplificação ou pré-ênfase. Assim, nas frequências acima de 2,1 kHz, as amplitudes de interferências sofrem atenuação quase linear com a frequência.

O método PDE de redução de ruído não se limita à difusão FM. Também é usado na gravação de áudio e na gravação (analógica) fonográfica, nas quais o ruído de chiado também se concentra nas frequências altas. Um som de chiado claro é causado por irregularidades no material de gravação. Sistemas **Dolby de redução de ruído** para fitas de áudio têm o mesmo princípio de funcionamento, embora o sistema Dolby-A seja um pouco mais elaborado. Nos sistemas Dolby-B e C, a banda é dividida em duas sub-bandas (abaixo e acima de 3 kHz, em vez de 2,1 kHz). No sistema Dolby-A, projetado para uso comercial, as bandas são divididas em quatro sub-bandas (abaixo de 80 Hz, 80 Hz-3 kHz, 3-9 kHz e acima de 9 kHz). O grau de pré-ênfase é otimizado para cada banda.

Também podemos usar PDE na difusão AM para melhorar a SNR de saída. Contudo, na prática, isso não é feito por várias razões. Primeira, em AM, a amplitude do ruído de saída é constante com a frequência e não aumenta linearmente, como em FM. Portanto, a deênfase não resulta em um ganho tão grande em AM quanto no FM. Segunda, a introdução de PDE obrigaria modificações nos receptores já em uso. Terceira, o aumento da amplitude das componentes de alta frequência (pré-ênfase) aumentaria a interferência com estações adjacentes (esse problema não ocorre em FM). Além disso, um aumento na razão de desvio de frequência β nas altas frequências tornaria o projeto do detector mais difícil.

5.6 RECEPTORES ANALÓGICOS AM/FM SUPER-HETERÓDINOS

O receptor de rádio usado em sistemas de difusão AM e FM é chamado receptor **super-heteródino** (Fig. 5.17). O receptor consiste em uma seção de RF (radiofrequência), um conversor de frequência (Exemplo 4.2), um amplificador de frequência intermediária (FI), um detector de envelope e um amplificador de áudio.

A seção de RF consiste basicamente em um filtro sintonizável e em um amplificador que pega a estação desejada sintonizando o filtro à apropriada faixa de frequências. A próxima seção, o misturador (conversor) de frequências, traslada a portadora de ω_c a uma frequência FI fixa ω_{FI} (ver Exemplo 4.2 para conversão de frequências). Para esse fim, o receptor usa um oscilador local, cuja frequência f_{LO} é exatamente um valor f_{FI} maior que a frequência portadora de entrada, f_c; ou seja,

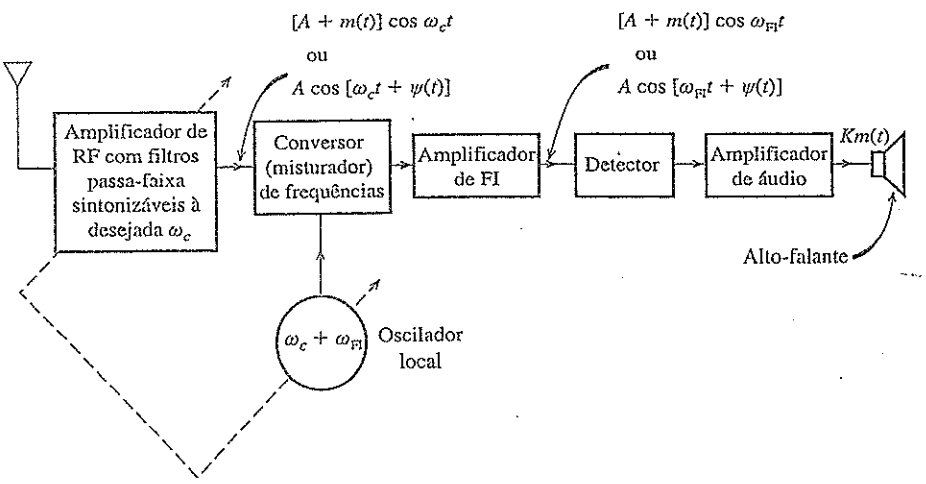
$$f_{LO} = f_c + f_{FI}$$

A sintonia simultânea do oscilador local e do filtro de RF sintonizável é feita por um único botão de sintonia. Capacitores de sintonia nos dois circuitos são agrupados e projetados de modo que a frequência de sintonia do oscilador local seja sempre f_{FI} Hz acima da frequência de sintonia f_c do filtro de RF. Isso significa que cada estação sintonizada é trasladada a uma frequência intermediária fixa de f_{FI} Hz pelo conversor de frequências, para processamento posterior na FI.

Essa estrutura do receptor super-heteródino é largamente utilizada na maioria dos sistemas de difusão. As frequências intermediárias são escolhidas em 455 kHz (rádio AM), 10,7 MHz (rádio FM) e 38 MHz (recepção de TV).

Como descoberto por Armstrong para sinais AM, a translação de todas as estações a uma frequência intermediária fixa (para AM, f_{FI} = 455 kHz) nos permite obter seletividade adequada. É difícil projetar filtros passa-faixa precisos com 10 kHz de largura de banda (espectro do sinal de áudio modulado) quando a frequência central f_c é muito alta. Isso é particularmente verdadeiro para o caso de filtros sintonizáveis. Assim, o filtro de RF não é capaz de oferecer seletividade adequada para evitar canais adjacentes. Contudo, quando trasladado a uma frequência intermediária por um conversor, o sinal é ainda amplificado por um amplificador de FI (em geral, um amplificador de três estágios), que tem boa seletividade. Isso é possível porque a frequência FI é razoavelmente baixa; além disso, a frequência central do amplificador de FI é fixa e sintonizada na fábrica. Portanto, a seção de FI é capaz de, efetivamente, eliminar a interferência entre canais adjacentes, devido à alta seletividade. Essa seção também amplifica o sinal para a detecção de envelope.

Figura 5.17 Receptor super-heteródino.



Na verdade, toda a seletividade é obtida na seção de FI; a seção de RF tem um papel secundário. A principal função da seção de RF é a supressão de frequência imagem. Como observado no Exemplo 4.2, a saída do misturador ou conversor consiste em componentes da diferença entre a frequência que chega ao receptor (f_c) e a do oscilador local (F_{LO}) (ou seja $f_{RF} = |f_{LO} - f_c|$). Agora, consideremos o exemplo de AM. Se a frequência portadora recebida for $f_c = 1000$ kHz, então $f_{LO} = f_c + f_{RF} = 1000 + 455 = 1455$ kHz. Contudo, outra portadora, com $f_c' = 1455 + 455 = 1910$ kHz, também será selecionada, pois a diferença $f_c' - f_{LO}$ é 455 kHz. Dizemos que a estação em 1910 kHz é a **imagem** da estação em 1000 kHz. Estações AM separadas em frequência por $2f_{FI} = 910$ kHz são denominadas **estações-imagem** e as duas apareceriam simultaneamente na saída de FI, se não fosse pela presença do filtro de RF na entrada do receptor. O filtro de RF pode ter seletividade pobre para estações adjacentes separadas por 10 kHz, mas tem seletividade razoável para estações separadas por 910 kHz. Assim, quando desejamos sintonizar uma estação em 1000 kHz, o filtro de RF, sintonizado em 1000 kHz, provê supressão adequada da estação-imagem em 1910 kHz.

Para converter a frequência portadora de entrada à FI, o receptor (Fig. 5.17) utiliza um oscilador local de frequência f_{LO} maior que a frequência portadora; por isso, recebe a denominação de receptor super-heteródino. Escolhemos f_{LO} maior que f_c porque isso resulta em menor razão de sintonia, considerando as frequências máxima e mínima de sintonia para o oscilador local. As frequências de difusão AM residem na faixa de 530 a 1710 kHz. A faixa da f_{LO} super-heteródina é de 1005 a 2055 kHz (razão de 2,045); a faixa da f_{LO} sub-heteródina é de 95 a 1145 kHz (razão de 12,05). É muito mais fácil projetar um oscilador sintonizável quando a razão de frequência é menor.

Não podemos deixar de ressaltar a importância do princípio super-heteródino na difusão de rádio e televisão. Inicialmente (antes de 1919), toda a seletividade para separar estações adjacentes estava no filtro de RF. Como, em geral, esse filtro tinha seletividade pobre, era necessário utilizar vários estágios (diversos circuitos ressonantes) em cascata para obter uma seletividade adequada. Nos primeiros receptores, cada filtro era sintonizado individualmente. A sintonia de uma estação, com a sincronização de todos os circuitos ressonantes, era uma tarefa tediosa e demorada. Isso foi facilitado com o agrupamento de capacitores variáveis, montados em um mesmo chassi, girado por um botão. Contudo, capacitores variáveis são volumosos e há um limite ao número de capacitores que podem ser agrupados. Esses fatores limitavam a seletividade de receptores. Em consequência, era necessária grande separação entre frequências portadoras adjacentes, o que resultava em menor número de bandas de frequências. O receptor super-heteródino permitiu a acomodação de muito mais estações de rádio.

5.7 SISTEMAS DE DIFUSÃO FM

A FCC alocou a faixa de frequências de 88 a 108 MHz para a difusão FM, com separação de 200 kHz entre estações adjacentes* e um desvio máximo de frequência $\Delta f = 75$ kHz.

Um receptor FM monofônico é idêntico ao receptor AM super-heteródino mostrado na Fig. 5.17, exceto que a frequência intermediária é 10,7 MHz, e o detector de envelope é substituído por uma PLL ou um discriminador de frequência seguido de um circuito de deênfase.

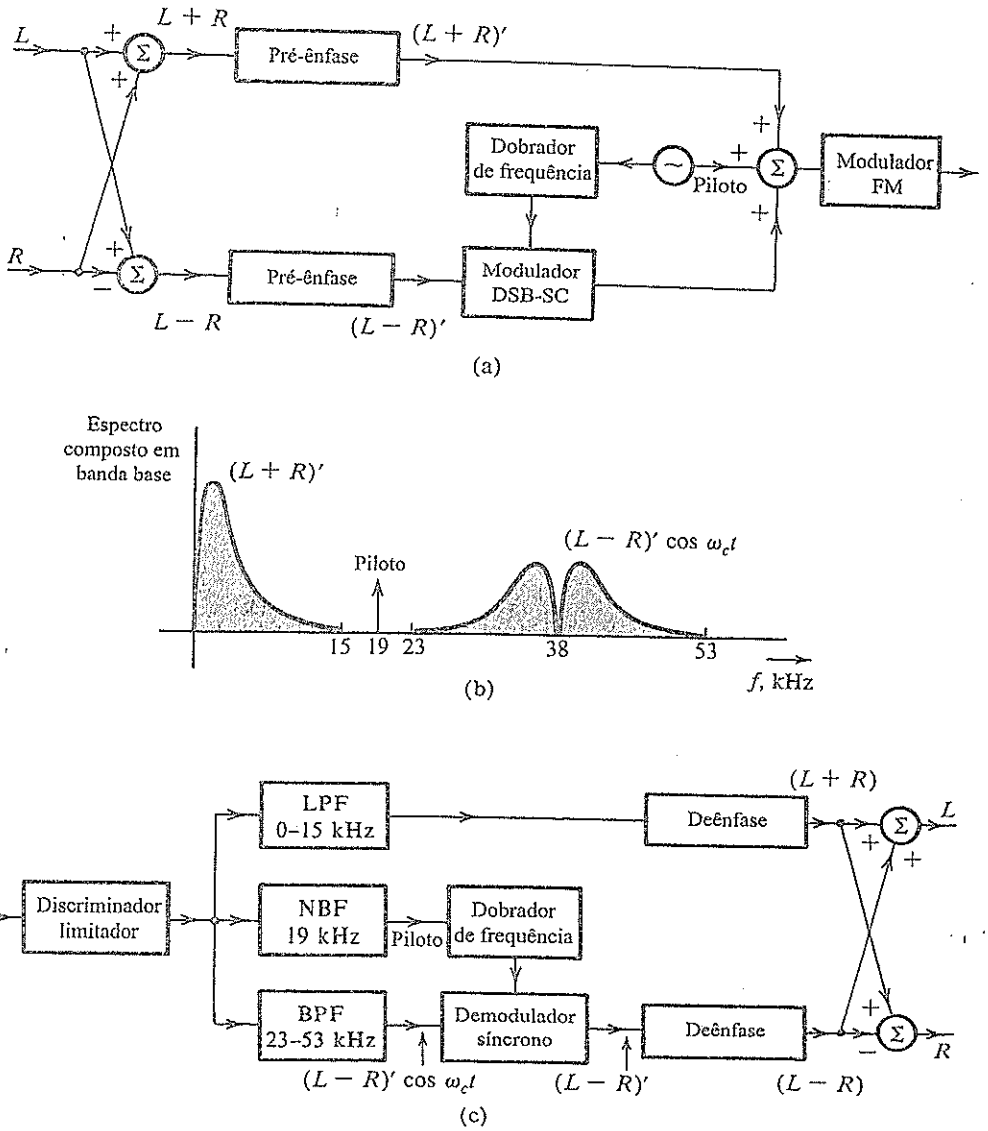
Inicialmente, as transmissões FM eram monofônicas. Transmissão FM estereofônica, na qual dois sinais de áudio – L (microfone da esquerda) e R (microfone da direita)** – são usados para um efeito mais natural, foi proposta posteriormente. A FCC estipulou que o sistema estereofônico devia ser compatível com o sistema monofônico original. Isso significava que os antigos receptores monofônicos deveriam ser capazes de receber o sinal $L + R$, e a largura de banda total de transmissão para os dois canais (L e R) deve permanecer em 200 kHz, com $\Delta f = 75$ kHz para os dois sinais combinados. Isso asseguraria que receptores antigos seriam capazes de receber tanto sinais monofônicos como estereofônicos, embora o efeito estéreo ficasse ausente.

---

*No Brasil, a Agência Nacional de Telecomunicações – ANATEL adota o mesmo plano de frequências. (N.T.)

**As letras L e R são herdadas do inglês: *Left* (esquerdo) e *Right* (direito). (N.T.)

Figura 5.18
(a) Transmissor FM estéreo.
(b) Espectro de um sinal estéreo em banda base.
(c) Receptor FM estéreo.



Um transmissor e um receptor para difusão estereofônica são mostrados na Fig. 5.18a e c. No transmissor, os dois sinais L e R são somados e subtraídos, fornecendo $L + R$ e $L - R$. Esses sinais passam por pré-ênfase. Após a pré-ênfase, o sinal $(L - R)'$ modula, em DSB-SC, uma portadora de 38 kHz, obtida do dobramento de um sinal de 19 kHz, usado como piloto. O sinal $(L + R)'$ é usado diretamente. Os três sinais (o terceiro é o piloto) formam um sinal em banda base compósto $m(t)$ (Fig. 5.18b),

$$m(t) = (L+R)' + (L-R)' \cos \omega_c t + \alpha \cos \frac{\omega_c t}{2} \qquad (5.33)$$

A razão para o uso do piloto de 19 kHz, em vez de 38 kHz, é a facilidade em separar o piloto de 19 kHz, pois não há componentes de sinal na faixa de 4 kHz nos dois lados dessa frequência.

O funcionamento do receptor (Fig. 5.18c) é autoexplicativo. Um receptor monofônico consiste apenas no braço superior do receptor estéreo e, portanto, recebe somente o sinal $L + R$. Esse é o sinal

de áudio completo, sem o efeito estéreo. O sistema, então, é compatível. O piloto é extraído e (depois de ter a frequência dobrada) usado para demodular coerentemente o sinal $(L-R)' \cos \omega_c t$.

Um aspecto interessante da transmissão estéreo é que a amplitude de pico do sinal composto $m(t)$ na Eq. (5.33) é praticamente igual à do sinal monofônico (se ignorarmos o piloto) e, portanto, o valor de Δf – que, na transmissão estereofônica, é proporcional à amplitude de pico do sinal – permanece praticamente igual ao do caso monofônico. Isso pode ser explicado pelo chamado efeito de **entrelaçamento** (*interleaving*), discutido a seguir.

Os sinais L' e R' são, em geral, muito parecidos. Assim, podemos supor que suas amplitudes de pico são iguais a A_p. Na pior das hipóteses, L' e R' alcançam seus valores de pico ao mesmo tempo, resultando em [Eq. (5.33)]

$$|m(t)|_{\text{máx}} = 2A_p(1-\alpha)$$

No caso monofônico, a amplitude de pico do sinal em banda base $(L+R)'$ é $2A_p$. Logo, as amplitudes de pico nos dois casos diferem apenas de α, a amplitude do piloto. Para levar isso em consideração, a amplitude de pico de áudio no caso estéreo é reduzida a 90% do valor total. Isso corresponde a uma redução na potência de sinal por uma razão $(0{,}9)^2 = 0{,}81$ ou 1 dB. Com isso, a SNR efetiva é reduzida em 1 dB, devido à inclusão do piloto.

5.8 EXERCÍCIOS COM O MATLAB

Nesta seção, usamos MATLAB para construir um exemplo de modulação e demodulação FM. O programa MATLAB é `ExampleFM.m`. Mais uma vez, aplicamos o mesmo sinal de mensagem $m_2(t)$. O coeficiente FM é $k_f = 80$ e o coeficiente PM, $k_p = \pi$. A frequência portadora permanece sendo 300 Hz. Os resultantes sinais FM e PM no domínio do tempo são mostrados na Fig. 5.19. As correspondentes respostas de frequência também são ilustradas na Fig. 5.19. As respostas no domínio da frequência mostram claramente as maiores larguras de banda dos sinais FM e PM, em comparação com o sinal modulado em amplitude.

```
% (ExampleFM.m)
% Este programa usa triangl.m para ilustrar modulação
% e demodulação em frequência
ts=1.e-4;

t=-0.04:ts:0.04;
Ta=0.01;
m_sig=triangl((t+0.01)/Ta)-triangl((t-0.01)/Ta);
Lfft=length(t);  Lfft=2^ceil(log2(Lfft));
M_fre=fftshift(fft(m_sig,Lfft));
freqm=(-Lfft/2:Lfft/2-1)/(Lfft*ts);
B_m=100;  %Bandwidth of the signal is B_m Hz.
% Design a simple lowpass filter with bandwidth B_m Hz.
h=fir1(80,[B_m*ts]);
%
kf=160*pi;
m_intg=kf*ts*cumsum(m_sig);
s_fm=cos(2*pi*300*t+m_intg);
s_pm=cos(2*pi*300*t+pi*m_sig);
Lfft=length(t);  Lfft=2^ceil(log2(Lfft)+1);
S_fm=fftshift(fft(s_fm,Lfft));
S_pm=fftshift(fft(s_pm,Lfft));
```

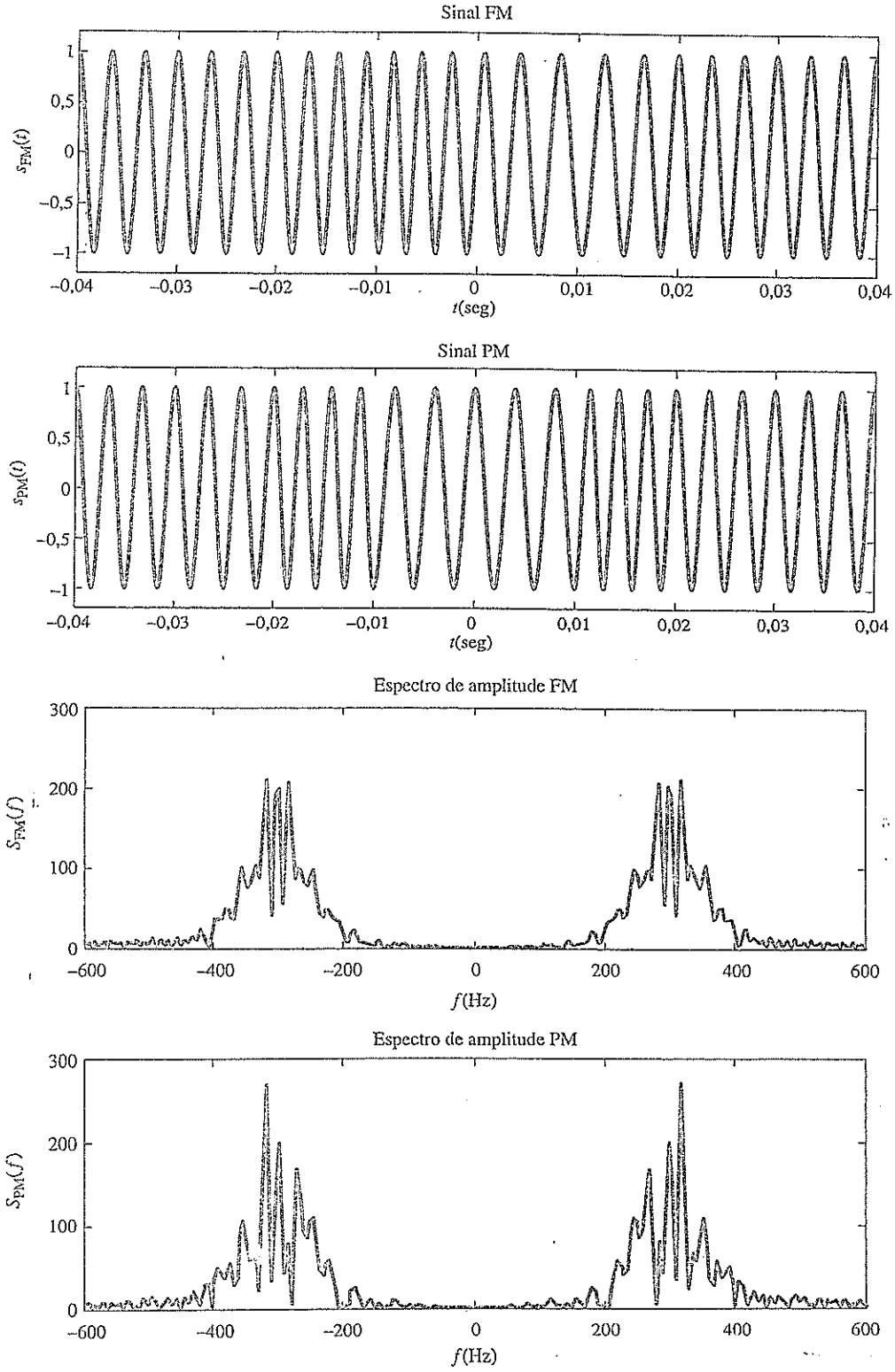


Figura 5.19
Sinais FM e PM nos domínios do tempo e da frequência.

```
freqs=(-Lfft/2:Lfft/2-1)/(Lfft*ts);

s_fmdem=diff([s_fm(1) s_fm])/ts/kf;
s_fmrec=s_fmdem.*(s_fmdem>0);
```

```
s_dec=filter(h,1,s_fmrec);

% Demodulação
% Usando um filtro passa-baixos ideal com largura de banda de 200 Hz

Trange1=[-0.04 0.04 -1.2 1.2];

figure(1)
subplot(211);m1=plot(t,m_sig);
axis(Trange1); set(m1,'Linewidth',2);
xlabel('{\it t} (seg)'); ylabel('{\it m}({\it t})');
title('Sinal de mensagem');
subplot(212);m2=plot(t,s_dec);
set(m2,'Linewidth',2);
xlabel('{\it t} (seg)'); ylabel('{\it m}_d({\it t})')
title('Sinal FM demodulado');

figure(2)
subplot(211);td1=plot(t,s_fm);
axis(Trange1); set(td1,'Linewidth',2);
xlabel('{\it t} (seg)'); ylabel('{\it s}_{\rm FM}({\it t})')
title('sinal FM');
subplot(212);td2=plot(t,s_pm);
axis(Trange1); set(td2,'Linewidth',2);
xlabel('{\it t} (seg)'); ylabel('{\it s}_{\rm PM}({\it t})')
title('Sinal PM');

figure(3)
subplot(211);fp1=plot(t,s_fmdem);
set(fp1,'Linewidth',2);
xlabel('{\it t} (seg)'); ylabel('{\it d s}_{\rm FM}({\it t})/dt')
title('derivada FM');
subplot(212);fp2=plot(t,s_fmrec);
set(fp2,'Linewidth',2);
xlabel('{\it t} (seg)');
title('derivada FM retificada');

Frange=[-600 600 0 300];
figure(4)
subplot(211);fd1=plot(freqs,abs(S_fm));
axis(Frange); set(fd1,'Linewidth',2);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{\rm FM}({\it f})')
title('espectro de amplitude FM');
subplot(212);fd2=plot(freqs,abs(S_pm));
axis(Frange); set(fd2,'Linewidth',2);
xlabel('{\it f} (Hz)'); ylabel('{\it S}_{\rm PM}({\it f})')
title('espectro de amplitude PM');
```

Para obter os resultados de demodulação, um diferenciador é, primeiro, aplicado para transformar o sinal modulado em frequência em um sinal modulado em amplitude e em frequência (Fig. 5.20). Com o uso do retificador para detecção de envelope, vemos que o sinal de mensagem segue de perto as variações do envelope da saída do retificador.

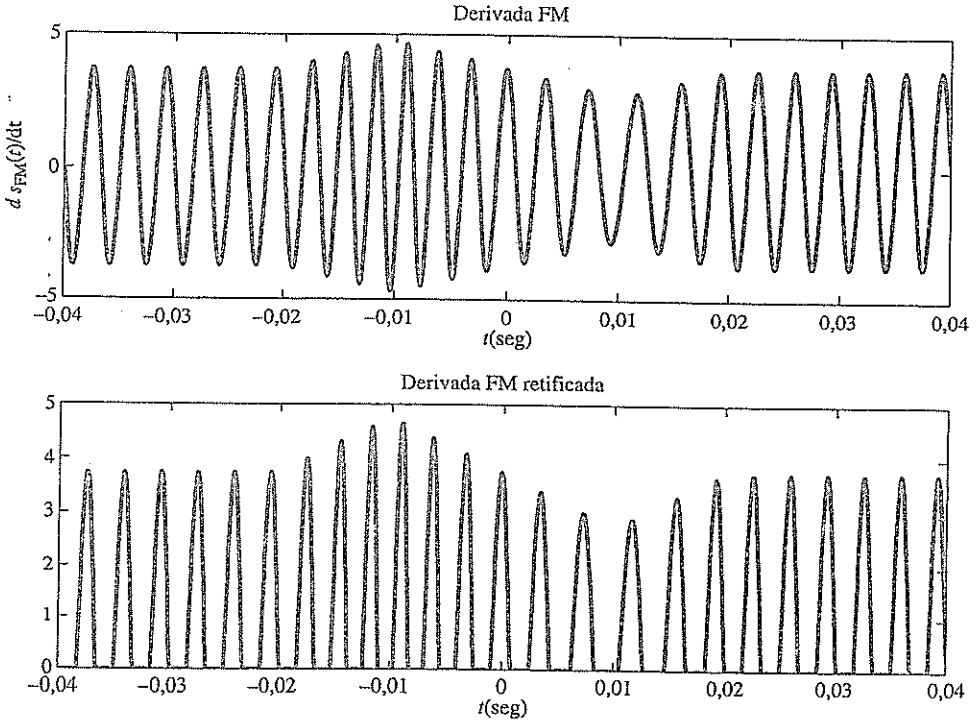


Figura 5.20 Sinais no demodulador: (a) após o diferenciador, (b) após o retificador.

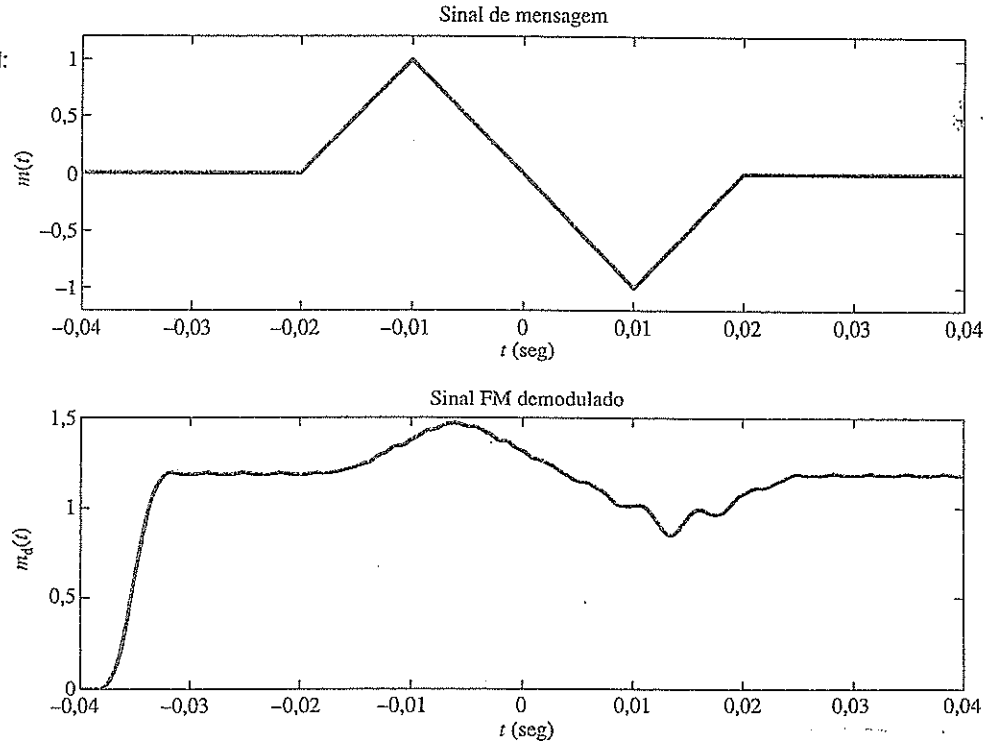


Figura 5.21 Modulação e demodulação FM: (a) mensagem original; (b) sinal recuperado.

Por fim, o sinal de saída do retificador é aplicado a um filtro passa-baixos com largura de banda de 100 Hz. Neste exemplo, por conta dos requisitos mais estritos de filtragem, usamos um filtro passa-baixos de ordem 80, com resposta ao impulso finita. Na Fig. 5.21, a saída do detector FM é comparada ao sinal de mensagem original.

Os resultados da demodulação FM mostram claramente algumas distorções. Primeiro, o tempo de resposta e o atraso do filtro passa-baixos de ordem mais alta são muito maiores. Segundo, a distorção durante a metade negativa da mensagem é mais severa porque o retificador gera poucos ciclos da meia senoide. Isso ocorre porque, quando o sinal de mensagem é negativo, a frequência instantânea do sinal FM é baixa. Como usamos uma frequência portadora de apenas 300 Hz, o efeito da baixa frequência instantânea é muito mais pronunciado. Se aplicássemos uma frequência portadora prática de 100 MHz, esse tipo de distorção seria totalmente desprezível.

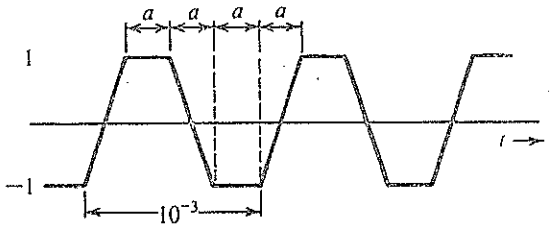
REFERÊNCIAS

1. J. Carson, "Notes on the Theory of Modulation," *Proc. IRE*, vol. 10, pp. 57–64, Feb. 1922.
2. J. Carson, "Reduction of Atmospheric Disturbances," *Proc. IRE*, vol. 16, July 1928.
3. E. H. Armstrong, "A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation," *Proc. IRE*, vol. 24, pp. 689–740, May 1936.
4. "A Revolution in Radio," *Fortune*, vol. 20, p. 116, Oct. 1939.
5. L. Lessing, *Man of High Fidelity: Edwin Howard Armstrong*, Lippincott, Philadelphia, 1956.
6. H. R. Slotten, "'Rainbow in the Sky': FM Radio Technical Superiority, and Regulatory Decision Making," Society for the History of Technology, 1996.
7. J. E. Brittain, "Electrical Engineering Hall of Fame—Edwin H. Armstrong," *Proc. IEEE*, vol 92, pp. 575–578, Mar. 2004.
8. W. B. Davenport, Jr., "Signal-to-Noise Ratios in Bandpass Limiters," *J. Appl. Phys.*, vol. 24, pp. 720–727, June 1953.
9. D. H. Sheingold, ed., *Nonlinear Circuits Handbook*, Analog Devices, Inc., Norwood, MA, 1974.
10. H. L. Krauss, C. W. Bostian, and F. H. Raab, *Solid-State Radio Engineering*, Wiley, New York, 1980.
11. 12. L. B. Arguimbau and R. B. Adler, *Vacuum Tube Circuits and Transistors*, Wiley, New York, 1964, p. 466.

EXERCÍCIOS

5.1-1 Esboce as formas de onda de $\varphi_{FM}(t)$ e $\varphi_{PM}(t)$ para o sinal modulante $m(t)$ mostrado na Fig. E.5.1-1, dados $\omega_c = 10^8$ rad/s, $k_f = 10^5$ e $k_p = 25$.

Figura E.5.1-1

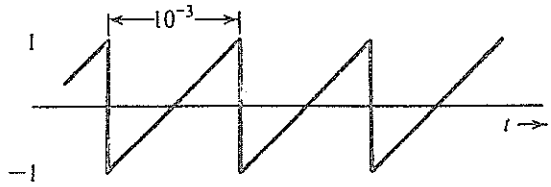


5.1-2 Um sinal em banda base $m(t)$ é o sinal dente de serra periódico mostrado na Fig. E.5.1-2.

(a) Esboce as formas de onda de $\varphi_{FM}(t)$ e $\varphi_{PM}(t)$ para este sinal $m(t)$, dados $\omega_c = 2\pi \times 10^6$ rad/s, $k_f = 2000\pi$ e $k_p = \pi/2$.

(b) Mostre que o sinal PM é equivalente a um sinal PM modulado por uma mensagem quadrada periódica. Explique por que, neste caso, é necessário usar $k_p < \pi$. [Note que o sinal PM tem uma frequência constante, mas tem descontinuidades de fase que correspondem às descontinuidades de $m(t)$.]

Figura E.5.1-2



5.1-3 No intervalo $|t| \leq 1$, um sinal modulado em ângulo é dado por

$$\varphi_{EM}(t) = 10 \cos 13.000\pi t$$

A frequência portadora é $\omega_c = 10.000\pi$ rad/s.

(a) Admitindo que este é um sinal PM com $k_p = 1000$, determine $m(t)$ no intervalo $|t| \leq 1$.

(b) Admitindo que este é um sinal FM com $k_f = 1000$, determine $m(t)$ no intervalo $|t| \leq 1$.

5.2-1 Para o sinal de mensagem

$$m(t) = 2\cos 100t + 18\cos 2000\pi t$$

(a) Escreva as expressões (sem desenhar gráficos) de $\varphi_{PM}(t)$ e $\varphi_{FM}(t)$ com $A = 10$, $\omega_c = 10^6$ rad/s, $k_f = 1000\pi$, $k_p = 1$. Para determinar $\varphi_{FM}(t)$, use a integral indefinida de $m(t)$, ou seja, considere o valor da integral em $t = -\infty$ como 0.

(b) Estime as larguras de banda de $\varphi_{FM}(t)$ e $\varphi_{PM}(t)$.

5.2-2 Um sinal modulado em ângulo com frequência central $\omega_c = 2\pi \times 10^6$ rad/s é descrito pela equação

$$\varphi_{EM}(t) = 10 \cos(\omega_c t + 0{,}1 \operatorname{sen} 2000\pi t)$$

(a) Calcule a potência do sinal modulado.

(b) Calcule o desvio de frequência Δf.

(c) Calcule o desvio de fase $\Delta\phi$.

(d) Estime a largura de banda de $\varphi_{EM}(t)$.

5.2-3 Repita o Exercício 5.2-2 para o caso em que

$$\varphi_{EM}(t) = 5\cos(\omega_c t + 20 \operatorname{sen} 1000\pi t + 10 \operatorname{sen} 2000\pi t)$$

5.2-4 Estime as larguras de banda de $\varphi_{PM}(t)$ e $\varphi_{FM}(t)$ no Exercício 5.1-1. Suponha que a largura de banda de $m(t)$ na Fig. E.5.1-1 seja igual à frequência do terceiro harmônico de $m(t)$.

5.2-5 Estime as larguras de banda de $\varphi_{PM}(t)$ e $\varphi_{FM}(t)$ no Exercício 5.1-2. Suponha que a largura de banda de $m(t)$ na Fig. E.5.1-2 seja igual à frequência do quinto harmônico de $m(t)$.

5.2-6 Dados $m(t) = \operatorname{sen} 2000\pi t$, $k_f = 200.000\pi$ e $k_p = 10$.

(a) Estime as larguras de banda de $\varphi_{FM}(t)$ e $\varphi_{PM}(t)$.

(b) Repita a parte (a) para o caso em que a amplitude do sinal de mensagem é dobrada.

(c) Repita a parte (a) para o caso em que a frequência do sinal de mensagem é dobrada.

(d) Comente a sensibilidade das larguras de banda de FM e PM em relação ao espectro de $m(t)$.

5.2-7 Dados $m(t) = e^{-t^2}$ $f_c = 10^4$ Hz, $k_f = 6000\pi$ e $k_p = 8000\pi$.

(a) Calcule Δf, o desvio de frequência para FM e PM.

(b) Estime as larguras de bandas das ondas FM e PM.

Sugestão: Determine $M(f)$ e sua largura de banda de 3 dB ($B << \Delta f$).

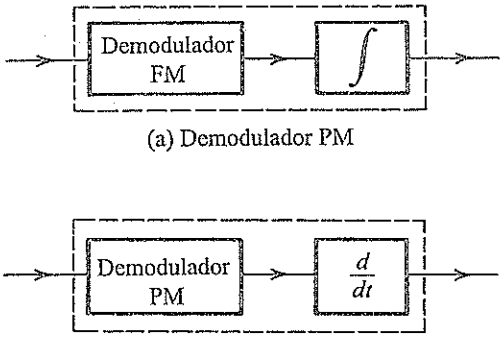
5.3-1 Projete (apenas o diagrama em blocos) um modulador FM indireto de Armstrong para gerar uma portadora FM com frequência de 98,1 MHz e $\Delta f = 75$ Hz. Há disponibilidade de um gerador de FM de banda estreita, com frequência portadora de 100 kHz e desvio de frequência $\Delta f = 10$ Hz. No estoque, há um oscilador de frequência ajustável na faixa de 10 a 11 MHz. Há, ainda, diversos dobradores, triplicadores e quintuplicadores de frequência.

5.3-2 Projete (apenas o diagrama em blocos) um modulador FM indireto de Armstrong para gerar uma portadora FM com frequência de 96 MHz e $\Delta f = 20$ Hz. Há disponibilidade de um gerador de FM de banda estreita, com $f_c = 200$ kHz e desvio de frequência Δf ajustável na faixa de 9 a 10 Hz. No estoque, há um oscilador de frequência ajustável na faixa de 9 a 10 MHz. Há, ainda, um filtro passa-faixa com frequência central arbitrária e dobradores de frequência.

5.4-1 (a) Mostre que, quando m(t) não tem descontinuidades do tipo degrau, um demodulador FM seguido por um integrador (Fig. E.5.4-1) forma um demodulador PM. Explique por que é necessário que o demodulador PM remova qualquer componente dc antes do integrador.

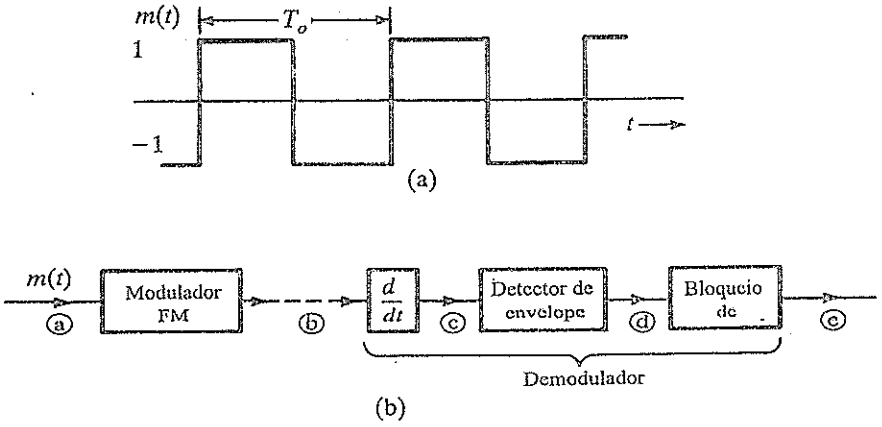
(b) Mostre que um demodulador PM seguido por um diferenciador (Fig. 5.4.1b) funciona como um demodulador FM mesmo que $m(t)$ tenha descontinuidades do tipo degrau ou que a saída do demodulador PM tenha componente dc.

Figura E. 5.4-1



5.4-2 Uma onda quadrada periódica $m(t)$ (Fig. E.5.4-2a) modula em frequência uma portadora $f_c = 10$ kHz, com $\Delta f = 1$ kHz. A amplitude da portadora é A. O resultante sinal FM é demodulado, como mostrado na Fig. E.5.4-2b, pelo método discutido na Seção 5.4 (Fig. 5.12). Esboce as formas de onda nos pontos b, c, d e e.

Figura E.5.4-2



5.4-3 Use a análise de PLL sob pequeno erro para mostrar que uma malha de primeira ordem $[H(s) = 1]$ não pode rastrear um sinal de entrada cuja frequência instantânea varia linearmente com o tempo $[\theta_i(t) = kt^2]$. Esse sinal pode ser rastreado, a menos de uma constante de fase, se $H(s) = (s + a)/s$. O sinal pode ser rastreado sem erro de fase caso $H(s) = (s^2 + as + b)/s^2$.

5.6-1 Um transmissor transmite um sinal AM com frequência portadora de 1500 kHz. Quando um receptor de rádio barato (cujo filtro passa-faixa do estágio de RF tem seletividade pobre) é sintonizado a 1500 kHz, o sinal é ouvido forte e claro. Esse mesmo sinal também é ouvido (não tão bem) em outra sintonia. Diga, com justificativa, em que frequência essa estação será ouvida. A frequência FI é de 455 kHz.

5.6-2 Considere um receptor FM super-heteródino projetado para receber a faixa de frequências de 1 a 30 MHz, com frequência FI de 8 MHz. Qual é a faixa de frequências geradas pelo oscilador local para esse receptor? Um sinal de entrada com frequência portadora de 10 MHz é recebido na sintonia de 10 MHz. Nessa sintonia do receptor, também é recebida interferência de um sinal com outra frequência portadora, caso o filtro passa-faixa do estágio de RF do receptor tenha seletividade pobre. Qual é a frequência portadora do sinal interferente?

5.6-2 Considere um receptor FM super-heteródino projetado para receber a faixa de frequências de 1 a 30 MHz, com frequência FI de 8 MHz. Qual é a faixa de frequências geradas pelo oscilador local para esse receptor? Um sinal de entrada com frequência portadora de 10 MHz é recebido na sintonia de 10 MHz. Nessa sintonia do receptor, também é recebida interferência de um sinal com outra frequência portadora, caso o filtro passa-faixa do estágio de RF do receptor tenha seletividade pobre. Qual é a frequência portadora do sinal interferente?

6 AMOSTRAGEM E CONVERSÃO ANALÓGICO-DIGITAL

Como discutido brevemente no Capítulo 1, sinais analógicos podem ser digitalizados por meio de amostragem e quantização. Essa conversão analógico-digital (A/D) estabelece o sustentáculo dos modernos sistemas de comunicação digital. No conversor A/D, a taxa de amostragem deve ser suficientemente grande para permitir que o sinal analógico seja reconstruído das amostras com precisão adequada. O **teorema da amostragem**, que é a base para a determinação da apropriada (sem perda) taxa de amostragem para um dado sinal, tem um papel essencial no processamento de sinais, teoria das comunicações e projeto de circuitos A/D.

6.1 TEOREMA DA AMOSTRAGEM

Primeiro, mostremos que um sinal $g(t)$ cujo espectro é limitado em banda a B Hz, ou seja,

$$G(f) = 0 \quad \text{para } |f| > B$$

pode ser reconstruído exatamente (sem qualquer erro) a partir de suas amostras em tempo discreto tomadas uniformemente a uma taxa de R amostras por segundo. A condição é que $R > 2B$. Em outras palavras, a mínima frequência de amostragem para recuperação perfeita do sinal é $f_s = 2B$ Hz.

Para provar o teorema da amostragem, consideremos um sinal $g(t)$ (Fig. 6.1a) cujo espectro seja limitado em banda a B Hz (Fig. 6.1b).* Por conveniência, espectros são mostrados como funções tanto de f como de ω. Amostrar $g(t)$ a uma taxa de f_s Hz significa tomar f_s amostras uniformes por segundo. Tal amostragem uniforme é efetuada com a multiplicação de $g(t)$ por um trem de impulsos $\delta_{T_s}(t)$ da Fig. 6.1c, que consiste em impulsos unitários repetidos periodicamente a cada T_s segundos, em que $T_s = 1/f_s$. Isso resulta no sinal amostrado $\overline{g}(t)$ ilustrado na Fig. 6.1d. O sinal amostrado consiste em impulsos espaçados a cada T_s segundos (o intervalo de amostragem). O n-ésimo impulso, localizado em $t = nT_s$, tem amplitude $g(nT_s)$, que é o valor de $g(t)$ em $t = nT_s$. Assim, a relação entre o sinal amostrado $\overline{g}(t)$ e o sinal analógico original $g(t)$ é

$$\overline{g}(t) = g(t)\delta_{T_s}(t) = \sum_n g(nT_s)\delta(t - nT_s) \tag{6.1}$$

*O espectro $G(f)$ na Fig. 6.1b é mostrado como real, por conveniência. Os argumentos são igualmente válidos para $G(f)$ complexo.

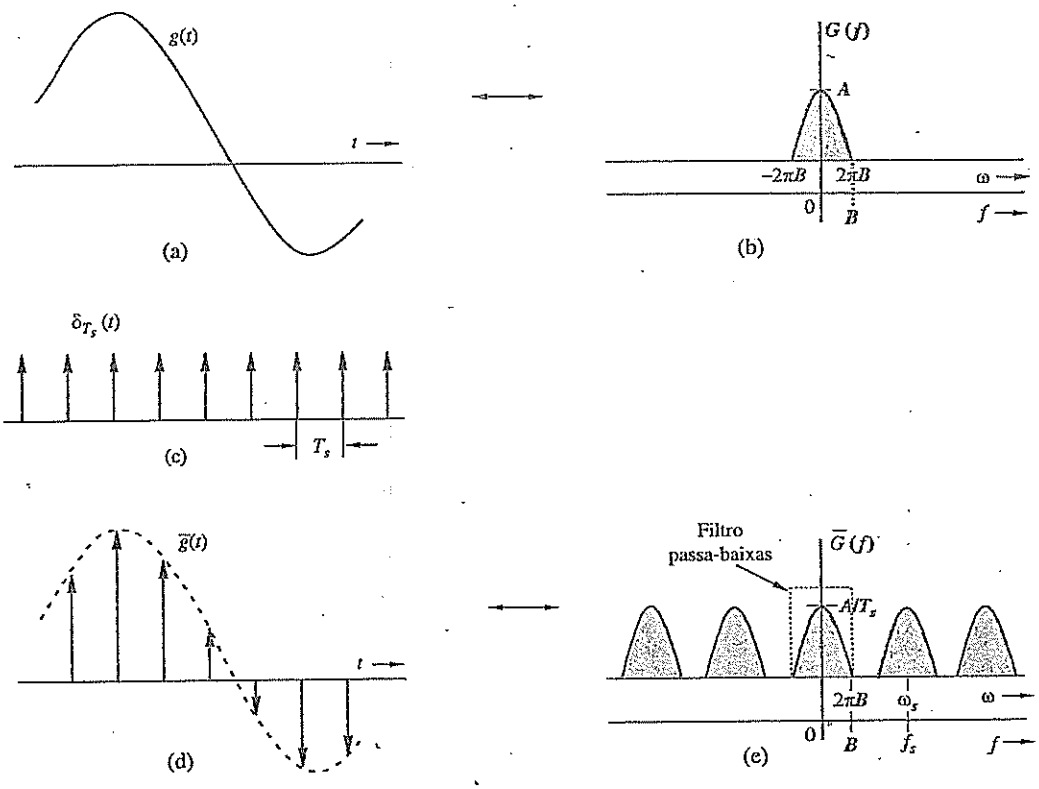


Figura 6.1 Sinal amostrado e seu espectro de Fourier.

Como o trem de impulsos $\delta_{T_s}(t)$ é um sinal periódico de período T_s, pode ser expresso como uma série de Fourier exponencial, já calculada no Exemplo 3.11, como

$$\delta_{T_s}(t) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} e^{jn\omega_s t} \qquad \omega_s = \frac{2\pi}{T_s} = 2\pi f_s \tag{6.2}$$

Logo,

$$\begin{aligned}\overline{g}(t) &= g(t)\delta_{T_s}(t)\\ &= \frac{1}{T_s}\sum_{n=-\infty}^{\infty} g(t)e^{jn2\pi f_s t}\end{aligned} \tag{6.3}$$

Para calcular $\overline{G}(f)$, a transformada de Fourier de $\overline{g}(t)$, tomamos a transformada de Fourier do somatório na Eq. (6.3). Com base na propriedade de translação de frequência, a transformada do n-ésimo termo é deslocada de nf_s. Portanto,

$$\overline{G}(f) = \frac{1}{T_s}\sum_{n=-\infty}^{\infty} G(f - nf_s) \tag{6.4}$$

Isso significa que o espectro $\overline{G}(f)$ consiste em $G(f)$ multiplicado por uma constante $1/T_s$ e repetido periodicamente a cada $f_s = 1/T_s$, como ilustrado na Fig. 6.1e.

Após a amostragem uniforme que gera um conjunto de amostras $\{g(kT_s)\}$, a questão vital se torna: **Será possível reconstruir $g(t)$ a partir de $\overline{g}(t)$ sem perda ou distorção?** Se formos capazes de reconstruir $g(t)$ de $\overline{g}(t)$, no domínio da frequência, equivalentemente, devemos ser capazes de recu-

perar $G(f)$ de $\overline{G}(f)$. Da Fig. 6.1, vemos que a recuperação perfeita é possível se não houver sobreposição entre as réplicas de $\overline{G}(f)$. A Fig. 6.1e mostra claramente que isso requer

$$f_s > 2B \tag{6.5}$$

O intervalo de amostragem é, então, $T_s = 1/f_s$. Portanto,

$$T_s < \frac{1}{2B} \tag{6.6}$$

Assim, desde que a frequência de amostragem f_s seja maior que o dobro da largura de banda B (em hertz), $\overline{G}(f)$ consistirá em repetições de $G(f)$ que não se sobrepõem. Quando isso é verdadeiro, a Fig. 6.1e mostra que $\overline{g}(t)$ pode ser recuperado de suas amostras $\overline{g}(t)$ aplicando o sinal amostrado $\overline{g}(t)$ a um filtro passa-baixas de largura de banda B. A mínima taxa de amostragem $f_s = 2B$ exigida para recuperar $g(t)$ de $\overline{g}(t)$ é denominada **taxa de Nyquist** para $g(t)$ e o correspondente intervalo de amostragem $T_s = 1/2B$, **intervalo de Nyquist** para o sinal passa-baixas $g(t)$.*

É necessário ressaltar um ponto importante relativo à possibilidade $f_s = 2B$ e a uma particular classe de sinais passa-baixas. Para um espectro de sinal genérico, provamos que a taxa de amostragem $f_s > 2B$. Contudo, caso o espectro $G(f)$ não contenha impulso (ou suas derivadas) na maior frequência B, não haverá sobreposição desde que a taxa de amostragem seja maior que ou igual à taxa de Nyquist, ou seja

$$f_s \geq 2B$$

Se, no entanto, $G(f)$ contiver um impulso na frequência mais alta $\pm B$, a igualdade deve ser removida ou ocorrerá sobreposição. Nesse caso, a taxa de amostragem f_s deve ser maior do que $2B$ Hz: Um exemplo conhecido é a senoide $g(t) = \text{sen } 2\pi B\,(t - t_0)$. Esse sinal é limitado à banda de B Hz, mas se as amostras são uniformemente tomadas à taxa $f_s = 2B$ (iniciando em $t = t_0$), todas serão nulas, e $g(t)$ não poderá ser recuperado de suas amostras de Nyquist. Portanto, para senoides, a condição $f_s > 2B$ deve ser satisfeita.

6.1.1 Reconstrução de Sinais a partir de Amostras Uniformes

O processo de reconstrução de um sinal do tempo contínuo $g(t)$ a partir de suas amostras também é conhecido com **interpolação**. Na Fig. 6.1, usamos uma prova construtiva para mostrar que um sinal $g(t)$ limitado em banda a B Hz pode ser reconstruído (interpolado) exatamente de suas amostras. Isso significa não apenas que amostragem uniforme feita acima da taxa de Nyquist preserva a informação do sinal, mas também que a simples aplicação do sinal amostrado a um filtro passa-baixas ideal de largura de banda B Hz reconstrói a mensagem original. Como visto na Eq. (6.3), o sinal amostrado contém uma componente $(1/T_s)g(t)$ e, para recuperar $g(t)$ [ou $G(f)$], o sinal amostrado

$$\overline{g}(t) = \sum g(nT_s)\delta(t - nT_s)$$

deve ser aplicado a um filtro passa-baixas ideal de largura de banda B Hz e ganho T_s. Um filtro ideal como esse tem a função de transferência

$$H(f) = T_s\,\Pi\left(\frac{\omega}{4\pi B}\right) = T_s\,\Pi\left(\frac{f}{2B}\right) \tag{6.7}$$

---

*O teorema enunciado (e provado) aqui se aplica a sinais passa-baixas. Um sinal passa-faixa, cujo espectro exista em uma banda de frequências $f_c - B/2 < |f| < f_c + B/2$, tem uma largura de banda B Hz. Um sinal desse tipo também pode ser recuperado exatamente de amostras tomadas à frequência de Nyquist $2B$ definida acima. Neste caso, em geral, o teorema da amostragem é mais complexo e faz uso de dois trens de amostragem uniforme entrelaçados, cada um na metade da taxa total de amostragem $R_s > B$. As referências 1 e 2 podem ser consultadas para mais detalhes.

Reconstrução Ideal

A Fig. 6.2a mostra a função de transferência do filtro ideal de interpolação, obtida na Eq. (6.7), para recuperar o sinal analógico de suas amostras uniformes. A resposta ao impulso desse filtro, transformada de Fourier inversa de $H(f)$, é

$$h(t) = 2BT_s \operatorname{sinc}(2\pi Bt) \tag{6.8}$$

Considerando o uso da taxa de amostragem de Nyquist, ou seja, $2BT_s = 1$, temos

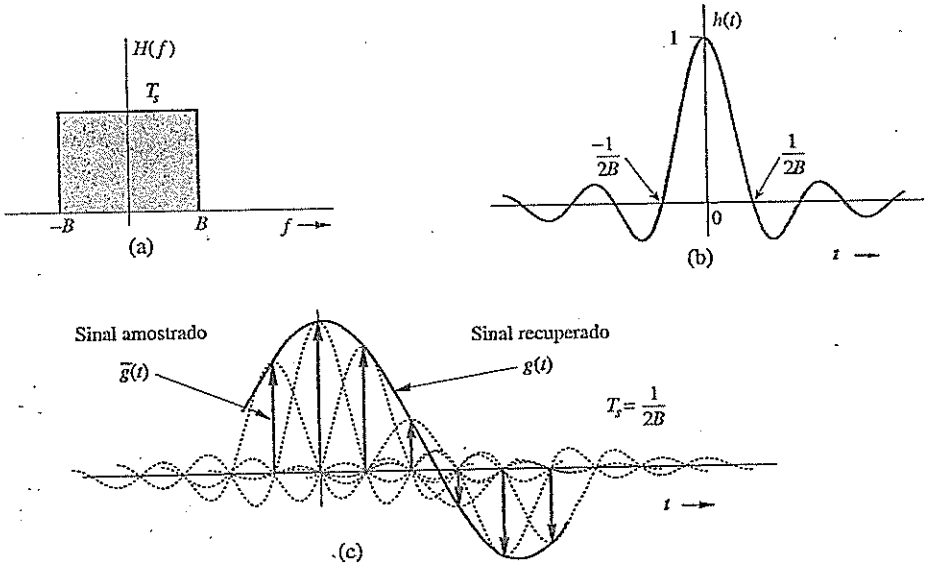
$$h(t) = \operatorname{sinc}(2\pi Bt) \tag{6.9}$$

Esta função $h(t)$ é mostrada na Fig. 6.2b. Um fato muito interessante a ser observado é que $h(t) = 0$ em todos os instantes de amostragem de Nyquist ($t = \pm n/2B$), exceto em $t = 0$. Quando o sinal amostrado $\overline{g}(t)$ é aplicado à entrada deste filtro, a saída é $g(t)$. Cada amostra em $\overline{g}(t)$, por ser um impulso, gera um pulso sinc de altura igual à intensidade da amostra, como ilustrado na Fig. 6.2c. O processo é idêntico ao mostrado na Fig. 6.6, exceto que $h(t)$ é um pulso sinc, em vez de um pulso retangular. A adição de pulsos sinc gerados por todas as amostras resulta em $g(t)$. A k-ésima amostra da entrada $\overline{g}(t)$ é o impulso $g(kT_s)\delta(t - kT_s)$; a saída do filtro para esse impulso é $g(kT_s)h(t - kT_s)$. Logo, a saída do filtro para $\overline{g}(t)$, que é $g(t)$, pode ser expressa como uma soma

$$\begin{aligned} g(t) &= \sum_k g(kT_s)h(t - kT_s) \\ &= \sum_k g(kT_s) \operatorname{sinc}[2\pi B(t - kT_s)] && (6.10a) \\ &= \sum_k g(kT_s) \operatorname{sinc}(2\pi Bt - k\pi) && (6.10b) \end{aligned}$$

A Eq. (6.10) é a **fórmula de interpolação**, que produz valores de $g(t)$ entre amostras como uma soma ponderada de todos os valores de amostras.

Figura 6.2 Interpolação ideal.



Exemplo 6.1 Determinemos o sinal $g(t)$ limitado em banda a B Hz e cujas amostras são

$$g(0) = 1 \quad \text{e} \quad g(\pm T_s) = g(\pm 2T_s) = g(\pm 3T_s) = \cdots = 0$$

em que o intervalo de amostragem T_s é o intervalo de Nyquist para $g(t)$, ou seja, $T_s = 1/2B$.

Usemos a fórmula de interpolação (6.10b) para construir $g(t)$ a partir das amostras. Como todas as amostras de Nyquist, exceto uma, são nulas, somente um termo (correspondente a $k = 0$) sobrevive na soma no lado direito da Eq. (6.10b). Assim,

$$g(t) = \text{sinc}\,(2\pi Bt) \tag{6.11}$$

Este sinal é mostrado na Fig. 6.3. Observemos que este é o único sinal que tem largura de banda B Hz e valores de amostras $g(0) = 1$ e $g(nT_s) = 0$ $(n \neq 0)$. Nenhum outro sinal satisfaz essas condições.

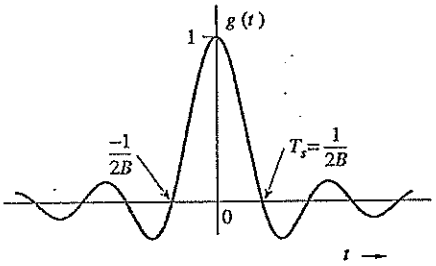


Figura 6.3 Sinal reconstruído das amostras de Nyquist no Exemplo 6.1.

Reconstrução (Interpolação) Prática de Sinais

Na Seção 3.5, vimos que um filtro passa-baixas ideal é não causal e, portanto, irrealizável. Isso também pode ser visto da natureza infinitamente longa do pulso de reconstrução sinc usado na reconstrução ideal da Eq. (6.10). Para aplicações práticas de reconstrução de sinais (por exemplo, em um reprodutor de CD), é necessário implementar um sistema realizável para a reconstrução de sinais a partir de amostras uniformes dos mesmos.

Para implementação prática, o pulso de reconstrução $p(t)$ deve ser de fácil geração. Por exemplo, podemos aplicar o pulso de reconstrução $p(t)$ mostrado na Fig. 6.4. No entanto, devemos primeiro usar o pulso não ideal de interpolação $p(t)$ para analisar a precisão do sinal reconstruído. Denotemos o sinal obtido da reconstrução por

$$\tilde{g}(t) \triangleq \sum_n g(nT_s)p(t - nT_s) \tag{6.12}$$

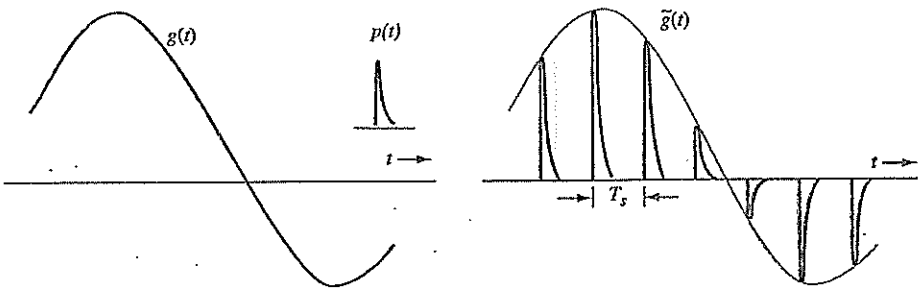


Figura 6.4 Pulso prático de reconstrução (interpolação).

Para determinar a relação entre este sinal e o sinal analógico original $g(t)$, das propriedades da convolução e da Eq. (6.1), vemos que

$$\begin{aligned}\tilde{g}(t) &= \sum_n g(nT_s)p(t-nT_s) = p(t) * \left[\sum_n g(nT_s)\delta(t-nT_s)\right] \\ &= p(t) * \bar{g}(t)\end{aligned} \tag{6.13a}$$

No domínio da frequência, a relação entre a reconstrução e o sinal analógico original é obtida com ajuda da Eq. (6.4)

$$\tilde{G}(f) = P(f)\frac{1}{T_s}\sum_n G(f-nf_s) \tag{6.13b}$$

Isso significa que o sinal $\tilde{g}(t)$ reconstruído com uso do pulso $p(t)$ consiste em múltiplas réplicas de $G(f)$, com frequências centrais nf_s e filtradas por $P(f)$. Para recuperar $g(t)$ completamente, filtragem adicional de $\tilde{g}(t)$ se torna necessária. Tais filtros são, em geral, referidos como equalizadores.

Denotemos a função de transferência do equalizador por $E(f)$. A reconstrução sem distorção requer

$$\begin{aligned}G(f) &= E(f)\tilde{G}(f) \\ &= E(f)P(f)\frac{1}{T_s}\sum_n G(f-nf_s)\end{aligned}$$

Esta relação indica claramente que o equalizador deve remover todas as réplicas deslocadas $G(f-nf_s)$ da soma, exceto o termo passa-baixas com $n=0$, ou seja,

$$E(f)P(f) = 0 \quad |f| > f_s - B \tag{6.14a}$$

Adicionalmente, a reconstrução sem distorção requer

$$E(f)P(f) = T_s \quad |f| < B \tag{6.14b}$$

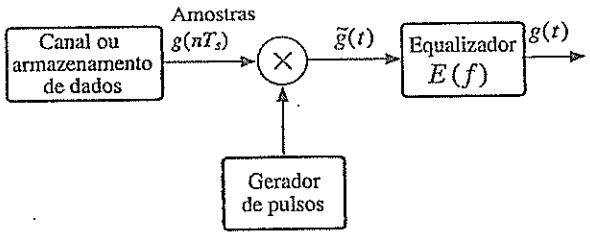


Figura 6.5 Reconstrução prática de sinal.

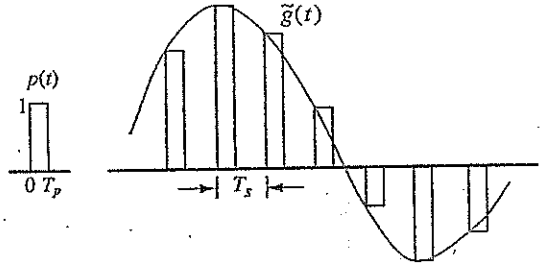


Figura 6.6 Interpolação simples por meio de pulsos retangulares.

O filtro equalizador $E(f)$ deve ser passa-baixas em natureza, para suprimir todo conteúdo de frequência acima de $f_s - B$ Hz, e deve ser o inverso de $P(f)$ na largura de banda de B Hz do sinal. A Fig. 6.5 ilustra o diagrama de um sistema prático de reconstrução que utiliza um equalizador desse tipo.

Consideremos, agora, um gerador de pulsos de interpolação muito simples que gera pulsos curtos (retenção de ordem zero – *zero order hold*). Como mostrado na Fig. 6.6,

$$p(t) = \Pi\left(\frac{t - 0{,}5T_p}{T_p}\right)$$

Este é um pulso retangular de altura unitária e duração T_p. A reconstrução, primeiro, gerará

$$\tilde{g}(t) = \sum_n g(nT_s)\,\Pi\left(\frac{t - nT_s - 0{,}5T_p}{T_p}\right)$$

A função de transferência do filtro $P(f)$ é a transformada de Fourier de $\Pi(t/T_p)$ deslocada de $0{,}5T_p$:

$$P(f) = T_p \operatorname{sinc}(\pi f T_p)\, e^{-j\pi f T_p} \tag{6.15}$$

Como resultado, a resposta de frequência do equalizador deve satisfazer

$$E(f) = \begin{cases} T_s/P(f) & |f| \le B \\ \text{Flexível} & B < |f| < (1/T_s - B) \\ 0 & |f| \ge (1/T_s - B) \end{cases}$$

É importante que determinemos se a resposta de banda passante do equalizador é realizável. Primeiro, podemos acrescentar outro atraso temporal à reconstrução, tal que

$$E(f) = T_s \cdot \frac{\pi f}{\operatorname{sen}(\pi f T_p)} e^{-j2\pi f t_0} \qquad |f| \le B \tag{6.16}$$

Para que um ganho de banda passante $E(f)$ seja bem definido, é imperativo que escolhamos uma pequena largura de pulso T_p, tal que

$$\frac{\operatorname{sen}(\pi f T_p)}{\pi f} \neq 0 \quad |f| \le B$$

Isso significa que o equalizador $E(f)$ não precisa alcançar ganho infinito. Se não for assim, o equalizador se tornará irrealizável. De modo equivalente,

$$T_p < 1/B$$

Portanto, desde que a largura do pulso retangular de reconstrução seja menor que $1/B$, deve ser possível projetar um filtro equalizador analógico para recuperar o sinal analógico original $g(t)$ do trem de pulsos de reconstrução não ideal. Obviamente, este é um requisito para um gerador de pulsos retangulares de reconstrução. Na prática, T_p pode ser escolhido muito pequeno, fornecendo a seguinte resposta de banda passante do equalizador:

$$E(f) = T_s \cdot \frac{\pi f}{\operatorname{sen}(\pi f T_p)} \approx \frac{T_s}{T_p} \qquad |f| \le B \tag{6.17}$$

Isso significa que resta pouca distorção quando pulsos retangulares muito curtos são usados na reconstrução de sinais. Esses casos tornam o projeto do equalizador desnecessário ou muito simples. Um exemplo é dado na forma de exercício com o MATLAB na Seção 6.9.

Podemos obter um resultado melhor que com o filtro de retenção de ordem zero se usarmos um filtro **de retenção de primeira ordem**, que resulta em interpolação linear no lugar da interpolação em degraus. O interpolador linear, cuja resposta ao impulso é um pulso triangular $\Delta(t/2T_s)$, leva a uma

interpolação em que os topos de amostras adjacentes são conectados por segmentos de reta (Exercício 6.1-7).

6.1.2 Questões Práticas Relativas à Amostragem e à Reconstrução de Sinais

Realizabilidade de Filtros de Reconstrução

Se um sinal for amostrado à taxa de Nyquist $f_s = 2B$ Hz, o espectro $\overline{G}(f)$ consiste em repetições de $G(f)$ sem qualquer separação entre ciclos adjacentes, como mostrado na Fig. 6.7a. Para recuperar $g(t)$ de $\overline{g}(t)$, é preciso aplicar o sinal amostrado $\overline{g}(t)$ a um filtro passa-baixas ideal (área pontilhada na Fig. 6.7a). Como visto na Seção 3.5, um filtro como este é irrealizável na prática; uma boa aproximação para o filtro pode ser obtida apenas com atraso temporal infinito na resposta. Isso significa que podemos recuperar o sinal $g(t)$ de suas amostras com atraso temporal infinito.

Uma solução prática a este problema consiste em amostrar o sinal a uma taxa maior que a de Nyquist ($f_s > 2B$ ou $\omega_s > 4\pi B$). O resultado é $\overline{G}(f)$ que, agora, consiste em repetições de $G(f)$ com separação finita entre ciclos adjacentes, como ilustrado na Fig. 6.7b. Assim, podemos recuperar $G(f)$ de $\overline{G}(f)$ [ou de $\tilde{G}(f)$] com o uso de um filtro passa-baixas com característica de corte gradual (área pontilhada na Fig. 6.7b). Contudo, mesmo neste caso, o ganho do filtro deve ser zero além do primeiro ciclo de $G(f)$ (Fig. 6.7b). Segundo o critério de Paley-Wiener, a realização deste filtro também é impossível. A única vantagem neste caso é a possibilidade de obter uma aproximação para o filtro com menor atraso temporal. Isto mostra que, na prática, é impossível recuperar exatamente um sinal limitado em banda $g(t)$ a partir de suas amostras, mesmo que a taxa de amostragem seja maior do que a de Nyquist. No entanto, à medida que a taxa de amostragem aumenta, o sinal recuperado se aproxima mais do sinal desejado.

A Perfídia do Mascaramento

Há outra dificuldade prática fundamental associada à reconstrução de um sinal a partir de suas amostras. O teorema da amostragem foi provado sob a hipótese de que o sinal $g(t)$ era limitado em banda. Entretanto, *todos os sinais práticos são limitados no tempo*; ou seja, têm duração ou largura finita. Podemos demonstrar (Exercício 6.1-8) que um sinal não pode ser simultaneamente limitado no tempo e em banda. Um sinal limitado no tempo não pode ser limitado em banda e vice-versa (mas um sinal pode ser simultaneamente não limitado no tempo e em banda). Fica claro que todos os sinais práticos, que são necessariamente limitados no tempo, não são limitados em banda, como ilustrado na

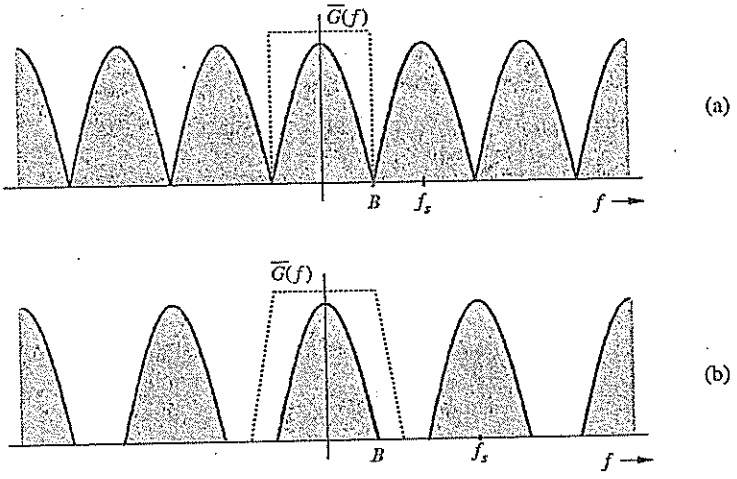


Figura 6.7 Espectros de um sinal amostrado; (a) à taxa de Nyquist; (b) a uma taxa maior que a de Nyquist.

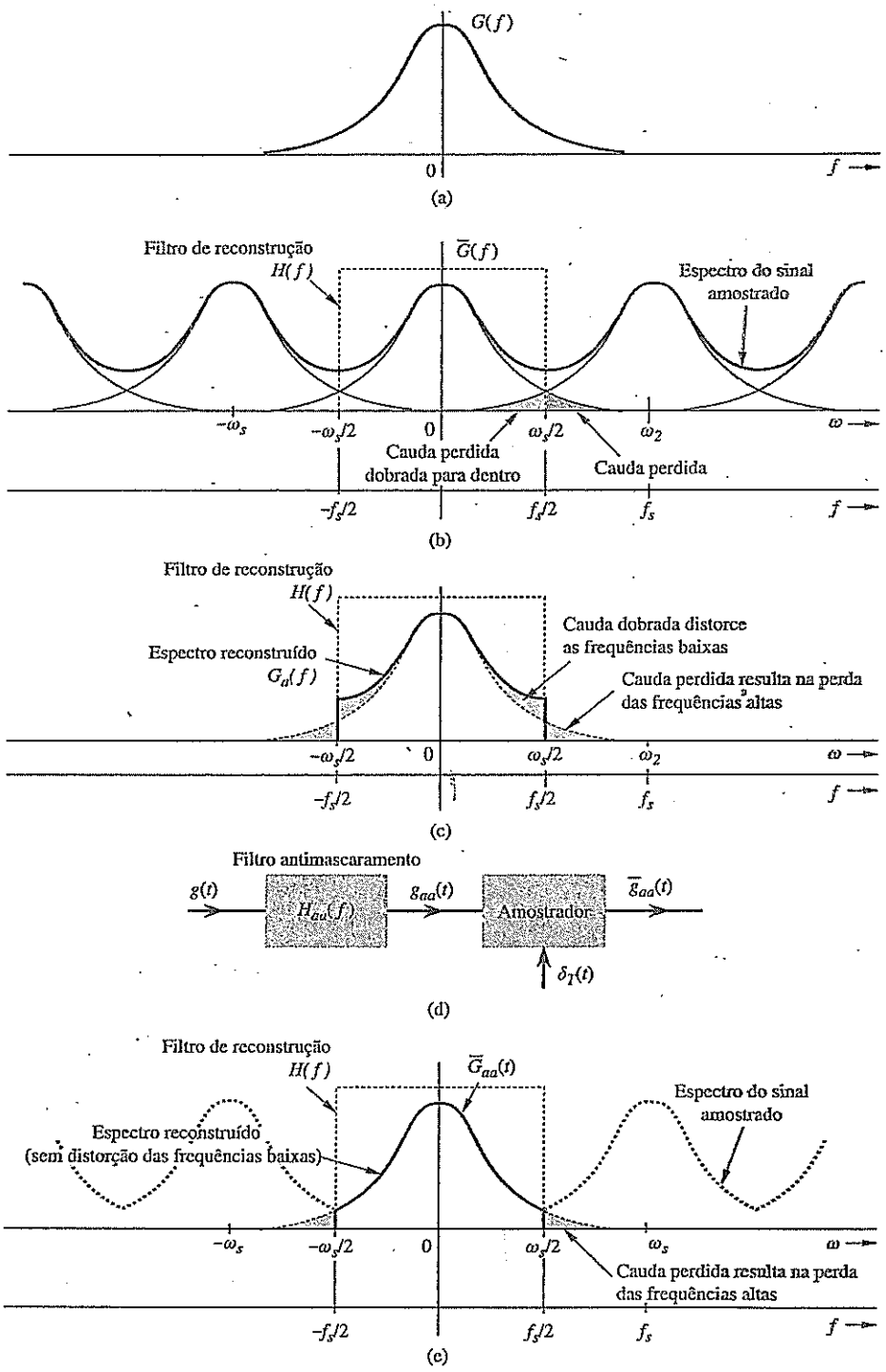


Figura 6.8
Efeito de mascaramento.
(a) Espectro de um sinal prático $g(t)$.
(b) Espectro de $g(t)$ amostrado.
(c) Espectro do sinal reconstruído.
(d) Esquema de amostragem com uso de filtro antimascaramento.
(e) Espectros dos sinais amostrado (linha pontilhada) e reconstruído (linha cheia) com filtro antimascaramento.

Fig. 6.8a; estes sinais têm largura de banda infinita e o espectro $\overline{G}(f)$ consiste em ciclos de $G(f)$ que se sobrepõem e se repetem a cada f_s Hz (frequência de amostragem), como indicado na Fig. 6.8b. Por causa da largura de banda infinita, a sobreposição de espectro é inevitável, qualquer que seja a taxa de amostragem. O aumento da taxa de amostragem reduz, mas não elimina a sobreposição de ciclos espectrais repetitivos. Devido à cauda da sobreposição, $\overline{G}(f)$ deixa de conter a informação completa

de $G(f)$ e não é mais possível, nem mesmo teoricamente, recuperar $g(t)$ exatamente do sinal amostrado $\bar{g}(t)$. Caso o sinal amostrado seja aplicado a um filtro passa-baixas ideal de frequência de corte $f_s/2$ Hz, a saída não será $G(f)$, mas $G_a(f)$ (Fig. 6.8c), que é uma versão distorcida de $G(f)$, devido a duas causas separadas:

1. A perda da cauda de $G(f)$ além de $|f| > f_s/2$ Hz.
2. A reaparição desta cauda invertida ou dobrada dentro do espectro.

Notemos que os espectros se cruzam na frequência $f_s/2 = 1/2T$ Hz, chamada frequência de *dobramento*. O espectro pode ser visto como se a cauda perdida se dobrasse para dentro na frequência de dobramento. Por exemplo, uma componente de frequência $(f_s/2) + f_x$ aparece "disfarçada" de uma componente de frequência baixa $(f_s/2) - f_x$. Assim, as componentes de frequências acima de $f_s/2$ reaparecem como componentes de frequências abaixo de $f_s/2$. Esta inversão da cauda, conhecida como *dobramento espectral* ou *mascaramento*, é ilustrada na Fig. 6.8b e, também, na Fig. 6.8c. No processo de mascaramento, não apenas são perdidas todas as componentes de frequências acima da frequência de dobramento $f_s/2$ Hz, mas as mesmas reaparecem (mascaradas) como componentes de frequências abaixo da frequência de dobramento $f_s/2$, como ilustrado na Fig. 6.8c.

O problema de mascaramento é análogo ao de um batalhão do exército em que um pelotão deserta para o lado inimigo, mas permanece aparentemente leal. O batalhão fica sujeito a risco duplo. Primeiro, perdeu o pelotão desertor como uma força de combate. Além disso, durante uma batalha, o batalhão enfrentará sabotagem causada pelos desertores e terá de usar um pelotão leal para neutralizar os desertores. Assim, a força armada perdeu dois pelotões, que deixaram de ter atividade produtiva.

Eliminação de Desertores: Filtro de Antimascaramento

Para o comandante do batalhão traído, a solução do problema seria óbvia. Assim que tomasse conhecimento da deserção, incapacitaria, de alguma forma, o pelotão desertor. Se fizer isso *antes do início da batalha*, o comandante perderá apenas um pelotão (o desertor).* Esta é uma solução parcial ao duplo problema de traição e sabotagem; uma solução que retifica o problema parcialmente e elimina metade das perdas.

Seguiremos exatamente esta estratégia. Os desertores em potencial são todas as componentes de frequências acima da frequência de dobramento $f_s/2 = 1/2T$ Hz. Devemos eliminar (suprimir) estas componentes de $g(t)$ *antes da amostragem*. Tal supressão das frequências altas pode ser efetuada por um filtro passa-baixas ideal com frequência de corte $f_s/2$ Hz, como mostrado na Fig. 6.8d. Esse filtro recebe a denominação de *filtro antimascaramento*. A Fig. 6.8d também mostra que a filtragem antimascaramento é feita antes da amostragem. A Fig. 6.8c mostra os espectros do sinal amostrado e do sinal reconstruído, $G_{aa}(f)$, quando o esquema antimascaramento é empregado. Um filtro antimascaramento basicamente limita o sinal $g(t)$ em banda a $f_s/2$ Hz. Dessa forma, perdemos apenas as componentes acima da frequência de dobramento $f_s/2$ Hz. As componentes suprimidas não podem reaparecer e corromper as componentes de frequências abaixo da frequência de dobramento. O uso de um filtro antimascaramento faz com que o espectro do sinal reconstruído seja $G_{aa}(f) = G(f)$ para $|f| < f_s/2$. Embora percamos o espectro além de $f_s/2$, o espectro para todas as frequências abaixo de $f_s/2$ permanece intacto. A distorção de mascaramento é reduzida à metade devido à eliminação do dobramento. Ressaltamos, mais uma vez, que a operação antimascaramento deve ser executada *antes da amostragem do sinal*.

Um filtro antimascaramento também ajuda a reduzir o ruído. Em geral, o ruído tem um espectro de banda larga e, sem antimascaramento, o próprio fenômeno de mascaramento faria que componentes de ruído fora da banda do sinal desejado aparecessem na banda do sinal. O antimascaramento suprime todo o espectro de ruído além da frequência $f_s/2$.

---

*A Fig. 6.8b mostra que, dentre um número infinito de ciclos repetitivos, apenas ciclos espectrais vizinhos se sobrepõem. Esta é uma imagem simplificada. Na verdade, todos os ciclos se sobrepõem e interagem com todos os outros, devido à largura infinita do espectro de qualquer sinal prático. Felizmente, todos os espectros práticos também devem decair nas frequências altas. Isso resulta em um grau insignificante de interferência de ciclos que não sejam vizinhos imediatos. Quando esta hipótese não se aplica, o cálculo de mascaramento se torna mais elaborado.

O filtro antimascaramento, por ser um filtro ideal, é irrealizável. Na prática, usamos um filtro de corte abrupto, que deixa um espectro residual altamente atenuado além da frequência de dobramento $f_s/2$.

Amostragem Força Sinais Não Limitados em Banda a Parecerem Limitados em Banda
A Fig. 6.8b mostra que o espectro de um sinal $\overline{g}(t)$ consiste em ciclos de $G(f)$ sobrepostos. Isso significa que $\overline{g}(t)$ são amostras sub-Nyquist de $g(t)$. Contudo, também podemos ver o espectro na Fig. 6.8b como o espectro $G_a(f)$ (Fig. 6.8c) repetido periodicamente a cada f_s Hz sem sobreposição. O espectro $G_a(f)$ é limitado em banda a $f_s/2$ Hz. Portanto, essas amostras (sub-Nyquist) de $g(t)$ são, na verdade, as amostras de Nyquist para o sinal $g_a(t)$. Em resumo, a amostragem de um sinal não limitado em banda $g(t)$ a uma taxa de f_s Hz faz com que as amostras pareçam ser as amostras de Nyquist de um outro sinal $g_a(t)$, que é limitado em banda a $f_s/2$ Hz. Em outras palavras, a amostragem faz um sinal não limitado em banda aparecer como um sinal limitado em banda $g_a(t)$, com largura de banda $f_s/2$ Hz. Uma conclusão semelhante se aplica quando $g(t)$ é um sinal limitado em banda, mas amostrado a uma taxa sub-Nyquist.

6.1.3 Máxima Taxa de Informação: Duas Porções de Informação por Segundo por Hertz

Um conhecimento da máxima taxa em que a informação pode ser transmitida ao longo de um canal de largura de banda B Hz é de fundamental importância em comunicações digitais. A seguir, deduziremos uma das relações mais básicas em comunicações, segundo a qual *um máximo de 2B porções independentes de informação por segundo pode ser transmitido, sem erro, em um canal de largura de banda B Hz.* Este resultado advém do teorema da amostragem.

Primeiro, o teorema da amostragem garante que um sinal de largura de banda de B Hz pode ser totalmente recuperado de amostras tomadas uniformemente a uma taxa de $2B$ amostras por segundo. Agora, precisamos mostrar que qualquer sequência de dados independentes a uma taxa de $2B$ Hz pode vir de amostras uniformes de um sinal passa-baixas de largura de banda B. Além disso, podemos construir esse sinal passa-baixas a partir da sequência de dados independentes.

Suponhamos que uma sequência de dados independentes seja denotada por $\{g_n\}$. A taxa correspondente é de $2B$ amostras por segundo. Portanto, sempre existe um sinal $g(t)$ (não necessariamente limitado em banda) tal que

$$g_n = g(nT_s) \qquad T_s = \frac{1}{2B}$$

Na Fig. 6.9a, ilustramos, mais uma vez, o efeito de amostrar um sinal não limitado em banda $g(t)$ a uma taxa $f_s = 2B$ Hz. Devido ao mascaramento, o sinal amostrado ideal é

$$\begin{aligned}\overline{g}(t) &= \sum_n g(nT_s)\delta(t - nT_s)\\ &= \sum_n g_a(nT_s)\delta(t - nT_s)\end{aligned}$$

em que $g_a(t)$ é o sinal passa-baixas mascarado cujas amostras $g_a(nT_s)$ são iguais às amostras de $g(nT_s)$. Em outras palavras, amostragem de um sinal $g(t)$ a uma taxa sub-Nyquist gera amostras que podem ser igualmente obtidas da amostragem de um sinal limitado em banda $g_a(t)$ à taxa de Nyquist. Assim, pela Fig. 6.9, demonstramos que as amostragens de $g(t)$ e $g_a(t)$ à taxa de $2B$ Hz geram a mesma sequência de dados independentes $\{g_n\}$:

$$g_n = g(nT_s) = g_a(nT_s) \qquad T_s = \frac{1}{2B} \tag{6.18}$$

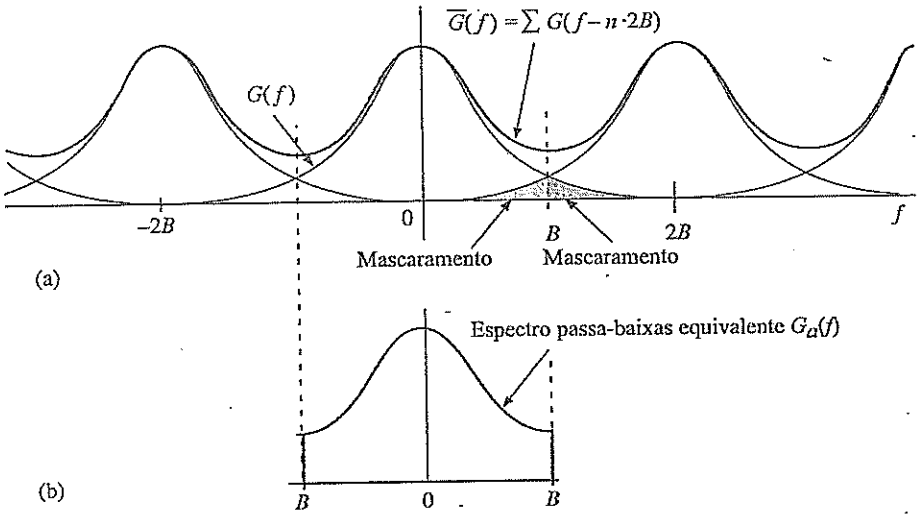


Figura 6.9 (a) Espectro de um sinal não limitado em banda e correspondente espectro amostrado $\overline{G}(f)$. (b) Espectro passa-baixas equivalente $G_a(f)$ construído das amostras uniformes de $g(t)$ à taxa de amostragem $2B$.

Ainda usando o teorema da amostragem, vemos que um sinal passa-baixas $g_a(t)$ com largura de banda B pode ser reconstruído de suas amostras uniformes [Eq. (6.10]

$$g_a(t) = \sum_n g_n \operatorname{sinc}\,(2\pi Bt - k\pi)$$

Presumindo que não haja ruído, esse sinal pode ser transmitido sem erro ao longo de um canal sem distorção e com largura de banda B Hz. No receptor, a sequência de dados $\{g_n\}$ pode ser recuperada das amostras de Nyquist da saída do canal sem distorção $g_a(t)$ como os dados de informação desejados.

Essa taxa teórica de comunicação supõe um canal sem ruído. Na prática, ruído de canal é inevitável e, em consequência, essa taxa causará alguns erros de detecção. No Capítulo 14, discutiremos a capacidade de Shannon, que permite determinar a taxa teórica de comunicação sem erro na presença de ruído.

6.1.4 Análise de Amostragem Prática Não Ideal

Até aqui, focamos principalmente a amostragem uniforme ideal, que pode usar um trem de impulsos ideais para extrair precisamente o valor de sinal $g(kT_s)$ no instante de tempo exato $t = kT_s$. Na prática, nenhum dispositivo físico é capaz de executar esta tarefa. Em consequência, precisamos considerar implementações mais práticas de amostragem. Essa análise é importante para um melhor entendimento dos erros que, em geral, ocorrem durante conversão A/D prática e seus efeitos na reconstrução de sinais.

Amostradores práticos colhem cada amostra de sinal em um curto período de tempo T_p em torno de $t = kT_s$. Em outras palavras, a cada T_s segundos, o dispositivo de amostragem colhe uma fotografia de duração T_p do sinal $g(t)$ sendo amostrado. Isso é como tirar uma sequência de fotografias de um velocista olímpico em uma corrida de 100 m rasos. Assim como uma câmera fotográfica gera uma imagem estática tomando uma média do cenário ao logo da janela T_p, o amostrador prático gera um valor de amostra em $t = kT_s$ tomando a média de valores do sinal $g(t)$ na janela T_p, ou seja,

$$g_1(kT_s) = \frac{1}{T_p}\int_{-T_p/2}^{T_p/2} g(kT_s + t)\,dt \tag{6.19a}$$

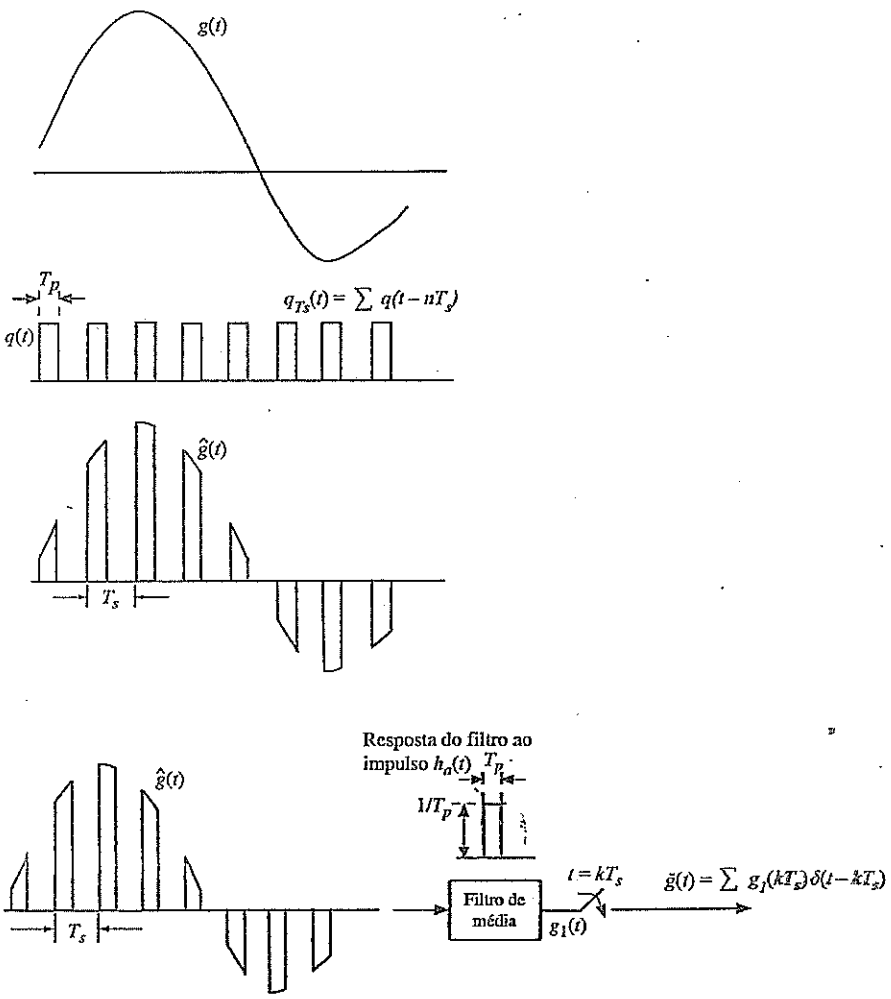


Figura 6.10 Ilustração de amostragem prática.

Dependendo do dispositivo, essa média pode ser ponderada por uma função de peso $q(t)$ que depende do dispositivo:

$$g_1(kT_s) = \frac{1}{T_p}\int_{-T_p/2}^{T_p/2} q(t)g(kT_s + t)\,dt \tag{6.19b}$$

Portanto, usamos a analogia com a câmera fotográfica para entender que amostradores práticos geram sinais amostrados da forma

$$\check{g}(t) = \sum g_1(kT_s)\delta(t - kT_s) \tag{6.20}$$

Agora, mostraremos a relação entre o sinal amostrado prático $\check{g}(t)$ e o sinal passa-baixas analógico original $g(t)$ no domínio da frequência.

Usaremos a Fig. 6.10 para ilustrar a relação entre $\check{g}(t)$ e $g(t)$ para o caso especial de ponderação uniforme. Isto significa que

$$q(t) = \begin{cases} 1 & |t| \le 0{,}5T_p \\ 0 & |t| > 0{,}5T_p \end{cases}$$

Como mostrado na Fig. 6.10, $g_1(t)$ pode ser obtido, de modo equivalente, usando primeiro "amostragem natural" para gerar as *fotografias* do sinal

$$\widehat{g}(t) = g(t) \cdot q_{T_s}(t) \tag{6.21}$$

em que

$$q_{T_s}(t) = \sum_{n=-\infty}^{\infty} q(t - nT_s)$$

A Fig. 6.10b ilustra o sinal-fotografia $\widehat{g}(t)$. Podemos, então, definir um filtro de média (*averaging filter*) com a seguinte resposta ao impulso

$$h_a(t) = \begin{cases} \dfrac{1}{T_p} & -\dfrac{T_p}{2} \le t < \dfrac{T_p}{2} \\ 0 & \text{para outros valores de } t \end{cases}$$

ou função de transferência

$$H_a(f) = \text{sinc}\left(\pi f T_p\right)$$

Aplicação do sinal-fotografia $\widehat{g}(t)$ ao filtro de média gera o sinal de saída

$$g_1(t) = h_a(t) * \widehat{g}(t)$$

Como ilustrado na Fig. 6.10c, o amostrador prático gera um sinal $\check{g}(t)$ por meio de amostragem da saída do filtro $g_1(kT_s)$. Portanto, usamos a Fig. 6.10c para estabelecer o processo equivalente para gerar amostras práticas de $g(t)$, que consiste em tirar fotografias, tomar a média e amostrar. Agora, podemos examinar as relações no domínio da frequência para analisar a distorção gerada por amostradores práticos.

Na análise a seguir, consideraremos uma função de peso genérica $q(t)$ sujeita à única condição

$$q(t) = 0, \qquad t \notin (-0{,}5T_p, 0{,}5T_p)$$

Primeiro, notemos que $q_{T_s}(t)$ é periódica. Portanto, sua série de Fourier pode ser escrita como

$$q_{T_s}(t) = \sum_{n=-\infty}^{\infty} Q_n e^{jn\omega_s t}$$

em que

$$Q_n = \frac{1}{T_s} \int_{-0{,}5T_p}^{0{,}5T_p} q(t) e^{-jn\omega_s t} dt$$

Assim, o sinal de saída do filtro de média é

$$\begin{aligned} g_1(t) &= h_a(t) * \left[g(t) q_{T_s}(t)\right] \\ &= h_a(t) * \sum_{n=-\infty}^{\infty} Q_n g(t) e^{jn\omega_s t} \end{aligned} \tag{6.22}$$

No domínio da frequência, temos

$$\begin{aligned} G_1(f) &= H(f) \cdot \sum_{n=-\infty}^{\infty} Q_n G(f - nf_s) \\ &= \text{sinc}\left(\pi f T_p\right) \sum_{n=-\infty}^{\infty} Q_n G(f - nf_s) \end{aligned} \tag{6.23}$$

Como

$$\check{g}(t) = \sum_k g_1(kT_s)\delta(t - kT_s)$$

podemos aplicar o teorema da amostragem para mostrar que

$$\begin{aligned}\check{G}(f) &= \frac{1}{T_s}\sum_m G_1(f + mf_s) \\ &= \frac{1}{T_s}\sum_m \operatorname{sinc}\left[\frac{(2\pi f + m2\pi f_s)T_p}{2}\right]\sum_n Q_n\, G(f + mf_s - nf_s) \\ &= \sum_\ell \left(\frac{1}{T_s}\sum_n Q_n \operatorname{sinc}\left[(\pi f + (n+\ell)\pi f_s)T_p\right]\right) G(f + \ell f_s) \qquad (6.24)\end{aligned}$$

A última igualdade veio da mudança do índice do somatório $\ell = m - n$.

Podemos definir a resposta em frequência

$$F_\ell(f) = \frac{1}{T_s}\sum_n Q_n \operatorname{sinc}\left[(\pi f + (n+\ell)\pi f_s)T_p\right]$$

Esta definição nos permite escrever

$$\check{G}(f) = \sum_\ell F_\ell(f) G_1(f + \ell f_s) \qquad (6.25)$$

Para o sinal passa-baixas $G(f)$ com largura de banda B Hz, a aplicação de um filtro passa-baixas ideal (interpolação) gera um sinal distorcido

$$F_0(f)G(f) \qquad (6.26a)$$

com

$$F_0(f) = \frac{1}{T_s}\sum_n Q_n \operatorname{sinc}\left[\pi(f + nf_s)T_p\right] \qquad (6.26b)$$

Podemos ver das Eqs. (6.25) e (6.26) que o sinal resultante de amostragem prática sempre contém uma distorção conhecida $F_0(f)$.

Além disso, o uso de um pulso prático de reconstrução $p(t)$, como na Eq. (6.12), gera distorções adicionais. Reconstruamos $g(t)$ usando as amostras práticas para gerar

$$\check{g}(t) = \sum_n g_1(nT_s)p(t - nT_s)$$

Da Eq. (6.13), obtemos a relação entre os espectros da reconstrução e da mensagem original $G(f)$ como

$$\check{G}(f) = P(f)\sum_n F_n(f)G(f + nf_s) \qquad (6.27)$$

Como $G(f)$ tem largura de banda de B Hz, precisaremos projetar um novo equalizador com função de transferência $E(f)$, de modo que a reconstrução não apresente distorção na largura de banda B, ou seja,

$$E(f)P(f)F_0(f) = \begin{cases} 1 & |f| < B \\ \text{Flexível} & B < |f| < f_s - B \\ 0 & |f| > f_s - B \end{cases} \qquad (6.28)$$

O equalizador simples pode ser projetado para compensar duas fontes de distorção: efeito de amostragem não ideal em $F_0(f)$ e efeito de reconstrução não ideal em $P(f)$. O projeto do equalizador é possível porque as duas distorções são conhecidas previamente.

6.1.5 Algumas Aplicações do Teorema da Amostragem

O teorema da amostragem é muito importante na análise, processamento e transmissão de sinais, pois nos permite substituir um sinal temporal contínuo por uma sequência discreta de números. O processamento de um sinal temporal contínuo é, portanto, equivalente ao processamento de uma sequência discreta de números. Isso nos leva diretamente à área de filtragem digital. No campo das comunicações, a transmissão de uma mensagem contínua se reduz à transmissão de uma sequência de números. Isso abre portas a muitas técnicas novas de comunicação de sinais temporais contínuos por trens de pulsos. O sinal temporal contínuo $g(t)$ é amostrado, e os valores das amostras são usados para modificar certos parâmetros de um trem de pulsos periódicos. Podemos variar as amplitudes (Fig. 6.11b), as larguras (Fig. 6.11c) ou as posições (Fig. 6.11d) dos pulsos proporcionalmente aos valores das amostras do sinal $g(t)$. Assim, podemos ter **modulação por amplitude de pulso** (PAM – *pulse amplitude modulation*), **modulação por largura de pulso** (PWM – *pulse width modulation*), ou **modulação por posição de pulso** (PPM – *pulse position modulation*). A mais importante forma de modulação por pulso hoje em dia é a modulação por **codificação de pulsos** (PCM – *pulse code modulation*), introduzida na Seção 1.2. Em todos estes casos, em vez de transmitir $g(t)$, transmitimos o sinal modulado em pulsos. No receptor, lemos a informação do sinal modulado em pulsos e reconstruímos o sinal analógico $g(t)$.

Uma das vantagens do uso de modulação em pulsos é permitir a transmissão simultânea de vários sinais por meio de compartilhamento do tempo – **multiplexação por divisão no tempo** (TDM – *time division multiplexing*). Como um sinal modulado em pulsos ocupa somente uma parte do canal temporal, podemos transmitir vários sinais modulados em pulsos em um mesmo canal entrelaçando-os. A Fig. 6.12 ilustra a TDM de dois sinais PAM. Dessa forma, podemos multiplexar diversos sinais em um mesmo canal com a redução da largura de cada pulso.

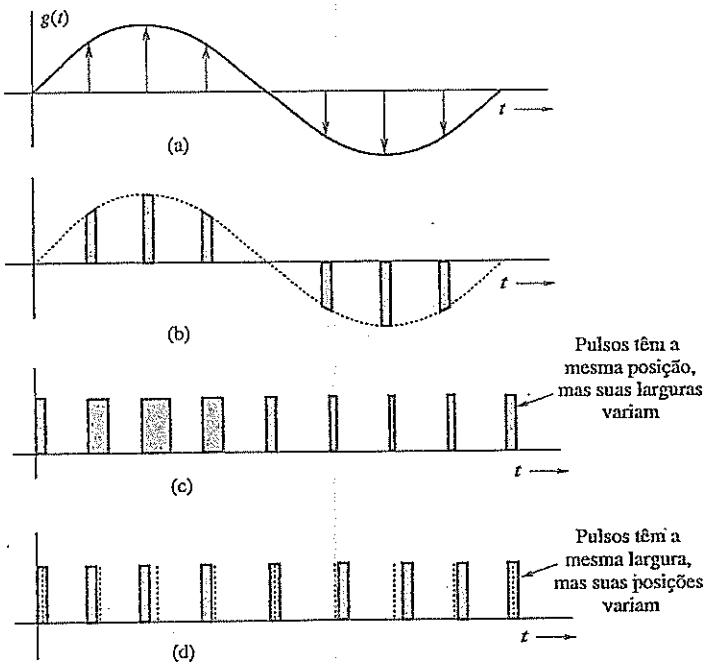


Figura 6.11 Sinais modulados em pulsos. (a) Sinal não modulado. (b) Sinal PAM. (c) Sinal PWM. (d) Sinal PPM.

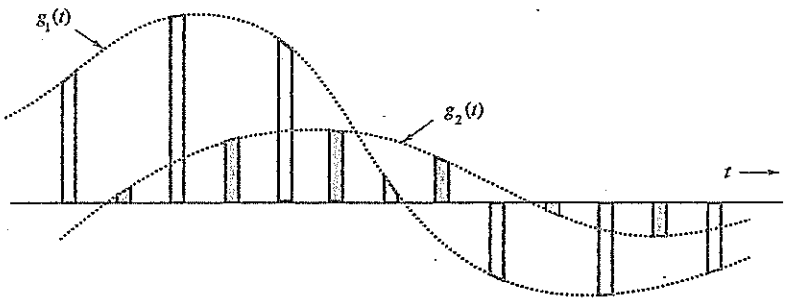


Figura 6.12 Multiplexação por divisão do tempo de dois sinais.

Outro método de transmissão simultânea de vários sinais em banda base é a multiplexação por divisão em frequência (FDM – *frequency division multiplexing*), discutida brevemente no Capítulo 4. Em FDM, vários sinais são multiplexados por compartilhamento da largura de banda do canal. O espectro de cada mensagem é deslocado a uma banda específica não ocupada por qualquer outro sinal. A informação de vários sinais é posicionada em bandas de frequências do canal que não se sobrepõem. De certa forma, TDM e FDM são duais uma da outra.

6.2 MODULAÇÃO POR CODIFICAÇÃO DE PULSO (PCM)

A PCM é a mais útil e mais largamente empregada das modulações em pulsos mencionadas. Como ilustrado na Fig. 6.13, a PCM é basicamente uma ferramenta para converter um sinal analógico em um sinal digital (conversão A/D). Um sinal **analógico** é caracterizado por uma amplitude que pode assumir qualquer valor em um intervalo contínuo. Isso significa que pode assumir um número infinito de valores. Um sinal **digital**, por sua vez, tem uma amplitude que pode assumir apenas um número finito de valores. Um sinal analógico pode ser convertido em um sinal digital através de amostragem e **quantização**, ou seja, aproximação de seu valor ao mais próximo dos números permitidos (ou **níveis de quantização**), como indicado na Fig. 6.14. As amplitudes do sinal analógico $m(t)$ ocorrem no intervalo $(-m_p, m_p)$, que é dividido em L subintervalos, cada um com largura $\Delta v = 2m_p/L$. A seguir, cada amplitude de amostra é aproximada pelo valor no ponto médio do subintervalo em que a amostra ocorre (Fig. 6.14, com $L = 16$). Cada amostra é, então, aproximada a um dos L números. Assim, o sinal é digitalizado, com amostras quantizadas que assumem um dos L valores. Um sinal deste tipo é conhecido como um **sinal digital L-ário**.

Figura 6.13 Diagrama de blocos de um sistema PCM.

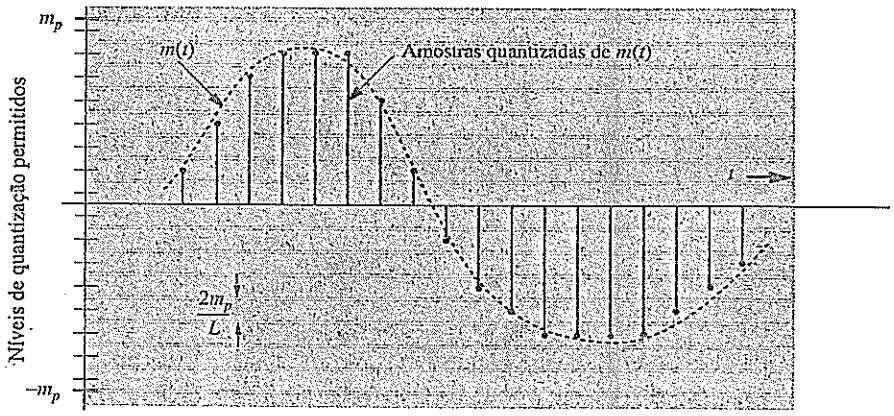


Figura 6.14 Quantização de um sinal analógico amostrado.

De um ponto de vista prático, um sinal digital binário (sinal que pode assumir apenas dois valores) é muito desejável, devido à sua simplicidade, economia, e facilidade de implementação. Podemos converter um sinal L-ário em um sinal binário através de codificação por pulsos. Essa codificação é ilustrada na Fig. 1.5 para o caso $L = 16$. Esta codificação, formada pela representação binária dos 16 dígitos decimais de 0 a 15, é conhecida como **código binário natural** (CBN). Outras maneiras de obtenção de códigos binários serão discutidas posteriormente. A cada um dos 16 níveis a serem transmitidos é alocado código binário de quatro dígitos. O sinal analógico $m(t)$ fica, então, convertido em um sinal digital (binário). Um dígito binário é denominado **bit** (de *binary digit*), por conveniência. Essa contração de "binary digit" em "bit" se tornou uma abreviação padrão na indústria e é usada em todo o livro.

Cada amostra no exemplo é, portanto, codificada por quatro bits. Para a transmissão destes dados binários, precisamos alocar uma forma diferente de pulso a cada um dos dois bits. Uma possibilidade é alocar um pulso negativo ao binário **0** e um pulso positivo ao binário **1** (Fig. 1.5), de modo que cada amostra seja transmitida por um grupo de quatro pulsos binários (codificação por pulsos). O sinal resultante é um sinal binário.

A largura de banda de um sinal de áudio é de cerca de 15 kHz. Contudo, para voz, testes subjetivos mostraram que a articulação (inteligibilidade) do sinal não é afetada se todas as componentes acima de 3400 Hz forem suprimidas.*[3] Como, em telefonia, o objetivo é a inteligibilidade e não alta fidelidade, as componentes acima de 3400 Hz são eliminadas por um filtro passa-baixas. O sinal resultante é, então, amostrado a uma taxa de 8000 amostras por segundo (8 kHz). Esta taxa é intencionalmente maior que a taxa de amostragem Nyquist de 6,8 kHz para que filtros realizáveis possam ser aplicados na reconstrução do sinal. Cada amostra é, por fim, quantizada em 256 níveis ($L = 256$), o que requer um grupo de oito pulsos binários para codificar cada amostra ($2^8 = 256$). Assim, um sinal de telefonia requer $8 \times 8.000 = 64.000$ pulsos binários por segundo.

Uma aplicação mais recente de PCM é o *compact disc* (CD). Esse é um caso de alta fidelidade e requer uma largura de banda de 20 kHz para o sinal de áudio. Embora a taxa de amostragem de Nyquist seja apenas 40 kHz, a taxa de amostragem usada é de 44,1 kHz, pelas razões mencionadas anteriormente. O sinal é quantizado em um número muito maior ($L = 65.536$) de níveis de quantização, e cada um é representado por 16 bits para reduzir o erro de quantização. Amostras codificadas em binário (1,4 milhão de bits/s) são, então, gravadas no CD.

6.2.1 Vantagens de Comunicação Digital

A seguir, são listadas algumas vantagens da comunicação digital em relação à comunicação analógica:

1. A comunicação digital é capaz de suportar muito mais ruído de canal e distorção, desde que o ruído e a distorção estejam dentro de certos limites, e é mais robusta do que a comunicação analógica. No caso de mensagens analógicas, qualquer distorção ou ruído, não importa quão pequenos sejam, alterará o sinal recebido.

2. A maior vantagem da comunicação digital em relação à analógica, no entanto, é viabilizar o uso de repetidores regeneradores. Em um sistema de comunicação analógica, um sinal de mensagem se torna progressivamente mais fraco à medida que viaja pelo canal, enquanto o ruído cumulativo de canal e a distorção se tornam progressivamente mais fortes. Por fim, o sinal é superado por ruído e distorção. A amplificação é de pouca utilidade, pois o sinal e o ruído são amplificados na mesma proporção. Em consequência, a distância ao longo da qual uma mensagem analógica pode ser transmitida é limitada pela potência de transmissão inicial. No caso de comunicação digital, uma longa distância de transmissão também pode acarretar ruído e interferências excessivos. O truque, no entanto, consiste em montar estações repetidoras ao longo da rota de transmissão em distâncias curtas o bastante para que pulsos de sinal possam ser detectados antes que o acúmulo de ruído e distorção destrua o sinal. Em cada estação repetidora, os pulsos são detectados e pulsos novos e limpos são transmitidos

*Componentes abaixo de 300 Hz também podem ser suprimidas sem afetar a articulação.

à próxima estação repetidora, que, por sua vez, duplica esse procedimento. Se o ruído e a distorção forem mantidos abaixo de certos limites (o que é possível, tendo em vista o pequeno espaçamento entre repetidores), os pulsos podem ser detectados corretamente.* Assim, mensagens digitais podem ser transmitidas por distâncias maiores e com mais confiabilidade do que mensagens analógicas. Na PCM, o erro mais significativo advém da quantização. Esse erro pode ser reduzido tanto quanto desejado, com o aumento do número de níveis de quantização; o preço dessa solução é pago com o aumento da largura debanda do meio de transmissão (canal).

3. A implementação em hardware digital é flexível e permite o uso de microprocessadores, comutadores digitais e circuitos integrados de larga escala.

4. Sinais digitais podem ser codificados para produzir taxas de erro extremamente pequenas e alta fidelidade, assim como privacidade.

5. A multiplexação de sinais digitais é mais fácil e eficiente que a de sinais analógicos.

6. A comunicação digital é inerentemente mais eficiente que a analógica no que diz respeito à troca de SNR por largura de banda.

7. O armazenamento de sinais digitais é relativamente simples e barato e permite a busca e seleção de informação em bases de dados eletrônicas distantes.

8. A reprodução com mensagens digitais pode ser extremamente confiável e sem deterioração. Mensagens analógicas, como fotocópias e filmes, por exemplo, perdem qualidade a cada estágio sucessivo de reprodução e devem ser transportadas fisicamente de um local distante a outro, o que, em geral, tem custo relativamente alto.

9. O custo de hardware digital continua tendo redução de 50% a cada dois ou três anos, enquanto o desempenho ou capacidade dobra no mesmo período. E não há um fim à vista para o empolgante e exponencial progresso da tecnologia digital. Em consequência, hoje, as tecnologias digitais dominam todas as áreas de comunicação e armazenamento de informação.

Nota Histórica

Para descrever prosódia, o antigo escritor indiano Pingala aplicou sofisticados conceitos matemáticos, produzindo a primeira descrição de um sistema numérico digital de que temos notícia, possivelmente no oitavo século a.C.[6] Outros, como R. Hall, em *Mathematics of Poetry* (Matemática da Poesia), localizam-no mais tarde, por volta de 200 a.C. Gottfried Wilhelm Leibniz (1646-1716) foi o primeiro matemático no ocidente a estabelecer sistematicamente a representação binária (usando 1s e 0s) para qualquer número. Ele sentiu uma significância espiritual em sua descoberta, acreditando que **1**, representando a unidade, era claramente um símbolo para Deus, enquanto **0** representava o nada. Ele raciocinou que, se todos os números podem ser representados meramente com o uso de **1** e **0**, isto provava, com certeza, que Deus criara o universo a partir do nada!

6.2.2 Quantização

Como mencionado anteriormente, sinais digitais têm uma variedade de fontes. Algumas, como computadores, são inerentemente digitais. Outras são analógicas, convertidas à forma digital por meio de diferentes técnicas, como PCM ou modulação delta (DM), que analisaremos em seguida. O resto desta seção apresenta uma discussão quantitativa da PCM e de seus vários aspectos, como quantização, codificação, sincronização e as necessárias largura de banda e SNR de transmissão.

Para quantização, limitamos a amplitude do sinal de mensagem $m(t)$ ao intervalo $(-m_p, m_p)$, como mostrado na Fig. 6.14. Notemos que m_p não é necessariamente a amplitude de pico $m(t)$. As amplitudes de $m(t)$ que ultrapassam $\pm m_p$ são simplesmente truncadas. Portanto, m_p não é um parâmetro do sinal $m(t)$; é o limite do quantizador. O intervalo de amplitudes $(-m_p, m_p)$ é dividido em L intervalos uniformemente espaçados, cada um com largura $\Delta v = 2m_p/L$. Um valor de amostra é aproximado pelo valor no ponto médio do intervalo em que a amostra ocorre (Fig. 6.14). As amostras quantizadas são

*O erro na detecção dos pulsos pode ser desprezado.

codificadas e transmitidas como pulsos binários. No receptor, alguns pulsos podem ser detectados erroneamente. Em consequência, há duas fontes de erro neste esquema: *erro de quantização* e *erro de detecção de pulso*. Em quase todos os esquemas práticos, o erro de detecção de pulsos é muito pequeno em comparação com o erro de quantização, e pode ser desprezado. Assim, na análise a seguir, admitiremos que o erro no sinal recebido é causado exclusivamente pela quantização.

Se $m(kT_s)$ for a k-ésima amostra do sinal $m(t)$ e $\hat{m}(kT_s)$ for a correspondente amostra quantizada, da fórmula de interpolação na Eq. (6.10), temos

$$m(t) = \sum_k m(kT_s)\ \text{sinc}\ (2\pi Bt - k\pi)$$

e

$$\hat{m}(t) = \sum_k \hat{m}(kT_s)\ \text{sinc}\ (2\pi Bt - k\pi)$$

em que $\hat{m}(t)$ é o sinal reconstruído das amostras quantizadas. A componente de distorção $q(t)$ no sinal reconstruído é $q(t) = \hat{m}(t) - m(t)$. Logo,

$$\begin{aligned} q(t) &= \sum_k [\hat{m}(kT_s) - m(kT_s)]\ \text{sinc}\ (2\pi Bt - k\pi) \\ &= \sum_k q(kT_s)\ \text{sinc}\ (2\pi Bt - k\pi) \end{aligned}$$

em que $q(kT_s)$ é o erro de quantização na k-ésima amostra. O sinal $q(t)$ é o sinal indesejado e, portanto, atua como um ruído, conhecido como **ruído de quantização**. Para calcular a potência ou valor quadrático médio de $q(t)$, temos

$$\begin{aligned} \overline{q^2(t)} &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} q^2(t)\, dt \\ &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \left[\sum_k q(kT_s)\ \text{sinc}\ (2\pi Bt - k\pi) \right]^2 dt \end{aligned} \tag{6.29a}$$

Podemos mostrar (ver o Exercício 3.7-4) que os sinais sinc $(2\pi Bt - m\pi)$ e sinc $(2\pi Bt - n\pi)$ são ortogonais, ou seja,

$$\int_{-\infty}^{\infty} \text{sinc}\ (2\pi Bt - m\pi)\ \text{sinc}\ (2\pi Bt - n\pi)\, dt = \begin{cases} 0 & m \neq n \\ \dfrac{1}{2B} & m = n \end{cases} \tag{6.29b}$$

Por conta desse resultado, a integral dos termos cruzados no lado direito da Eq. (6.29a) se anula, e obtemos

$$\begin{aligned} \overline{q^2(t)} &= \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \sum_k q^2(kT_s)\ \text{sinc}^2\ (2\pi Bt - k\pi)\, dt \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_k q^2(kT_s) \int_{-T/2}^{T/2} \text{sinc}^2\ (2\pi Bt - k\pi)\, dt \end{aligned}$$

pela relação de ortogonalidade (6.29b), temos

$$\overline{q^2(t)} = \lim_{T\to\infty} \frac{1}{2BT} \sum_k q^2(kT_s) \tag{6.30}$$

Como a taxa de amostragem é $2B$, o número total de amostras no intervalo de média T é $2BT$. Assim, o lado direito da Eq. (6.30) representa o valor médio, ou a média do quadrado do erro de quantização. Os níveis de quantização são separados por $\Delta v = 2m_p/L$. Como um valor de amostra é aproximado pelo ponto médio do subintervalo (de altura Δv) em que a amostra é feita, o máximo erro de quantização é $\pm\Delta v/2$. Assim, o erro de quantização está no intervalo $(-\Delta v/2, \Delta v/2)$, sendo

$$\Delta v = \frac{2m_p}{L} \tag{6.31}$$

Admitindo que o erro possa assumir qualquer valor no intervalo $(-\Delta v/2, \Delta v/2)$ com igual probabilidade, o erro de quantização quadrático médio $\overline{q^2}$ é dado por*

$$\begin{aligned}\overline{q^2} &= \frac{1}{\Delta v}\int_{-\Delta v/2}^{\Delta v/2} q^2\,dq \\ &= \frac{(\Delta v)^2}{12} \qquad (6.32)\\ &= \frac{m_p^2}{3L^2} \qquad (6.33)\end{aligned}$$

Como $\overline{q^2(t)}$ é o valor quadrático médio ou potência do ruído de quantização, será denotado por N_q.

$$N_q = \overline{q^2(t)} = \frac{m_p^2}{3L^2}$$

Supondo que o erro de detecção de pulsos no receptor seja desprezível, o sinal reconstruído $\hat{m}(t)$ na saída do receptor é

$$\hat{m}(t) = m(t) + q(t)$$

O sinal desejado na saída é $m(t)$ e o ruído (erro) de quantização, $q(t)$. Como a potência do sinal de mensagem $m(t)$ é $\overline{m^2(t)}$, então

$$S_o = \overline{m^2(t)}$$

$$N_o = N_q = \frac{m_p^2}{3L^2}$$

e

$$\frac{S_o}{N_o} = 3L^2\frac{\overline{m^2(t)}}{m_p^2} \tag{6.34}$$

Nesta equação, m_p é o valor de pico da amplitude que o quantizador pode aceitar e, portanto, é um parâmetro do quantizador. Isso significa que S_o/N_o, a SNR, é uma função linear da potência $\overline{m^2(t)}$ do sinal de mensagem (ver a Fig. 6.18, com $\mu = 0$).

---

*Aqueles que têm familiaridade com a teoria da probabilidade podem deduzir esse resultado diretamente observando que a densidade de probabilidade do erro de quantização q é $1/(2m_p/L) = L/2m_p$ no intervalo $|q| \le m_p/L$ e zero para qualquer outro valor de q. Logo,

$$\overline{q^2} = \int_{-m_p/L}^{m_p/L} q^2 p(q)\,dq = \int_{-m_p/L}^{m_p/L} \frac{L}{2m_p} q^2\,dq = \frac{m_p^2}{3L^2}$$

6.2.3 Princípio da Taxação Progressiva: Quantização Não Uniforme

Recordemos que S_o/N_o, a SNR, é uma indicação da qualidade do sinal recebido. Idealmente, gostaríamos de ter uma SNR constante (mesma qualidade) para todos os valores de potência $m^2(t)$ do sinal de mensagem. Infelizmente, a SNR é diretamente proporcional à potência $m^2(t)$ do sinal, que chega a variar em até 40 dB (uma relação de potência de 10^4) de alto-falante para alto-falante. A potência de sinal também pode variar devido a diferentes comprimentos dos circuitos de conexão. Isso indica que a SNR na Eq. (6.34) pode variar muito, dependendo do alto-falante e do comprimento do circuito. Mesmo para um dado alto-falante, a qualidade do sinal recebido sofrerá degradação apreciável se a pessoa falar com voz suave. Estatisticamente, as menores amplitudes predominam na voz, e as maiores amplitudes ocorrem com frequência muito menor. Isso significa que a SNR será baixa na maior parte do tempo.

A raiz desta dificuldade reside no fato de que os incrementos de quantização têm valor uniforme $\Delta v = 2m_p/L$. O ruído de quantização $N_q = (\Delta v)^2/12$ [Eq. (6.32)] é diretamente proporcional ao quadrado do incremento. O problema pode ser resolvido com o uso de incrementos menores para amplitudes menores (quantização não uniforme), como mostrado na Fig. 6.15a. O mesmo resultado é obtido se as amostras de sinal forem, primeiro, comprimidas e, depois, quantizadas com incremento uniforme. A característica entrada-saída de um compressor é ilustrada na Fig. 6.15b. Nessa figura, o eixo horizontal é o sinal de entrada normalizado (ou seja, a amplitude do sinal de entrada m dividida pelo valor de pico do sinal m_p). O eixo vertical é o sinal de saída y. O compressor mapeia incrementos Δm do sinal de entrada em incrementos maiores Δy para sinais menores; o contrário se passa para grandes sinais de entrada. Dessa forma, um dado intervalo Δm contém um maior número de incrementos (ou incrementos de menor tamanho) quando m é pequeno. O erro de quantização é menor para menores valores de potência do sinal de entrada. Uma característica logarítmica aproximada de compressão produz um ruído de quantização que é praticamente proporcional à potência de sinal $m^2(t)$, resultando em uma SNR que praticamente independe da potência do sinal de entrada em um grande intervalo dinâmico[5] (ver a Fig. 6.18). Essa abordagem de equalização da SNR parece semelhante ao uso de taxação progressiva de imposto de renda para equalizar os rendimentos da população. Os que falam alto e sinais mais fortes são penalizados com maiores incrementos de ruído Δv para compensar os que falam baixo e os sinais mais fracos.

Entre as diversas opções, duas leis de compressão têm sido aceitas como padrões desejáveis pela ITU-T:[6] a lei μ, usada na América do Norte e no Japão, e a lei A, usada na Europa, no resto do mundo e em rotas internacionais. As curvas das leis μ e A têm simetria ímpar em relação ao eixo vertical. A lei μ (para amplitudes positivas) é dada por

$$y = \frac{1}{\ln(1+\mu)} \ln\left(1 + \frac{\mu m}{m_p}\right) \qquad 0 \le \frac{m}{m_p} \le 1 \tag{6.35a}$$

A lei A (para amplitudes positivas) é dada por

$$y = \begin{cases} \dfrac{A}{1+\ln A}\left(\dfrac{m}{m_p}\right) & 0 \le \dfrac{m}{m_p} \le \dfrac{1}{A} \\[2ex] \dfrac{1}{1+\ln A}\left(1 + \ln\dfrac{A m}{m_p}\right) & \dfrac{1}{A} \le \dfrac{m}{m_p} \le 1 \end{cases} \tag{6.35b}$$

Estas características são ilustradas na Fig. 6.16.

O parâmetro de compressão μ (ou A) determina o grau de compressão. Para que S_o/N_o seja praticamente constante em um intervalo dinâmico de 40 dB para a potência do sinal de entrada, μ deve ser maior que 100. Os primeiros bancos de canais e outros terminais digitais norte-americanos usavam um valor $\mu = 100$, que produzia os melhores resultados para codificação em 7 bits (128 níveis). Um valor ótimo $\mu = 255$ tem sido usado em todos os terminais digitais norte-americanos de 8 bits (256 ní-

Figura 6.15
Quantização não uniforme.

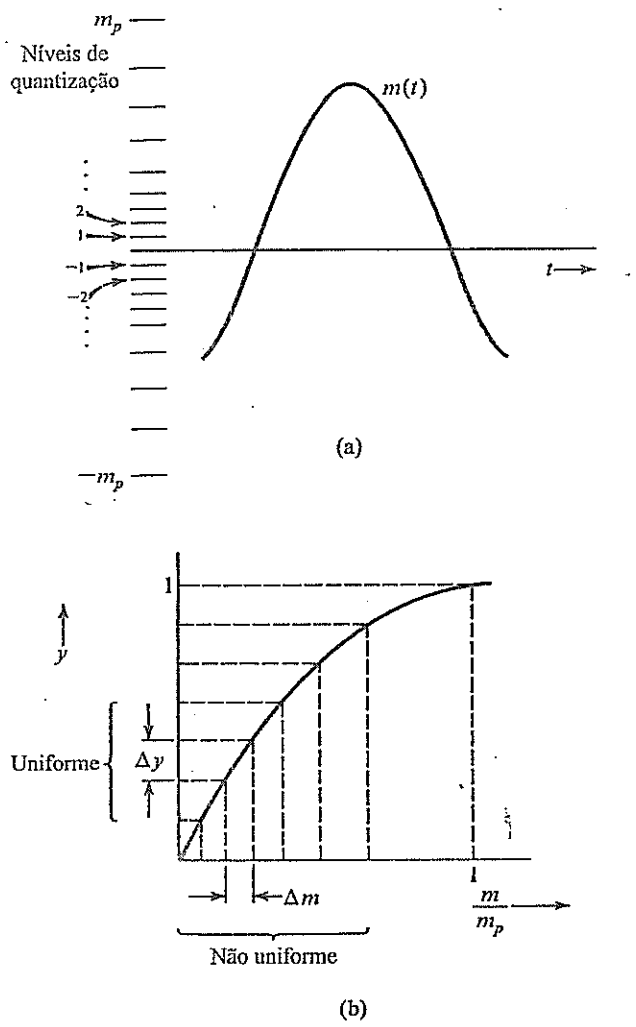
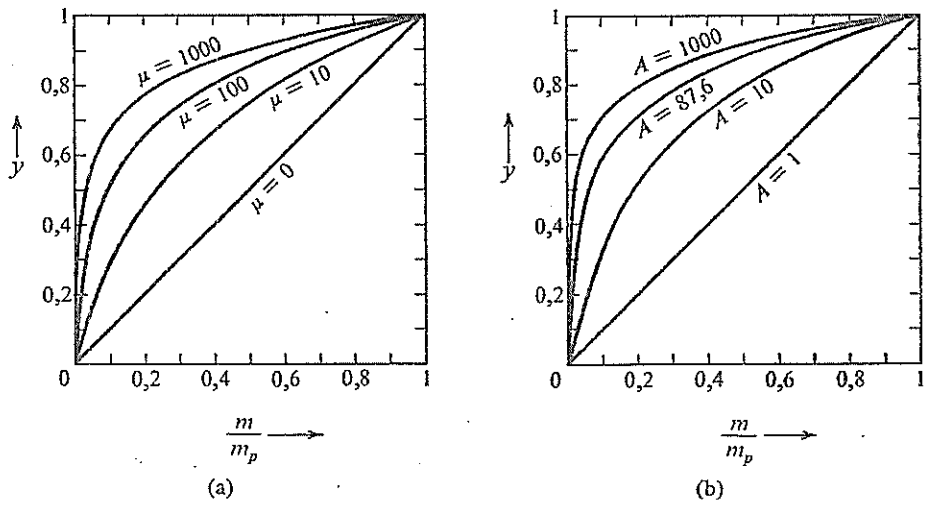


Figura 6.16
(a) Característica de lei μ. (b) Característica de lei A.



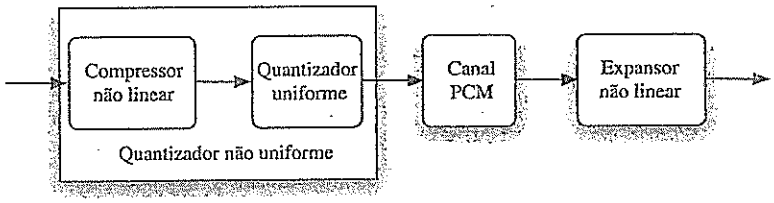


Figura 6.17 Utilização de compressor e expansor para quantização não uniforme.

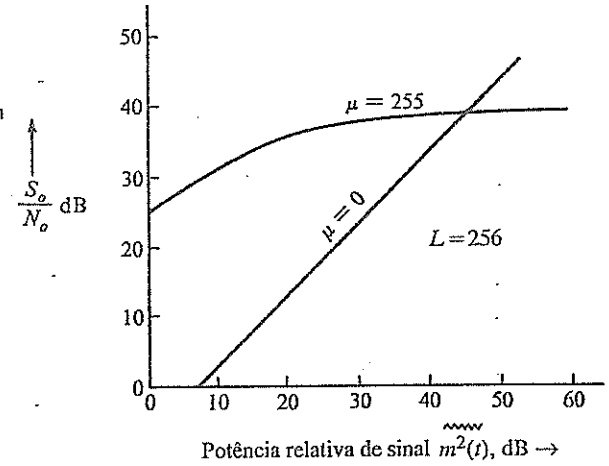


Figura 6.18 Razão entre sinal e ruído de quantização em PCM com e sem compressão.

veis), e o valor anterior de μ se tornou quase extinto. Para a lei A, um valor $A = 87{,}6$ produz resultados comparáveis e foi padronizado pela ITU-T.[6]

No receptor, as amostras comprimidas devem ser restauradas a seus valores originais; isso é feito com o uso de um expansor com característica complementar à do compressor. Juntos, compressor e expansor são denominados **compandor**. A Fig. 6.17 descreve o uso de compressor e expansor com um quantizador uniforme para obter quantização não uniforme.

De modo geral, a compressão temporal de um sinal aumenta a largura de banda do mesmo. No entanto, em PCM, o que é comprimido no tempo não é o sinal $m(t)$, mas suas amostras. Como, neste caso, nenhuma alteração é feita na escala de tempo e no número de amostras, o problema de aumento de largura de banda não ocorre. Quando um compandor de lei μ é usado, a SNR de saída é

$$\frac{S_o}{N_o} \simeq \frac{3L^2}{[\ln(1+\mu)]^2} \qquad \mu^2 \gg \frac{m_p^2}{m^2(t)} \tag{6.36}$$

A Fig. 6.18 mostra, para os casos $\mu = 255$ e $\mu = 0$ (quantização uniforme), a SNR de saída em função de $m^2(t)$ (potência do sinal de mensagem).

Compandor

Um compressor logarítmico pode ser realizado com um diodo semicondutor, pois a característica V-I de um diodo desse tipo tem, no primeiro quadrante, a forma desejada

$$V = \frac{KT}{q} \ln\left(1 + \frac{I}{I_s}\right)$$

Dois diodos casados em paralelo, com polaridades opostas, produzem uma característica desejada aproximada no primeiro e no terceiro quadrantes (ignorando a corrente de saturação). Na prática, resistores ajustáveis são posicionados em série com cada diodo, e um terceiro resistor variável é adicionado em paralelo. O ajuste de vários resistores possibilita que a característica resultante iguale a característica ideal em um número finito de pontos (em geral, sete).

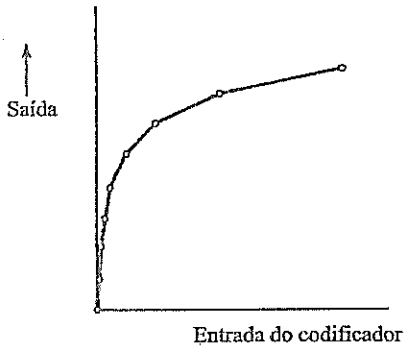


Figura 6.19 Característica linear por partes de compressor.

Uma abordagem alternativa consiste em usar uma aproximação linear por partes para a característica logarítmica. Uma aproximação com 15 segmentos (Fig. 6.19) à lei de 8 bits ($L = 256$), com $\mu = 255$, é largamente usada em banco de canais D2 usado em conjunção com o sistema de portadora T1. A aproximação segmentada é apenas marginalmente inferior em termos de SNR.[8] A aproximação linear por partes substituiu quase totalmente a anterior aproximação logarítmica à verdadeira característica $\mu = 255$, e se tornou o método preferencial adotado em padrões norte-americanos. Embora o conjunto de um verdadeiro compressor com $\mu = 255$ trabalhando com um expansor com $\mu = 255$ seja superior a dispositivos similares lineares por partes, um terminal digital com a característica verdadeira tem, em uma rede atual, de trabalhar com outros elementos de rede baseados na aproximação linear por partes. Essa combinação de características diferentes tem desempenho inferior ao obtido quando o compressor e o expansor operam segundo a mesma lei de compressão.

No formato padrão de arquivo de áudio usado por Sun, Unix e Java, o áudio em arquivos "au" pode ser modulado por codificação de pulsos ou comprimido com o padrão ITU-T G.711 usando a lei μ ou a lei A.[6] O compressor de lei μ (com $\mu = 255$) converte amostras PCM lineares assinadas de 14 bits a amostras logarítmicas de 8 bits, favorecendo economia de armazenagem. O compressor de lei A (com $A = 87{,}6$) converte amostras PCM lineares positivas ou negativas de 13 bits a amostras logarítmicas de 8 bits. Nos dois casos, com taxa de amostragem de 8000 Hz, o codificador G.77 cria, a partir de sinais de áudio, sequências de bit a 64 quilobits por segundo (kbit/s). Como as leis A e μ são mutuamente compatíveis, áudio gravado em arquivos "au" pode ser decodificado em qualquer dos dois formatos. Vale notar que o formato de áudio WAV da Microsoft também tem opções de compressão que usam a lei μ e a lei A.

Codificador PCM

A saída multiplexada PAM é aplicada à entrada do codificador, que quantiza e codifica cada amostra em um grupo de n dígitos binários. Embora exista uma variedade de codificadores,[7,10] discutiremos o codificador de **um dígito por vez** (*digit-at-a-time*), que faz n comparações sequenciais para gerar uma palavra de código de n bits. A amostra é comparada com uma tensão obtida por uma combinação de tensões de referência proporcionais a $2^7, 2^6, 2^5, ..., 2^0$. As tensões de referência são convenientemente geradas por um banco de resistores $R, 2R, 2^2R, ..., 2^7R$.

A codificação exige respostas a sucessivas perguntas, das quais a primeira é se a amostra está na metade superior ou inferior do intervalo permitido. O primeiro dígito de código 1 ou 0 é gerado, dependendo se a amostra estiver na metade superior ou inferior do intervalo. No segundo passo, outro dígito 1 ou 0 é gerado, dependendo se a amostra estiver na metade superior ou inferior do subintervalo em que está localizada. O processo continua até que o último dígito binário no código tenha sido gerado.

A decodificação é o inverso da codificação. Neste caso, cada um dos n dígitos é aplicado a um resistor de valor diferente. O k-ésimo dígito é aplicado a um resistor 2^kR. As correntes em todos os resistores são somadas. A soma é proporcional ao valor da amostra quantizada. Por exemplo, uma palavra

de código binário **10010110** produzirá uma corrente proporcional a $2^7 + 0 + 0 + 2^4 + 0 + 2^2 + 2^1 + 0 = 150$. Isso completa a conversão D/A.

6.2.4 Largura de Banda de Transmissão e SNR de Saída

Para PCM binária, alocamos um grupo distinto de n dígitos binários (bits) a cada um dos L níveis de quantização. Como a sequência de n dígitos binários pode ser organizada em 2^n padrões diferentes,

$$L = 2^n \qquad \text{ou} \qquad n = \log_2 L \tag{6.37}$$

cada amostra quantizada é, portanto, codificada em n bits. Como um sinal $m(t)$ limitado em banda a B Hz requer um mínimo de $2B$ amostras por segundo, precisamos de $2nB$ bits/s, ou seja, $2nB$ porções de informação por segundo. Tendo em vista que uma largura de banda unitária (1 Hz) pode transmitir um máximo de duas porções de informação por segundo (Seção 6.1.3), precisamos de uma largura de banda mínima de canal B_T Hz dada por

$$B_T = nB \text{ Hz} \tag{6.38}$$

Essa é a mínima largura de banda teórica para a transmissão de um sinal PCM. Nas Seções 7.2 e 7.3, mostraremos que, por motivos práticos, podemos usar larguras de banda maiores que esse mínimo teórico.

Exemplo 6.2 Um sinal $m(t)$ limitado em banda a 3 kHz é amostrado a uma taxa $33\frac{1}{3}\%$ maior que a taxa de Nyquist. O erro máximo aceitável na amplitude de amostras (máximo erro de quantização) é 0,5% da amplitude de pico m_p. As amostras quantizadas são codificadas em binário. Determinemos a mínima largura de banda de um canal necessária para a transmissão do sinal binário codificado. Se 24 desses sinais forem multiplexados por divisão no tempo, determinemos a mínima largura de banda necessária para a transmissão do sinal multiplexado.

A taxa de amostragem de Nyquist é $R_N = 2 \times 3000 = 6000$ Hz (amostras por segundo). A real taxa de amostragem é $R_A = 6000 \times (1\frac{1}{3}) = 8000$ Hz.

O incremento de quantização é Δv e o máximo erro de quantização é $\pm\Delta v/2$. Portanto, a partir da Eq. (6.31),

$$\frac{\Delta v}{2} = \frac{m_p}{L} = \frac{0{,}5}{100} m_p \Longrightarrow L = 200$$

Para codificação binária, L deve ser uma potência de 2. Assim, o próximo valor de L maior que 200 que é uma potência de 2 é $L = 256$.

Pela Eq. (6.37), precisamos de $n = \log_2 256 = 8$ bits por amostra. Precisaremos transmitir um total de $C = 8 \times 8000 = 64.000$ bits/s. Como podemos transmitir até 2 bits/s por hertz de largura de banda, necessitaremos de uma largura de banda mínima de transmissão $B_T = C/2 = 32$ kHz.

O sinal multiplexado tem um total de $C_M = 24 \times 64.000 = 1{,}536$ Mbit/s, o que requer uma largura de banda mínima de transmissão de $1{,}536/2 = 0{,}768$ MHz.

Aumento Exponencial da SNR de Saída

Da Eq. (6.37), $L^2 = 2^{2n}$, e a SNR de saída na Eq. (6.34) ou Eq. (6.36) pode ser expressa como

$$\frac{S_o}{N_o} = c(2)^{2n} \tag{6.39}$$

em que

$$c = \begin{cases} \dfrac{3\,\overline{m^2(t)}}{m_p^2} & \text{[caso não comprimido, Eq. (6.34)]} \\[2ex] \dfrac{3}{[\ln(1+\mu)]^2} & \text{[caso comprimido, Eq. (6.36)]} \end{cases}$$

A substituição da Eq. (6.38) na Eq. (6.39) resulta em

$$\frac{S_o}{N_o} = c(2)^{2B_T/B} \tag{6.40}$$

Da Eq. (6.40), observamos que a SNR aumenta exponencialmente com a largura de banda de transmissão B_T. Essa barganha de SNR por largura de banda é interessante e se aproxima do limite teórico. Um pequeno aumento na largura de banda resulta em um grande benefício em termos de SNR. Essa relação é vista claramente quando usamos a escala em decibéis para escrever a Eq. (6.39)

$$\begin{aligned} \left(\frac{S_o}{N_o}\right)_{\text{dB}} &= 10\log_{10}\left(\frac{S_o}{N_o}\right) \\ &= 10\log_{10}[c(2)^{2n}] \\ &= 10\log_{10} c + 2n\log_{10} 2 \\ &= (\alpha + 6n)\ \text{dB} \end{aligned} \tag{6.41}$$

na qual $\alpha = 10\log_{10} c$. Isso mostra que o aumento de n em 1 (acrescentar um bit na palavra de código) quadruplica a SNR de saída (aumento de 6 dB). Assim, se aumentarmos n de 8 para 9, a SNR quadruplica, mas a largura de banda de transmissão aumenta apenas de 32 kHz para 36 kHz (aumento de 12,5%). Isso mostra que, em PCM, a SNR pode ser controlada pela largura de banda de transmissão. Veremos, mais tarde, que isso também ocorre com modulações em frequência e em fase. Mas, nestes casos, para quadruplicar a SNR, é necessário dobrar a largura de banda. Neste ponto, PCM é muito superior a FM ou PM.

Exemplo 6.3 Um sinal $m(t)$ de largura de banda $B = 4$ kHz é transmitido com o uso de PCM compandido binário com $\mu = 100$. Comparemos o caso $L = 64$ com o caso $L = 256$, quanto à largura de banda de transmissão e SNR de saída.

Para $L = 64$, $n = 6$ e a largura de banda de transmissão é $nB = 24$ kHz,

$$\frac{S_o}{N_o} = (\alpha + 36)\ \text{dB}$$

$$\alpha = 10\log\frac{3}{[\ln(101)]^2} = -8{,}51$$

Logo,

$$\frac{S_o}{N_o} = 27{,}49\ \text{dB}$$

Para $L = 256$, $n = 8$ e a largura de banda de transmissão é $nB = 32$ kHz,

$$\frac{S_o}{N_o} = \alpha + 6n = 39{,}49\ \text{dB}$$

A diferença entre as duas SNRs é 12 dB, que é uma razão de 16. Assim, a SNR para $L = 256$ é 16 vezes a SNR para $L = 64$. A largura de banda no primeiro caso é apenas 33% maior que a no segundo.

Comentários sobre Unidades Logarítmicas

Unidades logarítmicas e escalas logarítmicas são convenientes quando uma variável tem um intervalo dinâmico grande. Este é o caso de variáveis de frequência e SNRs. Uma unidade logarítmica para a razão de potência é o decibel (dB), definido como 10 $\log_{10}$ (razão de potência). Assim, uma SNR é x dB, sendo

$$x = 10 \log_{10} \frac{S}{N}$$

Usamos a mesma unidade para expressar ganho ou perda de potência em certos meios de transmissão. Por exemplo, se, ao longo de um certo cabo, a potência de sinal for atenuada por um fator de 15, o ganho do cabo é

$$G = 10 \log_{10} \frac{1}{15} = -11{,}76 \text{ dB}$$

ou a atenuação (perda) do cabo, 11,76 dB.

Embora seja uma medida de razões de potência, decibel também é largamente utilizado como uma medida de potência. Por exemplo, uma potência de "100 watts" pode ser considerada com uma razão de 100 em relação a uma potência de 1 watt, e expressa em unidades de dBW como

$$P_{\text{dBW}} = 10 \log_{10} 100 = 20 \text{ dBW}$$

Assim, uma potência de 100 watts corresponde 20 dBW. De modo similar, potência medida em relação à potência de 1 mW é expressa em dBm. Por exemplo, uma potência de 100 watt é

$$P_{\text{dBm}} = 10 \log \frac{100 \text{ W}}{1 \text{ mW}} = 50 \text{ dBm}$$

6.3 TELEFONIA DIGITAL: PCM EM SISTEMAS DE PORTADORAS

Nota Histórica

Devido à indisponibilidade de dispositivos de comutação apropriados, mais de 20 anos se passaram entre a invenção da PCM e sua implementação. Válvulas a vácuo, usadas antes da invenção do transistor, não eram apenas volumosas, mas eram comutadores pobres e dissipavam muito calor. Sistemas que utilizam válvulas como comutadores são grandes, nada confiáveis e tendem a sofrer de superaquecimento. A PCM aguardava a invenção do transistor, um dispositivo pequeno, que consome pouca potência e é um comutador quase ideal.

Coincidentemente, na mesma época em que o transistor foi inventado, a demanda por serviços telefônicos aumentou tanto que o sistema existente ficou sobrecarregado, particularmente em grandes cidades. Não era simples a instalação de novos cabos subterrâneos, porque o espaço sob as ruas de muitas cidades já estava ocupado com outros serviços (água, gás, esgoto etc.) Além disso, a escavação de ruas e o resultante transtorno não eram bem vistos. Foi feita uma tentativa, em escala limitada, de aumentar a capacidade com a multiplexação por divisão em frequência de vários canais de voz por meio de modulação em amplitude. Infelizmente, os cabos haviam sido projetados principalmente para a faixa de frequências da voz (0-4 kHz) e eram extremamente afetados por ruído. Um problema adicional era a interferência entre pares de canais em um mesmo cabo, inaceitável em frequências altas. Ironicamente, a PCM – que exigia uma largura de banda muitas vezes maior que a necessária para sinais FM – era a solução. Isso se devia ao fato de que sistemas digitais com repetidores regeneradores

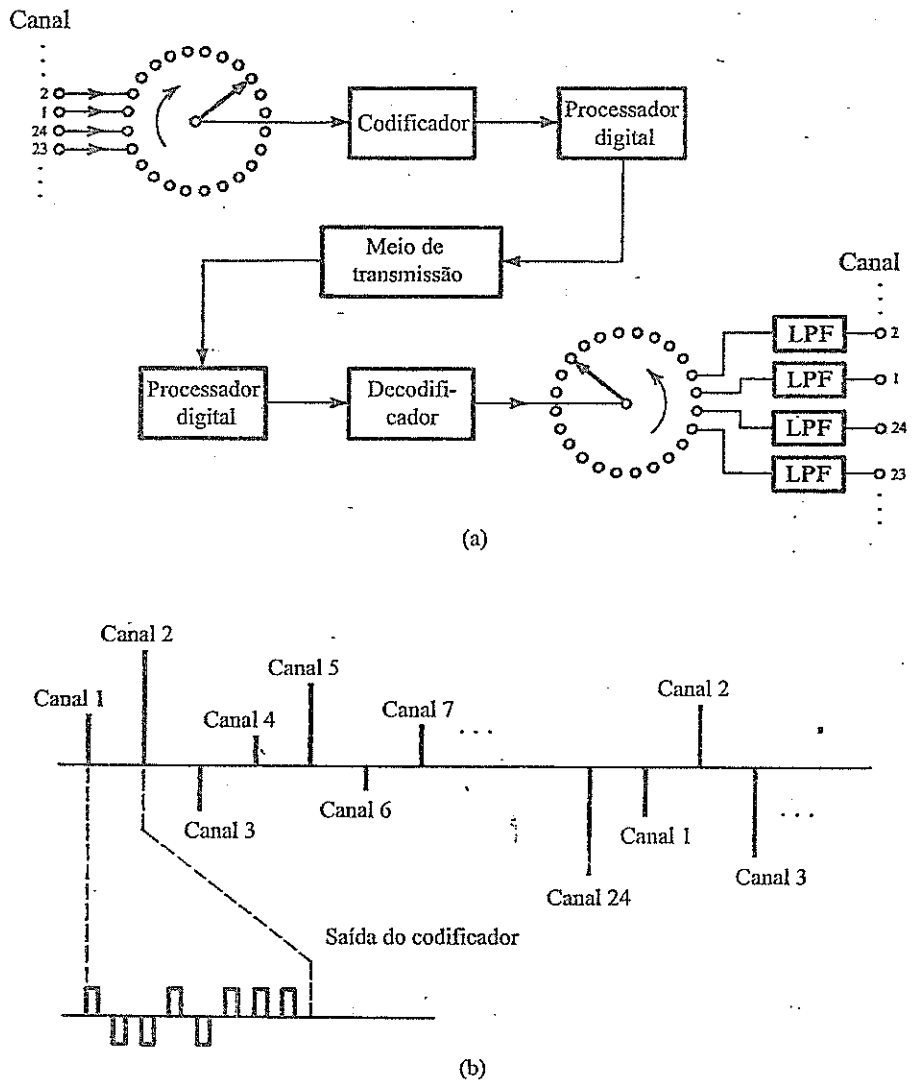


Figura 6.20 Sistema de portadora T1.

proximamente espaçados são capazes de trabalhar de modo satisfatório em linhas ruidosas que têm baixo desempenho em altas frequências.[9] Os repetidores, espaçados a cada 1,8 km aproximadamente, limpavam o sinal e regeneravam os pulsos antes que se tornassem demasiadamente distorcidos e ruidosos. Essa é a história do sistema de portadora T1 do grupo Bell System.[3,10] Um par de fios usado para transmitir um sinal de áudio com 4 kHz de largura de banda passou a transmitir 24 sinais telefônicos PCM multiplexados por divisão no tempo, com largura de banda total de 1,544 MHz.

Multiplexação por Divisão no Tempo T1

Um esquema de um sistema de portadora T1 é mostrado na Fig. 6.20a. Todos os 24 canais são amostrados em sequência. A saída do amostrador representa um sinal PAM multiplexado por divisão no tempo. O sinal PAM multiplexado é, então, aplicado à entrada de um codificador que quantiza cada amostra e a codifica em oito pulsos binários – uma palavra de código binária* (Fig. 6.20b). O sinal, agora convertido à forma digital, é enviado ao longo do meio de transmissão. Repetidores regeneradores espaçados em cerca de 1,8 km detectam os pulsos e retransmitem pulsos novos. No receptor, o

*Em uma versão anterior, cada amostra era codificada por sete bits. Um bit adicional foi acrescentado para sinalização.

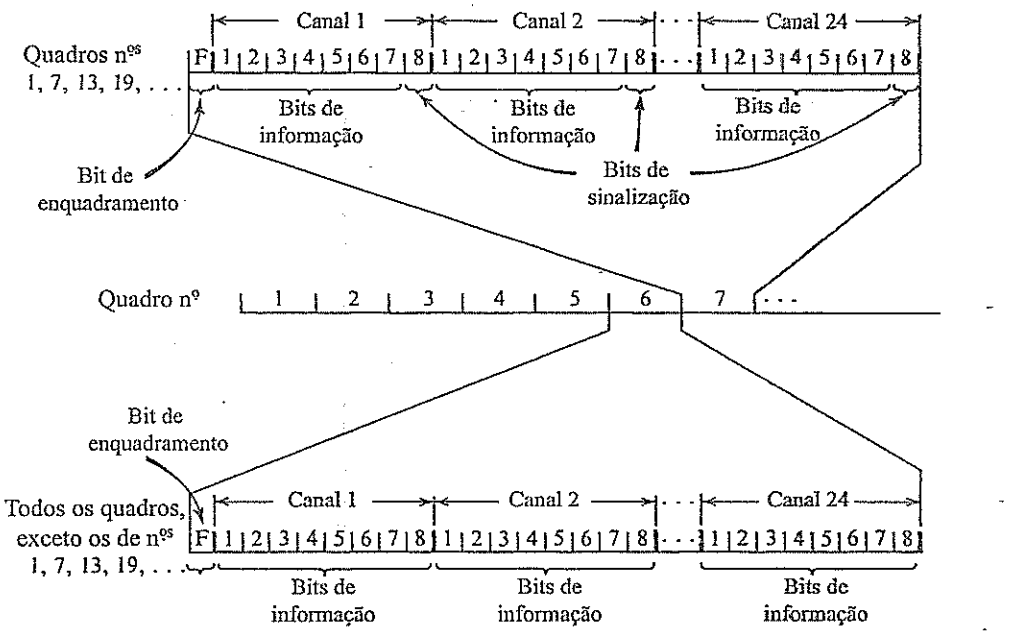


Figura 6.21 Formato de sinalização do sistema T1.

decodificador converte os pulsos binários em amostras (decodificação). As amostras são, então, demultiplexadas (ou seja, distribuídas a cada um dos 24 canais). O desejado sinal de áudio é reconstruído com a aplicação das amostras a um filtro passa-baixas em cada canal.

Os comutadores na Fig. 6.20 não são mecânicos: são circuitos comutadores eletrônicos de alta velocidade. Existem diversos esquemas para este propósito.[11] A amostragem é feita por portas (*gates*) eletrônicas (como um circuito ponte de diodos, como mostrado na Fig. 4.5a) abertas periodicamente por pulsos estreitos de 2 μs de duração. O sinal de 1,544 Mbit/s do sistema T1, chamado **nível de sinal digital 1 (DS1 – *digital signal level 1*)**, é progressivamente multiplexado para formar sinais de níveis mais altos DS2, DS3 e DS4, como descrito na Seção 6.4.

Depois de o grupo Bell System ter introduzido o sistema de portadora T1 nos Estados Unidos, dezenas de variações foram propostas ou adotadas em outros locais, até que a ITU-T padronizasse o sistema PCM de 30 canais, com uma taxa de 2,048 Mbit/s (em contraste com o sistema T1, de 24 canais e 1,544 Mbit/s). O sistema de 30 canais é usado em todo o mundo, exceto na América do Norte e no Japão. Devido à adoção do sistema de portadora T1 em larga escala na América do Norte e no Japão antes da padronização pela ITU-T, os dois padrões continuam em uso em diferentes partes do mundo, com interfaces apropriadas em conexões internacionais.

Sincronização e Sinalização

Palavras de código binárias correspondentes às amostras de cada um dos 24 canais são multiplexadas em uma sequência, como indicado na Fig. 6.21. Um segmento contendo uma palavra de código (correspondente a uma amostra) de cada um dos 24 canais recebe o nome de **quadro** (*frame*). Cada quadro tem $24 \times 8 = 192$ bits de informação. Como a taxa de amostragem é de 8000 amostras por segundo, cada quadro ocupa 125 μs. Para separar os bits de informação corretamente no receptor, é necessário ter certeza do ponto de início de cada quadro. Para isso, um **bit de enquadramento** (*framing bit*) é adicionado ao início de cada quadro. O número total de bits em um quadro, então, passa a 193. Bits de enquadramento são escolhidos de modo que uma sequência desses bits, um no início de cada quadro, forme um padrão especial, cuja ocorrência em um sinal de voz seja improvável.

A sequência formada com o primeiro bit de cada quadro é examinada pela lógica do terminal receptor. Caso a sequência não siga o dado padrão de código (padrão dos bits de enquadramen-

to), uma perda de sincronismo é detectada e a próxima posição é examinada para determinar se é o bit de enquadramento. A detecção leva de 0,4 a 6 ms e o reenquadramento, cerca de 50 ms (no pior caso).

Além de bits de informação e de enquadramento, precisamos transmitir bits de sinalização correspondentes aos pulsos de discagem, assim como sinais telefônicos de livre e ocupado. Quando canais desenvolvidos por esse sistema são usados na transmissão de sinais entre sistemas de comutação telefônica, os comutadores devem ser capazes de se comunicarem entre si, para que os canais sejam usados de forma eficaz. Como todos os oito bits agora são usados para transmissão, em vez dos sete da versão anterior,* o canal de sinalização provido pelo oitavo bit deixou de existir. Como basta um canal de sinalização de baixa velocidade, em vez de criar uma janela temporal adicional para esta informação, usamos um bit de informação (o bit menos significativo) a cada seis amostras de um sinal para transmitir essa informação. Isso significa que cada sexta amostra de um sinal de voz terá um possível erro correspondente ao bit menos significativo. Cada sexto quadro tem, portanto, $7 \times 24 = 168$ bits de informação, 24 bits de sinalização e 1 bit de enquadramento. Todos os outros quadros terão 192 bits de informação e 1 bit de enquadramento. Essa técnica é denominada codificação de bit 7⅚, e o canal de sinalização recebe o nome **sinalização por bit roubado** (*robbed-bit signaling*). A pequena degradação da SNR decorrente da alteração de um de cada seis quadros é considerada uma penalidade aceitável. Os bits de sinalização para cada sinal ocorrem a uma taxa de $8000/6 = 1333$ bit/s. O formato de quadro é mostrado na Fig. 6.21.

O antigo formato de quadros com sete bits requeria apenas identificação das fronteiras dos quadros para que cada canal pudesse ser localizado na sequência de bits. Quando a sinalização é superposta aos canais a cada seis quadros, torna-se necessário identificar, no receptor, os quadros que são quadros de sinalização. Uma nova estrutura de quadros, denominada **superquadro**, foi desenvolvida para cuidar disso. Os bits de enquadramento são transmitidos a 8 kbit/s como antes e ocupam o primeiro bit de cada quadro. Os bits de enquadramento formam um padrão especial, que se repete a cada 12 quadros: **100011011100**. Assim, esse padrão permite a identificação das fronteiras dos quadros como antes, mas também permite a determinação da localização do sexto e do décimo segundo quadro no superquadro. Vale notar que o superquadro descrito aqui tem 12 quadros de comprimento. Como dois bits por superquadro são disponíveis para sinalização para cada canal, é possível prover sinalização de quatro estados para um canal com o uso dos quatro padrões possíveis para os dois bits de sinalização: **00**, **01**, **10** e **11**. Embora a maioria das aplicações comutador a comutador na rede de telefonia exija apenas sinalização de dois estados, técnicas de sinalização de três e quatro estados são usadas em certas aplicações especiais.

Avanços na eletrônica digital e na teoria de codificação tornaram desnecessário o uso de todos os 8 kbits/s do canal de enquadramento em um sistema DS1 para implementar a tarefa de enquadramento. Uma nova estrutura de superquadro, denominada formato de **superquadro estendido** (ESF – *extended superframe*), foi introduzida na década de 1970 para tirar proveito da reduzida exigência de largura de banda para enquadramento. Um ESF tem 24 quadros de comprimento e carrega bits de sinalização a cada oito bits de cada canal nos quadros 6, 12, 18 e 24. Isso possibilita sinalização em dezesseis estados, que é usada em alguns casos, embora, com o formato de superquadro, a maioria das aplicações requeira somente sinalização em dois estados.

Os 8 kbit/s de *overhead* de capacidade (de enquadramento) do sinal ESF são divididos em três canais: 2 kbit/s para enquadramento, 2 kbit/s para um canal de detecção de erro por verificação cíclica de redundância (CRC-6 – *cyclic redundancy check*) e 4 kbit/s para um canal de dados. A detecção de erro de alta confiabilidade provida pelo padrão CRC-6 e o uso de um canal de dados para transportar informação sobre o desempenho de sinal como recebido pelo terminal distante fazem com que o formato ESF seja muito mais atraente para provedores de serviços que o antigo formato de superquadro. O Capítulo 14 apresenta mais detalhes sobre detecção de erro por CRC.

---

*Na versão anterior de T1, os níveis de quantização $L = 128$ requeriam apenas sete bits de informação. O oitavo bit era usado para sinalização.

O canal de enquadramento de 2 kbit/s do formato ESF transporta o padrão repetitivo **001011...**, um padrão que se repete a cada 24 quadros e é muito menos vulnerável a falsificação que padrões associados a formatos anteriores.

Por diversas razões, incluindo o desenvolvimento de nós comutadores de rede inteligente, a função da sinalização está sendo transferida dos canais que transportam as mensagens ou sinais de dados para redes de sinalização separadas, denominadas sistemas de **canal comum de sinalização entre centrais (CCIS – *common channel interoffice signaling*)**. O emprego universal desse sistema diminuirá a importância da sinalização por bit roubado, e todos os oito bits de cada mensagem (ou amostra) serão transmitidos na maioria das aplicações.

A Conferência sobre Administração Postal e Telegráfica Europeia (CEPT – *Conference on European Postal and Telegraph Administration*) padronizou um PCM com 256 janelas temporais por quadro. Cada quadro tem $30 \times 8 = 240$ bits de informação, correspondendo a 30 canais de voz (cada um com oito bits). Os restantes 16 bits por quadro são usados para sincronização e sinalização. Portanto, embora a taxa de bits seja 2,048 Mbit/s, o que corresponde a 32 canais de voz, somente 30 canais de voz são transmitidos.

6.4 MULTIPLEXAÇÃO DIGITAL

Vários sinais de baixas taxas de bits podem ser multiplexados, ou combinados, para formar um sinal a uma taxa de bits mais elevada, a ser transmitido em um meio de alta frequência. Como o meio é compartilhado no tempo por vários sinais, este é um caso de TDM (multiplexação por divisão no tempo). Os sinais dos diversos canais de entrada, ou tributários, podem ter naturezas variadas, como sinais de voz digitalizada (PCM), uma saída de computador, dados de telemetria ou um fac-símile digital. Não é necessário que as taxas de bits dos diversos tributários sejam iguais.

Consideremos, inicialmente, o caso em que todos os tributários têm a mesma taxa de bits. A multiplexação pode ser feita bit a bit (o que é conhecido como **entrelaçamento de bits**), como mostrado na Fig. 6.22a, ou palavra a palavra (o que é conhecido como entrelaçamento de palavras). A Fig. 6.22b mostra o entrelaçamento de palavras, formadas por quatro bits. A hierarquia digital norte-americana usa entrelaçamento de bits (exceto no nível mais baixo), no qual um bit por vez é tirado de cada um dos sinais a serem multiplexados. O entrelaçamento de bits, usado na construção do sinal DS1 e de sinais formatados para SONET*, consiste na inserção alternada de bytes dos canais a serem multiplexados.

A portadora T1, discutida na Seção 6.3, usa o entrelaçamento de palavras de oito bits. Quando as taxas de bits dos sinais de entrada não são iguais ao canal de taxa de bits mais elevada são alocadas proporcionalmente mais janelas temporais. A multiplexação de quatro canais consiste em três canais, B, C e D, com a mesma taxa de bits R e um (canal A) com taxa de bits $3R$. (Fig. 6.22c, d). Resultados similares podem ser obtidos com a combinação de palavras de comprimentos diferentes. Fica evidente que o comprimento mínimo do quadro multiplexado deve ser um múltiplo do mínimo múltiplo comum das taxas de bits dos sinais de entrada e, portanto, este tipo de esquema é prático somente quando existe uma relação simples entre as várias taxas. O caso de canais totalmente assíncronos é discutido posteriormente.

No terminal receptor, a sequência de dígitos de entrada deve ser dividida e distribuída ao canal de saída apropriado. Para tanto, o terminal receptor deve ser capaz de identificar cada bit corretamente. Isso requer que o sistema receptor tenha sincronismo temporal com o início de cada quadro, com cada janela (*slot*) em um quadro, e com cada bit em uma janela. Para isso, bits de sincronismo e de enquadramento são adicionados aos bits de dados. Esses bits adicionais são parte dos chamados **bits de *overhead***.

**synchronous optical networks*, ou redes ópticas síncronas, versão óptica da hierarquia digital síncrona, de uso padronizado na América do Norte. Existe um padrão internacional semelhante, SDH – *synchronous digital hierarchy*, utilizado no resto do mundo. (N.T.)

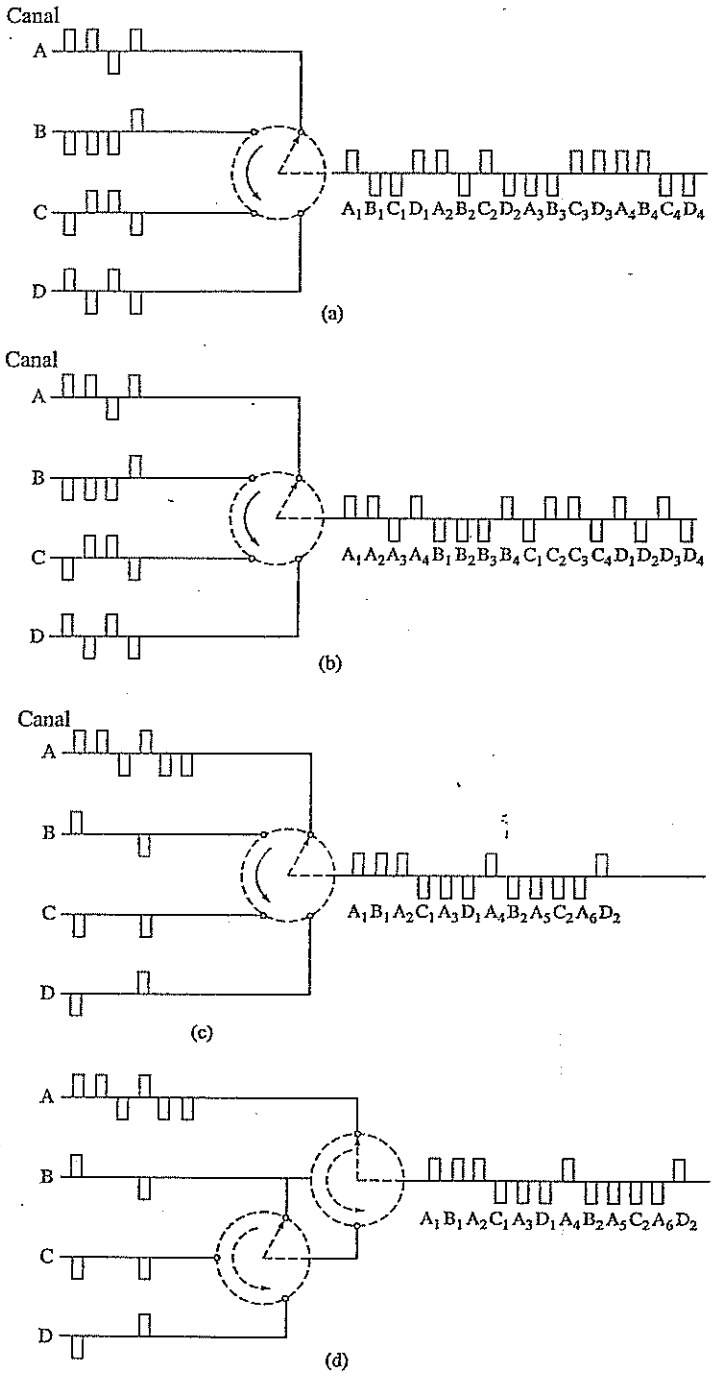


Figura 6.22 Multiplexação por divisão no tempo de sinais digitais: (a) entrelaçamento de dígitos; (b) entrelaçamento de palavras (ou bytes); (c) entrelaçamento de canais com diferentes taxas de bits; (d) esquema alternativo para (c).

6.4.1 Formato de Sinal

A Fig. 6.23 ilustra um formato típico, o do multiplexador DM1/2, em que há entrelaçamento bit a bit de quatro canais, cada um à taxa de 1,544 Mbit/s. O quadro principal (multiquadro) consiste em quatro subquadros. Cada subquadro tem seis bits de *overhead*; por exemplo, o subquadro 1 (primeira

Figura 6.23 Formato de multiplexador DM1/2.

M_0	[48]	C_A	[48]	F_0	[48]	C_A	[48]	C_A	[48]	F_1	[48]
M_1	[48]	C_B	[48]	F_0	[48]	C_B	[48]	C_B	[48]	F_1	[48]
M_1	[48]	C_C	[48]	F_0	[48]	C_C	[48]	C_C	[48]	F_1	[48]
M_1	[48]	C_D	[48]	F_0	[48]	C_D	[48]	C_D	[48]	F_1	[48]

linha na Fig. 6.23) tem bits de *overhead* M_0, C_A, F_0, C_A, C_A e F_1. Entre os bits de *overhead*, há 48 bits de dados entrelaçados, provenientes dos quatro canais (12 bits de dados de cada canal). Iniciamos com o bit de *overhead* M_0, seguido por 48 bits de dados multiplexados, adicionamos um segundo bit de *overhead* C_A, seguido por 48 bits de dados multiplexados, e assim por diante. Portanto, há um total de $48 \times 6 \times 4 = 1152$ bits de dados e $6 \times 4 = 24$ bits de *overhead* perfazendo um total de 1176 bits/quadro. A eficiência é de $1152/1176 \simeq 98\%$. Os bits de *overhead* com subscrito 0 são sempre **0** e aqueles com subscrito 1, sempre **1**. Assim, M_0, F_0 são todos **0**s, M_1, F_1 são todos **1**s. Os dígitos F são periódicos **010101**... e provêm o principal padrão de quadros que o multiplexador usa para obter sincronismo com quadro. Após alcançar sincronismo por meio deste padrão, o demultiplexador busca o padrão **0111** formado pelos bits de *overhead* $M_0M_1M_1M_1$. Isso identifica os quatro subquadros, cada um correspondendo a uma linha na Fig. 6.23. É possível, embora improvável, que bits de sinal também tenham um padrão **101010**... O receptor poderia ficar travado à sequência incorreta. A presença de $M_0M_1M_1M_1$ permite a confirmação da verdadeira sequência $F_0F_1F_0F_1$. Os bits C são usados para transmitir informação adicional sobre preenchimento de bits, como discutido mais adiante.

Na maioria dos casos, nem todos os canais de entrada são ativos todo o tempo: alguns transmitem dados, outros estão ociosos. Isso significa que o sistema é subutilizado. Para, em um dado momento, tirar proveito da inatividade de pelo menos um canal, podemos aceitar mais canais de entrada. Obviamente, isso envolve operações de comutação muito mais complicadas, além de um planejamento cuidadoso do sistema. Em qualquer situação de tráfego aleatório, não podemos garantir que a demanda por canais de transmissão não ultrapassará o número de canais disponíveis; contudo, por meio de uma estatística das fontes de sinal, é possível assegurar uma probabilidade baixa aceitável de que isso ocorra. Estruturas de multiplexação desse tipo foram desenvolvidas para sistemas de comunicação por satélite e são conhecidas como **sistemas de acesso múltiplo por divisão no tempo (TDMA – *time division multiple-access*)**.

Nos sistemas TDMA empregados em telefonia, os parâmetros de projeto são escolhidos de modo que qualquer condição de sobrecarga dure apenas uma fração de um segundo, o que garante desempenho aceitável para comunicação de voz. Para outros tipos de dados e telegrafia, atrasos de transmissão são irrelevantes. Portanto, em situação de sobrecarga, os dados de entrada podem ser armazenados e transmitidos posteriormente.

6.4.2 Canais Assíncronos e Preenchimento de Bits

Na discussão anterior, presumimos a existência de sincronismo entre todos os canais de entrada e o multiplexador. Isso é difícil de ocorrer, mesmo quando todos os canais têm iguais taxas nominais. Por exemplo, consideremos um cabo coaxial de 1000 km que transporta 2×10^8 m/s pulsos por segundo. Supondo que a velocidade nominal de propagação no cabo seja 2×10^8 m/s, o tempo de trânsito é de 1/200 segundo, e um milhão de pulsos estarão em trânsito. Se a temperatura do cabo aumentar em 0,6°C, a velocidade de propagação aumentará em cerca da 0,01%. Isso fará com que os pulsos em trânsito cheguem mais cedo, resultando em um aumento temporário na taxa de pulsos recebidos. Como os pulsos adicionais não podem ser acomodados no multiplexador, devem ser temporariamente armazenados no receptor. Se a temperatura do cabo cair, a taxa de pulsos recebidos também cairá, e o multiplexador terá janelas vazias, sem dados. Estas janelas devem ser preenchidas com dígitos arbitrários (**preenchimento de pulsos**).

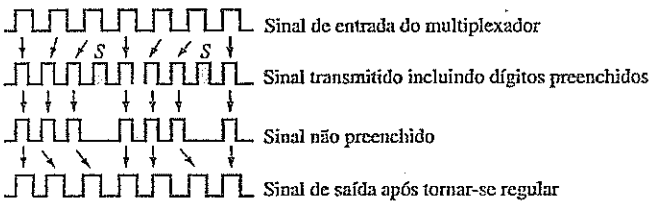


Figura 6.24 Preenchimento de pulsos.

Na rede norte-americana, sinais DS1 são, em geral, produzidos por osciladores a cristal em bancos de canais individuais ou outro terminal digital. Embora os osciladores sejam muito estáveis, não oscilarão exatamente na mesma frequência, dando origem a outra causa de assincronismo na rede.

Isso mostra que, mesmo em sistemas multiplexados síncronos, os dados raramente são recebidos a uma taxa síncrona. Sempre há necessidade de armazenagem (conhecida como **armazenagem elástica**) e preenchimento de pulsos (conhecido como **justificação**) para acomodar tal situação. Obviamente, esse método de armazenagem elástica e preenchimento de pulsos funcionará até mesmo quando os canais estiverem em assincronismo.

Há três variantes do esquema de preenchimento de pulsos: (1) preenchimento positivo de pulsos, (2) preenchimento negativo de pulsos e (3) preenchimento positivo/negativo de pulsos. No preenchimento positivo de pulsos, a taxa do multiplexador é mais alta que o necessário para acomodar todos os tributários de chegada em suas taxas máximas. Assim, as janelas temporais no sinal multiplexado ficarão disponíveis a uma taxa maior que a dos dados de entrada, de modo que os dados tributários tenderão a se atrasar (Fig. 6.24). Em algum momento, o sistema decidirá que esse atraso se tornou grande o bastante para exigir preenchimento de pulsos. A informação sobre as posições dos pulsos preenchidos é transmitida juntamente com os bits de *overhead*. A partir dos bits de *overhead*, o receptor toma conhecimento da posição do pulso preenchido e o elimina.

O preenchimento negativo de pulsos é um complemento do preenchimento positivo de pulsos. As janelas temporais no sinal multiplexado agora aparecem a uma taxa ligeiramente menor que as dos tributários, de modo que o sinal multiplexado não é capaz de acomodar todos os pulsos tributários. Informação a respeito de qualquer pulso que tenha ficado de fora e sua posição é transmitida pelos bits de *overhead*. O preenchimento positivo/negativo de pulsos é uma combinação dos dois esquemas anteriores: usamos preenchimento positivo de pulsos em alguns momentos e preenchimento negativo em outros. Toda essa informação é enviada através dos bits de *overhead*.

Os dígitos C na Fig. 6.23 são usados para transmitir informação de preenchimento. Apenas um bit de preenchimento por canal de entrada é permitido por quadro. Isso é o suficiente para acomodar variações esperadas na taxa de sinal de entrada. Os bits C_A transportam informação sobre preenchimento no canal A, bits C_B transportam informação sobre preenchimento no canal B, e assim por diante. A inserção de um pulso de preenchimento em um subquadro qualquer é denotada com a fixação dos valores dos três Cs na linha correspondente em **1**. Se não houver preenchimento, valores dos três Cs são fixados em **0**s. Se um bit de preenchimento for inserido, este será o primeiro bit de informação associado ao canal imediatamente após o bit F_1, ou seja, o primeiro desses bits na última sequência de 48 bits no subquadro em questão. Para o primeiro subquadro, o bit de preenchimento vem imediatamente após o bit F_1. Para o segundo subquadro, o bit de preenchimento será o segundo bit após o bit F_1, e assim por diante.

6.4.3 Hierarquia Digital Plesiócrona (Quase Síncrona)

A seguir, apresentaremos a hierarquia digital desenvolvida pelo grupo Bell System e atualmente incluída nos padrões ANSI para telecomunicações (Fig. 6.25). A hierarquia digital norte-americana é implementada na América do Norte e no Japão.

Duas categorias principais de multiplexadores são usadas na prática. A primeira é usada para combinar canais de baixa taxa de dados, e multiplexa canais de taxas de até 9.600 bit/s em um sinal com

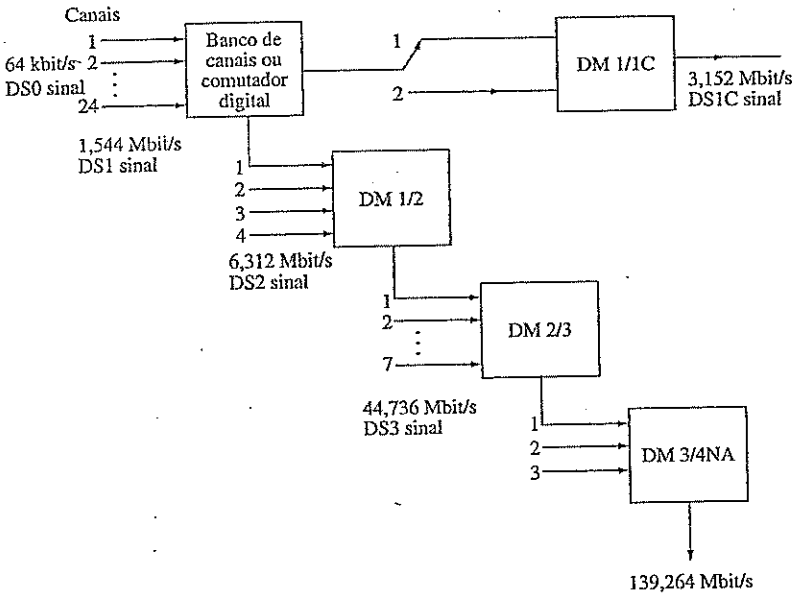


Figura 6.25 Hierarquia digital norte-americana (sistema da AT&T).

taxa de dados de até 64 kbit/s. O sinal multiplexado, denominado **"nível de sinal digital 0" (DS0)** na hierarquia norte-americana, é transmitido em um canal de classe de voz. A segunda categoria de multiplexadores é voltada para uma taxa de bits muito mais elevada.

Há quatro ordens, ou níveis, de multiplexação. O primeiro nível é o **multiplexador T1** ou **banco de canais**, consistindo em 24 canais de 64 kbit/s cada um. A saída deste multiplexador é um sinal **DS1 (nível digital 1)**, a uma taxa de 1,544 Mbit/s. Quatro sinais DS1 são multiplexados por um multiplexador DM1/2 para produzir um sinal DS2 à taxa de 6,321 Mbit/s. Sete sinais DS2 são multiplexados por um multiplexador DM2/3 para produzir um sinal DS3 à taxa de 44,736 Mbit/s. Por fim, três sinais DS3 são multiplexados por um multiplexador DM3/4NA para produzir um sinal DS4NA à taxa de 132,264 Mbit/s. Há, ainda, uma hierarquia de multiplexação a uma taxa inferior, conhecida como **sistema de dados digitais (DDS – *digital data system*)**, que provê padrões para a multiplexação de sinais digitais a taxas baixas, como 2,4 kbit/s, em um sinal DS0 para transmissão na rede.

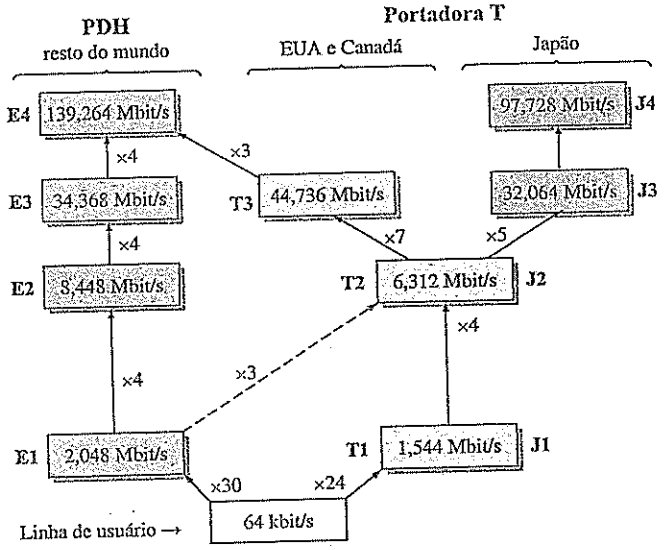


Figura 6.26 Hierarquia digital plesiócrona (PDH – *plesiochronous digital hierarchy*), segundo a Recomendação G.704 de ITU-T.

Não é necessário que as entradas de um multiplexador T1 fiquem restritas somente a canais de voz digitalizados. Qualquer sinal digital de 64 kbit/s e formato apropriado pode ser transmitido. O caso dos níveis mais altos é similar. Por exemplo, não é necessário que todos os canais de entrada do multiplexador DM1/2 tenham sido obtidos da multiplexação de 24 canais de 64 kbit/s cada um. Alguns podem ser sinais digitais de 1,544 Mbit/s e formato apropriado.

Na Europa e em muitas outras partes do mundo, foi adotada outra hierarquia, recomendada pela ITU como um padrão. Essa hierarquia, baseada na multiplexação de 30 canais telefônicos de 64 kbit/s (canais E-0) em uma portadora E-1 de 2,048 Mbit/s (30 canais), é ilustrada na Fig. 6.26. A partir do nível básico E-1, quatro linhas de nível mais baixo formam, progressivamente, uma linha de nível elevado, gerando uma linha E-1 com vazão de dados de 8,448 Mbit/s, uma linha E-3 com 34,368 Mbit/s de vazão de dados, uma linha E-4 com vazão de dados de 139,264 Mbit/s, e uma linha E-5 com vazão de dados de 565,148 Mbit/s. Como deve haver interconexão de redes distintas nos três sistemas em uso no mundo (norte-americano, japonês e outro), a Fig. 6.26 demonstra a relação entre os sistemas e pontos das interfaces comuns.

6.5 MODULAÇÃO POR CODIFICAÇÃO DE PULSO DIFERENCIAL (DPCM)

A PCM não é um sistema muito eficiente, pois gera um número excessivo de bits e requer grande largura de banda de transmissão. Vários conceitos diferentes foram propostos para melhorar a eficiência de codificação da conversão A/D. Em geral, esses conceitos exploram características da fonte do sinal. A DPCM (*differential pulse code modulation*) é um deles.

Em mensagens analógicas, podemos ter uma boa ideia do valor de uma amostra a partir dos valores de amostras anteriores. Em outras palavras, os valores das amostras não são independentes e, em geral, há um alto grau de redundância nas amostras de Nyquist. A exploração adequada dessa redundância permite a codificação de um sinal com um número menor de bits. Consideremos um esquema simples: em vez de transmitir os valores das amostras, transmitimos a diferença entre valores de amostras sucessivas. Assim, se $m[k]$ for a k-ésima amostra, em vez de transmitirmos $m[k]$, transmitimos a diferença $d[k] = m[k] - m[k-1]$. No receptor, o conhecimento de $d[k]$ e de diversos valores de amostras anteriores $m[k-1]$ permite a reconstrução de $m[k]$. Ou seja, a partir do conhecimento de $d[k]$, podemos reconstruir $m[k]$ iterativamente no receptor. A diferença entre valores de amostras sucessivas é, em geral, muito menor que os valores das amostras. Assim, a diferença de pico m_p dos valores transmitidos é consideravelmente reduzido. Como o intervalo de quantização é $\Delta v = m_p/L$, para um dado L (ou n), isso também reduz o tamanho Δv do intervalo de quantização, o que reduz o ruído de quantização, dado por $\Delta v^2/12$. Por conseguinte, para um dado n (ou largura de banda de transmissão), podemos aumentar a SNR ou, para uma dada SNR, podemos reduzir n (ou a largura de banda de transmissão).

Podemos melhorar este esquema se fizermos uma estimativa (ou previsão) do valor $m[k]$ da k-ésima amostra a partir do conhecimento de valores de várias amostras anteriores. Se a estimativa for $\hat{m}[k]$, podemos transmitir a diferença (erro de predição) $d[k] = m[k] - \hat{m}[k]$. No receptor, determinamos o valor da estimativa $\hat{m}[k]$ a partir de valores de amostras anteriores e, então, geramos $m[k]$ somando $d[k]$ à estimativa $\hat{m}[k]$. Assim, reconstruímos as amostras no receptor iterativamente. Se a predição for boa, o valor predito (estimado) $\hat{m}[k]$ será próximo de $m[k]$ e a diferença entre eles (o erro de predição) $d[k]$ será ainda menor que a diferença entre valores de amostras sucessivas. Em consequência, esse esquema, conhecido como **PCM diferencial (DPCM)**, é superior à simples predição descrita no parágrafo anterior, que é um caso especial de DPCM. em que a estimativa do valor de uma amostra é tomado como o valor da amostra anterior, ou seja, $\hat{m}[k] = m[k-1]$.

Espíritos de Taylor, Maclaurin e Wiener

Antes de descrever os princípios de DPCM, discutiremos brevemente a abordagem à predição (estimação) de sinal. Aos iniciantes, a previsão do futuro parece algo misterioso, associado a videntes,

magos, médiuns e assemelhados, que podem clamar ajuda do mundo espiritual. Engenheiros eletricistas não têm a menor chance nessa atividade. Não é bem assim! Também podemos clamar ajuda dos espíritos de Taylor, Maclaurin, Wiener e de outros. Mais que isso, ao contrário dos espíritos de Shakespeare, quando chamados, nossos espíritos vêm.* Consideremos, por exemplo, um sinal $m(t)$, cujas derivadas de todas as ordens existem em t. Usando a série de Taylor para este sinal, podemos expressar $m(t + T_s)$ como

$$m(t + T_s) = m(t) + T_s\dot{m}(t) + \frac{T_s^2}{2!}\ddot{m}(t) + \frac{T_s^3}{3!}\dddot{m}(t) + \cdots \tag{6.42a}$$

$$\approx m(t) + T_s\dot{m}(t) \text{ para pequeno } T_s \tag{6.42b}$$

A Eq. (6.42a) mostra que, a partir do conhecimento do sinal e de suas derivadas no instante t, podemos prever um valor futuro do sinal em $t + T_s$. Na verdade, mesmo que conheçamos somente a primeira derivada, ainda podemos prever este valor aproximadamente, como na Eq. (6.42b). Denotemos a k-ésima amostra de $m(t)$ por $m[k]$, ou seja, $m(kT_s) = m[k]$, e $m(kT_s \pm T_s) = m[k \pm 1]$, e assim por diante. Tomando $t = kT_s$ na Eq. (6.42b) e usando a aproximação $\dot{m}[kT_s] \approx [m(kT_s) - m(kT_s - T_s)]/T_s$, obtemos

$$m[k+1] \approx m[k] + T_s\left[\frac{m[k] - m[k-1]}{T_s}\right]$$

$$= 2m[k] - m[k-1]$$

Isso mostra que podemos obter uma previsão grosseira do valor da amostra de ordem $(k + 1)$ a partir dos valores das duas amostras anteriores. A aproximação na Eq. (6.42b) melhora se adicionarmos mais termos à série no lado direito. Para determinar derivadas de ordens superiores na série, precisamos de mais amostras no passado. Quanto maior o número de amostras passadas, melhor será a predição. Em geral, podemos expressar a fórmula de predição como

$$m[k] \approx a_1 m[k-1] + a_2 m[k-2] + \cdots + a_N m[k-N] \tag{6.43}$$

O lado direito é $\hat{m}[k]$, o valor predito para $m[k]$. Logo,

$$\hat{m}[k] = a_1 m[k-1] + a_2 m[k-2] + \cdots + a_N m[k-N] \tag{6.44}$$

Essa é a equação para o preditor de ordem N. Em geral, um maior valor de N resulta em melhor predição. A saída desse filtro (preditor) é $\hat{m}[k]$, o valor predito para $m[k]$. A entrada consiste nas amostras anteriores $m[k-1]$, $m[k-2]$, ..., $m[k-N]$, embora seja comum dizer que a entrada é $m[k]$ e a saída, $\hat{m}[k]$. Vale observar que, no caso da predição de primeira ordem, essa equação se reduz a $\hat{m}[k] = m[k-1]$, resultado que advém da Eq. (6.42b), onde retemos somente o primeiro termo no lado direito. Isso significa que $a_1 = 1$ e que o preditor de primeira ordem é apenas um atraso temporal.

Delineamos, assim, um procedimento muito simples para o projeto do preditor. Em uma abordagem mais sofisticada, discutida na Seção 8.5, em que usamos o critério de mínimo erro quadrático médio para a melhor predição, os **coeficientes de predição** a_j na Eq. (6.44) são determinados da correlação estatística entre as várias amostras. O preditor descrito na Eq. (6.44) é denominado *preditor linear* e consiste basicamente em um filtro transversal (linha de atraso com derivação – *tapped delay line*), em que os ganhos das derivações (*taps*) são tomados como iguais aos coeficientes de predição, como ilustrado na Fig. 6.27.

---

*De Shakespeare, Henrique IV, Parte 1, Ato III, Cena I:

Glendower: *Posso evocar espíritos do abismo.*

Hotspur: *Isso, até eu, e assim qualquer pessoa;*

Mas eles vêm, no caso de os chamardes?

Da tradução disponível em http://www.ebooksbrasil.org/eLibris/henry4.html#12. (N.T.)

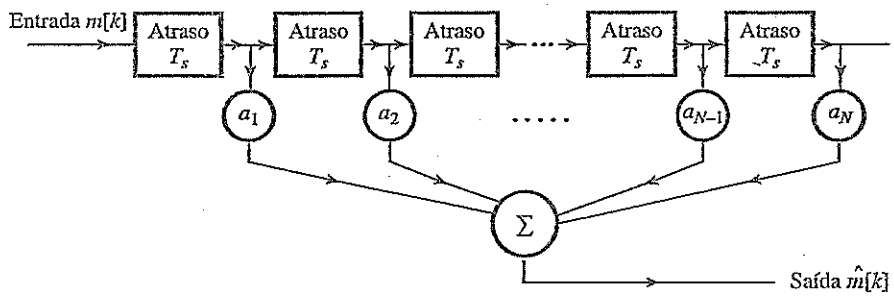


Figura 6.27 Filtro transversal (linha de atraso com derivação) usado como um preditor linear.

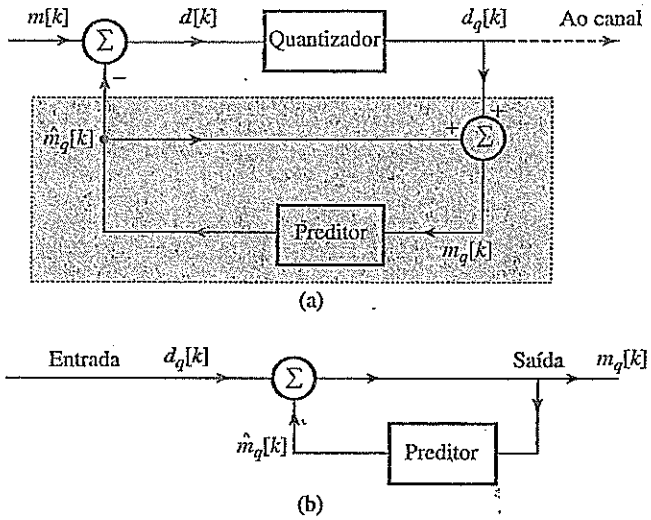


Figura 6.28 Sistema DPCM: (a) transmissor; (b) receptor.

Análise da DPCM

Como já mencionado, na DPCM, não transmitimos a amostra atual $m[k]$, mas $d[k]$ (a diferença entre $m[k]$ e seu valor predito $\hat{m}[k]$). No receptor, geramos $\hat{m}[k]$ a partir dos valores de amostras passadas, aos quais, para gerar $m[k]$, é somado o valor recebido $d[k]$. No entanto, há uma dificuldade associada a este esquema. No receptor, em vez das amostras passadas $m[k-1]$, $m[k-2]$, ..., e de $d[k]$, temos suas versões quantizadas $m_q[k-1]$, $m_q[k-2]$, ... Portanto, não podemos determinar $\hat{m}[k]$, mas apenas $\hat{m}_q[k]$, o valor estimado da amostra quantizada $m_q[k]$, em termos das amostras quantizadas $m_q[k-1]$, $m_q[k-2]$, ... Isso aumentará o erro na reconstrução. Nesses casos, uma estratégia melhor consiste em, no transmissor, determinar $\hat{m}_q[k]$, o valor estimado de $m_q[k]$ (em vez de $m[k]$) a partir das versões quantizadas $m_q[k-1]$, $m_q[k-2]$, ... A diferença $d[k] = m[k] - \hat{m}_q[k]$ é, então, transmitida via PCM. No receptor, podemos gerar $\hat{m}_q[k]$ e, a partir do valor $d[k]$ recebido, reconstruir $m_q[k]$.

A Fig. 6.28a ilustra um transmissor DPCM. Logo, mostraremos que a entrada do preditor é $m_q[k]$. Naturalmente, sua saída é $\hat{m}_q[k]$, o valor predito para $m[k]$. A diferença

$$d[k] = m[k] - \hat{m}_q[k] \tag{6.45}$$

é quantizada, produzindo

$$d_q[k] = d[k] + q[k] \tag{6.46}$$

em que $q[k]$ é o erro de quantização. A saída do preditor $\hat{m}_q[k]$ é realimentada à entrada, de modo que a entrada do preditor $m_q[k]$ se torna

$$\begin{aligned} m_q[k] &= \hat{m}_q[k] + d_q[k] \\ &= m[k] - d[k] + d_q[k] \\ &= m[k] + q[k] \end{aligned} \tag{6.47}$$

Isso mostra que $m_q[k]$ é uma versão quantizada de $m[k]$. A entrada do preditor é, de fato, $m_q[k]$, como presumido. O sinal quantizado $d_q[k]$ é, agora, transmitido através do canal. O receptor mostrado na Fig. 6.28b é idêntico à porção hachurada do transmissor. Nos dois casos, as entradas também são as mesmas, ou seja, $d_q[k]$. Portanto, a saída do preditor deve ser $\hat{m}_q[k]$ (igual à saída do preditor no transmissor). Assim, a saída do receptor (que é a entrada do preditor) também é a mesma, $m_q[k] = m[k] + q[k]$, como calculado na Eq. (6.47). Isso mostra que foi possível receber o sinal desejado $m[k]$ mais o erro de quantização $q[k]$. Este é o ruído de quantização associado ao sinal-diferença $d[k]$ que, em geral, é muito menor que $m[k]$. As amostras recebidas $m_q[k]$ são decodificadas e aplicadas a um filtro passa-baixas para conversão D/A.

Melhora da SNR

Para determinar a melhora da DPCM em relação à PCM, denotemos as amplitudes de pico de $m(t)$ e $d(t)$ por m_p e d_p, respectivamente. Se usarmos o mesmo valor de L nos dois casos, o incremento de quantização Δv em DPCM é reduzido por um fator d_p/m_p. Como a potência do ruído de quantização é $(\Delta v)^2/12$, na DPCM, o ruído de quantização é reduzido por um fator $(m_p/d_p)^2$, e a SNR é aumentada pelo mesmo fator. Além disso, a potência de sinal é proporcional ao quadrado do valor de pico (supondo que outras propriedades estatísticas sejam invariantes). Portanto, G_p (melhora da SNR devido à predição) é, pelo menos,

$$G_p = \frac{P_m}{P_d}$$

em que P_m e P_d são as potências de $m(t)$ e $d(t)$, respectivamente. Em decibéis, isso significa que a SNR aumenta em $10\log_{10}(P_m/P_d)$ dB. Portanto, a Eq. (6.41) se aplica à DPCM também com um valor de α aumentado por $10\log_{10}(P_m/P_d)$ dB. No Exemplo 8.24, é analisado um processador preditor de segunda ordem para sinais de voz. Para este caso, a melhora na SNR é calculada como 5,6 dB. Na prática, em casos como espectros de sinais de voz de curta duração e espectros de imagens de baixa atividade, a melhora da SNR pode chegar a 25 dB.[12] Alternativamente, para a mesma SNR, a taxa de bits para DPCM pode ser 3 a 4 bits por amostra ser menor que para PCM. Assim, o sistema telefônico que usa DPCM pode, em geral, operar a 32 ou, até mesmo, a 24 kbits/s.

6.6 PCM DIFERENCIAL ADAPTATIVA (ADPCM)

A DPCM adaptativa (ADPCM) pode melhorar ainda mais a eficiência da codificação DPCM com a incorporação de um quantizador adaptativo no codificador. A Fig. 6.29 ilustra a configuração básica de ADPCM. Por motivos práticos, o número de níveis de quantização L é fixo. Quando é aplicado um incremento de quantização fixo Δv, o erro de quantização ou é muito grande, quando Δv é grande, ou o quantizador não pode cobrir o necessário intervalo de variação do sinal, quando Δv é muito pequeno. Portanto, seria melhor que o incremento de quantização Δv fosse adaptável, de modo que Δv fosse grande ou pequeno dependendo se o erro de predição para quantização fosse grande ou pequeno.

É importante notar que o erro de predição quantizado $d_q[k]$ pode ser um bom indicador do tamanho do erro de predição. Por exemplo, quando as amostras do erro de predição quantizado variam muito próximas do maior valor positivo (ou maior valor negativo), isso indica que o erro de predição é grande e que Δv deve aumentar. Reciprocamente, se as amostras quantizadas oscilam próximas de zero, o erro de predição é pequeno e Δv deve diminuir. É importante que tanto o modulador como o receptor tenham acesso às mesmas amostras quantizadas. Portanto, o quantizador adaptativo e a reconstrução do receptor podem aplicar o mesmo algoritmo para ajustar o valor de Δv identicamente.

Em comparação com a DPCM, a ADPCM pode comprimir ainda mais o número de bits necessários para uma forma de onda. Por exemplo, na prática, é muito comum que uma sequência PCM de 8 bits seja codificada em uma sequência ADPCM de 4 bits à mesma taxa de amostragem. Isso representa uma redução de 2:1 na largura de banda ou na necessidade de armazenagem, sem praticamente qualquer perda.

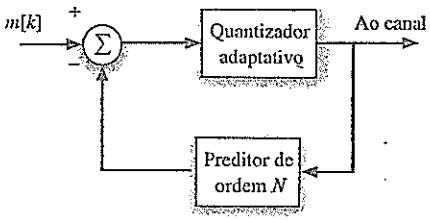


Figura 6.29 Codificador ADPCM usa um quantizador adaptativo controlado apenas pelos bits de saída do codificador.

O codificador ADPCM tem várias aplicações. O padrão ITU-T G.726 especifica um codificador e decodificador (denominado **codec**) de voz para amostras de voz tomadas a 8 kHz.[7] O preditor ADPCM G.726 usa um preditor de oitava ordem. Para diferentes níveis de quantização, o padrão G.726 especifica quatro taxas ADPCM diferentes: 16, 24, 32 e 40 kbit/s. Essas taxas correspondem a quatro diferentes representações de cada amostra de voz, a 2 bits, 3 bits, 4 bits e 5 bits, respectivamente, o que é equivalente a 4, 8, 16 e 32 níveis de quantização, respectivamente.

Os codificadores de voz ADPCM mais comuns usam 32 kbit/s. Na prática, há uma variedade de codecs de voz ADPCM. Além da especificação ITU-T G.726,[7] há o codec ADPCM OKI, o codec ADPCM Microsoft suportado por *players* do formato WAVE, e ADPCM da Associação de Multimídia Interativa (IMA – *Interactive Multimedia Association*), também conhecido como DVI ADPCM. O codec 32 kbit/s ADPCM de voz ITU-T G.726 é largamente utilizado em sistemas DECT (*digital enhanced cordless telecommunications* – telecomunicações sem fio digitais aprimoradas), que, por sua vez, encontra grande utilização em telefonia sem fio residencial e comercial. Projetado para uso em curta distância como um mecanismo de acesso à rede principal, DECT oferece comunicação sem fio para voz, fax, dados e multimídia. Atualmente, DECT é usado em mais de 100 países ao redor do mundo. Outro uso importante do codec ADPCM de 32 kbit/s é o Personal Handy-phone System (PHS) [sistema de telefonia portátil pessoal], também comercializado como Personal Access System (PAS) [sistema de acesso pessoal] e conhecido na China como *Xiaolingtong*.

O PHS é um sistema de telefonia móvel semelhante ao sistema celular, que opera na faixa de frequências de 1880 a 1930 MHz e é usado principalmente no Japão, China, Formosa e em outros locais na Ásia. Originalmente desenvolvido pela Laboratório NTT, no Japão, em 1989, o PHS tem implementação e utilização muito simples. Ao contrário das redes celulares, telefones e estações de base de PHS são de baixa potência e de curto alcance. O serviço é pejorativamente chamado de "celular do pobre" devido ao alcance limitado e pouca capacidade de *roaming*. As primeiras aplicações de PHS ocorreram no Japão, em 1995, (NTT-Personal, DDI-Pocket e ASTEL) e, desde então, praticamente desapareceu. Contudo, de modo surpreendente, o PHS está ressurgindo em mercados como China, Formosa, Vietnã, Bangladesh, Nigéria, Mali, Tanzânia e Honduras, onde o baixo custo de equipamentos e de utilização compensa as desvantagens do sistema. Na China, houve uma expansão explosiva e, em 2006, o número de assinantes era próximo de 80 milhões.

6.7 MODULAÇÃO DELTA

A correlação de amostras usada em DPCM é ainda mais explorada na **modulação delta (DM** – *delta modulation*) por meio de superamostragem (tipicamente, a uma taxa quatro vezes maior que a de Nyquist) do sinal em banda base. Isso aumenta a correlação entre amostras adjacentes, o que resulta em pequeno erro de predição, que pode ser codificado por apenas um bit ($L = 2$). Assim, DM é basicamente DPCM de 1 bit, ou seja, DPCM que utiliza somente dois níveis ($L = 2$) para a quantização de $m[k] - \hat{m}_q[k]$. Em comparação com a PCM (e DPCM), esse é um método de conversão A/D muito simples e barato. A palavra de código de 1 bit em DM torna desnecessário o enquadramento (*framing*) de palavra no transmissor e no receptor. Essa estratégia permite o uso de menor número de bits por amostra para a codificação de um sinal em banda base.

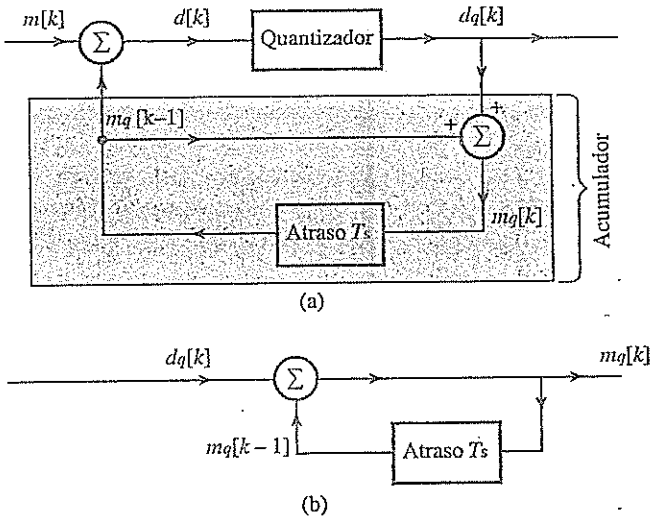


Figura 6.30 Modulação delta é um caso especial de DPCM.

Em DM, usamos um preditor de primeira ordem, que, como visto anteriormente, é apenas um atraso temporal T_s (intervalo de amostragem). Assim, o transmissor (modulador) e receptor (demodulador) DM são idênticos aos de DPCM na Fig. 6.28, com um atraso temporal para o preditor, como mostrado na Fig. 6.30, da qual podemos escrever:

$$m_q[k] = m_q[k-1] + d_q[k] \tag{6.48}$$

Logo,

$$m_q[k-1] = m_q[k-2] + d_q[k-1]$$

A substituição desta equação na Eq. (6.48) resulta em

$$m_q[k] = m_q[k-2] + d_q[k] + d_q[k-1]$$

Prosseguindo com esse processo iterativamente e admitindo condição inicial zero, ou seja, $m_q[0] = 0$, escrevemos

$$m_q[k] = \sum_{m=0}^{k} d_q[m] \tag{6.49}$$

Isso mostra que o receptor (demodulador) é apenas um acumulador (somador). Se a saída $d_q[k]$ for representada por impulsos, o acumulador (receptor) pode ser realizado por um integrador, pois sua saída é a soma das intensidades dos impulsos de entrada (soma das áreas sob os impulsos). Podemos, também, substituir a porção de realimentação do modulador (que é idêntico ao demodulador) pelo integrador. A saída do demodulador é $m_q[k]$, que, ao ser aplicada a um filtro passa-baixas, produz o sinal desejado, reconstruindo-o das amostras quantizadas.

A Fig. 6.31 mostra uma implementação prática do modulador e do demodulador delta. Como discutido anteriormente, o preditor de primeira ordem é substituído por um circuito integrador de baixo custo (como um integrador RC). O modulador (Fig. 6.31a) consiste em um comparador e um amostrador na rota direta, e em um amplificador-integrador na rota de realimentação. Examinemos o funcionamento deste modulador delta.

O sinal analógico $m(t)$ é comparado com o sinal de realimentação (que serve como um sinal preditor) $\hat{m}_q[t]$. O sinal de erro $d(t) = m(t) - \hat{m}_q[t]$ é aplicado a um comparador. Se $d(t)$ for positivo, a

saída do comparador é um sinal constante de amplitude E; se $d(t)$ for negativo, a saída do comparador é $-E$. Assim, a diferença é um sinal binário ($L = 2$), necessário para gerar DPCM de 1 bit. A saída do comparador é amostrada por um amostrador à taxa de f_s amostras por segundo, sendo f_s, tipicamente, muito maior que a taxa de Nyquist. O amostrador, então, produz um trem de pulsos estreitos $d_q[k]$ (para simular impulsos); quando $m(t) > \hat{m}_q[t]$, o pulso é positivo; quando $m(t) < \hat{m}_q[t]$, o pulso é negativo. Reparemos que cada amostra é codificada por apenas um pulso binário (DPCM de 1 bit), como necessário. O trem de pulsos $d_q[k]$ é o trem de pulsos com modulação delta (Fig. 6.31d). O sinal modulado $d_q[k]$ é amplificado e integrado na rota de realimentação para gerar $\hat{m}_q[t]$ (Fig. 6.31c), que tenta seguir $m(t)$.

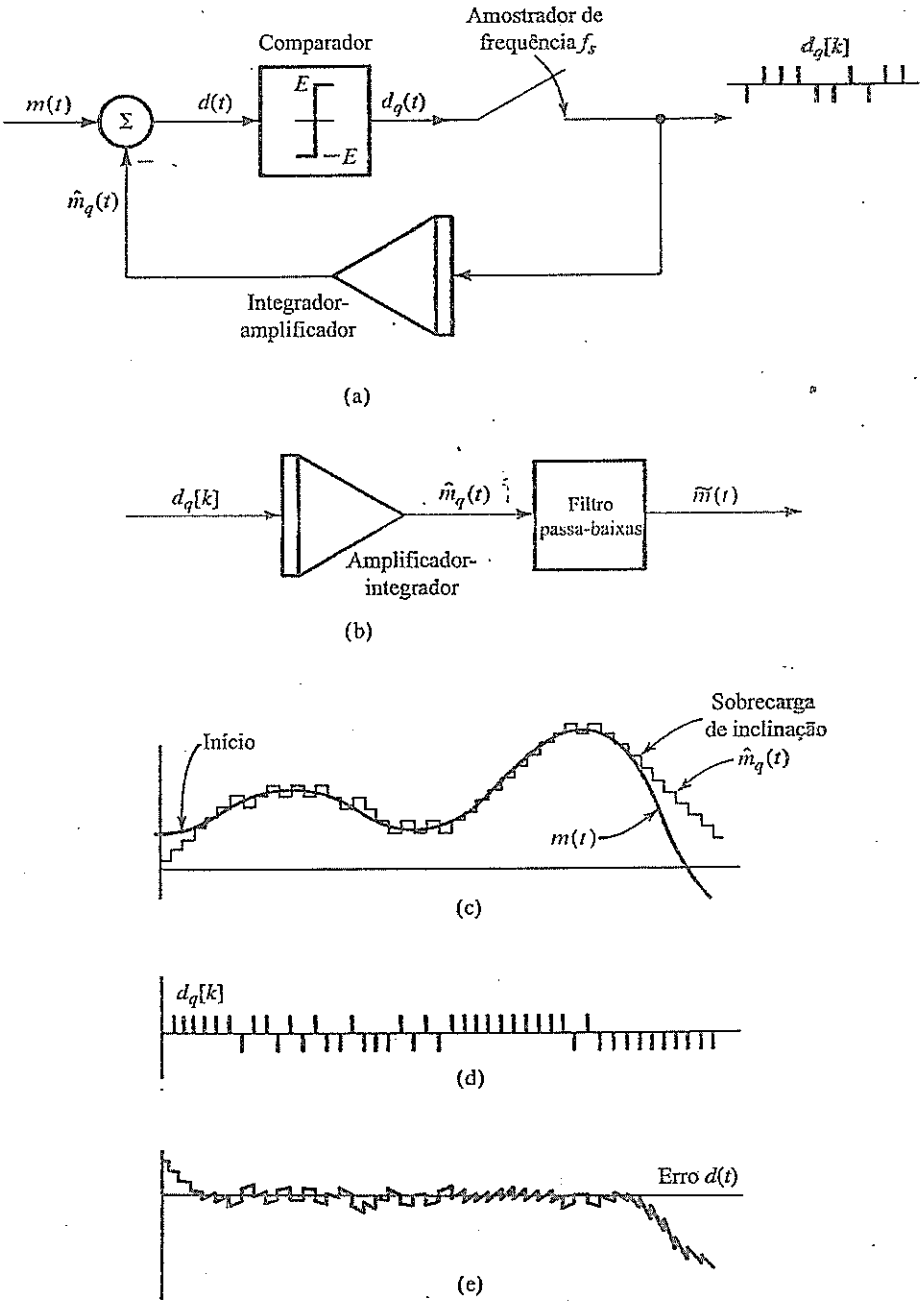


Figura 6.31 (a) Modulador delta; (b) demodulador delta; (c) sinal de mensagem *versus* sinal de saída do integrador; (d) trem de pulsos com modulação delta; (e) erros de modulação.

Para entender como isso funciona, notemos que cada pulso em $d_q[k]$ na entrada do integrador origina uma função degrau (positiva ou negativa, dependendo da polaridade do pulso) em $\hat{m}_q[t]$. Se, por exemplo, $m(t) > \hat{m}_q[t]$, é gerado um pulso positivo em $d_q[k]$, que dá origem a um degrau positivo em $\hat{m}_q[t]$, tentando, em pequenos passos a cada instante de amostragem, fazer $\hat{m}_q[t]$ igualar $m(t)$, como mostrado na Fig. 6.31c. Podemos observar que $\hat{m}_q[t]$ é uma espécie de aproximação em degraus de $m(t)$. Quando $\hat{m}_q[t]$ é aplicado a um filtro passa-baixas, a aspereza dos degraus é eliminada, e obtemos uma melhor aproximação para $m(t)$. No receptor, o demodulador consiste em um integrador-amplificador (idêntico ao da rota de realimentação do modulador) seguido por um filtro passa-baixas (Fig. 6.31b).

DM Transmite a Derivada de $m(t)$

Em PCM, as amostras do sinal analógico são quantizadas em L níveis , e esta informação é transmitida por n pulsos por amostras ($n = \log_2 L$). Uma pequena reflexão mostra que, em DM, o sinal modulado transporta informação sobre a diferença entre amostras sucessivas, e não sobre amostras de sinais. Se a diferença for positiva ou negativa, um pulso positivo ou negativo (respectivamente) é gerado no sinal modulado $d_q[k]$. Portanto, DM transporta informação sobre a derivada de $m(t)$, daí o nome "modulação delta". Isso também pode ser visto do fato de que a integração do sinal modulado em delta produz $\hat{m}_q[t]$, que é uma aproximação de $m(t)$.

Em PCM, a informação de cada amostra quantizada é transmitida por uma palavra de código de n bits, enquanto em DM a informação sobre a diferença entre amostras sucessivas é transmitida por uma palavra de código de 1 bit.

Limiar de Codificação e Sobrecarga

Efeitos de limiar e sobrecarga podem ser vistos claramente na Fig. 6.31c. Variações em $m(t)$ menores que o valor do degrau (limiar de codificação) são perdidos em DM. Além disso, se $m(t)$ variar demasiadamente rápido, ou seja, se $\dot{m}(t)$ for demasiadamente grande, $\hat{m}_q(t)$ não será capaz de seguir $m(t)$ e ocorrerá sobrecarga. Isso caracteriza a chamada sobrecarga de inclinação (*slope overload*), que dá origem ao ruído de sobrecarga de inclinação. Esse ruído é um dos fatores limitantes básicos ao desempenho de DM. Em DM, deveríamos esperar sobrecarga de inclinação em vez de sobrecarga de amplitude, pois DM transporta basicamente informação sobre $\dot{m}(t)$. A natureza granular do sinal de saída dá origem ao ruído granular, similar ao ruído de quantização. O ruído de sobrecarga de inclinação pode ser reduzido com o aumento de E (o incremento). Lamentavelmente, isso aumenta o ruído granular. Existe um valor ótimo de E, que produz o melhor equilíbrio, resultando no ruído mínimo. Esse valor ótimo de E depende da frequência de amostragem f_s e da natureza do sinal.[12]

A sobrecarga de inclinação ocorre quando $\hat{m}_q(t)$ não consegue seguir $m(t)$. Durante o intervalo de amostragem T_s, $\hat{m}_q(t)$ pode ser alterado em E, em que E é a altura do degrau. Assim, a máxima inclinação que $\hat{m}_q(t)$ é capaz de seguir é E/T_s ou Ef_s, em que f_s é a frequência de amostragem. Portanto, sobrecarga de inclinação não ocorre quando

$$|\dot{m}(t)| < Ef_s$$

Consideremos o caso da modulação por tom (mensagem senoidal):

$$m(t) = A\cos\omega t$$

A condição para não haver sobrecarga é

$$|\dot{m}(t)|_{\max} = \omega A < Ef_s \tag{6.50}$$

Logo, a máxima amplitude $A_{\max}$ deste sinal que pode ser tolerada sem sobrecarga é dada por

$$A_{\max} = \frac{Ef_s}{\omega} \tag{6.51}$$

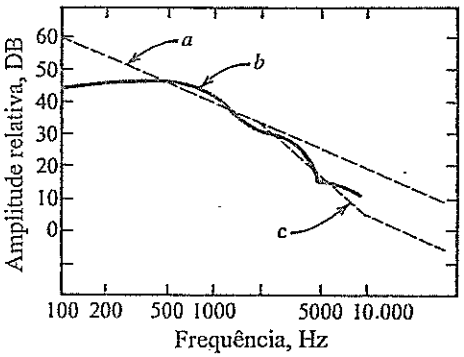


Figura 6.32 Espectro de sinal de voz.

A amplitude de sobrecarga do sinal modulante é inversamente proporcional à frequência ω. Para frequências modulantes mais altas, a sobrecarga ocorre para amplitudes menores. No caso de sinais de voz, que contêm todas as componentes de frequência de até (digamos) 4 kHz, o cálculo de A_{max} pela Eq. (6.51), com $\omega = 2\pi \times 4000$, fornece um valor demasiadamente conservador. Jager[13] mostrou que, para sinais de voz, A_{max} pode ser calculado usando $\omega_r \simeq 2\pi \times 800$ na Eq. (6.51),

$$[A_{max}]_{voz} \simeq \frac{Ef_s}{\omega_r} \tag{6.52}$$

Portanto, a máxima amplitude de sinal de voz $A_{máx}$ que pode ser usada sem causar sobrecarga de inclinação em DM é igual à máxima amplitude de um sinal senoidal com frequência de referência f_r ($f_r \simeq 800$ Hz) que pode ser usada sem causar sobrecarga de inclinação no mesmo sistema.

Felizmente, a amplitude do espectro de voz (assim como do sinal de vídeo de televisão) também cai com a frequência e segue a característica de sobrecarga de perto (curva *c*, Fig. 6.32). Por essa razão, DM é muito apropriada para sinais de voz (e de televisão). Na verdade, o espectro do sinal de voz (curva *b*) cai com $1/\omega$ até cerca de 2000 Hz e, além dessa frequência, passa a cair com $1/\omega^2$. Se tivéssemos usado uma integração dupla no circuito de realimentação, em vez de uma integração simples, A_{max}, na Eq. (6.51), seria proporcional a $1/\omega^2$. Assim, um melhor casamento entre o espectro de voz e as características de sobrecarga é obtido com o uso de integração simples até 2000 Hz e de integração dupla acima de 2000 Hz. Um circuito desse tipo (com integração dupla) tem resposta rápida, mas com uma tendência à instabilidade, que pode ser reduzida com o uso de algum preditor de ordem baixa juntamente com a integração dupla. Um integrador duplo pode ser construído com uma cascata de dois integradores *RC* passa-baixas, com constantes de tempo $R_1C_1 = 1/200\pi$ e $R_2C_2 = 1/4000\pi$. Isso resulta em integração simples de 100 a 2000 Hz e integração dupla além de 2000 Hz.

Modulação Sigma-Delta

Na discussão sobre limiar de codificação e sobrecarga, ilustramos que a essência de DM convencional é a codificação e transmissão da derivada do sinal de mensagem analógico. Assim, o receptor DM requer um integrador, como mostrado na Fig. 6.31 e, de modo equivalente, na Fig. 6.33a. Como a transmissão de sinal é inevitavelmente sujeita ao ruído de canal, tal ruído será integrado e se acumulará na saída do receptor, um fenômeno altamente indesejável e a maior deficiência de DM.

Para superar essa deficiência de DM, podemos fazer uma pequena modificação. Primeiro, podemos ver o sistema DM global como consistindo em um transmissor e um receptor aproximadamente sem distorção e lineares. Assim, um dos componentes seriais, o integrador $1/s$ do receptor, pode ser deslocado para a frente do transmissor (codificador) sem afetar as respostas finais do modulador e do demodulador, como mostrado na Fig. 6.33b. Por fim, os dois integradores podem ser fundidos em um só, após o subtrator, como indicado na Fig. 6.33c. Esse sistema modificado é conhecido como modulação sigma-delta (Σ-ΔM).

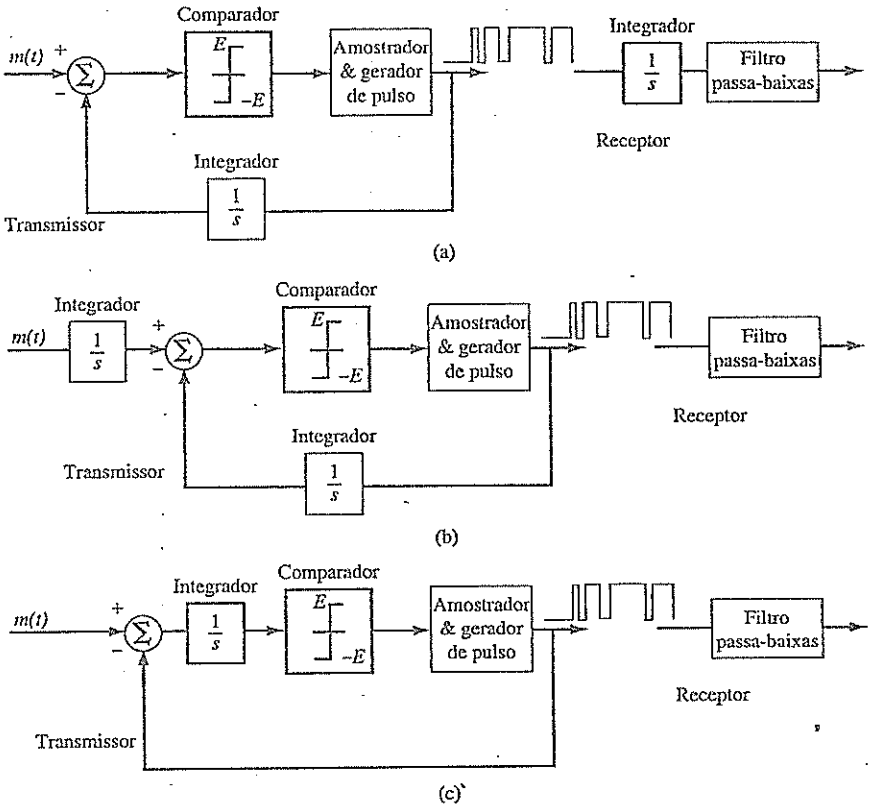


Figura 6.33
(a) Modulador delta convencional.
(b) Modulador Σ-Δ.
(c) Modulador Σ-Δ mais simples.

Como vimos no estudo de filtros de pré-ênfase e deênfase em FM, dado que o ruído de canal e o sinal de mensagem não seguem a mesma rota, a ordem dos componentes seriais no sistema global de modulação-demodulação pode ter diferentes efeitos na SNR. O aparente pequeno deslocamento do integrador $1/s$, na verdade, tem várias vantagens significativas:

- O ruído de canal não mais se acumula no demodulador.
- O importante conteúdo de baixa frequência da mensagem $m(t)$ sofre pré-ênfase no integrador $1/j\omega$. Isso ajuda muitos sinais práticos (como os de voz), cujas componentes de frequências baixas são mais importantes.
- O integrador, de fato, suaviza o sinal para codificação (Fig. 6.33b). Assim, a sobrecarga se torna menos provável.
- A natureza passa-baixas do integrador aumenta a correlação entre amostras sucessivas, resultando em menor erro de codificação.
- O demodulador é simplificado.

Modulação Delta Adaptativa (ADM)

A DM discutida até aqui tem uma séria desvantagem. A faixa dinâmica de amplitudes é demasiadamente pequena, devido aos efeitos de limiar e sobrecarga analisados anteriormente. Para superar esse problema, algum tipo de compressão de sinal se faz necessário. Em DM, um método adequado parece ser a adaptação (ajuste) do valor do degrau E segundo o nível da derivada do sinal de entrada. Por exemplo, na Fig. 6.31, quando o sinal $m(t)$ cai rapidamente, ocorre sobrecarga de inclinação. Se pudermos aumentar a altura do degrau durante este período, a sobrecarga seria evitada. Contudo, se

a inclinação do sinal $m(t)$ for pequena, uma redução na altura do degrau serviria para reduzir o nível de limiar e o ruído granular. A sobrecarga de inclinação faz com que $d_q[k]$ tenha vários pulsos sucessivos de mesma polaridade. Isso pede uma redução na altura do degrau. De modo similar, pulsos em $d_q[k]$ com alternância contínua de polaridade indicam variações de pequena amplitude, o que requer uma redução na altura do degrau. Em ADM (*adaptative delta modulation*), detectamos esses padrões de pulsos e ajustamos a altura do degrau automaticamente.[14] Isso resulta em uma faixa dinâmica muito maior para DM.

6.8 VOCODERS E COMPRESSÃO DE VÍDEO

PCM, DPCM, ADPCM, DM e Σ-ΔM são exemplos do que é conhecido como codificadores de fonte de forma de onda. Basicamente, codificadores de forma de onda não levam em consideração como foram gerados os sinais para digitalização. Assim, o grau de compressão alcançável por codificadores de forma de onda é altamente limitado pelo grau de correlação entre amostras sucessivas de sinal. Para uma fonte de sinal passa-baixas com largura de banda finita de B Hz, mesmo que aplicássemos a mínima taxa de amostragem de Nyquist, de $2B$ Hz, e codificação de 1 bit, a taxa de bits não pode ser inferior a $2B$ bit/s. Vários métodos foram introduzidos com sucesso para reduzir drasticamente a taxa de codificação de fonte de sinais de voz e de vídeo, muito importantes para nossas necessidades diárias de comunicação. Diferentemente de codificadores de forma de onda, os mais bem-sucedidos codificadores de voz e de vídeo são baseados em modelos fisiológicos humanos envolvidos na geração da voz e na percepção de vídeo. Aqui, descreveremos os princípios básicos da predição linear de codificadores de voz (conhecidos como vocoders) e o método de compressão de vídeo proposto pelo *Moving Picture Experts Group* – MPEG (Grupo de Especialistas em Imagens Animadas).

6.8.1 Vocoders com Codificação de Predição Linear

Modelos da Fala e Vocoders Baseados em Modelo

Vocoders com codificação de predição linear (LPC – *linear prediction coding*) são sistemas baseados em modelo. O modelo, por sua vez, é baseado no bom entendimento do mecanismo de produção da voz humana. A Fig. 6.34 apresenta uma ilustração seccional do aparato de produção da voz humana. Em resumo, a voz humana é produzida pela interação dos pulmões, cordas vocais e o trato de articulação, que consiste nas cavidades nasal e bucal. Com base nesse modelo fisiológico da fala, a voz humana pode ser dividida nas categorias de som *sonoro* (*voiced*) e *surdo* (*unvoiced*). Sons sonoros são feitos com cordas vocais vibrantes. Coloque um dedo no pomo-de-adão* enquanto fala e poderá sentir a vibração das cordas vocais quando pronunciar todas as vogais e algumas consoantes, como *g*, como em *gota*, *b*, como em *boa*, *n*, como em *não*. Sons surdos são feitos sem vibração das cordas vocais. Várias consoantes, como *k*, *p*, *t*, são surdas. Alguns exemplos de sons surdos são *r*, como em *roda*, *c*, como em *calo*, *p*, como em *pulo*.

Para a produção de sons sonoros, os pulmões expelem ar através da epiglote, fazendo vibrar as cordas vocais. As cordas vocais vibrantes interrompem o fluxo de ar e produzem uma onda de pressão quase periódica, que consiste em impulsos. Os impulsos da onda de pressão são denominados impulsos de timbre (*pitch*), e a frequência do sinal de pressão é chamada frequência de timbre ou frequência fundamental, como mostrado na Fig. 6.34b. Essa é a parte do sinal de voz que define o timbre da fala. Voz emitida em uma frequência de timbre constante soa monótona. Em geral, a frequência de timbre de um orador varia de modo quase contínuo, de sílaba para sílaba.

*Pequena protuberância na frente da garganta, formada pela maior cartilagem da laringe; em geral, mais saliente no homem que na mulher.

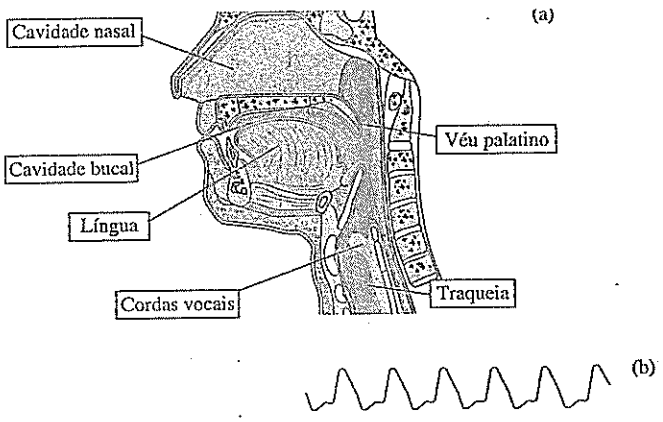


Figura 6.34
(a) Mecanismo de produção da voz humana.
(b) Típicos impulsos de pressão.

Em sons sonoros, os impulsos de timbre estimulam o ar no trato vocal (cavidades bucal e nasal). Em sons surdos, a excitação advém diretamente do fluxo de ar. Estudos detalhados[15–17] mostraram que, no caso de sons surdos, a excitação do trato vocal é mais parecido com um ruído de banda larga. Quando as cavidades no trato vocal ressoam sob excitação, radiam uma onda sonora, que é o sinal de voz. As duas cavidades formam ressoadores com frequências de ressonâncias características (frequências formantes). A alteração da forma (e, portanto, das características de ressonância) da cavidade bucal permite que diferentes sons sejam pronunciados. Surpreendentemente, esse trato de articulação (vocal) pode ser aproximado por um modelo que consiste em um filtro digital linear com função de transferência que tem somente polos

$$H(z) = \frac{g}{A(z)} = g \cdot \left(1 - \sum_{i=1}^{p} a_i z^{-i}\right)^{-1}$$

em que g é o fator de ganho, e $A(z)$ é conhecido como filtro preditor, como o filtro de realimentação usado em DPCM e ADPCM. Podemos interpretar a função do aparato de articulação vocal como um filtro de formatação espectral $H(z)$.

Modelos LPC

Com base neste modelo da fala humana, podemos determinar uma abordagem para codificação de voz diferente da codificação de forma de onda. Em vez de enviar as próprias amostras de sinal, vocoders baseados em modelo *analisam* os sinais de voz segmento a segmento, para determinar os parâmetros que melhor se ajustem ao sinal. Como mostrado na Fig. 6.35, após a análise do sinal de voz, para cada segmento do sinal, o transmissor envia ao receptor os necessários parâmetros (formantes) do modelo da voz. O receptor usa os parâmetros do modelo da voz para estabelecer um sintetizador de voz para regenerar os respectivos segmentos do sinal de voz. Em outras palavras, o que o usuário ouve no receptor consiste, na verdade, em sinais reproduzidos por uma *máquina sintetizadora* de voz artificial!

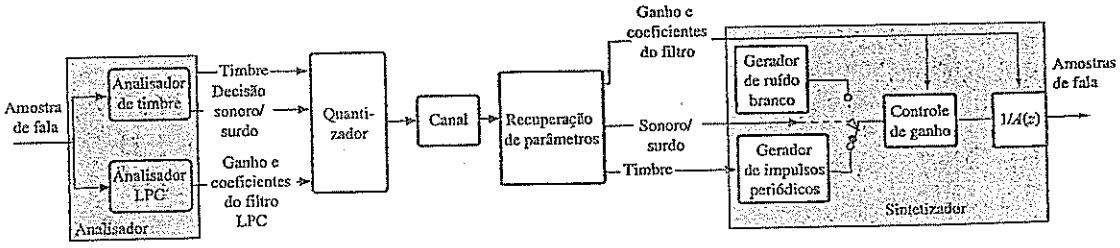


Figura 6.35
Análise e síntese de sinais de voz em codificador e decodificador LPC.

Tabela 6.1
Alocação de Bit de Quantização em Vocoder LPC-10

Período do Timbre	Sonoro/Surdo	Ganho g	Parâmetros do Filtro LP10, bits/coeficientes				
			$r_1 - r_4$	$r_5 - r_8$	r_9	r_{10}	
			5 bits	4 bits	3 bits	2 bits	Sonoro
6 bits	1 bit	5 bits	5 bits		*Não usado*		Surdo

Na análise de um segmento de voz amostrada (consistindo em múltiplas amostras), a análise de timbre determinará, primeiro, se a voz é sonora ou surda. Se o sinal for classificado como "sonoro", o analisador de timbre estimará a frequência do timbre (ou, o que é equivalente, o período do timbre). Além disso, o analisador de coeficientes de predição linear (LPC – *linear prediction coefficients*) estimará os coeficientes do filtro com polos apenas em $A(z)$. Como o erro de predição linear indica quão bem o filtro de predição linear se ajusta às amostras de voz, o analisador LPC pode determinar os coeficientes do filtro ótimo por meio da minimização do erro quadrático médio (MSE – *mean square error*) do erro de predição linear.[18, 19]

A transmissão direta dos parâmetros do filtro de predição linear (LP – *linear prediction*) não faz sentido, pois o filtro é muito sensível a erros nos parâmetros, devido aos ruídos de quantização e de canal. Pior ainda, o filtro LP pode até mesmo se tornar instável por causa de pequenos erros nos coeficientes. Na prática, a estabilidade desse filtro de predição linear com polos apenas pode ser assegurada com a utilização de filtro em estrutura de modular de treliça, por meio do bem conhecido algoritmo de Levinson-Durbin.[20, 21] Parâmetros do filtro de treliça, conhecidos como coeficientes de reflexão $\{r_k\}$, são menos sensíveis a erros de quantização e de ruído. A transmissão pode ser melhorada com o envio das razões logarítmicas de área (LAR – *Log-area ratios*), definidas como

$$o_k \triangleq \log \frac{1 + r_k}{1 - r_k}$$

ou com o envio de valores intermediários da recursão de Levinson-Durbin, conhecidos como coeficientes de reflexão parcial (PARCOR – *partial reflection coefficients*). Outra abordagem prática consiste em determinar os *pares de raias espectrais* (LSP – *line spectral pairs*) equivalentes como uma representação dos coeficientes do filtro LPC para transmissão nos canais. LSP tem a vantagem de baixa sensibilidade ao ruído de quantização.[22, 23] Desde que seja estável, o filtro LP só de polos de ordem p pode ser representado por p frequências de raias espectrais reais. Em cada representação, um filtro sintetizador de ordem p pode ser obtido pelo decodificador LPC a partir dos p coeficientes reais. Em geral, de 8 a 14 parâmetros LP bastam para a representação do trato vocal.

Podemos, agora, usar um exemplo especial de LPC para ilustrar a eficiência de codificação de vocoders baseados em modelo. No chamado vocoder LPC-10,* a fala é amostrada a 8 kHz. Cento e oitenta amostras (22,5 ms) formam um quadro LPC para transmissão.[24] Os bits por quadro de fala são alocados para quantizar o período do timbre, o sinalizador (*flag*) sonoro/surdo, o ganho do filtro e os 10 coeficientes do filtro, segundo a Tabela 6.1. Assim, cada quadro requer entre 32 (surdo) e 53 (sonoro) bits. A adição de bits de controle de quadro resulta em uma palavra de código com comprimento médio de 54 bits por quadro, ou uma taxa global de 2.400 bits/s.[24] Com base em testes subjetivos, esse simples codec LPC-10 tem baixo grau de opinião médio (MOS – *mean opinion score*), mas permite conexões de voz altamente inteligíveis. LPC-10 é parte de FS-1015, um padrão de codec de telefonia segura de baixa taxa desenvolvido pelo Departamento de Defesa do governo dos Estados Unidos em 1984. Um aprimoramento posterior de LPC-10 é conhecido como LPC-10(e).

Em comparação com codecs de forma de onda PCM de 64 kbit/s ou ADPCM de 32 kbit/s, vocoders LPC são muito mais eficientes e podem alcançar taxas de codificação de voz abaixo

*Assim chamado porque usa ordem $p = 10$. A ideia é alocar dois parâmetros a cada possível pico de frequência formante.

de 9,6 kbit/s. O exemplo de LPC-10 de 2,4 kbit/s permite digitalização de fala a uma taxa muito mais baixa que a taxa de amostragem de forma de onda de 8 kHz. A perda de qualidade da fala é uma barganha natural. Para um melhor entendimento da diferença entre vocoders de forma de onda e vocoders baseados em modelo, como LPC, podemos usar a analogia com um serviço de entrega de restaurante. Imaginemos uma família que viva no Alasca e deseje pedir um prato de um famoso restaurante de Nova York. Por motivos práticos, o restaurante teria de enviar pratos preparados crus e congelados, e a família, então, seguiria as instruções de cozimento. A comida, provavelmente, seria saborosa, mas a refeição perderia a fineza do *chef* original. Essa opção é parecida com a transmissão via PCM. O receptor tem os ingredientes básicos, mas deve tolerar o erro de quantização (que se manifesta pela ausência da fineza do *chef* de cozinha). Para reduzir o peso do transporte, a família tem a opção e pedir somente os ingredientes essenciais. Os ingredientes mais pesados e mais comuns (como arroz e batata) podem ser adquiridos localmente. Essa abordagem é como DPCM ou ADPCM, nas quais apenas a parte imprevisível da fala é transmitida. Por fim, a família pode simplesmente encomendar a receita do *chef*. Todos os ingredientes seriam comprados localmente e o prato também seria montado no local. Assim, a família alasquense poderia satisfazer seu desejo de comida gourmet sem receber um único item do restaurante de Nova York! O último cenário captura a ideia de vocoders baseados em modelo. Vocoders LPC basicamente transmitem a receita (ou seja, os parâmetros LPC) para síntese de voz no receptor.

Vocoders LP Práticos de Alta Qualidade

A simples síntese LPC de dois estados ilustrada na Fig. 6.35 descreve somente a ideia básica por trás de codecs de voz baseados em modelo. Na prática, a qualidade de vocoders LPC tem sido melhorada por meio de numerosos codecs mais elaborados. Com a adição de alguns bits, esses vocoders baseados em LP tentam melhorar a qualidade da voz de duas maneiras: por codificação do erro de predição residual e com melhora do sinal de excitação.

Os métodos mais bem-sucedidos pertencem à classe conhecida como vocoders de predição linear com excitação por código (CELP – *Code-Excited Linear Prediction*). Vocoders CELP usam um livro de códigos, tabela de típicos sinais de erro (ou resíduo) LP, definido *a priori* por projetistas. No transmissor, o analisador compara o resíduo real de predição para todas as entradas no livro de códigos, escolhe aquela que representa melhor casamento e adiciona o endereço (código) dessa entrada aos bits para transmissão. O sintetizador recebe esse código, recupera o correspondente resíduo do livro de códigos e o utiliza para modificar a saída sintetizada. Para que CELP funcione bem, o livro de códigos deve ser suficientemente extenso, o que requer mais bits para transmissão. O vocoder FS-1016 representa um aprimoramento do FS-1015 e oferece boa qualidade e fala de som natural, a 4,8 kbits/s.[25] Variantes mais modernas incluem o codec LPC RPE-LTP (*regular pulse excitation, long- term prediction* – excitação normal por pulsos, predição de longo prazo) usado em sistemas celulares GSM, CELP algébrico (ACELP), CELP relaxado (RCELP), Qualcomm CELP (QCELP) de sistemas celulares CDMA e predição linear com excitação por soma vetorial (VSELP – *vector-sum excited linear prediction*). As taxas de dados variam de 1,2 kbit/s a 13 kbit/s (GSM em taxa completa). Estes vocoders formam a base de muitos dos modernos vocoders para sistemas celulares, voz sobre protocolo IP (VoIP) e outros padrões da ITU-T da série G.

Compressão de Vídeo

A digitalização de vídeo e de televisão representa um enorme desafio. Devido à grande largura de banda (cerca de 4,2 MHz), o uso de amostragem e quantização diretas resulta em sinal de vídeo não comprimido de aproximadamente 150 Mbit/s. Em consequência, o modesto grau de compressão oferecido por técnicas como ADPCM e codificação em sub-banda[26, 27] é insuficiente. A chave para a compressão de vídeo está associada à percepção visual humana.

Um grande esforço de pesquisa e desenvolvimento resultou em métodos que reduzem drasticamente a largura de banda digital necessária à transmissão de vídeo. As primeiras técnicas de compressão de sinais de vídeo produziam taxas da ordem de 45 Mbit/s (DS3). Contudo, para emergentes

tecnologias de entrega de vídeo, como HFC, ADSL, HDTV e outras, um maior grau de compressão era necessário. MPEG tratou desse problema e desenvolveu novas técnicas de compressão, que ofereciam qualidade de vídeo de rede ou VCR com graus de compressão muito mais elevados. O MPEG é um esforço conjunto das Organizações Internacionais de Padronização (ISO – *International Standards Organizations*), do Comitê Eletrotécnico Internacional (IEC – *International Electrotechnical Committee*) e do Comitê X3L3 do Instituto Nacional Americano de Padronização (ANSI – *American National Standards Institute*).[28, 29] O MPEG mantém um web site muito informativo, que provê extensa informação sobre as tecnologias MPEG e JPEG e sobre padrões (http://www.mpeg.org/index.html/). Mantém, ainda, um fórum industrial para a promoção dos produtos da organização (http://www.m4if.org/).

O conceito de compressão de vídeo digital se baseia no fato de que, em média, um número relativamente pequeno de pixels muda de um quadro a outro. Portanto, se apenas as alterações forem transmitidas, a largura de banda de transmissão pode ser reduzida drasticamente. A digitalização permite a recuperação sem ruído de sinais analógicos e melhora a qualidade da imagem no receptor. A compressão reduz a largura de banda necessária à transmissão e o espaço para armazenagem de um programa de vídeo e, em consequência, expande a capacidade do canal. Sem compressão, a armazenagem de um programa de vídeo NTSC de duas horas de duração exige aproximadamente 100 gigabytes, muito além da capacidade de qualquer disco DVD.

Há três padrões principais MPEG em uso:

MPEG-1: Usado para vídeo de qualidade VCR e armazenagem em CD de vídeo (ou VCD), a uma taxa de 1,5 Mbit/s. Esses VCDs são muito populares em países da Ásia (excetuando o Japão). Decodificadores MPEG-1 estão disponíveis na maioria dos computadores. O VCD também é um formato muito popular para karaokê.

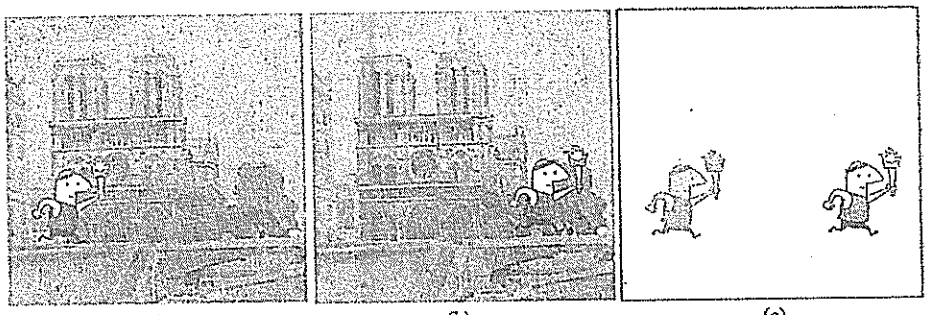
MPEG-2: Este padrão suporta diversas aplicações de codificação de vídeo para transmissão, com qualidade variando de VCR a TV de alta definição (HDTV – *high definition television*), dependendo da taxa de dados. O MPEG-2 oferece uma taxa de compressão de vídeo de 50:1, e é um formato extremamente popular, usado em DVD, HDTV, difusão de vídeo digital terrestre (DVB-T – *terrestrial digital video broadcasting*) e difusão de vídeo digital por satélite (DVB-S).

MPEG-4: Este padrão provê transmissão contínua (*streaming*) de conteúdo multimídia (áudio, visual ou audiovisual) em diferentes larguras de banda, incluindo internet. O MPEG-4 é suportado pelo Media Player de Microsoft Windows, Real Networks, Quicktime e IPod da Apple. Recentemente, o MPEG-4 convergiu a um padrão ITU-T conhecido com H.264, a ser discutido mais adiante.

O poder da compressão de vídeo é fabuloso. Para efeito de comparação, a difusão de televisão NTSC no formato digital requer de 45 a 120 Mbit/s, enquanto MPEG-2 requer de 1,5 a 15 Mbit/s. A HDTV, por sua vez, exige 800 Mbit/s sem compressão; com compressão MPEG-2, a transmissão de HDTV pode ser feita a 19,39 Mbit/s.

Há dois tipos de compressão MPEG, que eliminam redundâncias em sinais audiovisuais não perceptíveis pelo ouvido ou visão humanos:

1. Vídeo
   - Compressão temporal ou *entre quadros* através da predição de movimento entre quadros e remoção de redundância entre quadros.
   - Compressão espacial ou intraquadro, que forma um identificador de bloco para um grupo de pixels com características iguais (cor, intensidade etc.), para cada quadro. Somente o identificador de bloco é transmitido.
2. Áudio, que utiliza um modelo psicoacústico de efeitos mascaradores.



(a) (b) (c)

Figura 6.36
(a) Quadro 1.
(b) Quadro 2.
(c) Informação transferida entre quadros 1 e 2.

A base da compressão de vídeo é a remoção de redundâncias no fluxo (*stream*) de sinal de vídeo. Como exemplo de redundância entre quadros, consideremos a Fig. 6.36a e b. Na Fig. 6.36a, o corredor está na posição A e na Fig. 6.36b, na posição B. Reparemos que o fundo (catedral, prédios e ponte) permanece essencialmente inalterado de um quadro a outro. A Fig. 6.36c representa a informação não redundante para transmissão; ou seja, o que sofreu alteração entre os dois quadros. A imagem do corredor à esquerda representa os blocos do quadro 1 que são substituídos pelo fundo do quadro 2. A imagem do corredor à direita representa os blocos do quadro 1 que substituirão o fundo no quadro 2.

A compressão de vídeo tem início com um codificador, que converte o sinal de vídeo analógico da câmera de vídeo ao formato digital pixel a pixel. Cada quadro de vídeo é dividido em blocos de 8×8 pixels, que são analisados pelo codificador para determinar que blocos devem ser transmitidos, ou seja, que blocos têm alterações significativas de um quadro a outro. Esse processo se dá em duas etapas:

1. Estimação e compensação de movimento. Aqui, um estimador de movimento identifica as áreas ou grupos de blocos de um quadro anterior que casa áreas correspondentes no quadro atual e envia a magnitude e direção do deslocamento a um preditor no decodificador. A informação sobre a diferença entre quadros é chamada de resíduo.
2. Transformação do resíduo quadro a quadro em uma forma mais compacta.

O sinal de resíduo codificado é transformado em uma forma mais compacta por meio de uma transformada de cosseno discreta (DCT – *discrete cosine transform*) (ver Seção 6.5.2 em Haskel *et al.*[28]), que usa um valor numérico para representar cada pixel e normaliza esse valor para transmissão mais eficiente. A DCT é da forma

$$F(j,k) = \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} f(n,m)\cos\left[\frac{(2n+1)j\pi}{2N}\right]\cos\left[\frac{(2m+1)k\pi}{2N}\right]$$

em que $f(n, m)$ é o valor alocado ao bloco na posição (n, m). A transformada inversa é

$$f(n,m) = \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} F(j,k)\cos\left[\frac{(2n+1)j\pi}{2N}\right]\cos\left[\frac{(2m+1)k\pi}{2N}\right]$$

Para um bloco de 8×8, a DCT é multiplicada pela expressão $C(j)C(k)/4$, em que

$$C(x) = \begin{cases} \dfrac{1}{\sqrt{2}} & \text{para } x = 0 \\ 1 & \text{ao contrário} \end{cases}$$

As Tabelas 6.2 e 6.3 listam os valores dos blocos de pixels antes e depois da DCT. Podemos notar, na Tabela 6.3, que há relativamente poucos elementos significativos; ou seja, elementos com valores significativos em relação aos valores centrados na posição 0, 0. Assim, a maioria dos valores na matriz pode ser admitida como zero e, após a transformada inversa, os valores originais são reproduzidos com alta precisão. Este processo em muito reduz a quantidade de dados que devem ser transmitidos, talvez por um fator médio de 8 a 10. Reparemos que o tamanho do resíduo pode ser igual ao de um bloco individual ou, no outro extremo, ao de toda a imagem.

Tabela 6.2
Bloco de Resíduo de 8 × 8 Pixels

	n							
	158	158	158	163	161	161	162	162
	157	157	157	162	163	161	162	162
	157	157	157	160	161	161	161	161
	155	155	155	162	162	161	160	159
m	159	159	159	160	160	162	161	159
	156	156	156	158	163	160	155	150
	156	156	156	159	156	153	151	144
	155	155	155	155	153	149	144	139

Tabela 6.3
Coeficientes da Transformação DCT
do Bloco de Resíduo de 8 × 8 Pixels

	i							
	1259,6	1,0	−12,1	5,2	2,1	1,7	−2,7	−1,3
	22,6	−17,5	6,2	−3,2	2,9	−0,1	−0,4	−1,2
	−10,9	9,3	−1,6	−1,5	0,2	0,9	−0,6	0,1
	7,1	−1,9	−0,2	1,5	−0,9	−0,1	0,0	0,3
k	−0,6	0,8	1,5	−1,6	−0,1	0,7	0,6	−1,3
	−1,8	−0,2	−1,6	−0,3	0,8	1,5	−1,0	−1,0
	−1,3	0,4	−0,3	1,5	−0,5	−1,7	1,1	0,8
	2,6	1,6	3,8	−1,8	−1,9	1,2	0,6	−0,4

Tabela 6.4
Coeficientes Normalizados e
Quantizados da DCT do Resíduo

	jn							
	21	0	−1	0	0	0	0	0
	2	−1	0	0	0	0	0	0
	−1	1	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0

Os valores transmitidos da matriz de um bloco (Tabela 6.4) são normalizados, de modo que os valores na matriz do bloco sejam menores que 1. A matriz resultante é, então, quantizada para produzir a Tabela 6.4. A normalização é feita por uma matriz dinâmica de valores multiplicativos, que são aplicados à matriz transformada elemento a elemento. A matriz normalizada na Tabela 6.4 é a informação de bloco transmitida ao decodificador. A matriz desnormalizada listada na Tabela 6.5 e o resíduo reconstruído (por transformação inversa) na Tabela 6.6 são determinados pelo decodificador. O processo de transformação segue um padrão de zigue-zague, como ilustrado na Fig. 6.37.

Tabela 6.5
Coeficientes DCT Desnormalizados

				jn				
	1260	0	−12	0	0	0	0	0
	23	−18	0	0	0	0	0	0
	−11	10	0	0	0	0	0	0
k	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0

Tabela 6.6
Coeficientes da DCT
Inversa do Resíduo Reconstruído

				n				
	158	158	158	163	161	161	162	162
	157	157	157	162	163	161	162	162
	157	157	157	160	161	161	161	161
m	155	155	155	162	162	161	160	159
	159	159	159	160	160	162	161	159
	156	156	156	158	163	160	155	150
	156	156	156	159	156	153	151	144
	155	155	155	155	153	149	144	139

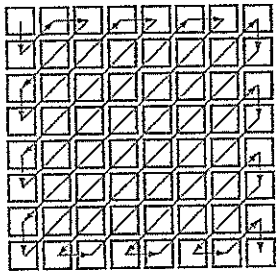


Figura 6.37 Padrão em zigue-zague de varredura de coeficientes DCT.

O MPEG utiliza estimação e compensação de movimento para remover redundâncias temporais (quadro a quadro) de forma própria. Usa três tipos de quadros: quadro I ou intraquadro (às vezes chamado de quadro codificado independentemente ou intracodificado), quadro P ou predito e quando B ou quadro de predição bidirecional. Quadros P são preditos de quadros I. Quadros B são preditos bidirecionalmente, de quadros passados ou futuros. Um quadro I e um ou mais quadros P e B formam o padrão básico de processamento MPEG, chamado de grupo de imagens (GOP – *group of pictures*). A maioria dos quadros em uma imagem comprimida MPEG é de quadros B. Os quadros I provêm a referência inicial para que a diferença entre quadros dispare o processo de codificação MPEG. Reparemos que o aspecto bidirecional do procedimento introduz um atraso na transmissão dos quadros. Isso ocorre porque o GOP é transmitido como uma unidade e, portanto, a transmissão não pode ter início até que o GOP esteja completo (Fig. 6.38). Os detalhes do procedimento estão além do escopo deste texto. Numerosos livros cobrem este assunto em detalhe. Além disso, várias referências sobre compressão MPEG e HDTV podem ser encontradas na internet.

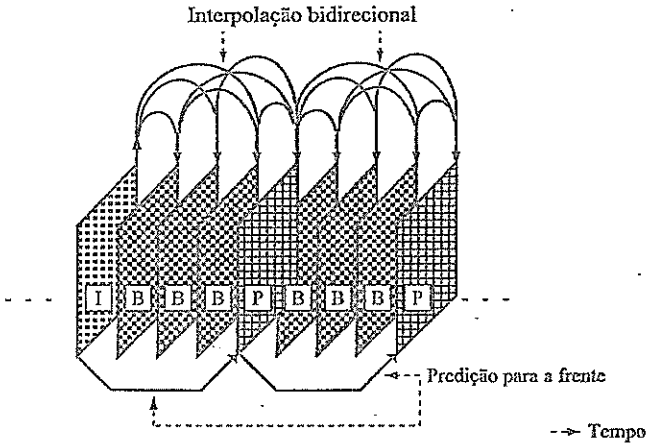


Figura 6.38 Estrutura temporal de quadros MPEG.

Outros Padrões de Compressão de Vídeo

Devemos mencionar que, além do MPEG, há um esforço paralelo da ITU-T para padronizar a codificação de vídeo. Esses padrões aplicam conceitos similares aos estudados para compressão de vídeo. Atualmente, os mais conhecidos padrões ITU-T para compressão de vídeo são os da série H.26x, incluindo H.261, H.263 e H.264. O padrão H.261 foi desenvolvido para transmissão de vídeo a uma taxa de múltiplos de 64 kbit/s para aplicações como videotelefonia e videoconferência. Assim como a compressão MPEG, o H.261 usa predição temporal por compensação de movimento.

O H.263 foi projetado para aplicações em codificação a taxas de bits muito baixas, como videoconferência. Este padrão utiliza a estrutura DCT com compensação de movimento.[30] Baseado no H.261, o H.263 foi otimizado para codificação a taxas de bits muito baixas e sua eficiência de codificação é maior que a do H.261. Flash Video, um formato extremamente popular para compartilhamento de vídeo em muitos web sites, como YouTube e MySpace, emprega uma variante do H.263 conhecida como codec Sorenson Spark.

O padrão H.264 representa, na verdade, a convergência entre ITU-T e MPEG, sendo um esforço conjunto dos dois grupos. Também conhecido como MPEG-4 Parte 10, o H.264 tem melhor desempenho que o MPEG-2 e reduz a taxa de dados em quase 50%. Esse versátil padrão suporta aplicações de vídeo em diferentes graus de largura de banda e qualidade, incluindo serviço de telefonia móvel em 50-60 kbit/s, vídeo com padrão internet a 1-2 Mbit/s e vídeo de alta definição a 5-8 Mbit/s. O H.264 também é suportado em diversos outros produtos e aplicações, como IPod, difusão direta de TV por satélite, algumas TV terrestres regionais, Mac OS X (Tiger) e Playstation Portable da Sony.

Nota sobre Televisão de Alta Definição (HDTV)

A televisão de alta definição (HDTV), que emprega o padrão MPEG-2 para compressão de vídeo, é uma das funções de televisão avançada (ATV – *advanced television*) com vídeo comprimido de 525 linhas para difusão direta por satélite (DBS – *direct broadcast satellite*) ou por cabo. O conceito de HDTV surgiu no final da década de 1970. Os primeiros desenvolvimentos se deram principalmente no Japão, com base em um sistema analógico. Em meados dos anos 1980, ficou claro que a largura de banda exigida pelo sistema analógico era excessiva. Foram, então, iniciados trabalhos para um sistema digital que utilizasse a largura de banda de 6 MHz do sistema de televisão NTSC. No início da década de 1990, foram propostos sete sistemas digitais, mas testes indicaram que nenhum seria satisfatório. Em consequência, em 1993, a FCC sugeriu a formação de uma "Grande Aliança" (GA – *Grand Alliance*) industrial para desenvolver um padrão HDTV. Em dezembro de 1997, o padrão A/53 para difusão de televisão, proposto pelo Comitê de Sistemas Avançados de Televisão (ATSC – *Advanced Television Systems Committee*), foi finalizado pela FCC nos Estados Unidos.

O padrão de HDTV do GA tem por base uma razão de aspecto de 16:9 (razão de aspecto de cinema), em vez da razão de aspecto de 4:3 do sistema NTSC de televisão. A HDTV utiliza compressão MPEG-2 a 19,39 Mbit/s e formato de modulação digital denominado 8-VSB (banda lateral vestigial), que usa um símbolo com oito níveis de amplitude para representar 3 bits de informação. A transmissão é feita em blocos de 207 bytes, incluindo 20 bytes de paridade para código de correção de erro para a frente Reed-Solomon. O formato de pacote com os restantes 187 bytes é um subconjunto do protocolo MPEG-2 e inclui cabeçalhos para sincronização, comutação e outros controles de transmissão.

O Comitê de Sistemas Avançados de Televisão, sucessor da Grande Aliança, prossegue com o desenvolvimento de padrões e recomendações para HDTV, que podem ser encontrados, assim como uma variedade de outras informações, em http://www.atsc.org/.

6.9 EXERCÍCIOS COM O MATLAB

Nos exercícios com o MATLAB desta seção, fornecemos exemplos de amostragem de sinal, reconstrução de um sinal a partir de amostras, quantização uniforme, modulação por modificação de pulsos (PCM) e modulação delta (DM).

Amostragem e Reconstrução de Sinais Passa-Baixas

No exemplo de amostragem, primeiro, construímos um sinal $g(t)$ com duas componentes senoidais de 1 segundo de duração; suas frequências são 1 e 3 Hz. Reparemos, contudo, que a duração finita do sinal implica que o mesmo não é limitado em banda, embora a maior parte do conteúdo esteja em uma largura de banda de 5 Hz. Por isso, selecionamos uma frequência de amostragem de 50 Hz, muito maior que a mínima frequência de Nyquist, de 6 Hz. O programa MATLAB, `Exsample.m`, implementa a amostragem e a reconstrução do sinal. A Fig. 6.39 ilustra o sinal original, suas amostras uniformes à taxa de amostragem de 50 Hz e a resposta de frequência do sinal amostrado. Segundo a análise na Seção 6.1, o espectro do sinal amostrado $g_T(t)$ consiste na repetição do espectro do sinal original a cada 50 Hz.

```
% (Exsample.m)
% Exemplo de amostragem, quantização e retenção de ordem zero
clear;clf;
td=0.002;        %taxa de amostragem original 500 Hz
t=[0:td:1.];     %intervalo temporal de 1 segundo
xsig=sin(2*pi*t)-sin(6*pi*t);   % senoides 1Hz+3Hz
Lsig=length(xsig);

ts=0.02;         %nova taxa de amostragem = 50Hz.
Nfactor=ts/td;
% envia o sinal por meio de um quantizador uniforme de 16 níveis
[s_out,sq_out,sqh_out,Delta,SQNR]=sampandquant(xsig,16,td,ts);
%   recebe 3 sinais:
%       1. sinal amostrado: s_out
%       2. sinal amostrado e quantizado: sq_out
%       3. sinal amostrado, quantizado e com retenção de ordem zero:sqh_out
%
%   calcula transformadas de Fourier
 Lfft=2^ceil(log2(Lsig)+1);
 Fmax=1/(2*td);
```

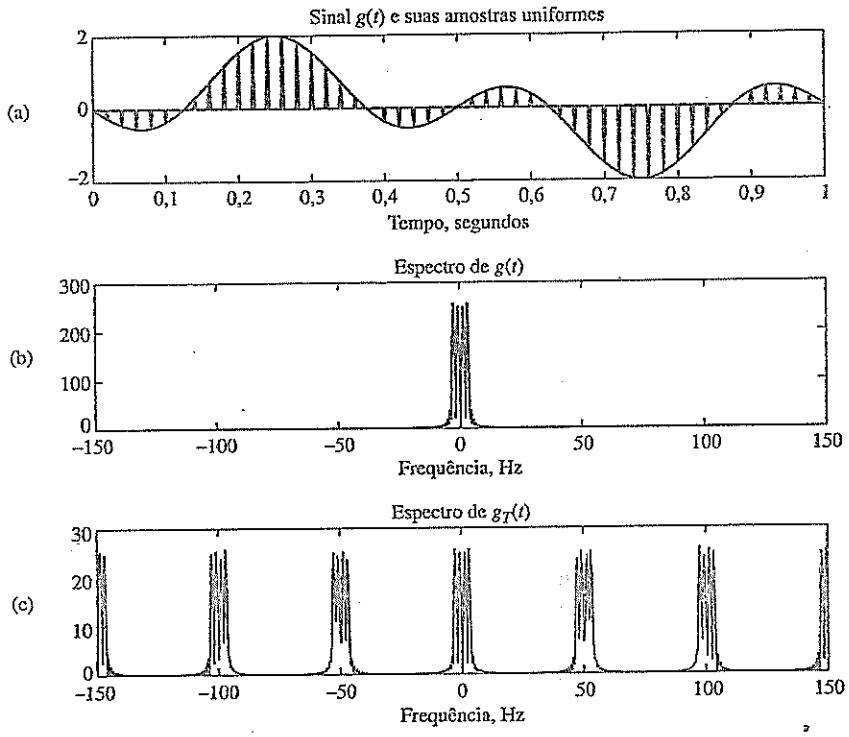


Figura 6.39 Relação entre sinal original e sinal amostrado uniformemente ideal, nos domínios do tempo (a) e da frequência (b, c).

```
 Faxis=linspace(-Fmax,Fmax,Lfft);
 Xsig=fftshift(fft(xsig,Lfft));
 S_out=fftshift(fft(s_out,Lfft));
%   Exemplos de amostragem e reconstrução usando
%       a) trem de impulsos ideais por LPF
%       b) reconstrução com pulso retangular por meio de LPF
%   traça gráfico do sinal original e do sinal amostrado nos domínios do
%   tempo e da frequência
figure(1);
subplot(311); sfig1a=plot(t,xsig,'k');
hold on; sfig1b=plot(t,s_out(1:Lsig),'b'); hold off;
set(sfig1a,'Linewidth',2); set(sfig1b,'Linewidth',2.);
xlabel('tempo, segundos)');
title('sinal {\it g}({\it t}) e suas amostras uniformes');
subplot(312); sfig1c=plot(Faxis,abs(Xsig));
xlabel('frequência (Hz)');
axis([-150 150 0 300])
set(sfig1c,'Linewidth',1);  title('Espectro de {\it g}({\it t})');
subplot(313); sfig1d=plot(Faxis,abs(S_out));
xlabel('frequência (Hz)');
axis([-150 150 0 300/Nfactor])
set(sfig1c,'Linewidth',1);  title('Espectro de {\it g}_T({\it t})');
%   calcula o sinal reconstruído a partir de amostragem ideal e
%   LPF (filtro passa-baixas) ideal
```

```
%   Máxima largura do LPF é igual a BW=floor((Lfft/Nfactor)/2);
 BW=10;                  %Largura de banda não é maior que 10Hz.
 H_lpf=zeros(1,Lfft);H_lpf(Lfft/2-BW:Lfft/2+BW-1)=1;    %ideal LPF
 S_recv=Nfactor*S_out.*H_lpf;        % filtragem ideal
 s_recv=real(ifft(fftshift(S_recv)));  % domínio da frequência reconstruído
 s_recv=s_recv(1:Lsig);                % domínio do tempo reconstruído
%   traça gráfico do sinal reconstruído idealmente nos domínios do
%   tempo e da frequência
figure(2)
subplot(211); sfig2a=plot(Faxis,abs(S_recv));
xlabel('frequência (Hz)');
axis([-150 150 0 300]);
title('Espectro de filtragem ideal (reconstrução)');
subplot(212); sfig2b=plot(t,xsig,'k-.',t,s_recv(1:Lsig),'b');
legend('Sinal original','Sinal reconstruído');
xlabel('Tempo, segundos');
title('Sinal original versus sinal reconstruído idealmente');
set(sfig2b,'Linewidth',2);
%   reconstrução não ideal
 ZOH=ones(1,Nfactor);
 s_ni=kron(downsample(s_out,Nfactor),ZOH);
 S_ni=fftshift(fft(s_ni,Lfft));
 S_recv2=S_ni.*H_lpf;        % filtragem ideal
 s_recv2=real(ifft(fftshift(S_recv2)));   % domínio da frequência reconstruído
 s_recv2=s_recv2(1:Lsig);                 % domínio do tempo reconstruído
%  traça gráfico do sinal reconstruído idealmente nos domínios do tempo
%  e da frequência
figure(3)
subplot(211); sfig3a=plot(t,xsig,'b',t,s_ni(1:Lsig),'b');
xlabel('Tempo, segundos');
title('sinal original versus reconstrução com pulso retangular');
subplot(212); sfig3b=plot(t,xsig,'b',t,s_recv2(1:Lsig),'b--');
legend('sinal original','Reconstrução LPF');
xlabel('Tempo, segundos');
set(sfig3a,'Linewidth',2); set(sfig3b,'Linewidth',2);
title('sinal original versus reconstrução com pulso retangular após LPF');
```

Para construir o sinal original $g(t)$ do trem de impulsos de amostragem $g_\delta(t)$, aplicamos um filtro passa-baixas ideal com largura de banda de 10 Hz no domínio da frequência. Isso corresponde à interpolação usando a função sinc ideal, como mostrado na Seção 6.1.1. O espectro resultante, como ilustrado na Fig. 6. 40, é quase idêntico ao espectro da mensagem original $g(t)$. Além disso, as formas de onda no domínio do tempo também são comparadas na Fig. 6.40 e mostram um casamento quase perfeito.

Em nosso último exercício sobre amostragem e reconstrução, dado no mesmo programa, usamos um pulso retangular de largura T_s (período de amostragem) para reconstruir o sinal original a partir das amostras (Fig. 6.41). Um filtro passa-baixas é aplicado à reconstrução retangular, e o resultado também é mostrado na Fig. 6.41. A comparação com o sinal original deixa claro que o sinal recuperado ainda é muito próximo do sinal original $g(t)$. Isso se deve ao fato de termos escolhido uma alta taxa de amostragem, tal que $T_p = T_s$ fosse tão pequeno que a aproximação na Eq. (6.17) fosse válida. Certamente, com base em nossa análise, com aplicação do filtro de equalização passa-baixas da Eq. (6.16), o erro de reconstrução pode ser muito reduzido.

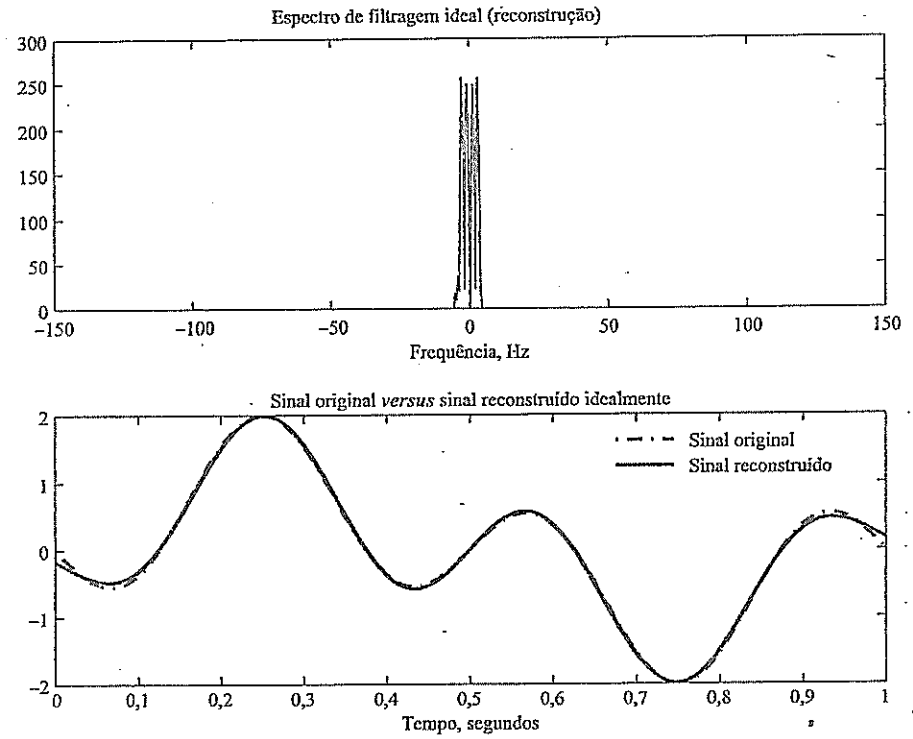


Figura 6.40 Espectro e forma de onda do sinal reconstruído com aplicação de amostragem ideal por impulsos e filtro passa-baixas ideal para reconstrução.

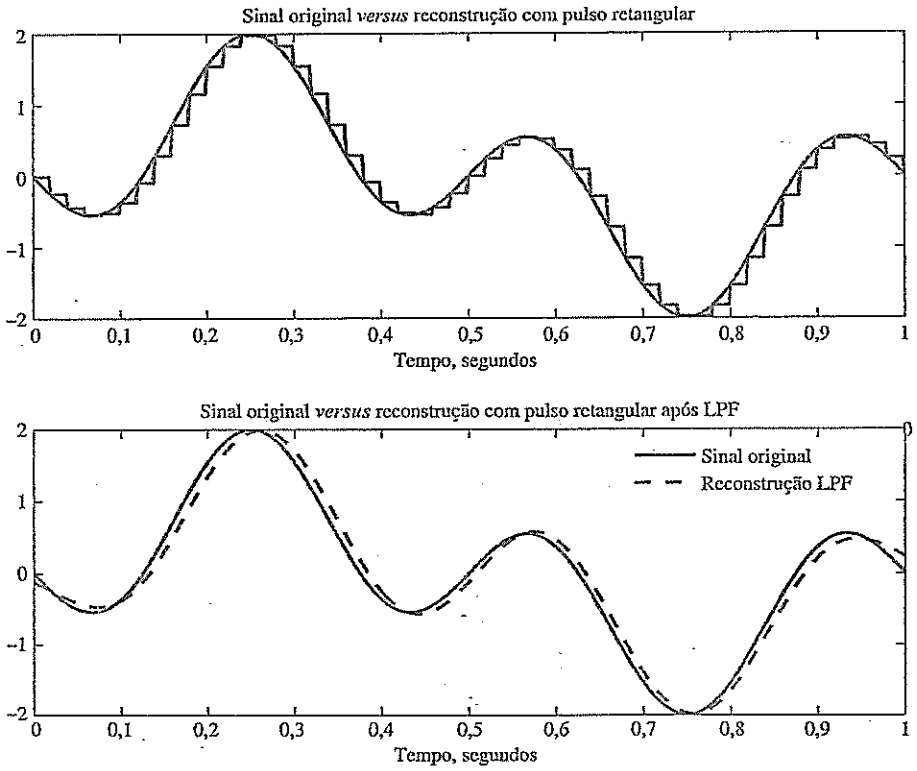


Figura 6.41 Espectro e forma de onda do sinal reconstruído com aplicação de pulso retangular de reconstrução (Fig. 6.6) seguido por LPF sem equalização.

Exemplo de PCM

A quantização uniforme de um sinal analógico usando L níveis de quantização pode ser implementada com a função MATLAB `uniquan.m`.

```
% (uniquan.m)
function [q_out,Delta,SQNR]=uniquan(sig_in,L)
%   Uso
%   [q_out,Delta,SQNR]=uniquan(sig_in,L)
%   L   -    número de níveis de quantização uniforme
%   sig_in -    vetor para sinal de entrada
%   Saídas de função:
%             q_out - saída quantizada
%             Delta - intervalo de quantização
%             SQNR  - real relação sinal-ruído de quantização
sig_pmax=max(sig_in);    % determina pico positivo
sig_nmax=min(sig_in);    % determina pico negativo
Delta=(sig_pmax-sig_nmax)/L;    % intervalo de quantização
q_level=sig_nmax+Delta/2:Delta:sig_pmax-Delta/2;  % define Q níveis
L_sig=length(sig_in);    % determina comprimento do sinal
sigp=(sig_in-sig_nmax)/Delta+1/2;  % converte à faixa de 1/2 a L+1/2
qindex=round(sigp);      % arredonda a 1, 2, ... L níveis
qindex=min(qindex,L);    % elimina L+1 como uma possibilidade rara
q_out=q_level(qindex);   % usa vetor index para gerar saída
SQNR=20*log10(norm(sig_in)/norm(sig_in-q_out)); %valor da SQNR
end
```

A função `sampandquant.m` executa amostragem e quantização uniforme simultaneamente. O período de amostragem `ts` é necessário, assim como o número L de níveis de quantização, para gerar a saída amostrada `s_out`, a saída amostrada e quantizada `sq_out` e o sinal após amostragem, quantização e retenção de ordem zero `sqh_out`.

```
% (sampandquant.m)
function [s_out,sq_out,sqh_out,Delta,SQNR]=sampandquant(sig_in,L,td,ts)
%   Uso
%       [s_out,sq_out,sqh_out,Delta,SQNR]=sampandquant(sig_in,L,td,fs)
%   L  -    número de níveis de quantização uniforme
%   sig_in -   vetor para sinal de entrada
%   td  -   período original de amostragem de sinal de sig_in
%   ts  -   novo período de amostragem
% NOTA:  td*fs deve ser um inteiro positivo;
%   Saídas de função:
%         s_out - saída amostrada
%         sq_out - saída amostrada e quantizada
%         sqh_out- saída amostrada, quantizada e retida
%          Delta - intervalo de quantização
%          SQNR - real relação sinal-ruído de quantização

if (rem(ts/td,1)==0)
nfac=round(ts/td);
p_zoh=ones(1,nfac);
s_out=downsample(sig_in,nfac);
```

```
[sq_out,Delta,SQNR]=uniquan(s_out,L);
s_out=upsample(s_out,nfac);
sqh_out=kron(sq_out,p_zoh);
sq_out=upsample(sq_out,nfac);
else
    warning('Erro! ts/td não é um inteiro!');
    s_out=[];sq_out=[];sqh_out=[];Delta=[];SQNR=[];
end
end
```

O programa MATLAB ExPCM.m fornece um exemplo numérico que usa essas duas funções MATLAB para gerar sinais PCM.

```
% (ExPCM.m)
% Exemplo de amostragem, quantização e retenção de ordem zero
clear;clf;
td=0.002;        %taxa de amostragem original 500 Hz
t=[0:td:1.];     %intervalo temporal de 1 segundo
xsig=sin(2*pi*t)-sin(6*pi*t);   % sinoides de 1Hz+3Hz
Lsig=length(xsig);
Lfft=2^ceil(log2(Lsig)+1);
Xsig=fftshift(fft(xsig,Lfft));
Fmax=1/(2*td);
Faxis=linspace(-Fmax,Fmax,Lfft);
ts=0.02;         %nova taxa de amostragem = 50Hz.
Nfact=ts/td;
% envia o sinal por meio de um quantizador uniforme de 16 níveis
[s_out,sq_out,sqh_out1,Delta,SQNR]=sampandquant(xsig,16,td,ts);
%   sinal PCM obtido, que é
%       - amostrado, quantizado e com retenção de ordem zero: sqh_out
%   traça gráfico do sinal original e do sinal PCM no domínio do tempo
figure(1);
subplot(211);sfig1=plot(t,xsig,'k',t,sqh_out1(1:Lsig),'b');
set(sfig1,'Linewidth',2);
title('Sinal {\it g}({\it t}) e o correspondente sinal PCM de 16 níveis')
xlabel('tempo, segundos');
% envia o sinal por meio de um quantizador uniforme de 16 níveis
[s_out,sq_out,sqh_out2,Delta,SQNR]=sampandquant(xsig,4,td,ts);
%   sinal PCM obtido, que é
%       - amostrado, quantizado e com retenção de ordem zero: sqh_out
%   traça gráfico do sinal original e do sinal PCM no domínio do tempo
subplot(212);sfig2=plot(t,xsig,'k',t,sqh_out2(1:Lsig),'b');
set(sfig2,'Linewidth',2);
title('Sinal {\it g}({\it t}) e o correspondente sinal PCM de 4 níveis')
xlabel('tempo, segundos');

Lfft=2^ceil(log2(Lsig)+1);
Fmax=1/(2*td);
Faxis=linspace(-Fmax,Fmax,Lfft);
SQH1=fftshift(fft(sqh_out1,Lfft));
SQH2=fftshift(fft(sqh_out2,Lfft));
```

```
% Agora, usa LPF para filtrar os dois sinais PCM
 BW=10;                  %Largura de banda não é maior que 10 Hz.
 H_lpf=zeros(1,Lfft);H_lpf(Lfft/2-BW:Lfft/2+BW-1)=1;  %ideal LPF
 S1_recv=SQH1.*H_lpf;                  % filtragem ideal
 s_recv1=real(ifft(fftshift(S1_recv))); % domínio da frequência reconstruído
 s_recv1=s_recv1(1:Lsig);               % domínio do tempo reconstruído
  S2_recv=SQH2.*H_lpf;                  % filtragem ideal
 s_recv2=real(ifft(fftshift(S2_recv))); % domínio da frequência reconstruído
 s_recv2=s_recv2(1:Lsig);               % domínio do tempo reconstruído
 % Traça gráfico dos sinais filtrados versus sinal original
figure(2)
 subplot(211);sfig3=plot(t,xsig,'b-',t,s_recv1,'b-.');
 legend('original','reconstruído')
set(sfig3,'Linewidth',2);
title('Sinal {\it g}({\it t}) e o correspondente sinal PCM filtrado de 16 níveis')
xlabel('tempo, segundos');
subplot(212);sfig4=plot(t,xsig,'b-',t,s_recv2(1:Lsig),'b-.');
 legend('original','reconstruído')
set(sfig4,'Linewidth',2);
title('sinal {\it g}({\it t}) e o correspondente sinal PCM filtrado de 4 níveis')
xlabel('tempo, segundos');
```

No primeiro exemplo, mantivemos a frequência de amostragem de 50 Hz e utilizamos $L = 16$ níveis de quantização uniforme. O resultante sinal PCM é mostrado na Fig. 6.42. Este sinal PCM pode ser aplicado a um filtro passa-baixas no receptor e comparado com o sinal de mensagem original, como na Fig. 6.43. Podemos observar que sinal recuperado é muito próximo do sinal original $g(t)$.

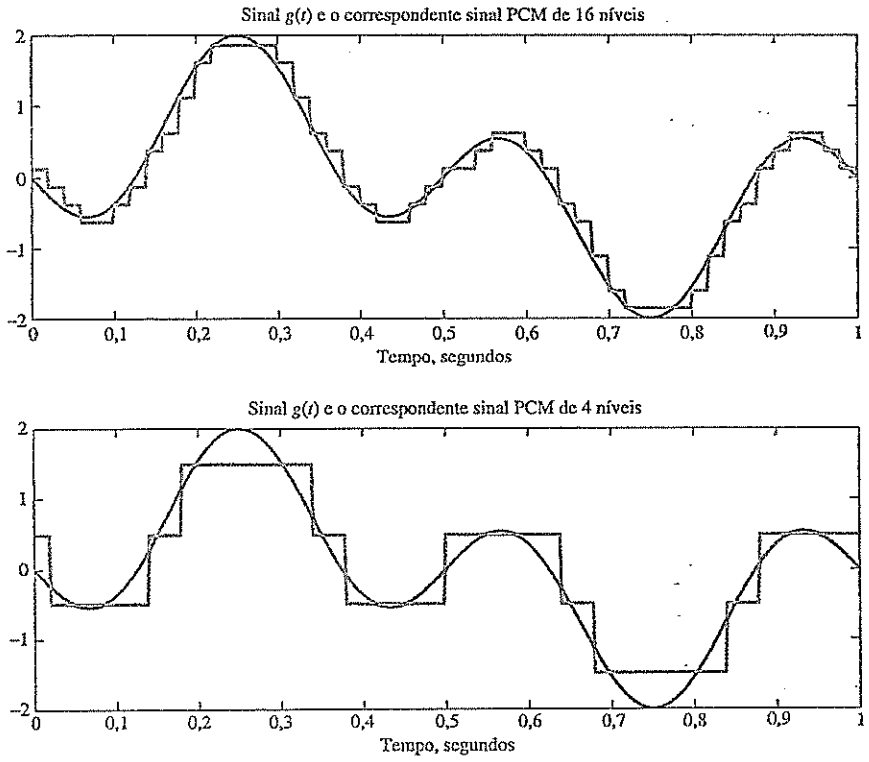


Figura 6.42 Sinal original e sinais PCM obtidos com diferentes níveis de quantização.

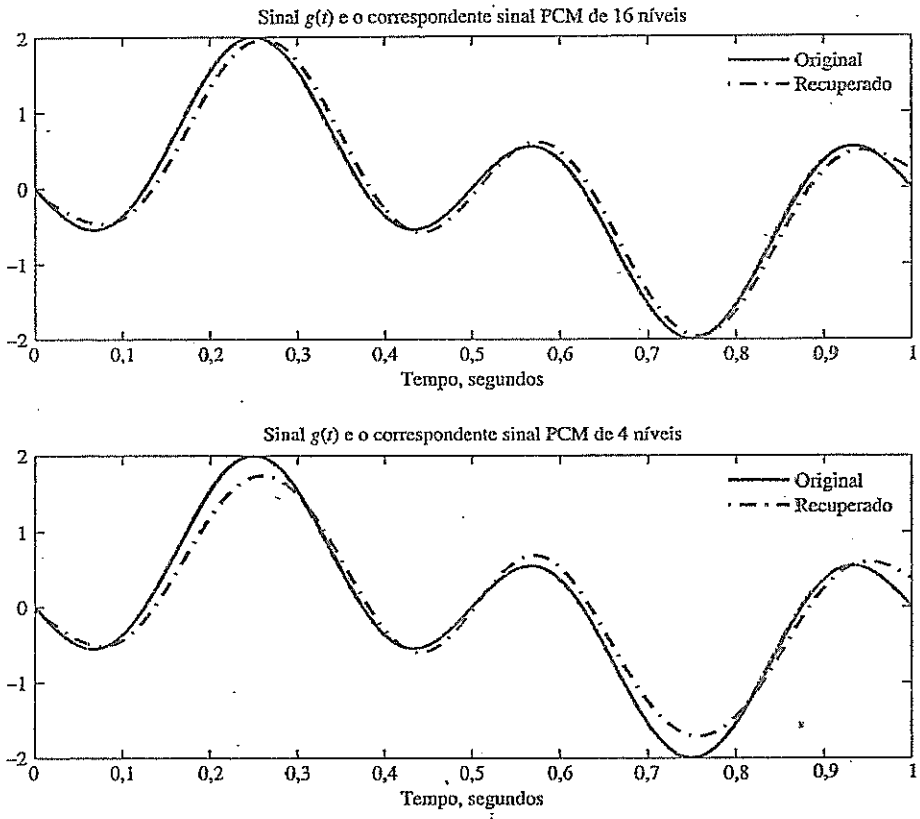


Figura 6.43 Comparação entre o sinal original e os sinais PCM após filtragem passa-baixas para recuperar a mensagem original.

Para ilustrar o efeito da quantização, aplicamos $L = 4$ níveis de quantização PCM. O sinal resultante PCM também é mostrado na Fig. 6.42. O correspondente sinal recuperado é mostrado na Fig. 6.43. Fica claro que o menor número de níveis de quantização ($L = 4$) resulta em maior erro de aproximação.

Modulação Delta

Em vez de aplicar PCM, ilustremos o efeito prático da seleção do incremento Δ no projeto do codificador DM. A função básica para a implementação de DM é dada em `deltamod.m`.

```
% (deltamod.m)
function s_DMout= deltamod(sig_in,Delta,td,ts)
%   Uso
%       s_DMout = deltamod(xsig,Delta,td,ts))
%   Delta  -    Incremento DM
%   sig_in   -    vetor para sinal de entrada
%   td  -      período original de amostragem de sinal de sig_in
%   ts  -      novo período de amostragem
% NOTA:  td*fs deve ser um inteiro positivo;
%   Saídas de função:
%            s_DMout    - saída DM amostrada
if (rem(ts/td,1)==0)
nfac=round(ts/td);
p_zoh=ones(1,nfac);
```

```
s_down=downsample(sig_in,nfac);
Num_it=length(s_down);
s_DMout(1)=-Delta/2;
for k=2:Num_it
    xvar=s_DMout(k-1);
    s_DMout(k)=xvar+Delta*sign(s_down(k-1)-xvar);
end
s_DMout=kron(s_DMout,p_zoh);
else
    warning('Erro! ts/td não é um inteiro!');
    s_DMout=[];
end
end
```

Para gerar sinais DM com diferentes incrementos, aplicamos o mesmo sinal $g(t)$ usado no exemplo PCM. O programa MATLAB ExDM.m aplica três valores de incrementos: $\Delta_1 = 0{,}2$, $\Delta_2 = 2\Delta_1$ e $\Delta_3 = 4\Delta_1$.

```
%   (ExDM.m)
% Exemplo de amostragem, quantização e retenção de ordem zero
clear;clf;
td=0.002;        %taxa de amostragem original 500 Hz
t=[0:td:1.];     %intervalo temporal de 1 segundo
xsig=sin(2*pi*t)-sin(6*pi*t);   % senoides de 1Hz+3Hz
Lsig=length(xsig);
ts=0.02;         %nova taxa de amostragem = 50Hz.
Nfact=ts/td;
% envia o sinal por meio de um quantizador uniforme de 16 níveis
Delta1=0.2;      % Primeiro, seleciona pequeno Delta = 0,2 em DM
s_DMout1=deltamod(xsig,Delta1,td,ts);
%   sinal DM obtido
%   traça gráfico do sinal original e do sinal DM no domínio do tempo
figure(1);
subplot(311);sfig1=plot(t,xsig,'k',t,s_DMout1(1:Lsig),'b');
set(sfig1,'Linewidth',2);
title('Sinal {\it g}({\it t}) e sinal DM')
xlabel('tempo, segundos)'); axis([0 1 -2.2 2.2]);
%
% Aplica DM novamente, dobrando o valor de Delta
Delta2=2*Delta1; %
s_DMout2=deltamod(xsig,Delta2,td,ts);
%   sinal DM obtido
%   traça gráfico do sinal original e do sinal DM no domínio do tempo
subplot(312);sfig2=plot(t,xsig,'k',t,s_DMout2(1:Lsig),'b');
set(sfig2,'Linewidth',2);
title('Sinal {\it g}({\it t}) e sinal DM com incremento dobrado')
xlabel('tempo, segundos'); axis([0 1 -2.2 2.2]);
%
Delta3=2*Delta2;            % Dobra o incremento DM novamente.
s_DMout3=deltamod(xsig,Delta3,td,ts);
%   traça gráfico do sinal original e do sinal DM no domínio do tempo
```

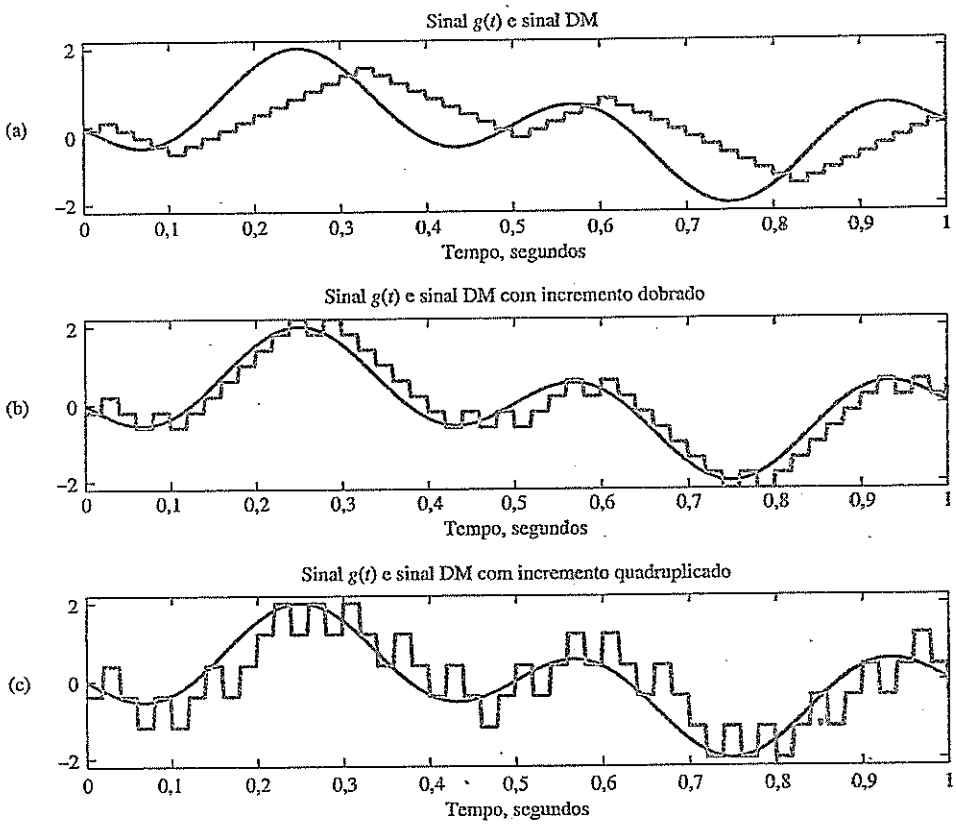


Figura 6.44 Exemplos de saída de modulação delta com três valores de incremento: (a) pequeno incremento resulta em sobrecarga; (b) incremento de valor razoável; (c) incremento grande causa grande erro de quantização.

```
subplot(313);sfig3=plot(t,xsig,'k',t,s_DMout3(1:Lsig),'b');
set(sfig3,'Linewidth',2);
title('Sinal {\it g}({\it t}) e sinal DM com incremento
    quadruplicado')
xlabel('tempo, segundos'); axis([0 1 -2.2 2.2]);
```

Para ilustrar o efeito de DM, os sinais resultantes do codificador DM são mostrados na Fig. 6.44. Este exemplo mostra claramente que, quando o valor do incremento é demasiadamente pequeno (Δ_1), há um grande efeito de sobrecarga, pois o sinal original varia tão rápido que o pequeno incremento não é capaz de acompanhar. Dobrar o incremento DM resolve o problema de sobrecarga neste exemplo. Contudo, a quadruplicação do valor do incremento (Δ_3) produziria erro de quantização desnecessariamente grande. Este exemplo, portanto, confirma nossa análise anterior, que indicou que a seleção do valor do incremento é crítica.

REFERÊNCIAS

1. D. A. Linden, "A discussion of sampling theorem," *Proc. IRE*, vol. 47, no.7, pp. 1219–1226, July 1959.
2. H. P. Kramer, "A Generalized Sampling Theorem," *J. Math. Phys.*, vol. 38, pp. 68–72, 1959.
3. W. R. Bennett, *Introduction to Signal Transmission*, McGraw-Hill, New York, 1970.
4. W. S. Anglin and J. Lambek, *The Heritage of Thales*, Springer, Berlin, 1995.

5. B. Smith, "Instantaneous Companding of Quantized Signals," *Bell Syst. Tech. J.*, vol. 36, pp. 653–709, May 1957.
6. ITU-T Standard Recommendation G.711, English, 1989.
7. ITU-T Standard Recommendation G.726, English, 1990.
8. C. L. Dammann, L. D. McDaniel, and C. L. Maddox, "D-2 Channel Bank Multiplexing and Coding," *Bell Syst. Tech. J.*, vol. 51, pp. 1675–1700, Oct. 1972.
9. K.W. Cattermole, *Principles of Pulse-Code Modulation*, Ilife, England, 1969.
10. Bell Telephone Laboratories, *Transmission Systems for Communication*, 4th ed., Bell, Murray Hill, NJ, 1970.
11. E. L. Gruenberg, *Handbook of Telemetry and Remote Control*, McGraw-Hill, New York, 1967.
12. J. B. O'Neal, Jr., " Delta Modulation Quantizing Noise: Analytical and Computer Simulation Results for Gaussian and Television Input Signals," *Bell Syst. Tech. J.*, pp. 117–141, Jan. 1966.
13. F. de Jager, "Delta Modulation, a Method of PCM Transmission Using the 1-Unit Code," *Philips Res. Rep.*, no. 7, pp. 442–466, 1952.
14. A. Tomozawa and H. Kaneko, "Companded Delta Modulation for Telephone Transmission," *IEEE Trans. Commun. Technol.*, vol. CT-16, pp. 149–157, Feb. 1968.
15. B. S. Atal, "Predictive Coding of Speech Signals at Low Bit Rates," *IEEE Trans. Commun.*, vol. COMM-30, pp. 600–614, 1982.
16. J. P. Campbell and T. E. Tremain, "Voiced/Unvoiced Classification of Speech with Applications to the U.S. Government LPC-10E Algorithm," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Tokyo, pp. 473–476, 1986.
17. A. Gersho, "Advances in Speech and Audio Compression," *Proc. IEEE*, vol. 82, pp. 900–918, 1994.
18. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Englewood Cliffs, NJ, 1978.
19. Lajos Hanzo, Jason Woodward, and Clare Sommerville, *Voice Compression andCommunications*, Wiley, Hoboken; NJ, 2001.
20. N. Levinson, "TheWiener rms Error Criterion in Filter Design and Prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.
21. A. H. Sayed, *Fundamentals of Adaptive Filtering*,Wiley-IEEE Press, Hoboken, NJ, 2003.
22. J. Y. Stein, *Digital Signal Processing: A Computer Science Perspective*,Wiley, Hoboken, NJ, 2000.
23. K. K. Paliwal and B.W. Kleijn, "Quantization of LPC Parameters," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, Eds. Elsevier Science, Amsterdam, 1995.
24. T. E. Tremain, "The Government Standard Linear Predictive Coding Algorithm LPC-10," *Speech Technol.*, 40–49, 1982.
25. M. R. Schroeder and B. S. Atal, "Code-Excited Linear Prediction (CELP): High-Quality Speech at Very Low Bit Rates," in Proc. *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, vol. 10, pp. 937–940, 1985.
26. S. Mallat, "ATheory of Multiresolution Signal Decomposition: TheWavelet Representation," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. 11, pp. 674–693, 1989.
27. M. J. Smith and T. P. Barnwell, "Exact Reconstruction for Tree Structured Sub-Band Coders," *IEEE Trans. Acoustics, Speech, Signal Process.*, vol. 34, no. 3, pp. 431–441, 1986.
28. B. G. Haskel, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman & Hall, New York, 1996.
29. J. L. Mitchell,W. B. Pennebaker, C.E Fogg, and D. J. LeGall, *MPEG Video Compression Standard*, Chapman & Hall, New York, 1996.
30. ITU-T Recommendation H.263, Video Coding for Low Bit Rate Communication.

EXERCÍCIOS

6.1-1 A Fig. E.6.1-1 mostra os espectros de Fourier dos sinais $g_1(t)$ e $g_2(t)$. Determine o intervalo de Nyquist e a taxa de amostragem para os sinais $g_1(t)$, $g_2(t)$, $g_1^2(t)$, $g_2^m(t)$ e $g_1(t)g_2(t)$.

Sugestão: Use convolução em frequência e a propriedade de largura da convolução.

Figura E.6.1-1

6.1-2 Determine a taxa de amostragem e o intervalo de amostragem de Nyquist para os sinais:

(a) sinc $(100\pi t)$
(b) sinc^2 $(100\pi t)$
(c) sinc $(100\pi t)$ + sinc $(50\pi t)$
(d) sinc $(100\pi t)$ + 3 sinc^2 $(60\pi t)$
(e) sinc $(50\pi t)$ sinc $(100\pi t)$

6.1-3 Um sinal $g(t)$ limitado em banda a B Hz é amostrado por um trem de pulsos periódicos $p_{T_s}(t)$, consistindo em um pulso retangular de largura $1/8B$ segundo (centrado na origem) que é repetido à taxa de Nyquist ($2B$ pulsos por segundo). Mostre que o sinal amostrado $\overline{g}(t)$ é dado por

$$\overline{g}(t) = \frac{1}{4}g(t) + \sum_{n=1}^{\infty} \frac{2}{n\pi} \operatorname{sen}\left(\frac{n\pi}{4}\right) g(t) \cos 4n\pi Bt$$

Mostre que o sinal $g(t)$ pode ser recuperado com a aplicação de $\overline{g}(t)$ a um filtro passa-baixas ideal de largura de banda B Hz e ganho 4.

6.1-4 Um sinal $g(t) = \text{sinc}^2\,(5\pi t)$ é amostrado (usando impulsos uniformemente espaçados) a uma taxa de (i) 5 Hz; (ii) 10 Hz; (iii) 20 Hz. Para cada caso,

(a) Esboce o gráfico do sinal amostrado no domínio do tempo.

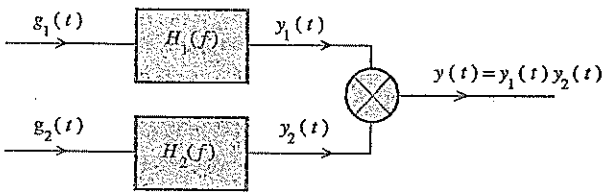
(b) Esboce o gráfico do espectro do sinal amostrado.

(c) Explique se é possível recuperar o sinal $g(t)$ do sinal amostrado.

(d) Se o sinal amostrado for aplicado a um filtro passa-baixas ideal de largura de banda 5 Hz, esboce o gráfico do espectro do sinal de saída.

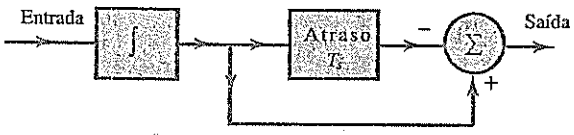
6.1-5 Os sinais $g_1(t) = 10^4\Pi(10^4 t)$ e $g_2(t) = \delta(t)$ são aplicados às entradas de filtros passa-baixas ideais $H_1(f) = \Pi(f/20.000)$ e $H_2(f) = \Pi(f/10.000)$ (Fig. E.6.1-5). As saídas $y_1(t)$ e $y_2(t)$ desses filtros são multiplicadas para produzir o sinal $y(t) = y_1(t)y_2(t)$. Determine a taxa de Nyquist para $y_1(t)$, $y_2(t)$ e $y(t)$. Use a propriedade da convolução e a propriedade de largura da convolução para determinar a largura de banda de $y_1(t)y_2(t)$. Veja, também, o Exercício 6.1-1.

Figura E.6.1-5



6.1-6 Um circuito de retenção de ordem zero (Fig. E.6.1-6) é, muitas vezes, usado para reconstruir um sinal $g(t)$ a partir de suas amostras.

Figura E.6.1-6



(a) Determine a resposta ao impulso unitário desse circuito.

(b) Determine a função de transferência $H(f)$ e esboce o gráfico de $|H(f)|$.

(c) Mostre que, quando um sinal amostrado $\overline{g}(t)$ é aplicado à entrada desse circuito, a saída é uma aproximação em degraus de $g(t)$. O intervalo de amostragem é T_s.

6.1-7 (a) Um circuito de retenção de primeira ordem também pode ser usado para reconstruir um sinal $g(t)$ de suas amostras. A resposta ao impulso desse circuito é

$$h(t) = \Delta\left(\frac{t}{2T_s}\right)$$

em que T_s é o intervalo de amostragem. Considere um típico sinal amostrado $\overline{g}(t)$ e mostre que esse circuito efetua interpolação linear. Em outras palavras, a saída do filtro consiste em amostras cujos picos são conectados por segmentos de linha reta. Siga o procedimento discutido na Seção 6.1.1 (Fig. 6.2b).

(b) Determine a função de transferência e a resposta de amplitude desse filtro e compare-a com a do filtro ideal necessário à reconstrução do sinal.

(c) Por ser não causal, esse filtro é irrealizável. Sugira uma modificação que torne esse filtro realizável. Como tal modificação afetaria a reconstrução de $g(t)$ de suas amostras? Como afetaria a resposta de frequência do filtro?

6.1-8 Prove que um sinal não pode ser simultaneamente limitado no tempo e em largura de banda.

Sugestão: Mostre que a hipótese oposta leva a uma contradição. Admita que um sinal seja simultaneamente limitado no tempo e em largura de banda, de modo que $G(f) = 0$ para $|f| > B$. Neste caso, $G(f) = G(f)\,\Pi(f/2B')$, para $B' > B$. Isso significa que $g(t)$ é igual a $g(t) * 2B'\,\mathrm{sinc}(2\pi B't)$. Mostre que a última função não pode ser limitada no tempo.

6.2-1 O Código Padrão Americano para Intercâmbio de Informação (ASCII – *American Standart Code for Information Interchange*) tem 128 caracteres, que são codificados em binário. Se um certo computador gerar 100.000 caracteres por segundo, determine:

(a) O número de bits (dígitos binários) necessários para a representação de cada caractere.

(b) O número de bits por segundo necessário para transmitir a saída do computador e a mínima largura de banda exigida para a transmissão do sinal.

(c) Para a capacidade de detecção de um erro, um bit (bit de paridade) é adicionado ao código de cada caractere. Modifique suas respostas aos itens (a) e (b) tendo em vista essa informação.

6.2-2 Um *compact disc* (CD) grava sinais de áudio digitalmente, com uso de PCM. Admita que a largura de banda do sinal de áudio seja igual a 15 kHz.

(a) Se as amostras de Nyquist forem uniformemente quantizadas em $L = 65.536$ níveis e, então, codificadas em binário, determine o número de dígitos binários necessários para codificar uma amostra.

(b) Se o sinal de áudio tiver potência média de 0,1 watt e tensão de pico de 1 volt, determine a resultante relação sinal-ruído de quantização (SQNR) da saída quantizada uniformemente na parte (a).

(c) Determine o número de dígitos binários por segundo (bit/s) requerido para a codificação do sinal de áudio.

(d) Pelas razões práticas discutidas no texto, sinais são amostrados a uma taxa bem acima da taxa de Nyquist. CDs usados na vida real empregam 44.100 amostras por segundo. Se $L = 65.536$, determine o número de bits por segundo necessário para a codificação do sinal, e a mínima largura de banda requerida para a transmissão do sinal codificado.

6.2-3 Um sinal de televisão (vídeo e áudio) tem largura de banda de 4,5 MHz. Esse sinal é amostrado e codificado em binário para produzir um sinal PCM.

(a) Determine a taxa de amostragem quando o sinal for amostrado a uma taxa 20% acima da taxa de Nyquist.

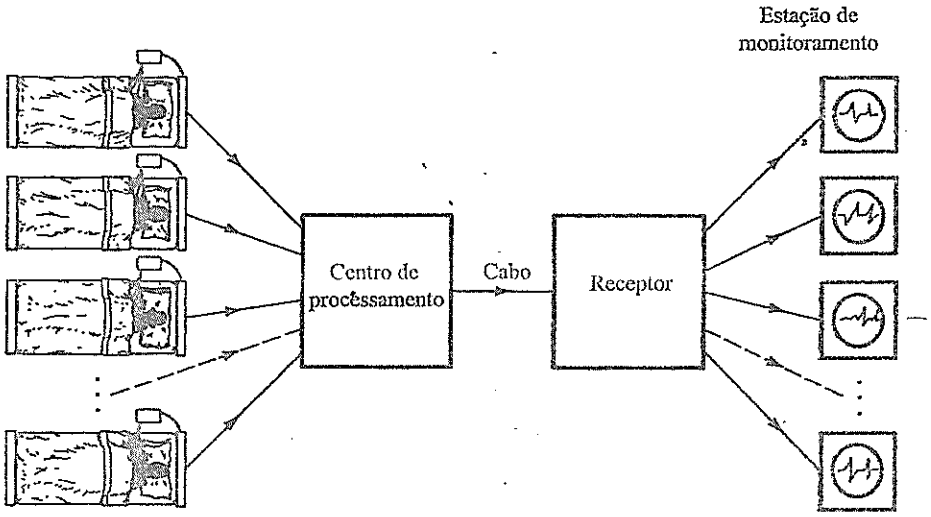
(b) Se as amostras forem quantizadas em 1024 níveis, determine o número de pulsos binários necessários à codificação de cada amostra.

(c) Determine a taxa de pulsos binários (bits por segundo) do sinal codificado em binário e a mínima largura de banda necessária à transmissão do sinal.

6.2-4 Cinco sinais de telemetria, cada um com largura de banda de 240 Hz, devem ser transmitidos simultaneamente por PCM binário. Os sinais devem ser amostrados a uma taxa pelo menos 20% acima da taxa de Nyquist. O enquadramento e a sincronização requerem 0,5% bits adicionais. Um codificador PCM é usado para converter esses sinais antes que sejam multiplexados no tempo em um fluxo de dados. Determine a mínima taxa de bits (bit/s) possível que deve ser transmitida e a mínima largura de banda necessária à transmissão do sinal multiplexado.

6.2-5 Desejamos montar uma estação central para o monitoramento simultâneo de eletrocardiogramas (ECG) de 10 pacientes hospitalizados. Os dados dos 10 pacientes são levados ao centro de processamento por meio de fios, e são amostrados, quantizados, codificados em binário e multiplexados por divisão no tempo. Os dados multiplexados são, então, transmitidos à estação de monitoramento (Fig. E.6.2-5). A largura de banda do sinal ECG é de 100 Hz. O máximo erro aceitável em amplitudes de amostras é de 0,25% da amplitude de pico do sinal. A taxa de amostragem deve ser pelo menos o dobro da taxa de Nyquist. Determine a mínima largura de banda do cabo necessária para transmitir estes dados.

Figura E.6.2-5

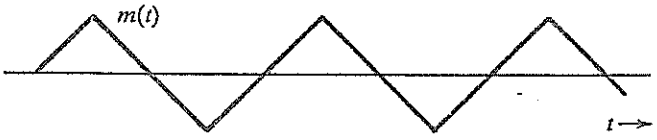


6.2-6 Um sinal de mensagem $m(t)$ é transmitido por PCM binário sem compressão. Se a necessária SQNR for de pelo menos 47 dB, determine o mínimo valor requerido para $L = 2^n$, supondo que $m(t)$ seja senoidal. Determine a real SQNR obtida com este valor mínimo de L.

6.2-7 Repita o Exercício E.6.2-6 para o sinal $m(t)$ mostrado na Fig. E.6.2-7.

Sugestão: A potência de um sinal periódico é a média de sua energia em um ciclo. Neste caso, como a amplitude do sinal assume o mesmo valor a cada quarto de ciclo, a potência também pode ser obtida da energia média em um quarto de ciclo.

Figura E.6.2-7



6.2-8 Para um sinal PCM, determine L quando o parâmetro de compressão $\mu = 100$ e a mínima SNR necessária é de 45 dB. Determine, com este valor de L, a SQNR da saída. Lembre-se de que L deve ser uma potência de 2, ou seja, $L = 2^n$, para um sinal PCM binário.

6.2-9 Um sinal limitado em banda a 1 MHz é amostrado a uma taxa 50% maior que a taxa de Nyquist e, com o uso de um quantizador de lei μ, é quantizado em 256 níveis empregando $\mu = 255$.

(a) Determine a relação sinal-ruído de quantização (SQNR).

(b) A SQNR (qualidade do sinal recebido) calculada na parte (a) era insatisfatória, e deve ser aumentada em pelo menos 10 dB. Caso fosse observado que uma taxa de amostragem 20% acima da de Nyquist era suficiente, você seria capaz de obter a desejada SQNR sem aumentar largura de banda de transmissão?

6.2-10 Foi determinado que a SQNR de saída de um PCM de 10 bits estava 30 dB abaixo do necessário. Para alcançar a desejada SNR de 42 dB, foi aumentado o número de níveis de quantização L. Determine o aumento relativo na largura de banda de transmissão acarretado pelo aumento de L.

6.4-1 Em um certo sistema de telemetria, há quatro sinais analógicos $m_1(t)$, $m_2(t)$, $m_3(t)$ e $m_4(t)$. A largura de banda de $m_1(t)$ é de 3,6 kHz; a largura de banda de cada um dos sinais restantes é de 1,4 kHz. Cada sinal deve ser amostrado a taxa não inferior à correspondente taxa de Nyquist, e os sinais devem ser multiplexados palavra a palavra. Isso pode ser feito com a multiplexação de amostras PAM dos quatro sinais, seguida da codificação binária das amostras multiplexadas (como no caso da portadora T1 na Fig. 6.20a). Sugira um esquema de multiplexação adequado a este propósito. Qual é a frequência do comutador (em rotações por segundo)? *Nota*: Em alguns casos, pode ser necessário amostrar algum(ns) sinal(is) a taxa(s) superior(es) à(s) correspondente(s) taxa(s) de Nyquist.

6.4-2 Repita o Exercício 6.4-1 para o caso de quatro sinais $m_1(t)$, $m_2(t)$, $m_3(t)$ e $m_4(t)$ com larguras de banda de 1200, 700, 300 e 200 Hz, respectivamente.

Sugestão: Primeiro multiplexe $m_2(t)$, $m_3(t)$ e $m_4(t)$ e, depois, multiplexe o sinal composto com $m_1(t)$.

6.4-3 Um sinal $m_1(t)$ é limitado em banda a 3,6 kHz, e três outros sinais $m_2(t)$, $m_3(t)$ e $m_4(t)$ são limitados em banda a 1,2 kHz cada um. Esses sinais são amostrados à taxa de Nyquist e codificados em binário com 512 níveis de quantização ($L = 512$). Sugira um esquema adequado de multiplexação bit a bit (como na Fig. 6.12). Qual é a frequência do comutador (em rotações por segundo) e qual é a taxa de bits de saída?

6.7-1 Em um sistema DM de integração simples, o sinal de voz é amostrado a uma taxa de 64 kHz, como em PCM. A máxima amplitude normalizada de sinal é $A_{max} = 1$.

(a) Determine o valor mínimo do incremento σ para evitar sobrecarga.

(b) Determine a potência de ruído granular N_o se a largura de banda do sinal de voz for 3,4 kHz.

(c) Admitindo que o sinal de voz seja senoidal, determine S_o e a SNR.

(d) Admitindo que a amplitude do sinal de voz seja uniformemente distribuída no intervalo $(-1, 1)$, determine S_o e a SNR.

(e) Determine a mínima largura de banda de transmissão.

7 PRINCÍPIOS DE TRANSMISSÃO DE DADOS DIGITAIS

Durante a maior parte do século vinte, uma parcela significativa de sistemas de comunicação era de forma analógica. Contudo, ao final da década de 1990, o formato digital começou a dominar a maioria das aplicações. Não é necessário muito esforço para que percebamos a contínua migração de comunicação analógica para digital: do áudio em fita cassete para MP3 e CD, da TV analógica NTSC ou PAL-M para a digital HDTV, da telefonia tradicional para VoIP, e do vídeo em VHS para DVD. Na verdade, até o último refúgio analógico da radiodifusão enfrenta um poderoso competidor digital, na forma de rádio por satélite. Dada a importância dominadora de sistemas de comunicação digital em nosso dia a dia, nunca é cedo demais para estudar os princípios básicos e vários aspectos de comunicação de dados digitais, como faremos neste capítulo.

Este capítulo trata do problema da transmissão de dados digitais em um canal. Portanto, consideraremos que as mensagens iniciais são digitais. Iniciaremos com a consideração do caso binário, em que os dados consistem em somente dois símbolos: **1** e **0**. Aloquemos uma forma de onda (pulso) distinta a cada um desses símbolos. A resultante sequência de pulsos é transmitida através do canal. No receptor, esses pulsos são detectados e convertidos aos dados binários (**1**s e **0**s).

7.1 SISTEMAS DE COMUNICAÇÃO DIGITAL

Um sistema de comunicação digital consiste em diversos componentes, como indicado na Fig. 7.1. Nesta seção, delinearemos, de forma conceitual, a funcionalidade de cada componente em sistemas de comunicação. Os detalhes de análise e projeto serão apresentados posteriormente no capítulo, em seções dedicadas.

7.1.1 Fonte

A entrada de um sistema digital toma a forma de uma sequência de dígitos. A entrada pode ser a saída de um conjunto de dados, de um computador, um sinal de áudio digitalizado (PCM, DM ou LPC), facsímile digital, HDTV, dados de telemetria, e assim por diante. Embora a maior parte da discussão neste capítulo seja restrita ao caso binário (esquemas de comunicação que usam apenas dois símbolos), o caso mais geral de comunicação M-ária, que utiliza M símbolos, também será discutido nas Seções 7.7 e 7.9.

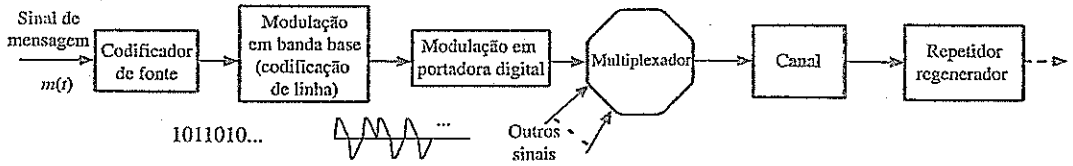


Figura 7.1 Blocos fundamentais de um sistema de comunicação digital.

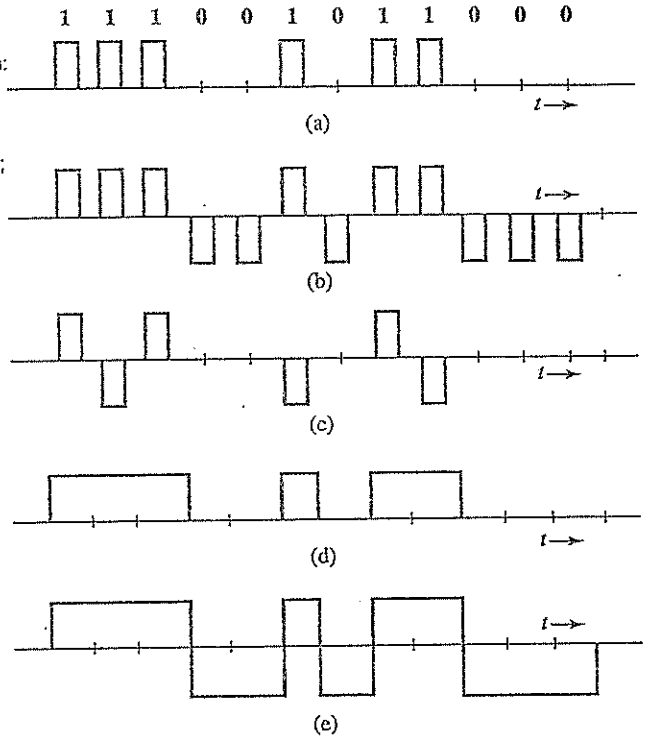


Figura 7.2
Exemplos de códigos de linha:
(a) on-off (RZ);
(b) polar (RZ);
(c) bipolar (RZ);
(d) on-off (NRZ);
(e) polar (NRZ).

7.1.2 Codificador de Linha

A saída digital de um codificador de fonte é convertida (ou codificada) em pulsos elétricos (formas de onda) para fins de transmissão no canal. Esse processo recebe a denominação **codificação de linha** ou **codificação de transmissão**. Há muitas formas possíveis de alocar formas de onda (pulsos) aos dados digitais. No caso binário (2 símbolos), por exemplo, a codificação de linha conceitualmente mais simples é **on-off** (liga-desliga), em que **1** é transmitido por um pulso $p(t)$ e **0** é transmitido por ausência de pulso (sinal nulo), como mostrado na Fig. 7.2a. Outro código de uso muito comum é o **polar**, em que **1** é transmitido por um pulso $p(t)$ e **0** é transmitido por um pulso $-p(t)$ (Fig. 7.2b). O esquema polar é o código com a maior eficiência de potência, pois, para um dado grau de imunidade ao ruído (probabilidade de erro), requer a menor potência. Outro código popular em PCM é o **bipolar**, também conhecido como **pseudoternário** ou **inversão alternada de sinal** (AMI – *alternate mark inversion*), em que **0** é codificado por ausência de pulso e **1** é codificado por um pulso $p(t)$ ou $-p(t)$, dependendo se o **1** anterior foi codificado por $-p(t)$ ou $p(t)$. Em resumo, pulsos que representam **1**s sucessivos têm sinais alternados, como indicado na Fig. 7.2c. Esse código tem a vantagem de que, se ocorrer *um único* erro na detecção de pulsos, a sequência de pulsos recebida violará a regra bipolar e o erro poderá ser detectado imediatamente (embora não possa ser corrigido).*

Outro código de linha que, inicialmente, pareceu promissor é o duobinário (e duobinário modificado), proposto por Lender.[1,2] Esse código é melhor que o bipolar em termos de eficiência de largura de banda. Sua versão mais proeminente, o código de linha *duobinário modificado*, tem sido aplicada em canais de leitura de hard disc drives, em transmissão óptica a 10 Gbit/s para redes metropolitanas, e na

*Isso admite não mais que um erro em sequência. Múltiplos erros em sequência podem cancelar os respectivos efeitos e, assim, não serem detectados. Contudo, a probabilidade de múltiplos erros é muito menor que a de apenas um erro. Mesmo para um erro, não podemos identificar sua localização exatamente. Portanto, esse código detecta a presença de erros isolados, mas não é capaz de corrigi-los.

primeira geração de modems para rede digital de serviços integradas (ISDN–*integrated services digital networks*). Detalhes de códigos de linha duobinários serão discutidos mais adiante neste capítulo.

Até aqui, em nossa discussão, usamos pulsos de meia largura somente para facilitar a ilustração. Podemos selecionar outras larguras. Pulsos de largura completa são usados em algumas aplicações. Sempre que pulsos de largura completa são utilizados, a amplitude do pulso é mantida em um valor constante durante toda a duração do pulso (isto é, a amplitude não retorna a zero antes do início do pulso seguinte). Por essa razão, tais esquemas são denominados **sem retorno ao zero** ou esquemas **NRZ** (*non-return-to-zero*), em contraste com esquemas com **retorno ao zero** ou **RZ** (*return-to-zero*) (Fig. 7.2a-c). A Fig. 7.2d mostra um sinal NRZ on-off e a Fig. 7.2e, um sinal NRZ polar.

7.1.3 Multiplexador

Em termos gerais, a capacidade de um canal físico (por exemplo, cabo coaxial, fibra óptica) para transmitir dados é muito maior que a taxa de dados de uma fonte. Para utilizar essa capacidade de modo eficaz, combinamos várias fontes por meio de um multiplexador digital. A multiplexação digital pode ser feita por divisão em frequência ou por divisão no tempo, como já discutimos. A divisão em código (a ser discutida no Capítulo 11) também é uma alternativa prática e eficiente. Um canal físico é, portanto, compartilhado por várias mensagens simultâneas.

7.1.4 Repetidor Regenerador

Repetidores regeneradores são usados em intervalos regularmente espaçados, ao longo de uma linha de transmissão digital, para detectar o sinal digital que chega e regenerar novos pulsos "limpos" para transmissão em mais um trecho da linha. Esse processo periódico elimina e, portanto, combate, o acúmulo de ruído e distorção de sinal na extensão da rota de transmissão. A capacidade de eliminação de ruído e efeitos de distorção de sinal desses repetidores regeneradores é uma das maiores vantagens de sistemas de comunicação digital em relação aos sistemas analógicos.

Se forem transmitidos pulsos a uma taxa de R_b pulsos por segundo, é necessário ter informação de temporização periódica – o sinal de relógio na frequência R_b Hz – para amostrar os pulsos que chegam ao receptor. Essa informação de temporização pode ser extraída do próprio sinal recebido, desde que o código de linha seja escolhido de forma adequada. A retificação do sinal polar *RZ* na Fig. 7.2b, por exemplo, produz um sinal de relógio periódico de frequência R_b Hz, que contém o desejado sinal periódico de temporização de frequência R_b Hz. Quando esse sinal é aplicado a um circuito ressonan-

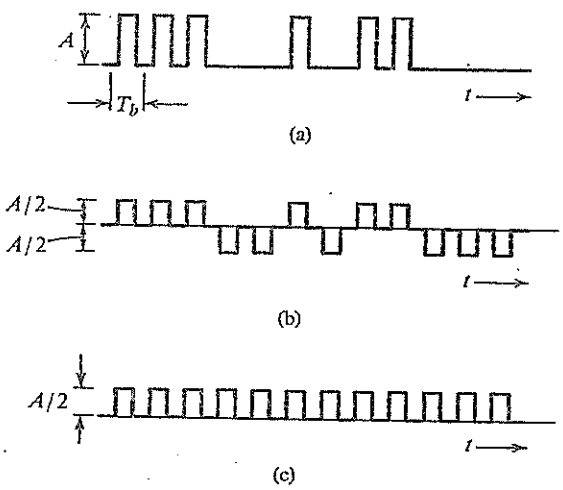


Figura 7.3 Sinal on-off (a) é a soma de um sinal polar randômico (b) com um sinal periódico na frequência do relógio (c).

te sintonizado na frequência R_b, a saída, que é uma senoide de frequência R_b Hz, pode ser usada para a temporização. O sinal on-off pode ser expresso como a soma de um sinal periódico (na frequência do relógio) e um sinal polar ou randômico, como mostrado na Fig. 7.3. Devido à presença da componente periódica, podemos extrair a informação de temporização desse sinal com o uso de um circuito ressonante sintonizado na frequência do relógio. Um sinal bipolar, ao ser retificado, se torna um sinal on-off. Portanto, a informação de temporização pode ser extraída da mesma forma como para um sinal on-off.

O sinal de temporização (saída do circuito ressonante) é sensível ao padrão de bit de chegada. No caso on-off ou polar, um 0 é transmitido por "ausência de pulso". Assim, se houver um grande número de 0s em uma sequência (ausência de pulsos), não haverá sinal na entrada do circuito ressonante e sua saída senoidal passa a decair, originando erro na informação de temporização. Mais adiante, discutiremos formas de solucionar esse problema. Dizemos que código de linha é **transparente** quando o padrão de bits não afeta a precisão da informação de temporização. O esquema polar *RZ* (em que cada bit é transmitido por algum pulso) é transparente; os esquemas on-off e bipolar não são transparentes.

7.2 CODIFICAÇÃO DE LINHA

Dados digitais podem ser transmitidos por vários **códigos de linha** ou de **transmissão**. Vimos exemplos de códigos on-off, polar e bipolar. Cada código de linha tem suas vantagens e desvantagens. Entre outras propriedades desejáveis, um código de linha deve ter as seguintes:

- *Largura de banda de transmissão* tão pequena quanto possível.
- *Eficiência de potência.* Para dadas largura de banda e taxa de erros de detecção, a potência transmitida deve ser tão baixa quanto possível.
- *Capacidade de detecção e correção de erros.* É desejável que erros sejam detectados e, de preferência, corrigidos. No caso bipolar, por exemplo, um erro isolado causará violação da regra bipolar e poderá ser facilmente detectado. Códigos corretores de erros serão discutidos em detalhe no Capítulo 14.
- *Densidade espectral de potência favorável.* É desejável ter densidade espectral de potência (PSD) zero em $f = 0$ (dc), pois acoplamento e transformadores ac são usados nos repetidores.* Potência significativa nas componentes de frequências baixas também deve ser evitada, pois causa ondulação dc (*dc wander*) na sequência de pulsos quando acoplamento ac é usado.
- *Conteúdo de temporização adequado.* Deve ser possível extrair a informação de temporização ou de relógio do sinal.
- *Transparência.* A transmissão correta de um sinal digital deve ser possível, independentemente do padrão de 1s e 0s. Vimos, nos casos de códigos on-off e bipolar, que uma longa sequência de 0s pode causar problema para a extração de temporização. Um código é transparente se os dados forem codificados de modo que, para qualquer sequência possível de bits, o sinal codificado possa ser detectado com fidelidade.

7.2.1 PSD de Diversos Códigos de Linha

No Exemplo 3.19, discutimos o procedimento para determinar a PSD de um trem de pulsos polares. Usemos um procedimento similar para determinar uma expressão geral para PSD de sinais resultantes da modulação em banda base (codificação de linha), como ilustrado na Fig. 7.1. Em particular, aplicamos diretamente a relação entre PSD e a função de autocorrelação do sinal da modulação em banda base, dada na Seção 3.8 [Eq. (3.85)].

Na discussão a seguir, consideramos um pulso genérico $p(t)$, cuja transformada de Fourier é $P(f)$. Denotamos o símbolo do código de linha no instante de tempo k por a_k. Se a taxa de transmissão for $R_b = 1/T_b$ pulsos por segundo, o código de linha gera um trem de pulsos construído a partir do pulso

* Acoplamento ac é necessário porque as rotas de providas por pares de cabo entre estações repetidoras são usadas para transmitir a potência necessária à operação dos repetidores.

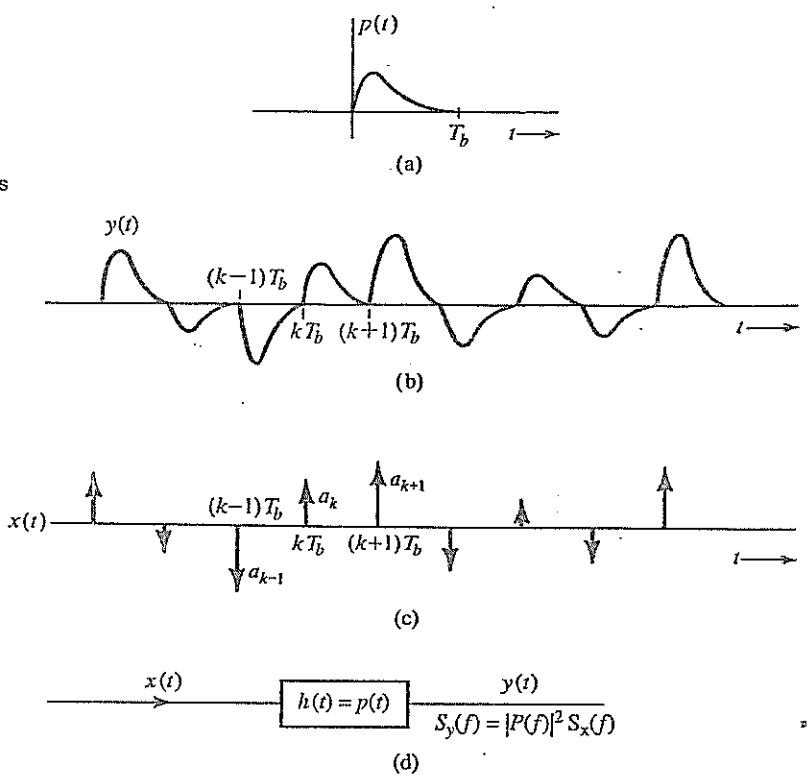


Figura 7.4
Sinal modulado em amplitude por sequência aleatória de pulsos e sua geração por um trem de impulsos PAM.

básico $p(t)$, com amplitude a_k e iniciando no tempo $t = kT_b$; em outras palavras, o k-ésimo símbolo é transmitido como $a_k p(t - kT_b)$. A Fig. 7.4a fornece uma ilustração de um pulso especial $p(t)$, enquanto a Fig. 7.4b mostra o correspondente trem de pulsos gerado pelo codificador de linha na banda base. Como mostrado na Fig. 7.4b, contando uma sucessão de transmissões de símbolos espaçados por T_b segundos, o sinal em banda base é um trem de pulsos da forma

$$y(t) = \sum a_k p(t - kT_b) \tag{7.1}$$

Reparamos que o codificador de linha determina o símbolo $\{a_k\}$ como a amplitude do pulso $p(t - kT_b)$.

Os valores de a_k são aleatórios e dependem da entrada do codificador de linha e do próprio código de linha; $y(t)$ é um sinal modulado em amplitude de pulso (sinal PAM). Os códigos de linha on-off, polar e bipolar, são casos especiais desse trem de pulsos $y(t)$, em que a_k assume os valores 0, 1 ou −1 de modo aleatório, sujeito a algumas condições. Podemos, portanto, analisar muitos códigos de linha segundo a PSD de $y(t)$. Infelizmente, a PSD de $y(t)$ depende de a_k e de $p(t)$. Se a forma do pulso $p(t)$ for alterada, pode ser necessário refazer todo o cálculo da PSD. Essa dificuldade pode ser contornada com o simples artifício de selecionar um sinal PAM $x(t)$ que use um impulso unitário como o pulso básico $p(t)$ (Fig. 7.4c). Os impulsos ocorrem a intervalos T_b e a intensidade (área) do k-ésimo impulso é a_k. Se $x(t)$ for aplicado à entrada de um filtro cuja resposta ao impulso seja $h(t) = p(t)$ (Fig. 7.4d), a saída será o trem de pulsos $y(t)$ na Fig. 7.4b. Além disso, aplicação da Eq. (3.92) permite calcular a PSD de $y(t)$ como

$$S_y(f) = |P(f)|^2 S_x(f)$$

Essa relação nos permite determinar $S_y(f)$, a PSD de um código de linha que corresponde a uma forma de onda $p(t)$ qualquer, uma vez que conheçamos $S_x(f)$. A generalidade dessa abordagem a torna atrativa.

Agora, precisamos determinar $\mathcal{R}_x(\tau)$, a função de autocorrelação temporal do trem de impulsos $x(t)$. Isso pode ser feito com certa facilidade se considerarmos o impulso como uma forma limite de pulsos retangulares, como mostrado na Fig. 7.5a. Cada pulso tem uma largura $\epsilon \to 0$, e a altura do k-ésimo pulso é dada por

$$h_k = \frac{a_k}{\epsilon} \to \infty$$

Com isso, garantimos que a intensidade do k-ésimo impulso seja a_k, ou

$$\epsilon h_k = a_k$$

Se designarmos o correspondente trem de pulsos retangulares por $\hat{x}(t)$, então, por definição [Eq. (3.82) na Seção 3.8],

$$\mathcal{R}_{\hat{x}}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \hat{x}(t)\hat{x}(t-\tau)\, dt \tag{7.2}$$

Como $\mathcal{R}_{\hat{x}}(\tau)$ é uma função par de τ [Eq. (3.83)], basta que consideremos somente valores positivos de τ. Inicialmente, consideremos o caso $\tau < \epsilon$. Nesse caso, a integral na Eq. (7.2) é a área sob o sinal $\hat{x}(t)$ multiplicado por $\hat{x}(t)$ atrasado por τ ($\tau < \epsilon$). Como visto na Fig. 7.5b, a área associada ao k-ésimo pulso é $h_k^2(\epsilon - \tau)$, e

$$\begin{aligned}\mathcal{R}_{\hat{x}} &= \lim_{T\to\infty} \frac{1}{T} \sum_k h_k^2(\epsilon - \tau) \\ &= \lim_{T\to\infty} \frac{1}{T} \sum_k a_k^2 \left(\frac{\epsilon - \tau}{\epsilon^2}\right) \\ &= \frac{R_0}{\epsilon T_b}\left(1 - \frac{\tau}{\epsilon}\right)\end{aligned} \tag{7.3a}$$

em que

$$R_0 = \lim_{T\to\infty} \frac{T_b}{T} \sum_k a_k^2 \tag{7.3b}$$

Durante o intervalo de média T ($T \to \infty$), há N pulsos ($N \to \infty$), em que

$$N = \frac{T}{T_b} \tag{7.4}$$

Da Eq. (7.3b),

$$R_0 = \lim_{N\to\infty} \frac{1}{N} \sum_k a_k^2 \tag{7.5}$$

Observemos que a soma é feita com N pulsos. Portanto, R_0 é a média temporal do quadrado das amplitudes a_k do pulso. Usando a notação de média temporal, podemos expressar R_0 como

$$R_0 = \lim_{N\to\infty} \frac{1}{N} \sum_k a_k^2 = \overline{a_k^2} \tag{7.6}$$

Também sabemos que $\mathcal{R}_{\hat{x}}(\tau)$ é uma função de τ [ver Eq. (3.83)]. Logo, a Eq. (7.3) pode ser expressa como

$$\mathcal{R}_{\hat{x}}(\tau) = \frac{R_0}{\epsilon T_b}\left(1 - \frac{|\tau|}{\epsilon}\right) \qquad |\tau| < \epsilon \tag{7.7}$$

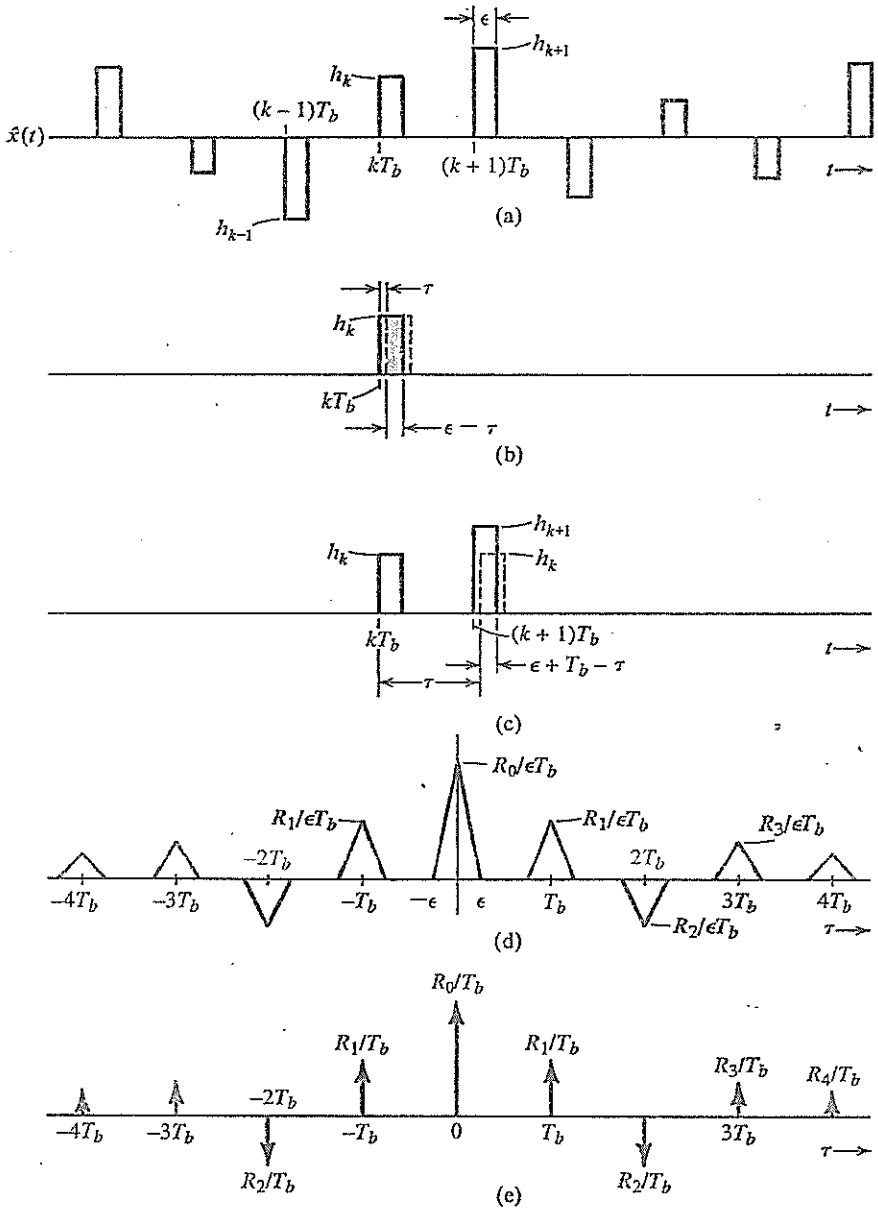


Figura 7.5 Dedução da PSD de um sinal PAM aleatório com largura de pulso muito estreita ϵ e altura $h_k = a_k/\epsilon$.

Este é um pulso retangular de altura $R_0/\epsilon T_b$, largura 2ϵ e centrado em $\tau = 0$ (Fig. 7.5d). Isso é esperado, pois, à medida que τ aumenta além de ϵ, não há sobreposição entre o sinal atrasado $\hat{x}(t-\tau)$ e $\hat{x}(t)$; logo, $\mathcal{R}_{\hat{x}}(\tau) = 0$, como visto na Fig. 7.5d. Contudo, à medida que aumentamos τ ainda mais, observamos que o k-ésimo pulso de $\hat{x}(t-\tau)$ começará a se sobrepor ao $(k+1)$-ésimo pulso de $\hat{x}(t)$, quando τ se aproximar de T_b (Fig. 7.5c). Repetindo esse argumento, vemos que $\mathcal{R}_{\hat{x}}(\tau)$ terá outro pulso retangular de largura 2ϵ centrado em $\tau = T_b$, com altura $R_1/\epsilon T_b$, em que

$$
\begin{aligned}
R_1 &= \lim_{T\to\infty} \frac{T_b}{T} \sum_k a_k a_{k+1} \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_k a_k a_{k+1} \\
&= \overline{a_k a_{k+1}}
\end{aligned}
$$

Observemos que R_1 é obtido da multiplicação de cada intensidade de pulso (a_k) pela intensidade de seu vizinho imediato (a_{k+1}), da soma de todos esses produtos e, então, da divisão pelo número total de pulsos. Isso é claramente a média temporal do produto $a_k a_{k+1}$ e o resultado, em nossa notação, é $\overline{a_k a_{k+1}}$. Algo similar ocorre em torno de $\tau = 2T_b, 3T_b, \ldots$. Assim, $\mathcal{R}_{\hat{x}}(\tau)$ consiste em uma sequência de pulsos triangulares de largura 2ϵ, centrados em $\tau = 0, \pm T_b, \pm 2T_b, \ldots$. A altura dos pulsos centrados em $\pm nT_b$ é R_n/T_b, em que

$$\begin{aligned} R_n &= \lim_{T\to\infty} \frac{T_b}{T} \sum_k a_k a_{k+n} \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_k a_k a_{k+n} \\ &= \overline{a_k a_{k+n}} \end{aligned}$$

R_n é, essencialmente, a função de autocorrelação discreta dos símbolos do código de linha $\{a_k\}$.

Para calcular $\mathcal{R}_x(\tau)$, fazemos $\epsilon \to 0$ em $\mathcal{R}_{\hat{x}}(\tau)$. À medida que $\epsilon \to 0$, a largura de cada pulso triangular $\to 0$ e a altura $\to \infty$, de modo que a área permanece finita. Assim, no limite $\epsilon \to 0$, os pulsos triangulares se tornam impulsos. Para o n-ésimo pulso centrado em nT_b, a altura é $R_n/\epsilon T_b$ e a área, R_n/T_b. Portanto, (Fig. 7.5e),

$$\mathcal{R}_x(\tau) = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} R_n \, \delta(\tau - nT_b) \tag{7.8}$$

A PSD $S_x(f)$ é a transformada de Fourier de $\mathcal{R}_x(\tau)$. Logo,

$$S_x(f) = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} R_n e^{-jn2\pi f T_b} \tag{7.9}$$

Lembrando que $R_{-n} = R_n$, [pois $\mathcal{R}(\tau)$ é uma função para de τ], temos

$$S_x(f) = \frac{1}{T_b} \left[R_o + 2 \sum_{n=1}^{\infty} R_n \cos n2\pi f T_b \right] \tag{7.10}$$

A entrada $x(t)$ do filtro com resposta ao impulso $h(t) = p(t)$ resulta na saída $y(t)$, como mostrado na Fig. 7.4d. Se $p(t) \Longleftrightarrow P(f)$, a função de transferência do filtro é $H(f) = P(f)$ e, segundo a Eq. (3.91),

$$S_y(f) = |P(f)|^2 S_x(f) \tag{7.11a}$$

$$= \frac{|P(f)|^2}{T_b} \left[\sum_{n=-\infty}^{\infty} R_n e^{-jn2\pi f T_b} \right] \tag{7.11b}$$

$$= \frac{|P(f)|^2}{T_b} \left[R_o + 2 \sum_{n=1}^{\infty} R_n \cos n2\pi f T_b \right] \tag{7.11c}$$

Portanto, a PSD de um código de linha é totalmente caracterizada pelo correspondente R_n e pela seleção de forma de pulso de $P(f)$. A seguir, usaremos esse resultado geral para determinar as PSDs de vários códigos de linha específicos, a partir da determinação da autocorrelação entre símbolos.

7.2.2 Sinalização Polar

Na sinalização polar, 1 é transmitido por um pulso $p(t)$ e 0, por um pulso $-p(t)$. Nesse caso, a_k tem igual probabilidade de ser 1 ou -1, e a_k^2 será sempre 1. Assim,

$$R_0 = \lim_{N\to\infty} \frac{1}{N} \sum_k a_k^2$$

Há N pulsos, e $a_k^2 = 1$ para cada um, de modo que a soma no lado direito da equação anterior é N. Logo,

$$R_0 = \lim_{N\to\infty} \frac{1}{N}(N) = 1 \tag{7.12a}$$

Além disso, tanto a_k como a_{k+1} são 1 ou −1. Logo, $a_k a_{k+1}$ é 1 ou −1. Como a amplitude a_k tem, em média, igual probabilidade de ser 1 e −1, considerando N termos, o produto $a_k a_{k+1}$ é igual a 1 para $N/2$ termos e igual a −1 para os restantes $N/2$ termos. Portanto,

Possíveis valores de $a_k a_{k+1}$

a_{k+1} \ a_k	−1	+1
−1	1	−1
+1	−1	1

$$R_1 = \lim_{N\to\infty} \frac{1}{N}\left[\frac{N}{2}(1) + \frac{N}{2}(-1)\right] = 0 \tag{7.12b}$$

Com esse raciocínio, podemos ver que o produto $a_k a_{k+n}$ também tem igual probabilidade de ser 1 ou −1. Assim,

$$R_n = 0 \qquad n \geq 1 \tag{7.12c}$$

Portanto, da Eq. (7.11c),

$$\begin{aligned} S_y(f) &= \frac{|P(f)|^2}{T_b} R_0 \\ &= \frac{|P(f)|^2}{T_b} \end{aligned} \tag{7.13}$$

Para efeitos de comparação de vários esquemas, consideremos uma forma de pulso *específica*. Seja $p(t)$ um pulso retangular de largura $T_b/2$ (pulso retangular de meia largura), ou seja,

$$p(t) = \Pi\left(\frac{t}{T_b/2}\right) = \Pi\left(\frac{2t}{T_b}\right)$$

e

$$P(f) = \frac{T_b}{2}\,\mathrm{sinc}\left(\frac{\pi f T_b}{2}\right) \tag{7.14}$$

Portanto,

$$S_y(f) = \frac{T_b}{4}\,\mathrm{sinc}^2\left(\frac{\pi f T_b}{2}\right) \tag{7.15}$$

A Fig. 7.6 mostra o espectro $S_y(f)$. Fica claro que a maior parte da potência do sinal polar se concentra nas frequências mais baixas. Teoricamente, o espectro se torna muito estreito à medida que a frequência aumenta, mas sua largura jamais iguala zero acima de uma dada frequência. Para termos uma medida representativa de largura de banda, consideramos a *primeira frequência de nulo não dc* como a **largura de banda essencial.***

*A rigor, a primeira frequência de nulo acima de dc nem sempre é uma boa medida da largura de banda de um sinal. A quantidade de sinal contida no lóbulo principal (primeiro) da PSD é que determina se o primeiro nulo não dc será uma medida significativa de largura de banda, como veremos mais adiante, ao compararmos a PSD de vários códigos de linha (Fig. 7.9). Na maioria dos casos práticos, essa aproximação é aceitável para os códigos de linha e formas de pulsos usados comumente.

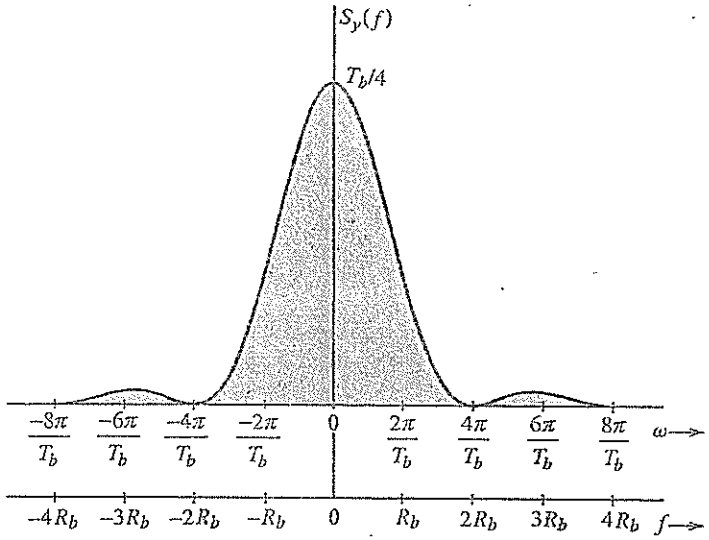


Figura 7.6 Densidade espectral de potência de um sinal polar.

Do espectro do sinal polar, a largura de banda essencial do sinal é calculada como $2R_b$ Hz (em que R_b é a frequência do relógio). Isso é 4 vezes a largura de banda teórica (largura de banda de Nyquist) exigida para transmitir R_b pulsos por segundo. O aumento da largura do pulso reduz a largura de banda (expansão no domínio do tempo resulta em compressão no domínio da frequência). Para um pulso de largura completa* (máxima largura de pulso possível), a largura de banda essencial é a metade, ou seja, R_b Hz. Esta, no entanto, ainda é o dobro da largura de banda teórica. Assim, sinalização polar não é a mais eficiente do ponto de vista de largura de banda.

Uma segunda deficiência da sinalização polar é não ter capacidade de detecção ou correção de erro. Outra desvantagem da sinalização polar é ter PSD não nula em dc ($f = 0$). Isso impede o uso de acoplamento ac durante a transmissão. O modo de acoplamento ac é muito importante na prática, pois permite que transformadores e capacitores de bloqueio auxiliem no casamento de impedância e na remoção de polarização, assim como a alimentação de potência dos repetidores em linha por meio dos pares de cabo. Mais adiante, mostraremos como a PSD de um código de linha pode ser forçada a zero em dc por meio de escolha adequada da forma de $p(t)$.

Quanto aos aspectos positivos, a sinalização polar é o esquema mais eficiente do ponto de vista de consumo de potência. Para uma dada potência, podemos mostrar que a probabilidade de detecção de erro em um esquema polar é a menor dentre todas as técnicas de sinalização (Capítulo 10). A sinalização polar também é transparente, pois sempre há algum pulso (positivo ou negativo), qualquer que seja a sequência de bits. Não há uma componente discreta do relógio no espectro do sinal polar. Contudo, a retificação do sinal polar *RZ* produz um sinal periódico na frequência do relógio e pode ser usado prontamente para extrair a temporização.

7.2.3 Construção de um Nulo DC na PSD via Formatação de Pulso

Como $S_y(f)$, a PSD de um código de linha, contém um fator $|P(f)|^2$, podemos forçar a PSD a ter um nulo dc se selecionarmos um pulso $p(t)$ tal que $P(f)$ seja zero em dc ($f = 0$). Como

$$P(f) = \int_{-\infty}^{\infty} p(t)e^{-j2\pi ft}\,dt$$

*O esquema que usa pulso de largura completa $p(t) = \Pi(t/T_b)$ é um exemplo de um esquema sem retorno ao zero (NRZ). O esquema de meia largura de pulso, por sua vez, é um exemplo de um esquema com retorno ao zero (RZ).

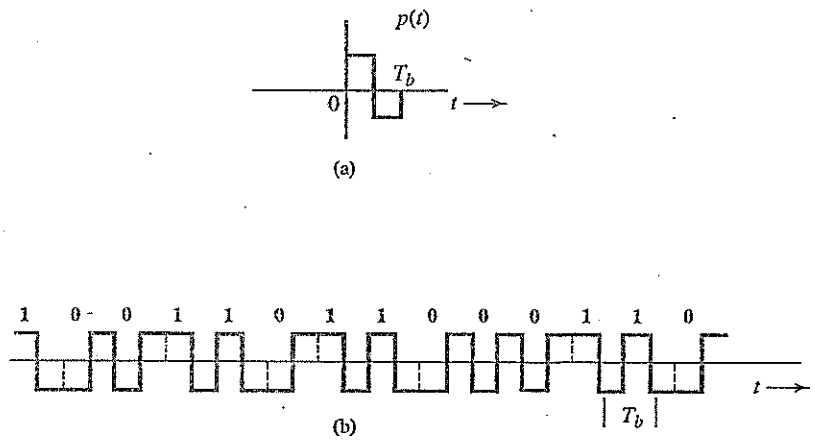


Figura 7.7 Sinal bifásico (Manchester ou de fase alternada). (a) Pulso básico para sinalização Manchester. (b) Forma de onda transmitida para sequência binária de dados usando sinalização Manchester.

temos

$$P(0) = \int_{-\infty}^{\infty} p(t)\, dt$$

Portanto, se a área sob $p(t)$ for feita igual a zero, $P(0)$ será zero, e teremos um nulo dc na PSD. Para um pulso retangular, uma possível forma de $p(t)$ para obter este efeito é mostrada na Fig. 7.7a. Quando usamos esse pulso com codificação de linha polar, o sinal resultante é conhecido como sinal em **código Manchester, bifásico,** ou de **fase alternada.** O leitor pode usar a Eq. (7.13) para mostrar que, para esse pulso, a PSD do código de linha Manchester tem um nulo dc (Exercício 7.2-2).

7.2.4 Sinalização On-Off

Na sinalização on-off, um **1** é transmitido por um pulso $p(t)$ e um **0**, pela ausência de pulso. Assim, a amplitude de pulso a_k tem igual probabilidade de ser 1 ou 0. Dentre N pulsos no intervalo de T segundos, a_k é 1 para $N/2$ pulsos e 0 para os restantes $N/2$ pulsos, em média. Portanto,

$$R_0 = \lim_{N\to\infty} \frac{1}{N}\left[\frac{N}{2}(1)^2 + \frac{N}{2}(0)^2\right] = \frac{1}{2} \tag{7.16}$$

Para calcular o valor de R_n, precisamos considerar o produto $a_k a_{k+n}$. Como a_k e a_{k+n} têm igual probabilidade de serem 1 ou 0, o produto $a_k a_{k+n}$ tem igual probabilidade de ser 1×1, 1×0, 0×1 ou 0×0, ou seja, 1, 0, 0, 0. Portanto, em média, o produto $a_k a_{k+n}$ é igual a 1 para $N/4$ termos e 0 para $3N/4$ termos, e

$$R_n = \lim_{N\to\infty} \frac{1}{N}\left[\frac{N}{4}(1) + \frac{3N}{4}(0)\right] = \frac{1}{4} \qquad n \geq 1 \tag{7.17}$$

Portanto [Eq. (7.9)],

$$S_x(f) = \frac{1}{2T_b} + \frac{1}{4T_b} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} e^{-jn2\pi f T_b} \tag{7.18a}$$

$$= \frac{1}{4T_b} + \frac{1}{4T_b} \sum_{n=-\infty}^{\infty} e^{-jn2\pi f T_b} \tag{7.18b}$$

A Eq. (7.18b) é obtida da Eq. (7.18a) dividindo em dois o termo $1/2T_b$ correspondente a R_0: $1/4T_b$ fora do somatório e $1/4T_b$ dentro do somatório (correspondendo a $n = 0$). Agora, usamos a fórmula (ver a nota de rodapé para a prova*)

$$\sum_{n=-\infty}^{\infty} e^{-jn2\pi fT_b} = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)$$

A substituição deste resultado na Eq. (7.18b) leva a

$$S_x(f) = \frac{1}{4T_b} + \frac{1}{4T_b^2} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right) \tag{7.19a}$$

e à desejada PSD da forma de onda on-off $y(t)$ [da Eq. (7.11a)]

$$S_y(f) = \frac{|P(f)|^2}{4T_b}\left[1 + \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)\right] \tag{7.19b}$$

Reparemos que, ao contrário do espectro contínuo da PSD da sinalização polar, a PSD do sinal on-off na Eq. (7.19b) tem uma parte discreta adicional, que pode ser anulada se a forma do pulso for escolhida de modo que

$$P\left(\frac{n}{T_b}\right) = 0 \qquad n = 0, \pm 1, \ldots$$

Tomando como exemplo o caso de um pulso retangular de meia largura [Eq. (7.14)],

$$S_y(f) = \frac{T_b}{16} \operatorname{sinc}^2\left(\frac{\pi f T_b}{2}\right)\left[1 + \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)\right] \tag{7.20}$$

A resultante PSD é mostrada na Fig. 7.8. A componente contínua do espectro é $(T_b/16)\operatorname{sinc}^2(\pi f T_b/2)$. Isso é igual (exceto por um fator de escala) ao espectro do sinal polar [Eq. (7.15)]. A componente discreta é representada pelo produto de um trem de impulsos pela componente contínua $(T_b/16)\operatorname{sinc}^2(\pi f T_b/2)$. Assim, essa componente aparece como impulsos periódicos, tendo a componente contínua como envelope. Além disso, a frequência de repetição dos impulsos é a frequência do relógio $R_b = 1/T_b$, pois sua frequência fundamental é $2\pi/T_b$ rad/s ou $1/T_b$ Hz. Esse é um resultado lógico, pois, como mostra a Fig. 7.3, um sinal on-off pode ser expresso como a soma de uma componente polar com uma componente periódica. A componente polar $y_1(t)$ é exatamente a metade do sinal polar discutido anteriormente. Logo, a PSD dessa componente é um quarto da PSD na Eq. (7.15). A componente periódica tem a frequência do relógio R_b, e consiste em componentes discretas de frequência R_b e seus harmônicos.

---

*O trem de impulsos na Fig. 3.23a do Exemplo 3.11 é $\delta_{T_b}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_b)$. Além disso, a série de Fourier para esse trem de impulsos é obtida da Eq. (2.67) como

$$\sum_{n=-\infty}^{\infty} \delta(t - nT_b) = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} e^{jn2\pi R_b t} \qquad R_b = \frac{1}{T_b}$$

Tomamos a transformada de Fourier dos dois lados dessa equação, e usamos o fato de que $\delta(t - nT_b) \Longleftrightarrow e^{-jn2\pi fT_b}$ e $e^{jn2\pi R_b t} \Longleftrightarrow \delta(f - nR_b)$. Isso leva a

$$\sum_{n=-\infty}^{\infty} e^{-jn2\pi fT_b} = \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T_b}\right)$$

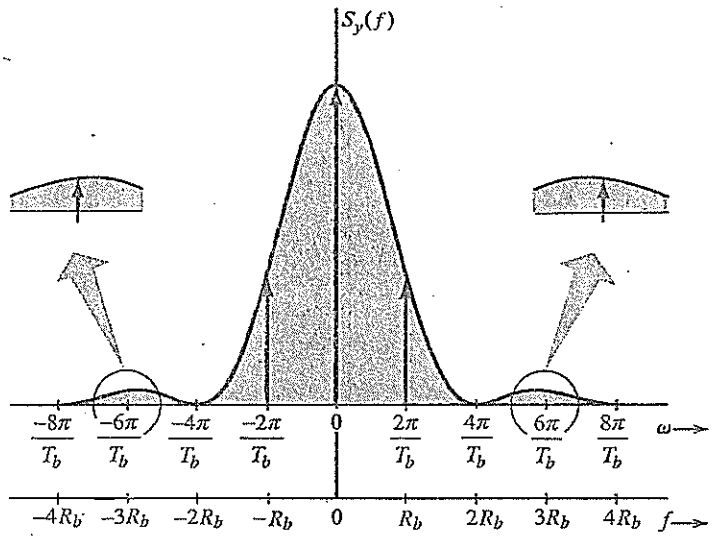


Figura 7.8 Densidade espectral de potência (PSD) de um sinal on-off.

A sinalização on-off não tem muito do que se gabar. Para uma dada potência transmitida, tem menos imunidade ao ruído do que o esquema polar, que usa um pulso positivo para **1** e um pulso negativo para **0**. Isso decorre do fato de que a imunidade ao ruído depende da diferença entre as amplitudes que representam **1** e **0**. Portanto, para a mesma imunidade, se a sinalização on-off usar pulsos de amplitude 2 e 0, a sinalização polar precisa apenas de pulsos com amplitudes 1 e −1. É simples mostrar que a sinalização on-off requer o dobro da potência da sinalização polar. Se um pulso de amplitude 1 ou −1 tiver energia E, o pulso de amplitude 2 tem energia $(2)^2E = 4E$. Como $1/T_b$ dígitos são transmitidos por segundo, a potência do sinal polar é $(E)(1/T_b) = E/T_b$. Para o caso on-off, a energia de cada pulso é $4E$, embora, em média, um pulso seja transmitido na metade do tempo, enquanto nada é transmitido na outra metade. Assim, a potência média de sinal on-off é

$$\frac{1}{T_b}\left(4E\frac{1}{2} + 0 \cdot \frac{1}{2}\right) = \frac{2E}{T_b}$$

que é o dobro da potência requerida pela sinalização polar. Além disso, ao contrário do caso polar, a sinalização on-off não é transparente. Uma longa sequência de **0**s (ou offs) causa ausência de sinal e pode levar a erros na extração da temporização. Adicionalmente, todas as desvantagens da sinalização polar (por exemplo, excessiva largura de banda de transmissão, espectro de potência não nulo em dc, incapacidade de detecção ou correção de erros) também estão presentes na sinalização on-off.

7.2.5 Sinalização Bipolar

O esquema de sinalização usado em PCM para redes de telefonia é denominado bipolar (pseudoternário ou inversão alternada de sinal). Um **0** é transmitido pela ausência de pulso e um **1**, por um pulso $p(t)$ ou $-p(t)$, dependendo se o **1** anterior foi transmitido por $-p(t)$ ou $p(t)$. Com pulsos consecutivos alternados, podemos evitar ondulação dc e, assim, causar um nulo na PSD. A sinalização bipolar, na verdade, usa três símbolos [$p(t)$, 0 e $-p(t)$] e, portanto, é uma sinalização ternária, e não binária.

Para calcular a PSD, temos

$$R_o = \lim_{N\to\infty} \frac{1}{N} \sum_k a_k^2$$

Em média, metade dos a_ks são 0 e a outra metade é 1 ou −1, com $a_k^2 = 1$. Logo,

$$R_o = \lim_{N\to\infty} \frac{1}{N}\left[\frac{N}{2}(\pm 1)^2 + \frac{N}{2}(0)^2\right] = \frac{1}{2}$$

Para calcular R_1, consideremos o produto das amplitudes de pulsos $a_k a_{k+1}$. Quatro sequências de dois bits têm igual probabilidade de ocorrência: **11, 10, 01, 00**. Como o bit **0** é codificado por ausência de pulso ($a_k = 0$), o produto $a_k a_{k+1}$ é zero para as três últimas sequências. Isso significa, em média, que $3N/4$ combinações têm $a_k a_{k+1} = 0$ e apenas $N/4$ combinações têm $a_k a_{k+1}$ não nulo. Devido à regra bipolar, a sequência de bits **11** pode ser codificada somente por pulsos de polaridades opostas. Isso significa que o produto $a_k a_{k+1} = -1$ para $N/4$ combinações. Portanto,

$$R_1 = \lim_{N\to\infty} \frac{1}{N}\left[\frac{N}{4}(-1) + \frac{3N}{4}(0)\right] = -\frac{1}{4}$$

Para calcular R_2 de modo similar, devemos analisar o produto $a_k a_{k+2}$. Para isso, precisamos considerar todas as possíveis combinações de três bits em sequência. Há oito combinações com igual probabilidade de ocorrência: **111, 101, 110, 100, 011, 010, 001, 000**. As últimas seis combinações têm 0 como o primeiro ou o último bit; logo, $a_k a_{k+2} = 0$ para todas essas seis combinações. As duas primeiras são as únicas combinações que têm $a_k a_{k+2}$ não nulo. Dada a regra bipolar, o primeiro e o último bits na combinação **111** têm a mesma polaridade, resultando em $a_k a_{k+2} = 1$. Contudo, em **101**, o primeiro e o último bits têm polaridades opostas, resultando em $a_k a_{k+2} = -1$. Assim, em média, $a_k a_{k+2} = 1$ para $N/8$ termos, −1 para $N/8$ termos e 0 para $3N/4$ termos. Logo,

$$R_2 = \lim_{N\to\infty} \frac{1}{N}\left[\frac{N}{8}(1) + \frac{N}{8}(-1) + \frac{3N}{8}(0)\right] = 0$$

Em geral,

$$R_n = \lim_{N\to\infty} \frac{1}{N}\sum_k a_k a_{k+n}$$

Para $n > 2$, o produto $a_k a_{k+n}$ pode ser 1, −1 ou 0. Além disso, um número igual de combinações tem valores 1 e −1. Isso causa $R_n = 0$. Portanto,

$$R_n = 0 \qquad n > 1$$

e [ver Eq. (7.11c)]

$$S_y(f) = \frac{|P(f)|^2}{2T_b}[1 - \cos 2\pi f T_b] \tag{7.21a}$$

$$= \frac{|P(f)|^2}{T_b}\,\text{sen}^2\,(\pi f T_b) \tag{7.21b}$$

Reparemos que $S_y(f) = 0$ para $f = 0$ (dc), qualquer que seja $P(f)$. Logo, a PSD tem um nulo dc, o que é desejável para acoplamento ac. Além disso, $\text{sen}^2\,(\pi f T_b) = 0$ em $f = 1/T_b$, ou seja, em $f = 1/T_b = R_b$ Hz. Portanto, independentemente de $P(f)$, fica garantida a largura de banda R_b Hz para o primeiro nulo não dc. Para o caso de pulsos de meia largura,

$$S_y(f) = \frac{T_b}{4}\,\text{sinc}^2\left(\frac{\pi f T_b}{2}\right)\,\text{sen}^2\,(\pi f T_b) \tag{7.22}$$

Isso é ilustrado na Fig. 7.9. A largura de banda essencial do sinal é R_b ($R_b = 1/T_b$), que é a metade da largura de banda do esquema polar que usa o mesmo pulso de meia largura ou da sinalização on-off e o dobro da mínima largura de banda teórica. Observamos que pudemos calcular a largura de banda R_b para o caso polar (ou on-off) com pulso de largura completa. Para o caso bipolar, a largura de banda é R_b Hz, tanto para pulso de meia largura como de largura completa.

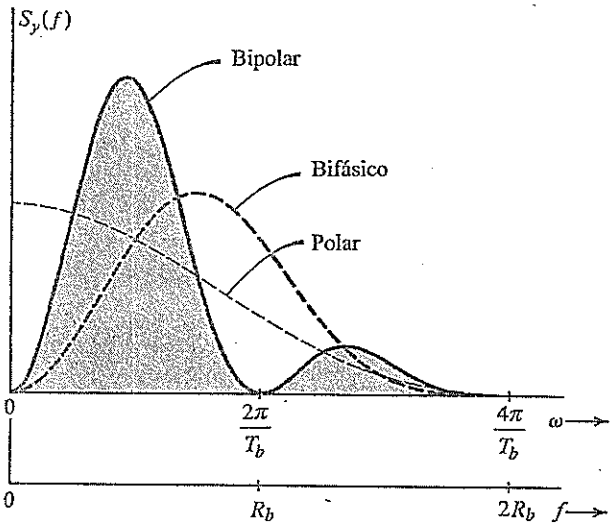


Figura 7.9 PSD de sinais bipolar, polar e bifásico normalizados para iguais potências. Pulsos retangulares de meia largura são usados.

A sinalização bipolar tem diversas vantagens: (1) seu espectro tem um nulo dc; (2) sua largura de banda não é excessiva, (3) tem capacidade de detecção de erro isolado. Isso se deve ao fato de que mesmo um erro isolado viola a regra de pulsos alternados e será detectado imediatamente. Se um sinal bipolar for retificado, obtemos um sinal on-off que tem uma componente discreta na frequência do relógio. Entre as desvantagens de um sinal bipolar está a exigência do dobro da potência (3 dB) usada por um sinal polar. Isso decorre do fato de a detecção ser essencialmente equivalente à da sinalização on-off, sob o ponto de vista de detecção: a distinção é entre $+p(t)$ ou $-p(t)$ e 0, e não entre $\pm p(t)$.

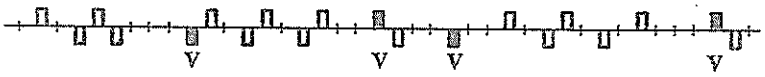
Outra desvantagem da sinalização bipolar é não ser transparente. Na prática, vários esquemas substitutos foram usados para evitar que longas de zeros lógicos permitissem que sinais de relógio se desviassem. A seguir, discutiremos dois desses esquemas.

Figura 7.10 (a) Sinal HDB3 e (b) sua PSD.

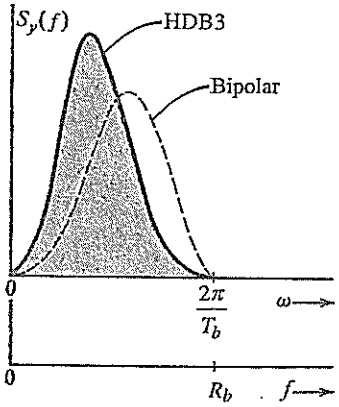
Dígitos de entrada **0 1 0 1 1 1 0 0 0 0 1 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 1 0 1 0 1 0 0 0 0 1**

Dígitos codificados **0 1 0 1 1 1 [0 0 0 V] 1 0 1 1 0 1 [1 0 0 V] [1 0 0 V] 0 0 1 0 1 1 0 1 0 1 [0 0 0 V] 1**

Forma de onda transmitida



(a)



(b)

Sinalização Bipolar de Alta Densidade (HDB)

O esquema HDB (*high-density bipolar*) é um padrão ITU-T (antigo CCITT – Comitê Consultivo Internacional de Telefonia e Telegrafia). Nesse esquema, o problema de falta de transparência da sinalização bipolar é eliminado com a adição de pulsos quando o número de pulsos consecutivos **0**s ultrapassa N. Esse código modificado é denominado **codificação bipolar de alta densidade** (HDBN), em que N pode assumir qualquer valor 1, 2, 3, ... O código HDB mais importante é o formato HDB3, adotado como padrão internacional.

A ideia básica do código HDBN é que, na ocorrência de uma sequência de $N+1$ zeros, esse grupo de zeros é substituído por uma das sequências especiais de $N+1$ dígitos binários. Para aumentar o conteúdo de temporização do sinal, as sequências são escolhidas de modo a incluir alguns **1**s binários. Os **1**s incluídos deliberadamente violam a regra bipolar para facilitar a identificação da sequência subsequente. Na codificação HDB3, por exemplo, as sequências especiais usadas são **000V** e **B00V**, em que **B** = **1**, que respeita a regra bipolar, e **V** = **1**, que viola a regra bipolar. A escolha da sequência **000V** ou **B00V** é feita de forma que os pulsos **V** consecutivos têm sinais alternados para evitar ondulação dc e manter o nulo dc na PSD. Isso requer que a sequência **B00V** seja usada quando há um número par de **1**s após a última sequência especial, e a sequência **000V** é usada quando há um número ímpar de **1**s após a última sequência. A Fig. 7.10a mostra um exemplo desse código. Reparemos que, na sequência **B00V**, **B** e **V** são codificados pelo mesmo pulso. O decodificador deve verificar duas coisas: as violações bipolares e o número de **0**s que antecede cada violação, para determinar se o **1** anterior também era uma substituição.

Apesar das deliberadas violações da regra bipolar, a sinalização HDB mantém a capacidade de detecção de erro. Qualquer erro isolado inserirá uma violação bipolar espúria (ou removerá uma das violações deliberadas). Isso se tornará aparente quando, na próxima violação, a alternância de violações não desaparecer. Isso também mostra que violações deliberadas podem ser detectadas, apesar de erros isolados. A Fig. 7.10b mostra a PSD de HDB3 juntamente com a de um sinal bipolar, para facilitar a comparação.[3]

Sinalização Binária com Substituição de N Zeros (BNZS)

Uma classe de códigos de linha similar a HDBN é a **binária com substituição de N zeros**, ou BNZS (*binary with N zero substitution*). Nesse código, se ocorrerem N zeros em sequência, estes são substituídos por uma das duas sequências especiais que contêm alguns **1**s para aumentar o conteúdo de temporização. Há violações deliberadas da regra bipolar, assim como em HDBN. A codificação binária com substituição de oito zeros (B8ZS) é usada em sinais DS1 da hierarquia de telefonia digital, discutidos no Capítulo 6. Essa codificação substitui quaisquer sequências de oito zeros por uma sequência de uns e zeros que contém duas violações bipolares. É improvável que uma sequência desse tipo seja afetada por erros, e qualquer uma que seja recebida é substituída por uma sequência de oito zeros antes da decodificação. A sequência empregada como substituição consiste nos padrões **000VB0VB**. Do mesmo modo, no código B6ZS usado em sinais DS2, uma sequência de seis zeros é substituída por **0VB0VB**; um código B3ZS é aplicado a sinais DS3. O código B3ZS é um pouco mais complexo que os outros, pois usa **B0V** ou **00V**, sendo a escolha feita de modo que o número de pulsos **B** entre pulsos **V** consecutivos seja ímpar. Esses códigos BNZS, com $N = 3$, 6 ou 8, envolvem violações bipolares e, portanto, devem ser cuidadosamente substituídos pelas correspondentes sequências de zero no receptor.

Há muitos outros códigos (de linha) de transmissão, em número demasiadamente grande para serem listados aqui. Uma lista de códigos e referências apropriadas pode ser encontrada em Bylanski e Ingram.[3]

7.3 FORMATAÇÃO DE PULSO

A PSD $S_y(f)$ de um sinal digital $y(t)$ pode ser controlada por uma escolha do código de linha ou por $P(f)$, a forma do pulso. Na última seção, discutimos como a PSD é controlada por um código de linha.

Na presente seção, examinaremos como $S_y(f)$ é influenciada pela forma do pulso $p(t)$, e aprenderemos como formatar um pulso $p(t)$ para obter uma $S_y(f)$ desejada. A PSD $S_y(f)$ é forte e diretamente influenciada pela forma do pulso $p(t)$, pois $S_y(f)$ contém o termo $|P(f)|^2$. Assim, em comparação com a natureza do código de linha, a forma do pulso é um fator mais direto e poderoso para formatar a PSD $S_y(f)$.

7.3.1 Interferências Intersimbólicas (ISI) e Efeitos

Na última seção, por conveniência, usamos um simples pulso retangular de meia largura $p(t)$. A rigor, nesse caso, a largura de banda $S_y(f)$ é infinita, pois $P(f)$ tem largura de banda infinita. Contudo, vimos que a largura de banda essencial de $S_y(f)$ era finita. Por exemplo, a maior parte da potência de um sinal bipolar está contida na largura de banda essencial entre 0 e R_b Hz. Notemos, no entanto, que a PSD é pequena, mas não nula, no intervalo $f > R_b$ Hz. Portanto, quando um sinal desse tipo é transmitido em um canal de largura de banda R_b Hz, uma porção significativa de seu espectro é transmitida, enquanto uma parcela pequena é suprimida. Nas Seções 3.5 e 3.6, vimos que uma distorção espectral como essa tendia a espalhar o pulso (dispersão). Espalhamento de um pulso além da janela de tempo T_b a ele alocada causará interferência com pulsos vizinhos. Isso é conhecido como **interferência intersimbólica** ou **ISI** (*intersymbol interference*).

ISI *não* é ruído. ISI é causada por canais não ideais que não são livres de distorção em toda a largura de banda do sinal. No caso de um pulso retangular de meia largura, a largura de banda do sinal é, estritamente falando, infinita. A ISI, como uma manifestação de distorção do canal, pode causar erros na detecção do pulso, caso seja suficientemente grande.

Para resolver a dificuldade associada à ISI, redefinamos nosso problema. Precisamos transmitir um pulso a cada T_b segundos, sendo $a_k p(t - kT_b)$ o k-ésimo pulso. O canal tem largura de banda finita, e devemos detectar a amplitude de pulso a_k corretamente (ou seja, sem ISI). Em nossa discussão até aqui, consideramos pulsos limitados no tempo. Como tais pulsos não podem ser limitados em frequência, parte de seus espectros é suprimida por um canal de largura de banda finita. Isso causa distorção dos pulsos (que se espalham) e, em consequência, a ISI. Podemos, inicialmente, tentar resolver essa dificuldade usando pulsos que tenham largura de banda finita, de modo que possam ser transmitidos intactos pelo canal de banda finita. No entanto, pulsos limitados em frequência não podem ser limitados no tempo. Obviamente, vários pulsos se sobreporão e causarão ISI. Portanto, quer usemos pulsos limitados no tempo ou pulsos limitados em frequência, parece que a ISI não pode ser evitada: é inerente à transmissão em largura de banda finita. Por sorte, existe uma escapatória desse beco sem saída. Amplitudes de pulso podem ser detectadas corretamente apesar de espalhamento (ou sobreposição) dos pulsos, desde que não haja ISI nos instantes de tomada de decisão. Isso pode ser alcançado por meio de pulsos limitados em frequência com forma apropriada. Para eliminar ISI, Nyquist propôs três diferentes critérios para formatação de pulsos,[4] sendo permitida sobreposição de pulsos. Todavia, os pulsos são formatados de modo a causar interferência zero (ou controlada) em todos os outros pulsos nos instantes de tomada de decisão. Assim, ao limitar a exigência de não interferência somente nos instantes de tomada de decisão, eliminamos a necessidade de não sobreposição total dos pulsos. Consideraremos apenas os dois primeiros critérios. O terceiro é menos útil,[5] e, por conseguinte, não será analisado aqui.

7.3.2 Primeiro Critério de Nyquist para ISI Nula

No primeiro método, Nyquist obtém ISI zero ao escolher uma forma de pulso que tenha amplitude não nula no centro (digamos, em $t = 0$) e amplitudes nulas em $t = \pm nT_b$ ($n = 1, 2, 3, \ldots$), em que T_b é a separação entre pulsos transmitidos sucessivos (Fig. 7.11a). Assim,

$$p(t) = \begin{cases} 1 & t = 0 \\ 0 & t = \pm nT_b \end{cases} \qquad \left(T_b = \frac{1}{R_b}\right) \tag{7.23}$$

Um pulso que satisfaça esse critério causa ISI zero nos centros de todos os outros pulsos, ou instantes de sinalização, como mostrado na Fig. 7.11a, na qual ilustramos vários pulsos sucessivos (linha tra-

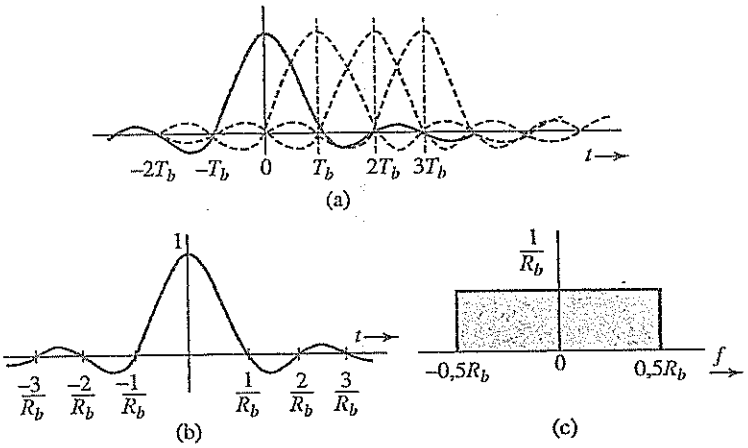


Figura 7.11 Pulso com mínima largura de banda que satisfaz o primeiro critério de Nyquist e seu espectro.

cejada), centrados em $t = 0, T_b, 2T_b, 3T_b, \ldots$ ($T_b = 1/R_b$). Por conveniência, mostramos todos os pulsos como positivos.* Fica claro que, desta figura, as amostras em $t = 0, T_b, 2T_b, 3T_b, \ldots$ consistem na amplitude de apenas um pulso (centrado no instante de amostragem), sem interferência dos outros pulsos.

A transmissão de R_b bit/s requer uma mínima largura de banda teórica de $R_b/2$ Hz. Seria muito bom se a mínima largura de banda de um pulso que satisfizesse o critério de Nyquist fosse $R_b/2$ Hz. Será possível determinar um pulso $p(t)$ desse tipo? Já resolvemos esse problema (Exemplo 6.1 com $B = R_b/2$), e mostramos que existe um (e somente um) pulso que atende o critério de Nyquist (7.23) e tem largura de banda de $R_b/2$ Hz. Esse pulso $p(t) = \text{sinc}\,(\pi R_b t)$ (Fig. 7.11b) tem a propriedade

$$\text{sinc}\,(\pi R_b t) = \begin{cases} 1 & t = 0 \\ 0 & t = \pm n T_b \end{cases} \qquad \left(T_b = \frac{1}{R_b}\right) \tag{7.24a}$$

Além disso, a transformada de Fourier desse pulso é

$$P(f) = \frac{1}{R_b}\Pi\left(\frac{f}{R_b}\right) \tag{7.24b}$$

que tem largura de banda $R_b/2$ Hz, como visto na Fig. 7.11c. Podemos usar esse pulso para transmitir uma taxa de R_b pulsos por segundo sem ISI, em uma largura de banda de apenas $R_b/2$.

Esse esquema mostra que podemos alcançar o limite teórico de desempenho se usarmos um pulso sinc. Lamentavelmente, esse pulso é impraticável, pois tem início em $-\infty$. Teremos de esperar um tempo infinito para gerá-lo. Qualquer tentativa de truncá-lo aumentaria sua largura de banda além dos $R_b/2$ Hz. Contudo, mesmo se esse pulso fosse realizável, teria uma característica indesejável: decairia a uma taxa demasiadamente lenta de $1/t$. Isto causaria alguns sérios problemas práticos. Por exemplo, se a taxa de dados nominal R_b bit/s requerida para esse esquema variasse um pouco, a amplitude de um pulso não se anularia nos centros dos outros pulsos. Como os pulsos decaem com $1/t$ apenas, a interferência cumulativa no centro de um dado pulso devido a todos os outros seria da forma $\sum(1/n)$. É um fato bem conhecido que essa série infinita não converge e pode produzir um valor muito alto. Um resultado similar ocorreria se tudo fosse perfeito no transmissor, mas a taxa de amostragem no receptor se desvia da taxa de R_b Hz. A mesma coisa aconteceria se os instantes de amostragem se desviassem um pouco devido à incerteza temporal nos pulsos, o que é inevitável mesmo nos sistemas mais sofisticados. Portanto, esse esquema falhará, a menos que tudo seja perfeito, o que é uma impossibi-

*Na verdade, um pulso correspondendo a 0 seria negativo. Contudo, considerar todos os pulsos como positivos não afeta o raciocínio. Mostrar pulsos negativos tornaria a figura desnecessariamente confusa.

lidade prática. E tudo isso decorre do fato de sinc $(\pi R_b t)$ decair de modo demasiadamente lento (com $1/t$). A solução consiste em encontrar um pulso $p(t)$ que satisfaça a Eq. (7.23), mas decaia mais rapidamente que $1/t$. Nyquist mostrou que um pulso como esse requer uma largura de banda $kR_b/2$, com $1 \leq k \leq 2$.

Isso pode ser provado da seguinte forma: consideremos $p(t) \Longleftrightarrow P(f)$, em que a largura de banda de $P(f)$ está no intervalo $(R_b/2, R_b)$ (Fig. 7.12a). O desejado pulso $p(t)$ satisfaz a Eq. (7.23). Se amostrarmos $p(t)$ a cada T_b segundos, multiplicando-o por $\delta_{T_b}(t)$ (um trem de impulsos), devido à propriedade (7.23), todas as amostras, exceto a feita na origem, serão iguais a zero. Assim, o sinal amostrado $\overline{p}(t)$ será

$$\overline{p}(t) = p(t)\delta_{T_b}(t) = \delta(t) \tag{7.25}$$

Pela análise da Eq. (6.4) do Capítulo 6, sabemos que o espectro de um sinal amostrado $\overline{p}(t)$ é $(1/T_b$ vezes) o espectro de $p(t)$ repetido periodicamente à frequência de amostragem R_b. Portanto, a transformada de Fourier dos dois lados da Eq. (7.25) produz

$$\frac{1}{T_b}\sum_{n=-\infty}^{\infty} P(f - nR_b) = 1 \qquad \text{em que} \qquad R_b = \frac{1}{T_b} \tag{7.26}$$

ou

$$\sum_{n=-\infty}^{\infty} P(f - nR_b) = T_b \tag{7.27}$$

Assim, a soma dos espectros formados com a repetição de $P(f)$ à frequência R_b é uma constante T_b, como mostrado na Fig. 7.12b.*

Consideremos o espectro na Fig. 7.12b, em um intervalo $0 < f < R_b$. Esse intervalo envolve apenas dois termos, $P(f)$ e $P(f - R_b)$, do somatório na Eq. (7.27). Logo,

$$P(f) + P(f - R_b) = T_b \qquad 0 < f < R_b$$

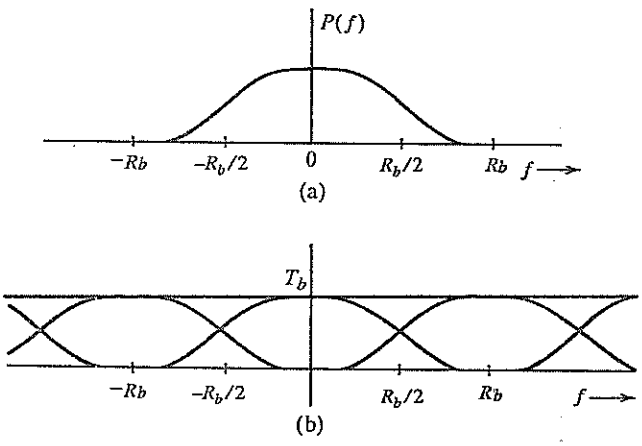


Figura 7.12 Dedução do critério Nyquist para ISI zero.

*Observemos que, se $R_b > 2B$, em que B é a largura de banda (em hertz) de $P(f)$, as repetições de $P(f)$ não se sobrepõem, e a condição (7.27) não pode ser satisfeita. Para $R_b = 2B$, a condição é satisfeita somente para o caso ideal passa-baixas $P(f)$ [$p(t) = \text{sinc}\,(\pi R_b t)$], que não é realizável. Logo, devemos ter $B > R_b/2$.

Seja $x = f - R_b/2$; portanto,

$$P(x + 0{,}5R_b) + P(x - 0{,}5R_b) = T_b \qquad |x| < 0{,}5R_b \tag{7.28a}$$

ou

$$P\left(x + \frac{R_b}{2}\right) + P\left(x - \frac{R_b}{2}\right) = T_b \qquad |x| < 0{,}5R_b \tag{7.28b}$$

Usando a propriedade da simetria conjugada [Eq. (3.11)] na Eq. (7.28), temos

$$P\left(\frac{R_b}{2} + x\right) + P^*\left(\frac{R_b}{2} - x\right) = T_b \qquad |x| < 0{,}5R_b \tag{7.29}$$

Se escolhermos $P(f)$ com valores reais e positivos, basta que $|P(f)|$ satisfaça a Eq. (7.29). Como $|P(f)|$ é real, a Eq. (7.29) implica em

$$\left|P\left(\frac{R_b}{2} + x\right)\right| + \left|P\left(\frac{R_b}{2} - x\right)\right| = T_b \qquad |x| < 0{,}5R_b \tag{7.30}$$

Portanto, $|P(f)|$ deve ter a forma mostrada na Fig. 7.13. Essa curva tem simetria ímpar em relação ao conjunto de eixos que se cruzam no ponto α [ponto na curva de $|P(f)|$ em $f = R_b/2$]. Isso requer

$$|P(0{,}5R_b)| = 0{,}5|P(0)|$$

A largura de banda de $P(f)$ em hertz é $0{,}5R_b + f_x$, em que f_x é a largura de banda além da mínima largura de banda $R_b/2$. Seja r a razão entre o excesso de largura de banda f_x e a mínima largura de banda teórica $R_b/2$:

$$\begin{aligned} r &= \frac{\text{excesso de largura de banda}}{\text{mínima largura de banda teórica}} \\ &= \frac{f_x}{0{,}5R_b} \\ &= 2f_xT_b \end{aligned} \tag{7.31}$$

Observemos que f_x não pode ser maior que $R_b/2$,

$$0 \le r \le 1 \tag{7.32}$$

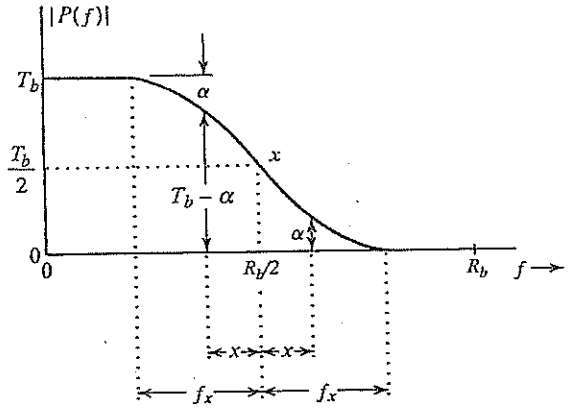


Figura 7.13 Espectro vestigial (cosseno levantado).

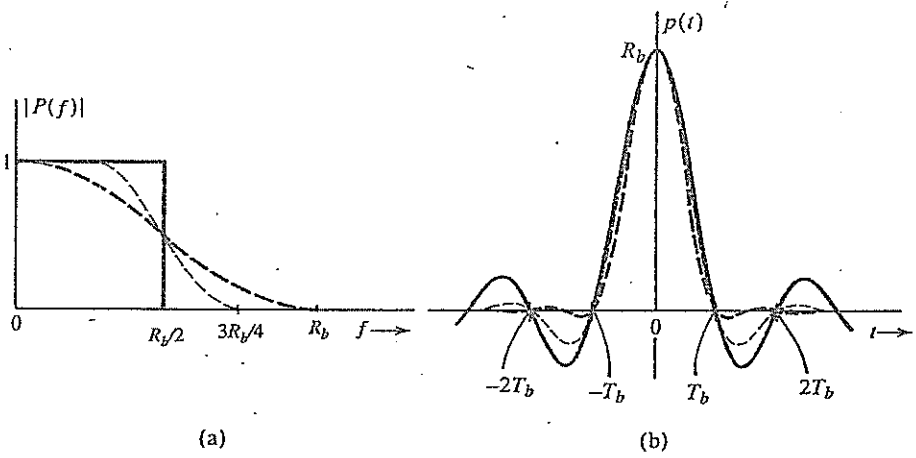


Figura 7.14 Pulsos que satisfazem o primeiro critério de Nyquist: linha cheia, $f_x = 0$ ($r = 0$) ideal; linha tracejada fina, $f_x = R_b/4$; linha tracejada grossa, $f_x = R_b/2$ ($r = 1$).

Em termos da frequência f, mínima largura de banda teórica é $R_b/2$ Hz, e o excesso de largura de banda é $f_x = rR_b/2$. Portanto, a largura de banda de $P(f)$ é

$$B_T = \frac{R_b}{2} + \frac{rR_b}{2} = \frac{(1+r)R_b}{2} \tag{7.33}$$

A constante r é denominada **fator de decaimento** (*roll-off*), e é também expressa em porcentagem. Por exemplo, se $P(f)$ for um espectro que segue o primeiro critério de Nyquist e cuja largura de banda seja 50% maior que o mínimo teórico, seu fator de decaimento será $r = 0{,}5$ ou 50%.

A modulação em banda lateral vestigial discutida na Seção 4.5 [Eq. (4.26)] requer um filtro cuja resposta de amplitude tenha essas mesmas características. Por essa razão, nos referimos ao espectro de $P(f)$ nas Eqs. (7.29) e (7.30) como **espectro vestigial**. O pulso $p(t)$ na Eq. (7.23) tem ISI zero nos centros de todos os outros pulsos transmitidos à taxa de R_b pulsos por segundo. Um pulso $p(t)$ que cause ISI zero nos centros de todos os outros pulsos (ou instantes de sinalização) é o pulso do primeiro critério de Nyquist. Mostramos que um pulso com espectro vestigial [Eq. (7.29) ou Eq. (7.30)] satisfaz o primeiro critério de Nyquist para ISI zero.

Como $0 \le r < 1$, a largura de banda de $P(f)$ fica restrita ao intervalo entre $R_b/2$ e R_b Hz. O pulso $p(t)$ pode ser gerado como a resposta ao impulso unitário de um filtro com função de transferência $P(f)$. Contudo, como $P(f) = 0$ em uma banda de frequências, o critério de Paley-Wiener é violado e, portanto, o filtro é irrealizável. Todavia, a característica de decaimento vestigial é gradual e pode ser bem aproximada por um filtro prático. Uma família de espectros que satisfazem o primeiro critério de Nyquist é

$$P(f) = \begin{cases} 1, & |f| < \dfrac{R_b}{2} - f_x \\ \dfrac{1}{2}\left[1 - \operatorname{sen} \pi\left(\dfrac{f - R_b/2}{2f_x}\right)\right] & \left|f - \dfrac{R_b}{2}\right| < f_x \\ 0, & |f| > \dfrac{R_b}{2} + f_x \end{cases} \tag{7.34}$$

A Fig. 7.14a mostra três curvas dessa família, correspondendo a $f_x = 0$ ($r = 0$), $f_x = R_b/4$ ($r = 0{,}5$) e $f_x = R_b/2$ ($r = 1$). As respectivas respostas ao impulso são mostradas na Fig. 7.14b. Pode ser visto que o aumento de f_x (ou de r) melhora $p(t)$; ou seja, um decaimento mais gradual reduz a natureza oscilatória de $p(t)$ e faz com que $p(t)$ decaia mais rapidamente no domínio do tempo. Para o caso do valor máximo $f_x = R_b/2$ ($r = 1$), a Eq. (7.34) se reduz a

$$P(f) = \frac{1}{2}(1 + \cos \pi f T_b)\,\Pi\left(\frac{f}{2R_b}\right) \tag{7.35a}$$

$$= \cos^2\left(\frac{\pi f T_b}{2}\right)\Pi\left(\frac{f T_b}{2}\right) \tag{7.35b}$$

A característica da Eq. (7.34) é conhecida na literatura como **cosseno levantado**, pois representa um cosseno levantado por sua amplitude de pico. A Eq. (7.35) também é conhecida como característica de **decaimento de cosseno amplo** (*full-cosine roll-off*). A transformada de Fourier inversa desse espectro é prontamente calculada como (Exercício 7.3-8)

$$p(t) = R_b \frac{\cos \pi R_b t}{1 - 4R_b^2 t^2} \operatorname{sinc}(\pi R_b t) \tag{7.36}$$

Esse pulso é mostrado na Fig. 7.14b ($r = 1$). Podemos fazer várias observações importantes sobre o pulso cosseno levantado. Primeira, esse pulso, cuja largura de banda é R_b Hz, tem valor R_b em $t = 0$ e é zero não apenas em todos os outros instantes de amostragens, mas também nos pontos médios entre todos os instantes de sinalização. Segunda, o pulso decai rapidamente, com $1/t^3$. Como resultado, o pulso cosseno levantado é relativamente insensível a desvios de R_b, da taxa de amostragem, incerteza temporal e assim por diante, Além disso, o filtro de geração de pulso com função de transferência $P(f)$ [Eq. (7.35b)] é realizável em boa aproximação. A característica de fase desse filtro é praticamente linear, de modo que nenhuma equalização de fase adicional se faz necessária.

Vale lembrar que os pulsos recebidos na entrada do detector é que devem ter a forma para ISI zero. Na prática, como o canal não é ideal (sem distorção), os pulsos transmitidos devem ser formatados para que, após passarem pelo canal com função de transferência $H_c(f)$, sejam recebidos com a forma adequada (como pulsos cosseno levantado). Portanto, o pulso transmitido $p_i(t)$ deve satisfazer

$$P_i(f)H_c(f) = P(f)$$

em que $P(f)$ tem o espectro vestigial da Eq. (7.30). Por conveniência, a função de transferência $H_c(f)$ como um canal também pode incluir um filtro receptor projetado para rejeitar interferências e outros ruídos fora da banda.

Exemplo 7.1 Determinemos a taxa de transmissão de pulsos em termos da largura de banda de transmissão B_T e do fator de decaimento r. Consideremos um esquema que use o primeiro critério de Nyquist.

Da Eq. (7.33)

$$R_b = \frac{2}{1+r} B_T$$

Como $0 \le r \le 1$, a taxa de transmissão de pulsos varia de $2B_T$ a B_T, dependendo da escolha de r. Um menor valor de r produz uma taxa de sinalização mais alta. Contudo, o pulso $p(t)$ decai lentamente, criando o mesmo problema discutido em relação ao pulso sinc. Para o pulso cosseno levantado $r = 1$ e $R_b = B_T$, alcançamos a metade da máxima taxa teórica. Mas, o pulso decai mais rapidamente, com $1/t^3$ e é menos vulnerável à ISI.

7.3.3 ISI Controlada ou Sinalização de Resposta Parcial

Pulsos que seguem o critério de Nyquist requerem uma largura de banda maior que o mínimo teórico. Se desejarmos reduzir a largura de banda do pulso, devemos encontrar uma forma de alargar o pulso $p(t)$ (mais largo o pulso, menor sua largura de banda). O aumento da largura do pulso pode resultar em interferência (ISI) em pulsos vizinhos. No entanto, na transmissão binária com apenas dois símbolos, pode ser possível remover ou cancelar um grau conhecido e controlado de ISI, pois somente uns poucos padrões de interferência são possíveis.

Consideremos um pulso especificado por (Fig. 7.15):

$$p(nT_b) = \begin{cases} 1 & n = 0, 1 \\ 0 & \text{para todos outros valores de } n \end{cases} \tag{7.37}$$

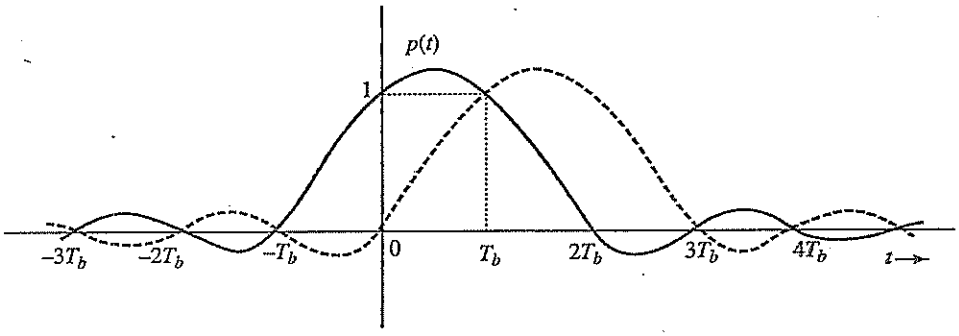


Figura 7.15 Comunicação usando ISI controlada ou o segundo critério de Nyquist.

Isso leva a um grau conhecido e controlado de ISI causada pelo k-ésimo pulso no próximo pulso transmitido. Usemos sinalização polar com esse pulso. Assim, **1** é transmitido por $p(t)$ e **0**, pelo pulso $-p(t)$. O sinal recebido é amostrado em $t = nT_b$, e o pulso $p(t)$ tem valor zero em todos os n, exceto em $n = 0$ e 1, quando vale 1 (Fig. 7.15). Fica claro que esse pulso causa ISI zero em todos os outros pulsos, exceto o que o sucede. Consideremos dois pulsos sucessivos localizados em 0 e T_b, respectivamente. Se os dois pulsos forem positivos, o valor amostrado do sinal resultante em $t = T_b$ será 2. Se os dois pulsos forem negativos, o valor amostrado será –2. Contudo, se os dois pulsos tiverem polaridades opostas, o valor amostrado será zero. Como somente esses três valores são possíveis, o valor da amostra do sinal permite que tomemos a decisão correta nos instantes de amostragem. A regra de decisão é a seguinte: se o valor da amostra for positivo, o bit corrente é **1** e o bit anterior também é **1**. Se o valor da amostra for negativo, o bit atual é **0** e o bit anterior também é **0**. Se o valor da amostra for zero, o bit atual é o oposto do anterior. O conhecimento do bit anterior permite a determinação do bit corrente.

A Tabela 7.1 mostra uma sequência de bits transmitidos, os valores das amostras do sinal recebido $x(t)$ (admitindo a ausência de erros devido a ruído do canal) e a decisão do detector. Esse exemplo também indica a propriedade de detecção de erro desse esquema. Um exame das amostras da forma de onda $y(t)$ na Tabela 7.1 revela que sempre há um número par de amostras de valor zero entre duas amostras de valores completos de mesma polaridade, e um número ímpar de amostras de valor zero entre duas amostras de valores completos de polaridades opostas. Assim, o primeiro valor amostrado de $x(t)$ é 2, e o próximo valor completo de amostra (da quarta amostra) é 2. Entre essas duas amostras de valores completos de mesma polaridade, há um número par (ou seja, 2) de amostras de valor zero. Se um dos valores de amostra for detectado erroneamente, essa regra será violada e o erro, detectado.

Tabela 7.1
Bits transmitidos e amostras recebidas na sinalização com ISI controlada

Sequência de informação	1	1	0	1	1	0	0	0	1	0	1	1	1
Amostra $y(kT_b)$	1	2	0	0	2	0	–2	–2	0	0	0	2	2
Sequência detectada	1	1	0	1	1	0	0	0	1	0	1	1	1

O pulso $p(t)$ vai a zero em $t = -T_b$ e $2T_b$, resultando em uma largura de pulso (do primeiro lóbulo) 50% maior que a do pulso do primeiro critério. Tal alargamento do pulso no domínio do tempo reduz a largura de banda do mesmo. Esse é o segundo critério proposto por Nyquist. Esse esquema de ISI controlada também é conhecido como esquema de **resposta parcial** ou **correlativa**. Um pulso que satisfaça o segundo critério na Eq. (7.37) também é conhecido como **pulso duobinário**.

7.3.4 Exemplo de Pulso Duobinário

Se restringirmos a largura de pulso a $R_b/2$, seguindo o procedimento do Exemplo 7.1, podemos mostrar que (Exercício 7.3-9) somente o pulso $p(t)$ dado a seguir satisfaz a exigência na Eq. (7.37) para o pulso duobinário:

$$p(t) = \frac{\operatorname{sen}(\pi R_b t)}{\pi R_b t(1 - R_b t)} \tag{7.38}$$

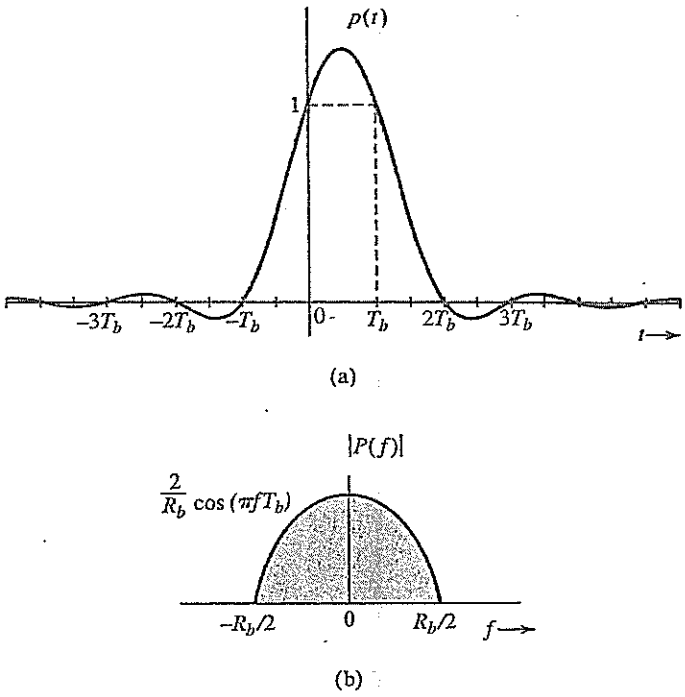


Figura 7.16 (a) Mínima largura de banda de pulso que satisfaz o critério de pulso duobinário e (b) seu espectro de amplitude.

A transformada de Fourier $P(f)$ do pulso $p(t)$ é dada por (Exercício 7.3-9):

$$P(f) = \frac{2}{R_b} \cos\left(\frac{\pi f}{R_b}\right) \Pi\left(\frac{f}{R_b}\right) e^{-j\pi f/R_b} \tag{7.39}$$

O pulso $p(t)$ e seu espectro de amplitude $|P(f)|$ são mostrados na Fig. 7.16.* Esse pulso transmite dados binários à taxa de R_b bit/s e tem a mínima largura de banda teórica de $R_b/2$ Hz. A Eq. (7.38) mostra que esse pulso decai rapidamente com o tempo, com $1/t^2$. Esse pulso não é idealmente realizável, pois $p(t)$ é não causal e tem duração infinita [pois $P(f)$ é limitado em frequência]. Contudo, decai rapidamente (com $1/t^2$) e, portanto, pode ser bem aproximado.

Pode parecer surpreendente que possamos alcançar a taxa teórica com o pulso duobinário. Na verdade, é uma ilusão. A taxa teórica de transmissão é 2 porções independentes de informação por segundo por hertz de largura de banda. Alcançamos essa taxa para informação binária. Aqui está o problema! Uma porção de informação binária não se qualifica como uma porção independente de informação, pois não pode assumir um valor arbitrário: deve ser selecionada de um conjunto finito. O pulso duobinário falharia se os pulsos fossem verdadeiramente porções independentes de informação, ou seja, se os pulsos pudessem assumir amplitudes arbitrárias. O esquema funciona somente porque os pulsos binários assumem valores finitos conhecidos e, portanto, existe apenas um número finito (conhecido) de padrões de interferência entre pulsos, o que permite a determinação correta das amplitudes dos pulsos, apesar da ocorrência de interferência.

7.3.5 Relações de Pulsos entre Sinalizações com ISI Zero, Duobinária e Duobinária Modificada

Agora, podemos estabelecer a relação simples entre um pulso $p_a(t)$ que satisfaz o primeiro critério de Nyquist (ISI zero) e um pulso duobinário $p_b(t)$ (com ISI controlada). Das Eqs. (7.23) e (7.37), fica

*O espectro de fase é linear: $\theta_p(f) = -\pi f T_b$.

claro que $p_a(kT_b)$ e $p_b(kT_b)$ diferem somente para $k = 1$. Para todos os outros valores do inteiro k, esses pulsos têm amostras com valores iguais. Portanto, podemos construir um pulso $p_b(t)$ de $p_a(t)$, da seguinte forma

$$p_b(t) = p_a(t) + p_a(t - T_b)$$

Essa adição é a sinalização com ISI "controlada" ou resposta parcial, deliberadamente introduzida para reduzir a exigência de largura de banda. Para determinar o efeito da sinalização "duobinária" na largura de banda espectral, consideremos a relação entre os dois pulsos no domínio da frequência:

$$P_b(f) = P_a(f)[1 + e^{-j2\pi fT_b}] \tag{7.40a}$$

$$|P_b(f)| = |P_a(f)|\sqrt{2(1 + \cos(2\pi fT_b)}2\,|\cos(\pi fT_b)| \tag{7.40b}$$

Podemos ver que a sinalização de resposta parcial, na verdade, força um nulo de $P(f)$ em $2\pi fT_b = \pi$ ou $f = 0{,}5/T_b$. Portanto, conceitualmente, podemos ver como a sinalização de resposta parcial oferece uma oportunidade adicional para reformatar a PSD ou largura de banda de transmissão. De fato, a sinalização duobinária, ao forçar um nulo na frequência $0{,}5/T_b$, faz com que sua largura de banda essencial seja a mínima largura de banda de transmissão necessária para uma taxa de dados de $1/T_b$ (como discutido na Seção 6.1.3).

Na verdade, muitos canais físicos, como uma gravação magnética, têm ganho zero em dc. Por tanto, não faz sentido que o sinal em banda base tenha qualquer componente dc em sua PSD. Uma modificação da sinalização de resposta parcial é comumente adotada para forçar um nulo em dc. Um exemplo notável é a chamada sinalização **duobinária modificada**, que requer

$$p_c(nT_b) = \begin{cases} 1 & n = -1 \\ -1 & n = 1 \\ 0 & \text{para todos outros inteiros } n \end{cases} \tag{7.41}$$

Um raciocínio similar indica que $p_c(t)$ pode ser gerado de qualquer pulso $p_a(t)$ que satisfaça o primeiro critério de Nyquist, via

$$p_c(t) = p_a(t + T_b) - p_a(t - T_b)$$

De modo equivalente, no domínio da frequência, o pulso duobinário é

$$P_c(f) = 2jP_a(f)\,\text{sen}\,(2\pi fT_b)$$

que usa sen $(2\pi fT_b)$ para forçar um nulo em dc e se conformar à limitação física do canal.

7.3.6 Detecção de Sinalização Duobinária e Codificação Diferencial

O diagrama básico do transmissor para o método de ISI controlada da sinalização duobinária é mostrado na Fig. 7.17. Agora, examinemos em mais detalhe a relação entre todos os símbolos de dados na banda base e o procedimento de detecção. Para o bit de mensagem binária $I_k = 0$ ou 1, símbolos polares são simplesmente

$$a_k = 2I_k - 1$$

Com ISI controlada, as amostras do sinal de transmissão $y(t)$ são

$$y(kT_b) = b_k = a_k + a_{k-1} \tag{7.42}$$

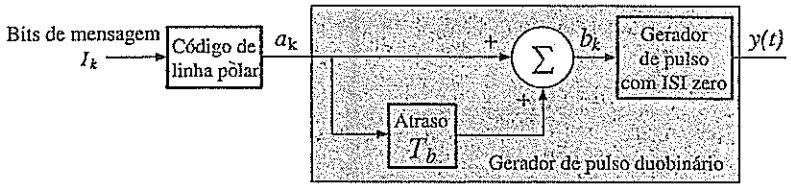


Figura 7.17 Sinalização duobinária equivalente.

Para o receptor, a questão é como *detectar* I_k de $y(kT_b)$ ou b_k. Essa questão pode ser respondida considerando, primeiro, todos os valores possíveis de b_k ou $y(kT_b)$. Como $a_k = \pm 1$, $b_k = 0, \pm 2$. Da Eq. (7.42), fica evidente que

$$\begin{array}{lll} b_k = 2 & \Rightarrow \quad a_k = 1 & \text{ou } I_k = 1 \\ b_k = -2 & \Rightarrow \quad a_k = -1 & \text{ou } I_k = 0 \\ b_k = 0 & \Rightarrow \quad a_k = -a_{k-1} & \text{ou } I_k = 1 - I_{k-1} \end{array} \tag{7.43}$$

Portanto, um simples detector de sinalização duobinária deve, primeiro, detectar todos os bits I_k correspondentes a $b_k = \pm 2$. Os remanescentes $\{b_k\}$ são amostras de valor zero que implicam transição; ou seja, o dígito corrente é **1**, e o anterior, **0** ou vice-versa. Isso significa que a detecção de dígitos deve ser baseada no dígito anterior. Um exemplo dessa detecção dígito a dígito foi mostrado na Tabela 7.1. A desvantagem do método de detecção na Eq. (7.43) é que, quando $y(kT_b) = 0$, a decisão para o bit corrente depende da decisão para o bit anterior. Se o dígito anterior tiver sido detectado de modo incorreto, o erro tende a se propagar, até que surja um valor de amostra ± 2. Para mitigar esse problema de propagação de erro, aplicamos um mecanismo eficiente, conhecido como **codificação diferencial**.

A Fig. 7.18 ilustra um gerador de sinal duobinário, no qual há um codificador diferencial adicional antes da geração do pulso de resposta parcial. Como mostrado na Fig. 7.18, a codificação diferencial é uma etapa muito simples que modifica a relação entre o código de linha e os bits de mensagem. A codificação diferencial gera uma nova sequência binária

$$p_k = I_k \oplus p_{k-1} \quad \text{módulo } 2$$

supondo que o estado inicial do pré-codificador seja $p_0 = 0$ ou $p_0 = 1$. A saída do pré-codificador é aplicada à entrada de um codificador de linha polar e gera

$$a_k = 2p_k - 1$$

Devido à sinalização duobinária $b_k = a_k + a_{k-1}$ e ao gerador de pulso com ISI zero, as amostras do sinal recebido $y(t)$ sem ruído se tornam

$$\begin{aligned} y(kT_b) = b_k &= a_k + a_{k-1} \\ &= 2(p_k + p_{k-1}) - 2 \\ &= 2\,(p_{k-1} \oplus I_k + p_{k-1} - 1) \\ &= \begin{cases} 2(1 - I_k) & p_{k-1} = 1 \\ 2(I_k - 1) & p_{k-1} = 0 \end{cases} \end{aligned} \tag{7.44}$$

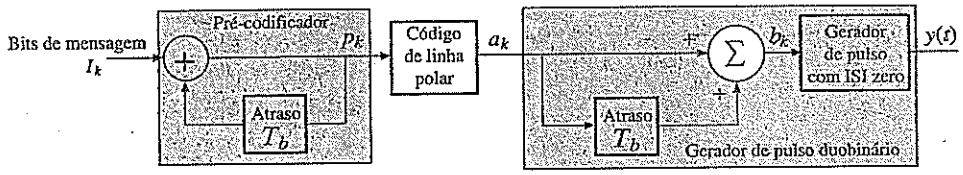


Figura 7.18 Sinalização duobinária com codificação diferencial.

A partir da Eq. (7.44), podemos resumir a relação direta entre os bits de mensagem e os valores de amostras como

$$y(kT_b) = \begin{cases} 0 & I_k = 1 \\ \pm 2 & I_k = 0 \end{cases} \tag{7.45}$$

Essa relação funciona como uma base para um algoritmo de detecção símbolo a símbolo. Em poucas palavras, o algoritmo de decisão se baseia na amostra corrente $y(kT_b)$; quando não há ruído, $y(kT_b) = b_k$ e a decisão do receptor é

$$I_k = \frac{2 - |y(kT_b)|}{2} \tag{7.46}$$

Portanto, a incorporação da codificação diferencial à sinalização duobinária não apenas simplifica a regra de decisão, mas também torna a decisão independente do dígito anterior e elimina a propagação de erro. Na Tabela 7.2, o exemplo da Tabela 7.1 é recalculado com a codificação diferencial. A relação de decodificação da Eq. (7.45) fica clara nesse exemplo.

Tabela 7.2
Sinalização duobinária com codificação diferencial

Tempo k	0	1	2	3	4	5	6	7	8	9	10	11	12	13
I_k		1	1	0	1	1	0	0	0	1	0	1	1	1
p_k	0	1	0	0	1	0	0	0	0	1	1	0	1	0
a_k	−1	1	−1	−1	1	−1	−1	−1	−1	1	1	−1	1	−1
b_k		0	0	−2	0	0	−2	−2	−2	0	2	0	0	0
Bits detectados		1	1	0	1	1	0	0	0	1	0	1	1	1

A codificação diferencial definida para símbolos de informação binária pode ser convenientemente generalizada para símbolos não binários. Quando os símbolos de informação I_k são M-ários, a única modificação a ser feita no bloco de codificação diferencial é a substituição de "módulo 2" por "módulo M". De modo similar, outra sinalização de resposta parcial generalizada, como a duobinária modificada, também está sujeita ao problema de propagação de erro na detecção. Um tipo adequado de codificação diferencial pode ser adotado para evitar a propagação de erro.

7.3.7 Geração de Pulsos

Um pulso $p(t)$ que satisfaça um critério de Nyquist pode ser gerado como a resposta ao impulso unitário de um filtro com função de transferência $P(f)$. Isso nem sempre é fácil. Um método melhor consiste em gerar a forma de onda diretamente, com o uso de um filtro transversal (linha de retardo com derivação) discutido a seguir. O pulso $p(t)$ a ser gerado é amostrado com um intervalo de amostragem T_s suficientemente pequeno (Fig. 7.19a), e os ganhos da derivação do filtro são ajustados proporcionalmente aos valores amostrados em sequência, como indicado na Fig. 7.19b. Quando um pulso retangular estreito, de largura T_s (intervalo de amostragem), é aplicado à entrada do filtro transversal, a saída será uma aproximação em degrau de $p(t)$. Essa saída, ao ser aplicada a um filtro passa-baixas, é suavizada. A aproximação pode ser melhorada com a redução do intervalo de amostragem T_s.

Vale ressaltar, mais uma vez, que os pulsos que chegam à entrada do detector no receptor devem atender o desejado critério de Nyquist. Portanto, os pulsos transmitidos devem ser formatados de modo que, após passagem pelo canal, sejam recebidos no formato (de Nyquist) desejado. Na prática, no entanto, não é necessário que pulsos sejam rigidamente formatados no transmissor. A formatação final pode ser efetuada por um equalizador no receptor, como discutido mais adiante (Seção 7.5).

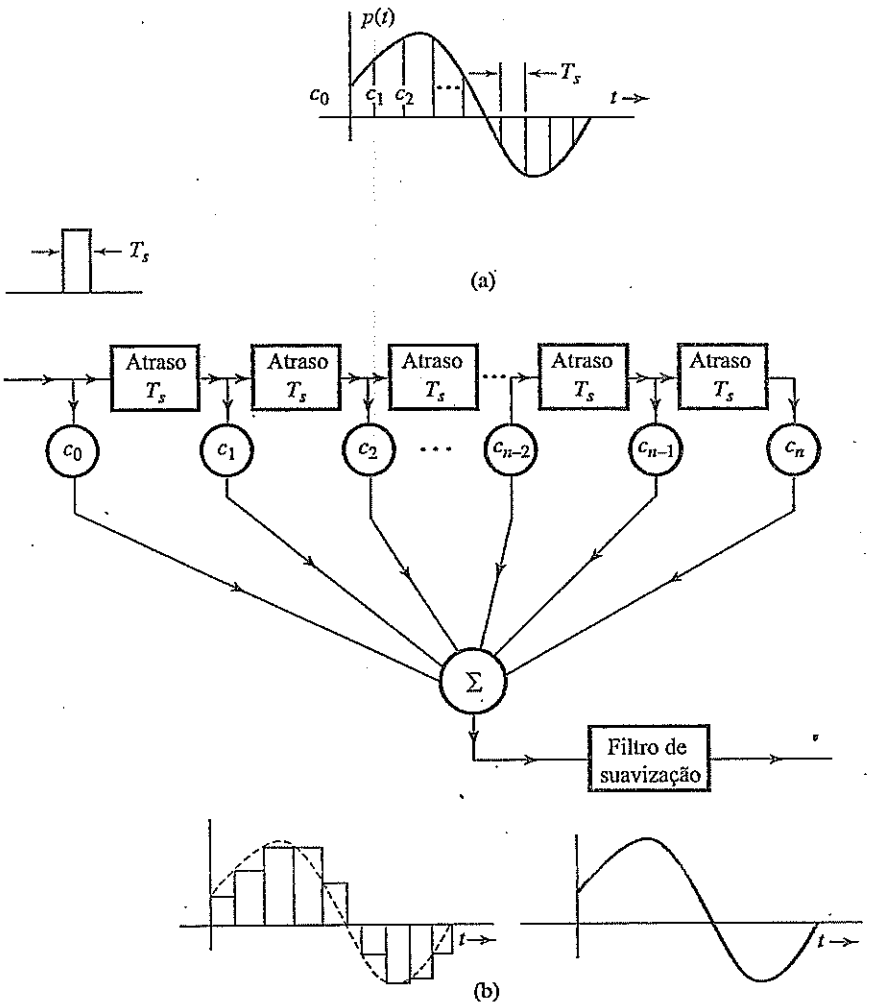


Figura 7.19 Geração de pulso por filtro transversal.

7.4 EMBARALHAMENTO (*SCRAMBLING*) DE DADOS

Em geral, um embaralhador tende aumentar o caráter aleatório de dados através da remoção de sequências de **1**s ou **0**s. O embaralhamento pode ser útil na extração da temporização, por remover longas sequências de **0**s de dados binários. Contudo, embaralhadores são usados principalmente para evitar acesso não autorizado aos dados, sendo otimizados para esse fim. Essa otimização pode resultar na geração de uma longa sequência de zeros nos dados. A rede digital deve ser capaz de tratar de tais longas sequências de zeros com o emprego de técnicas de substituição discutidas na Seção 7.2.

A Fig. 7.20 mostra típicos embaralhador e desembaralhador. O embaralhador consiste em um registrador de deslocamento com realimentação (*feedback shift register*), e o correspondente desembaralhador tem um registrador de deslocamento com antecipação (*feedforward shift register*), como indicado na Fig. 7.20. Cada estágio do registrador de deslocamento atrasa um bit em uma unidade. Para analisar o embaralhador e o correspondente desembaralhador, consideremos uma sequência de saída T do embaralhador (Fig. 7.20a). Se S for a sequência de entrada do embaralhador,

$$S \oplus D^3T \oplus D^5T = T \tag{7.47}$$

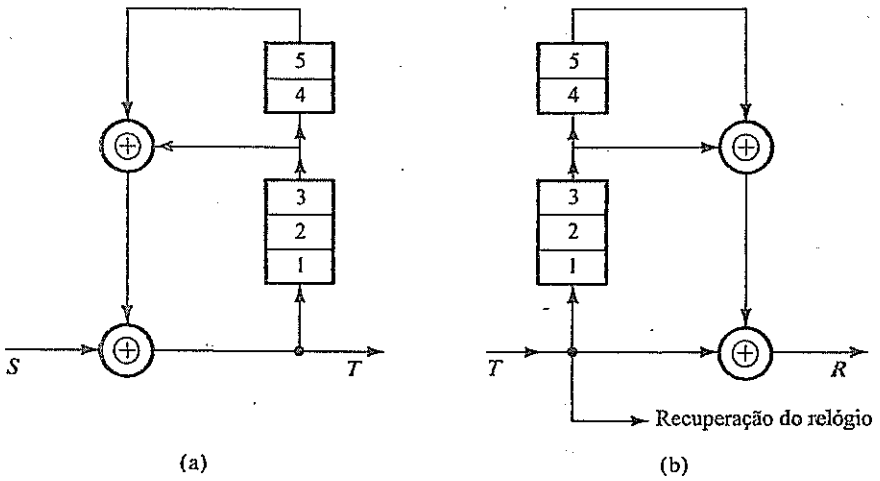


Figura 7.20 (a) Embaralhador. (b) Desembaralhador.

em que D representa o operador de atraso; ou seja, D^nT é a sequência T atrasada por n unidades. Agora, recordando que a soma em módulo 2, de qualquer sequência, com ela própria produz uma sequência toda de **0**s. Somando $(D^3 \oplus D^5)$ aos dois lados da Eq. (7.47), obtemos,

$$\begin{aligned} S &= T \oplus (D^3 \oplus D^5)T \\ &= [1 \oplus (D^3 \oplus D^5)]T \\ &= (1 \oplus F)T \end{aligned} \tag{7.48}$$

em que $F = D^3 \oplus D^5$.

Para projetar o desembaralhador no receptor, iniciamos com T, a sequência recebida no desembaralhador. Da Eq. (7.48), temos

$$T \oplus FT = T \oplus (D^3 \oplus D^5)T = S$$

Essa equação, na qual regeneramos a sequência de entrada S a partir da sequência recebida T, é prontamente implementada pelo desembaralhador mostrado na Fig. 7.20b.

Reparemos que um erro de detecção isolado na sequência recebida T afetará três bits de saída em R. Logo, o embaralhamento tem a desvantagem de causar múltiplos erros para um único erro de bit recebido.

Exemplo 7.2 A sequência de dados **101010100000111** é alimentada ao embaralhador na Fig. 7.20a. Determinemos a saída T do embaralhador, admitindo que o conteúdo inicial dos registradores seja zero.

Da Fig. 7.20a, observamos que, inicialmente, $T = S$, a sequência entra no registrador e sai como $(D^3 \oplus D^5)S = FS$ pela rota de realimentação. Essa nova sequência FS entra novamente no registrador e retorna como F^2S, e assim por diante. Portanto,

$$\begin{aligned} T &= S \oplus FS \oplus F^2S \oplus F^3S \oplus \cdots \\ &= (1 \oplus F \oplus F^2 \oplus F^3 \oplus \cdots)S \end{aligned} \tag{7.49}$$

Reconhecendo

$$F = D^3 \oplus D^5$$

temos

$$F^2 = (D^3 \oplus D^5)(D^3 \oplus D^5) = D^6 \oplus D^{10} \oplus D^8 \oplus D^8$$

Como a soma em módulo 2 de qualquer sequência com ela própria é zero, $D^8 \oplus D^8 = 0$ e

$$F^2 = D^6 \oplus D^{10}$$

Do mesmo modo,

$$F^3 = (D^6 \oplus D^{10})(D^3 \oplus D^5) = D^9 \oplus D^{11} \oplus D^{13} \oplus D^{15}$$

e assim por diante. Logo [Eq. (7.49)],

$$T = (1 \oplus D^3 \oplus D^5 \oplus D^6 \oplus D^9 \oplus D^{10} \oplus D^{11} \oplus D^{12} \oplus D^{13} \oplus D^{15} \cdots)S$$

Como D^nS é simplesmente a sequência S atrasada em n bits, vários termos na equação anterior correspondem às seguintes sequências:

$$\begin{aligned}
S &= \mathbf{101010100000111} \\
D^3S &= \mathbf{000101010100000111} \\
D^5S &= \mathbf{00000101010100000111} \\
D^6S &= \mathbf{000000101010100000111} \\
D^9S &= \mathbf{000000000101010100000111} \\
D^{10}S &= \mathbf{0000000000101010100000111} \\
D^{11}S &= \mathbf{00000000000101010100000111} \\
D^{12}S &= \mathbf{000000000000101010100000111} \\
D^{13}S &= \mathbf{0000000000000101010100000111} \\
\underline{D^{15}S} &= \underline{\mathbf{000000000000000}}\mathbf{101010100000111} \\
T &= \mathbf{101110001101001}
\end{aligned}$$

Notemos que a sequência de entrada contém a sequência periódica **10101010**..., assim como uma longa sequência de **0**s. A saída do embaralhador remove a componente periódica, assim como a longa sequência de **0**s. A sequência de entrada tem 15 dígitos. A saída do embaralhador é mostrada somente até o 15º dígito, pois todos os outros dígitos além do 15º dependem dos dígitos de entrada além do 15º, que não foram dados.

O leitor pode comprovar que a saída do desembaralhador é, de fato, S, se a sequência T anterior for aplicada à sua entrada.

7.5 RECEPTORES DIGITAIS E REPETIDORES REGENERATIVOS

Basicamente, um receptor ou repetidor regenerativo executa três funções: (1) reformata pulsos recebidos por meio de um equalizador, (2) extrai informação de temporização necessária para amostrar os pulsos recebidos nos instantes ótimos, e (3) decide que símbolo foi detectado, com base em amostras de pulsos. O repetidor mostrado na Fig. 7.21 consiste em um receptor mais um "regenerador". Um receptor completo pode incluir ainda a provisão para separar potência dc e sinais ac. Isso é normalmente feito por acoplamento ac dos sinais por transformador, e desviando dc do transformador para o circuito de alimentação de potência.*

*O repetidor, em geral, inclui circuito para proteger a eletrônica do regenerador de transientes de alta tensão induzidos por surtos de potência ou descargas elétricas. Enrolamentos especiais de transformador podem ser fornecidos para acoplar sinais espúrios a cabos dedicados.

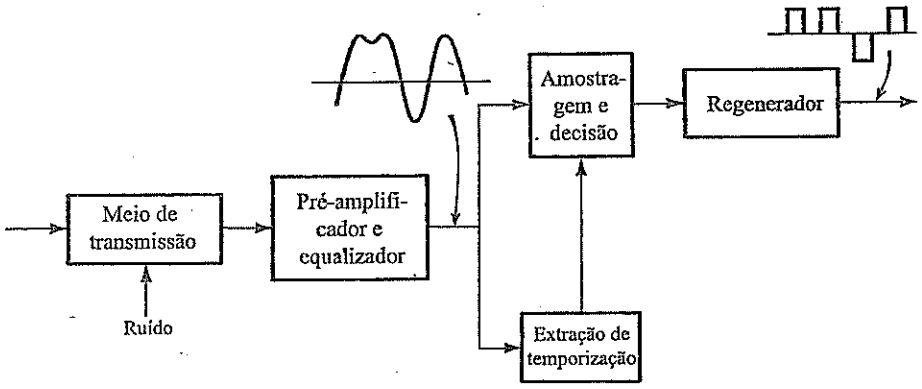


Figura 7.21 Repetidor regenerador.

7.5.1 Equalizadores

Um trem de pulsos é atenuado e distorcido pelo meio de transmissão. A atenuação pode ser compensada pelo pré-amplificador, enquanto a distorção é compensada pelo equalizador. A distorção de canal tem a forma de dispersão, causada pela atenuação de certas *componentes em frequências críticas* do trem de pulsos de dados. Teoricamente, um equalizador deve ter uma característica de frequência igual ao inverso da característica do meio de transmissão. Isso restaurará as componentes em frequências críticas e eliminará a dispersão dos pulsos. Lamentavelmente, isso também realça o ruído de canal recebido, pois amplifica as componentes desse nas frequências críticas. Esse fenômeno indesejável é conhecido como *amplificação de ruído*.

Para sinais digitais, no entanto, a equalização completa não é necessária, pois o detector precisa tomar decisões relativamente simples – como decidir se o pulso é positivo ou negativo (ou se o pulso está presente ou ausente). Portanto, considerável dispersão de pulsos pode ser tolerada. Dispersão de pulsos resulta em ISI e, por conseguinte, em aumento no erro de detecção. O aumento de ruído resultante da equalização (que amplifica as frequências altas) também aumenta a probabilidade de erro de detecção. Por essa razão, o projeto de um equalizador ótimo envolve um inevitável equilíbrio entre redução de ISI e redução de ruído de canal. Uma escolha acertada das características de equalização é uma questão central em todos os sistemas de comunicação digital bem projetados.[6]

Equalizador com Forçamento a Zero

A eliminação ou minimização da ISI (interferência) entre pulsos vizinhos para todo t não é, de fato, necessária. Basta eliminar ou minimizar a interferência entre pulsos vizinhos em seus respectivos *instantes de amostragem*, pois a decisão do receptor é baseada somente em valores de amostras. Esse tipo de equalização (abrandada) pode ser feito com equalizadores que usem a estrutura de filtro transversal discutida anteriormente. Ao contrário de filtros tradicionais, equalizadores a filtros transversais são facilmente ajustáveis para compensar efeitos de canais diferentes ou que variem lentamente no tempo. O objetivo de projeto é forçar que o pulso de saída do equalizador tenha ISI zero nos instantes de amostragem (tomada de decisão). Em outras palavras, os pulsos de saída do equalizador satisfazem o critério de Nyquist ou o de ISI controlada. O atraso temporal T entre derivações sucessivas é escolhido como T_b, o intervalo entre pulsos.

Inicialmente, tomemos os ganhos das derivações $c_0 = 1$ e $c_k = 0$, para todos os valores de k, no filtro transversal na Fig. 7.22a. Com isso, a saída do filtro será igual à entrada atrasada por NT_b. Para um pulso $p_r(t)$ (Fig. 7.22b) na entrada do filtro transversal com essa especificação de derivação, a saída do filtro $p_o(t)$ será exatamente $p_r(t - NT_b)$, ou seja, $p_r(t)$ atrasado por NT_b. Esse atraso não tem efeito prático em nosso sistema de comunicação e não é relevante para nossa discussão. Assim, por conveniência, ignoraremos esse atraso. Isso significa que $p_r(t)$ na Fig. 7.22b também representa a saída do filtro

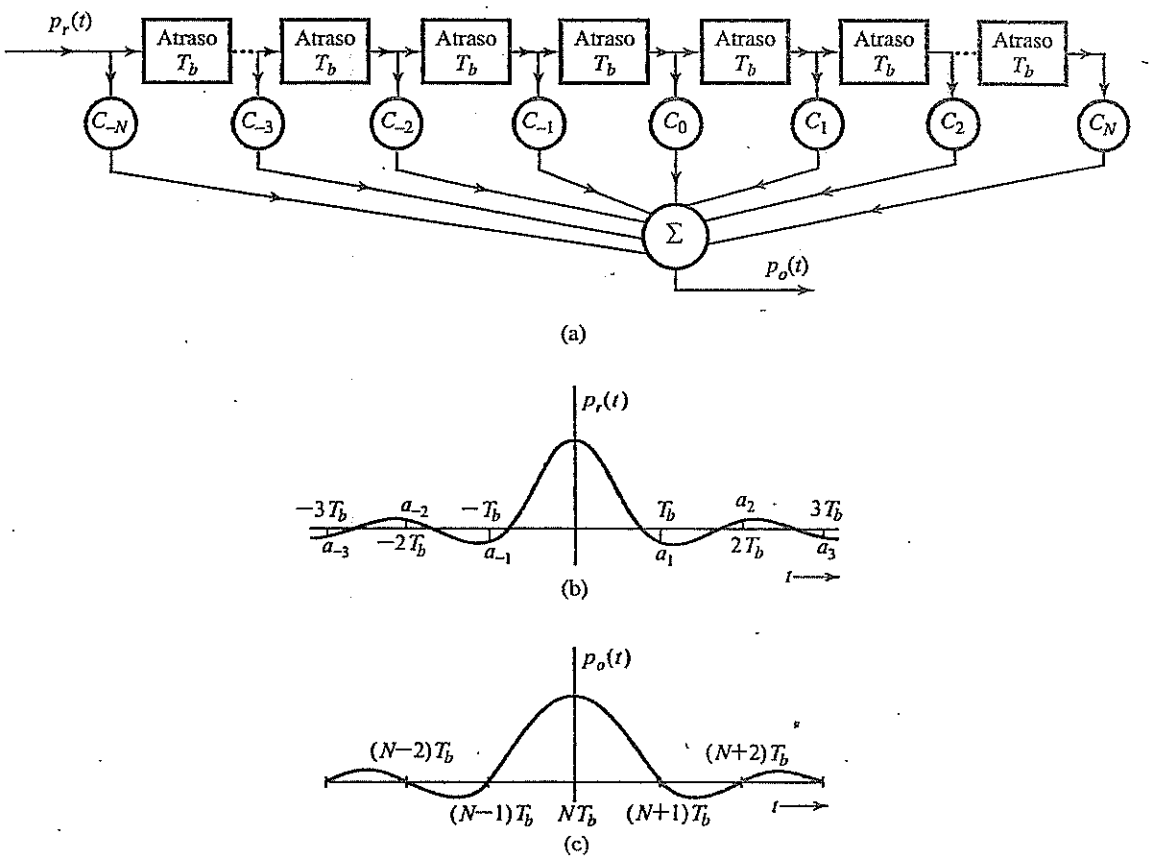


Figura 7.22 Análise de equalizador com forçamento a zero.

$p_o(t)$ para essa configuração de derivações ($c_0 = 1$ e $c_k = 0$, $k \neq 0$). Exigimos que esse pulso satisfaça o critério de Nyquist ou o critério de ISI controlada, como for o caso. Da Fig. 7.22b, vemos que as amplitudes de pulso a_1, a_{-1}, e a_2 em T_b, $-T_b$ e $2T_b$, respectivamente, não são desprezíveis. Ajustando os ganhos das derivações (c_k), geramos pulsos deslocados adicionais de amplitudes apropriadas que forçarão o resultante pulso de saída a ter os valores desejados em $t = 0, \pm T_b, \pm 2T_b, \ldots$

A saída $p_o(t)$ (Fig. 7.22c) é a soma de pulsos da forma $c_k p_r(t - kT_b)$ (ignorando o atraso NT_b). Assim,

$$p_o(t) = \sum_{n=-N}^{N} c_n p_r(t - nT_b) \tag{7.50}$$

As amostras de $p_o(t)$ em $t = kT_b$ são

$$p_o(kT_b) = \sum_{n=-N}^{N} c_n p_r(kT_b - nT_b) \qquad k = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{7.51a}$$

Usando uma notação mais conveniente $p_r[k]$ para denotar $p_r(kT_b)$, e $p_o[k]$ para denotar $p_o(kT_b)$, a Eq. (7.51a) pode ser expressa como

$$p_o[k] = \sum_{n=-N}^{N} c_n p_r[k - n] \qquad k = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{7.51b}$$

O primeiro critério de Nyquist requer amostras $p_o[k] = 0$ para $k \neq 0$ e $p_o[k] = 1$ para $k = 0$. Substituindo esses valores na Eq. (7.51b), obtemos um conjunto de infinitas equações simultâneas em termos de $2N + 1$ variáveis. Obviamente, a solução dessas equações não é possível. Contudo, se especificarmos os valores de $p_o[k]$ somente em $2N + 1$ pontos como

$$p_o[k] = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1,\ \pm 2, \ldots, \pm N \end{cases} \tag{7.52}$$

existirá uma única solução. Isso assegura que um pulso terá interferência zero nos instantes de amostragem de N pulsos anteriores e N pulsos posteriores. Como as amplitudes dos pulsos decaem rapidamente, em geral, interferência além do N-ésimo pulso não é significativa para $N > 2$. A substituição da condição (7.52) na Eq. (7.51b) produz um conjunto de $2N + 1$ equações simultâneas para $2N + 1$ incógnitas. Essas $2N + 1$ equações podem ser reescritas em forma matricial como

$$\underbrace{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{p}_o} = \underbrace{\begin{bmatrix} p_r[0] & p_r[-1] & \cdots & p_r[-2N+1] & p_r[-2N] \\ p_r[1] & p_r[0] & \cdots & p_r[-2N+2] & p_r[-2N+1] \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ p_r[2N-1] & p_r[2N-2] & \ddots & p_r[0] & p_r[-1] \\ p_r[2N] & p_r[2N-1] & \cdots & p_r[1] & p_r[0] \end{bmatrix}}_{\mathbf{P}_r} \cdot \underbrace{\begin{bmatrix} c_{-N} \\ c_{-N+1} \\ \vdots \\ c_{-1} \\ c_0 \\ c_1 \\ \vdots \\ c_{N-1} \\ c_N \end{bmatrix}}_{\mathbf{c}} \tag{7.53}$$

Nessa expressão compacta, a matriz $\mathbf{P}_r$ de ordem $(2N + 1) \times (2N + 1)$ tem entradas idênticas ao longo de todas as diagonais. Esse tipo de matriz é conhecida como matriz de Toeplitz e é comumente encontrado na descrição de relações convolutivas. Uma matriz de Toeplitz é totalmente determinada pelas primeiras linha e coluna, tem algumas propriedades interessantes e admite algoritmos mais simples para o cálculo de sua inversa (por exemplo, os métodos de Trench[7]). Os ganhos c_k das derivações podem ser obtidos da solução dessa equação, com o cálculo da inversa da matriz $\mathbf{P}_r$,

$$\mathbf{c} = \mathbf{P}_r^{-1}\mathbf{p}_o$$

Exemplo 7.3 Para o pulso recebido $p_r(t)$ na Fig. 7.22b, sejam

$$p_r[0] = 1$$
$$p_r[1] = -0{,}3 \qquad p_r[2] = 0{,}1$$
$$p_r[-1] = -0{,}2 \qquad p_r[-2] = 0{,}05$$

Projetemos um equalizador de três derivações ($N = 1$).

Substituindo os valores anteriores na Eq. (7.53), obtemos

$$\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & -0{,}2 & 0{,}05 \\ -0{,}3 & 1 & -0{,}2 \\ 0{,}1 & -0{,}3 & 1 \end{bmatrix} \begin{bmatrix} c_{-1} \\ c_0 \\ c_1 \end{bmatrix} \tag{7.54}$$

A solução desse conjunto de equações produz $c_{-1} = 0{,}210$, $c_0 = 1{,}13$ e $c_1 = 0{,}318$. Essa configuração de derivações assegura $p_0[0] = 1$ e $p_0[-1] = p_0[1] = 0$. O gráfico da saída ideal $p_o(t)$ é mostrado na Fig. 7.22c.

Vale notar que o equalizador determinado da Eq. (7.53) pode garantir somente a condição de ISI zero da Eq. (7.52). Em outras palavras, a ISI é zero somente para $k = 0, \pm 1, \ldots, \pm N$. Na verdade, para outros valores de k, é muito comum que as amostras $p_o(kT_b) \neq 0$, indicando ISI residual. Por exemplo, consideremos o problema de equalizador do Exemplo 7.3. As amostras do pulso equalizado têm ISI zero para $k = -1, 0, 1$. Contudo, de

$$p_o[k] = \sum_{n=-N}^{N} c_n p_r[k-n]$$

podemos ver que os parâmetros do equalizado de três derivações com forçamento a zero resultarão em

$$p_o[-3] = 0{,}010 \quad p_o[-2] = 0{,}0145 \quad p_o[2] = 0{,}0176$$
$$p_o[3] = 0{,}0318 \quad p_o[k] = 0 \quad k = 0, \pm 1, \pm 4, \ldots$$

Fica claro, portanto, que nem toda ISI foi removida, devido a essas quatro amostras não nulas do pulso de saída do equalizador. Na verdade, como temos apenas $2N + 1$ ($N = 1$ no Exemplo 7.3) parâmetros no equalizador, é impossível forçar $p_o[k] = 0$ para todo k, a menos que $N = \infty$. Isso significa que não poderemos projetar um equalizador prático com um número finito de derivações que alcance ISI zero ideal. Todavia, quando N é suficientemente grande, os valores de amostras não nulas residuais serão pequenos, indicando que a maior parte da ISI foi suprimida.

Método do Mínimo Erro Quadrático Médio

Na prática, uma abordagem alternativa consiste em minimizar a diferença quadrática média entre a resposta de saída do equalizador $p_o[k]$ e a resposta desejada com ISI zero. Isso é conhecido como método do mínimo erro quadrático médio (MMSE – *minimum mean square error*) para o projeto de equalizadores com filtros transversais. O método MMSE não tenta forçar que as amostras de pulsos sejam zero em $2N$ pontos, mas busca minimizar o erro quadrático médio considerando um conjunto de amostras de saída. O método envolve mais equações simultâneas. Assim, devemos determinar os valores de derivações do equalizador para minimizar o erro quadrático (médio) em uma janela maior $[-K, K]$:

$$\text{MSE} \triangleq \frac{1}{2K+1} \sum_{k=-K}^{K} (p_o[k] - \delta[k])^2$$

em que usamos uma função conhecida como delta de Kronecker

$$\delta[k] = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

A solução desse problema de minimização é convenientemente representada em forma matricial como

$$\mathbf{c} = \mathbf{P}_r^{\dagger} \mathbf{p}_o$$

em que $\mathbf{P}_r^{\dagger}$ representa a pseudoinversa de Moore-Penrose da matriz não quadrada $\mathbf{P}_r$ de ordem $(2K + 1) \times (2N + 1)$. O projeto MMSE leva a um equalizador mais robusto para a redução de ISI.

Equalização Adaptativa e Outros Equalizadores Mais Genéricos

A estrutura do filtro equalizador descrita aqui tem a forma mais simples. Sistemas de comunicação digital práticos, muitas vezes, aplicam estruturas equalizadoras mais sofisticadas e algoritmos de equalização mais avançados.[6] Devido às necessárias ferramentas probabilísticas, adiaremos a discussão desse tema especializado de equalização para o Capítulo 12.

7.5.2 Extração de Temporização

O sinal digital recebido deve ser amostrado em instantes precisos. Isso requer um sinal de relógio no receptor em sincronismo com o sinal do relógio no transmissor (**sincronização de símbolo** ou **de bit**), atrasado pela resposta do canal. Há três métodos genéricos de sincronização:

1. Derivação de um padrão primário ou secundário (por exemplo, transmissor e receptor escravizados à fonte mestra de temporização).
2. Transmissão de um sinal de sincronização separado (relógio piloto).
3. Autossincronização, em que a informação de temporização é extraída do próprio sinal recebido.

Devido ao alto custo, o primeiro método é adequado a grandes volumes de dados e sistemas de comunicação de alta velocidade. O segundo método, em que parte da capacidade do canal é usada para transmitir a informação de temporização, é adequado a situações em que a capacidade disponível é grande em comparação com a taxa de dados, e quando há disponibilidade de potência de transmissão. O terceiro é um método muito eficiente de extração de temporização ou recuperação de relógio, pois a temporização é extraída do próprio sinal de mensagem recebido. Um exemplo do método de autossincronização será discutido a seguir.

Já mostramos que um sinal digital, como o sinal on-off (Fig. 7.3a), contém uma componente discreta da frequência do relógio (Fig. 7.3c). Portanto, quando o sinal binário on-off é aplicado a um circuito ressonante sintonizado na frequência do relógio, o sinal de saída é o desejado sinal do relógio.

Nem todos os sinais binários contêm uma componente discreta da frequência do relógio. Por exemplo, um sinal bipolar não tem componente discreta de qualquer frequência [Eq. (7.21) ou Fig. 7.9]. Nesses casos, a extração da temporização pode ser possível com o uso de um *dispositivo não linear* para gerar um tom de frequência relacionado ao relógio de temporização. No caso bipolar, por exemplo, uma simples retificação converte um sinal bipolar em um sinal on-off, que pode ser prontamente utilizado para extrair a temporização.

Pequenos desvios aleatórios dos pulsos recebidos em relação à localização ideal (conhecidos como **incerteza temporal**) sempre estão presentes, mesmo nos mais sofisticados sistemas. Embora a fonte emita pulsos nos instantes corretos, operações subsequentes durante a transmissão (como deslocamento Doppler) tendem a forçar os pulsos a se desviarem de suas posições originais. O Q do circui-

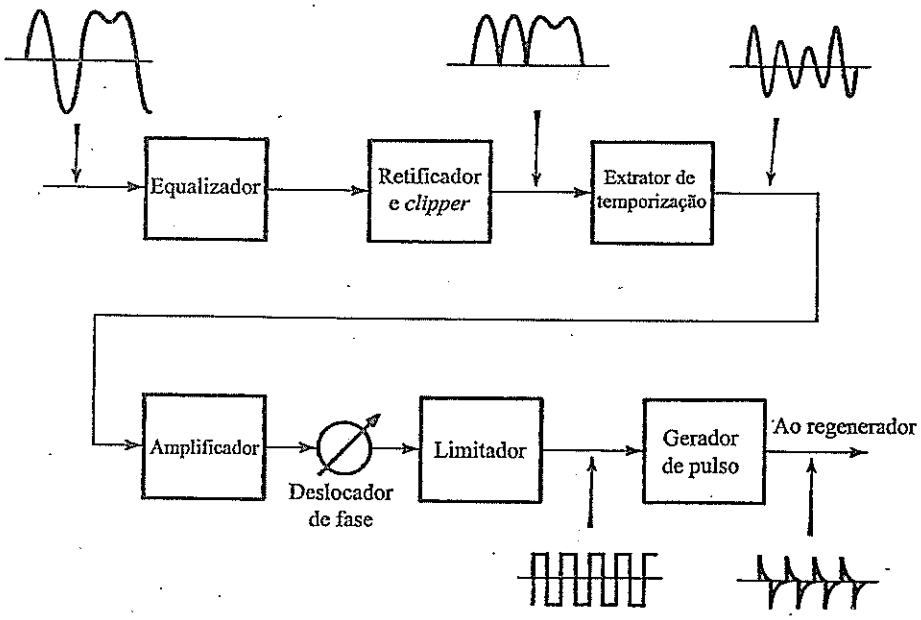


Figura 7.23 Extração de temporização.

to sintonizado usado para extração da temporização deve ser grande o bastante para prover adequada supressão da incerteza temporal, tornando-a suficientemente pequena para atender os requisitos de estabilidade. Durante os intervalos em que não há pulos na entrada, a oscilação continua, devido ao efeito de volante (*flywheel effect*) do circuito de alto Q. Não obstante, a saída do oscilador é sensível ao padrão de entrada; por exemplo, durante uma longa sequência de **1**s, a amplitude de saída aumentará, ao passo que durante uma longa sequência de **0**s, diminuirá. Isso introduz incerteza adicional ao sinal de temporização extraído.

O completo extrator de temporização e gerador de pulso de temporização para o caso bipolar são ilustrados na Fig. 7.23. A saída senoidal do oscilador (extração de temporização) é aplicada ao deslocador de fase, que ajusta a fase do sinal de temporização de modo que os pulsos de temporização ocorram nos pontos máximos. Esse método é usado para recuperar o relógio em cada um dos regeneradores em um sistema PCM. As incertezas introduzidas por sucessivos regeneradores se somam e, após um certo número de regeneradores, se torna necessário o uso de um regenerador com um sistema mais sofisticado de recuperação do relógio, como um sistema de *malha de captura de fase* (PLL – *phase locked loop*).

Incerteza temporal

Variações nas posições dos pulsos ou instantes de amostragem causam incerteza temporal (*timing jitter*). Isso advém de várias causas, algumas das quais dependem do padrão de pulsos sendo transmitido, enquanto outras não. As primeiras são cumulativas ao longo da cadeia de repetidores regenerativos, pois todos os repetidores são afetados da mesma forma. As outras formas de incerteza são aleatórias de regenerador para regenerador e, portanto, tende a haver um cancelamento parcial de efeitos mútuos ao longo de um enlace de grande distância. Formas aleatórias de incerteza temporal são causadas por ruído, interferências e desvio de sintonia de circuitos de relógio. Incerteza dependente de padrão resulta de desvios de relógio, conversão amplitude-fase em circuitos de relógio e ISI, que altera a posição do pico do sinal de entrada segundo o padrão. É possível mostrar que o valor rms da incerteza ao longo de uma cadeia de N repetidores aumenta com $\sqrt{N}$.

O acúmulo de incerteza em um enlace digital pode ser reduzido com o emprego de armazenagem elástica (*buffer*) e temporização da sequência de dígitos sob o controle de uma malha de captura de fase altamente estável. Em um longo enlace digital, se faz necessária a redução de incertezas a cada 300 km, para manter a incerteza máxima dentro de limites razoáveis.

7.5.3 Detecção de Erro

Uma vez que a transmissão tenha passado pelo equalizador, a detecção pode ser feita no detector que amostra o sinal recebido com base no relógio fornecido pelo extrator de temporização. O sinal recebido no detector consiste no trem de pulsos equalizados mais um ruído de canal aleatório. O ruído pode causar erro na detecção de pulso. Consideremos, por exemplo, o caso da transmissão polar que usa um pulso básico $p(t)$ (Fig. 7.24a). Esse pulso tem uma amplitude de pico A_p. Um típico trem de pulsos recebidos é mostrado na Fig. 7.24b. Os pulsos são amostrados em seus valores de pico. Se o ruído estivesse ausente, a amostra do pulso positivo (correspondendo a **1**) seria A_p e a do pulso negativo (correspondendo a **0**), $-A_p$.* Devido ao ruído, os valores dessas amostras passariam a $\pm A_p + n$, em que n é a amplitude do ruído aleatório (Fig. 7.24b). Dada a simetria da situação, o limiar de detecção é zero; ou seja, se o valor da amostra de pulso for positivo, o dígito é detectado como **1**; se o valor da amostra for negativo, o dígito é detectado como **0**.

A decisão do detector em declarar **1** ou **0** poderia ser tomada prontamente a partir da amostra de pulso, não fosse o valor do ruído n aleatório, o que significa que seu valor exato é imprevisível. O valor do ruído pode ser grande ou pequeno, assim como pode ser negativo ou positivo. É possível que **1** seja transmitido e que, no instante de amostragem, n tenha um valor negativo grande. Isso tornaria o valor da amostra $A_p + n$ pequeno ou, até mesmo, negativo. Se, por outro lado, um **0** for transmitido e

*Assumindo ISI zero.

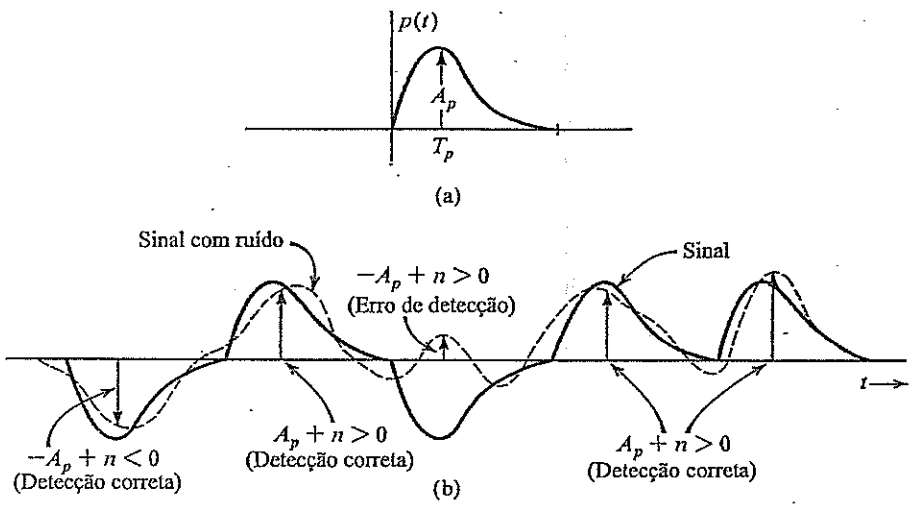


Figura 7.24 Probabilidade de erro na detecção por limiar.

n tiver um valor positivo grande no instante de amostragem, o valor da amostra $-A_p + n$ pode ser positivo e o dígito, detectado como **1**. Isso fica claro na Fig. 7.24b.

O desempenho de um sistema de comunicação digital é, em geral, especificado pelo número médio de erros de detecção. Por exemplo, se dois telefones celulares (receptores) no mesmo local tentam detectar a mesma transmissão da estação radiobase, o telefone celular com menor erros de detecção será o melhor receptor. É provável que esse telefone tenha menor perda de chamadas e menos dificuldade para receber uma fala clara. Contudo, como o ruído é aleatório, algumas vezes um telefone celular pode ser melhor; outras, o segundo telefone celular pode ter menos erros. A real medida do desempenho do receptor é a razão média entre o número de erros e o número total de dados transmitidos. Assim, a comparação razoável de desempenho é a probabilidade de ocorrência de erro de detecção, ou **probabilidade de erro de detecção.**

Como a análise precisa e o cálculo dessa probabilidade de erro requer conhecimento e ferramentas da teoria da probabilidade, adiaremos a análise de erros até introdução de probabilidade no Capítulo 8. Mais adiante, no Capítulo 10, discutiremos em detalhe a análise de probabilidade de erro de diferentes sistemas de comunicação digital, considerando diversos modelos de ruído, assim como projetos de sistemas para combater ruídos variados. Por exemplo, o ruído gaussiano pode caracterizar o ruído aleatório de canal proveniente de efeitos térmicos e interferência (*cross talk*) intrassistema. Detectores ótimos podem ser projetados para minimizar a probabilidade de erro devido a ruído gaussiano. Contudo, transientes de comutação, descargas elétricas, chaveamento de carga em linhas de alta tensão, e outros eventos singulares causam pulsos de ruído de alta intensidade e curta duração, que contaminam os pares de cabo que transportam sinais digitais. Esses pulsos, denominados coletivamente **ruído impulsional,** são de difícil combate e constituem a prevalecente fonte de erros em ambientes externos aos sistemas digitais. Erros, portanto, praticamente jamais ocorrem de modo isolado, mas em rajadas de até centenas de cada vez. Para corrigir uma rajada de erros, lançamos mão de **códigos especiais para corrigir erros em rajadas,** descritos no Capítulo 14.

7.6 DIAGRAMAS DE OLHO: UMA FERRAMENTA ÚTIL

Na seção anterior, estudamos o efeito de ruído e ISI de canal na detecção de transmissões digitais. Descrevemos, também, o projeto de equalizadores para compensar distorção de canal e explicamos o processo de extração de temporização. A seguir, apresentamos uma prática ferramenta de engenharia

conhecida como **diagrama de olho**. O diagrama de olho é de fácil geração, e aplicado por engenheiros a sinais recebidos, pois facilita o exame visual da gravidade da ISI, da precisão da extração de temporização, da imunidade ao ruído e de outros fatores importantes.

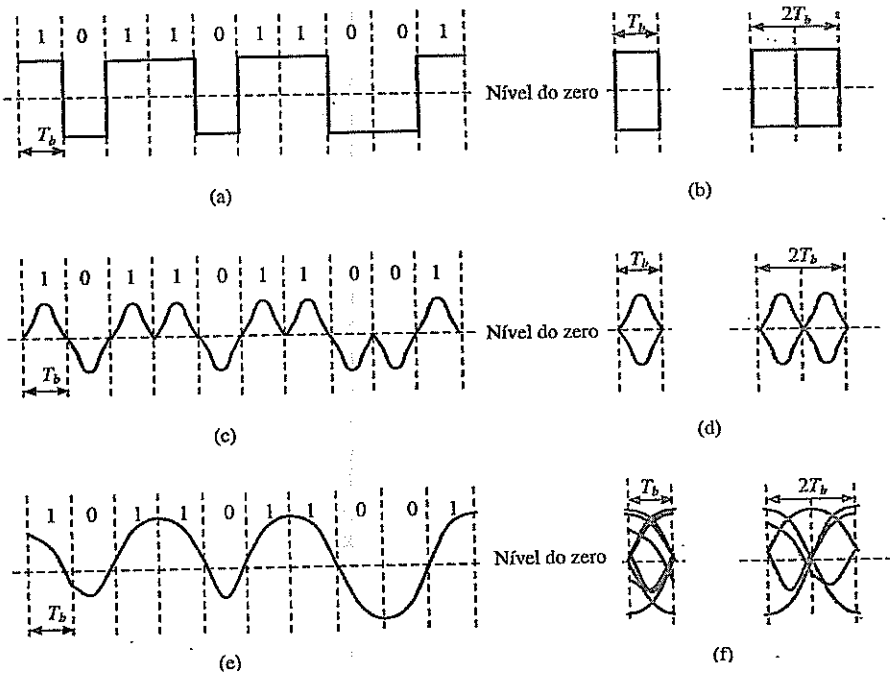
Para gerar um diagrama de olho, precisamos apenas de um osciloscópio básico. Dado um sinal em banda base na saída do canal

$$y(t) = \sum a_k p(t - kT_b)$$

o mesmo pode ser aplicado à entrada vertical do osciloscópio. A base de tempo do osciloscópio é disparada à mesma taxa $1/T_b$ de chegada de pulsos, produzindo uma varredura que dura exatamente T_b segundos, a duração de um símbolo de dado transmitido a_k. O osciloscópio mostra a superposição de diversos traços de comprimento T_b da saída do canal $y(t)$. O que aparece no osciloscópio são simplesmente cortes do sinal de entrada (entrada vertical), feitos a cada T_b e superpostos uns aos outros. O padrão resultante no osciloscópio tem a aparência de um olho humano, daí a denominação diagrama de olho. De modo mais genérico, também podemos aplicar uma varredura temporal com duração de m intervalos de símbolos ou mT_b. O padrão no osciloscópio passa, então, a ser composto por cortes do sinal de entrada (entrada vertical), feitos a cada mT_b e superpostos uns aos outros. O osciloscópio exibirá um diagrama de olho com largura mT_b e tem o formato de m olhos em uma linha horizontal.

A seguir, apresentamos um exemplo. Consideremos a transmissão de um sinal binário por pulsos polares NRZ (Fig. 7.25a). Os correspondentes diagramas de olho são mostrados na Fig. 7.25b para bases de dados T_b e $2T_b$, respectivamente. Nesse exemplo, o canal tem largura de banda infinita para passar o pulso NRZ e não há distorção de canal. Assim, obtemos diagramas de olho com olho(s) totalmente *aberto(s)*. Podemos, ainda, considerar uma saída de canal com o mesmo código de linha polar, mas um formato diferente de pulsos (RZ), como mostrado na Fig. 7.25c. Os resultantes diagramas de olho são mostrados na Fig. 7.25d. Nesse caso, o olho está bem aberto apenas no ponto médio do intervalo de duração de cada pulso. Com extração de temporização adequada, para melhor imunidade ao ruído no ponto de decisão (Seção 7.5.3), o receptor deve amostrar o sinal recebido exatamente no ponto médio, em que o olho está totalmente aberto. O ponto médio do olho representa o melhor instante de amostragem para cada pulso, pois a amplitude é máxima, sem interferência de pulsos vizinhos (ISI zero).

Figura 7.25
Diagrama de olho.



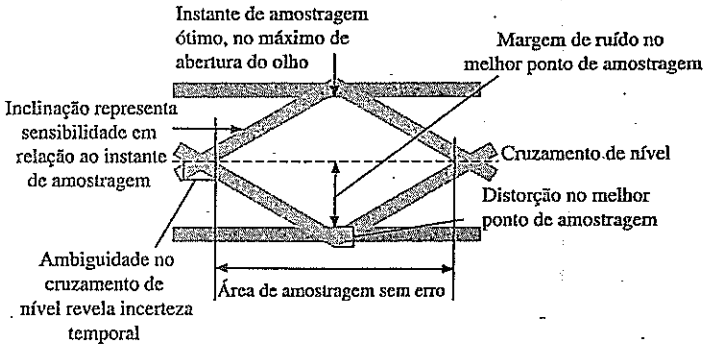


Figura 7.26 Leitura de um diagrama de olho.

Consideremos, agora, um canal com distorção ou com largura de banda finita, ou ambos. Após passagem pelo canal não ideal, o sinal polar NRZ da Fig. 7.25a se torna a forma da onda na Fig. 7.25e. Os pulsos de sinal recebidos não são mais retangulares, mas arredondados, distorcidos e alargados. Os diagramas de olho não são mais totalmente abertos, como mostrado na Fig. 7.25f. Nesse caso, a ISI não é zero. Portanto, em cada traço, os valores do pulso nos respectivos instantes de amostragem se desviarão dos valores ideais por um grau variável, causando um borrão e resultando em um padrão de olho parcialmente fechado.

Na presença de ruído de canal, o olho tenderá a se fechar em todos os casos. Ruído fraco causa um fechamento proporcionalmente pequeno do olho. O limiar de decisão, em relação ao símbolo (**1** ou **0**) transmitido, é o ponto médio do olho.* Observemos que, para ISI zero, o sistema pode tolerar ruído de até metade da abertura vertical do olho. Qualquer valor de ruído maior que esse limite pode causar erro de decisão caso seu sinal seja oposto ao do símbolo de dado. Como a ISI reduz a abertura do olho, também reduz a tolerância ao ruído. O diagrama de olho é usado, ainda, para determinar especificação ótima de derivações do equalizador. Derivações são ajustadas para obter o máximo de aberturas vertical e horizontal de olho.

O diagrama de olho é uma ferramenta muito eficaz para análise de sinais em experimentos em tempo real. Não apenas é de operação simples, mas fornece informação rica e importante sobre a qualidade e suscetibilidade do sinal digital recebido. Do típico diagrama de olho ilustrado na Fig. 7.26, podemos extrair diversas medidas importantes relativas à qualidade do sinal:

- *Ponto de máxima abertura.* O grau de abertura no instante de amostragem e decisão indica a quantidade de ruído que o detector pode tolerar sem cometer um erro. Essa quantidade é conhecida como *margem de ruído*. O instante de máxima abertura do olho indica o instante ótimo de amostragem e decisão.
- *Sensibilidade à incerteza temporal.* A largura do olho indica o intervalo de tempo em que uma decisão correta pode ser feita, sendo desejável ter um olho com máxima abertura horizontal. Se o instante de tomada de decisão se desviar do instante em que o olho tem máxima abertura vertical, a margem de tolerância ao ruído fica reduzida. Isso causa maior probabilidade de erro na detecção de pulsos. A inclinação do olho mostra quão rápido a tolerância ao ruído é reduzida e, em consequência, a sensibilidade da tolerância da decisão ao ruído em relação a variações no instante de amostragem. A inclinação demonstra efeitos da incerteza temporal.
- *Incerteza (temporal) no cruzamento de nível.* Tipicamente, receptores práticos extraem informação de temporização sobre a taxa de pulso e sobre o relógio de amostragem a partir do cruzamento de nível (zero) da forma de onda do sinal recebido. A variação do cruzamento de nível pode ser vista da largura das esquinas do olho. Essa medida fornece informação sobre a incerteza temporal a que o receptor estará sujeito.

*Isso se aplica ao caso de decisão em dois níveis [por exemplo, quando $p(t)$ e $-p(t)$] são usados para **1** e **0**, respectivamente]. No caso de decisão em três níveis (por exemplo, sinalização bipolar), haverá dois limiares.

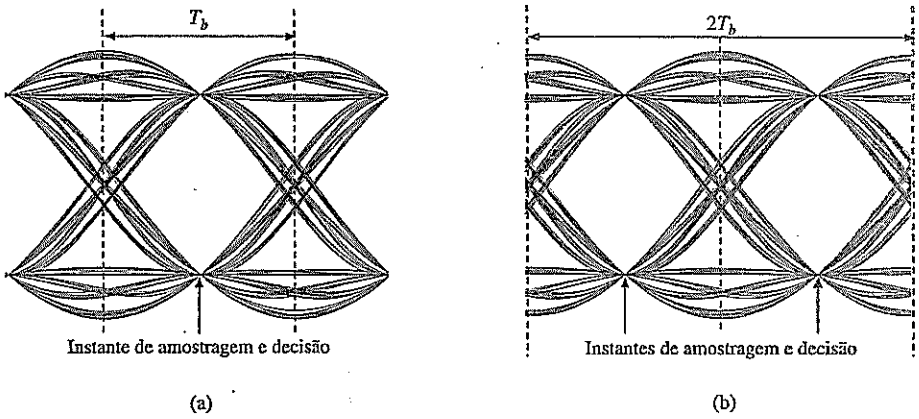


Figura 7.27 Diagramas de olho de um sistema de sinalização polar usando um pulso cosseno levantado, com fator de decaimento $r = 0{,}5$; (a) largura de $2T_b$, 2 períodos de símbolo, e deslocamento temporal $T_b/2$; (b) sem deslocamento temporal.

Por fim, vejamos um exemplo prático de diagrama de olho para uma forma de onda de sinalização polar. Para esse caso, selecionemos um pulso com decaimento cossenoidal que satisfaça o primeiro critério de Nyquist para ISI zero. Escolhamos o fator de decaimento como $r = 0{,}5$. O diagrama de olho é mostrado na Fig. 7.27, com uma base de tempo $2T_b$. Na verdade, para um mesmo sinal, o diagrama de olho pode adquirir aparências distintas, dependendo dos valores dos deslocamentos temporais aplicados (pontos iniciais). A Fig. 7.27a ilustra o diagrama de olho dessa forma de onda de sinalização polar para um deslocamento temporal $T_b/2$, enquanto a Fig. 7.27b mostra o diagrama de olho sem a aplicação de deslocamento temporal. Uma comparação dos dois diagramas deixa claro que os mesmo guardam uma simples relação de deslocamento circular horizontal. Observando a máxima abertura de olho, vemos que esse sinal em banda base tem ISI zero, confirmando a principal característica do pulso cosseno levantado. Como o primeiro critério de Nyquist não impõe nenhuma condição ao cruzamento do zero do pulso, o diagrama de olho indica a probabilidade de ocorrência de incerteza temporal.

7.7 PAM: SINALIZAÇÃO *M*-ÁRIA EM BANDA BASE PARA TAXA DE DADOS MAIS ELEVADA

Independentemente do código de linha usado, modulações binárias em banda base têm algo em comum: todas transmitem um bit de informação no intervalo de T_b segundos, ou seja, a uma taxa de $1/T_b$ bits por segundo. Caso o transmissor deseje enviar seus bits a uma taxa muito mais elevada, T_b deve ser reduzido. Por exemplo, para aumentar a taxa de bits por um fator M, T_b deve ser reduzido pelo mesmo fator; contudo, um alto preço será pago em termos de largura de banda. Como demonstramos na Fig. 7.9, a largura de banda de uma modulação em banda base é proporcional à taxa de pulsos $1/T_b$. A redução de T_b por um fator M certamente aumentará a necessária largura de banda de canal pelo mesmo fator M. Afortunadamente, a redução de T_b não é a única forma de aumentar a taxa de dados. Uma solução muito prática consiste em permitir que cada pulso transporte múltiplos bits. A seguir, explicaremos esse conceito.

Para que cada símbolo de transmissão transporte mais bits no intervalo de tempo T_b, devem existir mais de dois símbolos. Aumentando o número de símbolos para M, asseguramos que a informação transmitida por símbolo também aumenta com M. Por exemplo, quando $M = 4$ (sinalização 4-ária ou quaternária), temos quatro símbolos básicos, ou pulsos, disponíveis para comunicação (Fig. 7.28a). Uma sequência de dois dígitos binários pode ser transmitida por apenas um símbolo 4-ário,

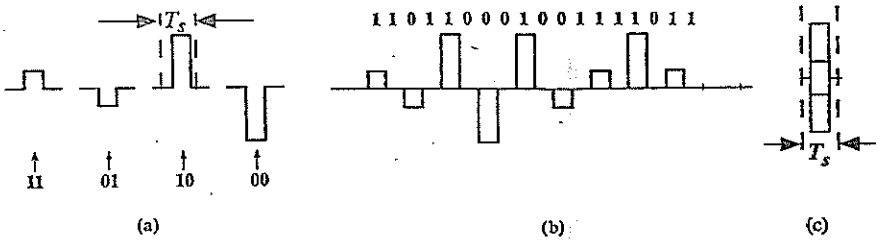


Figura 7.28 Sinalização PAM 4-ária: (a) quatro símbolos RZ; (b) transmissão em banda base; (c) diagrama de olho RZ 4-ário.

pois uma sequência de dois bits pode formar apenas quatro possíveis sequências (**11, 10, 01** e **00**). Como dispomos de quatro símbolos distintos, podemos alocar um deles a cada uma dessas combinações (Fig. 7.28a). Cada símbolo ocupa um intervalo de tempo T_s. Um exemplo de sinalização para uma sequência curta é dada na Fig. 7.28b e o diagrama de olho 4-ário é mostrado na Fig. 7.28c.

Essa sinalização nos permite transmitir cada par de bits por um pulso 4-ário (Fig. 7.28b). Logo, para transmitir n bits, precisamos somente de $(n/2)$ pulsos 4-ários. Isso significa que um símbolo 4-ário pode transmitir a informação de dois dígitos binários. Além disso, como três bits formam $2 \times 2 \times 2 = 8$ combinações, um grupo de três bits pode ser transmitido por um símbolo 8-ário. De modo similar, um grupo de quatro bits podem ser transmitidos por um símbolo 16-ário. Em geral, a informação I_M transmitida por um símbolo M-ário é

$$I_M = \log_2 M \text{ bits} \tag{7.55}$$

Isso significa que podemos aumentar a taxa de transmissão de informação aumentando o valor de M.

Essa sinalização M-ária especial é conhecida como **modulação por amplitude de pulso (PAM –** *pulse amplitude modulation*), pois a informação de dados é transportada pela variação da amplitude do pulso. Vale observar que a modulação por amplitude de pulso é apenas uma das várias possibilidades de sinalização M-ária. Há um número infinito de possibilidades. Todavia, apenas algumas poucas são, de fato, eficazes no combate ao ruído e na redução de largura de banda e consumo de potência. Uma discussão mais detalhada de outros esquemas de sinalização M-ária será apresentada um pouco mais adiante, na Seção 7.9.

Na maioria dos projetos de sistemas, sempre há um preço a ser pago por ganho possível. O preço pago por PAM para aumentar a taxa de dados é o consumo de potência. À medida que M aumenta, a potência transmitida também aumenta com M. Isso se deve ao fato de que, para termos imunidade ao ruído, a mínima separação entre amplitudes de pulsos deve ser comparável à do pulso binário. Portanto, amplitudes de pulso aumentam com M (Fig. 7.28). Podemos mostrar que a potência transmitida aumenta com M^2 (Exercício 7.7-5). Assim, para aumentar a taxa de comunicação por um fator $\log_2 M$, a potência necessária aumenta com M^2. Como a largura de banda de transmissão depende somente da taxa de pulsos, e não das amplitudes dos pulsos, a largura de banda independe de M. Para ilustrar esse ponto, usemos o exemplo de análise de PSD a seguir.

Exemplo 7.4 Determinemos a PSD da sinalização quaternária (4-ária) em banda base dada na Fig. 7.28, considerando que os bits de mensagem **1** e **0** têm igual probabilidade de ocorrência.

O código de linha 4-ário tem quatro símbolos distintos, correspondentes às quatro combinações de dois bits de mensagem. Um possível mapeamento é

$$a_k = \begin{cases} -3 & \text{bits de mensagem } 00 \\ -1 & \text{bits de mensagem } 01 \\ +1 & \text{bits de mensagem } 10 \\ +3 & \text{bits de mensagem } 11 \end{cases} \tag{7.56}$$

Portanto, todos os quatro valores de a_k têm igual probabilidade de ocorrência, cada um com uma chance de 1 em 4. Recordemos que

$$R_0 = \lim_{N\to\infty} \frac{1}{N} \sum_k a_k^2$$

No somatório, 1/4 de a_k será ± 1 e ± 3. Logo,

$$R_0 = \lim_{N\to\infty} \frac{1}{N} \left[\frac{N}{4}(-3)^2 + \frac{N}{4}(-1)^2 + +\frac{N}{4}(1)^2 + +\frac{N}{4}(3)^2 \right] = 5$$

Para $n > 0$, devemos determinar

$$R_n = \lim_{N\to\infty} \frac{1}{N} \sum_k a_k a_{k+n}$$

Para calcular esse valor médio, construamos uma tabela com todos os possíveis valores do produto $a_k a_{k+n}$:

Possíveis valores de $a_k a_{k+n}$

a_{k+n} \ a_k	-3	1	$+1$	$+3$
-3	9	3	-3	-9
-1	3	1	-1	-3
$+1$	-3	-1	1	3
$+3$	-9	-3	3	9

Da listagem anterior de todos os possíveis produtos $a_k a_{k+n}$, vemos que cada produto no somatório $a_k a_{k+n}$ pode assumir qualquer um dos seguintes seis valores: ± 1, ± 3, ± 9. Observamos que (± 1, ± 9) têm igual probabilidade de ocorrência (1 em 8). Por sua vez, ± 3 também têm igual probabilidade de ocorrência (1 em 4). Assim, obtemos

$$R_n = \lim_{N\to\infty} \frac{1}{N} \left[\frac{N}{8}(-9) + \frac{N}{8}(+9) + \frac{N}{8}(-1) + \frac{N}{8}(+1) + \frac{N}{4}(-3) + \frac{N}{4}(+3) \right] = 0$$

Logo,

$$S_x(f) = \frac{5}{T_s} \quad \Longrightarrow \quad S_y(f) = \frac{5}{T_s} |P(f)|^2$$

Portanto, o código de linha M-ário gera a mesma forma de PSD que a sinalização polar binária. A única diferença é que utiliza 5 vezes a potência de sinal original.

Embora a maioria da rede terrestre de telefonia digital empregue codificação binária, a porção da malha de assinante da rede digital de serviços integrados (ISDN) usa o código quaternário, 2B1Q, semelhante à Fig. 7.28a. São usados pulsos NRZ para transmitir 160 kbit/s de dados a uma taxa de **baud** (taxa de pulsos ou de símbolos) de 80 kbit/s. Dos vários códigos de linha examinados pelo comitê de padronização de ANSI, 2B1Q é o que provê maior redução da taxa de baud na planta de cabos locais, ambiente ruidoso e sujeito a interferências.

Formatação de Pulso e Diagramas de Olho em PAM: Neste caso, podemos usar pulsos que satisfaçam o critério de Nyquist, pois tais pulsos têm zero ISI nos pontos de amostragem e, portanto, suas amplitudes podem ser detectadas corretamente se forem amostradas nos centros dos pulsos. Também podemos usar ISI controlada (sinalização de resposta parcial) para sinalização M-ária.[8]

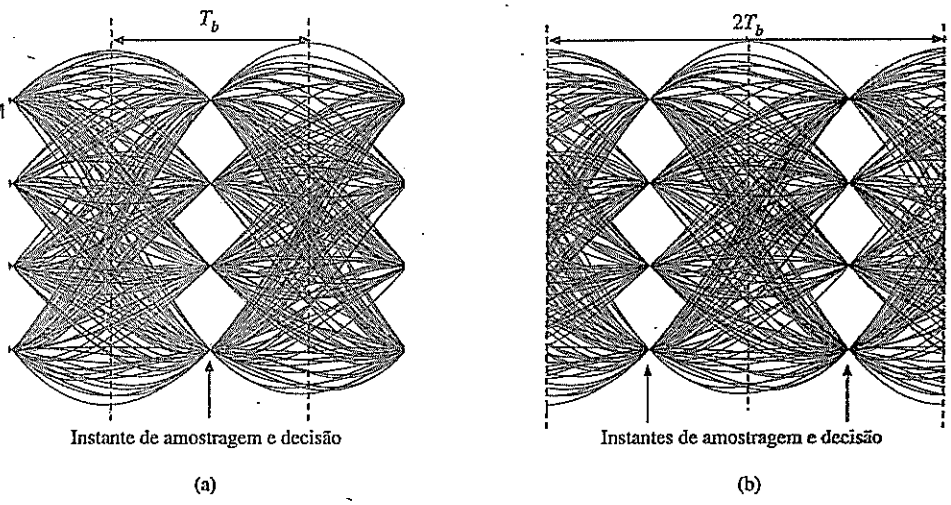


Figura 7.29 Diagramas de olho de um sistema de sinalização PAM 4-ária usando pulso cosseno levantado com fator de decaimento $r = 0{,}5$: (a) com dois períodos de símbolo $2T_b$ e deslocamento temporal de $T_b/2$; (b) sem deslocamento temporal.

Diagramas de olho também podem ser gerados para PAM M-ária usando o mesmo método empregado em modulações binárias. Devido à sinalização em múltiplos níveis, o diagrama de olho deve ter M níveis nos instantes de amostragem ótima, mesmo quando a ISI for zero. Aqui, geramos o exemplo prático de diagrama de olho para um sinal PAM de quatro níveis que usa o mesmo pulso cosseno levantado com fator de decaimento $r = 0{,}5$, como no diagrama de olho na Fig. 7.27. Os correspondentes diagramas de olho com deslocamentos temporais de $T_b/2$ e 0 são dados nas Fig. 7.29 a e b, respectivamente. Mais uma vez, nenhuma ISI é observada nos instantes de amostragem. Os diagramas de olho mostram claramente quatro valores de sinais igualmente separados, sem ISI nos pontos de amostragem ótima.

7.8 SISTEMAS DIGITAIS COM PORTADORA

Até aqui, discutimos sistemas digitais em banda base, em que sinais são transmitidos diretamente, sem nenhum deslocamento em frequência. Como sinais em banda base têm muita potência nas frequências baixas, são adequados à transmissão por um par de fios ou por um cabo coaxial. Uma boa parte da comunicação moderna é conduzida dessa forma. Contudo, sinais em banda base não podem ser transmitidos em enlaces de rádio ou de satélites, pois isso exigiria antenas demasiadamente grandes, e impraticáveis, para irradiar eficientemente o espectro de baixa frequência desses sinais. Portanto, para essas aplicações, o espectro do sinal deve ser deslocado a uma faixa de frequências altas. Um deslocamento do espectro para frequências mais altas também se faz necessário para transmitir várias mensagens simultaneamente, com compartilhamento da grande largura de banda de um meio de transmissão. Como vimos no Capítulo 4, o espectro de um sinal pode ser deslocado para uma frequência mais alta usando o sinal em banda base para modular uma senoide (portadora) de alta frequência.

Na transmissão e recepção de sinais digitais com portadora, precisamos de um modulador e de um demodulador para transmitir e receber dados. Para comunicação em duas direções (duplex), esses dois dispositivos, modulador e demodulador, são comumente empacotados em uma unidade que recebe o nome de **modem**.

7.8.1 Modulações Binárias Básicas com Portadora

Há duas formas básicas de modulação de portadora: modulação em amplitude e modulação em ângulo. Na modulação em amplitude, a amplitude da portadora é variada em proporção ao sinal modulante (ou seja, o sinal em banda base). Isso é ilustrado na Fig. 7.30. Uma portadora não modulada cos $\omega_c t$ é

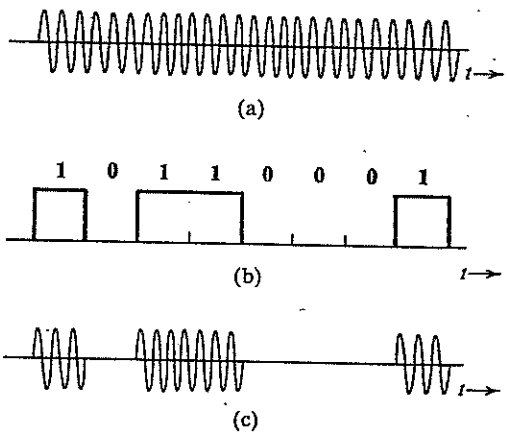


Figura 7.30
(a) Portadora cos $\omega_c t$.
(b) Sinal modulante $m(t)$.
(c) ASK: sinal modulado $m(t)$ cos $\omega_c t$.

mostrada na Fig. 7.30a. O sinal on-off em banda base $m(t)$ (sinal modulante) é mostrado na Fig. 7.30b. Segundo a Eq. (7.1), o sinal modulante pode ser escrito como

$$m(t) = \sum a_k p(t - kT_b), \qquad \text{em que} \quad p(t) = \Pi\left(\frac{t - T_b/2}{T_b}\right)$$

O código de linha $a_k = 0, 1$ é on-off. Quando a amplitude da portadora é variada em proporção a $m(t)$, podemos escrever o sinal modulado com portadora como

$$\varphi_{\text{ASK}}(t) = m(t) \cos \omega_c t \tag{7.57}$$

que é mostrado na Fig. 7.30c. Reparemos que o sinal modulado ainda é um sinal on-off. Esse esquema de modulação para transmissão de dados binários é conhecido como **chaveamento on-off** (OOK – *on-off keying*) ou **modulação por chaveamento de amplitude** (ASK – *amplitude shift keying*).

O sinal em banda base $m(t)$ pode, obviamente, usar um pulso $p(t)$ diferente do retangular mostrado no exemplo da Fig. 7.30. Isso gerará um sinal ASK que não tem amplitude constante durante a transmissão de **1** ($a_k = 1$).

Se o sinal em banda base $m(t)$ fosse polar (Fig. 7.31a), o correspondente sinal modulado $m(t)$ cos $\omega_c t$ teria a aparência ilustrada na Fig. 7.31b. Nesse caso, se $p(t)$ for o pulso básico, transmitimos **1** por um pulso $p(t)$ cos $\omega_c t$ e **0**, por $-p(t)$ cos $\omega_c t = p(t)$ cos $(\omega_c t + \pi)$. Os dois pulsos, portanto, estão defasados de π radianos. A informação reside na fase ou no sinal do pulso. Por essa razão, esse esquema é conhecido como **chaveamento por deslocamento de fase (PSK** – *phase shift keying*). Reparemos que a transmissão ainda é polar. Na verdade, assim como ASK, o sinal PSK modulado com portadora tem a mesma forma

$$\varphi_{\text{PSK}}(t) = m(t) \cos \omega_c t \qquad m(t) = \sum a_k p(t - kT_b) \tag{7.58}$$

com a diferença que o código de linha é polar, $a_k = \pm 1$.

Quando dados são transmitidos por meio da variação da frequência, temos o caso de **chaveamento por deslocamento de frequência (FSK** – *frequency shift keying*), como ilustrado na Fig. 7.31c. Um **0** é transmitido por um pulso de frequência ω_{c_0} e um **1**, por um pulso de frequência ω_{c_1}. A informação sobre o dado transmitido reside na frequência portadora. O sinal FSK pode ser visto como a soma de dois sinais ASK entrelaçados, um com frequência portadora ω_{c_0} e outro com frequência portadora ω_{c_1}. Podemos usar a expressão de ASK binário da Eq. (7.57) e escrever o sinal FSK como

$$\varphi_{\text{FSK}}(t) = \sum a_k p(t - kT_b) \cos \omega_{c_1} t + \sum (1 - a_k) p(t - kT_b) \cos \omega_{c_0} t$$

em que $a_k = 0, 1$ é on-off. Assim, o sinal FSK é uma superposição de dois sinais AM com diferentes frequências portadoras e amplitudes diferentes, mas complementares.

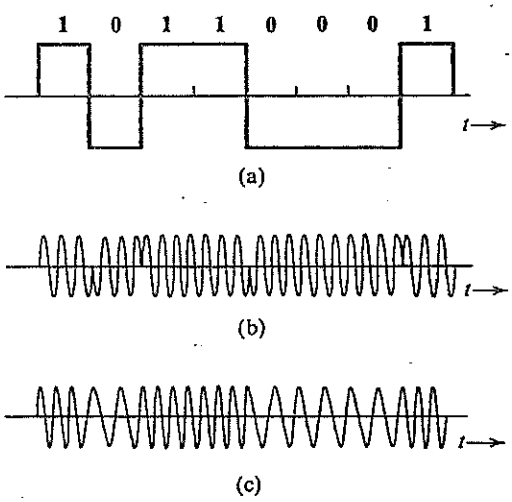


Figura 7.31 (a) Sinal modulante $m(t)$. (b) PSK: sinal modulado $m(t)$ $\cos\omega_c t$. (c) FSK: sinal modulado.

Na prática, a ASK como um esquema on-off é comumente usado em sistemas de comunicação por fibra óptica, na forma de modulação da intensidade do laser. A PSK é comumente utilizado em sistemas de comunicação por satélite e também foi usado nos primeiros modems telefônicos (2400 e 4800 bit/s). Quanto à FSK, em 1962, a AT&T desenvolveu um dos primeiros modems para linha telefônica, chamado 103A, e usava a FSK para transmitir 300 bit/s em duas frequências, 1070 e 1270 Hz; a recepção era FSK em 2025 e 2225 Hz.

7.8.2 PSD de Modulação Digital com Portadora

Acabamos de mostrar que os esquemas de modulação binária com portadora ASK, PSK e FSK podem ser escritos na forma $m(t)\cos\omega_c t$. Para determinar a PSD dos sinais ASK, PSK e FSK, é conveniente que, primeiro, determinemos a relação entre a PSD de $m(t)$ e a PSD do sinal modulado

$$\varphi(t) = m(t)\cos\omega_c t$$

Recordemos, da Eq. (3.80), que a PSD de $\varphi(t)$ é

$$S_\varphi(f) = \lim_{T\to\infty} \frac{|\Psi_T(f)|^2}{T}$$

em que Ψ_T é a transformada de Fourier do sinal truncado

$$\begin{aligned}\varphi_T(t) &= \varphi(t)[u(t+T/2) - u(t-T/2)]\\ &= m(t)[u(t+T/2) - u(t-T/2)]\cos\omega_c t\\ &= m_T(t)\cos\omega_c t\end{aligned} \tag{7.59}$$

Aqui, $m_T(t)$ é o sinal em banda base truncado, com transformada de Fourier $M_T(f)$. Aplicando a propriedade de translação em frequência [Eq. (3.36)], temos

$$\Psi_T(f) = \frac{1}{2}[M_T(f-f_c) + M_T(f+f_c)]$$

Por conseguinte, a PSD do sinal modulado com portadora $\varphi(t)$ é

$$S_\varphi(f) = \lim_{T\to\infty} \frac{1}{4}\frac{|M_T(f+f_c) + M_T(f-f_c)|^2}{T}$$

Como $M(f)$ é um sinal em banda base, $M_T(f+f_c)$ e $M_T(f-f_c)$ têm sobreposição zero à medida que $T \to \infty$, desde que f_c seja maior que a largura de banda de $M(f)$. Portanto, concluímos que

$$\begin{aligned} S_\varphi(f) &= \lim_{T\to\infty} \frac{1}{4}\left[\frac{|M_T(f+f_c)|^2}{T} + \frac{|M_T(f-f_c)|^2}{T}\right] \\ &= \frac{1}{4}S_M(f+f_c) + \frac{1}{4}S_M(f-f_c) \end{aligned} \tag{7.60}$$

Em outras palavras, para uma frequência portadora escolhida de modo apropriado, a modulação causa um deslocamento na PSD do sinal em banda base.

O sinal ASK na Fig. 7.30c se encaixa nesse modelo, com $m(t)$ sendo um sinal on-off (usando um pulso NRZ ou de largura completa). Portanto, a PSD do sinal ASK é igual à de um sinal on-off (Fig. 7.4b) deslocada para $\pm f_c$, como mostrado na Fig. 7.32a. Recordemos que, usando um pulso retangular de largura completa $p(t)$,

$$P\left(\frac{n}{T_b}\right) = 0 \qquad n = \pm 1, \pm 2, \ldots$$

Nesse caso, a PSD do sinal on-off na Fig. 7.30b não tem componentes discretas, exceto em dc. Em consequência, o espectro ASK tem componentes discretas apenas em ω_c.

O sinal PSK também se encaixa nessa descrição de modulação, em que $m(t)$ é um sinal polar que usa um pulso NRZ de largura completa. Portanto, a PSD de um sinal PSK é igual à do sinal polar em banda base deslocada para $\pm\omega_c$, como mostrado na Fig. 7.32b. Reparemos que essa PSD tem a mesma forma (com diferente fator de escala) que a PSD do sinal ASK, sem suas componentes discretas.

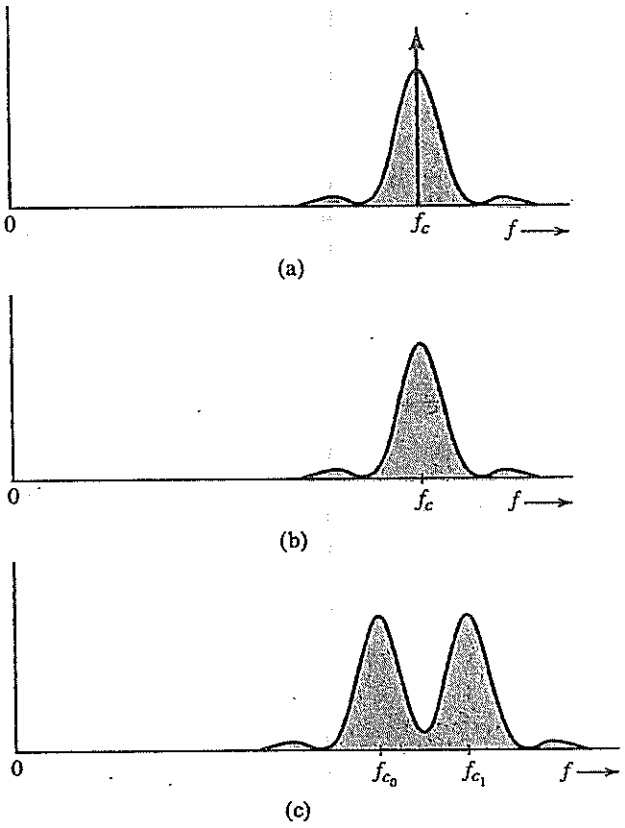


Figura 7.32 PSD de (a) ASK, (b) PSK e (c) FSK.

Por fim, mostramos que o sinal FSK pode ser visto como a soma de dois sinais ASK entrelaçados usando o pulso de largura completa. Portanto, o espectro FSK é a soma de dois espectros ASK nas frequências ω_{c_0} e ω_{c_1}, como mostrado na Fig. 7.32c. Podemos mostrar que, com escolha adequada de ω_{c_0} e ω_{c_1}, e com manutenção da continuidade de fase durante a translação em frequência, as componentes discretas podem ser eliminadas em ω_{c_0} e ω_{c_1}. Com isso, nenhuma componente discreta aparece nesse espectro. É importante ressaltar que a largura de banda de FSK é maior que às de ASK ou PSK.

Como observado anteriormente, a sinalização polar é o esquema mais eficiente do ponto de vista de potência. PSK, por ser polar, requer 3 dB de potência menor que ASK (ou FSK), para uma mesma imunidade ao ruído, ou seja, para uma mesma probabilidade de erro na detecção de pulsos.

É claro que podemos também modular o esquema bipolar ou qualquer outro discutido anteriormente. Reparemos que, na Fig. 7.30 e na Fig. 7.31, usamos pulso retangular NRZ apenas para facilitar a ilustração. Na prática, pulsos em banda base podem ser formatados especificamente para eliminar a ISI.

7.8.3 Relações entre Modulações Analógica e Digital com Portadora

Existe uma relação natural e clara entre ASK e AM, pois a mensagem de informação é diretamente refletida na variação da amplitude do sinal modulado. Devido à sua amplitude não negativa, a ASK é, em essência, um sinal AM com índice de modulação $\mu = 1$. Existe uma conexão similar entre FSK e FM. FSK é simplesmente um sinal FM com um número limitado de frequências instantâneas.

A relação entre PSK e modulação em ângulo é um pouco mais sutil. Para PSK, o sinal modulado pode ser escrito como

$$\varphi_{\text{PSK}}(t) = A\cos(\omega_c t + \theta_k) \qquad kT_b \le t < kT_b + T_b$$

Esse sinal pode, portanto, ser relacionado a PM. Contudo, um exame mais detalhado do sinal PSK revela que, devido à fase constante θ_k, sua frequência instantânea não varia. Na verdade, fazendo $a_k = A\cos\theta_k$ e $b_k = -A\,\text{sen}\,\theta_k$, podemos reescrever o sinal PSK como

$$\begin{aligned}\varphi_{\text{PSK}}(t) &= A\cos\theta_k\cos\omega_c t - A\,\text{sen}\,\theta_k\,\text{sen}\,\omega_c t \\ &= a_k\cos\omega_c t + b_k\,\text{sen}\,\omega_c t \qquad kT_b \le t < kT_b + T_b\end{aligned} \tag{7.61}$$

Da Eq. (7.61), reconhecemos uma forte semelhança com a representação do sinal QAM dada na Seção 4.4. Portanto, uma modulação digital PSK guarda uma relação próxima com o sinal analógico QAM. Em particular, para PSK binário, $\theta = 0, \pi$. Assim, PSK binário pode ser escrito como

$$\pm A\cos\omega_c t$$

Isso é, efetivamente, uma manifestação digital da modulação em amplitude DSB-SC. Na verdade, como veremos mais adiante, se permitirmos que a_k assuma valores em múltiplos níveis e tomarmos $b_k = 0$, podemos gerar outra modulação digital com portadora conhecida como modulação por amplitude de pulso (ou PAM), que pode transportar múltiplos bits durante cada intervalo de tempo T_b de modulação.

Como estudado no Capítulo 4, a modulação em amplitude DSB-SC tem melhor eficiência de potência que AM. Portanto, a PSK binária tem melhor eficiência de potência que a ASK. Em termos de uso de largura de banda, podemos ver, de suas relações com modulações analógicas, que a ASK e a PSK têm idênticas larguras de banda, enquanto a FSK requer maior largura de banda. Essas observações corroboram, de modo intuitivo, os resultados de PSD na Fig. 7.32.

7.8.4 Demodulação

A demodulação de sinais modulados digitais é similar à de sinais modulados analógicos. Devido às relações entre ASK e AM, entre FSK e FM e entre PSK e QAM (ou DSB-SC AM), as diferentes técnicas empregadas para modulações analógicas podem ser diretamente aplicadas aos equivalentes digitais.

Detecção ASK

Assim como a AM, a ASK (Fig. 7.30c) pode ser demodulada de forma coerente (para detecção síncrona) ou incoerente (para detecção de envelope). O detector coerente requer equipamento mais elaborado e tem desempenho superior, especialmente quando a potência de sinal (e, portanto, a SNR) é baixa. Para SNR mais alta, o detector de envelope tem desempenho quase tão bom como o detector coerente. Por conseguinte, a detecção coerente não é muito utilizada para ASK, pois violaria o propósito do esquema (simplicidade de detecção). Se pudermos tirar proveito de um detector síncrono, é preferível o uso de PSK, que tem melhor eficiência de potência que ASK.

Detecção FSK

Novamente, o sinal FSK binário pode ser visto como dois sinais ASK entrelaçados, com frequências portadoras ω_{c_0} e ω_{c_1} (Fig. 7.32c). Portanto, a FSK pode ser detectada de forma coerente ou incoerente. Na detecção incoerente, o sinal que chega é aplicado a um par de filtros sintonizados em ω_{c_0} e ω_{c_1}. Cada filtro é seguido por um detector de envelope (Fig. 7.33a). As saídas dos dois detectores de envelope são amostradas e comparadas. Se um **0** for transmitido por um pulso de frequência ω_{c_0}, esse pulso aparecerá na saída do filtro sintonizado em ω_{c_0}. Praticamente nenhum sinal aparecerá na saída do filtro sintonizado em ω_{c_1}. Assim, a amostra da saída do detector de envelope que segue o filtro ω_{c_0} será maior que a amostra da saída do detector de envelope que segue o filtro ω_{c_1}, e o receptor decide que um **0** foi transmitido. No caso de um **1**, ocorre o oposto.

É óbvio que a FSK também pode ser detectada coerentemente com a geração de duas frequências de referência ω_{c_0} e ω_{c_1}, para os dois demoduladores, para a demodulação do sinal recebido e a

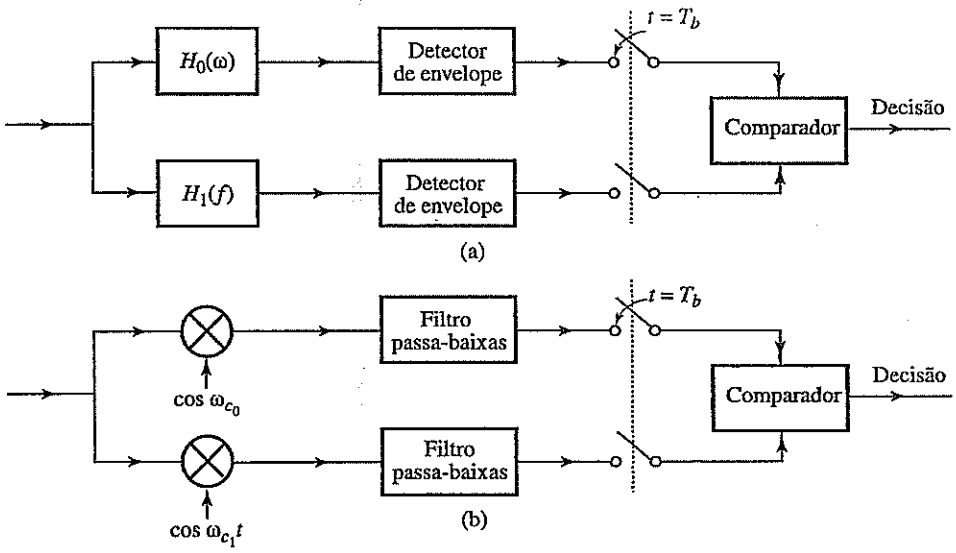


Figura 7.33 (a) Detecção incoerente de FSK. (b) Detecção coerente de FSK.

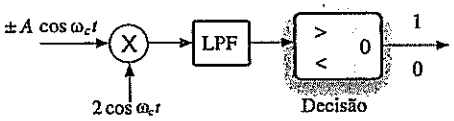


Figura 7.34 Detector PSK binário coerente (similar a um demodulador DSB-SC).

subsequente comparação das saídas dos dois demoduladores, como ilustrado na Fig. 7.33b. Assim, o detector FSK coerente deve gerar duas portadoras em sincronia com as portadoras de modulação. Novamente, esse demodulador complexo viola o propósito da FSK, projetado principalmente para detecção mais simples e não coerente. Na prática, a detecção FSK coerente não é utilizada.

Detecção PSK

Em PSK binário, um **1** é transmitido por um pulso $A \cos \omega_c t$ e um **0**, por um pulso $-A \cos \omega_c t$ (Fig. 7.31b). A informação em sinais PSK reside na fase da portadora. Assim como em DSB-SC, esses sinais não podem ser demodulados por detecção de envelope, pois o envelope é o mesmo para **1** e **0** (Fig. 7.31b). O detector coerente da modulação PSK binária é mostrado na Fig. 7.34. A detecção coerente é similar à usada para sinais analógicos. Métodos de aquisição de portadora foram discutidos na Seção 4.8.

PSK Diferencial

Embora a detecção de envelope não possa ser usada para PSK, ainda é possível explorar o número finito de valores de fase de modulação para detecção incoerente. De fato, sinais PSK podem ser demodulados de forma incoerente por meio de um engenhoso método conhecido como **PSK diferencial** ou DPSK. O princípio da detecção diferencial é que o receptor detecte a mudança de fase relativa entre sucessivas fases moduladas θ_k e θ_{k-1}. Como os valores de fase em PSK são finitos (iguais a 0 e π em PSK binário), o transmissor pode codificar o dado de informação na diferença de fase $\theta_k - \theta_{k-1}$. Por exemplo, uma diferença de fase zero representa **0**, enquanto uma diferença de fase π significa **1**.

Essa técnica é conhecida como **codificação diferencial** (antes da modulação). Em um código diferencial, um **0** é codificado pelo mesmo pulso usado para codificar o bit de dado anterior (sem transição), e **1** é codificado pelo negativo do pulso usado para codificar o bit de dado anterior (transição). A codificação diferencial é de simples implementação, como mostrado na Fig. 7.35b. Reparemos que a adição é em módulo 2. O sinal codificado é mostrado na Fig. 7.35b. Desse modo, uma transição na sequência de pulsos do código de linha indica **1** e nenhuma transição, **0**. O sinal modulado consiste em pulsos

$$A \cos(\omega_c t + \theta_k) = \pm A \cos \omega_c t$$

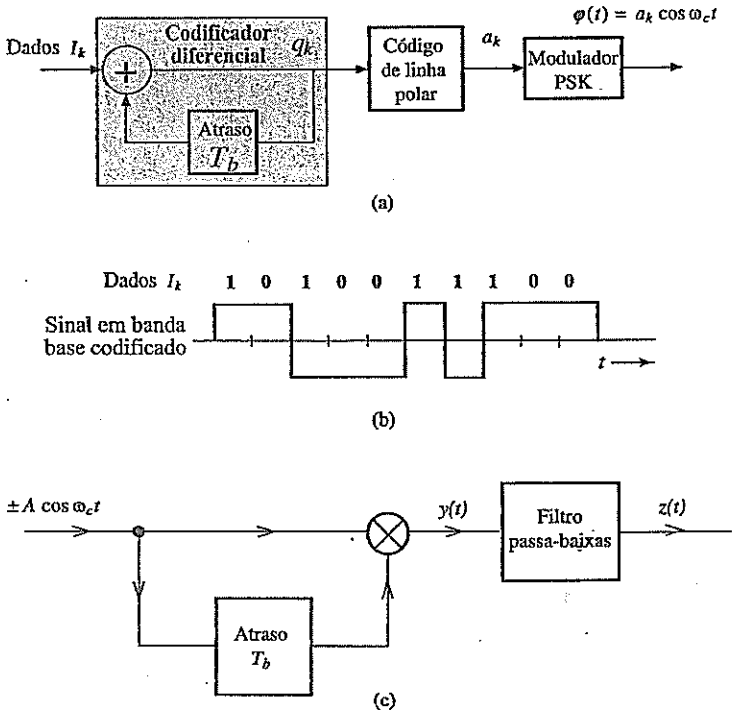


Figura 7.35 (a) Codificação diferencial; (b) sinal codificado; (c) receptor PSK diferencial.

Se o bit de dado for **0**, o pulso corrente e o anterior têm a mesma polaridade ou fase; os dois pulsos são iguais a $A \cos \omega_c t$ ou iguais a $-A \cos \omega_c t$. Se o bit de dado for **1**, o pulso corrente e o anterior têm polaridades ou fases opostas; se o pulso corrente for $A \cos \omega_c t$, o anterior é $-A \cos \omega_c t$, ou vice-versa.

Na demodulação da DPSK (Fig. 7.35c), evitamos a geração de uma portadora local observando que o próprio sinal modulado recebido é uma portadora ($\pm A \cos \omega_c t$), com uma possível ambiguidade de sinal. Para demodulação, no lugar da portadora, podemos usar o sinal recebido atrasado por T_b (um intervalo de bit). Se o pulso recebido for igual ao anterior, o produto dos dois será $y(t) = A^2 \cos^2 \omega_c t = (A^2/2)(1 + \cos 2\omega_c t)$, e a saída do filtro passa-baixas, $z(t) = A^2/2$. Imediatamente, detectamos o pulso corrente como um **0**. Se o pulso recebido e o anterior tiverem polaridades opostas, $y(t) = -A^2 \cos^2 \omega_c t$ e $z(t) = -A^2/2$, e o bit corrente é detectado como **1**. A Tabela 7.3 ilustra um exemplo específico de codificação e decodificação.

Tabela 7.3
Codificação diferencial e detecção de DPSL binário

Tempo k	0	1	2	3	4	5	6	7	8	9	10
I_k		1	0	1	0	0	1	1	1	0	0
q_k	0	1	1	0	0	0	1	0	1	1	1
Código de linha a_k	–1	1	1	–1	–1	–1	1	–1	1	1	1
θ_k	π	0	0	π	π	π	0	π	0	0	0
$\theta_k - \theta_{k-1}$		π		π			π	π	π	0	0
Bits detectados		1	0	1	0	0	1	1	1	0	0

Em termos de complexidade de demodulação, ASK, FSK e DPSK podem ser todos detectados incoerentemente, sem uma portadora síncrona no receptor. A PSK, por sua vez, deve ser detectada coerentemente. Detecção incoerente, no entanto, tem um preço em termos de imunidade ao ruído. Do ponto de vista de imunidade ao ruído, a PSK coerente é superior a todos os outros esquemas. A PSK também requer menor largura de banda do que a FSK (Fig. 7.32). Uma discussão quantitativa sobre esse tema será feita no Capítulo 10.

7.9 MODULAÇÃO DIGITAL *M*-ÁRIA COM PORTADORA

As modulações digitais binárias ASK, FSK e PSK com portadora transmitem um bit de informação por intervalo de T_b segundos, o que corresponde a uma taxa de bits de $1/T_b$ bit/s. Assim como no caso da transmissão de sinais digitais em banda base, maior taxa de transmissão de bits pode ser alcançada com a redução de T_b ou com aplicação de sinalização *M*-ária; a primeira opção requer mais largura de banda e a segunda, mais potência. Na maioria dos sistemas de comunicação, largura de banda é estritamente limitada. Assim, para conservar largura de banda, uma forma eficaz de aumentar a taxa de transmissão de dados consiste em generalizar a modulação binária com o emprego de sinalização *M*-ária. Especificamente, podemos aplicar modulações ASK de *M* níveis, FSK de *M* frequências e PSK de *M* fases.

ASK *M*-ária e Detecção Incoerente

A ASK *M*-ária é uma generalização muito simples da modulação ASK binária. Em vez de enviar apenas

$$\varphi(t) = 0 \text{ para } \mathbf{0} \quad \text{e} \quad \varphi(t) = A \cos \omega_c t \text{ para } \mathbf{1}$$

a modulação ASK *M*-ária pode enviar $\log_2 M$ bits por vez ao transmitir, por exemplo,

$$\varphi(t) = 0, A \cos \omega_c t, 2A \cos \omega_c t, \ldots, (M-1)A \cos \omega_c t$$

Esse ainda é um sinal AM que usa M amplitudes distintas e um índice de modulação $\mu = 1$. A largura de banda é igual à do sinal ASK binário, enquanto a potência cresce proporcionalmente com M^2. A demodulação pode ser feita por detecção de envelope ou por detecção coerente.

FSK M-ária e Sinalização Ortogonal

A FSK M-ária é gerada com a seleção de uma senoide do conjunto $\{A \cos 2\pi f_i t, i = 1,..., M\}$ para transmitir um padrão particular de $\log_2 M$ bits. Em geral, com FSK, podemos projetar um incremento de frequência δf e definir

$$f_m = f_1 + (m-1)\delta f \qquad m = 1, 2, \ldots, M$$

Para essa modulação FSK com igual separação entre frequências, o desvio de frequência (na análise do sinal FM) é

$$\Delta f = \frac{f_M - f_1}{2} = \frac{1}{2}(M-1)\delta f$$

Fica claro que a seleção do conjunto de frequências $\{f_i\}$ determina o desempenho e a largura de banda da modulação FSK. Se δf for escolhido demasiadamente grande, o sinal FSK M-ário usará excessiva largura de banda. Se, por outro lado, δf for escolhido demasiadamente pequeno, no intervalo de tempo de T_b segundos praticamente não haverá distinção entre diferentes símbolos FSK e o receptor será incapaz de distinguir os símbolos de modo confiável. Assim, grande valor de δf leva a desperdício de largura de banda, enquanto pequeno δf favorece a suscetibilidade a erros de detecção devido ao ruído de transmissão e à interferência.

O projeto da modulação FSK M-ária requer a determinação de um valor suficientemente pequeno de δf, de modo que cada símbolo FSK $A \cos \omega_c t$ seja distinguível de todos os outros símbolos FSK. Uma solução para esse problema de projeto de sinais FSK pode ser encontrada na discussão do espaço ortogonal de sinais na Seção 2.6.2. Se pudermos projetar símbolos FSK que sejam ortogonais em T_b com a seleção de um pequeno valor para δf (ou Δf), os sinais FSK serão verdadeiramente distintos no intervalo T_b e o consumo de largura de banda será pequeno.

Para determinar o mínimo valor de δf que leva a um conjunto ortogonal de sinais FSK, a condição de ortogonalidade, segundo a Seção 2.6.2, requer que

$$\int_0^{T_b} A\cos(2\pi f_m t)\, A\cos(2\pi f_n t)\, dt = 0 \qquad m \neq n \tag{7.62}$$

Podemos usar essa condição para determinar o mínimo δf. Primeiro, calculemos

$$\begin{aligned}\int_0^{T_b} A\cos(2\pi f_m t)\, A\cos(2\pi f_n t)\, dt &= \frac{A^2}{2}\int_0^{T_b} [\cos 2\pi(f_m + f_n)t + \cos 2\pi(f_m - f_n)t]\, dt \\ &= \frac{A^2}{2}T_b \frac{\operatorname{sen} 2\pi(f_m + f_n)T_b}{2\pi(f_m + f_n)T_b} + \frac{A^2}{2}T_b \frac{\operatorname{sen} 2\pi(f_m - f_n)T_b}{2\pi(f_m - f_n)T_b}\end{aligned} \tag{7.63}$$

Como, em modulações práticas, $(f_m + f_n)T_b$ é muito grande (em geral, não menor que 10^3), o primeiro termo na Eq. (7.63) é efetivamente zero e desprezível. Assim, a condição de ortogonalidade se reduz à condição que, para quaisquer inteiros $m \neq n$,

$$\frac{A^2}{2}\frac{\operatorname{sen} 2\pi(f_m - f_n)T_b}{2\pi(f_m - f_n)} = 0$$

Como $f_m = f_1 + (m-1)\delta f$, para mútua ortogonalidade, temos

$$\operatorname{sen}[2\pi(m-n)\delta f T_b] = 0 \qquad m \neq n$$

Dessa condição, fica claro que o menor δf que satisfaz a condição de mútua ortogonalidade é

$$\delta f = \frac{1}{2T_b}\ \text{Hz}$$

Essa escolha de mínima separação de frequência é conhecida como FSK *de mínimo deslocamento*. Por formar um conjunto ortogonal de símbolos, essa escolha também é conhecida como sinalização ortogonal.

Podemos descrever a modulação FSK *de mínimo deslocamento* geometricamente, aplicando o conceito de funções de base ortogonais da Seção 2.6. Sejam

$$\psi_i(t) = \sqrt{\frac{2}{T_b}} \cos 2\pi \left(f_1 + \frac{i-1}{2T_b}\right) t \qquad i = 1, 2, \ldots, M$$

Podemos comprovar com facilidade que

$$\int_0^{T_b} \psi_m(t)\psi_n(t)\, dt = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases}$$

Assim, cada símbolo FSK pode ser escrito como

$$A \cos 2\pi f_m t = A\sqrt{\frac{T_b}{2}}\psi_m(t) \qquad m = 1, 2, \ldots, M$$

A relação geométrica entre os dois símbolos FSK para $M = 2$ é capturada na Fig. 7.36.

A demodulação de sinais FSK M-ários segue os mesmos princípios que a demodulação de sinais FSK binários. Com a generalização dos demoduladores FSK binários na Fig. 7.33, podemos aplicar um banco de M detectores coerentes ou incoerentes para sinais FSK M-ários antes de tomar uma decisão com base no ramo detector mais intenso.

Anteriormente, na análise da PSD de modulações em banda base, mostramos que, no intervalo de símbolo T_b, a largura de banda de um sinal digital em banda base podia ser aproximada por $1/T_b$. Portanto, para FSK de deslocamento mínimo, $\Delta f = (M-1)/(4T_b)$ e, segundo a regra de Carson, a largura de banda é, aproximadamente,

$$2(\Delta f + B) = \frac{M-3}{2T_b}$$

Na verdade, podemos mostrar que, em geral, a largura de banda de um esquema M-ário ortogonal é M vezes a do esquema binário [Seção 10.7, Eq. (10.123)]. Portanto, em um esquema M-ário ortogonal, a taxa de comunicação aumenta por um fator $\log_2 M$, ao custo de um aumento na largura de banda de transmissão por um fator M. No esquema ortogonal, para uma comparável imunidade ao ruído, a potência de transmissão praticamente independe de M. Por conseguinte, em contraste com a modulação ASK M-ária, a FSK M-ária não requer maior potência de transmissão. Contudo, seu requisito de largura de banda aumenta quase linearmente com M (em comparação com FSK binária ou ASK M-ária).

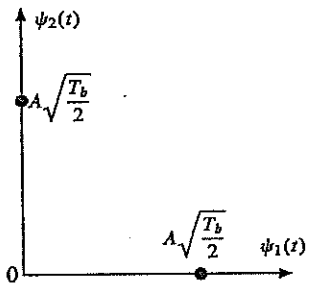


Figura 7.36 Símbolos FSK binários no espaço ortogonal bidimensional de sinais.

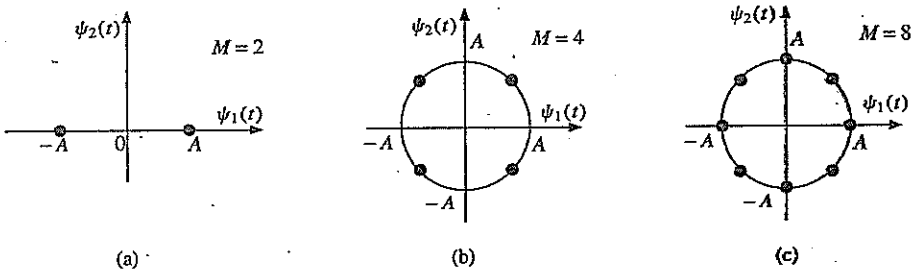


Figura 7.37
Símbolos PSK M-ário no espaço ortogonal de sinais: (a) $M = 2$; (b) $M = 4$; (c) $M = 8$.

PSK, PAM e QAM M-árias

Com uma pequena modificação na Eq. (7.61), sinais PSK genéricos podem ser escritos no formato

$$\varphi_{\text{PSK}}(t) = a_m\sqrt{\frac{2}{T_b}}\cos\omega_c t + b_m\sqrt{\frac{2}{T_b}}\,\text{sen}\,\omega_c t \qquad 0 \le t < T_b \tag{7.64a}$$

em que $a_m = A\cos\theta_m$ e $b_m = -A\,\text{sen}\,\theta_m$. Na verdade, com base na análise na Seção 2.6, $\sqrt{2/T_b}\cos\omega_c t$ e $\sqrt{2/T_b}\,\text{sen}\,\omega_c t$ são mutuamente ortogonais. Além disso, são normalizados no intervalo $[0, T_b]$. Como resultado, podemos representar todos os símbolos PSK em um sinal bidimensional de sinais com funções de base

$$\psi_1(t) = \sqrt{\frac{2}{T_b}}\cos\omega_c t \qquad \psi_2(t) = \sqrt{\frac{2}{T_b}}\,\text{sen}\,\omega_c t$$

tais que

$$\varphi_{\text{PSK}}(t) = a_m\psi_1(t) + b_m\psi_2(t) \tag{7.64b}$$

Podemos ilustrar geometricamente a relação entre símbolos PSK no espaço de sinais (Fig. 7.37). A Eq. (7.64) significa que modulações PSK podem ser representadas como uma modulação QAM. Na verdade, como o sinal é PSK, os pontos de sinal devem satisfazer a condição especial

$$\begin{aligned} a_m^2 + b_m^2 &= A^2\cos^2\theta_m + (-A)^2\text{sen}^2\,\theta_m \\ &= A^2 = \text{constante} \end{aligned} \tag{7.64c}$$

Em outras palavras, todos os pontos de sinal devem estar em uma circunferência de raio A. Na prática, todos os pontos de sinal são escolhidos para serem igualmente espaçados, com o propósito de obter melhor imunidade ao ruído. Portanto, para a sinalização PSK M-ária, os ângulos são escolhidos como uniformemente espaçados:

$$\theta_m = \theta_0 + \frac{2\pi}{M}(m-1) \qquad m = 1, 2, \ldots, M$$

A sinalização PSK especial com $M = 4$ é um formato de modulação digital extremamente popular e poderoso.* Esse sinal é a soma de dois sinais PSK binários de mesma frequência, um com portadora $\cos\omega_c t$ (em fase) e outro com portadora $\text{sen}\,\omega_c t$ (em quadratura). Por essa razão, este sinal também é conhecido como **PSK em quadratura (QPSK** – *quadrature PSK*). Podemos transmitir e receber os dois sinais no mesmo canal, dobrando a taxa de transmissão.

*QPSK tem diversas variações práticas, incluindo QPSK offset.

Para generalizar PSK e alcançar taxa de dados ainda mais alta, podemos ver que a representação PSK da Eq. (7.64) é um caso especial da modulação em amplitude em quadratura (QAM) discutida no Capítulo 4 (Fig. 4.19). A única diferença reside na exigência de PSK de que o sinal modulado tenha uma magnitude (módulo) constante A. Na verdade, o formato de modulação QAM, muito mais flexível e geral, também pode ser convenientemente usado para modulação digital. O sinal transmitido por um sistema QAM M-ário pode ser escrito como

$$\begin{aligned} p_i(t) &= a_i p(t) \cos \omega_c t + b_i p(t) \operatorname{sen} \omega_c t \\ &= r_i p(t) \cos(\omega_c t - \theta_i) \qquad i = 1, 2, \ldots, M \end{aligned}$$

em que

$$r_i = \sqrt{a_i^2 + b_i^2} \qquad \text{e} \qquad \theta_i = \tan^{-1} \frac{b_i}{a_i} \tag{7.65}$$

e $p(t)$ é um pulso em banda base adequadamente formatado. A escolha mais simples para $p(t)$ seria um pulso retangular

$$p(t) = \sqrt{\frac{2}{T_b}}\,[u(t) - u(t - T_b)]$$

Pulsos melhores podem, certamente, ser aplicados para conservar largura de banda.

A Fig. 7.38 mostra o modulador e o demodulador QAM. Cada um dos dois sinais $m_1(t)$ e $m_2(t)$ é uma sequência de pulsos $\sqrt{M}$-ários em banda base. Os dois sinais são modulados por duas portadoras de mesma frequência, mas em quadratura de fase. O sinal QAM digital $p_i(t)$ pode ser gerado por meio de QAM com $m_1(t) = a_i p(t)$ e $m_2(t) = b_i p(t)$. Os sinais $m_1(t)$ e $m_2(t)$ são sinais PAM em banda base. O diagrama de olho do sinal QAM consiste na componente em fase $m_1(t)$ e na componente em quadradura $m_2(t)$. Ambas exibem o diagrama de olho do sinal PAM M-ário em banda base, como discutido anteriormente na Seção 7.6.

A representação geométrica de sinais QAM M-ários pode ser obtida como extensão do espaço de sinais PSK com a simples remoção da condição de módulo constante na Eq. (7.64c). Uma escolha prática e muito popular de r_i e θ_i para $M = 16$ é ilustrada graficamente na Fig. 7.38b. O pulso transmitido $p_i(t)$ pode assumir 16 formas distintas e, portanto, é um pulso 16-ário. Como $M = 16$, cada pulso pode transmitir a informação de $\log_2 16 = 4$ dígitos binários. Isso pode ser feito da seguinte forma: há 16 possíveis sequências de quatro dígitos binários e há 16 combinações (a_i, b_i) na Fig. 7.38b. Assim, cada possível sequência de quatro bits é transmitida por um particular (a_i, b_i) ou (r_i, θ_i). Portanto, um pulso de sinal $r_i p(t) \cos(\omega_c t - \theta_i)$ transmite quatro bits. Em comparação com PSK binário (ou BPSK), a taxa de bits do sinal QAM 16-ário é quatro vezes maior, sem aumentar a largura de banda. A taxa de transmissão pode ser aumentada ainda mais com o aumento do valor de M.

Modulação e demodulação podem ser efetuadas com o sistema ilustrado na Fig. 7.38a. As entradas são $m_1(t) = a_i p(t)$ e $m_2(t) = b_i p(t)$. As duas saídas do demodulador são $a_i p(t)$ e $b_i p(t)$. O conhecimento de (a_i, b_i) permite a determinação de quatro bits transmitidos. Uma análise mais detalhada de sinais QAM 16-ários em canal ruidoso é feita na Seção 10.6 [Eq. (10.104)]. O valor prático da sinalização QAM-16 fica evidente quando consideramos a grande variedade de suas aplicações. Na verdade, a QAM-16 é usada em modems de dados telefônicos/fax V.32 (9600 bit/s), em modems de cabo de alta velocidade e na moderna difusão da televisão digital via satélite.

Reparemos que, se desabilitarmos a sequência de dados que modula sen $\omega_c t$ em QAM, todos os pontos de sinalização poderão ser reduzidos a uma única dimensão. Fazendo $m_2(t) = 0$, a modulação QAM passa a

$$p_i(t) = a_i p(t) \cos \omega_c t, \qquad t \in [0, T_b]$$

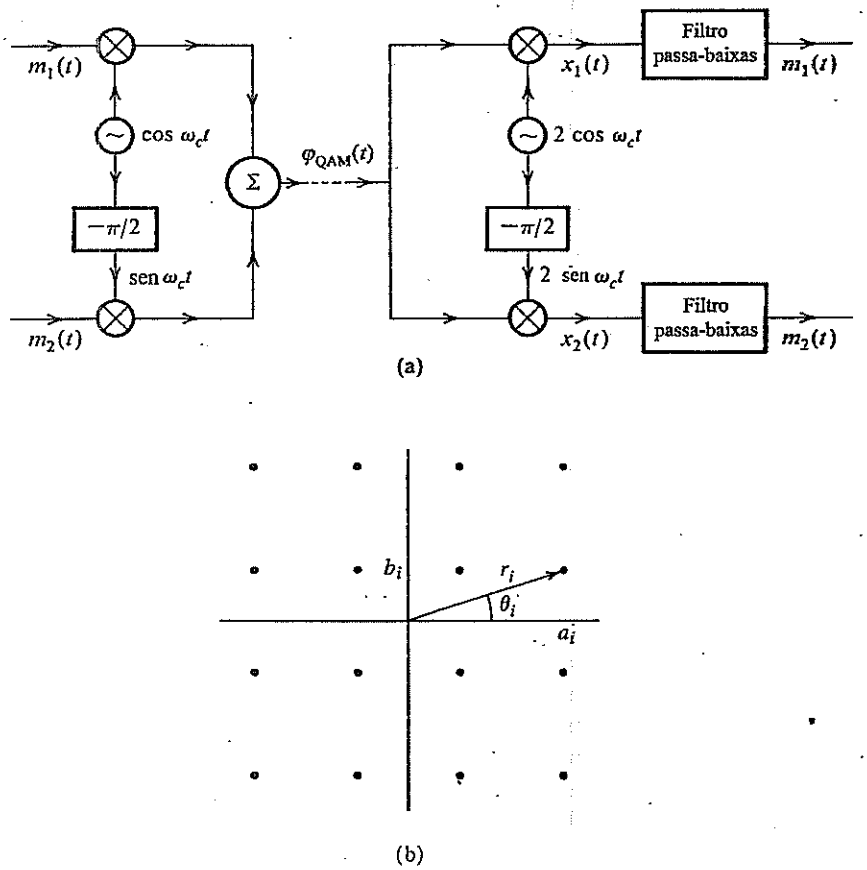


Figura 7.38 (a) QAM ou multiplexação em quadratura e (b) QAM de 16 pontos ($M = 16$).

Que se degenera em modulação por amplitude de pulso ou PAM. Uma comparação da expressão do sinal $p_i(t)$ com a do sinal DSB-SC analógico deixa claro que PAM é a versão digital de DSB-SC. Assim como um sinal QAM analógico é formado pela superposição de dois sinais com modulação em amplitude DSB-SC em quadratura de fase, o sinal QAM digital consiste em dois sinais PAM, cada um com $\sqrt{M}$ níveis de sinalização. Como na relação entre as modulações DSB-SC analógica e QAM, a PAM requer a mesma largura de banda que QAM. No entanto, a PAM é muito menos eficiente, pois necessita de M níveis de sinalização de modulação em uma dimensão, enquanto QAM requer apenas com $\sqrt{M}$ níveis de sinalização em cada uma das duas dimensões QAM ortogonais.

Negociação entre Potência e Largura de Banda

No Capítulo 10, discutiremos outros tipos de sinalização M-ária. A natureza da negociação entre largura de banda de transmissão e potência transmitida (ou SNR) depende da escolha do esquema M-ário. Por exemplo, na sinalização ortogonal, a potência transmitida praticamente independe de M, mas a largura de banda de transmissão aumenta com M. No caso PAM, em contraste, a potência transmitida aumenta aproximadamente com M^2, enquanto a largura de banda permanece constante. Assim, a sinalização M-ária nos permite grande flexibilidade na negociação entre potência de sinal (ou SNR) e largura de banda de transmissão. A escolha do sistema apropriado depende de circunstâncias particulares. Por exemplo, o uso de sinalização QAM é apropriado se a largura de banda for limitada (como em linha telefônicas), e o uso de sinalização ortogonal é adequado quando há restrição de potência (como em comunicações espaciais).

7.10 EXERCÍCIOS COM O MATLAB

Nesta seção, apresentamos programas MATLAB para gerar diagramas de olho. O primeiro passo consiste na especificação das formas básicas de pulso em PAM. Os quatro programas curtos dados a seguir são usados para gerar pulsos NRZ, RZ, meia senoide e cosseno levantado.

```
% (pnrz.m)
% gera um pulso retangular de largura T
% Uso function pout=pnrz(T);
function pout=prect(T);
pout=ones(1,T);
end
```

```
% (prz.m)
% gera um pulso retangular de largura T/2
% Uso function pout=prz(T);
function pout=prz(T);
pout=[zeros(1,T/4) ones(1,T/2) zeros(1,T/4)];
end
```

```
% (psine.m)
% gera pulso senoidal de largura T
%
function pout=psine(T);
pout=sin(pi*[0:T-1]/T);
end
```

```
% (prcos.m)
% Uso y=prcos(rollfac,length, T)
function y=prcos(rollfac,length, T)
% rollfac = 0 a 1 é o fator de decaimento
% length é o comprimento unilateral do pulso em termos de T
% length = 2T+1;
% T é a taxa de superamostragem
y=rcosfir(rollfac, length, T,1, 'normal');
end
```

O primeiro programa (binary_eye.m) usa quatro diferentes formas de pulso para gerar diagramas de olho de sinalização polar binária.

```
% (binary_eye.m)
% gera e traça diagrama de olho
%
clear;clf;
```

```
data = sign(randn(1,400));    % Gera 400 bits aleatórios
Tau=64;                       % Define o período de símbolo
dataup=upsample(data, Tau);   % Gera trem de impulsos
yrz=conv(dataup,prz(Tau));    % Sinal polar com retorno ao zero
yrz=yrz(1:end-Tau+1);
ynrz=conv(dataup,pnrz(Tau)); % Sinal polar sem retorno ao zero
ynrz=ynrz(1:end-Tau+1);
ysine=conv(dataup,psine(Tau)); % Sinal polar de meia senoide
ysine=ysine(1:end-Tau+1);
Td=4;                 % trunca cosseno levantada em 4 períodos
yrcos=conv(dataup,prcos(0.5,Td,Tau)); % fator de decaimento = 0,5
yrcos=yrcos(2*Td*Tau:end-2*Td*Tau+1); % gera trem de pulsos RC
eye1=eyediagram(yrz,2*Tau,Tau,Tau/2);title('Diagrama de olho RZ');
eye2=eyediagram(ynrz,2*Tau,Tau,Tau/2);title('Diagrama de olho NRZ');
eye3=eyediagram(ysine,2*Tau,Tau,Tau/2);title('Diagrama de olho de
meia senoide');
eye4=eyediagram(yrcos,2*Tau,Tau); title('Diagrama de olho de cosseno
levantado);
```

O segundo programa (Mary-eye.m) usa quatro diferentes formas de pulso para gerar diagramas de sinalização PAM de quatro níveis.

```
% (Mary_eye.m)
% gera e traça diagrama de olho
%
%
clear;clf;
data = sign(randn(1,400))+2* sign(randn(1,400)); % Gera 400 símbo-
los PAM
Tau=64;                       % Define o período de símbolo
dataup=upsample(data, Tau);   % Gera trem de impulsos
yrz=conv(dataup,prz(Tau));    % Sinal polar com retorno ao zero
yrz=yrz(1:end-Tau+1);
ynrz=conv(dataup,pnrz(Tau)); % Sinal polar sem retorno ao zero
ynrz=ynrz(1:end-Tau+1);
ysine=conv(dataup,psine(Tau)); % Sinal polar de meia senoide
ysine=ysine(1:end-Tau+1);
Td=4;                 % trunca cosseno levantada em 4 períodos
yrcos=conv(dataup,prcos(0.5,Td,Tau)); % fator de decaimento = 0,5
yrcos=yrcos(2*Td*Tau:end-2*Td*Tau+1); % gera trem de pulsos RC
eye1=eyediagram(yrz,2*Tau,Tau,Tau/2);title('Diagrama de olho RZ');
eye2=eyediagram(ynrz,2*Tau,Tau,Tau/2);title('Diagrama de olho NRZ');
eye3=eyediagram(ysine,2*Tau,Tau,Tau/2);title('Diagrama de olho de meia
senoide');
eye4=eyediagram(yrcos,2*Tau,Tau); title('Diagrama de olho de cosseno
levantado');
```

REFERÊNCIAS

1. A. Lender, "Duobinary Technique for High Speed Data Transmission," *IEEE Trans. Commun. Electron.*, vol. CE-82, pp. 214–218, May 1963.

2. A. Lender, "Correlative Level Coding for Binary-Data Transmission," *IEEE Spectrum*, vol. 3, no. 2, pp. 104–115, Feb. 1966.
3. P. Bylanski and D. G. W. Ingram, *Digital Transmission Systems*, Peter Peregrinus Ltd., Hertshire, England, 1976.
4. H. Nyquist, "Certain Topics in Telegraph Transmission Theory," *AIEE Trans.*, vol. 47, p. 817, April 1928.
5. E. D. Sunde, *Communication Systems Engineering Technology*, Wiley, New York, 1969.
6. R.W. Lucky and H. R. Rudin, "Generalized Automatic Equalization for Communication Channels," *IEEE Int. Commun.* Conf., vol. 22, 1966.
7. W. F. Trench, "An Algorithm for the Inversion of Finite Toeplitz Matrices," *J. SIAM*, vol. 12, pp. 515–522, Sept. 1964.
8. A. Lender, Chapter 7, in *Digital Communications: Microwave Applications*, K. Feher, Ed., Prentice-Hall, Englewood Cliffs, NJ, 1981.

EXERCÍCIOS

7.2-1 Considere um pulso retangular de largura completa

$$p(t) = \Pi(t/T_b)$$

(a) Determine a PSD para as sinalizações polar, on-off e bipolar.

(b) Esboce o gráfico de cada PSD e determine sua largura de banda. Para cada sinalização, compare a largura de banda com a do caso em que $p(t)$ é um pulso retangular de meia largura.

7.2-2 **(a)** Uma sequência aleatória de dados binários **110100101**... é transmitida com o uso do código de linha Manchester (bifásico) com o pulso $p(t)$ mostrado na Fig. 7.7a. Esboce o gráfico da forma de onda $y(t)$.

(b) Calcule $S_y(f)$, a PSD de um sinal Manchester (bifásico) na parte **(a)**, supondo que **1** e **0** têm igual probabilidade de ocorrência. Esboce o gráfico desta PSD e determine sua largura de banda.

7.2-3 Para a forma de pulso

$$p(t) = \Pi\left(\frac{t}{0{,}5T_b}\right)$$

use o código diferencial (Fig. 7.18) para determinar a PSD para um sinal binário. Determine a PSD $S_y(f)$.

7.2-4 O código de linha **duobinário** proposto por Lender também é terciário, como o bipolar, mas requer somente metade da largura de banda de bipolar. Na prática, o código duobinário é realizado indiretamente com o uso de uma forma especial de pulso, como discutido na Seção 7.3 (Fig. 7.18). Nesse código, um **0** é transmitido por ausência de pulso e um **1**, por um pulso $p(t)$ ou $-p(t)$ segundo a regra: um **1** é codificado pelo mesmo pulso usado para o **1** anterior, se houver um número par de **0**s entre eles; um **1** é codificado por um pulso com polaridade oposta à do pulso usado para o **1** anterior, se houver um número ímpar de **0**s entre eles. Um número 0 é considerado um número par. Como o bipolar, esse código também tem uma capacidade de detecção de erro isolado, pois a recepção correta implica que, entre sucessivos pulsos de mesma polaridade, deve ocorrer um número par de **0**s, e entre sucessivos pulsos de polaridades opostas, um número ímpar de **0**s.

(a) Admitindo um pulso retangular de meia largura, esboce o gráfico do sinal duobinário $y(t)$ para a sequência binária aleatória

1110001101001010...

(b) Determine R_0, R_1 e R_2 para este código. Admita (se preferir, prove) que $R_n = 0$ para todo $n > 2$. Determine a PSD para esse código de linha (considerando um pulso de meia largura) e esboce seu gráfico. Mostre que sua largura de banda é $R_b/2$, metade da largura de banda do código bipolar.

7.3-1 Dados, a uma taxa de 6 kbit/s, devem ser transmitidos em uma linha privada com largura de banda de 4 kHz, usando pulsos que satisfaçam o critério de Nyquist. Determine o máximo valor do fator de decaimento r que pode ser usado.

7.3-2 Em um certo sistema de telemetria, há oito medidas analógicas, cada uma com largura de banda de 2 kHz. Amostras desses sinais são multiplexadas por divisão no tempo, quantizadas e codificadas em binário. O erro nas amplitudes das amostras não pode ultrapassar 1% da amplitude de pico.

(a) Determine L, o número de níveis de quantização.

(b) Determine a largura de banda de transmissão B_T se forem usados pulsos que satisfaçam o critério de Nyquist, com fator de decaimento $r = 0{,}2$. A taxa de amostragem deve ser pelo menos 25% acima da taxa de Nyquist.

7.3-3 Uma linha privada de telefonia, com largura de banda de 3 kHz, é usada para transmitir dados binários. Calcule da taxa de dados (em bits por segundo) que pode ser transmitida se for usado:

(a) Sinal polar com pulsos retangulares de meia largura.

(b) Sinal polar com pulsos retangulares de largura completa.

(c) Sinal polar com pulsos que satisfaçam o critério de Nyquist, com fator de decaimento $r = 0{,}25$.

(d) Sinal bipolar com pulsos retangulares de meia largura.

(e) Sinal bipolar com pulsos retangulares de largura completa.

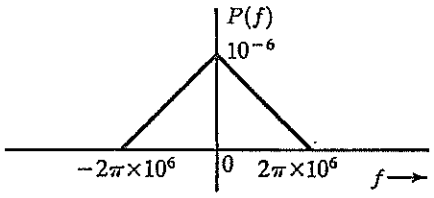
7.3-4 A transformada de Fourier $P(f)$ do pulso básico $p(t)$ usado em certos sistemas de comunicação binária é mostrado na Fig. E.7.3-4.

(a) A partir da forma de $P(f)$, determine a que taxa de pulsos esse pulso satisfaria o critério de Nyquist.

(b) Determine $p(t)$ e comprove que esse pulso satisfaz (ou não) o critério de Nyquist.

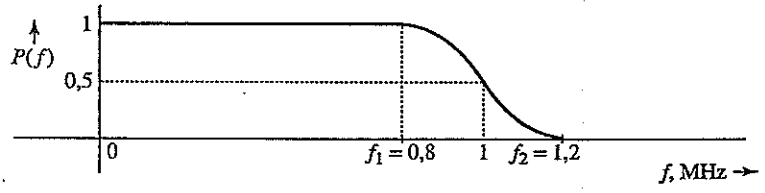
(c) Se o pulso não satisfizer o critério de Nyquist, qual é a taxa de transmissão (em bits por segundo) e qual é o valor do fator de decaimento?

Figura E.7.3-4



7.3-5 Um pulso $p(t)$ cujo espectro $P(f)$ é mostrado na Fig. E.7.3-5 satisfaz o critério de Nyquist. Se $f_1 = 0{,}8$ MHz e $f_2 = 1{,}2$ MHz, determine a máxima taxa em que dados binários podem ser transmitidos por esse pulso usando o critério de Nyquist. Qual é o valor do fator de decaimento?

Figura E.7.3-5



7.3-6 Dados binários, à taxa de 1 Mbit/s, devem ser transmitidos com uso de pulsos que satisfaçam o critério de Nyquist, sendo $P(f)$ mostrada na Fig. E.7.3-5. As frequências f_1 e f_2 do espectro são ajustáveis. O canal disponível para transmissão desses dados tem largura de banda de 700 kHz. Determine f_1 e f_2 e o fator de decaimento.

7.3-7 Mostre que a transformada de Fourier inversa de $P(f)$ na Eq. (7.39) é, de fato, o pulso $p(t)$ dado na Eq. (7.38), que satisfaz o segundo critério de Nyquist.

Sugestão: Use a Eq. (3.32) para determinar a transformada inversa de Fourier de $P(f)$ na Eq. (7.39) e expresse sinc (x) na forma sen x/x.

7.3-8 Mostre que a transformada de Fourier inversa de $P(f)$ (o espectro do pulso cosseno levantado) na Eq. (7.35) é o pulso $p(t)$ dado na Eq. (7.36).

Sugestão: Use a Eq. (3.32) para determinar a transformada inversa de Fourier de $P(f)$ na Eq. (7.35) e expresse sinc (x) na forma sen x/x.

7.3-9 Mostre que existe um (e somente um) pulso $p(t)$ de largura de banda $R_b/2$ Hz que satisfaz o segundo critério de Nyquist [Eq. (7.37)]. Mostre que esse pulso é dado por

$$p(t) = [\text{sinc}\,(\pi R_b t) + \text{sinc}\,[\pi R_b(t - T_b)]] = \frac{\text{sen}\,(\pi R_b t)}{\pi R_b t(1 - R_b t)}$$

e que sua transformada de Fourier é $P(f)$ dada na Eq. (7.39).

Sugestão: Para um pulso de largura de banda $R_b/2$ Hz, o intervalo de Nyquist é $1/R_b = T_b$, e a condição (7.37) dá os valores das amostras de Nyquist em $t = \pm nT_b$. Para construir $p(t)$, use a fórmula de interpolação [Eq. (6.10)] com $B = R_b/2$, $T_s = T_b$. Na determinação de $P(f)$, use $\left(1 + e^{-j2\pi fT_b}\right) = e^{-j\pi fT_b}\left(e^{j\pi fT_b} + e^{-j\pi fT_b}\right)$.

7.3-10 Em uma transmissão de dados binários usando pulsos duobinários, valores de amostras foram lidos como:

120 –2 –200 –202002000 –2

(a) Determine se há qualquer erro na detecção.

(b) Se não houver erro de detecção, determine a sequência de bits recebida.

7.3-11 Em uma transmissão de dados binários usando pulsos duobinários, valores de amostras foram lidos como:

12000 –200 –20200 –20220 –2

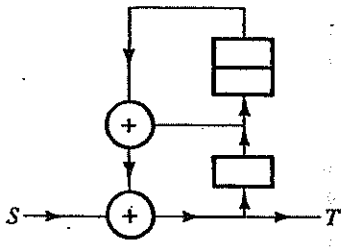
(a) Determine se há qualquer erro na detecção.

(b) Você é capaz de deduzir a correta sequência de dígitos transmitida? Há mais de uma sequência correta possível. Forneça tantas sequências corretas quanto possível, admitindo que a ocorrência de mais de um erro de detecção é extremamente provável.

7.4-1 No Exemplo 7.2, quando a sequência S = **101010100000111** foi aplicada à entrada do desembaralhador na Fig. 7.20a, a saída T foi determinada como **101110001101001**. Comprove que, quando essa sequência T é aplicada à entrada do desembaralhador na Fig. 7.20b, a saída é a sequência original de entrada, S = **101010100000111**.

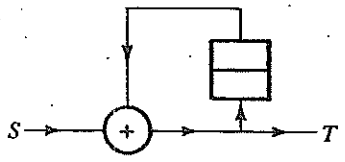
7.4-2 Projete um desembaralhador para o embaralhador da Fig. E.7.4-2. Se uma sequência S = **101010100000111** for aplicada à entrada do embaralhador, determine a sequência de saída T. Comprove que, se essa sequência T for aplicada à entrada do desembaralhador, a saída será a sequência S.

Figura E.7.4-2



7.4-3 Repita o Exercício 7.4-2 para o caso em que o desembaralhador mostrado na Fig. E.7.4-3 é concatenado com o embaralhador na Fig. E.7.4-2 para formar um embaralhador composto.

Figura E.7.4-3



7.5-1 Em certo sistema de comunicação binária que usa pulsos que satisfazem o critério de Nyquist, um pulso recebido $p_r(t)$ (Fig. 7.22a) tem os seguintes valores de amostras não nulas

$$p_r(0) = 1$$
$$p_r(T_b) = 0{,}1 \qquad p_r(-T_b) = 0{,}3$$
$$p_r(2T_b) = -0{,}02 \qquad p_r(-2T_b) = -0{,}07$$

(a) Determine a configuração das derivações de um equalizador de três derivações e forçamento a zero.

(b) Usando o equalizador na parte (a), determine a ISI residual não nula.

7.7-1 Em um esquema PAM com $M = 16$,

(a) Determine a mínima largura de banda necessária para transmitir dados a uma taxa de 12.000 bits/s com ISI zero.

(b) Determine a largura de banda de transmissão se pulsos que satisfazem o critério de Nyquist com fator de decaimento $r = 0{,}2$ forem usados para transmitir os dados.

7.7-2 Um sinal de áudio de largura de banda de 4 kHz é amostrado a uma taxa 25% acima da taxa de Nyquist, e quantizado. O erro de quantização não deve exceder 0,1% da amplitude de pico do sinal. As resultantes amostras quantizadas são, então, codificadas e transmitidas por pulsos 4-ários.

(a) Determine o número mínimo de pulsos 4-ários para codificar cada amostra.

(b) Determine a largura mínima de banda necessária à transmissão destes dados com ISI zero.

(c) Se pulsos 4-ários que satisfaçam o critério de Nyquist com 25% de decaimento forem usados para transmitir estes dados, determine a largura de banda de transmissão.

7.7-3 Dados binários são transmitidos em um certo canal a uma taxa de R_b bit/s. Para reduzir a largura da banda, foi decidido o uso de sinalização PAM 16-ária para a transmissão desses dados.

(a) Qual é o fator de redução de largura de banda?

(b) Qual é o fator de aumento da potência de transmissão, admitindo que a mínima separação entre amplitudes de pulso seja a mesma nos dois casos.

Sugestão: Tome amplitudes de pulso como $\pm A/2, \pm 3A/2, \pm 5A/2, \pm 7A/2, \ldots, \pm 15A/2$, de modo que a mínima separação entre os vários níveis de amplitude seja A (como no caso binário, com pulsos $\pm A/2$). Suponha que todos os 16 níveis tenham igual probabilidade de ocorrência. Recorde, ainda, que a multiplicação de um pulso por uma constante k aumenta a energia do mesmo k^2 vezes.

7.7-4 Um sinal de áudio de largura de banda de 10 kHz é amostrado à taxa de 24 kHz, quantizado em 256 níveis e codificado por pulsos PAM M-ários que satisfazem o critério de Nyquist com fator de decaimento $r = 0{,}2$. Uma largura de banda de 30 kHz é disponível para a transmissão dos dados. Determine o melhor valor de M.

7.7-5 Considere o caso de transmissão binária via sinalização polar que usa pulsos retangulares de meia largura, de amplitudes $A/2$ e $-A/2$. A taxa de dados é R_b bit/s.

(a) Qual é a mínima largura de banda e a potência transmitida.

(b) Esses dados devem ser transmitidos por pulsos retangulares M-ários de meia largura e amplitudes

$$\pm A/2,\ \pm 3A/2,\ \pm 5A/2, \ldots, \pm[(M-1)/2]A$$

Repare que, para manter a mesma imunidade ao ruído, a mínima separação entre amplitudes de pulso é A. Se todos os pulsos M-ários tiverem igual probabilidade de ocorrência, mostre que a potência transmitida será

$$P = \frac{(M^2 - 1)A^2}{24 \log_2 M}$$

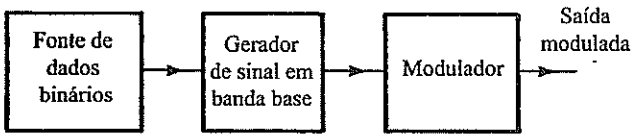
Determine, ainda, a largura de banda.

7.8-1 A Fig. E7.8-1 mostra um esquema de transmissão de dados binários. O gerador de sinal em banda base usa pulsos de largura completa e sinalização polar. A taxa de dados é 1 Mbit/s.

(a) Se o modulador gerar um sinal PSK, qual é a largura de banda da saída modulada?

(b) Se o modulador gerar FSK com diferença $f_{c1} - f_{c0} = 100$ kHz (Fig. 7.32c), determine a largura de banda do sinal modulado.

Figura E.7.8-1



7.8-2 Refaça o Exercício 7.8-1 com pulsos de largura completa que satisfazem o critério de Nyquist com $r = 0{,}2$.

7.8-3 Refaça o Exercício 7.8-1 para um esquema multiamplitude com $M = 4$ (sinalização PAM com pulsos de largura completa). Em FSK [Exercício 7.8-1, parte **(b)**], admita que níveis sucessivos de amplitude sejam transmitidos por frequências separadas por 100 kHz.

APÊNDICE A

ORTOGONALIDADE DE ALGUNS CONJUNTOS DE SINAIS

A.1 Ortogonalidade do Conjunto de Sinais Trigonométricos

Consideremos a integral I definida por

$$I = \int_{T_0} \cos n\omega_0 t \cos m\omega_0 t \, dt \tag{A.1a}$$

em que $\int_{T_0}$ significa integração em qualquer intervalo contíguo de $T_0 = 2\pi/\omega_0$ segundos. Usando uma identidade trigonométrica (Apêndice E), a Eq. (A.1a) pode ser expressa como

$$I = \frac{1}{2}\left[\int_{T_0} \cos (n+m)\omega_0 t \, dt + \int_{T_0} \cos (n-m)\omega_0 t \, dt\right] \tag{A.1b}$$

Como $\cos \omega_0 t$ executa um ciclo completo em qualquer intervalo de T_0 segundos, $\cos (n+m)\omega_0 t$ executa $(n+m)$ ciclos completos em qualquer intervalo com duração de T_0 segundos. Portanto, a primeira integral na Eq. (A.1b), que representa a área sob $(n+m)$ ciclos completos de uma senoide, é igual a zero. O mesmo argumento mostra que a segunda integral na Eq. (A.1b) também é zero, exceto quando $n = m$. Portanto, I na Eq. (A.1b) é zero para todo $n \neq m$. Quando $n = m$, a primeira integral na Eq. (A.1b) continua igual a zero, mas a segunda integral fornece

$$I = \frac{1}{2}\int_{T_0} dt = \frac{T_0}{2}$$

Assim,

$$\int_{T_0} \cos n\omega_0 t \cos m\omega_0 t \, dt = \begin{cases} 0 & n \neq m \\ \frac{T_0}{2} & m = n \neq 0 \end{cases} \tag{A.2a}$$

Podemos usar argumento similar para mostrar que

$$\int_{T_0} \operatorname{sen} n\omega_0 t \operatorname{sen} m\omega_0 t \, dt = \begin{cases} 0, & n \neq m \\ \frac{T_0}{2} & n = m \neq 0 \end{cases} \tag{A.2b}$$

e

$$\int_{T_0} \operatorname{sen} n\omega_0 t \cos m\omega_0 t \, dt = 0 \quad \text{para todo } n \text{ e } m \tag{A.2c}$$

A.2 Ortogonalidade do Conjunto de Sinais Exponenciais

O conjunto de exponenciais $e^{jn\omega_0 t}$ ($n = 0, \pm 1, \pm 2, ...$) é ortogonal em qualquer intervalo de duração T_0, ou seja

$$\int_{T_0} e^{jm\omega_0 t}(e^{jn\omega_0 t})^* \, dt = \int_{T_0} e^{j(m-n)\omega_0 t} \, dt = \begin{cases} 0 & m \neq n \\ T_0 & m = n \end{cases} \tag{A.3}$$

Seja I a integral no lado esquerdo da Eq. (A.3):

$$\begin{aligned} I &= \int_{T_0} e^{jm\omega_0 t}(e^{jn\omega_0 t})^* \, dt \\ &= \int_{T_0} e^{j(m-n)\omega_0 t} \, dt \end{aligned} \tag{A.4}$$

O caso $m = n$ é trivial: o integrando é a unidade, de modo que $I = T_0$. Todavia, quando $m \times n$,

$$\begin{aligned} I &= \frac{1}{j(m-n)\omega_0} \, e^{j(m-n)\omega_0 t}\Big|_{t_1}^{t_1+T_0} \\ &= \frac{1}{j(m-n)\omega_0} e^{j(m-n)\omega_0 t_1}[e^{j(m-n)\omega_0 T_0} - 1] = 0 \end{aligned}$$

O último resultado advém de $\omega_0 T_0 = 2\pi$ e $e^{j2\pi k} = 1$ para todos os valores inteiros de k.

APÊNDICE B

DESIGUALDADE DE CAUCHY-SCHWARZ

Provemos a seguinte desigualdade de Cauchy-Schwarz para um par de sinais de energia finita $f(t)$ e $g(t)$:

$$\left[\int_a^b f(t)g(t)\,dt\right]^2 \leq \left[\int_a^b f^2(t)\,dt\right]\left[\int_a^b g^2(t)\,dt\right] \tag{B.1}$$

em que a igualdade ocorre somente se $g(t) = cf(t)$, sendo c uma constante arbitrária.

A desigualdade de Cauchy-Schwarz para funções de valores complexos e energia finita $X(\omega)$ e $Y(\omega)$ é dada por

$$\left|\int_{-\infty}^{\infty} X(\omega)Y(\omega)\,d\omega\right|^2 \leq \int_{-\infty}^{\infty} |X(\omega)|^2 d\omega \int_{-\infty}^{\infty} |Y(\omega)|^2\,d\omega \tag{B.2}$$

em que a igualdade ocorre somente se $Y(\omega) = cX^*(\omega)$, sendo c uma constante arbitrária.

Podemos provar a Eq. (B.1) da seguinte maneira: para qualquer valor real de λ, sabemos que

$$\int_a^b [\lambda f(t) - g(t)]^2\,dt \geq 0 \tag{B.3}$$

ou

$$\lambda^2 \int_a^b f^2(t)\,dt - 2\lambda \int_a^b f(t)g(t)\,dt + \int_a^b g^2(t)\,dt \geq 0 \tag{B.4}$$

Como essa equação quadrática em λ é não negativa para qualquer valor de λ, seu discriminante deve ser não positivo, e a Eq. (B.1) segue. Se o discriminante for zero, então, para algum valor de $\lambda = c$, a forma quadrática se torna igual a zero. Isso é possível somente se $cf(t) - g(t) = 0$, e o resultado segue.

Para provar a Eq. (B.2), observemos que $|X(\omega)|$ e $|Y(\omega)|$ são funções reais, de modo que a desigualdade na Eq. (B.1) se aplica. Logo,

$$\left[\int_a^b |X(\omega)Y(\omega)|\,d\omega\right]^2 \leq \int_a^b |X(\omega)|^2 d\omega \int_a^b |Y(\omega)|^2\,d\omega \tag{B.5}$$

em que a igualdade ocorre somente se $|Y(\omega)| = c|X(\omega)|$, sendo c uma constante arbitrária. Agora, recordemos que

$$\left|\int_a^b X(\omega)Y(\omega)\,d\omega\right| \leq \int_a^b |X(\omega)||Y(\omega)|\,d\omega = \int_a^b |X(\omega)Y(\omega)|\,d\omega \tag{B.6}$$

em que a igualdade ocorre somente se $Y(\omega) = cX^*(\omega)$, sendo c uma constante arbitrária. A Eq. (B.2) segue imediatamente das Eq. (B.5) e (B.6).

APÊNDICE C

ORTOGONALIZAÇÃO DE GRAM-SCHMIDT DE UM CONJUNTO DE VETORES

Definimos a dimensionalidade de um espaço vetorial como o máximo número de vetores independentes no espaço. Assim, em um espaço de N dimensões, não pode haver mais que N vetores independentes. Alternativamente, sempre é possível encontrar um conjunto de N vetores independentes. Uma vez que um conjunto tenha sido escolhido, qualquer vetor nesse espaço pode ser expresso em termos dos vetores nesse conjunto (como uma combinação linear dos mesmos). Esse conjunto forma o que comumente chamamos de conjunto de base, que define o sistema de coordenadas. Tal conjunto de N vetores independentes não é, de modo algum, único. O leitor deve ter familiaridade com essa propriedade no espaço físico de três dimensões, no qual é possível encontrar um número infinito de conjuntos independentes de três vetores. Isso fica claro pelo fato de ser possível definir um número infinito de sistemas de coordenadas. Um conjunto ortogonal, no entanto, é de interesse especial, por ser de tratamento mais fácil que um conjunto não ortogonal. Se tivermos um conjunto de N vetores independentes, a partir deles, é possível obter outro conjunto de N vetores independentes que é ortogonal. Isso é feito com o procedimento de ortogonalização de Gram-Schmidt.

Na dedução a seguir, usamos o resultado [obtido na Eq. (2.27)] que diz que a projeção (ou componente) de um vetor $\mathbf{x}_2$ na direção de outro vetor $\mathbf{x}_1$ (Fig. C.1) é $c_{12}\mathbf{x}_1$, em que

$$c_{12} = \frac{< \mathbf{x}_1, \mathbf{x}_2 >}{\|\mathbf{x}_1\|^2}\mathbf{y}_1 \tag{C.1}$$

O erro nessa aproximação é o vetor $\mathbf{x}_2 - c_{12}\mathbf{x}_1$, ou seja,

$$\text{vetor de erro} = \mathbf{x}_2 - \frac{< \mathbf{x}_1, \mathbf{x}_2 >}{\|\mathbf{x}_1\|^2}\mathbf{x}_1 \tag{C.2}$$

O vetor de erro, mostrado tracejado na Fig. C.1, é ortogonal ao vetor $\mathbf{x}_1$.

Para um entendimento físico desse procedimento, consideremos um caso simples de espaço bidimensional. Sejam $\mathbf{x}_1$ e $\mathbf{x}_2$ dois vetores independentes em um espaço bidimensional (Fig. C.1). Desejamos gerar, a partir de $\mathbf{x}_1$ e $\mathbf{x}_2$, um novo conjunto de dois vetores ortogonais $\mathbf{y}_1$ e $\mathbf{y}_2$. Por conveniência, escolhamos

$$\mathbf{y}_1 = \mathbf{x}_1 \tag{C.3}$$

Agora, busquemos um vetor $\mathbf{y}_2$ que seja ortogonal a $\mathbf{y}_1$ (e a $\mathbf{x}_1$). A Fig. C.1 mostra que o vetor de erro na aproximação de $\mathbf{x}_2$ por $\mathbf{y}_1$ (linhas tracejadas) é ortogonal a $\mathbf{y}_1$ e pode ser tomado como $\mathbf{y}_2$.

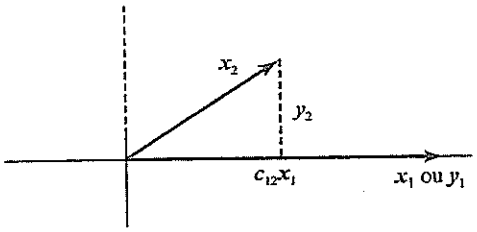


Figura C.1 Procedimento de Gram-Schmidt para um caso bidimensional.

Logo,

$$\begin{aligned} \mathbf{y}_2 &= \mathbf{x}_2 - \frac{<\mathbf{x}_1, \mathbf{x}_2>}{\|\mathbf{x}_1\|^2}\mathbf{x}_1 \\ &= \mathbf{x}_2 - \frac{<\mathbf{y}_1, \mathbf{x}_2>}{\|\mathbf{y}_1\|^2}\mathbf{y}_1 \end{aligned} \tag{C.4}$$

As Eqs. (C.3) e (C.4) fornecem o desejado conjunto ortogonal. Reparemos que esse conjunto não é único. Há um número infinito de conjuntos de vetores ortogonais $(\mathbf{y}_1, \mathbf{y}_2)$ que podem ser gerados de $(\mathbf{x}_1, \mathbf{x}_2)$. Em nossa dedução, poderíamos ter começado com $\mathbf{y} = \mathbf{x}_2$, em vez de $\mathbf{y}_1 = \mathbf{x}_1$. Esse ponto de partida produziria um conjunto totalmente diferente.

O leitor pode estender esses resultados ao caso tridimensional. Se vetores $\mathbf{x}_1$, $\mathbf{x}_2$ e $\mathbf{x}_3$ formarem um conjunto independente nesse espaço, podemos formar vetores $\mathbf{y}_1$ e $\mathbf{y}_2$ como nas Eqs. (C.3) e (C.4). Para determinar $\mathbf{y}_3$, aproximamos $\mathbf{x}_3$ em termos dos vetores $\mathbf{y}_1$ e $\mathbf{y}_2$. O erro nessa aproximação deve ser ortogonal a $\mathbf{y}_1$ e a $\mathbf{y}_2$ e, portanto, pode ser tomado como o vetor ortogonal $\mathbf{y}_3$. Assim,

$$\begin{aligned} \mathbf{y}_3 &= \mathbf{x}_3 - \text{soma das projeções de } \mathbf{x}_3 \text{ em } \mathbf{y}_1 \text{ e } \mathbf{y}_2 \\ &= \mathbf{x}_3 - \frac{<\mathbf{y}_1, \mathbf{x}_3>}{\|\mathbf{y}_1\|^2}\mathbf{y}_1 - \frac{<\mathbf{y}_2, \mathbf{x}_3>}{\|\mathbf{y}_2\|^2}\mathbf{y}_2 \end{aligned} \tag{C.5}$$

Esses resultados podem ser estendidos a um espaço de N dimensões. Em geral, dados N vetores independentes $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, se procedermos por linhas semelhantes às anteriores, podemos obter um conjunto ortogonal $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$, em que

$$\mathbf{y}_1 = \mathbf{x}_1 \tag{C.6}$$

e

$$\mathbf{y}_j = \mathbf{x}_j - \sum_{k=1}^{j-1} \frac{<\mathbf{y}_k, \mathbf{x}_j>}{\|\mathbf{y}_k\|^2}\mathbf{y}_k \qquad j = 2, 3, \ldots, N \tag{C.7}$$

Reparemos que esse é um dos infinitos conjuntos ortogonais que podem ser obtidos a partir de $\mathbf{x}_1$, $\mathbf{x}_2, \ldots, \mathbf{x}_N$. Além disso, esse não é um conjunto ortonormal. O conjunto ortonormal $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_N$ pode ser obtido por normalização dos comprimentos dos respectivos vetores:

$$\hat{\mathbf{y}}_k = \frac{\mathbf{y}_k}{\|\mathbf{y}_k\|}$$

Podemos aplicar esses conceitos ao espaço de sinais, pois existe uma correspondência biunívoca entre sinais e vetores. Se tivermos N sinais independentes $x_1(t), x_2(t), \ldots, x_N(t)$, podemos formar um conjunto de N sinais ortogonais $y_1(t), y_2(t), \ldots, y_N(t)$,

$$\begin{aligned} y_1(t) &= x(t) \\ y_j(t) &= x_j(t) - \sum_{k=1}^{j-1} c_{kj} y_k(t) \qquad j = 2, 3, \ldots, N \end{aligned} \tag{C.8}$$

em que

$$c_{kj} = \frac{\int y_k(t)x_j(t)\,dt}{\int y_k^2(t)\,dt} \tag{C.9}$$

Reparemos que esse é um dos infinitos conjuntos ortogonais que podem ser formados a partir do conjunto $x_1(t), x_2(t),..., x_N(t)$. O conjunto pode ser normalizado dividindo cada sinal $y_j(t)$ por sua energia.

Exemplo C.1 Os sinais exponenciais

$$g_1(t) = e^{-pt}u(t)$$
$$g_2(t) = e^{-2pt}u(t)$$
$$\vdots$$
$$g_N(t) = e^{-Npt}u(t)$$

formam um conjunto de sinais independentes no espaço de N dimensões, em que N pode ser um inteiro qualquer. Todavia, esse conjunto não é ortogonal. Podemos usar o procedimento de Gram-Schmidt para obter um conjunto ortogonal para esse espaço. Seja $y_1(t), y_2(t),..., y_N(t)$ o desejado conjunto de base ortogonal; escolhamos

$$y_1(t) = g_1(t) = e^{-pt}u(t)$$

Das Eqs. (C.8) e (C.9), temos

$$y_2(t) = x_2(t) - c_{12}y_1(t)$$

em que

$$\begin{aligned} c_{12} &= \frac{\int_{-\infty}^{\infty} y_1(t)x_2(t)\,dt}{\int_{-\infty}^{\infty} y_1^2(t)\,dt} \\ &= \frac{\int_0^{\infty} e^{-pt}e^{-2pt}\,dt}{\int_0^{\infty} e^{-2pt}\,dt} \\ &= \frac{2}{3} \end{aligned}$$

Logo,

$$y_2(t) = (e^{-2pt} - \tfrac{2}{3}e^{-pt})u(t) \tag{C.10}$$

Do mesmo modo, prosseguimos com a determinação das funções restantes $y_3(t),..., y_N(t)$, e assim por diante. O leitor pode comprovar que esse é um conjunto mutuamente ortogonal.

APÊNDICE D

PROPRIEDADES BÁSICAS MATRIZES E OPERAÇÕES COM MATRIZES

D.1 Notação

Um vetor-coluna x $n \times 1$ consiste em n entradas e é formado por

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{D.1a}$$

O transposto de x é um vetor-linha, representado por

$$x^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \tag{D.1b}$$

O transposto conjugado de x também é um vetor-linha, escrito como

$$x^H = (x^*)^T = \begin{bmatrix} x_1^* & x_2^* & \cdots & x_n^* \end{bmatrix} \tag{D.1c}$$

x^H também é conhecido como hermitiano de x.

Uma matriz $m \times n$ consiste em n vetores-coluna

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \tag{D.2a}$$

$$= \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \tag{D.2b}$$

Também definimos a transposta e a hermitiana de A como, respectivamente,

$$A^T = \begin{bmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & \cdots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{m,n} \end{bmatrix} \qquad A^H = \begin{bmatrix} a_{1,1}^* & a_{2,1}^* & \cdots & a_{m,1}^* \\ a_{1,2}^* & a_{2,2}^* & \cdots & a_{m,2}^* \\ \vdots & \vdots & \cdots & \vdots \\ a_{1,n}^* & a_{2,n}^* & \cdots & a_{m,n}^* \end{bmatrix} \tag{D.2c}$$

- Se $A^T = A$, dizemos que A é uma matriz simétrica.
- Se $A^H = A$, dizemos que A é uma matriz hermitiana.
- Se $A^H = A$ e A tiver apenas entradas reais, será hermitiana e simétrica.

D.2 Produto de Matrizes e Suas Propriedades

Para uma matriz A $m \times n$ e uma matriz B $n \times \ell$, com

$$B = \begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,\ell} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,\ell} \\ \vdots & \vdots & \cdots & \vdots \\ b_{n,1} & b_{n,2} & \cdots & b_{n,\ell} \end{bmatrix} \tag{D.3}$$

a matriz resultante do produto $C = A \cdot B$ tem dimensão $m \times \ell$ e é dada por

$$C = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & a_{1,\ell} \\ c_{2,1} & c_{2,2} & \cdots & a_{2,\ell} \\ \vdots & \vdots & \cdots & \vdots \\ c_{m,1} & c_{m,2} & \cdots & a_{m,\ell} \end{bmatrix} \quad \text{em que} \quad c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j} \tag{D.4}$$

Em geral, $AB \neq BA$. Na verdade, os produtos podem nem mesmo ser bem definidos. Para que possamos multiplicar A e B, o número de colunas de A deve ser igual ao número de linhas de B.

Em particular, o produto de um vetor-linha por um vetor-coluna é

$$y^H x = \sum_{k=1}^{n} y_k^* x_k \tag{D.5a}$$

$$= < x, y > \tag{D.5b}$$

Portanto, $x^H x = \|x\|^2$.

Dois vetores x e y são ortogonais de $y^H x = x^H y = 0$.

O produto de matrizes tem diversas propriedades de uso frequente:

$$A(B + C) = AB + AC \tag{D.6a}$$

$$A(BC) = (AB)C \tag{D.6b}$$

$$(AB)^* = A^* B^* \tag{D.6c}$$

$$(AB)^T = B^T A^T \tag{D.6d}$$

$$(AB)^H = B^H A^H \tag{D.6e}$$

D.3 Matrizes Identidade e Diagonal

Uma matriz quadrada $n \times n$ é diagonal se todas as entradas fora da diagonal forem zero:

$$D = \text{diag}\,(d_1, d_2, \ldots, d_n) \tag{D.7a}$$

$$= \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & 0 & d_{n-1} & 0 \\ 0 & 0 & \cdots & 0 & d_n \end{bmatrix} \tag{D.7b}$$

Uma matriz identidade I_n tem entradas diagonais unitárias:

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & 1 \end{bmatrix}_{n \times n}$$

Para uma matriz quadrada A $n \times n$, se existir uma matriz quadrada B $n \times n$ tal que

$$BA = AB = I_n$$

então

$$B = A^{-1} \tag{D.9}$$

é a inversa de A. Por exemplo, dada uma matriz diagonal,

$$D = \text{diag}\,(d_1, d_2, \ldots, d_n)$$

$$D^{-1} = \text{diag}\left(\frac{1}{d_1}, \frac{1}{d_2}, \ldots, \frac{1}{d_n}\right)$$

D.4 Determinante de Matriz Quadrada

O **determinante** de uma matriz quadrada A $n \times n$ é definido recursivamente por

$$\det(A) = \sum_{i=1}^{n} a_{i,j}(-1)^{i+j} M_{i,j} \tag{D.10}$$

em que $M_{i,j}$ é um matriz $(n-1) \times (n-1)$, conhecida como o **menor** de A, obtida pela eliminação da i-ésima linha e j-ésima coluna de A. Especificamente, para uma matriz 2×2,

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Como base na definição de determinante, para um escalar α,

$$\det(\alpha A) = \alpha^n \det(A) \tag{D.11a}$$

$$\det\left(A^T\right) = \det(A) \tag{D.11b}$$

Para uma matriz identidade,

$$\det(I) = 1 \tag{D.11c}$$

Para duas matrizes quadradas A e B,

$$\det(AB) = \det(A)\det(B) \tag{D.11d}$$

Portanto,

$$\det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right) = 1 \tag{D.11e}$$

Para uma matriz A $m \times n$ e uma matriz B $n \times m$, temos

$$\det(I_m + AB) = \det(I_n + BA) \tag{D.12}$$

D.5 Traço

O traço de uma matriz quadrada A é a soma de suas entradas diagonais:

$$\mathrm{Tr}\,(A) = \sum_{i=1}^{n} a_{i,i} \tag{D.13}$$

Para uma matriz A $m \times n$ e uma matriz B $n \times m$, temos

$$\mathrm{Tr}\,(AB) = \mathrm{Tr}\,(BA) \tag{D.14}$$

D.6 Autodecomposição

Se uma matriz quadrada A $n \times n$ for hermitiana, a equação

$$Au = \lambda u \tag{D.15}$$

específica um autovalor λ e o associado autovetor u.

Quando A é hermitiana, seus autovalores são números reais. Além disso, A pode ser decomposta como

$$A = U \Lambda U^H \tag{D.16}$$

em que a matriz

$$U = [u_1 \; u_2 \; \cdots \; u_n] \tag{D.17}$$

consiste em autovetores ortogonais, tal que

$$UU^H = I_n \tag{D.18}$$

Matrizes que satisfazem essa igualdade são denominadas matrizes unitárias.

Além disso, a matriz diagonal

$$\Lambda = \mathrm{diag}\,(\lambda_1 \; \lambda_2 \; \cdots \; \lambda_n) \tag{D.19}$$

consiste nos correspondentes autovalores de A.

Como

$$U^H U = UU^H = I_n \tag{D.20}$$

também podemos escrever

$$U^H A U = \Lambda \tag{D.21}$$

Os autovalores de A têm características muito úteis. Em particular,

$$\det\,(A) = \prod_{i=1}^{n} \lambda_i \tag{D.22a}$$

$$\mathrm{Trace}\,(A) = \sum_{i=1}^{n} \lambda_i \tag{D.22b}$$

D.7 Matrizes Quadradas Hermitianas Especiais

Seja uma matriz A $n \times n$ hermitiana. A é **positiva definida** se, para qualquer vetor $x \neq 0$ $n \times 1$, tivermos

$$x^H A x > 0 \tag{D.23}$$

A é **positiva semidefinida** se, para qualquer vetor x $n \times 1$, tivermos

$$x^H A x \geq 0 \tag{D.24}$$

A é **negativa definida** se, para qualquer vetor $x \neq 0$ $n \times 1$, tivermos

$$x^H A x < 0 \tag{D.25}$$

A é positiva definida se e somente se todos seus autovalores forem positivos.

APÊNDICE E

MISCELÂNEA

E.1 Regra de L'Hôpital

Caso $\lim f(x)/g(x)$ resulte na forma indeterminada 0/0 ou ∞/∞,

$$\lim \frac{f(x)}{g(x)} = \lim \frac{\dot{f}(x)}{\dot{g}(x)} \tag{E.1}$$

E.2 Séries de Taylor e de Maclaurin

$$f(x) = f(a) + \frac{(x-a)}{1!}\dot{f}(a) + \frac{(x-a)^2}{2!}\ddot{f}(a) + \cdots$$

$$f(x) = f(0) + \frac{x}{1!}\dot{f}(0) + \frac{x^2}{2!}\ddot{f}(0) + \cdots$$

E.3 Séries de Potências

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots$$

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \cdots$$

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \frac{17x^7}{315} + \cdots \qquad x^2 < \frac{\pi^2}{4}$$

$$Q(x) = \frac{e^{-x^2/2}}{x\sqrt{2\pi}}\left(1 - \frac{1}{x^2} + \frac{1\cdot 3}{x^4} - \frac{1\cdot 3\cdot 5}{x^6} + \cdots\right)$$

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots + \binom{n}{k}x^k + \cdots + x^n$$

$$\approx 1 + nx \qquad |x| \ll 1$$

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \qquad |x| < 1$$

E.4 Somas

$$\sum_{m=0}^{k} r^m = \frac{r^{k+1}-1}{r-1} \qquad r \neq 1$$

$$\sum_{m=M}^{N} r^m = \frac{r^{N+1}-r^M}{r-1} \qquad r \neq 1$$

$$\sum_{m=0}^{k} \left(\frac{a}{b}\right)^m = \frac{a^{k+1}-b^{k+1}}{b^k(a-b)} \qquad a \neq b$$

E.5 Números Complexos

$$e^{\pm j\pi/2} = \pm j$$

$$e^{\pm jn\pi} = \begin{cases} 1 & n \text{ par} \\ -1 & n \text{ ímpar} \end{cases}$$

$$e^{\pm j\theta} = \cos\theta \pm j\,\mathrm{sen}\,\theta$$

$$a + jb = re^{j\theta} \qquad r = \sqrt{a^2+b^2}, \quad \theta = \tan^{-1}\left(\frac{b}{a}\right)$$

$$(re^{j\theta})^k = r^k e^{jk\theta}$$

$$(r_1 e^{j\theta_1})(r_2 e^{j\theta_2}) = r_1 r_2 e^{j(\theta_1+\theta_2)}$$

E.6 Identidades Trigonométricas

$$e^{\pm jx} = \cos x \pm j\,\mathrm{sen}\,x$$

$$\cos x = \frac{1}{2}(e^{jx} + e^{-jx})$$

$$\mathrm{sen}\,x = \frac{1}{2j}(e^{jx} - e^{-jx})$$

$$\cos\left(x \pm \frac{\pi}{2}\right) = \mp\,\mathrm{sen}\,x$$

$$\mathrm{sen}\left(x \pm \frac{\pi}{2}\right) = \pm\cos x$$

$$2\,\mathrm{sen}\,x\cos x = \mathrm{sen}\,2x$$

$$\mathrm{sen}^2 x + \cos^2 x = 1$$

$$\cos^2 x - \mathrm{sen}^2 x = \cos 2x$$

$$\cos^2 x = \frac{1}{2}(1 + \cos 2x)$$

$$\mathrm{sen}^2 x = \frac{1}{2}(1 - \cos 2x)$$

$$\cos^3 x = \frac{1}{4}(3\cos x + \cos 3x)$$

$$\mathrm{sen}^3 x = \frac{1}{4}(3\,\mathrm{sen}\,x - \mathrm{sen}\,3x)$$

$$\mathrm{sen}\,(x \pm y) = \mathrm{sen}\,x\cos y \pm \cos x\,\mathrm{sen}\,y$$

$$\cos(x \pm y) = \cos x \cos y \mp \operatorname{sen} x \operatorname{sen} y$$

$$\tan(x \pm y) = \frac{\tan x \pm \tan y}{1 \mp \tan x \tan y}$$

$$\operatorname{sen} x \operatorname{sen} y = \frac{1}{2}[\cos(x-y) - \cos(x+y)]$$

$$\cos x \cos y = \frac{1}{2}[\cos(x-y) + \cos(x+y)]$$

$$\operatorname{sen} x \cos y = \frac{1}{2}[\operatorname{sen}(x-y) + \operatorname{sen}(x+y)]$$

$$a \cos x + b \operatorname{sen} x = C \cos(x+\theta)$$

$$\text{em que } C = \sqrt{a^2+b^2} \quad \text{e} \quad \theta = \tan^{-1}\left(\frac{-b}{a}\right)$$

E.7 Integrais Indefinidas

$$\int u\,dv = uv - \int v\,du$$

$$\int f(x)\dot{g}(x)\,dx = f(x)g(x) - \int \dot{f}(x)g(x)\,dx$$

$$\int \operatorname{sen} ax\,dx = -\frac{1}{a}\cos ax \qquad \int \cos ax\,dx = \frac{1}{a}\operatorname{sen} ax$$

$$\int \operatorname{sen}^2 ax\,dx = \frac{x}{2} - \frac{\operatorname{sen} 2ax}{4a} \qquad \int \cos^2 ax\,dx = \frac{x}{2} + \frac{\operatorname{sen} 2ax}{4a}$$

$$\int x \operatorname{sen} ax\,dx = \frac{1}{a^2}(\operatorname{sen} ax - ax\cos ax)$$

$$\int x \cos ax\,dx = \frac{1}{a^2}(\cos ax + ax \operatorname{sen} ax)$$

$$\int x^2 \operatorname{sen} ax\,dx = \frac{1}{a^3}(2ax \operatorname{sen} ax + 2\cos ax - a^2x^2\cos ax)$$

$$\int x^2 \cos ax\,dx = \frac{1}{a^3}(2ax\cos ax - 2\operatorname{sen} ax + a^2x^2 \operatorname{sen} ax)$$

$$\int \operatorname{sen} ax \operatorname{sen} bx\,dx = \frac{\operatorname{sen}(a-b)x}{2(a-b)} - \frac{\operatorname{sen}(a+b)x}{2(a+b)} \qquad a^2 \neq b^2$$

$$\int \operatorname{sen} ax \cos bx\,dx = -\left[\frac{\cos(a-b)x}{2(a-b)} + \frac{\cos(a+b)x}{2(a+b)}\right] \qquad a^2 \neq b^2$$

$$\int \cos ax \cos bx\,dx = \frac{\operatorname{sen}(a-b)x}{2(a-b)} + \frac{\operatorname{sen}(a+b)x}{2(a+b)} \qquad a^2 \neq b^2$$

$$\int e^{ax}\,dx = \frac{1}{a}e^{ax}$$

$$\int xe^{ax}\,dx = \frac{e^{ax}}{a^2}(ax-1)$$

$$\int x^2e^{ax}\,dx = \frac{e^{ax}}{a^3}(a^2x^2 - 2ax + 2)$$

$$\int e^{ax} \operatorname{sen} bx\,dx = \frac{e^{ax}}{a^2+b^2}(a \operatorname{sen} bx - b\cos bx)$$

$$\int e^{ax} \cos bx \, dx = \frac{e^{ax}}{a^2 + b^2}(a \cos bx + b \operatorname{sen} bx)$$

$$\int \frac{1}{x^2 + a^2} \, dx = \frac{1}{a} \tan^{-1} \frac{x}{a}$$

$$\int \frac{x}{x^2 + a^2} \, dx = \frac{1}{2} \ln (x^2 + a^2)$$

ÍNDICE

A

ADSL (*asymmetric digital subscriber line*), 654
AFH (*adaptive frequency hopping*), 574
AGC (*automatic gain control*), 96
Alargamento, 92
Aleatoriedade, 13
Algoritmo
de Viterbi, 771
com saída suave, 783
soma-produto para decodificação LDPC, 795
AM (*amplitude modulation*), 10, 131
geração de sinais, 145
AMI (*alternate mark inversion*), 303
Amostra(s)
elemento, 365
espaço de, 365
ponto de, 365
Amostragem
aleatória da saída de um gerador de onda senoidal, 397
conversão analógico-digital e, 232-301
da função impulso unitário, propriedades de, 25
prática não ideal, análise de, 243
teorema da, 6, 232
Amplitude, modulações e demodulações em, 130-186
AMPS, 4
Análise
de desempenho de sistema(s)
binários genéricos, 477
de comunicações digitais, 468-566
de QPSK, 527
QAM *M*-ária, 512
Ângulo
generalizado, 188
modulações e demodulações em, 187-231
Antimascaramento, filtro de, 241
Armazenagem elástica, 266
Armstrong
Edwin H., 203, 204
método indireto de, 208
ARQ (*automatic repeat request*), 746
Árvore de código, 767, 775
ASK (*amplitude shift keying*), 345
binária, 481
detecção, 349
Atraso
de envelope, 88
de fase, 88
de grupo, 88
Autocorrelação
função de, 34
método de, 107
Autodecomposição, 820
AWGN (*additive white gaussian noise*), 9, 495
Axiomas básicos, 377

B

Banco de canais, 267
Banda
base, 130
de transmissão, largura da, 257
largura da, 8
lateral
simples (SSB), 148
detecção de sinais, 153
representação no domínio do tempo, 149
vestigial (VSB), 155
possante, 664
Barganha entre
largura de banda e potência em sinais *M*-ários ortogonais, 523
potência e largura de banda, 511
Base
de sinais, 487
ortogonal
completa, 35
determinação de um conjunto de, 489
Baud, taxa de, 343
Bayes
receptor de, 533
regra de, 370
de decisão, 533
teorema de, 376
Bell, Alexander Graham, 15
BER (taxa de erro de bit), 510
de sinalização ortogonal, 522
Bernoulli, repetições de, 371
Bessel, funções de, 43

BIBO (*bounded-input–bounded-output systems*), 85
Bit(s), 7, 249, 683
de enquadramento, 261
de quantização em *vocoder* LPC-10, alocação, 280
entrelaçamento, 263
transmitidos, 324
Bloco de pixels, 284
Bluetooth, 573
BPSK, 480
Branly, Edouard, 15
Broadcasting, 132
BSC (*binary symetric channel*), 372, 691
comunicação sem erro em um, comprovação, 699

C

Canal(is), 3
assíncronos, 265
AWGN limitado em banda, capacidade de um, 708
capacidade do, 8, 9, 692
de um canal discreto sem memória, 694
medida da, 699
por segundo, 699
com desvanecimento, 96, 659
com ruído gaussiano branco aditivo, 9
comum de sinalização entre centrais, 263
contínuo
comprovação de comunicação sem erro em um, 712
sem memória, capacidade de um, 701
de comunicação, distorção de sinal em um, 91
de largura de banda infinita, capacidade de, 711
distorção causada por não linearidade do, 93
em fase, 154
em quadratura, 154
estimação de, 637
matriz de, 695
MIMO, capacidade de, 723
receptor, equalização do, 620
seletivo em frequência, 3
capacidade, 719
Captura, efeito de, 218
Carregamento
de potência para máxima capacidade, 720
ótimo de potência
como enchimento com água, interpretação do, 722
em OFDM/DMT, 723
CCIS (*common channel interoffice signaling*), 263
CDMA (*code division multiple-access*), 571
de DSSS, múltiplo acesso por, 582
de Hollywood a, 575
em redes de telefonia celular, 595
no sistema de posicionamento global, 598
potência em, controle de, 588
cdmaOne(IS-95), 596
Central de comutação de telefonia móvel, 595
Chappe, Claude, 15
Chaveamento (de)
amplitude, modulação por, 345
fase, 193
on-off, 345
por deslocamento
de amplitude, 536
de fase, 345
de frequência, 345, 483, 539
mínimo, 484
Circuito(s)
balanceados, 136
regenerativo, 204
super-heteródino, 204
Codificação(ões)
bipolar de alta densidade, 317
conjunta fonte-canal, 14
de fonte, 12, 13
de transmissão, 303
diferencial, 326, 350, 351
digitais de fonte, 12
limiar de, 275
para correção de erros, 12, 13
Codificador
convolucional, 767
com comprimento de restrição, 766
de dígito por vez, 256
de linha, 303
PCM, 256
Código(s)
0001, 14
00011, 14
árvore de, 767
binário
10010110, 257
natural (CBN), 249
cíclicos, 753
geração, 759
matriz geradora de, 758
sistemáticos, 756
com teste de paridade simples, 691
combinação de, 778
compacto, 688
concatenados, 780
convolucionais, 743, 766
sistemático recursivo, 770
corretores de erro, 743-809
exemplos, alguns, 745
de blocos, 743
lineares, 746, 748
de Bose-Chaudhuri-Hocquenghen, 761
de correção de erro à frente, 570
de DSSS, múltiplo acesso por divisão de código, 582
de fase alternada, 312
de Hamming, 746
construção de, 752
de Huffmann, 687
de linha, 305
duobinário modificado, 303
PSD de diversos, 305
de Reed-Solomon, 761
de transmissão, 305
eficiência de, 688
entrelaçamento de, 778
lineares, 748
Manchester, 312
Morse, 4

para corrigir erros em rajadas, especiais, 338
para detecção e correção de erros em rajada, 765
perfeitos, 746
-produto, 779
sistemático, 746
taxa de, 744
-turbo, 784, 788
decodificação iterativa para, 790
verificadores de
paridade de baixa densidade, 792
redundância cíclica para detecção de erros, 761
Coeficiente(s)
de correlação, 32, 406
de predição, 269
de reflexão parcial, 280
normalizados e quantizados da DCT do resíduo, 284
Coerência em fase, 166
Colina do Telégrafo, 15
Combinação de códigos, 778
Compandor, 255
Complemento, 365
Componente(s)
de um sinal, 28
de um vetor na direção de outro vetor, 27
em quadratura, 454
Compressão de vídeo, 278, 281
Comprimento máximo, sequência de, 578
Computador mainframe, 18
Comunicação(ões)
binária na presença de ruído, 497
digital(is)
com canais sujeitos a distorção linear, 617, 680
vantagens, 249
elétrica, 1
em banda de base *versus* comunicação por portadora, 130
OFDM (multiportadora), 640
por espalhamento espectral, 567-616
por fibras ópticas, 15
por portadora, 131
sem erro em um canal ruidoso, 691
sistemas de, 1, 2
Condições de Dirichlet, 64
Conjunto(s)
completo, 35
ortogonal, 36
de sinais, 43
Constante
fator de decaimento, 322
r, 322
Contador de frequência, 215
Controle
automático de ganho, 96
de potência *versus* MUD, 598
Conversão(ões)
analógico-digital, 6
ascendente, 140
de canais com desvanecimento seletivo, 662
descendente, 140
Conversor de frequências, 139
ascendente, 568
Convolução
no domínio
da frequência, 82
do tempo, 82
teorema da, 81
Correlação, 404
coeficiente de, 32, 406
cruzada, 33
de sinais, 32
funções, 33
Cosseno levantado, 323
Covariância, 404
CPFSK (*continuous phase FSK*), 484
CRC (*cyclic redundancy check codes*), 761
códigos, 762
Critério(s)
de Paley-Winer, 90
Cross talk, 338

D

DAB (*digital audio broadcasting*), 657
Dado(s)
digitais, princípios de transmissão, 302-363
embaralhamento de, 329
DDS (*digital data system*), 267
Decaimento
de cosseno amplo, 323
fator de, 322
Decisão
procedimento de, 500
regiões de, 497, 504
Decodificação, 749, 759
com realimentação, 776
de máxima verossimilhança, 771
de Viterbi, 772
iterativa para códigos-turbo, 790
LDPC por troca de bits, 795
sequencial, 774
suave, 780
Deênfase
filtros de, 219
na difusão FM, 219
Delta de Dirac, 24
Demodulação(ões), 11, 12
de sinais FM, 213
FM via PLL, 215
modulação e
em amplitude, 138-186
em ângulo, 187-231
Demodulador(es)
coerentes, 140
de frequência práticos, 214
homódinos, 140
síncrones, 140
Densidade(s)
condicionais, 393
de probabilidade
gaussiana, 385
normal, 385
de Rayleigh, 394
de Rice, 461
espectral, 65
de energia, 97, 98, 102
da entrada e da saída, 103
de potência, 105, 429
cruzada, 443
sinais, 104
marginais, 392
riceana, 461
Descontinuidade, pontos de, 114, 192

Desempenho, critérios úteis de, 532
Desertores, eliminação de, 241
Desigualdade
de Cauchy-Schwarz, 33, 812
de Chebyshev, 403
de Hadamard, 729
Deslocamento
Doppler, 659
registrador de, 748
Desvanecimento
canais com, 96, 659
de canal seletivo em frequência, 660
plano, canais com, 661
seletivo de frequência, 95
Desvio-padrão, 399
Detecção, 10
coerente, 134
de envelope de sinais VSB+C, 167
de inclinação, 214, 215
de sinalização duobinária, 326
não coerente, 536
ótima de PSK DSSS, 577
por limiar, 389
síncrona, 134
Detector
de cruzamento zero, 215
de fase, 161
de razão, 215
linear ótimo para sinalização polar binária, 468
retificador, 145
DFE (*decision feedback equalizer*), 639
erro em, propagação de, 640
DFT, 110
cálculo, 114
Diagrama
de espalhamento, 404
de olho, 338, 339
em PAM, 343
de transição de estados, representação, 769
em treliças, 770
de códigos de blocos, 777
Difusão
de áudio digital, 657
digital, 656
Dígito(s)
binário, 7
de verificação de paridade, 747
Dirichlet, condições de, 64
Discriminador balanceado, 215
Dispersão, 92
Distância de Hamming, 692, 747
Distorção(ões)
causada por
não linearidade do canal, 93
por efeitos de multipercurso, 95
de sinal
durante a transmissão, 86
em um canal de comunicação, 91
em canal variante no tempo devido a mobilidade, 659
em sinais de áudio e de vídeo, natureza da, 89
lineares, 3, 91
em canais sem fio de multipercurso, 617
não linear, efeitos de, 216
transmissão sem, 87
Distribuição conjunta, 391
Dividir para conquistar, 375
DM (*delta modulation*), 131, 272
DMT (*discrete multitone*), 650
aplicações práticas, 654
carregamento de bits nas subportadoras em, 652
Dobramento, frequência de, 114
Downconversion, 140
DPCM (*differential pulse code modulation*), 268
análise da, 270
DS1 (*digital signal level 1*), 261
DSB-SC (*double-sideband supressed carrier*), 132
DSSS (*direct sequence spread spectrum*), 577
resiliência de, características, 580
sob interferências de banda
estreita, análise, 581
larga, análise, 582
Dualidade
propriedade, 73
tempo-frequência, 71
Duas porções de informação por segundo por Hertz, 242

E

Embaralhamento de dados, 329
Encurtamento de canal, 653
Energia
da soma de sinais ortogonais, 31
de sinal(is), 488
modulados, 101
densidade espectral de, 102
e densidade espectral de energia de sinais, 96
sinais de, 24
unidades de, 20
Enquadramento, bit de, 261
Ensemble, 421
estatística de, por que precisamos, 424
Entradas, 19
Entrelaçador, 778
aleatório, 779
de blocos, 779
Entrelaçamento
de bits, 263
de códigos, 778
de palavras, 263
profundidade de, 779
Entropia
condicional, 695
de ruído gaussiano branco limitado em banda, 705
de uma fonte, 684
diferencial, 702
Envelope, 142
condições para detecção de, 143
de sinais VSB+C, detecção, 156
detector de, 146
Equação de Shannon, 716
Equalização
adaptativa, 335
cega, 658
do canal do receptor, 620
linear com espaçamento T, 626
TSE com forçamento a zero, 627
projeto com base em MMSE, 629
OFDM, 649

Equalizador(es)
com alimentação para a frente, 638
com forçamento a zero, 332
com realimentação de decisão, 638, 639
lineares fracionalmente espaçados, 633
configuração, 635
FSE MMSE, 637
MMSE de comprimento finito, 631
Erro(s)
codificação para correção de, 13
códigos corretores de, 743-809
correção de, efeitos da, 762
de bit, probabilidade de, 476
de mascaramento, 113
detecção de, 337
probabilidade de, 504
arbitrariamente pequena, 743
de receptores ótimos, expressão geral para, 517
quadrático médio, 280
total em PCM, 402
redundância para correção de, 744
vetor de, 28
ESF (*extended superframe*), 262
Esfera de *D* dimensões, volume de uma, 713
Espaço, 19
de amostras, 365
de sinais, 487
análise de detecção ótima no, 485
de detecção, dimensionalidade do, 497
interpretação geométrica no, 505
geométrico de sinais, 485
ortogonal de sinais, 35
vetorial ortogonal, 34
Espalhamento
diagrama de, 404
espectral
comunicações por, 567-616
no GPS, por quê?, 599
por saltos de frequência, sistemas de, 567
por sequência direta, 577
Espectro
de sinal de televisão, 158
de Fourier exponencial, 38, 39
vestigial, 322
Esquema(s)
de resposta
correlativa, 324
parcial, 324
NRZ, 304
sem retorno ao zero, 304
Estações-imagem, 222
Estimação
da sequência de máxima verossimilhança, 623
complexidade, 626
de canal, 637
quadrática média linear, 407
Evento(s), 364, 365
disjuntos, 366
independentes, 369-371
mutuamente exclusivos, 366, 371
nulo, 365
união de, 365
Expansão de *Karhunen-Löeve*, 490, 491
Experimento, 364, 365
Explosão de cor, 155
Exponencial da SNR de saída, aumento, 257
Expressão geral para a probabilidade de erro de receptores ótimos, 517
Extensões de segunda e terceira ordens, 690

F

Faixa
de aquisição, 166
de captura, 166
de retenção, 166
de travamento, 166
Falso(a)
alarme, 535
largada, 187
rejeição, 535
Faraday, Michael, 15
Fase, chaveamento de, 193
Fator de decaimento, 322
FDM (*frequency division multiplexing*), 12, 130,160
FEC (*forward error correction codes*), 570
FFT (*fast Fourier transform*), 115
FFW (*feedforward*), 638
Filtragem, 119
ótima, 447
Filtro(s)
antimascaramento *versus* filtro casado, 241, 621
casado, 470, 471
passa-faixa como receptor coerente, 482
de reconstrução, realizabilidade de, 239
de retenção de primeira ordem, 238
de Wiener-Hopf, 447
ideais *versus* filtros práticos, 89
ótimo, 448
receptor ótimo, 470
transversal, 270
VSB complementar, 156
FM (*frequency modulation*), 11
de banda
estreita, 194
larga, 195
análise, 195
Fonte, 2, 302
codificação de, 686
entropia de uma, 684
sem memória, 684
Forçamento a zero, 627
Formatação de pulso, 317, 343
Fórmula de interpolação, 235
Fourier
espectro exponencial, 38, 39
integral de, 62
série exponencial, 37
transformada de, direta, 62
Frame, 261
Framing bit, 261
Frequência
chaveamento de, 192
contador de, 215
de dobramento, 114
desvanecimento seletivo em, 95

instantânea, 188
livre, 162
modulação em, 187
multiplicadores de, 208
negativa, o que significa, 93
relativa, 366
resolução de, 114
FSK (*frequency shift keying*), 192
de fase contínua, 484
detecção, 349
M-ária, 352
Função(ões)
de amostra, 421
de autocorrelação, 34, 108
de um processo aleatório, 424
temporal, 102
de sinais de potência, 105
de base, 36
para um processo aleatório, determinação de, 490
de Bessel, 43
modificada de ordem zero, 461
de correlação, 33
cruzada, 442
de densidade espectral de potência, 108
de Laguerre, 43
de transferência, 86
de um filtro passa-baixas, 92
de Walsh, 43
degrau unitário, 26
densidade de
de Rayleigh, 395
probabilidade, 383, 384
distribuição cumulativa, 382, 383
impulso unitário, propriedade de amostragem da, 25
-kernel, 490
-núcleo, 490
por um impulso, multiplicação de uma, 25
$Q(x)$, 387
retangular unitária, 66
sinc sinc(x), 67
triangular unitária, 66

G

Gauss, Carl Friedrich, 385
GPS (*global positioning system*), 598
como funciona, 599
o que é?, 598
por que espalhamento espectral no, 599
Gráfico
bipartido, 794
de Tanner, 794
Grupo
básico de canais, 161
de imagens, 285
mestre, 161

H

Hadamard, desigualdade de, 729
Hamming
código de, 746
distância de, 692, 747
limite de, 745
Hartley, 684
HDB (*high-density bipolar*), 317
HDTV, 286
Heliógrafos, 15
Hertz, Heinrich, 15
Hierarquia digital plesiócrona, 266
Hilbert, transformada de, 148

I

Identidades trigonométricas, 823
IEEE
802.11, 573
802.11b
modulação e fator de espalhamento na transmissão, 601
na banda ISM de 2,4 GHz, alocação de canais no padrão, 601
Impulso(s)
de pressão, 279
de timbre, 278
multiplicação de uma função por um, 25
unitário, 25
sinal, 24
Imunidade
da modulação em ângulo a não linearidades, 216
de sinais digitais ao ruído, 4
Incerteza temporal, 336, 337
no cruzamento de nível, 340
sensibilidade à, 340
Inclinação, sobrecarga de, 275
Independência *versus* descorrelação, 406
Índice de modulação, 197
Informação
extrínseca, 792
média por mensagem, 684
medida de
no senso da engenharia, 682
pelo senso comum, 681
mútua, 695, 706
teoria da, introdução a, 681-742
unidade de, 684
Integral(is)
de Fourier, 62
definidas, 824
Interferência, 216
cocanal, 155
devido a ruído de canal, 218
efeito, 217
intersimbólicas, 318
Interpolação
fórmula, 235
ideal, 235
Interpretação
de engenharia da entropia, 685
intuitiva, 685
Interseção, 366
Inversão alternada de sinal, 303
ISDN (*integrated services digital net works*), 304
ISI (*intersymbol interference*), 92, 318
controlada, 323
em comunicação sem fio, 620
Isolação de uma causa particular, 376

J

Jamming, 567
Justificação, 266

K

Karhunen-Löeve, expansão de, 490

L

Laguerre, funções de, 43
Lamarr
-Antheil, patente, 576
Hedy, 575
Lâmpadas de Aldis, 15
LAR (*log-area ratios*), 280
Largura da banda, 8
do produto de dois sinais, 82
essencial, 310
de um sinal, 98
LDPC (*low-density parity check codes*), 793
irregulares, 794
regulares, 794
Legendre, polinômios de, 43
Leis básicas, 377
Liga-desliga, 303
Limiar
de detecção binária, 469
direção por, 389
ótimo, 475
Limitador passa-faixa, 207
Limite
central, teorema do, 412
de Amin, 745
Linha de assinante digital assimétrica, 654
Lodge, Oliver, 15
Lógica
dedutiva, 377
indutiva, 377
LP (*linear prediction*), 280
LPC (*linear prediction coding*), 278
modelos, 279
LSP (*line spectral pairs*), 280

M

Maclaurin, série de, 93
MAI (*multiple-access interference*), 582
não ortogonal, aproximação gaussiana, 585
Malha
de captura de fase. *Veja também* PLL, 161
de costas, 167
de primeira ordem, análise de, 164
travada, 162
Mascaramento
efeito de, 240
erro de, 113
perfídia do, 239
MATLAB, exercícios com, 43
amostragem e reconstrução de sinais passa-baixas, 287
cálculo de transformada de Fourier, 115
capacidade de canal
e informação mútua, 733
MIMO, 735
CDMA de multiusuário em ambiente próximo-distante, 610
código
de Hamming, 801
Huffman, 731
comunicação FSK FHSS sujeita a interferência de banda estreita, 603
decodificação do código de blocos linear, 798
detecção
FSK não coerente, 553
não coerente de PSK diferencial binária, 555
diagramas de olho, 357
equalização
com realimentação de decisão, 659
linear de QAM-16, 662
filtragem, 119
modulação
delta, 294
e demodulação
AM, 172
DSB-SC, 168
FM, 224
QAM, 177
SSB-SC, 175
QAM-16, 551
PCM, exemplo de, 291
sinalização
binária *on-off*, 548
polar binária com pulsos diferentes, 543
sistema CDMA-DS multiusuário, 607
transmissão
DSSS de QPSK, 605
OFDM de sinais QAM, 672
Matriz(es)
básicas, propriedades, 817
de canal, 695
de verificação de paridade, 749
geradora, 747
de códigos cíclicos, 758
identidade e diagonal, 818
operações com, 817
produto de, 818
quadrada(s)
determinante, 819
hermitianas especiais, 821
Média(s)
da soma, 398
de uma função de uma variável aleatória, 396
do produto de duas funções, 398
estatísticas, 395
Mensagem(ns)
analógicas, 3
binária, 4
digitais, 3
M-ária, 4
Método
de autocorrelação, 107
de deslocamento de fase, 152
de filtragem seletiva, 152
do mínimo erro quadrático médio, 335
indireto de Armstrong, 208
quadrático, 166
Weaver, 153
MFSK. *Veja também* Sinalização multitom, 521
não coerente, 540
receptor coerente, 521
MIMO (*multiple-input–multiple-output*), 662
canais, capacidade de, 723

Minimax
conceito, explicação, 535
receptor, 534
Misturador de frequências, 139
MMSE (*minimum mean square error*) *versus* ZF, 335, 627, 632
Modelo uma entrada, múltiplas saídas, 634
Modem(s), 344
para redes de área local sem fio, 18
Modulação(ões), 10
binárias básicas com portadora, 344
com portadora, 131
delta, 131, 272
adaptativa, 277
digital(is)
com portadora, receptores coerentes para, 480
M-ária com portadora, 351
e demodulações
em amplitude, 130-186
comunicação em banda *versus* comunicação por portadora, 130
em ângulo, 187-231
em amplitude, 10, 131, 141
de pulso, 247
em quadratura, 147, 153, 155
geração de sinais, 145
em ângulo, 131
de banda estreita, aproximação para, 194
imunidade a não linearidades, 216
propriedades da, 212
em amplitude, 131
banda lateral vestigial, 155
com banda lateral dupla, 132
com eficiência de largura de banda, 147
em banda lateral
dupla com portadora suprimida, 132
simples, 147
em exponencial, 189
em fase, 11, 197
em frequência, 11, 187, 189. *Veja também* FM
por tom, análise espectral, 198
exponencial, 189
índice de, 197
multitons discretos, 650
não linear, 187
por amplitude de pulso, 342
por chaveamento de amplitude, 345
por codificação de pulsos, 7, 131, 248
diferencial, 268. *Veja também* DPCM
por tom, 134, 144, 150
sigma-delta, 276
Modulador(es)
balanceados, 136
chaveados, 137
DSB-SC não linear, 136
duplamente balanceado, 139
em anel, 138
em ponte de diodos, 138
multiplicadores, 136
não lineares, 136
por largura do pulso, 247
por posição de pulso, 247
Momento central, 398
MPSK, 515
MUD (*multiuser detection*), 589
com realimentação de decisão, 593
ótima, 590
Multipercurso, 617
distorção causada por efeitos de, 95
Multiplexação, 12
digital, 263
em quadratura, 154
por divisão
em frequência, 12, 160
no tempo, 12, 247
T1, 260
Multiplexador, 304
DM 12, formato de, 265
Múltiplo(s)
acesso por divisão de código, 571
de DSSS, 582
sinais, transmissão simultânea de, 12
Multiusuário(s)
detecção de, 589

N

NBFM (*narrowband FM*), 195
geração de, 205
NBPM (*narrowband PM*), 195
Nós de redes
grandes, 5
regenerativos, 5
Notação, 817
NRZ (*non-return-to-zero*), 304
Nulo DC no PSD via formatação de pulso, construção de um, 311
Números complexos, 823
Nyquist
intervalo de, 234
primeiro critério para ISI nula, 318
taxa de, 234

O

Oersted, Hans Christian, 15
OFDM (modulação por divisão em frequências ortogonais), 617
aplicações práticas de, 654
história, lista curta, 657
canal, ruído em, 646
com preenchimento zero, 648
comunicações, 640
princípios, 641
redundância de prefixo cíclico em, 649
Onda(s)
FM
formas de, 191, 192
geração de, 205
modulada em ângulo
banda de, largura, 193
potência, 190
moduladoras digitais, forma de, 481
PM, formas de, 191, 192
On-off, 303
OOK (*on-off keying*), 345
Operações com
matrizes, 817
transformada de Fourier, propriedades de, 85
Ortogonalidade
de alguns conjuntos de sinais, 810
do conjunto de sinais

exponenciais, 811
trigonométricos, 810
Ortogonalização de Gram-Schmidt, 489
de um conjunto de vetores, 814
procedimento de, 490
Oscilador controlado por tensão, 161, 162

P

Palavras
de código, 3, 682
entrelaçamento de, 263
Paley-Winer, critério de, 90
PAM (*pulse amplitude modulation*), 131, 247, 508
aleatório, trem de pulsos, 437
de diversos códigos de linha, 305
de modulação digital com portadora, 346
sinalização *M*-ária em banda base para taxa de dados mais elevada, 341
PARCOR (*partial reflection coefficients*), 280
Pares de raios espectrais, 280
Paridade
de baixa densidade, 793
dígitos de verificação de, 747
matriz de verificação de, 749
Parseval, teorema de, 36, 37
PCM (*pulse coded modulation*), 7
diferencial adaptativa (ADPCM), 271
em sistemas de portadoras, 259
exemplo de, 8
repetidor, probabilidade de erro em, 372
PDF(s) (*probability density function*), 384
condicionais em PAM, 509
Peneiramento, 25
Peso da palavra de código, 747
Piconet, 574
Picorrede, 574
Pitch, 278
PLL (*phase locked loop*), 161
comportamentos de, generalização, 165
em pequeno erro, análise de, 163
funcionamento básico, 162
PM (*phase modulation*), 11, 131
de banda estreita, 195
PN (*pseudonoise*), 568
sequência, geração da, 578
Polinômio(s)
de Chebyshev, 43
de Hermite, 43
de Jacobi, 43
de Legendre, 43
gerador, 758
Ponto de máxima abertura, 340
Popov, Alexander, 15
Portadora, 10
em DSB-SC, aquisição de, 166
em SSB-SC, aquisição de, 168
local, sincronização da, 159
senoidal, 131, 187
sistemas digitais com, 344
suprimida, 132, 134
Potência(s)
da banda lateral, 144
da portadora, 144
de sinal, 8, 9, 104
e seu valor quadrado médio, 106
de um processo aleatório, 433
de uma onda modulada em ângulo, 190
densidade espectral de, 429
em CDMA, controle de, 588
PPM (*pulse position modulation*), 131
Pré-ênfase
filtros de, 219
na difusão FM, 219
Predição
coeficiente de, 269
linear excitada por código de Qualcomm, 596
Preditor linear, 269
Preenchimento
com zero, 112
OFDM com, 648
de bits, 265
de pulsos, 265, 266
Prefixo cíclico, 653
Princípio da taxação progressiva, 253
Probabilidade
conceito, 364
condicional(is), 369, 380
regra de multiplicação para, 374
de erro, 504
de bit, 476
de detecção, 338
em repetidor PCM, 372
teoria
axiomática da, 377
fundamentos, 364–420
total, teorema da, 375
Processo(s)
aleatórios, 364
análise espectral e, 421–467
classificação, 426
de ruído branco, decomposição vetorial de, 489
estacionários, 426
no sentido mais amplo, 426
função(ões)
de autocorrelação de um, 424
de base para um, determinação, 490
gaussiano(s)
"branco" passa-faixa, 459
propriedades de, 494
independentes, 442
múltiplos, 442
não estacionários, 426
passa-faixa, 454
potência de um, 433
soma de, 445
transmissão por meio de sistemas lineares, 443
binário aleatório, 435
conjuntamente estacionários, 442
de *handoff*, 595
de ruído branco, representação geométrica, 491
descorrelacionados, 442
estacionários no sentido amplo, 426, 427
ergódigos, 428
fracamente estacionários, 426, 427
incoerentes, 442
ortogonais, 442
Produto escalar, 488
Projeto com dados finitos, 632

Propriedade de amostragem, 25
Próximo-distante, problema, 586
resistência ao, 588
PSD (*potencial spectral density*), 105, 430
Pseudoternário, 303
PSK (*phase shift keying*), 345
diferencial, 350
detecção da, 542
diferencialmente coerente, 541
direção, 350
Pulso
duobinário, 324
formatação de, 317, 343
geração de, 328
preenchimento de, 265, 266
quadrado, 3
relações de, 325
retangular, 66
e correspondente espectro de Fourier, 68
sinc, 67
triangular, 66

Q

QAM (*quadrature amplitude modulation*), 147
modelos genéricos, 619
QCELP (*Qualcomm code-excited linear prediction*), 596
Quadro, 261
Quantização, 6, 250
não uniforme, 253, 254
níveis de, 248
ruído de, 251

R

Radiação(ões)
facilidade de, 11
intergaláctica, 3
Radiofrequência (RF), 10
Rainey, P. M., 7
Raiz de cosseno levantado, 624
Rayleigh
densidade de, 394
função, 395
Razões logarítmicas de área, 280
Receptor(es)
analógicos AM/FM
super-heterodinos, 221
binários ótimos equivalentes, 477
coerentes para modulações digitais com portadora, 480
com realimentação de decisão, 593
de Bayes generalizado, 533
de detecção ótima, 468
de máxima verossimilhança, 533, 534, 590
de mínimo erro quadrático médio, 591
descorrelator, 591
digitais, 331
linear ótimo, análise de, 474
minimax, 534
MUD com realimentação e decisão, 594
Reciprocidade entre duração de um sinal e sua largura de banda, 75
Recursão de Levinson-Durbin, 280
Rede(s)
de área local sem fio, 573
de telefonia celular, CDMA em, 595
digital de serviços integradas, 304
regenerativa, 5
Redundância, 12, 13
de prefixo cíclico em OFDM, 649
para correção de erro, 744
Reeves, A. H., 7
Regiões de decisão, 497, 504
ótimas, determinação de, 507
Registrador de deslocamento, 748
Regra (de)
Bayes, 370
Carson, 197
decisão de Bayes, 533
L'Hôpital, 822
multiplicação para probabilidades condicionais, 374
Regularidade estatística, 366
Relação(ões)
entre FM e PM, 190
entre modulações analógica e digital com portadora, 348
sinal-ruído, 6, 8
Repetição(ões), 693
de Bernoulli, 371
solicitação automática de, 746
Repetidor(es)
regenerador, 304
regenerativos, 331
sem distorção, viabilidade de, 5
Representação em quadratura, não unicidade da, 458
Resolução de frequência, 114
Resposta em frequência, 86
Resultados, 364, 365
Rice, densidade de, 461
Ripple, 146
Rms (*root mean square*), 20
Rota
ótima, 771
sobrevivente, 771
RSC (*Recursive Systematic Convolutional code*), 770
Ruído(s), 3
branco
decomposição vetorial de processos aleatórios de, 489
estacionário, 491
processos de, representação geométrica, 491
de canal
colorido, 532
interferência devido a, 218
não branco, 532
de quantização, 251
Dolby de redução de, 220
em sinal senoidal, 460
em um circuito resistivo, cálculo, 446
gaussiano branco, 492
aditivo, 9, 495
caso especial de, 476
receptor ótimo para canais com, 495
dimensionalidade do espaço de sinais de detecção, 497
espaço de sinais e procedimentos de decisão, 500
regiões de decisão e probabilidade de erro, 504
representações geométricas, 495
sinalização multiamplitude, 508

impulsional, 338
imunidade de sinais digitais ao, 4
ortogonal ao espaço de sinais, eliminação do, 499
térmico, 444

S

Saídas, 19
Salto(s)
em frequência
adaptativas, 574
em sincronia, 572
sistemas de espalhamento espectral por, 567
lento, 568
Senso comum, 685
Sequência
de comprimento máximo, 578
de máxima verossimilhança, estimação da, 623
Série(s)
de Fourier
exponencial, 37
generalizada, 36
teorema de Parseval na, 43
de Maclaurin, 93, 822
de potências, 822
de Taylor, 822
Serviços celulares 3G, 597
Shannon, 10
equação de, 716
Shift register, 748
Simetria
conjugada, propriedade de, 63
hermitiana, 63
SIMO (modelo uma entrada, múltiplas saídas), 634
Sinal(is)
a partir de amostras uniformes, reconstrução de, 234
aleatórios, 24
AM, demodulação de, 145
análise e transmissão de, 59-129
analógicos, 22
em tempo contínuo, 23
antipodais, 530
aperiódicos, 23
representação através da integral de Fourier, 59
base de, 485, 487
bifásico, 312
binários, 32
causal, 26
classificação de, 21
com energia finita, 20
com potência finita, 20
componentes de um, 28
conjunto(s)
equivalentes de, 525
ortogonal de, 34, 519
PSK quaternários, 526
conversão analógica-digital de um, 6
correlação de, 32
de base, 36
de mensagem, 2
com deslocamento
não nulo, 143
nulo, 143
de mínima energia, conjunto de, 527
de potência, 24
densidade espectral de, 104
função de autocorrelação temporal de, 105
de radiofrequência, 10
de voz, espectro de, 276
decomposição de um, 28
determinísticos, 24, 364
digitais, 22
L-ário, 248
digital 1, nível de, 261
digital 0, nível de, 267
do relógio no transmissor, 336
em banda base, 2
em tempo
contínuo, 21, 22
discreto, 21, 32
energia de um, 19, 488
espaço de, 487
geométrico, 485
formato de, 264
impulso unitário, 24
interpolação prática de, 236
M-ário, 22
modulado(s), 132
em pulsos, 247
energia de, 101
MSK, 515
multiplexado, 267
on-off, 304
periódicos, 23
potência de um, 20
quantizado, 6
senoidal, 131
simplex, 531
conjunto de, 531
tamanho de um, 19
transmissão simultânea de, 12
transortogonais, conjunto de, 530
vetores e, analogia entre, 26
vivos, 15
Sinalização, 261
binária
com substituição de *N* zeros (BNZS), 317
genérica, 474
bipolar, 314
de alta densidade. *Veja também* HDB, 317
de resposta parcial, 323
duobinária
com codificação diferencial, 327, 328
detecção, 326
modificada, 326
multiamplitude, 508
multitom, 521
on-off, 312, 479
ortogonal, 352, 480, 719
PAM 4-ária, 342
polar, 309, 477
binária, detector linear ótimo para, 468
por bit roubado, 262
SINCGARS, 575
Sincronização, 261
da portadora local, 159
de bit, 336
de símbolo, 336

Síndrome, 749
treliças de, diagrama em, 777
Sistema(s)
analógicos
de VHS, 4
em banda base, análise de desempenho, 449
BIBO, 85
CDMA DSSS práticos modernos, 594
celulares
por que CDMA em, 595
codificados e não codificados, comparação entre, 762
de acesso múltiplo por divisão no tempo, 265
de comunicação, 1, 2
com múltiplas entradas e múltiplas saídas, 723
digital, 302
análise de desempenho de, 468 566
em banda base, 130
M-ário, 496
modelo de, 450
por portadora, 130
práticos, 716
de dados digitais, 267
de difusão FM, 222
de espalhamento espectral por saltos em frequência, 567, 568
aplicações, 573
assíncrono, 572
com acesso de múltiplos usuários, desempenho, 571
de Hollywood a CDMA, 575
de IEEE 802.11 a Bluetooth, 573
multiusuário e desempenho, 570
principais especificações em 802.11 e de Bluetooth, 574
SINCGARS, 575
de modulação SSB, 151
de multiplexação por divisão em frequência, 160
de posicionamento global (GPS), 598
CDMA no, 598
de pré-ênfase\deênfase, ótimos, 451
digitais com portadora, 344
JTRS, 575
linear, transmissão de sinal em um, 85
múltiplas entradas e múltiplas saídas, 662
passa-tudo *versus* sistemas sem distorção, 87
pseudorruído, 567
RZ (*return-to-zero*), 304
sinais processados por, 19
T1, formato de sinalização do, 261
Slotted frequency hopping, 572
SNR (*signal-to-noise ratio*), 6, 161
de saída, 257
aumento exponencial da, 257
melhora da, 271
Sobrecarga
limiar de, 275
Solicitação automática de repetição, 746
Soma
de processos aleatórios, 445
de variáveis aleatórias, 410
gaussianas, 411
independentes, variância de uma, 401
média da, 398
SOVA (*soft-output Viterbi algorithms*), 783
Spread spectrum, 567
SSB (*single-sideband*), 147
Subportadoras, 160
Supergrupo, 161
Superposição, teorema da, 64
Superquadrado estendido, 262

T

Taxa(s)
de baud, 343
de código, 744
de dados, 1, 9
de erro de bit, 476, 510
de informação, máxima, 242
de Nyquist, 234
Taxação progressiva, princípio, 253
TDM (*time division multiplexing*), 12, 160
TDMA (*time division multiple-access*), 265
Tecnologia(s)
de vídeo digital, 4
DVD, 4
Telecomunicações
eventos dos dois últimos séculos das, 16
modernas, breve revisão histórica de, 14
Telefonia digital, 259
Telegrafia semafórica, conceito de, 15
Televisão de alta definição, 286
Tempo, 19
dilatação no, propriedade, 74
domínio do
diferenciação, 83
integração, 83
translação no, propriedade de, 76
Temporização, extração de, 336
Teorema(s)
da amostragem, 6, 232
algumas aplicações, 247
da codificação de fonte, 687
da convolução, 81
da probabilidade total, 375
da superposição, 64
de Bayes, 376
de codificação de canal ruidoso, 743
de código de blocos linear cíclico, 754
de Parseval, 36, 37, 97
na série de Fourier, 43
de Wiener-Khintchine, 432
do limite central, 412
para a média de amostras, 413
para a soma de variáveis aleatórias independentes, 414
Teoria axiomática da probabilidade, 377
Tesla, Nikola, 15
Timbre, impulsos de, 278
Tom
modulação por, 144
puro, 144
Traço, 820
Transdutor
de entrada, 2
de saída, 3
Transferência, função de, 86

Transformada(s)
de algumas funções úteis, 66
de cosseno discreta, 283
de Fourier
cálculo, 115
numérico, 110
dilatação da, propriedade de, 75
direta, 62
existência, 64
indireta, 62
interpretação, 64
linearidade, 64
operações com, propriedades de, 85
propriedades, 71
rápida (FFT), 114, 114
tabela de, 72
de Hilbert, 148
Transmissão
de sinal em um sistema linear, 85
distorção do sinal durante a, 86
em banda lateral simples, 148
em canais simétricos binários, 691
sem distorção, 87
simultânea de múltiplos sinais, 12
Transmissor, 3
com conhecimento do canal, 727
facilidade de, 11
FM indireto de Armstrong, diagrama de blocos do, 209
sem conhecimento do canal, 725
Travamento
em fase, 166
faixa de, 166
Treliças de síndrome, diagrama em, 777
Trem de pulsos PAM aleatório, 437

U

Unidades logarítmicas, 259
Upconversion, 140

V

Valor(es)
esperado, 396
médios estatístico, 395
quadrático médio, 20
da soma de variáveis descorrelacionadas, 406
do erro
causado por ruído de canal em PCM, 400
de quantização uniforme em PCM, 399
Variável(is)
aleatórias, 378
contínua, 379, 383
discretas, 378
gaussiana, 385
independentes, 393
soma de, 410
uso de *n*, para estimar uma variável aleatória, 408
VCO (*voltage-controlled oscillator*), 161
Velocidade instantânea, 188
Verossimilhança relativa, 781
Vetor(es)
de base, 486
de erro, 28
de peso mínimo, 750
normalizado, 486
ortogonais, 486
Viabilidade de repetidores regenerativos sem distorção, 5
Vídeo, compressão de, 278, 281
Viterbi
algoritmo de, 771
decodificação de, 772
Vocoders, 278
QCELP, 596
LP práticos de alta qualidade, 281
Voz humana, mecanismo de produção da, 279
VSB (*vestigial sideband*), 155
na difusão de televisão, uso de, 158

W

Walsh, funções de, 43
WBFM (*wideband FM*), 195
Weaver, método de, 153
WLAN (*wireless local area network*), 18, 573

Z

ZF (*zero forcing*), 627
configuração, 636