

APONTAMENTOS DA DISCIPLINA DE

Fundamentos das Telecomunicações

DO CURSO DE ENGENHARIA DE SISTEMAS E INFORMÁTICA

Vasco Freitas

**Universidade do Minho
1995**

Índice

Prefácio	1
1 Introdução	3
1.1 Informação, Mensagens e Sinais	4
1.2 Elementos de um Sistema de Comunicação	6
1.3 Sinais de Comunicação	8
1.4 Limitações Fundamentais	11
1.5 Modulação e Codificação	15
1.5.1 Técnicas de Modulação	16
1.5.2 Vantagens e Aplicações da Modulação	19
1.5.3 Técnicas de Codificação	21
1.6 Problemas	23
2 Análise de Sinais	25
2.1 Sinais periódicos: Espectros de Linhas	26
2.1.1 Fasores e Espectros de Linhas	26
2.1.2 Sinais Periódicos e Potência Média	30
2.1.3 Série de Fourier	31
2.1.4 Sequência de pulsos rectangulares	34
2.1.5 Teorema da Potência	36
2.2 Sinais não-periódicos: Espectros Contínuos	38

2.2.1	A Transformada de Fourier	38
2.2.2	Pulso rectangular	40
2.2.3	Teorema da Energia	42
2.2.4	Largura de Banda	42
2.3	Modulação	44
2.3.1	Modulação de Amplitude	44
2.3.2	Modulação de Frequência	46
2.4	Problemas	47
3	Análise de Sistemas	51
3.1	Transmissão e Filtragem de Sinais	51
3.2	Sistemas lineares e invariantes no tempo	51
3.3	Função de transferência	53
3.4	Sistemas de primeira ordem	55
3.5	Largura de banda de transmissão	57
3.6	Análise com diagramas de blocos	58
3.7	Distorção do sinal na transmissão	59
3.7.1	Transmissão isenta de distorção	60
3.7.2	Distorção Linear e Equalização	61
3.8	Perdas de Transmissão e Decibéis	63
3.8.1	Ganho de Potência	63
3.8.2	Perdas de Transmissão e Repetidores	65
3.9	Filtros	67
3.9.1	Filtros ideais	67
3.9.2	Filtros Reais	68
3.10	Resposta temporal de sistemas	70
3.10.1	Resposta ao escalão	71
3.10.2	Ritmo máximo de símbolos digitais	73
3.10.3	Domínio do tempo versus domínio da frequência	74
3.11	Problemas	75

4	Ruído e Erros	79
4.1	Sinais aleatórios	80
4.2	Ruído	83
4.2.1	O ruído térmico	84
4.2.2	Ruído branco e gaussiano	86
4.2.3	Largura de banda equivalente de ruído	87
4.3	Erros	89
4.3.1	Regeneração do sinal digital	89
4.3.2	Probabilidade de erro	91
4.4	Problemas	97
5	Digitalização	99
5.1	Teoria da Amostragem	100
5.1.1	A operação de amostragem	100
5.1.2	Teorema da Amostragem	102
5.1.3	Amostragem na prática	105
5.2	Quantização	107
5.2.1	Quantização uniforme	108
5.2.2	Ruído de quantização	109
5.2.3	Quantização não-uniforme	110
5.3	Conversão analógico a digital	113
5.3.1	Modulação de Impulso Codificada, MIC ou PCM	113
5.3.2	Ruído em PCM	118
5.3.3	Modulação Delta	120
5.4	Problemas	126
6	Multiplexagem	129
6.1	Multiplexagem por divisão do tempo, TDM	130
6.1.1	Sincronização	131
6.1.2	Conceito de canal virtual	133

6.1.3	Técnicas TDM	134
6.2	TDM síncrono	134
6.2.1	Organização das tramas	135
6.2.2	Alinhamento das tramas	136
6.2.3	Sinalização	137
6.2.4	Sistemas de multiplexagem PCM	138
6.2.5	Hierarquia de multiplexagem PDH	143
6.2.6	Hierarquia de multiplexagem SDH e SONET . . .	151
6.3	TDM estatístico	156
6.3.1	Características do tráfego gerado por computadores	156
6.3.2	Multiplexagem estatística	158
6.3.3	Desempenho	160
6.4	Multiplexagem por divisão da frequência	168
6.4.1	Método	168
6.4.2	Hierarquias FDM	170
6.5	Problemas	172
7	Códigos de Linha	175
7.1	Códigos de banda base	175
7.1.1	Não Retorno a Zero (NRZ)	178
7.1.2	Bipolar AMI	180
7.1.3	Bifásico Manchester	181
7.1.4	Códigos Scrambling	183
7.1.5	Códigos de redução de ritmo, mBnL	186
7.2	Códigos de banda de canal	188
7.2.1	Técnicas fundamentais de codificação	188
7.2.2	Codificação <i>M</i> -ária	191
7.2.3	Técnicas híbridas	192
7.3	Problemas	194

8	Teoria da Informação	195
8.1	Medida da Informação	197
8.1.1	Informação própria	198
8.1.2	Entropia e Débito de informação	199
8.2	Codificação da fonte discreta sem memória	202
	Construção do código	206
	Códigos de Huffman	206
8.3	Fontes com memória	207
8.4	Transmissão de informação: o canal	209
8.4.1	Canais discretos	210
8.4.2	Canais contínuos	217
8.5	Problemas	221
9	Códigos para Control de Erros	225
9.1	Tipos de erros	225
9.2	Tipos de códigos	226
9.3	Códigos Lineares de Bloco	227
9.3.1	Distância de Hamming	228
9.3.2	Capacidade de control de erros dos códigos de bloco	229
9.3.3	Códigos de Hamming	231
9.4	Códigos Cíclicos Binários	232
9.4.1	Estrutura algébrica dos códigos cíclicos	232
9.4.2	Geração de códigos cíclicos sistemáticos	234
9.4.3	Codificação com registos de deslocamento	237
9.4.4	O síndrome: detecção e correcção de erros	240
9.4.5	Códigos BCH	242
9.5	Técnicas de Correcção de Erros	244
9.5.1	Correcção Progressiva e ARQ	244
9.5.2	Probabilidade de Erro na Correcção Progressiva .	245
9.6	Problemas	246

Prefácio

A disciplina das telecomunicações constituiu, até muito recentemente, uma área relativamente especializada da engenharia electrotécnica. Porém, dos extraordinários progressos que se têm vindo a verificar quer na tecnologia dos computadores e da programação quer nas tecnologias das telecomunicações, resultou uma sinergia pela qual os sistemas de processamento e de acesso à informação são projectados e realizados através de computadores interligados entre si por sistemas de telecomunicações, não se concebendo hoje em dia que um computador não possua a capacidade de comunicar com outros.

Isto passa-se não só com computadores ou sistemas de informação mas também com uma grande diversidade de outros equipamentos que incorporem alguma forma de *inteligência*, como por exemplo, os terminais de vídeo controlados por microprocessador, as estações de trabalho para a computação pessoal ou actividades de laser, até à instrumentação inteligente integrada em maquinaria industrial, os terminais de ponto de venda (*Point Of Sale*, *POS*), etc. Uma componente muito importante a considerar pois, na concepção dos equipamentos de computação e das aplicações, diz respeito às tecnologias e aos serviços de telecomunicações a utilizar que lhes permitam comunicar com outros dispositivos ou aplicações. Os temas da comunicação de dados, das redes de computadores e dos teleserviços multimédia que são afins àquelas aplicações, tornam-se pois tópicos essenciais em todos os cursos de engenharia informática, engenharia de computadores ou de engenharia de software.

Em muitas aplicações, a simples existência de meios físicos que permitam que dois sistemas troquem informação através de uma rede de computadores, só resolve parte do problema. Em aplicações que envolvam um conjunto de sistemas de computação heterogéneos geograficamente distribuídos que necessitem trocar dados entre si através de uma rede,

problemas tais como a existência de diferentes sistemas operativos e portanto de diferentes sistemas de ficheiros e possivelmente diferentes formatos para a representação dos dados e dos programas, têm de ser devidamente abordados e resolvidos para que esses sistemas possam efectivamente comunicar de uma forma *aberta* e eficaz. É este o objectivo dos vários *protocolos* de comunicações orientados às aplicações, associados ao Modelo de Referência para a Interligação de Sistemas Abertos criado pela ISO (International Standards Organization) ou ao modelo Internet, também e mais vulgarmente conhecido por *pilha* TCP/IP. É, portanto, indispensável, ao abordar as aplicações de redes de computadores, compreender o que são e como funcionam os *interfaces* e os *protocolos* desde o nível físico até aos níveis mais altos da comunicação entre aplicações.

Em suma, é importante que um engenheiro informático conheça não só as diferentes técnicas de telecomunicações utilizadas na actualidade como também os respectivos fundamentos tanto teóricos como práticos.

Este curso semestral destina-se, essencialmente, a abordar pela primeira vez esses fundamentos. Tratará, de uma forma introdutória, o nível mais *baixo* das telecomunicações, ou seja, a teoria os sinais de informação e a sua codificação, a teoria dos sistemas de telecomunicações, os efeitos indesejáveis do ruído na transmissão dos sinais, a teoria da digitalização das mensagens e a teoria da informação.

Podem encontrar-se apresentações destes fundamentos, com maior ou menor desenvolvimento ou com ênfase diferente, em qualquer livro sobre sistemas de telecomunicações, comunicação de dados, comunicações por computador, redes de computadores, etc. Os seguintes títulos constituem uma selecção de referências bibliográficas:

- A. Bruce Carlson, *Communication Systems: An introduction to signals and noise in electrical communications*, McGraw-Hill, 3^a edição, 1986
- G. Langley, *Telecommunications Primer*, Pitman, 3^a edição, 1990
- W. Stallings, *Data and Computer Communications*, MacMillan, 4^a edição, 1994

Capítulo 1

Introdução

Com a transmissão da primeira mensagem de telégrafo através de uma linha de 16 Km, efectuada por Samuel Morse em 1838, teve início a era das comunicações eléctricas. Dessa altura até ao presente, a engenharia das comunicações avançou até ao ponto de se poder assistir na televisão à emissão directa de naves espaciais a partir de Marte. O telefone, a rádio e a televisão tornaram-se parte integrante da vida moderna. Texto, dados, voz e imagem percorrem o globo terrestre em fracções de segundo através de circuitos de longa distância e de satélites. Os computadores trocam informação com outros computadores através das *redes de dados* com cobertura intercontinental. Ao longo das próximas décadas assistir-se-á, sem dúvida, a uma evolução ainda mais extraordinária da tecnologia e dos serviços de telecomunicações devido ao crescente envolvimento das tecnologias da informática, sinergia que está na origem do que se vem designando por *teleinformática* ou *telemática*.

É pois muito importante não só que os técnicos de telecomunicações tenham conhecimentos profundos sobre computadores como também que os engenheiros de informática possuam noções sobre os fundamentos das comunicações eléctricas à distância, as suas possibilidades e limitações e o estado desta tecnologia.

Um sistema de comunicações tem por objectivo o transporte de informação desde a sua fonte até um destino situado à distância. Os domínios de aplicação dos sistemas de comunicação são tão variados que não nos é praticável referir a todos eles num curso introdutório. Torna-se também impossível, neste contexto, apresentar e discutir em detalhe todas as

partes que constituem estes sistemas. Nem é necessário. Um sistema de comunicação típico envolve inúmeros sub-sistemas, ou componentes, cuja análise e projecto se enquadram, tradicionalmente, no âmbito dos vários ramos da engenharia electrotécnica: os circuitos e a electrónica, o electromagnetismo, o *hardware* dos computadores, etc. O estudo detalhado destes componentes obscureceria a questão essencial segundo a qual um sistema de comunicação é um todo integrado cuja funcionalidade excede a soma das partes. Trataremos pois o tema das comunicações dum ponto de vista geral e integrado embora procurando não perder o rigor.

Reconhecendo que todos os sistemas de comunicação têm por objectivo fundamental a *transferência de informação* procuraremos isolar os princípios e os problemas fundamentais que se apresentam ao transporte de informação.

Na primeira parte deste curso abordaremos o *nível físico*, isto é, a representação da informação e a sua transmissão sob a forma de sinais eléctricos. Examiná-lo-emos com a profundidade suficiente que permita compreender o que se passa na realidade física e as limitações fundamentais que se enfrentam. Na segunda parte do curso, abordaremos o *nível lógico* das comunicações, isto é, partindo do princípio que o sub-sistema de transmissão físico existe e é fiável, trataremos da comunicação *fim-a-fim*, dos problemas que se colocam a este nível e das técnicas para os resolver. Esta será uma abordagem mais algorítmica, orientada à realização de serviços de aplicações de comunicações por via de *software* e de sistemas de computação.

1.1 Informação, Mensagens e Sinais

O conceito de *informação* é crucial em comunicações. Embora possa ser quantificado com precisão, o termo *informação* tem conotações semânticas e filosóficas de difícil definição. Estas dificuldades podem ser evitadas se em lugar de informação lidarmos com *mensagens*. A mensagem é definida como a manifestação física da informação tal como é produzida pela fonte. Qualquer que seja a forma em que a mensagem é apresentada, o objectivo de um sistema de comunicação é o de reproduzir, no destino, uma cópia aceitável dessa mensagem.

Existem muitos tipos de fontes de informação, incluindo pessoas e máquinas, e as mensagens aparecem sob a forma mais variada. Podem, no en-

tanto, identificar-se duas categorias distintas de mensagens: as *analógicas* e as *digitais*. A cada uma delas, aplica-se um critério distinto de avaliação do sucesso ou desempenho (*performance*) da comunicação.

Uma mensagem *analógica* é representada por uma grandeza física que varia no tempo de uma maneira *suave* e *contínua*. A pressão acústica produzida pela fala, a posição angular de um instrumento de medida ou a intensidade da luz num ponto de uma imagem de televisão são exemplos de mensagens analógicas. Dado que a informação reside numa forma de onda que varia no tempo, um sistema de comunicação analógico deve entregar esta forma de onda com um determinado grau de *fidelidade*.

Uma mensagem *digital* é representada por uma sequência ordenada de símbolos escolhidos de entre um conjunto, normalmente finito. As letras do alfabeto, uma listagem das leituras horárias de temperatura ou as teclas premidas a um terminal de computador constituem exemplos de símbolos das mensagens digitais. Como a informação está contida nos símbolos, e na sua sequência, um sistema de comunicação digital deverá entregar esses símbolos com um determinado grau de *certeza* num determinado intervalo de tempo.

Quer sejam analógicas quer sejam digitais, poucas são as fontes de mensagens inerentemente eléctricas. Consequentemente, a maior parte dos sistemas de comunicação possuem *transdutores* de entrada e de saída, como mostra a figura 1.1. O transdutor de entrada converte a mensagem

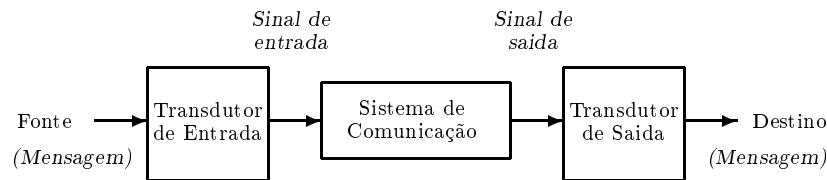


Figura 1.1: Sistema de comunicação incluindo transdutores

num *sinal* eléctrico, por exemplo uma tensão ou uma corrente, e o transdutor de saída converte o sinal à saída do sistema, na forma mais apropriada de mensagem a ser entregue no destino. Num sistema de comunicação telefónico os transdutores serão o microfone à entrada e o altifalante, ou auscultador, à saída. Consideraremos de aqui em diante que existem transdutores apropriados tanto à entrada como à saída e concentraremos a atenção na função de *transmissão do sinal*. Nestas condições, os ter-

mos *senal* e *mensagem* serão utilizados com o mesmo significado dado que o sinal, tal como a mensagem, constituem a representação física da informação.

1.2 Elementos de um Sistema de Comunicação

A figura 1.2 representa os elementos de um sistema de comunicação onde se omitiram os transdutores e se incluíram as contaminações indesejáveis, genericamente designadas por *ruído*. Existem três partes essenciais num

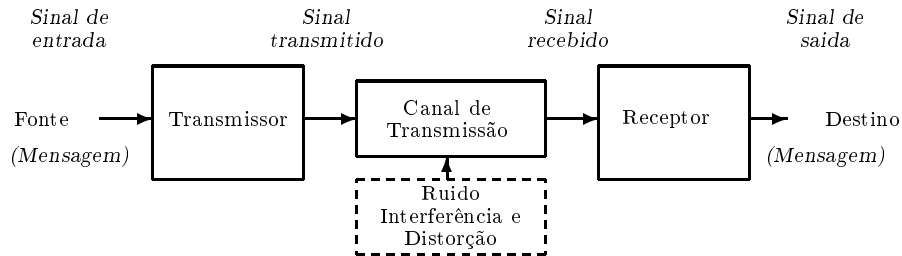


Figura 1.2: Elementos de um sistema de comunicação

sistema de comunicação: o transmissor, o canal de transmissão e o receptor. Cada uma destas três partes desempenha um papel específico na transmissão do sinal:

O *transmissor* processa o sinal de entrada de modo a produzir um sinal bem adaptado às características do canal de transmissão. O processamento do sinal para a transmissão quase sempre envolve a *modulação* e a *codificação de linha*.

O *canal de transmissão* é o meio físico que o sinal percorre desde a fonte até ao destino. Pode ser constituído por um par de fios condutores, cabo coaxial, o espaço livre, fibras ópticas, etc, todos eles servindo de suporte à transmissão das ondas electromagnéticas. Todos estes meios físicos introduzem no sistema *perdas* de transmissão ou *atenuação* do sinal, e assim, a potência do sinal decresce progressivamente com a distância.

O *receptor* opera sobre o sinal de saída do canal preparando-o para a entrega ao transdutor de destino. As operações no receptor consistem na *amplificação*, para compensar as perdas, na *desmodulação* e na *descodificação*, efectuando o processamento inverso do do transmissor. A

filtragem é uma outra função importante do receptor, como se verá mais adiante. Durante a transmissão do sinal aparecem vários efeitos indesejáveis. A atenuação é indesejável dado que reduz a *força* do sinal no receptor. Mais sérios, contudo, são a distorção, a interferência e o ruído que aparecem como alterações à forma do sinal. Embora tais contaminações possam ocorrer em qualquer ponto do sistema, a convenção que se adopta é a de as atribuir inteiramente ao canal, considerando-se portanto como ideais, tanto o transmissor como o receptor. A figura 1.2 reflecte esta convenção.

A *distorção* é a perturbação da forma de onda causada por uma *resposta* imperfeita do sistema ao próprio sinal. A distorção desaparece quando o sinal desaparece o que não acontece com a interferência e com o ruído.

A *interferência* é a contaminação por sinais estranhos ou externos ao sistema, produzidos por fontes de origem humana (não-naturais), tais como outras transmissões, as linhas de transporte de energia eléctrica, as máquinas eléctricas, os circuitos de comutação, etc. A interferência ocorre mais frequentemente em sistemas com canais rádio cujas antenas de recepção interceptam vários sinais simultaneamente. A interferência de rádio-frequência (RF) também aparece em cabos se os fios, ou o receptor, captam sinais radiados por fontes próximas. A interferência pode ser eliminada através de filtragem apropriada na medida em que os sinais interferentes ocupem uma banda de frequências diferente da do sinal desejado.

O *ruído* consiste em sinais aleatórios e imprevisíveis produzidos por processos naturais tanto internos como externos ao sistema de comunicação. Quando tais variações aleatórias se sobrepõem a um sinal portador de informação, a mensagem pode corromper-se parcialmente ou mesmo perder-se completamente. A filtragem reduz a contaminação pelo ruído mas este não pode ser totalmente eliminado. Este ruído residual é inevitável constitui uma das limitações fundamentais dos sistemas de comunicação eléctricos.

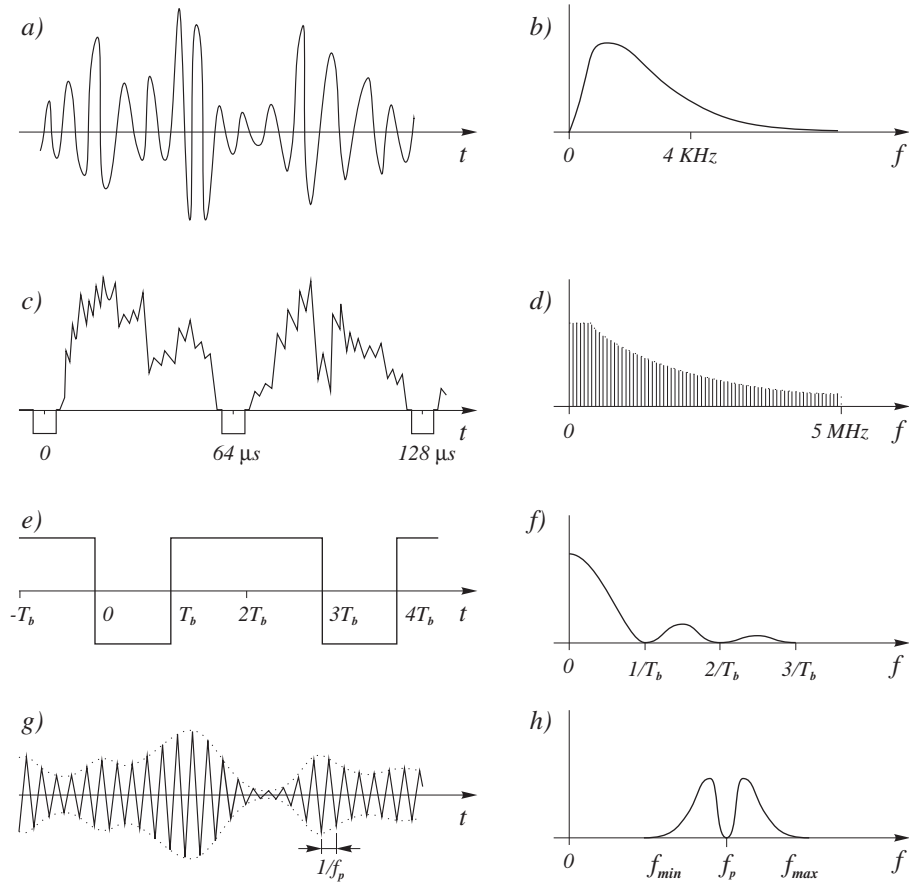
No sistema da figura 1.2 a transmissão efectua-se num só sentido, o que se designa por transmissão **simplex** (SX). A transmissão nos dois sentidos requer, obviamente, um transmissor e um receptor de cada lado. Designa-se por sistema **full-duplex** (FDX), ou simplesmente **duplex** aquele que permite a transmissão simultânea em ambos os sentidos. Se a transmissão fôr possível em ambos os sentidos mas não simultaneamente, o sistema designa-se por **half-duplex** (HDX), ou **meio-duplex**.

1.3 Sinais de Comunicação

A forma de onda de tensão produzida por um microfone em resultado da pressão acústica da fala constitui um exemplo de um sinal eléctrico. A forma que a onda de voz terá, depende da mensagem, isto é, do que foi dito, das características da voz da pessoa que falou e das características do microfone. A figura 1.3(a) mostra uma forma típica deste sinal. A tensão eléctrica varia continuamente no tempo numa maneira imprevisível sendo a função do sistema de comunicação produzir uma réplica tão fiel quanto possível deste sinal no receptor. Por exemplo, o sinal pode ser transportado até ao receptor através de um par de fios mas, se as distâncias forem grandes, o sinal será rapidamente atenuado e a acumulação de interferência e distorção pode torná-lo completamente irreconhecível no destino.

Todos os meios de transmissão possuem características próprias afectando os sinais de maneira diferente. O espaço livre, por exemplo, atenua em poucos metros os sinais de baixa frequência dissipando-os completamente enquanto que os de frequência mais elevada percorrem com muito menor atenuação vários quilómetros. Torna-se portanto necessário *modular* o sinal de voz para a banda de frequência mais adaptada ao canal (ou meio) de transmissão a utilizar. Neste caso, em que o meio é o espaço livre, ter-se-á de adoptar um sinal portador de rádio-frequência para transportar o sinal de voz. É este tipo de modificações ao sinal da fonte, que o tornam compatível com o canal disponível, que o transmissor tem de efectuar.

Na prática não é necessário, nem possível, reproduzir exactamente o sinal no receptor. Basta apenas uma aproximação. Este ponto é muito importante e influencia significativamente o projecto de sistemas de comunicação. O sinal de voz, por exemplo, pode ser passado através de um filtro que lhe retire certas componentes de frequência e permanecer ainda perfeitamente inteligível. Pode fazer-se uma ideia de quais as componentes do sinal que mais importância têm na sua preservação, examinando a distribuição da potência do sinal às várias frequências como se mostra na figura 1.3(b). Esta forma de representação do sinal é designada de *espectro de potência* ou *função densidade espectral de potência* e será estudada no próximo capítulo. Basta por agora referir que um sistema de comunicação deve possibilitar uma boa transmissão na banda de frequências na qual o espectro de potência do sinal é mais significativo. Qual o nível de potência que podemos considerar significativo, é uma



- (a) Forma de onda de áudio (b) Espectro do sinal de áudio
(c) Forma de onda de vídeo (d) Espectro do sinal de vídeo
(e) Forma de onda binária (f) Espectro do sinal binário
(g) Sinal de banda passante (h) Espectro de banda passante.

Figura 1.3: Formas de onda de sinais típicos e respectivos espectros

questão em aberto. Dizer que um sinal de voz foi ou não significativamente alterado é uma questão subjectiva mas, na realidade, a qualidade de um canal de comunicação pode necessitar ser avaliado na base de testes subjectivos.

Através de tais testes subjectivos chegou-se à conclusão de que, se um canal de comunicação "deixa passar", relativamente inalteradas, as componentes do sinal com frequências no intervalo crítico de 300 Hz a 4.3 KHz, então o sinal de voz é inteiramente inteligível. Por este motivo os sistemas telefónicos possuem filtros destinados a limitar o sinal de voz a esta banda de frequências. Dado que os filtros, na prática, são imperfeitos, os sinais telefónicos de voz são tratados como sendo sinais de 4 KHz na banda de base, isto é, contendo frequências desde 0 Hz até 4000 Hz. À amplitude do intervalo de frequências em que um sinal tem componentes significativas chama-se *largura de banda* do sinal.

O sinal de voz, embora importante, é só um de entre a grande variedade de sinais transportadores de informação. Os sinais de música de alta fidelidade contêm componentes de frequência significativos ao longo de todo o intervalo audível que se estende, aproximadamente, desde os 20 Hz aos 20 KHz. Pode obter-se uma boa reprodução tratando os sinais de música como sendo de banda de base e com uma largura de banda de 15 KHz.

A televisão constitui outro exemplo. O requisito de largura de banda de um sinal de vídeo é de 5 MHz (5×10^6 Hz) em banda de base, para sinais da qualidade da radiodifusão. Tem algum sentido, portanto, o provérbio que diz, *uma imagem vale mais do que mil palavras*. Faremos a prova desta asserção quando tratarmos da teoria da informação.

A figura 1.3(c) ilustra uma forma de onda de um sinal de televisão, correspondente ao varrimento de uma linha de écran. A amplitude do sinal corresponde à intensidade luminosa encontrada nessa linha. A figura 1.3(e) mostra um sinal binário típico em transmissão de dados ou em transmissão digitalizada. A tensão varia no tempo entre dois valores discretos. O intervalo mais pequeno entre transições, T_s , é o intervalo de sinalização, ou tempo de símbolo (neste caso em que o símbolo é binário, o tempo de símbolo designa-se por T_b), sendo $r_s = \frac{1}{T_s}$ o *ritmo de símbolos* em *símbolos por segundo*, unidade que se designa também por **baud** (neste caso, igualmente, são dígitos binários por segundo, $r_b = \frac{1}{T_b}$).

Pode verificar-se através do espectro de potência que a maior parte da

potência do sinal está concentrada nas frequências abaixo de $\frac{1}{T_b}$ Hz. Deve esperar-se, portanto que a largura de banda de transmissão exigida a um sistema de transmissão de dados binário seja proporcional ou talvez da mesma ordem de grandeza do ritmo binário. Basta notar, de momento, que se fôr necessário enviar uma grande quantidade de dados num curto intervalo de tempo é necessário que o sinal possua um ritmo elevado e portanto, também, um sistema de elevada largura de banda. É normal terem-se ritmos de 140 Mbit/s nas redes de telecomunicações existindo sistemas experimentais que operam a ritmos superiores a 1 Gbit/s (Gigabit/s = 10^9 bit/s).

Os sinais a que nos vimos referindo são todos do tipo *banda de base*, isto é, as suas componentes de frequência estão concentradas numa banda de frequências que se estende desde 0 Hz, o sinal contínuo, até uma frequência máxima f_{max} à qual corresponde, aproximadamente o mesmo valor da largura de banda, B ($B \approx f_{max}$). Nem todos os sinais são deste tipo. Alguns são, como se costuma designar, de *banda passante*, concentrados numa banda de frequências que se estende desde uma frequência mais baixa, f_{min} , até uma mais elevada, f_{max} , possuindo uma largura de banda $B \approx f_{max} - f_{min}$. A figura 1.3(g) representa a forma de onda de um sinal de banda passante. Os sinais deste tipo são obtidos dos sinais de banda de base através da operação de *modulação*. Como já se referiu, o objectivo desta operação é o de permitir que a mensagem possa ser transmitida através de um canal de banda passante.

1.4 Limitações Fundamentais

A engenharia das telecomunicações confronta-se com dois tipos gerais de limitações. Por um lado existem os *problemas tecnológicos*, tão diversos como a disponibilidade de *hardware*, os factores económicos, a regulamentação oficial, etc. São problemas de viabilidade que, em teoria, podem ser resolvidos, embora por vezes as soluções perfeitas possam não ser práticas. Por outro lado, existem as *limitações físicas fundamentais*, isto é, as que são ditadas pelas leis da natureza. São estas limitações que em última instância determinam o que pode ou não pode ser realizado, independentemente dos problemas tecnológicos.

As limitações fundamentais à transmissão da informação por meios eléctricos são a *largura de banda* e o *ruído*.

O conceito de largura de banda aplica-se tanto aos sinais como aos sistemas de transmissão que os transportam, constituindo uma medida de *velocidade* com se pode transferir informação desde a fonte até ao destino. Quando um sinal possui variações rápidas no tempo o seu *espectro* contém frequências elevadas e o sinal possuirá eventualmente uma grande largura de banda. De igual modo, a facilidade com que um sistema de comunicação consegue reproduzir, ou "acompanhar", as variações do sinal de entrada é uma consequência da sua *característica de frequência* traduzida por uma *largura de banda de transmissão*, B_T . Ora, todos os sistemas eléctricos são constituídos por componentes que armazenam e transferem energia não o podendo fazer instantâneamente. Consequentemente, todo o sistema de transmissão possui uma largura de banda finita (B_T) a qual limita a sua *resposta* à variação dos sinais.

As comunicações em condições de tempo real necessitam suficiente largura de banda de transmissão para comportar o espectro do sinal, caso contrário, o sinal sofrerá distorção. Como se referiu anteriormente, um sinal de vídeo de TV exige uma largura de banda de vários Megahertz ($B_T \approx 5$ MHz) enquanto que as variações muito mais lentas de um sinal de voz se encaixam numa banda de cerca de $B_T \approx 4$ KHz. Um sinal digital (fig. 1.3(e)) com um ritmo de símbolos $r_s = \frac{1}{T_s}$ simb/s exige do sistema de transmissão uma largura de banda proporcional a r_s . Justificar-se-á mais adiante que esta largura de banda não poderá ser inferior a $\frac{r_s}{2}$, isto é,

$$B_T \geq \frac{r_s}{2} \text{ Hz} \quad (1.1)$$

Convém acrescentar que os sinais digitais não são necessariamente binários. A mensagem duma fonte discreta pode ser representada por um sinal com mais do que dois níveis lógicos. A figura 1.4 mostra um sinal quaternário no qual a tensão varia discretamente entre quatro níveis distintos. Duma maneira geral, cada nível representa um *símbolo*. Tal como

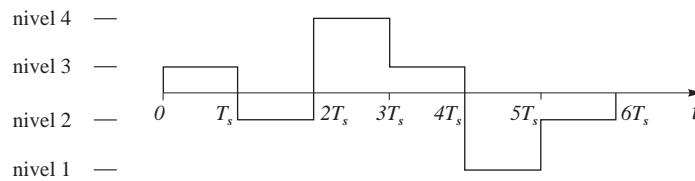


Figura 1.4: Forma de onda de um sinal quaternário

se disse para o sinal binário, o intervalo de tempo entre dois símbolos consecutivos representa o tempo do símbolo, T_s , e o seu inverso, $\frac{1}{T_s}$, é o ritmo de símbolos que é medido em **baud** (símbolos por segundo). Enquanto que no sinal binário o símbolo pode ser representado por um dígito binário, neste sinal o símbolo é representado por um dígito quaternário. Torna-se evidente que o sinal quaternário é capaz de transportar duas vezes mais informação do que o sinal binário com o mesmo ritmo de símbolos. Efectivamente, se o sinal binário contiver 1 *bit* de informação por cada símbolo, o sinal quaternário conterá 2 *bits* de informação por símbolo. Este raciocínio leva-nos a introduzir o *bit* como unidade de medida relativa de informação. A *Teoria da Informação*, que trataremos em capítulo mais adiante, utiliza o termo *bit* para designar a unidade de medida (absoluta) de informação, distintamente da sua utilização como abreviatura de dígito binário. Por esta razão e para distinguir as duas situações em que o termo pode ser utilizado, se adopta o termo *símbolo* para designar um elemento do sinal discreto mesmo quando esse elemento pertence a um conjunto binário.

No caso da transmissão sem limitações de tempo real, a largura de banda disponível determina a velocidade máxima a que o sinal pode ser transmitido. O tempo necessário para transmitir uma dada quantidade de informação é, consequentemente, proporcional a B_T .

O ruído impõe uma segunda limitação à transmissão de informação. A justificação para a inevitabilidade do ruído é-nos dada pela teoria cinética das partículas. A temperaturas superiores a zero absoluto, a energia térmica faz com que as partículas materiais microscópicas fiquem possuídas de movimento aleatório. O movimento aleatório de partículas carregadas eléctricamente, tais como os electrões, dá origem a correntes e tensões aleatórias designadas por *ruído térmico*. Existem outros tipos de ruído, mas o térmico aparece em todos os sistemas de comunicação.

Caso não houvesse ruído, a capacidade de transmissão de um sistema de comunicação seria tão grande quanto se desejasse – teóricamente infinita. Uma maneira de realizar esta capacidade ilimitada seria a de aumentar indefinidamente o número de níveis distintos do sinal discreto da figura 1.4. Mas a presença do ruído introduz tensões aleatórias que se adicionam ao sinal. Se o intervalo entre níveis for mais pequeno do que a amplitude de variação do ruído ao longo de todo o percurso da transmissão, a sobreposição dos dois sinais dará origem a erros como a figura 1.5 exemplifica. Neste exemplo, o facto da tensão de ruído entre os instantes

$2T_s$ e $3T_s$ fazer com que o sinal permaneça num nível errado durante um certo tempo pode, ou não, ser considerado como erro. Tudo depende da decisão que o receptor tomar. Olhando para a representação gráfica

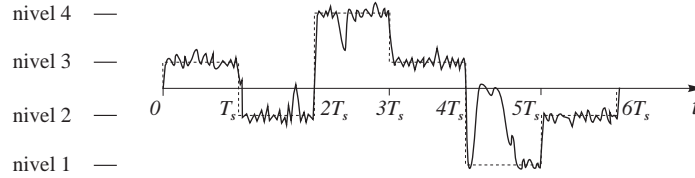


Figura 1.5: Sinal quaternário afectado de ruído

do sinal resultante (a traço contínuo), mesmo desconhecendo a forma de onda transmitida (a tracejado) é mais provável que o nível certo seja o 4 e não o 3 pois o sinal permanece a maior parte do tempo do símbolo junto ao nível 4. Mas quanto ao que se passa entre os instantes $4T_s$ e $5T_s$ já a decisão não será tão óbvia. Teríamos certamente dificuldade em decidir entre os níveis 1 e 2. Se o receptor adoptar uma regra de decisão tal que o valor tomado como correcto é o valor do nível no instante médio do intervalo do símbolo (obtido por amostragem), então o sinal recebido terá dois símbolos errados. Se, por outro lado, a decisão do receptor for tomada na base do valor médio da amplitude do sinal durante o intervalo do símbolo, então, entre $2T_s$ e $3T_s$ o receptor detectará (decidirá a favor de) o nível 4 e não o nível 3 e, entre $4T_s$ e $5T_s$ detectará o nível 2. O sinal recebido teria um símbolo errado. É claro que o receptor não tem possibilidade de saber isto. Simplesmente, a segunda regra de decisão parece ser melhor do que a primeira. O valor médio é dado pelo integral do sinal no intervalo do símbolo dividido pela amplitude deste intervalo. Em lugar do integral do sinal costuma considerar-se o integral do quadrado do sinal, do que resulta uma quantidade com significado físico muito mais expressivo: a sua potência média, S .

De igual forma, o ruído é considerado um sinal aleatório, $n(t)$, possuindo uma determinada densidade de potência, η Watt/Hz praticamente constante em todo o domínio das frequências (espectro electromagnético), a que corresponde uma potência média total na banda do sinal de $N = \eta B$ Watts.

A influência do ruído num sinal de informação mede-se em termos da *razão de potências sinal-ruído*, $\frac{S}{N}$. A potência do ruído térmico é normal-

mente pequena pelo que a razão sinal-ruído, $\frac{S}{N}$, pode ser suficientemente elevada que o ruído passe despercebido. Para pequenos valores de $\frac{S}{N}$, contudo, o ruído degrada a fidelidade das comunicações analógicas e dá origem a erros nas comunicações digitais. Estes problemas tornam-se críticos nas comunicações a longas distâncias quando as perdas de transmissão reduzem a potência do sinal recebido para níveis da ordem de grandeza da potência do ruído. Qualquer amplificação efectuada no receptor não resolve o problema porque o ruído será amplificado juntamente com o sinal.

Tomando ambas estas limitações em consideração, ou seja, a largura de banda do sistema de comunicação e o ruído nele presente, o matemático Claude Shannon demonstrou, em 1948, que cada sistema de comunicação é caracterizado por uma quantidade C , a *capacidade do canal*¹, dada por

$$C = B_T \log \left(1 + \frac{S}{N} \right) \quad (1.2)$$

e que o ritmo de informação não pode exceder esta *capacidade*². A relação 1.2, conhecida por Lei de Hartley-Shannon, estabelece um limite superior ao desempenho (*performance*) de um sistema de comunicação com uma dada largura de banda e um dado valor da razão sinal-ruído. Ela será justificada quando se abordar a Teoria da Informação.

1.5 Modulação e Codificação

O transmissor efectua duas operações extremamente importantes destinadas a permitir uma transmissão de informação rentável e fiável: a *modulação* e a *codificação*. Discuti-las-emos em seguida com um pouco mais de pormenor antes de as tratarmos mais rigorosamente em capítulos posteriores.

¹o termo *canal* é utilizado, no contexto da Teoria da Informação, no seu sentido mais lato significando uma *via* de comunicação, isto é, um sistema de comunicação simplex completo e não sómente o canal físico ou meio de transmissão

²a base a adoptar para a determinação do logaritmo depende da unidade que se escolher para a medida da informação

1.5.1 Técnicas de Modulação

A modulação envolve duas formas de onda: o sinal *modulante* que representa a mensagem e a onda *portadora*. Um modulador tem por função alterar sistematicamente a onda portadora de acordo com as variações da amplitude do sinal modulante. É desta forma que a onda modulada resultante "transporta" a informação contida na mensagem. Em geral, pretende-se que a modulação seja uma operação inversível de forma a permitir a recuperação da mensagem sob a forma do sinal modulante original. A operação inversa é a *desmodulação*.

O sinal da figura 1.3(g) é a onda modulada resultante da modulação de uma onda portadora sinusoidal por um sinal modulante analógico. A figura 1.6 ilustra este processo. O sinal (c) foi obtido adicionando uma constante ao sinal modulante (a) seguido de uma multiplicação pela onda portadora (b), isto é, a onda modulada foi obtida variando a amplitude da portadora de acordo com a amplitude do sinal modulante. Trata-se da familiar modulação de amplitude (AM) utilizada em radiodifusão. A

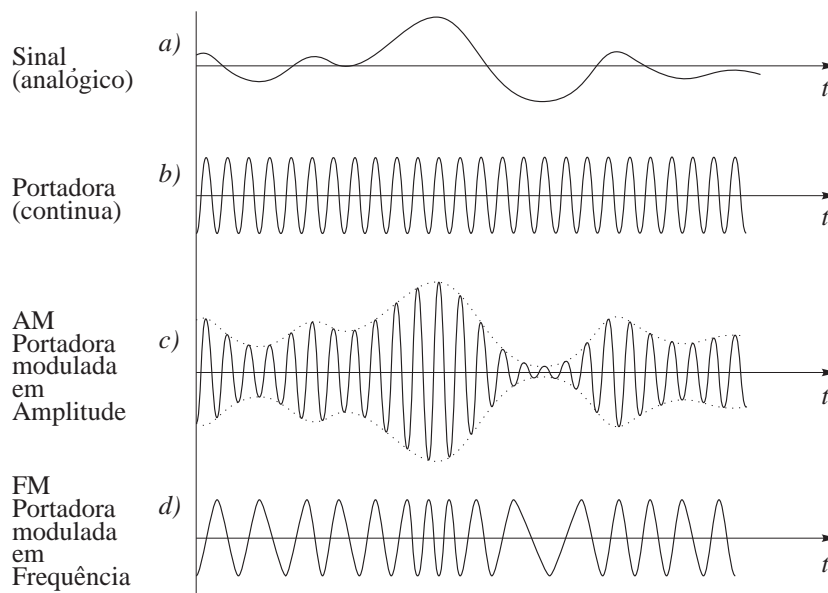


Figura 1.6: Formas de onda em modulação analógica de onda contínua

mensagem pode igualmente ser impressa sobre a portadora por modula-

ção de frequência (FM) ou por modulação de fase (PM) consoante é a frequência ou a fase da portadora que se faz variar com a amplitude do sinal modulante. O sinal (d) na figura 1.6 representa a onda (a) modulada em frequência.

Estes três processos de modulação de portadora sinusoidal são designados de *modulação de onda contínua*. O processo da fala constitui uma modulação de onda contínua. As cordas vocais geram ondas portadoras de várias frequências que ao passar pela cavidade bucal são moduladas em amplitude através das acções musculares que fazem variar o volume e a geometria desta cavidade. O que se ouve não é mais do que um sinal acústico modulado em amplitude, semelhante a um sinal AM.

Estes casos constituem o que se designa por modulação analógica de onda contínua porque o sinal modulante é analógico. Quando o sinal modulante é digital a modulação designa-se por modulação digital de onda contínua. A figura 1.7 mostra um sinal binário e as suas formas moduladas em amplitude (ASK, Amplitude Shift Keying também chamda OOK, On-Off Keying), em frequência (FSK, Frequency Shift Keying) e em fase (PSK, Phase Shift Keying). A maior parte dos sistemas de transmissão a

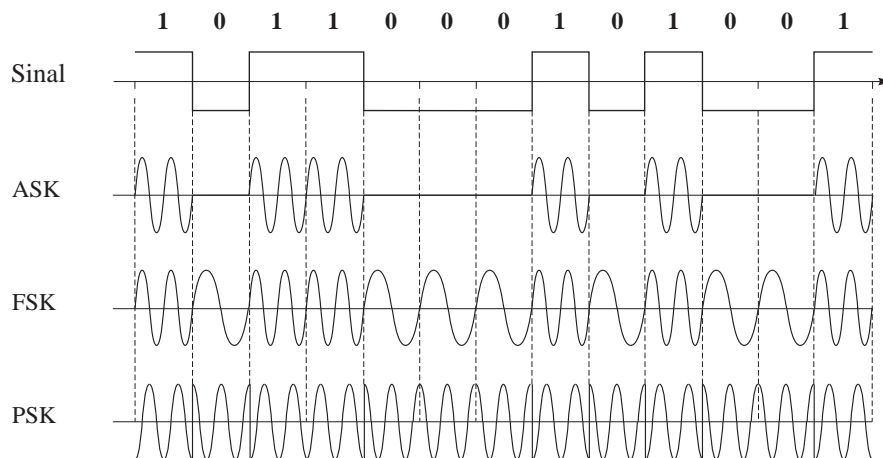


Figura 1.7: Formas de onda em modulação digital de onda contínua

longa distância empregam modulação de onda contínua com frequências da portadora muito maiores do que a maior componente de frequência do sinal modulante. O espectro do sinal modulado consiste portanto numa

banda de frequências concentradas à volta da frequência da portadora conforme mostra a figura 1.3(h). Diz-se, por isso, que a modulação de onda contínua efectua uma *translação de frequência* do espectro do sinal modulante. Em radiodifusão AM, por exemplo, o espectro da mensagem estende-se tipicamente desde 100 Hz até 5 KHz. Se a frequência da portadora fôr de 600 KHz, o espectro da portadora modulada situa-se entre 595 KHz e 605 KHz.

Um outro método de modulação é o chamado *modulação de impulso* em que a portadora é uma sequência periódica de pulsos³ de curta duração. A figura 1.8 mostra a forma de onda do sinal modulante (a), e a forma de onda resultante (c) da modulação em amplitude de uma sequência periódica (b) de pulsos de amplitude constante, duração τ e período T_a . A sequência de pulsos de amplitude variável assim obtida designa-se por sinal PAM (*Pulse Amplitude Modulation*).

Interessa realçar que o sinal PAM consiste numa sequência de amostras extraídas do sinal analógico. Por este motivo, a modulação de impulso em amplitude toma o nome de *amostragem*, ou *sampling* na designação inglesa. A *amostragem* constitui uma técnica importante de processa-

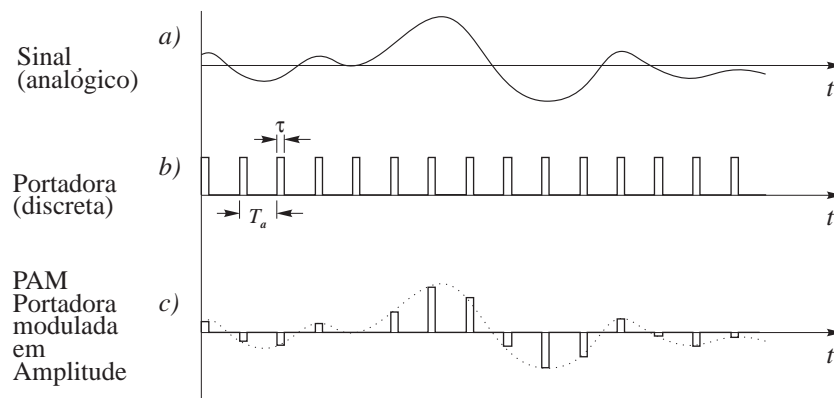


Figura 1.8: Modulação de impulso em amplitude

mento do sinal. Se a amostragem fôr efectuada em certas condições, é

³embora se designe por modulação de *impulso* o sinal que é modulado é uma sequência de *pulsos*. O pulso é uma aproximação física do impulso, isto é, tem uma duração não-nula e uma amplitude positiva limitada. Reservaremos o termo *impulso* para designar o pulso de duração nula.

possível *reconstruir* completamente a forma de onda contínua original a partir das amostras periódicas. A digitalização das fontes contínuas é feita por este processo. Mas a modulação de impulso (amostragem), só por si, não produz a translação de frequência necessária à transmissão do sinal. É pois necessário combinar a modulação de impulso com uma subsequente modulação de onda contínua.

A maior parte dos sistemas de transmissão digitais efectuem ainda uma operação intermédia àquelas duas – a *codificação*.

A *codificação* consiste na transformação de cada uma das amostras do sinal, isto é, cada pulso PAM, num número binário. Esta transformação é conseguida através duma operação de *quantização* seguida de *conversão analógico-digital*. O número binário assim obtido é representado por um sinal (eléctrico) binário o qual vai modular uma onda contínua para efeitos de transmissão.

1.5.2 Vantagens e Aplicações da Modulação

O objectivo primeiro da modulação é o de gerar um sinal adaptado às características do canal físico de transmissão. A modulação apresenta várias vantagens práticas algumas das quais se discutem a seguir.

Modulação para rentabilização da transmissão

A transmissão de sinais a grandes distâncias envolve ondas electromagnéticas progressivas, guiadas ou não, num meio de transmissão. O rendimento de um determinado meio físico de transmissão depende da frequência do sinal que o percorre. Manipulando convenientemente a propriedade de translação de frequência da modulação de onda contínua, o sinal representativo da mensagem pode ser sobreposto a uma portadora cuja frequência tenha sido seleccionada para o meio desejado.

Como exemplo refram os a transmissão de um sinal desde uma antena de emissão até uma antena de recepção em linha de vista uma da outra. Comprova-se que para que haja radiação na emissão e portanto propagação através do espaço livre, a dimensão física da antena tem de ser pelo menos $\frac{1}{10}$ do comprimento de onda do sinal a transmitir. O comprimento de onda, λ , está relacionado com a frequência, f , pela expressão

$$f \cdot \lambda = \nu \quad (1.3)$$

em que ν é a velocidade de propagação das ondas electromagnéticas no meio físico considerado. No caso do espaço livre esta velocidade é praticamente igual à velocidade da luz no vácuo $\nu \approx c \approx 3 \times 10^8$ m/s (e nos meios condutores $\nu \approx 2 \times 10^8$ m/s). Se tentássemos transmitir um sinal de áudio sem modulação, o qual contém componentes de frequência a partir dos 100 Hz necessitaríamos de uma antena com cerca de 300 Km de comprimento, o que é, naturalmente, impraticável. Modulando o sinal numa portadora de 100 MHz (100×10^6 Hz), tal como em radiodifusão de FM, a dimensão suficiente da antena é da ordem de 1 metro.

A figura 1.9 mostra as zonas do espectro electromagnético apropriadas à transmissão de sinais. A figura inclui as frequências e os comprimentos de onda no espaço livre, as designações das bandas normalizadas pelo CCIR⁴ e os modos de propagação.

Modulação para a multiplexagem

A multiplexagem é uma operação que combina vários sinais a fim de serem transmitidos simultaneamente através do mesmo canal. A multiplexagem por divisão de frequência (FDM, Frequency Division Multiplexing) utiliza a modulação de onda contínua para colocar cada sinal sobre uma portadora de frequência diferente. No destino, os sinais são separados por filtragem. A multiplexagem por divisão de tempo (TDM, Time Division Multiplexing) utiliza a modulação de impulso para colocar amostras de sinais diferentes intercaladas no tempo. Consideremos dois sinais modulados em impulso pelo processo indicado na figura 1.8. Os intervalos entre pulsos de um dos sinais podem ser ocupados pelas amostras de outros sinais.

No receptor, as amostras correspondentes a cada um dos sinais têm de ser separadas de modo a se poder reconstituir cada um dos sinais originais. Esta separação é efectuada por circuitos lógicos. Um cabo coaxial é capaz de transportar cerca de 1 800 sinais de voz multiplexados. A multiplexagem permite pois aumentar o rendimento das comunicações na medida em que permite utilizar ao máximo a capacidade de transmissão de informação disponível do canal físico utilizado.

⁴CCIR –Consulting Committee for International Radio, é um dos comités da International Telecommunications Union (ITU), organização internacional para a normalização em telecomunicações

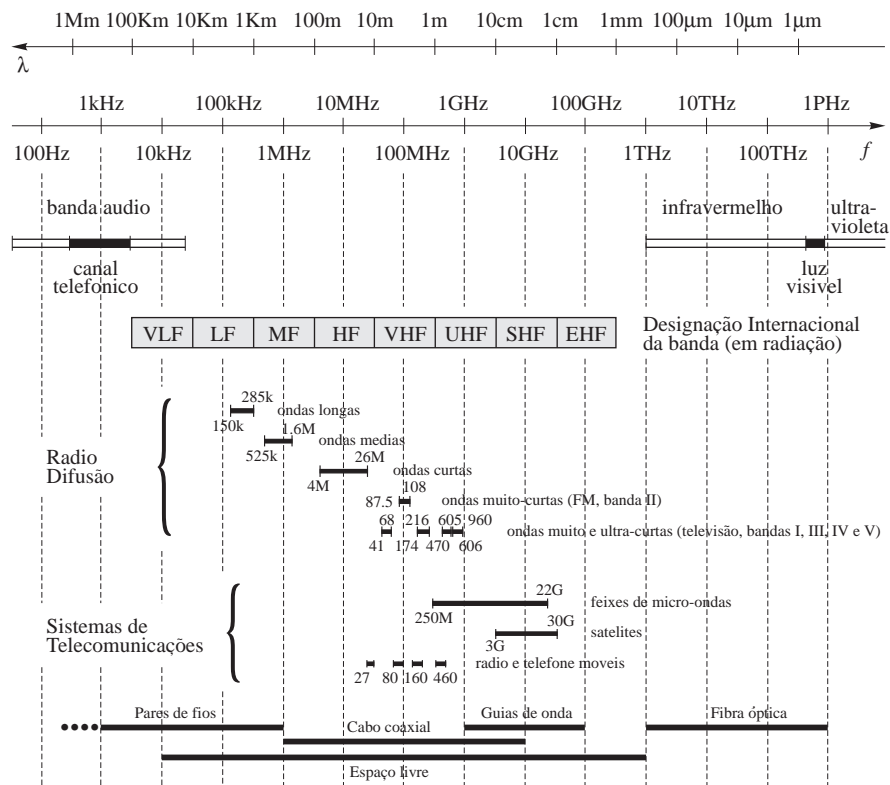


Figura 1.9: Utilização do espectro electromagnético

1.5.3 Técnicas de Codificação

Descreveu-se a modulação como uma operação de *processamento do sinal*. A *codificação* é uma operação de *processamento do símbolo* destinada a melhorar a comunicação quando a informação é digital e pode ser representada por símbolos discretos. Tanto a codificação como a modulação podem tornar-se necessárias, em conjunto, para se conseguir uma transmissão fiável a longa distância.

A operação de *codificação* transforma uma mensagem digital numa nova sequência de símbolos. A *descodificação* converte a sequência codificada na mensagem original, provavelmente com alguns erros, que surgiram na transmissão. A maior parte dos métodos de codificação envolve circuitos

lógicos manipulando símbolos binários.

Considere-se um terminal de computador ou uma outra qualquer fonte discreta cujo alfabeto contém $M \gg 2$ símbolos. A transmissão não-codificada de uma mensagem desta fonte necessitaria de M formas de onda diferentes, uma para cada símbolo. Alternativamente, cada símbolo pode ser representado por uma *palavra de código binária*, um número de k dígitos binários. Dado que existem 2^k palavras diferentes com k dígitos binários, são necessários $k \geq \log_2 M$ dígitos por cada palavra de código para codificar os M símbolos da fonte. Se a fonte produzir r_s símbolos por segundo, o código binário é produzido a um ritmo de $r_b = kr_s$ dígitos por segundo e a largura de banda necessária para a transmissão é k vezes maior do que seria necessária para a transmissão do sinal não-codificado.

Apesar de requerer uma maior largura de banda, a codificação em binário de uma fonte M -ária apresenta duas vantagens. Em primeiro lugar um sinal binário só necessita de duas formas de onda distintas e os circuitos electrónicos serão menos complicados. Em segundo lugar o ruído afecta menos um sinal binário do que um sinal composto por M formas de onda diferentes. Trata-se, portanto, de um método de codificação essencialmente destinado a reduzir o ruído por alargamento da banda do sinal.

A *codificação para controlo de erros* vai um passo mais além no sentido da redução do ruído por alargamento da banda. Adicionando *dígitos de verificação* suplementares a cada palavra de código binária é possível detectar ou mesmo corrigir a maior parte dos erros. Este tipo de codificação aumenta tanto a largura de banda como a complexidade do *hardware*, desvantagens que são compensadas ao obter-se uma comunicação isenta de erros apesar de uma razão sinal-ruído baixa. A utilização do conhecido dígito de verificação de *paridade* constitui um exemplo, embora o mais rudimentar, deste tipo de codificação.

Considere-se agora a situação oposta, isto é, o caso de uma fonte discreta e binária e um sistema de comunicação com uma razão sinal-ruído adequada mas pequena largura de banda. Esta situação é típica em redes de comunicação por computador que utilizam canais telefónicos. A codificação de blocos de k dígitos binários em símbolos M -ários *reduz* a largura de banda do sinal de um factor $k = \log_2 M$, permitindo portanto aumentar o ritmo de informação num canal de largura de banda limitada. Existem ainda outras técnicas mais sofisticadas, de natureza estatística, designadas genericamente por *codificação da fonte* com as quais se con-

segue *comprimir* ainda mais a largura de banda do sinal. Basicamente, estas técnicas tentam reduzir ou mesmo eliminar a *redundância* dos sinais ou mensagens que são produzidos pela fonte.

Finalmente interessa referir que os benefícios da codificação digital podem ser incorporados em sistemas de comunicação *analógicos* através da utilização de conversão analógico-digital tais como a modulação de impulso codificado (MIC), mais conhecido por PCM (Pulse Code Modulation). Um sinal PCM é gerado por amostragem do sinal analógico, dando origem ao sinal PAM, seguido de quantização das amplitudes e da sua conversão em números binários. O sistema PCM constitui a tecnologia mais utilizada na digitalização de sistemas de comunicação analógicos, tais como o telefone e a televisão. A manipulação dos sinais pode ser efectuada com microprocessadores em lugar de circuitos electrónicos analógicos que são, em geral, mais complexos.

1.6 Problemas

- 1.1 – Considere a expressão de lei de Hartley-Shannon dada pela equação 1.2. O valor da potência média total do ruído, N , representa sómente a parte do ruído presente no meio ambiente do sistema de comunicação que *passa* até ao destino pelo facto do sistema possuir uma largura de banda de valor B_T Hz. Isto é, se o ruído no canal de transmissão se apresentar com um valor η Watt/Hz, a potência média do ruído no destino é de $N = \eta \cdot B_T$ Watt. Considere que num sistema de comunicação se utiliza um sinal eléctrico com uma potência média tal que $\frac{S}{\eta} = 10^4$.

Determine o ritmo máximo de transmissão de informação (capacidade do canal) pelo sistema de comunicação se este possuir uma largura de banda de:

- a) $B_T = 1$ KHz
- b) $B_T = 10$ KHz
- c) $B_T = 100$ KHz

Em que unidades poderá ser apresentado o resultado?

- 1.2 – Considere um sistema de transmissão com uma largura de banda $B_T = 3$ KHz. Determine o valor mínimo que a razão de potências

sinál-ruído, $\frac{S}{N}$, deve ter para se obter uma transmissão fiável de informação com ritmos de

- a) 2 400 *bits por segundo*
- b) 4 800 *bits por segundo*
- c) 9 600 *bits por segundo*

Apresente os resultados em *decibéis*.

- 1.3 – Um sistema de transmissão possui largura de banda $B_T = 1 \text{ KHz}$. O ruído eléctrico tem uma potência média uniforme em toda a banda do espectro electromagnético, com valor $\eta = 1 \mu\text{W/Hz}$.

Determine o valor mínimo que a potência do sinal, S , deve ter para se conseguir uma transmissão de informação fiável às velocidades de

- a) 100 *bits/s*
- b) 1 000 *bits/s*
- c) 10 000 *bits/s*

Apresente os resultados em *miliWatt*

- 1.4 – Considere a figura 1.9 onde se representam as bandas de frequência atribuídas aos vários serviços de comunicações. Dê uma ideia dos comprimentos dos elementos de antena para a emissão e recepção de

- a) Televisão na banda I (VHF)
- b) Televisão na banda III (UHF)
- c) Rádio em FM (banda II, UHF)
- d) Satélites

- 1.5 – Um terminal de computador possuindo um alfabeto de 256 caracteres, transmite pela sua porta de saída a uma velocidade de 1 200 caracteres por segundo. A linha de transmissão não garante inexistência de erros acima dos 5 000 símbolos eléctricos por segundo.

- a) Indique como poderia fazer a codificação dos caracteres para a transmissão nessa linha;
- b) Se o alfabeto do terminal fosse o CCITT IA 5 (eq ASCII) esquematize a forma de onda na linha correspondente à transmissão da sequência de caracteres **AR**;
- c) Qual seria o ritmo de símbolos na linha?

fim do capítulo 1

Capítulo 2

Análise de Sinais

Os sinais de comunicação eléctricos são quantidades variáveis no tempo, voltagens ou correntes. Embora costumem, como grandeza física que são, ser representados por funções no domínio do tempo, são também susceptíveis de representação no *domínio da frequência* no qual podem ser considerados como uma sobreposição de componentes sinusoidais de várias frequências, amplitudes e fases. Esta descrição do sinal no domínio da frequência é designada de *espectro* do sinal.

A análise espectral de sinais utiliza as séries e as transformadas de Fourier da análise matemática, formalismos que constituem uma das ferramentas fundamentais da engenharia de telecomunicações. Estes formalismos permitem tratar grandes classes de sinais que possuam propriedades semelhantes no domínio da frequência independentemente dos detalhes que os distingam no domínio do tempo. A representação dos sinais do ponto de vista espectral torna-se muito mais sugestiva ao estudar-se a sua interação com os sistemas de transmissão os quais podem ser caracterizados por uma função de transferência, também definida como uma função da frequência.

Este capítulo debruça-se sobre a análise espectral dos sinais e dedica uma atenção especial à interpretação das suas propriedades no domínio da frequência. Examina os *espectros de linhas* baseados no desenvolvimento em série de Fourier dos sinais periódicos e os *espectros contínuos* baseados na transformada de Fourier dos sinais não-periódicos.

2.1 Sinais periódicos: Espectros de Linhas

O domínio da frequência vai ser introduzido e interpretado em termos de fasores rotativos. Começa-se por definir o espectro de linhas de um sinal sinusoidal. É depois obtido o espectro de um sinal periódico qualquer, de potência média finita, a partir da sua representação em série de Fourier.

2.1.1 Fasores e Espectros de Linhas

Consideremos a forma de onda sinusoidal $v(t)$ representada na figura 2.1. Por convenção, representam-se as sinusoides por funções cosseno, isto é,

$$v(t) = A \cos(\omega_0 t + \phi) \quad (2.1)$$

onde A é o valor de pico ou amplitude, ω_0 é a frequência angular, em radianos por segundo, e ϕ o ângulo de fase. Este ângulo tem a ver com o instante escolhido para origem dos tempos e representa o facto do valor máximo estar desviado dessa origem de um valor $t = -\phi/\omega_0$. A equação 2.1 significa que $v(t)$ se repete indefinidamente com um período de repetição $T_0 = 2\pi/\omega_0$.

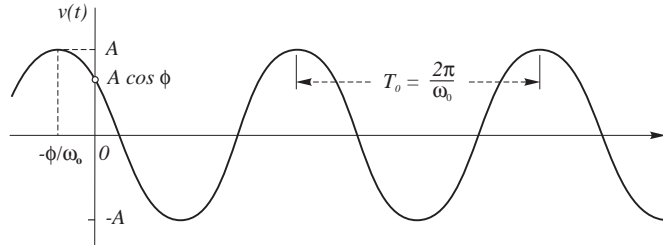


Figura 2.1: Forma de onda sinusoidal

O inverso do período é a frequência cíclica

$$f_0 = \frac{1}{T_0} = \frac{\omega_0}{2\pi} \quad (2.2)$$

medida em ciclos por segundo, ou Hertz (Hz). Evidentemente que nenhum sinal real existe indefinidamente mas a equação 2.1 constitui um *modelo* razoável para uma forma de onda sinusoidal com uma duração muito maior do que o seu período. Tal como em análise da corrente alterna

estacionária, a senoide pode ser representada no plano complexo por uma exponencial ou *fasor*. Esta representação baseia-se no teorema de Euler segundo o qual

$$e^{\pm j\theta} = \cos \theta \pm j \sin \theta \quad (2.3)$$

em que $j = \sqrt{-1}$ e θ é um ângulo arbitrário. Fazendo $\theta = \omega_0 t + \phi$ pode representar-se qualquer senoide como sendo a parte real de uma exponencial complexa

$$A \cos(\omega_0 t + \phi) = A \Re [e^{j(\omega_0 t + \phi)}] = \Re [A e^{j\omega_0 t} \cdot e^{j\phi}] \quad (2.4)$$

designada de *representação por fasor* porque o termo dentro do parêntesis recto pode ser considerado como um vector rotativo no plano complexo, conforme mostra a figura 2.2.

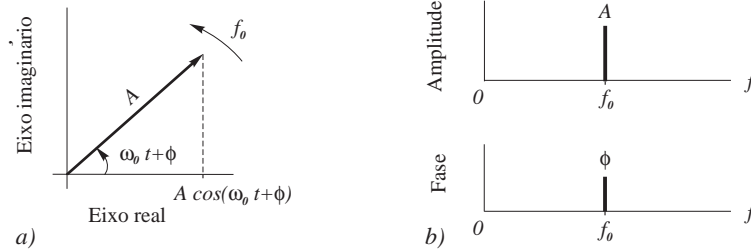


Figura 2.2: Representação de senoide por um fasor

O fasor tem comprimento A , roda em sentido retrógrado a f_0 rotações por segundo e no instante t_0 faz um ângulo de ϕ radianos com o eixo real positivo. A projecção do fasor no eixo real reproduz a senoide da equação 2.4. O fasor fica completamente definido por três parâmetros: a amplitude, o ângulo de fase e a frequência de rotação.

Para descrever o mesmo fasor no *domínio da frequência* torna-se necessário associar a amplitude e a fase à frequência. Assim, uma descrição conveniente no domínio da frequência será o *espectro de linha* da figura 2.2(b), o qual consiste em dois gráficos: o da amplitude em função da frequência e o da fase em função da frequência. Embora esta representação seja simples e pareça trivial, possui grande valor conceptual quando aplicada a sinais mais complexos. Na representação espectral adoptaremos as seguintes convenções:

- (i) A variável independente é a *frequência*, f em Hz. A frequência

angular ω , em radianos/seg, é uma notação sintética para o valor $2\pi f$

- (ii) Os ângulos de fase são medidos relativamente a funções *cosseno*. Os senos serão convertidos a cosseno através da identidade $\sin \omega t = \cos(\omega t - 90^\circ)$
- (iii) A amplitude é sempre uma quantidade *positiva*. Quando aparecerem sinais com amplitude negativa, esta será absorvida na fase, isto é,
 $-A \cos(\omega t) = A \cos(\omega t \pm 180^\circ)$
- (iv) Os ângulos de fase são expressos em *graus* embora ângulos tais como ωt sejam inerentemente em radianos.

A figura 2.3 ilustra estas convenções. Os gráficos da figura 2.3(b), chamados *espectros de linhas unilaterais*, podem ser construídos para qualquer combinação linear de sinusóides.

Uma outra representação espectral ainda mais valiosa, embora envolva frequências negativas, é o *espectro de linhas bilateral*. Esta representação decorre da equação 2.4, tendo em atenção que $\Re[z] = 1/2(z + z^*)$ e fazendo $z = A e^{j2\pi f_0 t} \cdot e^{j\phi}$ e $z^* = A e^{-j2\pi f_0 t} \cdot e^{-j\phi}$ tem-se

$$A \cos(2\pi f_0 t + \phi) = \frac{A}{2} e^{j2\pi f_0 t} \cdot e^{j\phi} + \frac{A}{2} e^{-j2\pi f_0 t} \cdot e^{-j\phi} \quad (2.5)$$

O lado direito da identidade 2.5 representa um par de fasores conjugados representados na figura 2.4 juntamente com o respectivo espectro de linhas. Os fasores têm amplitudes iguais e giram em sentidos opostos à mesma frequência f_0 . O fasor-soma situa-se sempre no eixo real e a sua amplitude é igual a $A \cos(2\pi f_0 t + \phi)$. O espectro de amplitude tem simetria par enquanto que o espectro de fase tem simetria impar pois os fasores são conjugados.

Deve notar-se que os espectros de linhas, tanto o unilateral como o bilateral, constituem tão-sómente representações pictóricas de sinusóides ou fasores funções do tempo. A figura 2.5 é a versão bilateral do espectro da figura 2.3.

Uma linha no espectro unilateral representa uma onda cosseno real, enquanto que uma linha no espectro bilateral representa uma exponencial complexa, donde, para se obter uma onda cosseno real, se deve adicionar o

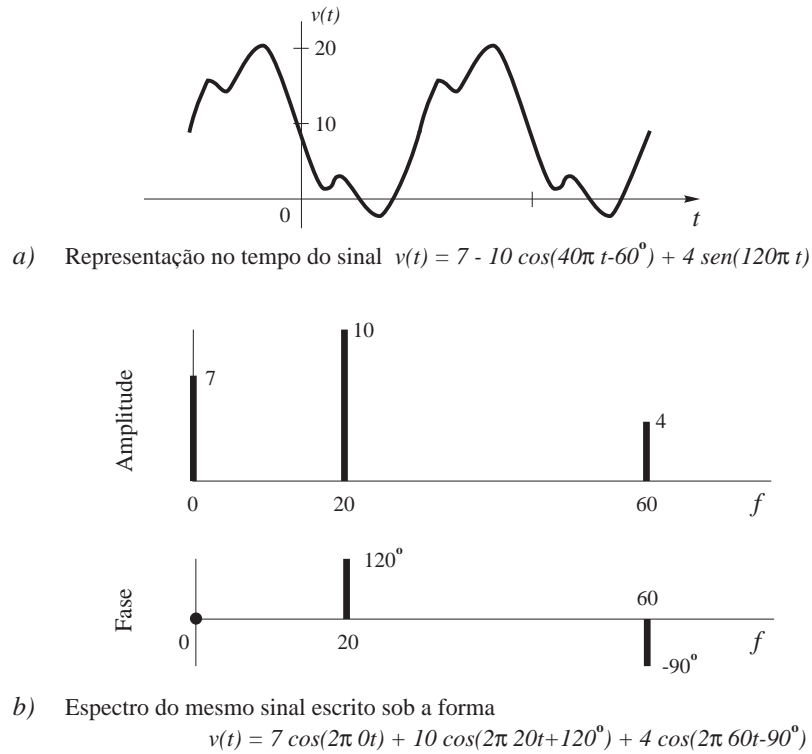


Figura 2.3: Representação de um sinal periódico e seu espectro de linhas

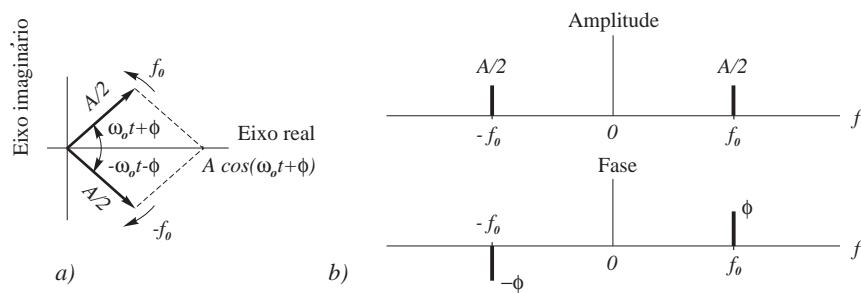


Figura 2.4: (a) Fasores conjugados; (b) Espectro de linhas bilateral

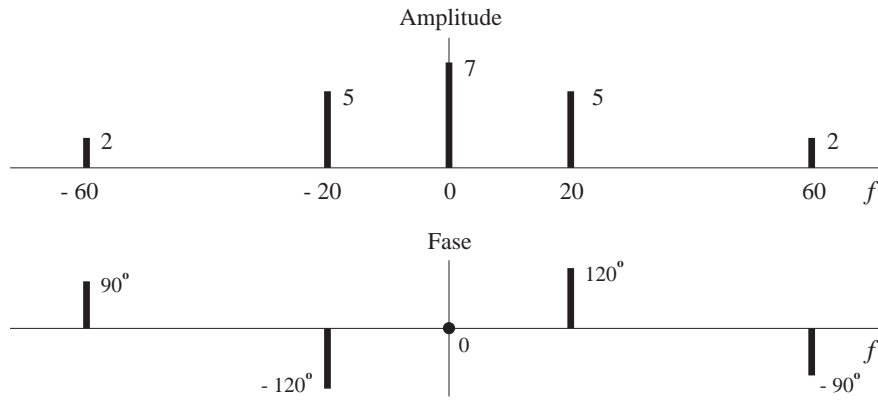


Figura 2.5: Versão bilateral do espectro da figura 2.4

termo conjugado. Assim, sempre que se fizer referência a um intervalo de frequência $[f_1, f_2]$ num espectro bilateral, está implicitamente nele contido o correspondente intervalo de frequências negativas, isto é, o intervalo é definido por $f_1 \leq |f| \leq f_2$.

Finalmente interessa notar que o espectro de amplitude, em qualquer das versões, fornece mais informação sobre o sinal do que o espectro de fase. Ambas as partes do espectro (amplitude e fase), porém, são necessárias à definição da função no domínio do tempo mas o espectro de amplitude, só por si, indica quais as frequências presentes no sinal e em que proporção. Por outras palavras, o espectro de amplitude *mostra o conteúdo em frequências* de um sinal.

2.1.2 Sinais Periódicos e Potência Média

As sinusoides e os fasores pertencem à classe dos *sinais periódicos*, isto é, os sinais que permanecem inalterados perante uma translação temporal de T_0 seg:

$$v(t \pm mT_0) = v(t) \quad -\infty < t < \infty$$

em que T_0 é o período do sinal. O sinal fica portanto completamente definido pelos valores que toma num qualquer intervalo de amplitude T_0 . Definamos agora alguns valores médios referentes a um sinal genérico $v(t)$.

O *valor médio* de $v(t)$ ao longo de todo o tempo é definido por

$$\langle v(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} v(t) dt \quad (2.6)$$

O integral dá a área sob a curva $v(t)$ num qualquer intervalo de duração T segundos que dividido por T dá o valor médio nesse intervalo. Fazendo o intervalo aumentar indefinidamente, $T \rightarrow \infty$, obtem-se o valor médio em todo o tempo. No caso de um sinal periódico, o valor médio para todo o tempo é igual ao valor médio num período

$$\langle v(t) \rangle = \frac{1}{T_0} \int_{t_1}^{t_1+T_0} v(t) dt = \frac{1}{T_0} \int_{T_0} v(t) dt \quad (2.7)$$

onde \int_{T_0} designa o integral calculado num intervalo qualquer de amplitude igual a um período.

Suponhamos que $v(t)$ é a voltagem aos terminais de uma resistência de valor R Ohms. O sinal $v(t)$ dará origem a uma corrente eléctrica $i(t) = \frac{v(t)}{R}$. Pode-se então calcular a potência média dissipada na resistência calculando o valor médio da potência instantânea $s_v(t) = v(t) \cdot i(t) = \frac{v^2(t)}{R}$. Na realidade, por vezes, não se sabe se um sinal é uma tensão ou uma corrente. Por isso se convencionou *normalizar* a potência supondo que $R = 1 \Omega$. Nestas condições, define-se *potência média*, S , associada a um sinal periódico arbitrário pela expressão

$$S = \langle |v(t)|^2 \rangle = \frac{1}{T_0} \int_{T_0} |v(t)|^2 dt \quad (2.8)$$

O valor de S será real e não-negativo. Quando o integral 2.8 existir e fôr limitado, isto é, se fôr $0 < S < \infty$, $v(t)$ é designado de sinal periódico *de potência*.

2.1.3 Série de Fourier

O sinal $v(t)$ da figura 2.3 foi obtido a partir da soma de uma parcela constante e de duas sinusoides. Indo agora em sentido contrário, procuremos decompôr os sinais periódicos em somas sinusoidais. A expansão de uma função em série exponencial de Fourier permite efectuar esta decomposição.

Seja $v(t)$ um sinal de potência de período $T_0 = \frac{1}{f_0}$. O seu desenvolvimento em série exponencial de Fourier é da forma

$$v(t) = \sum_{n=-\infty}^{+\infty} C_n e^{j2\pi n f_0 t} \quad n = 0, \pm 1, \pm 2, \dots \quad (2.9)$$

em que os coeficientes C_n da série são dados por

$$C_n = \frac{1}{T_0} \int_{T_0} v(t) e^{-j2\pi n f_0 t} dt \quad (2.10)$$

isto é, C_n é o valor médio do produto $v(t) \cdot e^{-j2\pi n f_0 t}$. Dado que, em geral, os coeficientes são números complexos, podem ser expressos na forma polar $C_n = |C_n| e^{j \arg C_n}$ onde $\arg C_n$ é o ângulo do complexo C_n . A equação 2.9 desenvolve um sinal periódico de potência numa soma infinita de fasores em que a n -ésima parcela é dada por

$$C_n e^{j2\pi n f_0 t} = |C_n| \cdot e^{j2\pi n f_0 t} \cdot e^{j \arg C_n}$$

A série existe e converge uniformemente se o sinal periódico $v(t)$ tiver um número finito de máximos, mínimos e descontinuidades por período e se $v(t)$ for absolutamente integrável possuindo área finita por período. Estas condições¹ são suficientes mas não estritamente necessárias. Uma condição alternativa é que $v(t)$ possua *quadrado somável* de forma que $|v(t)|^2$ tenha área finita por período, o que equivale a um sinal de potência. Pode dizer-se que todos os sinais com interesse que ocorrem na prática obedecem a estas condições.

O sinal $v(t)$ na equação 2.9 consiste numa soma de fasores de amplitude $|C_n|$ e ângulo $\arg C_n$ com frequências $n f_0 = 0, \pm f_0, \pm 2 f_0, \dots$ portanto a sua representação gráfica no domínio da frequência consiste num espectro de linhas bilateral definido pelos coeficientes da série. Para reforçar esta interpretação espectral também se escreve

$$C(n f_0) \equiv C_n$$

pelo que $|C(n f_0)|$ representará o *espectro de amplitude* como uma função de f e $\arg C(n f_0)$ representará o *espectro de fase*. Interessa realçar três propriedades importantes dos espectros dos sinais de potência:

¹condições de Dirichlet

- (i) Todas as frequências são múltiplas inteiras, *harmónicas*, da *frequência fundamental*, $f_0 = \frac{1}{T_0}$. Assim, as linhas espectrais estão uniformemente espaçadas de um valor igual a f_0
- (ii) A componente constante é igual ao *valor médio* do sinal, dado que para $n = 0$, a equação 2.10 dá

$$C_0 = \frac{1}{T_0} \int_{T_0} v(t) dt = \langle v(t) \rangle \quad (2.11)$$

- (iii) Se o sinal $v(t)$ for real, então

$$C_{-n} = C_n^* = |C_n| e^{-j \arg C_n} \quad (2.12)$$

o que se deduz a partir da equação 2.10 substituindo n por $-n$. Assim, tem-se $|C(-nf_0)| = |C(nf_0)|$ e $\arg C(-nf_0) = -\arg C(nf_0)$ o que significa que o espectro de amplitude tem simetria par (é simétrico relativamente ao eixo das ordenadas) e o espectro de fase tem simetria impar (é simétrico em relação à origem dos eixos).

Para sinais reais, a terceira propriedade permite reagrupar a série exponencial em *pares conjugados* excluindo a parcela C_0 . A equação 2.9 toma então a forma

$$v(t) = C_0 + \sum_{n=1}^{\infty} |2C_n| \cos(2\pi n f_0 t + \arg C_n) \quad (2.13)$$

que se designa por *série trigonométrica de Fourier* sugerindo um espectro sob a forma unilateral. É usual, porém, utilizar-se a série exponencial e o espectro bilateral.

Antes de considerarmos o exemplo na secção que se segue, interessa introduzir uma função que aparece frequentemente. O cálculo de C_n (2.10) envolve muitas vezes o cálculo do valor médio de um fasor, isto é, o integral

$$\frac{1}{T} \int_{-T/2}^{T/2} e^{j2\pi f t} dt = \frac{1}{j2\pi f T} \left(e^{j\pi f T} - e^{-j\pi f T} \right) = \frac{1}{\pi f T} \sin(\pi f T) \quad (2.14)$$

dá uma função do tipo $\frac{\sin(\pi\alpha)}{\pi\alpha}$ que passaremos a designar por $\text{sinc}(\alpha)$, isto é,

$$\text{sinc}(\alpha) \equiv \frac{\sin(\pi\alpha)}{\pi\alpha} \quad (2.15)$$

onde α é a variável independente. A figura 2.6 representa graficamente esta função cujo máximo absoluto ocorre em $\alpha = 0$ e tem zeros em todos os valores inteiros de α

$$\text{sinc}(\alpha) = \begin{cases} 1 & \text{se } \alpha = 0 \\ 0 & \text{se } \alpha = \pm 1, \pm 2, \dots \end{cases}$$

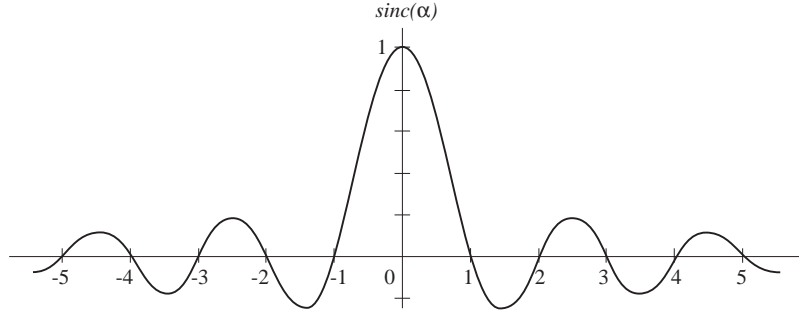


Figura 2.6: Representação da função $\text{sinc}(\alpha)$

2.1.4 Sequência de pulsos rectangulares

Consideremos a sequência de pulsos rectangulares da figura 2.7. Cada pulso tem amplitude A e duração τ . Existem descontinuidades nos extremos de cada pulso, $t = nT_0 \pm \tau/2$, portanto os valores de $v(t)$ são indefinidos nesses pontos de descontinuidade. Este facto realça a diferença

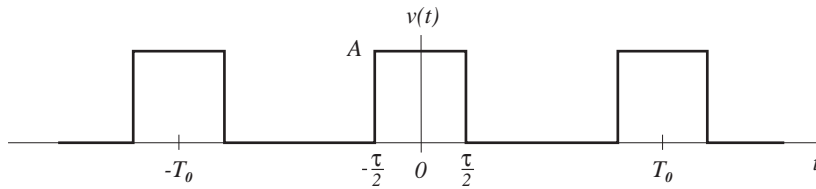


Figura 2.7: Sequência de pulsos rectangulares

entre um sinal com existência física e o correspondente modelo matemático, visto que o primeiro nunca apresenta transições instantâneas. Contudo, o modelo é razoável se os tempos de transição forem muito pequenos quando comparados com a duração dos pulsos.

Para calcular os coeficientes de Fourier, tomemos como limites de integração na equação 2.10 os limites do período central $t \in [-\frac{T_0}{2}, +\frac{T_0}{2}]$. Obtém-se sucessivamente

$$\begin{aligned} C_n &= \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} v(t) e^{-j2\pi n f_0 t} dt \\ C_n &= \frac{1}{T_0} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} A e^{-j2\pi n f_0 t} dt \\ C_n &= \frac{A}{-j2\pi n f_0 T_0} \left(e^{-j\pi n f_0 \tau} - e^{+j\pi n f_0 \tau} \right) \\ C_n &= A f_0 \frac{\sin(\pi n f_0 \tau)}{\pi n f_0} \end{aligned}$$

Multiplicando e dividindo por τ dá finalmente

$$C_n = A f_0 \tau \operatorname{sinc}(n f_0 \tau)$$

o que decorre imediatamente da definição 2.15 com $\alpha = n f_0 \tau$.

A figura 2.8 mostra o espectro de amplitude obtido a partir de $|C(n f_0)| = A f_0 \tau |\operatorname{sinc}(n f_0 \tau)|$ e para o caso em que $\frac{\tau}{T_0} = f_0 \tau = \frac{1}{4}$. As linhas espectrais situadas nos múltiplos inteiros de f_0 têm uma amplitude definida

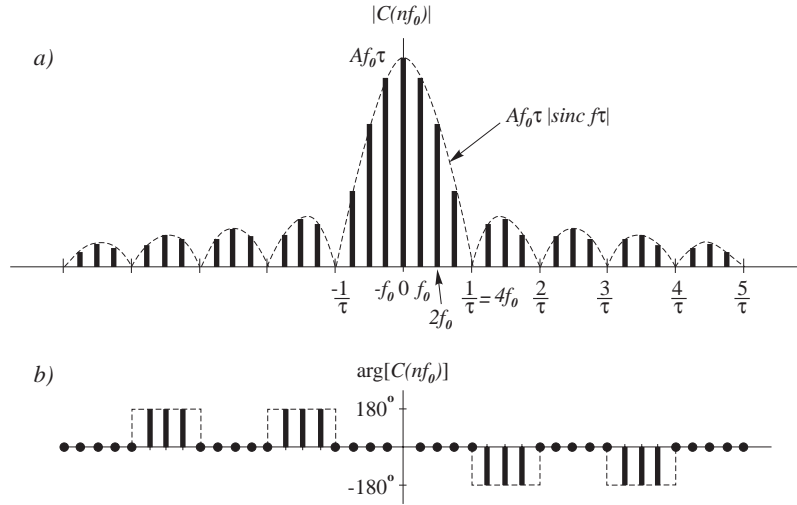


Figura 2.8: Espectro da sequência de pulsos rectangulares com $f_0\tau = 1/4$

pela função sinc, isto é, a "envolvente" das amplitudes é a função sinc que aparece a tracejado na figura 2.8. As linhas espectrais em $\pm 4f_0$, $\pm 8f_0$, etc, não existem dado que se situam precisamente sobre os múltiplos de $\frac{1}{\tau}$ onde a envolvente é nula. A componente constante tem amplitude $C(0) = \frac{A\tau}{T_0}$, valor que poderia ser obtido pelo cálculo do valor médio de $v(t)$, figura 2.7.

O espectro de fase na figura 2.8(b) obtem-se observando que C_n é sempre real mas por vezes negativo. Portanto $\arg C(nf_0)$ toma os valores 0° e $\pm 180^\circ$ dependendo do sinal de $\text{sinc}(nf_0\tau)$.

A sequência de pulsos rectangulares está pois decomposta nas suas componentes de frequência. Consideremos agora a operação inversa, isto é, a construção da sequência de pulsos a partir das sinusoides. Para tal escrevemos a série trigonométrica da equação 2.13 com $\frac{\tau}{T_0} = f_0\tau = \frac{1}{4}$ e portanto com $C(0) = \frac{A}{4}$ e $|2C_n| = \frac{2A}{4} |\text{sinc}(\frac{n}{4})| = \frac{2A}{n\pi} |\sin(\frac{n\pi}{4})|$. Vem

$$v(t) = \frac{A}{4} + \frac{\sqrt{2}A}{\pi} \cos(2\pi f_0 t) + \frac{A}{\pi} \cos(4\pi f_0 t) + \frac{\sqrt{2}A}{3\pi} \cos(6\pi f_0 t) + \dots$$

Somando as parcelas até à terceira harmónica dá a aproximação de $v(t)$ esquematizada na figura 2.9. Grosso modo, esta aproximação assemelha-se à sequência de pulsos, mas faltam-lhe as transições bruscas nos "cantos" dos rectângulos. Se se desejar uma melhor aproximação podem-se tomar mais harmónicas, por exemplo, até à 7ª, que é o caso da figura 2.9(b).

De notar que as harmónicas de ordem superior e de menor amplitude servem principalmente para aguçar os cantos da onda rectangular. Deve-se notar também que a série converge para o ponto médio de amplitude $\frac{A}{2}$ em $t = \pm \frac{\tau}{2}$ onde $v(t)$ tem as descontinuidades.

Exercício 2.1 *Esquematizar o espectro de amplitude de uma sequência de pulsos rectangulares para cada um dos seguintes casos:*

$$\tau = \frac{T_0}{5} \quad \tau = \frac{T_0}{2} \quad \tau = T_0$$

No último caso a sequência de pulsos degenera numa constante ao longo do tempo. Como é que este facto transparece no espectro?

2.1.5 Teorema da Potência

Este teorema relaciona a potência média S de um sinal periódico com os seus coeficientes de Fourier.

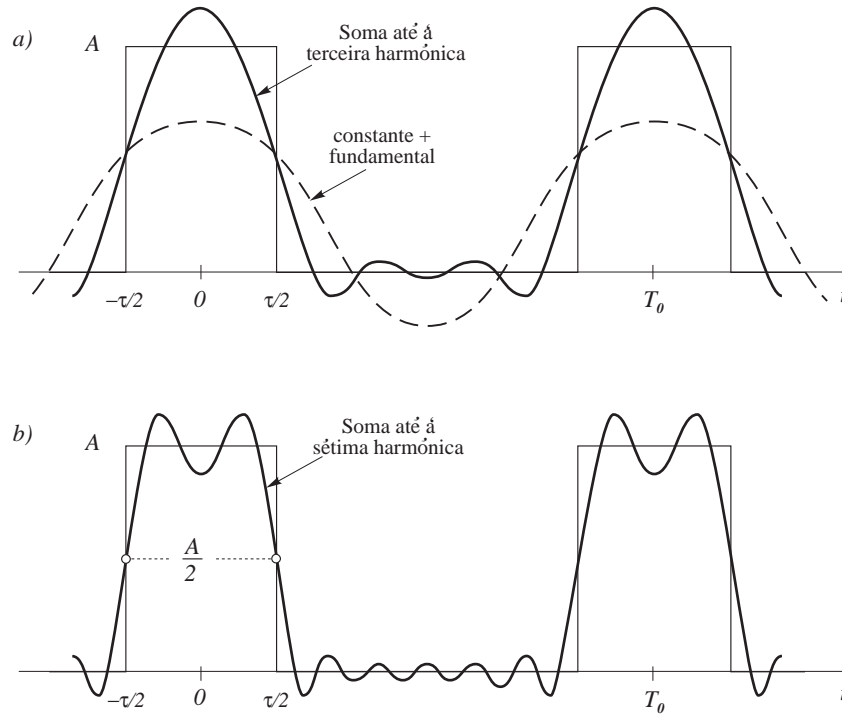


Figura 2.9: Reconstrução por série de Fourier de uma sequência de pulsos

Se se admitir que $v(t)$ é uma função complexa – o caso mais geral – pode escrever-se

$$|v(t)|^2 = v(t) \cdot v^*(t)$$

onde $v^*(t)$ é o complexo conjugado de $v(t)$. A potência média de $v(t)$ é

$$S = \frac{1}{T_0} \int_{T_0} |v(t)|^2 dt \quad \Leftrightarrow \quad S = \frac{1}{T_0} \int_{T_0} v(t) \cdot v^*(t) dt$$

Substituindo $v^*(t)$ pela sua série exponencial vem,

$$\begin{aligned} S &= \frac{1}{T_0} \int_{T_0} v(t) \left[\sum_{n=-\infty}^{+\infty} C_n^* e^{-j2\pi n f_0 t} \right] dt \\ S &= \sum_{n=-\infty}^{+\infty} \left[\frac{1}{T_0} \int_{T_0} v(t) e^{-j2\pi n f_0 t} dt \right] C_n^* \end{aligned}$$

em que o integral que figura no somatório é igual a C_n . Tem-se portanto,

$$\begin{aligned} S &= \sum_{n=-\infty}^{+\infty} C_n C_n^* \\ S &= \sum_{n=-\infty}^{+\infty} |C_n|^2 \end{aligned}$$

resultado que constitui o teorema da potência de Parseval.

A interpretação espectral deste resultado é simples: a potência média dum sinal pode ser determinada quadrando e adicionando as alturas $|C_n| \equiv |C(nf_0)|$ das linhas do espectro de amplitude. O teorema de Parseval implica a sobreposição da potência média dado que a potência média total de um sinal $v(t)$ é a soma das potências médias das suas componentes de frequência.

2.2 Sinais não-periódicos: Espectros Contínuos

Passemos dos sinais periódicos aos não-periódicos, ou seja, aos sinais que só existem durante um intervalo de tempo relativamente curto. Se um sinal não-periódico possui uma energia total finita mas não nula, será representado, no domínio da frequência, por um espectro contínuo que é dado pela transformada de Fourier desse sinal.

2.2.1 A Transformada de Fourier

A figura 2.10 mostra um sinal não-periódico típico. O pulso rectangular é *estritamente limitado no tempo* dado que $v(t)$ é identicamente nulo fora do intervalo $[-\tau/2, +\tau/2]$. Tais sinais são designados, de uma maneira geral, por *pulsos*. Em qualquer dos casos ao calcularem-se as médias de $v(t)$ ou de $|v(t)|^2$ para todo o tempo verifica-se que são iguais a zero. Consequentemente, em lugar de se falar em potência média, nos sinais não-periódicos, interessa considerar a sua energia total.

Se $v(t)$ representar uma tensão aos terminais de uma resistência R , a energia total transferida para essa resistência será dada pelo integral da potência instantânea $\frac{v^2(t)}{R}$. Definiremos portanto *energia normalizada*, E , de um sinal por

$$E \doteq \int_{-\infty}^{+\infty} |v(t)|^2 dt \quad (2.16)$$

A energia do sinal é dada pela área total sob a curva definida por $|v(t)|^2$. Por exemplo, a energia normalizada de um pulso rectangular de amplitude A é $E = A^2\tau$.

Se o integral 2.16 existir e fôr limitado, isto é, se $0 < E < \infty$, o sinal $v(t)$ tem uma energia bem definida e será chamado de *sinal de energia não-periódico*. A maior parte dos sinais limitados no tempo de interesse prático satisfazem esta condição que é necessária para que o sinal possa ser analisado por transformada de Fourier.

O espectro do sinal periódico foi definido pela função discreta C_n definida no conjunto domínio $\{\dots, -2f_0, -f_0, 0, +f_0, +2f_0, \dots\}$

$$C_n = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} v(t) e^{-j2\pi n f_0 t} dt$$

Um sinal não-periódico, por sua vez, pode ser considerado um sinal periódico de período infinito, definindo-se o seu espectro pelo limite de $T_0 C_n$ quando $T_0 \rightarrow \infty$, isto é

$$\lim_{T_0 \rightarrow \infty} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} v(t) e^{-j2\pi n f_0 t} dt = \int_{-\infty}^{+\infty} v(t) e^{-j2\pi f t} dt$$

onde, na passagem ao limite, $n f_0$ se transforma na variável contínua f , passando o espectro do sinal não-periódico $v(t)$ a estar definido no intervalo contínuo $]-\infty, +\infty[$. Designaremos essa função por $V(f)$, a qual representa a *Transformada de Fourier*, ou *espectro*, do sinal $v(t)$.

$$V(f) \doteq \mathcal{F}[v(t)] = \int_{-\infty}^{+\infty} v(t) e^{-j2\pi f t} dt \quad (2.17)$$

Assim, o sinal $v(t)$ já não é uma soma de componentes de frequência discreta mas o integral de uma função $V(f)$ num intervalo contínuo, função essa que tem as dimensões de uma densidade de amplitude, *Volts/Hz*. Isto é, enquanto o sinal periódico se obtinha por soma

$$v(t) = \sum_{n=-\infty}^{+\infty} C_n e^{j2\pi n f_0 t} \quad n = 0, \pm 1, \pm 2, \dots$$

o sinal não-periódico obtem-se por integração

$$v(t) \doteq \int_{-\infty}^{+\infty} V(f) e^{j2\pi f t} df \quad (2.18)$$

integral que define a transformada inversa $\mathcal{F}^{-1}[V(f)]$. As equações 2.17 e 2.18 correspondem respectivamente às equações 2.10 e 2.9 da secção 2.1 para os sinais periódicos.

$V(f)$ desempenha o mesmo papel para os sinais não-periódicos que C_n para os sinais periódicos. Um sinal não-periódico terá um *espectro contínuo* em lugar de um espectro de linhas. Interessa realçar três das propriedades mais importantes da Transformada de Fourier $V(f)$ de um sinal $v(t)$

- (i) $V(f)$ é uma função complexa de variável real. $|V(f)|$ é o espectro de amplitude e $\arg V(f)$ é o espectro de fase
- (ii) O valor de $V(f)$ para $f = 0$ é igual à área de $v(t)$ pois $V(0) = \int_{-\infty}^{\infty} v(t) dt$ e já C_0 era igual ao valor médio de $v(t)$
- (iii) Se o sinal $v(t)$ for real então $V(-f) = V^*(f)$ e $|V(-f)| = |V(f)|$ e $\arg V(-f) = -\arg V(f)$ pelo que o espectro de amplitude tem simetria par e o de fase tem simetria ímpar.

2.2.2 Pulso rectangular

Na secção 2.1.4 determinámos o espectro de linhas de uma sequência de pulsos rectangulares. Consideremos agora pulso rectangular da figura 2.10(a). Este modelo de sinal é tão frequente que se utiliza o símbolo Π para o representar. Adoptaremos a seguinte notação

$$\Pi\left(\frac{t}{\tau}\right) = \begin{cases} 1 & |t| < \frac{\tau}{2} \\ 0 & |t| > \frac{\tau}{2} \end{cases} \quad (2.19)$$

para simbolizar uma função rectangular com amplitude unitária e duração τ , centrada em $t = 0$. O pulso da figura será pois representado por

$$v(t) = A \Pi\left(\frac{t}{\tau}\right) \quad (2.20)$$

Inserindo este $v(t)$ na definição (2.17) da transformada, vem

$$\begin{aligned} V(f) &= \int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} A e^{-j2\pi ft} dt \\ V(f) &= \frac{A}{\pi f} \sin(\pi f \tau) \\ V(f) &= A\tau \operatorname{sinc}(f\tau) \end{aligned} \quad (2.21)$$

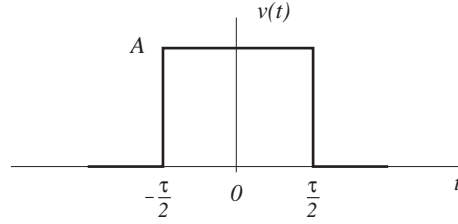


Figura 2.10: Pulso rectangular não-periódico

e tem-se $V(0) = A\tau$ que é exactamente a área do pulso. O espectro correspondente está esquematizado na figura 2.11. Comparando este espectro com o da figura 2.8 verifica-se que à medida que $T_0 \rightarrow \infty$, as linhas se aproximam indefinidamente passando o espectro a consistir na curva contínua que inicialmente era a envolvente das amplitudes das mesmas linhas.

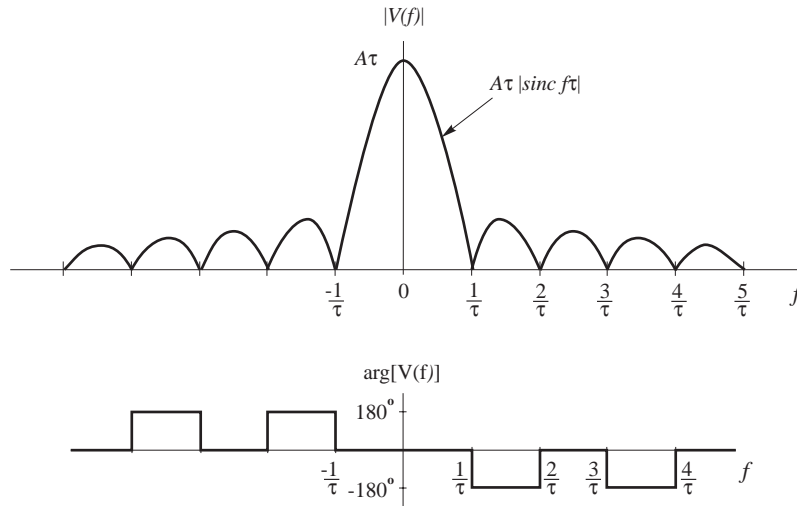


Figura 2.11: Espectro do pulso rectangular $V(f) = A\tau \text{sinc}(f\tau)$

Uma análise detalhada da figura 2.11 revela que a parte significativa do espectro se encontra no intervalo definido por $|f| < \frac{1}{\tau}$ dado que $|V(f)| \ll |V(0)|$ para $|f| > \frac{1}{\tau}$. Como se justificará mais adiante, na secção 2.2.4, podemos considerar o valor $\frac{1}{\tau}$ como uma medida da *largura espectral* do pulso rectangular de duração τ .

Se a largura do pulso for reduzida (redução de τ) a largura espectral é aumentada e vice-versa. Conclui-se então que pulsos curtos têm espectros largos e pulsos prolongados possuem espectros estreitos. Este fenómeno, chamado de *espalhamento recíproco* é uma propriedade geral de todos os sinais, quer sejam pulsos ou não, pois as variações temporais rápidas dão origem a componentes harmónicas de alta frequência enquanto que variações temporais mais lentas e suaves exibem menor conteúdo de altas frequências.

2.2.3 Teorema da Energia

O teorema da energia (ou teorema de Rayleigh) de sinais não-periódicos é análogo ao teorema da potência de Parseval para sinais periódicos. Diz que a energia, E , de um sinal $v(t)$ está relacionada com o seu espectro, $V(f)$, por

$$\begin{aligned} E &= \int_{-\infty}^{+\infty} V(f) \cdot V^*(f) df \\ E &= \int_{-\infty}^{+\infty} |V(f)|^2 df \end{aligned} \quad (2.22)$$

o que significa que a energia total de um sinal pode ser calculada integrando o quadrado do espectro de amplitude em $]-\infty, +\infty[$.

A importância da equação 2.22 não reside tanto no facto de se poder calcular o valor de E dessa forma, pois que normalmente o integral no domínio do tempo é mais fácil de calcular, isto é, $E = \int_{-\infty}^{+\infty} |v(t)|^2 dt$. A sua importância revela-se no facto de indicar que $|V(f)|^2$ representa a distribuição da energia do sinal no domínio da frequência designando-se, por isso, de *função de densidade espectral de energia*. Esta função aparece frequentemente representada pela letra G , isto é

$$G_v(f) \equiv |V(f)|^2$$

é a densidade espectral de energia do sinal v em *Joules/Hz*. A energia numa qualquer banda de frequência diferencial df *Hz* é igual a $|V(f)|^2 df$ *Joules*.

2.2.4 Largura de Banda

A interpretação dada acima a $|V(f)|^2$ fornece uma justificação quantitativa à noção de largura espectral no sentido em que a maior parte da

energia de um dado sinal deverá estar contida no intervalo de frequência que se tomar como sendo a *banda* do sinal, intervalo esse cuja amplitude define a *largura de banda* do sinal.

Para ilustrar este ponto considere-se a figura 2.12 que representa a densidade espectral de energia de um pulso rectangular cuja largura espectral se disse atrás ser de $\frac{1}{\tau}$.

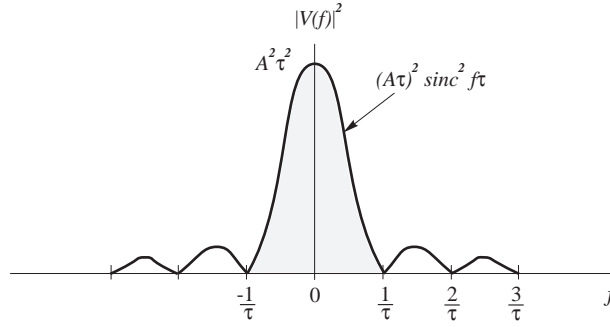


Figura 2.12: Densidade espectral de energia de um pulso rectangular

A energia, $E_{1/\tau}$, na banda $]-\frac{1}{\tau}, +\frac{1}{\tau}[$ é dada pela área sombreada na figura 2.12, situada entre o primeiro zero negativo e o primeiro zero positivo.

$$\begin{aligned} E_{1/\tau} &= \int_{-\frac{1}{\tau}}^{+\frac{1}{\tau}} |V(f)|^2 df \\ E_{1/\tau} &= \int_{-\frac{1}{\tau}}^{+\frac{1}{\tau}} (A\tau)^2 \text{sinc}^2(f\tau) df \\ E_{1/\tau} &= 0.92 A^2 \tau \end{aligned}$$

O cálculo do integral foi feito numericamente. A energia total do pulso é $E = \int_{-\infty}^{+\infty} |v(t)|^2 dt = A^2 \tau$, portanto, o intervalo espectral que na secção 2.2.2 se disse ter uma amplitude de $\frac{1}{\tau}$, contém mais de 90% da energia total do sinal.

Definição 2.1 —Largura de Banda de um sinal

Largura de Banda, B , de um sinal é a amplitude do menor intervalo espectral positivo que contém 90% da energia total do sinal (ou da sua potência média total, caso se trate de um sinal periódico).

2.3 Modulação

2.3.1 Modulação de Amplitude

A multiplicação de um sinal $v(t)$ por uma onda sinusoidal dá origem a um novo sinal $v_m(t)$ cujo espectro é o de $v(t)$ transladado na frequência de um valor igual à frequência do sinal sinusoidal. Este resultado, que é a uma propriedade da transformada de Fourier conhecida por *Teorema da modulação*, é extremamente importante em telecomunicações pelo que passamos a demonstrá-lo.

Seja $v(t) \leftrightarrow V(f)$, isto é, $V(f) = \mathcal{F}[v(t)]$.

Calculemos o espectro de $v_m(t) = v(t) \cdot \cos(2\pi f_p t)$:

$$\begin{aligned} V_m(f) &= \int_{-\infty}^{+\infty} v_m(t) e^{-j2\pi f t} dt \\ V_m(f) &= \int_{-\infty}^{+\infty} v(t) \cdot \cos(2\pi f_p t) \cdot e^{-j2\pi f t} dt \end{aligned}$$

como $\cos(2\pi f_p t) = \frac{e^{j2\pi f_p t} + e^{-j2\pi f_p t}}{2}$ tem-se

$$\begin{aligned} V_m(f) &= \frac{1}{2} \int_{-\infty}^{+\infty} v(t) \cdot e^{j2\pi f_p t} \cdot e^{-j2\pi f t} dt + \\ &\quad + \frac{1}{2} \int_{-\infty}^{+\infty} v(t) \cdot e^{-j2\pi f_p t} \cdot e^{-j2\pi f t} dt \\ V_m(f) &= \frac{1}{2} \underbrace{\int_{-\infty}^{+\infty} v(t) e^{-j2\pi(f-f_p)t} dt}_{(1)} + \frac{1}{2} \underbrace{\int_{-\infty}^{+\infty} v(t) e^{-j2\pi(f+f_p)t} dt}_{(2)} \end{aligned}$$

O integral (1) é o espectro de $v(t)$ transladado na frequência e centrado em f_p e o integral (2) é o espectro de $v(t)$ centrado em $-f_p$, donde resulta sinteticamente

$$v(t) \cdot \cos(2\pi f_p t) \leftrightarrow \frac{1}{2} [V(f - f_p) + V(f + f_p)] \quad (2.23)$$

O sinal $\cos(2\pi f_p t)$ chama-se sinal portador ou *portadora* e f_p é o valor da frequência portadora. Importa notar que a modulação, que neste caso é uma *modulação de amplitude*, duplica a largura espectral do sinal pois a parte das frequências negativas do espectro passa para o lado das frequências positivas.

2.3.2 Modulação de Frequência

O sinal sinusoidal da figura 2.14 possui frequência constante, f_p , excepto no intervalo $-1/f_p < t < 1/f_p$ em que a frequência muda para $2f_p$. Este é um sinal típico de *frequência modulada*, designadamente, uma modulação digital de frequência de onda contínua, em que é a frequência da portadora que varia de acordo com as variações da amplitude do sinal modulante que, no caso da figura, é um sinal digital com tempo de símbolo igual a $\frac{2}{f_p}$ seg. Interessa-nos conhecer o seu espectro.

Para efeitos de análise seja $\tau = 2/f_p$ e decomponhamos o sinal $v(t)$ numa soma de três parcelas da seguinte maneira:

$$v(t) = A \cos(2\pi f_p t) - A \Pi\left(\frac{t}{\tau}\right) \cos(2\pi f_p t) + A \Pi\left(\frac{t}{\tau}\right) \cos(4\pi f_p t)$$

A primeira parcela representa um sinal cosseno de frequência f_p , a se-

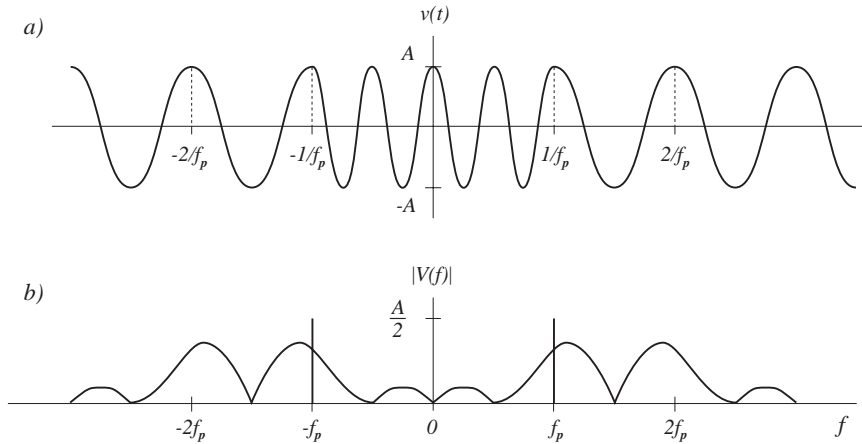


Figura 2.14: a) Sinal modulado em frequência e b) respectivo espectro

gunda retira-lhe a parte entre $-1/f_p$ e $+1/f_p$ e a terceira introduz neste intervalo um cosseno de frequência $2f_p$. Calculando o espectro (a trans-

formada de Fourier) de $v(t)$ parcela a parcela obtem-se

$$\begin{aligned} V(f) = & \mathcal{F}[A \cos(2\pi f_p t)] - \\ & - \frac{A\tau}{2} [\text{sinc}(f - f_p)\tau + \text{sinc}(f + f_p)\tau] + \\ & + \frac{A\tau}{2} [\text{sinc}(f - 2f_p)\tau + \text{sinc}(f + 2f_p)\tau] \end{aligned}$$

A primeira parcela é o espectro de uma sinusóide de amplitude A que é constituído por duas linhas de amplitude $A/2$ localizadas respectivamente em f_p e em $-f_p$. As duas parcelas seguintes, de acordo com o resultado da equação 2.23, são os espectros de dois pulsos rectangulares modulando em amplitude portadoras de frequências respectivamente f_p e $2f_p$.

A figura 2.14b) representa graficamente o espectro de amplitude desta modulação de frequência. Notar que $|V(f)|$ não é simétrica relativamente à frequência da portadora f_p .

Nesta figura pode verificar-se que o sinal modulado em frequência possui uma largura de banda que é *quatro vezes* maior do que a do sinal modulante (o sinal da informação).

2.4 Problemas

2.1 – A figura 2.15 representa dois sinais rectangulares bipolares periódicos, $v_1(t)$ e $v_2(t)$, onde cada rectângulo simboliza um dígito binário. O tempo do símbolo binário é de T_b seg em ambos os casos, a duração do rectângulo é, respectivamente, τ_1 s e τ_2 s e a sua amplitude é $\pm A$ Volt.

- Determine o espectro de cada um dos sinais.
- Represente graficamente os espectros de amplitude.
- Compare os espectros de amplitude obtidos em b) com o do sinal rectangular periódico unipolar da secção 2.1.4 quanto às respectivas componentes contínuas e harmónicas.

Sugestão: represente graficamente os espectros de amplitude dos três sinais na mesma escala de frequência. Considere $\tau = \tau_1$ e note que o período T_0 do sinal na figura 2.7 é diferente do destes

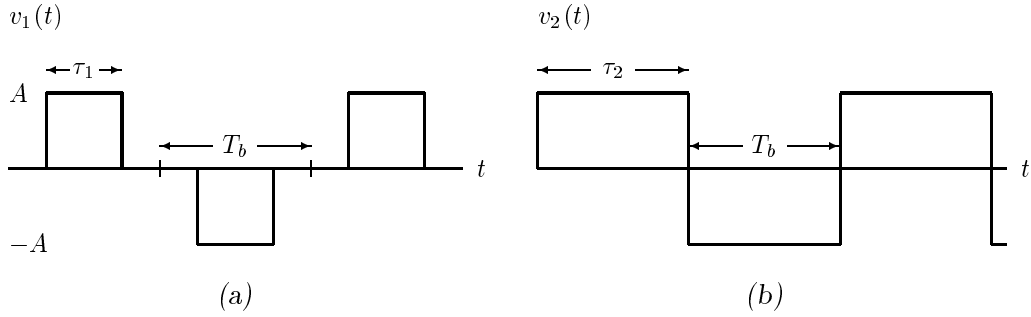


Figura 2.15: Sinais retangulares periódicos (a) RZ e (b) NRZ

2.2 – A figura 2.16 representa um sinal triangular bipolar periódico, $x(t)$, onde cada *pulso* triangular simboliza um dígito binário. O tempo de símbolo, T_b seg corresponde à duração do *pulso* que é de τ s.

- Determine uma expressão para o espectro deste sinal.
- Represente graficamente o espectro de amplitude, $|C_x(nf_0)|$ deste sinal. Indique no gráfico os valores das amplitudes.
- Compare o espectro de amplitude obtido com os dos sinais retangulares periódicos da figura 2.15 quanto às respectivas componentes contínuas e harmônicas.

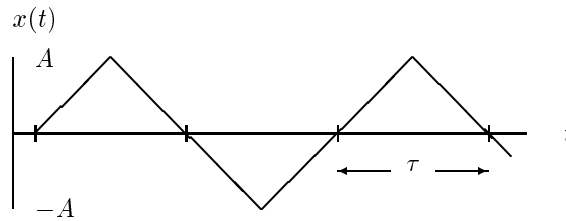


Figura 2.16: Sinal triangular bipolar periódico

2.3 – Represente graficamente as densidades espectrais de potência dos sinais das figuras 2.15 e 2.16 e determine a largura de banda a 90% da potência de cada um daqueles sinais.

2.4 – Considere o sinal periódico $v(t)$, de período T_0 seg, definido por

$$v(t) = \begin{cases} \sin(2\pi f_p t) & \text{para } 0 < t < \frac{T_0}{2} \\ -\sin(2\pi f_p t) & \text{para } -\frac{T_0}{2} < t < 0 \end{cases}$$

em que $f_p = \frac{2}{T_0}$.

- Esquematize graficamente três períodos deste sinal.
 - Determine uma expressão para o seu espectro, $V(f)$.
 - Discuta a possibilidade de utilização de formas de onda deste tipo para a codificação de linha de dígitos binários.
- 2.5 – A substituição de t por αt em $v(t)$ reflecte uma mudança da escala de tempos do sinal $v(t)$. O sinal $v(\alpha t)$ resultará *expandido* ou *comprimido* no tempo consoante $|\alpha| < 1$ ou $|\alpha| > 1$. Se α fôr negativo significa que se inverteu o sentido do tempo. Estas três situações correspondem, por exemplo, e respectivamente, à reprodução de uma gravação de audio ou de vídeo a uma velocidade mais baixa, mais elevada ou em sentido contrário. Esta mudança de escala no domínio do tempo implica uma *reciprocidade* no domínio da frequência definida por

$$v(\alpha t) \leftrightarrow \frac{1}{|\alpha|} V\left(\frac{f}{\alpha}\right) \quad \alpha \neq 0 \quad (2.24)$$

que é designada por lei do *espalhamento recíproco*. Utilize a definição da transformada de Fourier para mostrar a veracidade da relação 2.24.

2.6 – Determine o espectro do sinal

$$v(t) = \begin{cases} A \sin\left(\frac{6\pi t}{\tau}\right) & \text{para } |t| < \frac{\tau}{2} \\ 0 & \text{para } |t| \geq \frac{\tau}{2} \end{cases}$$

e represente graficamente o espectro de amplitude.

No gráfico, considere $\tau = \frac{1}{3} \times 10^{-4}$ s.

- 2.7 – Considere a modulação em frequência de uma portadora de frequência f_p , por uma sequência periódica de pulsos rectangulares de duração $\tau = \frac{4}{f_p}$ e de amplitude unitária com polaridade alternada. O desvio de frequência para cada polaridade é $0.5f_p$.

- a) Esboce a forma de onda de dois períodos do sinal modulado.
- b) Determine a densidade espectral de potência do sinal modulado e represente-a graficamente.
- c) O que conclui quanto à largura de banda deste sinal relativamente à que obteria por modulação de amplitude da mesma portadora.

fim do capítulo 2

Capítulo 3

Análise de Sistemas

3.1 Transmissão e Filtragem de Sinais

A *transmissão* de um sinal é o processo pelo qual uma forma de onda eléctrica transita de uma fonte para um destino, desejavelmente, sem sofrer alteração na sua forma (distorção). Por outro lado a *filtragem* de um sinal é uma operação que, propositadamente e a fim de atingir determinado objectivo, altera o espectro do sinal e, consequentemente, a sua forma, ou seja introduz no sinal uma distorção propositada. No entanto, tanto os *sistemas de transmissão* como os *filtros* são modelados de forma semelhante, normalmente por funções de relação entrada-saída. O sinal que se obtém à saída de um sistema como resultado da introdução de um determinado sinal na sua entrada designa-se por *resposta* do sistema a esse sinal de entrada.

Neste capítulo iremos caracterizar os *sistemas* em termos da função razão *resposta-entrada* designada por *função de transferência* do sistema utilizando-a depois para analisar os efeitos do sistema na transmissão ou filtragem dos sinais que por ele transitam.

3.2 Sistemas lineares e invariantes no tempo

A figura 3.1 representa um sistema sob a forma de uma *caixa preta* com um *sinal de entrada* $x(t)$ e um *sinal de saída* $y(t)$. No contexto das comunicações eléctricas, o sistema é normalmente um circuito com dois

pares de polos (ou portas) por exemplo uma linha de transmissão ou um cabo coaxial, um amplificador de potência de sinal, um filtro, etc ou uma combinação de elementos destes tipos. Na porta de entrada, é aplicado o sinal de entrada, de tensão ou de corrente, que dará origem a um sinal de saída também eléctrico, na porta de saída. Representaremos esta relação



Figura 3.1: Sistema visto como uma *caixa preta*

causal de entrada-resposta da seguinte forma

$$x(t) \rightarrow y(t)$$

Estamos interessados apenas numa classe importante de sistemas que é a dos sistemas *lineares e invariantes no tempo (LIT)*.

Um sistema é linear se, quando

$$\begin{aligned} x_1(t) &\rightarrow y_1(t) \\ x_2(t) &\rightarrow y_2(t) \end{aligned} \tag{3.1}$$

então

$$a x_1(t) + b x_2(t) \rightarrow a y_1(t) + b y_2(t) \tag{3.2}$$

em que a e b são constantes, independentes de t .

Um sistema é invariante no tempo se

$$x(t - t_1) \rightarrow y(t - t_1) \tag{3.3}$$

em que t_1 é uma constante. A propriedade de invariância no tempo significa que as características do sistema permanecem *fixas* ao longo do tempo, ou seja, um mesmo sinal de entrada aplicado à entrada do sistema t_1 segundos mais tarde produz a mesma saída também t_1 segundos mais tarde.

3.3 Função de transferência

Os elementos (componentes) eléctricos que constituem o sistema, quer sejam passivos quer sejam activos, farão com que o sinal à saída tenha uma forma diferente do da entrada. Independentemente do que estiver dentro da caixa preta, um sistema pode ser sempre caracterizado por uma relação *excitação-e-resposta* entre a entrada e a saída.

A questão fundamental que se coloca é a seguinte: Quais os $x(t)$ que passam pelo sistema sem alteração de forma? Isto é, quais os $x(t)$ tais que

$$y(t) = H \cdot x(t) \quad (3.4)$$

em que H é um escalar, independente de t . Os sinais $x_i(t)$ que satisfazem esta condição constituem as *funções próprias* ou *invariantes* do sistema. Vamos mostrar que se o sistema é LIT os invariantes são da forma

$$x_i(t) = e^{st} \quad (3.5)$$

em que s é uma constante, em geral complexa.

Suponhamos então que se introduz à entrada de um sistema LIT um sinal da forma indicada em (3.5) e procuremos determinar a forma da resposta correspondente $y_i(t)$, isto é,

$$x_i(t) = e^{st} \rightarrow y_i(t) \quad (3.6)$$

Se multiplicarmos o sinal $x_i(t)$ por uma constante e^{-st_1} , obter-se-á pela propriedade linear

$$e^{-st_1} \cdot x_i(t) = e^{s(t-t_1)} \rightarrow e^{-st_1} \cdot y_i(t) \quad (3.7)$$

e se aplicarmos $x_i(t)$ à entrada t_1 segundos mais tarde obter-se-á pela propriedade de invariancia no tempo

$$x_i(t - t_1) = e^{s(t-t_1)} \rightarrow y_i(t - t_1) \quad (3.8)$$

donde se conclui, de (3.7) e (3.8) que

$$e^{-st_1} \cdot y_i(t) = y_i(t - t_1) \quad (3.9)$$

cuja única solução é

$$y_i(t) = H \cdot e^{st} \quad (3.10)$$

com H independente de t mas eventualmente dependente da constante complexa $s = a + jb$. O resultado (3.10) mostra que tem de ser $x_i(t) = e^{st}$. Se na constante complexa se fizer $a = 0$ e $b = -2\pi f$, a família de sinais $x_i(t)$ pode ser escrita sob a forma

$$\begin{aligned} x_i(t) &= e^{-j2\pi ft} && \text{ou} \\ x_i(t) &= \cos(2\pi ft) - j \sin(2\pi ft) \end{aligned} \quad (3.11)$$

o que significa que as exponenciais complexas, ou seja, os sinais oscilatórios no tempo são invariantes do sistema LIT e *passam* através dele sem alteração de forma a menos de um factor multiplicativo constante H (que actua apenas na amplitude da oscilação). Embora independente de t , H é função de s ou seja, uma função da frequência $H = H(f)$ e designa-se *função de transferência* do sistema.

Portanto qualquer sinal vê cada uma das suas componentes espectrais *passar* no sistema sem alteração de forma mas com alteração de amplitude consoante a frequência. Isto é, se $x(t) \rightarrow y(t)$ a relação entre os respectivos espectros é

$$Y(f) = H(f) \cdot X(f) \quad (3.12)$$

dado que os espectros não são mais do que combinações lineares de funções circulares — as harmónicas dos sinais —, que são invariantes do sistema. Esta relação extremamente simples constitui a base da análise de sistemas no domínio da frequência. Diz-nos que o espectro do sinal de saída $Y(f)$ se obtém multiplicando o espectro do sinal de entrada $X(f)$ pela função de transferência $H(f)$ e é esta noção que a figura 3.2 representa. A função

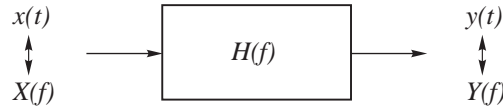


Figura 3.2: Resposta em frequência de um sistema LIT

$H(f)$ é também designada de *resposta em frequência* do sistema. Neste sentido, trata-se de uma resposta natural e não de uma resposta forçada, isto é, trata-se de uma característica intrínseca do próprio sistema. O módulo $|H(f)|$ representa a característica de amplitude do sistema e o

argumento a sua característica de fase e tem-se

$$|Y(f)| = |H(f)| \cdot |X(f)| \quad (3.13)$$

$$\arg Y(f) = \arg H(f) + \arg X(f) \quad (3.14)$$

Se $x(t)$ fôr um sinal de energia, então $y(t)$ será também um sinal de energia cuja densidade espectral de energia e energia total são dadas respectivamente por

$$\begin{aligned} |Y(f)|^2 &= |H(f)|^2 \cdot |X(f)|^2 \\ E_y &= \int_{-\infty}^{\infty} |H(f)|^2 |X(f)|^2 df \end{aligned} \quad (3.15)$$

3.4 Sistemas de primeira ordem

Alguns sistemas simples podem ser modelados, sob o ponto de vista entrada-saída, por um circuito eléctrico RC. É o caso, por exemplo, de uma linha de transmissão constituída por um par de fios condutores. Os condutores possuem uma resistência eléctrica distribuída ao longo do seu comprimento a qual pode ser modelada por uma resistência concentrada de valor R Ohms colocada em série entre a entrada e a saída. Pelo facto de se encontrarem próximos, separados ou pelo ar ou por um outro meio isolante, existe entre os condutores uma certa capacidade eléctrica, também distribuída, que se pode modelar por um condensador (concentrado) colocado em paralelo entre a entrada e a saída, de valor igual à capacidade total, C Farads. A figura 3.3 ilustra este modelo, ou seja, trata-se de um circuito eléctrico RC, passivo, sob a forma de uma malha com duas portas e com uma tensão eléctrica de entrada $x(t)$ e uma tensão eléctrica de saída, $y(t)$. Procuremos a equação que rege o funcionamento

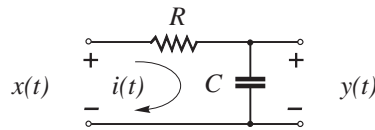


Figura 3.3: Modelo eléctrico de um sistema de primeira ordem

deste circuito. A tensão de entrada provocará uma corrente $i(t)$ na malha,

como se indica. Tomando em consideração as polaridades de referência indicadas, a tensão $x(t)$ será dada pela soma da queda de tensão na resistência com a queda de tensão no condensador, esta última sendo precisamente o valor de $y(t)$. A equação da malha é então

$$R i(t) + y(t) = x(t) \quad (3.16)$$

em que a tensão $y(t)$ é o produto da corrente pela impedância complexa do condensador, Z_C , se $y(t)$ e $i(t)$ forem invariantes do sistema

$$\begin{aligned} y_i(t) &= Z_C i_i(t) \\ y_i(t) &= \frac{1}{j2\pi f C} i_i(t) \quad \text{donde} \\ i_i(t) &= j2\pi f C y_i(t) \end{aligned} \quad (3.17)$$

Substituindo o valor de $i_i(t)$ dado em (3.17) na equação (3.16) obtém-se

$$j2\pi f R C y_i(t) + y_i(t) = x_i(t)$$

donde resulta imediatamente a expressão para a função de transferência do sistema modelado por este circuito

$$\frac{y_i(t)}{x_i(t)} = H(f) = \frac{1}{1 + j2\pi f R C} \quad (3.18)$$

A *característica de amplitude*, também designada por *resposta de ampli-*

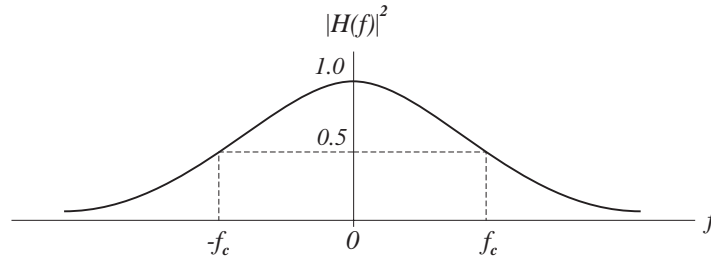


Figura 3.4: Característica de potência de um sistema de primeira ordem

tude, do sistema, é dada pelo módulo $|H(f)|$. Mas a característica mais significativa é a que relaciona as densidades de energia (ou as potências) dos sinais de entrada e saída, isto é $|H(f)|^2$, como mostra a equação 3.15

$$|H(f)|^2 = \frac{1}{1 + (2\pi f R C)^2} \quad (3.19)$$

a qual tem uma representação gráfica que se mostra na figura 3.4

3.5 Largura de banda de transmissão

Na figura 3.4 indica-se o valor da frequência, f_c , a partir da qual $|H(f)|^2$ vem reduzida de mais de $\frac{1}{2}$ do seu valor máximo.

No caso concreto deste sistema, todas as componentes de um sinal de entrada, de frequência superior a f_c vêm a sua energia (ou potência) *atenuadas* de um factor de valor superior a $\frac{1}{2}$ e todas as componentes de frequência inferior a f_c são atenuadas de um factor de valor inferior àquele.

A frequências f_c designam-se frequências de *corte* do sistema. Elas definem os limites da banda de transmissão, que é a banda útil do sistema.

Definição 3.1 — Banda de transmissão de um sistema

É o intervalo de frequências positivas no qual o sistema de transmissão não atenua a energia (ou a potência) das componentes de um sinal de entrada em mais do que $\frac{1}{2}$ da atenuação mínima desse sistema.

A banda de transmissão de um sistema também é designada de *banda passante* do sistema (ver figura 3.12).

Definição 3.2 — Largura de Banda de um sistema

É a amplitude da banda de transmissão desse sistema.

Para o caso do sistema de primeira ordem, caracterizado pela relação 3.18, as frequências de corte são dadas pelas raízes da equação

$$|H(f)|^2 = \frac{1}{2} |H(0)|^2$$

ou seja $\frac{1}{1 + (2\pi f_c RC)^2} = \frac{1}{2}$ donde $f_c = \pm \frac{1}{2\pi RC}$

pelo que, de acordo com a definição, a largura de banda de transmissão deste sistema será $B_T = |f_c - 0|$, ou seja

$$B_T = \frac{1}{2\pi RC} \quad (3.20)$$

A função de transferência do sistema de primeira ordem representado pela equação 3.18 pode então escrever-se da seguinte forma

$$H(f) = \frac{1}{1 + j \frac{f}{B_T}} \quad (3.21)$$

3.6 Análise com diagramas de blocos

Um sistema de comunicação é normalmente constituído por vários sub-sistemas componentes. Cada sub-sistema (LIT) possuirá uma função de transferência e a função de transferência do sistema total será uma composição destas. A tabela 3.1 mostra algumas funções de transferência *primitivas* que se podem deduzir aplicando a definição da transformada referida no capítulo 2. Sempre que os sub-sistemas em questão sejam

Tabela 3.1: Funções de Transferência de algumas operações temporais

Operação no domínio do tempo	F de Transferência
Multiplicação por Escalar	$y(t) = \pm K x(t)$ $H(f) = \pm K$
Diferenciação	$y(t) = \frac{dx(t)}{dt}$ $H(f) = j 2\pi f$
Integração	$y(t) = \int_{-\infty}^t x(\lambda) d\lambda$ $H(f) = \frac{1}{j 2\pi f}$
Atraso no Tempo	$y(t) = x(t - t_a)$ $H(f) = e^{-j 2\pi f t_a}$

descritos por funções de transferência individuais é possível e desejável obter a função de transferência do sistema global por composição daquelas. A figura 3.5 mostra as formas básicas de composição de sub-sistemas designadamente as composições paralela, em cascata e realimentada. As configurações mais complicadas podem ser analisadas por aplicação sucessiva destas regras elementares. É, no entanto, essencial, que nas funções de transferência dos sub-sistemas se tenham em consideração quaisquer interações ou efeitos de *carga* (eléctrica) entre eles de modo a representarem o comportamento do sub-sistema no contexto do sistema total. É fácil de concluir que as regras de composição se traduzem nas seguintes relações

$$\begin{array}{ll}
 H(f) = H_1(f) + H_2(f) & \text{Ligação em Paralelo} \\
 H(f) = H_1(f) \cdot H_2(f) & \text{Ligação em Cascata} \\
 H(f) = \frac{H_1(f)}{1 + H_1(f) \cdot H_2(f)} & \text{Ligação Realimentada}
 \end{array}$$

Este último caso é mais propriamente designado por realimentação *negativa* em contraste com a realimentação positiva na qual o sinal de saída é adicionado ao de entrada.

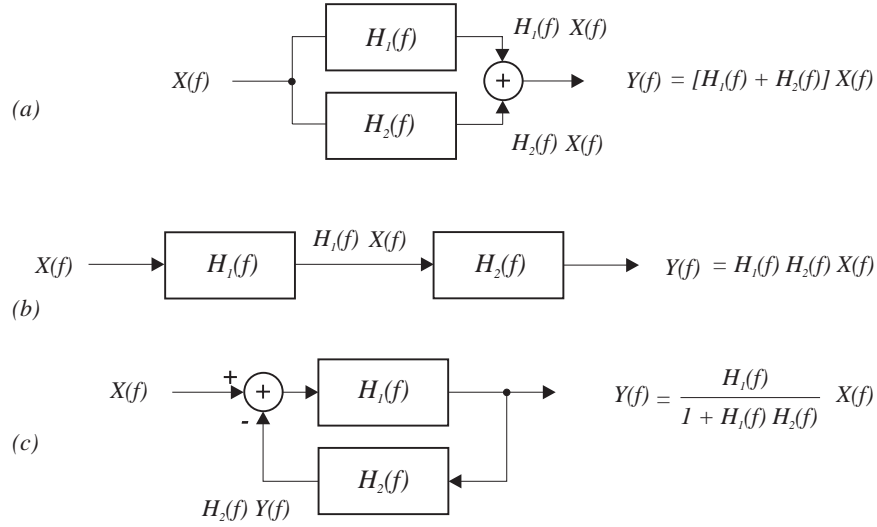


Figura 3.5: Composição: (a) Paralela; (b) em cascata; (c) realimentada

3.7 Distorção do sinal na transmissão

Um sistema de transmissão de sinal é o canal eléctrico entre uma fonte de informação e um destino. Os vários sistemas existentes possuem complexidade variável, desde um simples par de fios até aos sofisticados *links* ópticos de *laser*. Mas todos eles possuem dois atributos físicos particularmente relevantes em comunicações: a dissipação interna de potência, cujo efeito é o de reduzir o sinal de saída, e o armazenamento interno de energia, que é variável para cada frequência – precisamente o que a função de transferência caracteriza – cujo efeito é o de alterar a forma do sinal de saída.

Formularemos pois as condições para a transmissão do sinal sem distorção em sistemas LIT de modo a podermos lidar com as respectivas funções de transferência. Referiremos ainda alguns dos tipos de distorção em sistemas de transmissão.

3.7.1 Transmissão isenta de distorção

Uma transmissão isenta de distorção é aquela em que o sinal de saída apresenta a mesma *forma* que o sinal de entrada¹. Mais precisamente, dado um sinal de entrada $x(t)$, diremos que a resposta não se apresenta distorcida se só diferir da entrada por uma constante multiplicativa K e por um tempo de atraso constante e finito t_a o que, analiticamente, se exprime pela relação

$$y(t) = K x(t - t_a)$$

As propriedades de um sistema não distorativo podem ser deduzidas a partir do espectro da resposta

$$Y(f) = \mathcal{F}[y(t)] = K X(f) e^{-j2\pi t_a f}$$

como, por definição de função de transferência se tem $Y(f) = H(f) \cdot X(f)$, será

$$H(f) = K e^{-j2\pi t_a f} \quad (3.22)$$

por outras palavras, um sistema que transmita sem distorção deve possuir uma *característica de amplitude constante* e uma *característica de fase linear* de declive negativo

$$|H(f)| = |K| \quad \text{e} \quad \arg H(f) = -2\pi t_a f \quad (3.23)$$

Óbviamente, as condições representadas pelas equações 3.22 e 3.23 apenas se têm de verificar na banda de frequências onde o sinal possui conteúdo espectral significativo. De modo a clarificar este ponto, repare-se na figura 3.6 que mostra a densidade espectral de energia de um sinal típico de voz obtida experimentalmente. Dado que a densidade espectral é bastante pequena para $f < 200$ Hz ou $f > 4200$ Hz pode concluir-se que um sistema que satisfaça as equações 3.22 e 3.23 no intervalo $200 \leq |f| \leq 4200$ Hz transmitirá voz virtualmente sem distorção. Definem-se três tipos principais de distorção:

- (i) *Distorção de Amplitude*, que ocorre quando $|H(f)| \neq |K|$
- (ii) *Distorção de Atraso*, que ocorre quando $\arg H(f) \neq -2\pi t_a f$
- (iii) *Distorção não-linear*, que ocorre quando o sistema possui componentes não-lineares.

Os dois primeiros agrupam-se sob a designação geral de distorção *linear*.

¹qualquer que este seja e não apenas os invariantes

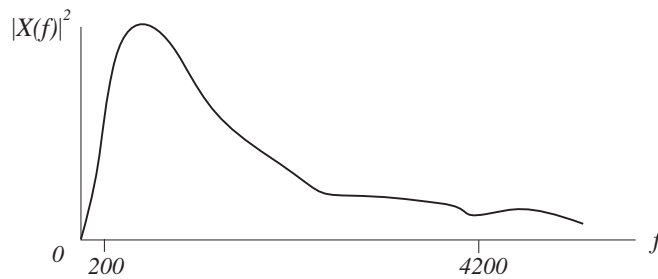


Figura 3.6: Densidade espectral de energia típica de um sinal de voz

3.7.2 Distorção Linear e Equalização

A distorção de amplitude é facilmente descrita no domínio da frequência. Uma distorção de amplitude significa que as amplitudes das componentes de frequência do sinal à saída não estão na proporção correcta, isto é, umas são mais atenuadas que outras. Dado que este fenómeno se deve a $|H(f)|$ não ser constante ao longo da frequência, a distorção de amplitude é algumas vezes designada de *distorção de frequência*. As formas mais comuns de distorção de amplitude são a atenuação ou a amplificação excessivas das componentes nos limites das baixas ou das altas frequências do espectro do sinal.

Na distorção de atraso as componentes de frequência sofrem atrasos não proporcionais à sua frequência. Sem haver distorção nas suas amplitudes, o efeito no sinal de saída é semelhante, isto é, a forma do sinal vem alterada, manifestando-se assim de forma semelhante à distorção de amplitude. Este efeito pode ser facilmente verificado se se sintetizar o sinal a partir das suas componentes tomando uma delas com um desvio de fase de, por exemplo, 90° . Facto interessante é o de o ouvido humano só ser sensível à distorção de amplitude e não à de atraso pelo que esta última não necessita de ser corrigida nos sistemas de áudio. Já nos sistemas de transmissão de dados ambas são importantes e necessitam ser corrigidas ou minimizadas visto que a correcta detecção do sinal de dados depende não só da sua forma como de uma correcta localização no tempo dos elementos do sinal digital.

A distorção linear – tanto a de amplitude como a de atraso – é teóricamente corrigível através de *equalizadores*.

A figura 3.7 esquematiza um equalizador $H_{eq}(f)$ em série (em cascata) com um canal distortivo $H_C(f)$. Dado que a função de transferência

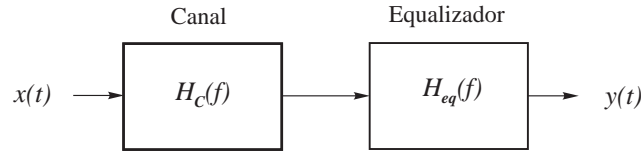


Figura 3.7: Canal terminando num equalizador

total é $H(f) = H_C(f) H_{eq}(f)$ a saída final estará isenta de distorção se $H_C(f) H_{eq}(f) = K e^{-j2\pi t_a f}$, onde K e t_a são constantes. Portanto, para corrigir a distorção linear, é necessário um equalizador com função de transferência

$$H_{eq}(f) = \frac{K e^{-j2\pi t_a f}}{H_C(f)} \quad (3.24)$$

para valores de f tais que $X(f) \neq 0$.

Rara é a situação em que é possível obter (ou projectar) um equalizador que satisfaça exactamente a equação 3.24 razão pela qual se disse que a equalização é uma correcção *teórica*. A técnica de equalização mais antiga é provavelmente aquela em que eram utilizadas *bobinas de carga* em linhas telefónicas. As bobinas constituíam indutores concentrados e eram colocados em série ao longo da linha, aproximadamente de quilómetro em quilómetro e o seu efeito era o de melhorar a característica de amplitude da linha como se mostra na figura 3.8 para uma linha carregada e não-carregada com bobinas (ver problema 3.7).

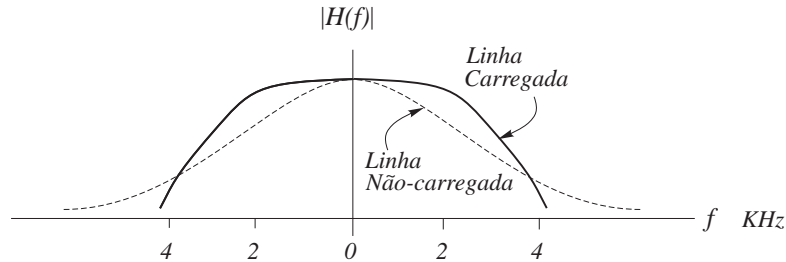


Figura 3.8: Característica de amplitude típica de uma linha telefónica

3.8 Perdas de Transmissão e Decibéis

Para além de distorcer o sinal, os sistemas de transmissão normalmente também reduzem a potência do sinal, ou seja, introduzem uma *perda de transmissão*. Embora esta perda possa ser compensada por amplificação de potência, o ruído, sempre presente ao longo do sistema, é também amplificado o que pode obviar à recuperação do sinal original. Começaremos por rever o conceito de *ganho* de potência e seguidamente o de *decibel* como medida de razão de potências.

3.8.1 Ganho de Potência

Suponhamos que a figura 3.9 representa um sistema LIT cujo sinal de entrada possui potência média P_e . Se o sistema não introduzir distorção a potência média do sinal à saída, P_s , será proporcional a P_e . O *ganho de potência* do sistema é então definido por

$$g \doteq \frac{P_s}{P_e} \quad (3.25)$$

que é um parâmetro constante do sistema. Os sistemas amplificadores

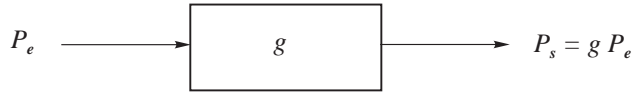


Figura 3.9: Sistema LIT com ganho de potência g .

normalmente possuem valores de g muito elevados pelo que tem sido prática corrente a utilização de uma escala logarítmica para representar e manipular esses valores. Assim, o ganho g em *decibéis* (dB) é definido por

$$g_{dB} \doteq 10 \log_{10} g \quad (3.26)$$

em que a letra B em dB se refere ao Bell em memória de Alexander Graham Bell que usou esta escala logarítmica pela primeira vez na medida de potências². As potências de 10 são pois convertidas em produtos por

²A unidade Bell, B, é uma ordem de grandeza superior à unidade decibel, dB, ié, seria $g_B \doteq \log_{10} g$

10, por exemplo $g = 10^m$ passa a ser $g_{dB} = m \times 10$ dB. O ganho de potência é sempre positivo mas podem ocorrer valores negativos em dB quando $0 < g \leq 1.0 = 10^0$ dando $-\infty < g_{dB} \leq 0$ dB. Notar que 0 dB corresponde a um ganho unitário, $g = 1$. Dado um valor em dB, o valor *linear* correspondente é

$$g = 10^{\frac{g_{dB}}{10}} \quad (3.27)$$

o que se obtém por inversão da equação 3.26. Enquanto que os decibéis representam sempre razões de potência, a própria potência de um sinal pode ser expressa em dB se se considerar relativa a uma potência fixa de referência e portanto dividida por essa referência. A potência de referência utilizada em telecomunicações para expressar potências em dB é o miliwatt e a correspondente *unidade* é designada por dBm:

$$P_{dBm} = 10 \log_{10} \frac{P}{1 \text{ mW}} \quad (3.28)$$

Assim, da equação 3.25 resulta

$$P_{s_{dBm}} = g_{dB} + P_{e_{dBm}} \quad (3.29)$$

que é uma forma bastante mais prática para os cálculos de balanços de potências já que as multiplicações e divisões se transformam em somas e subtracções. Os engenheiros de comunicações trabalham normalmente em dBm dado que as potências dos sinais à saída dos sistemas de transmissão são bastantes pequenas.

Consideremos agora um sistema descrito pela sua função de transferência $H(f)$. A uma entrada sinusoidal de amplitude A_x corresponde uma saída de amplitude $A_y = |H(f)| A_x$ sendo as potências normalizadas desses sinais $S_x = A_x^2/2$ e $S_y = A_y^2/2 = |H(f)|^2 S_x$. A razão das potências normalizadas (S_x e S_y) não é necessariamente igual à razão das potências *reais* (P_x e P_y) dos sinais da equação 3.25 excepto se o sistema (*real*) tiver a mesma impedância na entrada e na saída. Se fôr este o caso, então

$$g = |H(f)|^2 = K^2 \quad (3.30)$$

Se as impedâncias não forem iguais o ganho será proporcional a K^2 . O ganho também se aplica a sinais de energia, isto é, $E_y = g E_x$. Se o sistema fôr selectivo em frequência a equação 3.30 não é aplicável, no entanto, $|H(f)|^2$ ainda diz como é que o ganho varia em função da frequência. Nestas condições considera-se $|H(f)|_{dB}^2$ como sendo o *ganho relativo* do sistema

$$|H(f)|_{dB}^2 \doteq 10 \log_{10} |H(f)|^2 \quad (3.31)$$

Exercício 3.1

- a) Verificar que $P_{dBm} = P_{dBW} + 30 \text{ dB}$. (dBW lê-se dB -Watt).
 b) Mostrar que se $|H(f_c)|_{dB}^2 = -3 \text{ dB}$ então $|H(f_c)| \approx 1/\sqrt{2}$.

3.8.2 Perdas de Transmissão e Repetidores

Todo o meio de transmissão passivo envolve uma perda de potência, em lugar de ganho, dado que $P_s < P_e$. É preferível, portanto, trabalhar em termos de *perda* ou *atenuação* de transmissão que se designa pela letra L .³

$$L \doteq \frac{1}{g} = \frac{P_e}{P_s}$$

$$L_{dB} = -g_{dB} = 10 \log_{10} \frac{P_e}{P_s} \quad (3.32)$$

ou seja, $P_s = P_e/L$ e $P_{s_{dBm}} = P_{e_{dBm}} - L_{dB}$. No caso das linhas de transmissão, dos cabos coaxiais e dos guias de onda, a potência de saída decresce exponencialmente com a distância, relação que toma a seguinte forma

$$P_s = 10^{-\frac{\alpha d}{10}} P_e \quad (3.33)$$

onde d é a *distância* percorrida pelo sinal entre a entrada e a saída do meio de transmissão e α é o *coeficiente de atenuação*, em dB por unidade de comprimento. Da equação 3.32 resulta então que

$$L = 10^{\frac{\alpha d}{10}} \quad \text{e} \quad L_{dB} = \alpha d \quad (3.34)$$

o que mostra que a perda em dB é proporcional ao comprimento. A tabela 3.2 lista alguns valores típicos de α para vários meios de transmissão e frequências do sinal. É de notar que os valores da atenuação, quando expressos em dB, obscurecem um pouco a diminuição dramática com a distância que se verifica na potência do sinal, como se pode constatar no seguinte exemplo.

Exemplo 3.1

Suponhamos a transmissão de um sinal através de um cabo de 30 Km de comprimento com $\alpha = 3 \text{ dB/Km}$. A perda em dB é $L_{dB} = 3 \times 30 = 90 \text{ dB}$, logo $L = 10^9$ e $P_s = 10^{-9} P_e$. Duplicando a distância duplica a atenuação

³ L de Loss, em inglês, que significa Perda.

Tabela 3.2: Valores típicos de perdas de transmissão

Meio de transmissão	Frequência	Perda (dB/Km)
Par aberto de fios (0.3 cm \varnothing)	1 kHz	0.005
Par entrançado de fios (16 AWG)	10 kHz	2
	100 kHz	3
	300 kHz	6
	100 kHz	1
Cabo coaxial (1 cm \varnothing)	1 MHz	2
	3 MHz	4
	100 MHz	1.5
Cabo coaxial (15 cm \varnothing)	100 MHz	1.5
Guia de onda rect ^{lar} (5 \times 2.5 cm)	10 GHz	5
Fibra óptica	4×10^{14} Hz	10

total para 180 dB, de modo que $L = 10^{18}$ e $P_s = 10^{-18} P_e$. Esta perda é tão grande que seria necessária uma potência de entrada de um megawatt (10^6 W) para obter uma potência de saída de um picowatt (10^{-12} W)!

Grandes atenuações exigem *amplificação* que é feita ao longo do percurso através de *amplificadores repetidores*. A figura 3.10 exemplifica um sistema de transmissão por cabo com um amplificador de saída e um amplificador *repetidor* a cerca do meio do percurso. Dado que numa ligação em cascata os ganhos de potência se multiplicam, tem-se

$$P_s = (g_1 g_2 g_3 g_4) P_e = \frac{g_2 g_4}{L_1 L_3} P_e \quad (3.35)$$

que se transforma na equação em dB

$$P_{s_{dBm}} = (g_{2_{dB}} + g_{4_{dB}}) - (L_{1_{dB}} + L_{3_{dB}}) + P_{e_{dBm}} \quad (3.36)$$

Os repetidores são colocados em locais do trajecto por forma a garantir, por um lado, um nível mínimo para a potência do sinal à entrada do amplificador de saída, que este necessita face à potência do ruído presente – valor mínimo esse que é ditado pelas características do amplificador –, e por outro lado, que a potência do sinal em nenhum ponto do cabo ultrapassa um determinado valor limite máximo admissível no cabo, valor este que é ditado pelas suas características eléctricas.

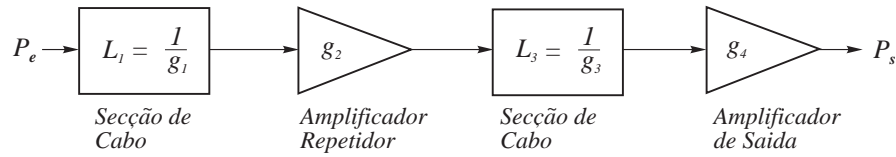


Figura 3.10: Sistema de transmissão por cabo com um amplificador repetidor

Exercício 3.2 Num sistema de transmissão por cabo, de 40 Km de comprimento, a potência de entrada é de $P_e = 2$ W, existindo um repetidor de 64 dB de ganho, a 24 Km da entrada. O cabo possui um coeficiente de atenuação $\alpha = 2.5$ dB/Km. Utilizar as equações de decibéis para determinar a potência do sinal:

- à entrada do repetidor;
- na saída final.

3.9 Filtros

Qualquer sistema de comunicação inclui um ou mais filtros com o objetivo de separar o sinal portador de informação de contaminações indesejáveis tais como a interferência, o ruído e as distorções.

Como se referiu no início deste capítulo, os filtros são modelados e comportam-se de maneira semelhante aos sistemas de transmissão, diferindo a sua designação apenas pela sua finalidade. O que se disser para os filtros é pois válido para qualquer sistema de transmissão.

3.9.1 Filtros ideais

Por definição, um filtro ideal é caracterizado por fornecer uma transmissão isenta de distorção em uma ou mais bandas de frequência bem determinadas e resposta nula nas restantes frequências. Em particular, a função de transferência de um filtro *passa-banda* ideal é

$$H(f) = \begin{cases} K e^{-j2\pi f t_a} & f_i \leq |f| \leq f_s \\ 0 & |f| > f_s \vee |f| < f_i \end{cases} \quad (3.37)$$

cujas características de amplitude está representada na figura 3.11. Os parâmetros f_i e f_s são as *frequências de corte*, respectivamente inferior e

superior, dado que eles delimitam a *banda passante*. A *largura de banda* do filtro é

$$B_T = f_s - f_i$$

a qual é medida na porção de frequências positivas da banda passante.

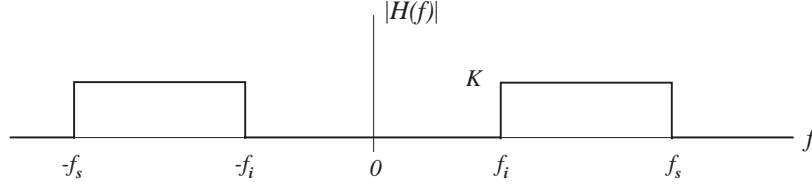


Figura 3.11: Característica de amplitude de um filtro passa-banda ideal

De igual forma, um filtro *passa-baixo* ideal é definido pela equação 3.37 com $f_i = 0$ o qual terá, portanto, $B_T = f_s$. E um filtro *passa-alto* ideal tem $f_i > 0$ e $f_s = \infty$.

Um filtro *rejeita-banda* permite transmissão isenta de distorção a todas as frequências excepto às da *banda de rejeição*, $f_i \leq |f| \leq f_s$, onde se tem $|H(f)| = 0$.

Interessa frizar que os filtros ideais são irrealizáveis, no sentido em que não é possível na prática obter transições abruptas (descontinuidades) da característica nas frequências limites da banda. Os filtros *reais*, portanto, diferem dos ideais nesta zona de transição.

3.9.2 Filtros Reais

A figura 3.12 mostra a característica de amplitude de um filtro passa-banda típico. Comparada com a do filtro ideal da figura 3.11 verifica-se uma banda passante onde $|H(f)|$ tem um valor relativamente elevado, mas não-constante, e bandas de rejeição onde $|H(f)|$ é bastante pequeno mas não nulo. Os pontos limites da banda passante são usualmente definidos por valores f_i e f_s tais que

$$|H(f_i)| = |H(f_s)| = \frac{1}{\sqrt{2}} |H(f)|_{max} = \frac{K}{\sqrt{2}} \quad (3.38)$$

de tal modo que $|H(f)|^2$ nunca seja inferior a $K^2/2$ para $f_i \leq |f| \leq f_s$. A largura de banda, $B_T = |f_s - f_i|$ é então designada de *largura de banda de*

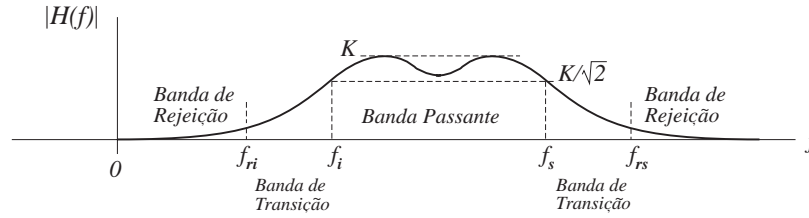


Figura 3.12: Característica de amplitude típica de um filtro passa-banda real

meia potência ou *largura de banda a 3 dB*. A largura da banda passante de um filtro é definida precisamente da mesma maneira que a largura da banda de transmissão a que se aludiu na secção 3.5, onde o sistema de primeira ordem ali exemplificado é um sistema passa-baixo que possui $f_i = 0$ e $f_s = f_c$ por frequências de corte. Por esta razão se está a usar B_T para designar ambas as larguras de banda.

De igual modo os limites das bandas de rejeição são definidos pelas frequências f_{ri} e f_{rs} a partir das quais $|H(f)|^2$ se mantém consistentemente abaixo de 10 % do seu valor máximo.

Entre a banda passante e as bandas de rejeição existem as *bandas de transição*, ou *bandas de corte*, onde se pode considerar que o filtro (ou o sistema de transmissão) nem deixa passar nem rejeita as correspondentes componentes de frequência do sinal. Estas bandas são definidas pelas frequências de corte, $f_c = \{f_i, f_s\}$, e de rejeição, $f_r = \{f_{ri}, f_{rs}\}$.

Assim, uma filtragem (ou transmissão) efectiva do sinal depende da existência de filtros com bandas de transição muito estreitas. Tais filtros (ou sistemas) são de ordem superior (ao de primeira ordem referido na secção 3.4). Faremos referência a uma classe conhecida por filtros (ou sistemas) de *Butterworth* de ordem- n .

O filtro de *Butterworth* de ordem- n é constituído por n elementos reactivos (condensadores e indutores). A sua característica de amplitude com $K = 1$ é da forma

$$|H(f)| = \frac{1}{\sqrt{1 + \left(\frac{f}{B_T}\right)^{2n}}} \quad (3.39)$$

Um filtro de Butterworth de primeira ordem ($n = 1$) possui a mesma característica do filtro passa-baixo RC exemplificado na figura 3.3 o qual constituiria uma fraca aproximação ao filtro ideal passa-baixo. Mas a

aproximação melhora se se aumentar o valor de n por adição de mais elementos ao circuito (ver Problema 3.7). A figura 3.13 mostra algumas características de potência para vários valores de n .

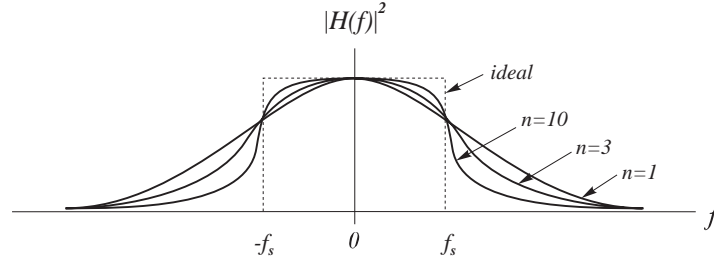


Figura 3.13: Características de filtros de Butterworth

Tendo o limite da banda de rejeição sido definido pelas frequências f_r tais que $|H(f_r)|^2 = -10$ dB, a largura da banda de transição, quando $n = 1$, é de $3 B_T - B_T = 2 B_T$, mas quando $n = 10$, ela tem somente o valor de $1.12 B_T - B_T = 0.12 B_T$. Claramente se verifica que $|H(f)|^2$ se aproxima da característica quadrada ideal, no limite, quando $n \rightarrow \infty$.

3.10 Resposta temporal de sistemas

Na secção 3.4, a partir de um modelo eléctrico simplificado de um sistema passa-baixo de 1^a ordem que é o circuito RC da figura 3.3 descrito pela equação 3.16, deduziu-se a relação entre a entrada e a saída indicada na equação 3.18. Pelo facto de na equação 3.16 se ter eliminado a corrente $i(t)$ substituindo-a por uma função de $y_i(t)$ e da impedância do condensador, esta última, por sua vez, uma função da frequência, obteve-se uma relação entrada-saída dependente apenas da frequência que é a função de transferência do sistema, $H(f)$. Esta função relaciona os espectros dos sinais de entrada e de saída, ou seja, as suas componentes harmónicas (os invariantes do sistema) mas não relaciona directamente os próprios sinais como funções do tempo, $x(t)$ e $y(t)$.

Nesta secção iremos obter relações deste último tipo, isto é, relações entre os sinais de entrada e de saída no domínio do tempo. Tentaremos assim verificar directamente qual a *forma* do sinal que se obtém à saída de um sistema como resposta a um sinal de entrada com determinada *forma*.

Esta análise, para o caso geral, é um tanto complicada de modo que restringir-nos-emos apenas ao caso de sinais digitais (discretos) e de sistemas de 1^a ordem.

3.10.1 Resposta ao escalão

Um sinal digital consiste numa sequência de escalões positivos e negativos de tensão entre amplitudes discretas (retângulos) que representam níveis lógicos e que se repetem a um ritmo r_s igual ao inverso do tempo de duração de cada amplitude (o nível lógico ou símbolo digital) que temos vindo a designar por τ . Dado que em geral se faz com que essa amplitude dure todo o tempo reservado ao símbolo, isto é, $\tau = T_s$, pode escrever-se $r_s = 1/T_s = 1/\tau$.

Determinemos então a resposta do sistema de 1^a ordem da figura 3.3 a um escalão de amplitude unitária, ou seja, ao sinal de entrada definido por

$$x(t) = \begin{cases} 0 & \text{para } t \leq 0 \\ 1 & \text{para } t > 0 \end{cases} \quad (3.40)$$

Partamos novamente da equação 3.16 mas agora eliminemos $i(t)$ substituindo-a por uma relação temporal⁴. Para tal basta considerar que a diferença de potencial instantânea aos terminais de um condensador, $y(t)$, é dada pelo quociente entre a carga instantânea, $q(t)$, e o valor da capacidade, C , sendo a corrente que nele passa, $i(t)$, definida pela variação da carga instantânea aos terminais do condensador. Teremos portanto

$$y(t) = \frac{q(t)}{C} \quad \text{e} \quad i(t) = \frac{dq(t)}{dt} \quad \text{donde} \quad i(t) = C \frac{dy(t)}{dt}$$

substituindo este valor de $i(t)$ na equação 3.16 obtem-se a equação que procuramos

$$RC \frac{dy(t)}{dt} + y(t) = x(t) \quad (3.41)$$

que é uma equação diferencial linear de primeira ordem, razão pela qual o sistema (circuito) da figura 3.3 se designa de sistema de primeira ordem.

Com $x(t)$ definido pela equação 3.40, uma condição de fronteira óbvia é ter-se $y(t) = 0$ para $t \leq 0$, ou seja, dizer-se que a saída é nula antes do escalão de entrada se manifestar.

⁴não se restringe agora $i(t)$ e $x(t)$ a serem invariantes do sistema

A solução da equação 3.41 nestas condições é

$$y(t) = 1 - e^{-\frac{t}{RC}} \quad \text{para } t > 0 \quad (3.42)$$

ou seja, um sinal que cresce exponencialmente para a amplitude $y(\infty) = 1$. De igual forma, a resposta ao escalão negativo de amplitude unitária

$$x(t) = \begin{cases} 1 & \text{para } t \leq 0 \\ 0 & \text{para } t > 0 \end{cases} \quad (3.43)$$

com a condição de fronteira $y(t) = 1$ para $t \leq 0$ será

$$y(t) = e^{-\frac{t}{RC}} \quad \text{para } t > 0 \quad (3.44)$$

Se se atender agora ao resultado obtido na secção 3.5 para a largura de banda deste sistema, expresso pela relação 3.20, as soluções 3.42 e 3.44 podem escrever-se, em função dessa largura de banda, respectivamente

$$\begin{array}{llll} \text{escalão positivo} & y_1(t) & = & 1 - e^{-2\pi B_T t} & \text{para } t > 0 \\ \text{escalão negativo} & y_2(t) & = & e^{-2\pi B_T t} & \text{para } t > 0 \end{array} \quad (3.45)$$

A figura 3.14 ilustra a forma da resposta para cada uma destas entradas em escalão. Em conclusão, quando a entrada é uma transição abrupta

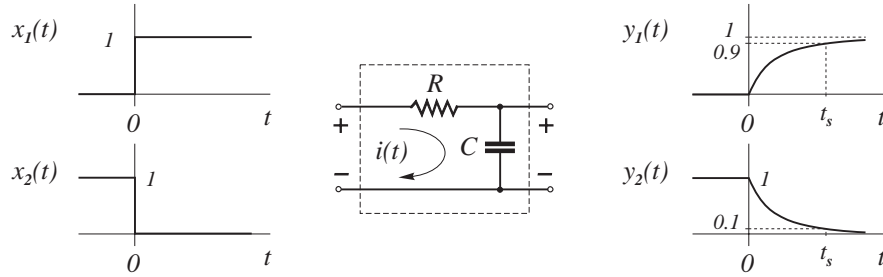


Figura 3.14: Resposta ao escalão de um sistema LIT PB de 1ª ordem

(instantânea) de tensão, o sinal de saída correspondente consiste numa transição exponencial entre o nível inicial e final da tensão de entrada com um *tempo de subida* (ou de *descida*) não nulo. E tal como se definiu a largura de banda de um sinal a 90% da sua energia ou potência média totais, assim também se considera que a resposta atingiu o nível final quando esta percorreu 90% do desnível do escalão à saída. O intervalo de tempo que decorre até ao instante em que isto acontece é designado de *tempo de subida* (ou de *descida*) do sinal, t_s .

3.10.2 Ritmo máximo de símbolos digitais

Consideremos agora a figura 3.15 na qual a entrada do sistema se apresenta uma sequência de pulsos representativos de um sinal digital de informação típico de uma transmissão de dados ou digitalizada. Por sobreposição, atendendo às soluções obtidas na secção anterior expressas pelas equações 3.45, a forma da resposta será a sequência de exponenciais crescentes e decrescentes indicada.

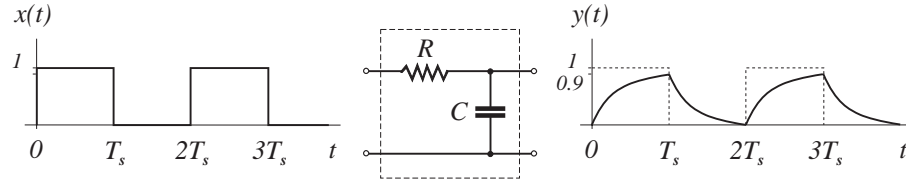


Figura 3.15: Resposta digital de um sistema LIT PB de primeira ordem

Tentemos determinar qual o ritmo máximo de pulsos (símbolos) que o sistema admite de modo a poderem ser recuperados à saída, isto é, de modo a que o receptor possa, inequivocamente, decidir da sua existência e do seu valor (amplitude, polaridade).

O aumento do ritmo de símbolos r_s obtém-se diminuindo a sua duração τ . E a questão que se coloca é a de saber até que ponto se pode diminuir esta duração por forma a que a saída represente inequivocamente a entrada. A resposta óbvia a esta questão será o menor τ desde que $\tau \geq t_s$.

Pela definição, o tempo de subida para o sistema LIT de 1ª ordem vem

$$1 - e^{-2\pi B_T t_s} = 0.9 \quad \text{donde} \quad (3.46)$$

$$t_s \approx \frac{0.37}{B_T} \quad (3.47)$$

Para sistemas LIT de ordem superior este tempo de subida aumenta ligeiramente e no limite, quando a ordem $n \rightarrow \infty$ (em que o sistema possui uma função de transferência ideal), pode verificar-se que o tempo de subida tem um valor

$$t_s \approx \frac{0.5}{B_T} \quad (3.48)$$

pelo que o ritmo teórico de símbolos digitais admissível num sistema LIT ideal tem de ser

$$\boxed{r_s \leq 2 B_T} \quad (3.49)$$

relação que foi invocada sem justificação no Capítulo 1. O ritmo de símbolos $r_s = 2B_T$ é designado de ritmo de Nyquist que foi quem demonstrou pela primeira vez este resultado. Tentativas de utilizar ritmos superiores a este limite dão origem ao que se designa por *interferência entre símbolos*. Na prática não é possível transmitir sinais digitais ao ritmo de Nyquist dado que nem os sistemas são ideais (ordem infinita e lineares) nem é possível originar sinais com transições de duração nula (escalões). Este resultado constitui, no entanto, uma referência fundamental.

3.10.3 Domínio do tempo versus domínio da frequência

A descrição de um sistema no domínio da frequência, feita pela sua função de transferência $H(f)$, é útil quando se pretende investigar as propriedades espectrais do sistema (banda de funcionamento, largura de banda, selectividade em frequência, etc) e bem assim as propriedades espectrais dos sinais resposta.

Neste domínio a relação entrada-saída é muito simples. É a relação multiplicativa da equação 3.12:

$$Y(f) = H(f) \cdot X(f)$$

Interessa trabalhar ou obter a resposta no domínio do tempo quando se pretende visualizar a *forma* dos sinais. No domínio do tempo esta relação é, como vimos, uma equação diferencial. Mas, se quisermos expressar os sinais resposta através de uma relação entre o sinal de entrada e uma característica do sistema como funções do tempo, de forma semelhante à relação 3.12 do domínio da frequência, ter-se-ia de determinar a função $h(t)$ que é a transformada inversa de $H(f)$, isto é, $h(t) \leftrightarrow H(f)$ e assim

$$y(t) = h(t) * x(t)$$

em que a relação $*$ não é uma multiplicação mas é agora uma *convolução*, uma relação integral do produto de $x(t)$ por $h(t)$ definida por

$$h(t) * x(t) = \int_{-\infty}^{\infty} h(\alpha) x(t - \alpha) d\alpha$$

Em geral é mais difícil trabalhar no domínio do tempo podendo-se chegar à *forma* das respostas neste domínio através do domínio da frequência calculando depois $y(t)$ por transformação inversa de $Y(f)$.

3.11 Problemas

3.1 – Considere um sistema de transmissão possuindo uma função de transferência, $H(f)$, dada por

$$H(f) = \frac{5}{1 + j\left(\frac{f}{4 \times 10^3}\right)^2}$$

- a) Classifique o sistema e represente graficamente a sua característica de amplitude.
 - b) Determine a largura de banda do sistema e a largura da banda de corte.
 - c) Considerando o sistema ideal com a largura de banda calculada em b), determine o ritmo máximo de símbolos, r_s , que o sistema admite.
- 3.2 – Suponha que a entrada $x(t)$ no sistema modelado na figura 3.3 é um pulso rectangular com 1 *Volt* de amplitude e 1 *ms* de duração. Suponha que a frequência de meia potência (-3 dB) do sistema é de 1 *KHz*. Determine:
- a) A energia total normalizada do sinal de entrada.
 - b) A densidade espectral de energia da resposta $y(t)$.
 - c) A energia total normalizada da resposta.

3.3 – Determinar e esquematizar a resposta $y(t)$ a um sinal de entrada

$$x(t) = \cos(2\pi f_0 t) - \frac{1}{3} \cos(6\pi f_0 t) + \frac{1}{5} \cos(10\pi f_0 t)$$

quando aplicado a um sistema passa-baixo de primeira ordem com largura de banda de transmissão $B_T = 3 f_0$.

3.4 – Determinar e esquematizar a resposta $y(t)$ quando o sinal de entrada do problema 3.3 é aplicado a um sistema passa-alto de primeira ordem com resposta em frequência

$$H(f) = \frac{jf}{B_T + jf} \quad \text{em que} \quad B_T = 3 f_0$$

- 3.5 – Considere um filtro RC passa-baixo de primeira ordem com função de transferência $H(f)$

$$H(f) = \frac{1}{1 + j\left(\frac{f}{B_T}\right)}$$

em que $R = 100 \text{ K}\Omega$ e $C = 200 \text{ pF}$.

- a) Calcule a atenuação, em *decibéis*, das componentes de frequência fundamental e da primeira harmônica não nula do sinal rectangular polar periódico NRZ ao ritmo de 8000 dig bin/s ao passar este filtro.
 - b) Qual a potência, em *dBm* e em *Watt*, da componente fundamental à saída, se à entrada ela é de -12 dBm ?
- 3.6 – Considere um sistema de comunicação digital cuja função de transferência $H(f)$ é razoavelmente aproximada por

$$H(f) = \frac{3.75 \times 10^3}{3.75 \times 10^3 - j9 \times 10^4 + jf}$$

- a) Esquematize a característica de amplitude do sistema e classifique-o.
 - b) Determine a largura de banda de transmissão do sistema se esta for definida a $\frac{1}{5}$ da potência.
 - c) Suponha agora que este sistema possui uma característica ideal com largura de banda igual à calculada em b). Esboce uma forma de onda para o sinal a ser transmitido através deste sistema correspondente à sequência de dígitos binários **101001**, ao ritmo máximo teórico.
- 3.7 – O circuito da figura 3.16 é uma realização de um filtro de Butterworth passa-baixo de segunda ordem com largura de banda $B_T = \frac{1}{2\pi\sqrt{LC}}$. Determinar uma expressão para $H(f)$ e obter a relação entre R , L e C .
- 3.8 – Calcular o valor mínimo da ordem n de um filtro de Butterworth, e $|H(2B_T)|^2$ em *dB* para esse filtro, tal que $|H(f)|^2 \geq -1 \text{ dB}$
- a) para $|f| < 0.8 B_T$
 - b) para $|f| < 0.9 B_T$

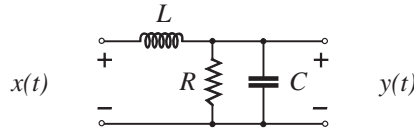


Figura 3.16: Filtro Butterworth de segunda ordem

3.9 – Considere que $R = \sqrt{\frac{L}{C}}$ no circuito da figura 3.16.

a) Mostre que

$$|H(f)|^2 = \left[1 - \left(\frac{f}{f_0} \right)^2 + \left(\frac{f}{f_0} \right)^4 \right]^{-1}$$

em que $f_0 = \frac{1}{2\pi\sqrt{LC}}$.

b) Determine a largura de banda a 3 dB em função de f_0 . Esquematize $|H(f)|$ e compare com a característica de um sistema de Butterworth de 2^ª ordem.

3.10 – Considere que no sistema com repetidores da figura 3.10 se tem $P_e = 0.5$ W, $\alpha = 2$ dB/Km e um comprimento total do percurso de 40 Km. Determine os ganhos dos amplificadores e a localização do repetidor de modo que $P_s = 100$ mW e que a potência do sinal à entrada de cada amplificador seja de 20μ W.

3.11 – Resolva o problema 3.10 com $P_e = 50$ mW e $P_s = 0.2$ W.

3.12 – Um sistema com repetidores, de 400 Km, consiste em m secções de cabo idênticas com $\alpha = 0.4$ dB/Km e m amplificadores idênticos com um ganho máximo de 30 dB. Determine o número de secções necessárias bem como o ganho por cada amplificador por forma a que se tenha $P_s = 50$ mW quando $P_e = 2$ W.

3.13 – Um sistema com repetidores, de 3000 Km, consiste em m secções de cabo idênticas com $\alpha = 0.6$ dB/Km e m amplificadores idênticos. Determine o número de secções necessárias e o ganho por amplificador por forma a que $P_s = P_e = 100$ mW e a que a potência de entrada em cada amplificador seja de pelo menos 25μ W.

- 3.14 – Determine e esquematize a resposta a uma sequência de pulsos rectangulares do filtro passa-alto (FPA) de primeira ordem modelado pelo circuito RC da figura 3.17.

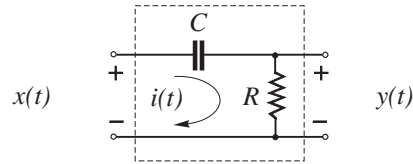


Figura 3.17: Modelo eléctrico de um sistema PA de primeira ordem

fim do capítulo 3

Capítulo 4

Ruído e Erros

Todos os sinais de comunicação úteis, sob o ponto de vista do receptor ou destinatário da informação, são aleatórios ou imprevisíveis. De facto, a *comunicação* não teria nenhum interesse se o destinatário tivesse conhecimento antecipado do comportamento exacto do sinal que irá receber dado que, se tal acontecesse, a *informação* recebida seria nula (como se justificará mais adiante) e não seria necessária a transmissão! Aquilo que o receptor conhece são as características gerais dos sinais utilizados na comunicação, como por exemplo a sua largura de banda, a densidade espectral de potência, os códigos e as técnicas de modulação utilizados para os símbolos da fonte, etc.

Assim, na impossibilidade de se lidar com descrições matemáticas determinísticas dos sinais de informação há que recorrer a descrições probabilísticas, em que os sinais são modelados por *processos aleatórios*.

Por outro lado, em qualquer sistema de transmissão, para além dos sinais de informação originados pela fonte, estão presentes outros sinais indesejáveis, genericamente designados por *ruído*, os quais não é possível eliminar totalmente.

O *ruído* é intrinsecamente aleatório dada a natureza dos fenómenos que o originam pelo que também devem ser descritos como *processos aleatórios*.

Mais propriamente, os sinais aleatórios são a manifestação de *processos aleatórios* ou *estocásticos* que têm lugar ao longo do tempo.

Neste capítulo faremos uma referência aos fundamentos da descrição de sinais através de processos aleatórios e em especial à descrição do *ruído*,

às suas características mais importantes e à forma como este afecta as comunicações.

4.1 Sinais aleatórios

A figura 4.1 mostra algumas das formas de onda, $s_1(t)$, $s_2(t)$, ..., $s_i(t)$, correspondentes à emissão de diferentes mensagens por uma fonte de informação. A mensagem concreta que é emitida em cada instância é pois desconhecida à priori, sendo portanto imprevisível a forma de onda (sinal) que irá ser produzida. O conjunto de todas as formas de onda

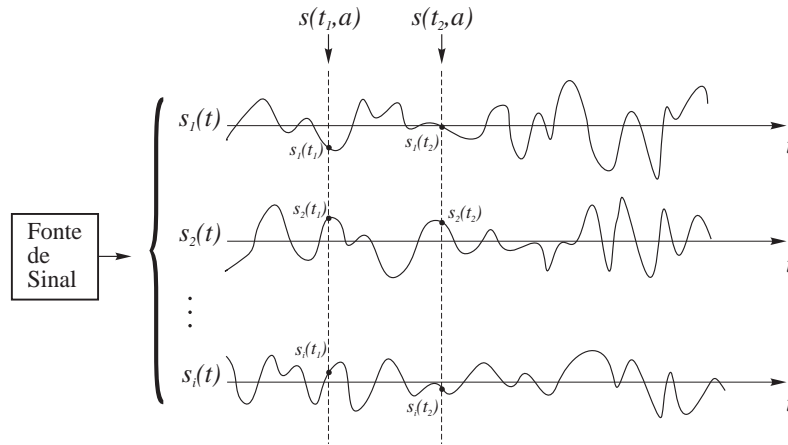


Figura 4.1: Formas de onda num conjunto $s(t, a)$

capazes de serem geradas pela fonte pode então ser representado formalmente pela função $s(t, a)$, em que cada elemento do conjunto, designado *função amostra*, corresponde a um determinado sinal, por exemplo, $s_i(t) = s(t, a_i)$.

O argumento fulcral que faz da função $s(t, a)$ um processo aleatório é a assumpção de que, quando se está a observar uma *função amostra*, não se sabe de qual das amostras a_i se trata. Portanto, num instante t_1 pode ocorrer *qualquer* valor do conjunto dos possíveis valores $s(t_1, a)$, o que significa que $s(t_1, a)$ constitui uma *variável aleatória*¹ que toma valores

¹apesar desta designação, uma *variável aleatória* não é nem variável nem aleatória! É uma *função* que toma valores numéricos bem definidos em consequência da ocorrência

definidos por $s(t, a_1), s(t, a_2), \dots, s(t, a_i)$ no instante $t = t_1$.

Do mesmo modo, $s(t_2, a)$ constitui outra *variável aleatória* que toma valores definidos por $s(t, a_1), s(t, a_2), \dots, s(t, a_i)$ no instante $t = t_2$, isto é, no conjunto $\{s(t_2, a_1), s(t_2, a_2), \dots, s(t_2, a_i)\}$.

Assim, um processo aleatório $s(t, a)$ que passamos a designar por $s(t)$, omitindo a referência a a , não é mais do que uma família de variáveis aleatórias $s(t_1), s(t_2), \dots, s(t_i)$, cujas funções de probabilidade descrevem o processo aleatório nos respectivos instantes de tempo.

Se, por exemplo, a *função de densidade de probabilidade (fdp)* de $s(t_1)$ for $p_1(s_1)$, em que a variável auxiliar s_1 não deve ser confundida com a função amostra $s_1(t) = s(t, a_1)$, a probabilidade do evento $s(t_1) \leq A$ pode ser obtida integrando $p_1(s_1)$ em $]-\infty, A]$ e o 1º momento, ou valor médio, $\overline{s(t_1)}$ pode ser obtido através da fórmula usual da esperança matemática de uma variável aleatória (eq 4.1).

Da mesma maneira, conhecida a *fdp* de $s(t_2)$, $p_2(s_2)$, se poderia calcular a probabilidade $P[s(t_2) \leq A]$ e a média $\overline{s(t_2)}$, cujos valores seriam, em geral, diferentes dos obtidos para $s(t_1)$.

Médias de conjunto

Designemos por $p(s, t)$ o conjunto das *fdp*, entendendo-se com esta notação que $p(s_1, t_1) = p_1(s_1)$, $p(s_2, t_2) = p_2(s_2)$, etc. A média estatística ou valor médio de $s(t)$ é dada por

$$\overline{s(t)} = E[s(t)] \doteq \int_{-\infty}^{\infty} s p(s, t) ds \quad (4.1)$$

em que, na operação de esperança, $E[\]$, t é uma constante. A equação 4.1 representa uma *média de conjunto*, isto é, uma média sobre o conjunto das formas de onda mantendo o tempo constante. Portanto, o valor da média de conjunto $\overline{s(t)}$ pode ser uma função do tempo pelo facto das médias $\overline{s(t_1)}, \overline{s(t_2)}, \dots, \overline{s(t_i)}$ poderem ser diferentes.

O mesmo se pode dizer para as restantes médias ou momentos de ordem superior. O n -ésimo momento de $s(t)$ será

$$\overline{s^n(t)} = E[s^n(t)] \doteq \int_{-\infty}^{\infty} s^n p(s, t) ds \quad (4.2)$$

de resultados de uma experiência que consiste na escolha, ao acaso, dos possíveis resultados dessa experiência

Processos estacionários e ergódicos

Um processo aleatório *estacionário* é aquele cujas características permanecem invariantes no tempo, isto é, qualquer translação da origem dos tempos para todo o conjunto de sinais-amostra $\{s(t, a_i)\}$ não afecta os valores das médias estatísticas. Neste caso tem-se $\overline{s^n(t_1)} = \overline{s^n(t_2)} = \dots = \overline{s^n(t_i)}$ e portanto $\overline{s^n(t)}$ não depende do tempo. Designadamente tem-se,

$$E[s(t)] = \bar{s} = m_s \quad (4.3)$$

$$E[s^2(t)] = \bar{s^2} = m_s^2 + \sigma_s^2 \quad (4.4)$$

Em que m_s e σ_s^2 são, respectivamente, a média e a variância² do processo aleatório $s(t)$.

Se num processo aleatório *todas* as médias de conjunto forem iguais às correspondentes médias temporais, o processo é designado de *ergódico*, isto é, ter-se-á

$$\langle s_i(t) \rangle = E[s(t)] \quad (4.5)$$

$$\langle s_i^2(t) \rangle = E[s^2(t)] \quad (4.6)$$

...

em que $\langle s_i(t) \rangle$ e $\langle s_i^2(t) \rangle$ são as já conhecidas médias temporais

$$\langle s_i(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s_i(t) dt \quad (4.7)$$

$$\langle s_i^2(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} s_i^2(t) dt = S \quad (4.8)$$

Se $0 < \bar{s^2} < \infty$, diz-se que o sinal aleatório é um sinal de potência com potência média $\bar{s^2}$. Assim, fica estabelecida uma relação entre as características da descrição aleatória e as da descrição temporal de um sinal, permitindo concluir que, se um sinal aleatório é estacionário e ergódico:

- (i) O valor médio m_s é igual à amplitude da *componente contínua* $\langle s(t) \rangle$.
- (ii) O quadrado da média m_s^2 é igual à potência normalizada da *componente contínua (DC)* $\langle s(t) \rangle^2$.

²a variância, ou 2º momento central, é dada por $\sigma_s^2 = \overline{(s - m_s)^2} = \bar{s^2} - m_s^2$

- (iii) O valor quadrático médio $\overline{s^2}$ é igual à *potência média total normalizada* $\langle s^2(t) \rangle = S$.
- (iv) A variância σ_s^2 é igual à *potência média normalizada das componentes variáveis no tempo* de $s(t)$, ou seja a potência AC.
- (v) O desvio padrão σ_s é igual à *raiz do valor quadrático médio*, ou seja, ao *valor eficaz*³ das componentes variáveis no tempo de $s(t)$.

Em conclusão, para efeitos da análise dos sinais de informação, a *função de densidade de probabilidade*, $p(s)$, de um sinal aleatório ergódico $s(t)$ substitui a sua descrição temporal. E, na realidade, verifica-se que os sinais de comunicação são razoavelmente bem modelados por processos estocásticos ergódicos.

4.2 Ruído

Os sinais eléctricos indesejáveis, genericamente designados de *ruído*, têm origens diversas podendo ser classificados como de *origem humana* ou de *origem natural*. O ruído de origem humana é causado, por exemplo, pela influência de outros sistemas de comunicação, de dispositivos de ignição e comutação eléctrica, etc. Os fenómenos que produzem o ruído natural são por exemplo as descargas atmosféricas, a radiação extra-terrestre e o ruído dos circuitos eléctricos. Um projecto bem concebido de um sistema de transmissão pode reduzir ou mesmo eliminar completamente os efeitos de certos tipos de ruído mas a presença de outros é mesmo inevitável o que impõe limitações fundamentais ao desempenho dos sistemas.

O ruído pode ser dividido em quatro categorias:

- (i) Ruído térmico
- (ii) Ruído de intermodulação
- (iii) *Crosstalk*
- (iv) Ruído impulsivo

³*root mean square (rms)*, $\sigma_s = \sqrt{\overline{s^2}}$

O ruído térmico é devido à agitação térmica dos electrões nos condutores, um movimento aleatório de partículas carregadas. É portanto, e fundamentalmente, uma função da temperatura a que o sistema se encontra.

O ruído de intermodulação acontece quando sinais com diferentes componentes de frequência partilham o mesmo meio de transmissão e interferem entre si. O seu efeito é a produção de sinais com frequências que são a soma ou a diferença das frequências que compõem os sinais originais.

O *crosstalk* pode ocorrer pelo acoplamento eléctrico ou electromagnético entre pares de fios próximos ou, mais raramente entre cabos coaxiais. Trata-se de um acoplamento indesejável entre os *percursos geográficos* dos sinais. Um exemplo típico é o da escuta de uma conversação telefónica estranha por *cruzamento de linhas*.

O ruído impulsivo é a designação da ocorrência irregular de *pulsos* ou *estalos* de curta duração e de relativamente grande amplitude (*spikes*). As suas causas são várias, como por exemplo, as perturbações electromagnéticas externas, as descargas atmosféricas, e as falhas ocasionais do próprio sistema de transmissão.

As primeiras três categorias de ruído possuem características (aleatórias) relativamente bem previsíveis ao passo que o ruído impulsivo não tem características regulares, não é contínuo.

O ruído impulsivo perturba pouco as comunicações analógicas. Uma transmissão telefónica pode ser corrompida por curtos *pulsos* ou *estalos* sem perda de inteligibilidade. Contudo, nas transmissões digitais, este tipo de ruído é a principal fonte de erro. Um *pulso* de ruído de 10 ms de duração não afectaria uma transmissão de voz analógica mas corromperia cerca de 50 símbolos de dados que estivessem a ser transmitidos digitalmente a 4 800 baud.

4.2.1 O ruído térmico

A teoria cinética das partículas diz que a energia média de uma partícula à temperatura absoluta de \mathcal{T} é proporcional a $k\mathcal{T}$ em que k é a constante de Boltzmann.

Quando uma resistência metálica de valor $R \Omega$ está a uma temperatura \mathcal{T} , o movimento aleatório dos electrões produz uma tensão aleatória de

ruído $n(t)$ aos seus terminais. De acordo com o teorema do limite central⁴, $n(t)$ possui uma *função de densidade de probabilidade gaussiana*, $p_N(n)$ de média nula e variância σ_n^2 , sendo

$$p_N(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-(n-m_n)^2/2\sigma_n^2} \quad (4.9)$$

$$m_n = \bar{n} = 0 \quad (4.10)$$

$$\sigma_n^2 = \overline{n^2} = \frac{2(\pi k \mathcal{T})^2}{3h} R \quad \text{Volt}^2 \quad (4.11)$$

onde a temperatura \mathcal{T} é medida em Kelvin e

$$\begin{aligned} k &= \text{constante de Boltzmann} = 1.38 \times 10^{-23} \quad \text{Joules/Kelvin} \\ h &= \text{constante de Planck} = 6.60 \times 10^{-34} \quad \text{Joules-segundo} \end{aligned}$$

As constantes presentes na equação 4.11 revelam que se trata de um resultado da mecânica quântica a qual consegue mostrar ainda que a *densidade espectral de potência* do ruído térmico produzido pela resistência de $R \Omega$ é dada por

$$|N(f)|_R^2 \approx 2 R k \mathcal{T} \left(1 - \frac{h|f|}{2k\mathcal{T}}\right) \quad \text{Volt}^2/\text{Hz} \quad \text{para } |f| \ll \frac{k\mathcal{T}}{h} \quad (4.12)$$

em que se utiliza o índice- R para indicar que se trata da densidade total do ruído produzido por uma resistência de valor $R \Omega$.

Se, por exemplo, a resistência se encontra a uma temperatura ambiente de 21°C, a que corresponde $\mathcal{T}_0 = 294 \text{ K}$, então $|N(f)|_R^2$ será essencialmente constante para $|f| < 0.1 \frac{k\mathcal{T}_0}{h} \approx 0.6 \times 10^{12} \text{ Hz}$. O limite superior desta banda de frequências situa-se na zona do infravermelho do espectro eletromagnético, substancialmente acima das frequências utilizadas nas transmissões eléctricas pelo que, para todos os efeitos práticos, se pode dizer que a densidade espectral de potência do ruído térmico é constante e igual a

$$|N(f)|_R^2 = 2 R k \mathcal{T} \quad \text{Volt}^2/\text{Hz} \quad (4.13)$$

No entanto, nem toda a potência de ruído produzida pela resistência é transferida, sob a forma de sinal de ruído, para o sistema de transmissão,

⁴O *teorema do limite central* diz essencialmente o seguinte: Se uma variável aleatória X resulta da soma de n componentes aleatórias independentes e se cada componente contribui muito pouco para a soma, então, a função de distribuição de probabilidade de X tende para uma distribuição *gaussiana* ou *normal* à medida que n aumenta, independentemente das distribuições de probabilidade das componentes individuais

porque parte dela é dissipada na própria resistência. Só uma parte dessa potência se manifesta no sistema e essa parte é máxima quando a resistência está ligada a uma carga, R_L , de igual valor⁵, $R_L = R$. Resultado que é aplicável à densidade de potência. A figura 4.2 mostra o circuito equivalente de Thévenin da resistência como fonte de ruído térmico, ligada a uma carga resistiva de valor $R_L = R$ que representa o sistema de transmissão influenciado pelo ruído produzido pela resistência. A tensão aos

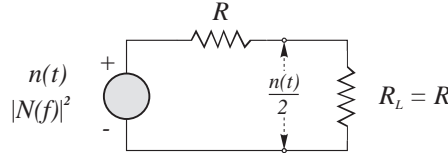


Figura 4.2: Circuito equivalente de Thévenin de uma fonte resistiva de ruído térmico e respectiva carga de máxima transferência de potência

terminais da carga, devido à divisão de tensão, é $\frac{n(t)}{2}$ e o seu valor eficaz é $V = \sqrt{\overline{(n(t)/2)^2}}$. A potência média transmitida à carga é pois V^2/R , ou seja, $\overline{n^2(t)}/4R$, em que $\overline{n^2(t)}$ é a potência média total produzida pela resistência. Assim, também, a densidade espectral de potência (máxima) do ruído térmico que se manifesta no sistema, será dada por

$$|N(f)|^2 = \frac{|N(f)|_R^2}{4R} = \frac{k\mathcal{T}}{2} \text{ Watt/Hz} \quad (4.14)$$

valor que depende apenas da temperatura (ver Prob 4.1).

4.2.2 Ruído branco e gaussiano

Para além do ruído térmico produzido pelos circuitos, de que a análise para uma resistência foi um exemplo, muitas outras fontes de ruído se caracterizam por uma *função de densidade de probabilidade gaussiana* e uma *densidade espectral de potência constante* ao longo de quase todo o espectro electromagnético útil.

⁵é um resultado conhecido da teoria dos circuitos eléctricos que a máxima transferência de potência para uma carga se dá quando a impedância da carga é o complexo conjugado da impedância da fonte de potência. Neste caso, o complexo conjugado de um valor puramente resistivo é igual a ele próprio.

O espectro do ruído em geral é caracterizado por possuir todas as componentes de frequência em igual proporção e, por esta razão e por analogia com a luz branca, é chamado de *ruído branco*. Em comunicações eléctricas o ruído *branco* e *gaussiano* é, na grande maioria das situações, um modelo aceitável para o ruído total presente na transmissão do sinal e manifesta-se de forma *aditiva*, no sentido de se ir adicionando ao sinal em vários pontos ao longo do percurso entre a fonte e o destino, sobrepondo-se aditivamente, tanto em *amplitude* como em *potência*, ao sinal de informação.

A combinação de todas as fontes de ruído deste tipo (e não só o de origem térmica) permite-nos escrever a sua densidade espectral sob a forma

$$|N(f)|^2 = \frac{\eta}{2} \quad \text{Watt/Hz} \quad (4.15)$$

em que o factor $1/2$ indica que metade da densidade de potência diz respeito ao sub-domínio negativo das frequências e a outra metade ao positivo e η quantifica não só a componente térmica ($k\mathcal{T}$) mas todo o ruído de origem natural. A figura 4.3 ilustra graficamente as duas funções características do ruído branco e gaussiano.

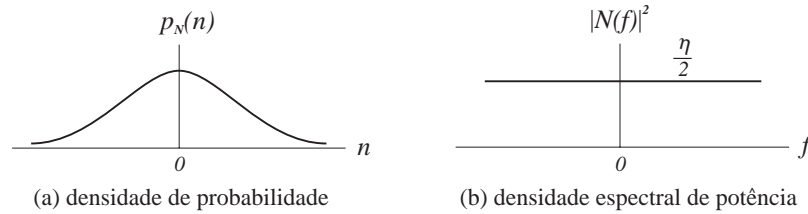


Figura 4.3: Características do ruído branco e gaussiano

4.2.3 Largura de banda equivalente de ruído

Considerar a densidade de potência do ruído como constante, levaria a concluir que a sua potência total no receptor seria infinita⁶ pelo que seria impossível detectar o sinal na presença de um tal valor do ruído. Porém assim não acontece dado que o sistema de transmissão tem sempre uma largura de banda limitada, limitando naturalmente aquela potência total.

⁶ pois $\int_{-\infty}^{\infty} |N(f)|^2 df = \frac{\eta}{2} \int_{-\infty}^{\infty} df = \infty$

Se considerarmos um sistema sem sinal de entrada, apenas o ruído se manifesta à saída com uma potência média

$$\begin{aligned}
 N &= \int_{-\infty}^{\infty} |N(f)|^2 |H(f)|^2 df \\
 &= \frac{\eta}{2} \int_{-\infty}^{\infty} |H(f)|^2 df \\
 &= \eta \int_0^{\infty} |H(f)|^2 df
 \end{aligned} \tag{4.16}$$

em que o integral na equação 4.16 tem as dimensões de Hz, ou seja, de uma largura de banda, podendo-se escrever

$$N = \eta g B_N \tag{4.17}$$

com

$$B_N \doteq \frac{1}{g} \int_0^{\infty} |H(f)|^2 df \tag{4.18}$$

e em que $g = |H(f)|_{\max}^2$ é o ganho máximo do sistema e B_N é a sua largura de banda equivalente de ruído ou, simplesmente, largura de banda de ruído.

Exêmplo 4.1 Considere-se o sistema de transmissão de 1ª ordem, com largura de banda a 3 dB igual a B_T , representado pela característica de potência

$$|H(f)|^2 = \frac{1}{1 + \left(\frac{f}{B_T}\right)^2}$$

Tem-se para este caso $g = 1$ e

$$\begin{aligned}
 B_N &= \int_0^{\infty} \frac{1}{1 + \left(\frac{f}{B_T}\right)^2} df \\
 &= B_T \int_0^{\infty} \frac{1/B_T}{1 + \left(\frac{f}{B_T}\right)^2} df \\
 &= B_T \left[\arctg \left(\frac{f}{B_T} \right) \right]_0^{\infty} \\
 &= B_T \frac{\pi}{2} \\
 B_N &\approx 1.57 B_T
 \end{aligned} \tag{4.19}$$

A figura 4.4 dá uma interpretação geométrica a B_N . Neste caso verifica-se que a largura de banda de ruído (B_N) é cerca de 50% maior do que a largura de banda a 3 dB (B_T).

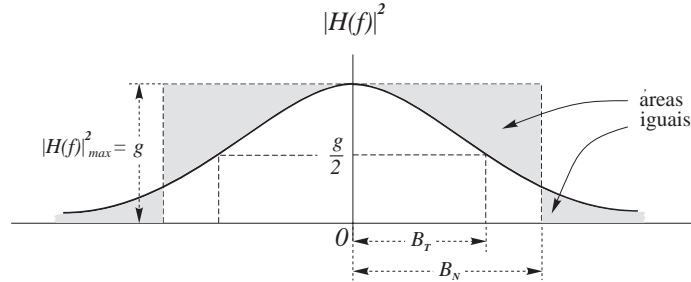


Figura 4.4: Largura de banda equivalente de ruído de um sistema PB

Assim, adopta-se para largura de banda de um sistema a seguinte

Definição 4.1 – Largura de banda equivalente de ruído

de um sistema, é a largura de banda do filtro ideal que deixa passar a mesma potência de ruído que esse sistema e tem o mesmo ganho máximo.

Se o sistema do exemplo 4.1 fosse mais selectivo⁷, isto é, com uma transição de corte mais abrupta, a largura de banda de ruído aproximar-se-ia da largura de banda a 3 dB (ver Prob 4.3). Na maior parte dos casos da prática toma-se, sem grande erro, $B_N \approx B_T$. Em resumo, se o ruído num sistema de transmissão com largura de banda B_T , é branco e gaussiano, ele será *filtrado* pelo sistema e manifestar-se-á no receptor com uma potência média

$$\boxed{N = \sigma_n^2 = \eta g B_T \text{ Watt}} \quad (4.20)$$

4.3 Erros

4.3.1 Regeneração do sinal digital

Suponhamos uma transmissão digital binária unipolar. Os símbolos transmitidos são pulsos rectangulares com T_s s de duração que podem tomar

⁷um sistema de ordem superior à 1ª possuindo, portanto, uma banda de transição mais estreita

apenas dois valores, isto é, o símbolo k terá amplitude $a_k = 0$ ou $a_k = A$ consoante represente, respectivamente, o valor lógico 0 ou o valor lógico 1. Seja t_k o instante dentro do intervalo do símbolo k em que o receptor toma uma decisão sobre o valor desse símbolo. Esta decisão é tomada com base no valor do sinal $y(t)$ à entrada do receptor no instante t_k

$$y(t_k) = a_k + n(t_k)$$

instante em que, para o efeito, é obtida uma *amostra* do sinal recebido cuja amplitude, $y(t_k)$, é *mantida*⁸ constante durante um tempo igual à duração do símbolo, T_s . O valor de $y(t_k)$ será a soma da amplitude do sinal com a amplitude do ruído naquele instante, previamente filtrada por um filtro passa-baixo de modo a limitar a potência total do ruído à entrada do receptor. O procedimento de decisão culmina com a comparação entre os sucessivos valores de $y(t_k)$ e uma *amplitude de limiar* fixa, V . A resposta do comparador será uma amplitude maior do que zero, indicando o valor lógico 1, ou zero, indicando o valor lógico 0, consoante $y(t_k) > V$ ou $y(t_k) < V$ respectivamente.

O comparador produz, assim, uma estimativa $\tilde{x}(t)$ do sinal que foi transmitido, $x(t)$. Estes procedimentos são ilustrados nas figuras 4.5 e 4.6 que representam os sub-sistemas componentes do receptor e as respectivas formas de onda e constituem o que se designa por *regeneração* do sinal.

Na realidade o regenerador comporta-se como um conversor analógico-digital ao converter a forma de onda analógica ruidosa $y(t)$ num sinal digital sem ruído, $\tilde{x}(t)$, ocasionalmente com erros.

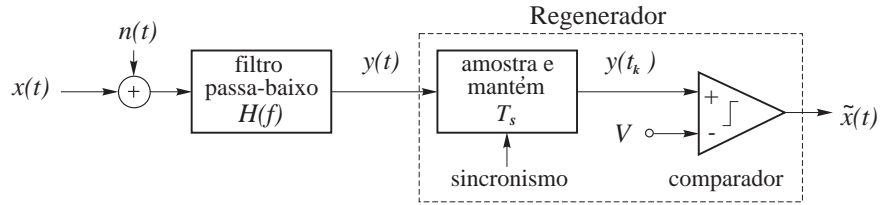


Figura 4.5: Receptor binário de banda de base

⁸trata-se da função *amostra-e-mantém* (ou *sample-and-hold*, S/H)

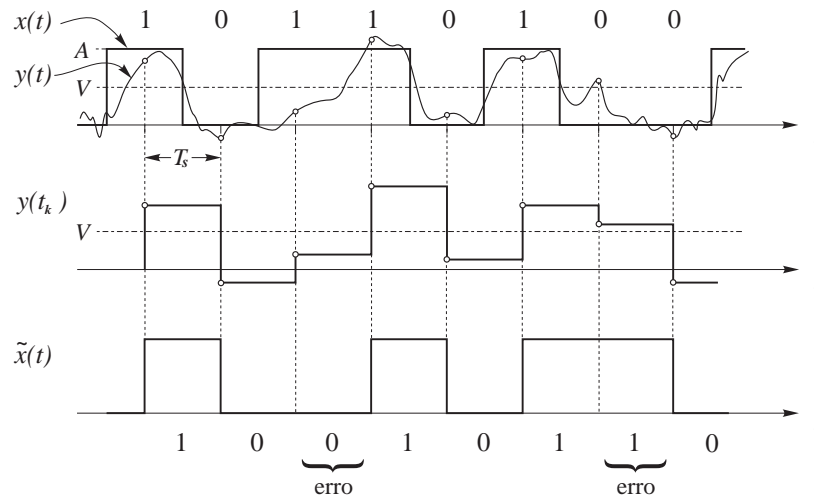


Figura 4.6: Regeneração de um sinal binário unipolar

4.3.2 Probabilidade de erro

Existe erro na transmissão digital quando a estimativa não coincide com o valor transmitido, isto é, quando $\tilde{x}(t_k) \neq x(t_k)$ o símbolo k está errado. Interessa conhecer a probabilidade de isto acontecer pois será uma *medida* importante da qualidade, ou desempenho, do sistema de transmissão digital. Prosseguindo na análise da transmissão unipolar, pode dizer-se, intuitivamente, que a amplitude de limiar, V , deve situar-se num valor intermédio $0 < V < A$. A regeneração dará origem a erro se, no instante da amostragem:

- o símbolo transmitido é o 0 ($a_k = 0$) e o ruído excede V ($n(t_k) > V$),
ou
- o símbolo transmitido é o 1 ($a_k = A$) e a soma do ruído com A é inferior a V ($n(t_k) + A < V$)

o que se traduz na seguinte expressão para a probabilidade de erro onde $P(X)$ representa a probabilidade do evento X :

$$P_e = P(a_k = 0) \cdot P(n(t_k) > V) + P(a_k = A) \cdot P(n(t_k) + A < V) \quad (4.21)$$

Dado que o ruído é estacionário $P(n(t_k) = x_0) = P(n(t) = x_0) = P(n = x_0) \quad \forall t, t_k$, isto é, a probabilidade do ruído ter um valor de amplitude igual a x_0 é independente do tempo, pelo que

$$P_e = P(a_k = 0) \cdot P(n > V) + P(a_k = A) \cdot P(n < V - A) \quad (4.22)$$

ou, escrita de uma forma mais condensada

$$P_e = P_0 \cdot P_{e0} + P_1 \cdot P_{e1} \quad (4.23)$$

Estamos a supôr que a ocorrência dos sucessivos valores dos símbolos transmitidos são estatisticamente independentes. Por outro lado espera-se que a ocorrência de 0s e 1s numa mensagem longa seja equiprovável, hipótese que é razoável admitir⁹, ou seja, $P(a_k = 0) = P(a_k = A) = \frac{1}{2}$, donde

$$P_e = \frac{1}{2} [P(n > V) + P(n < V - A)] \quad (4.24)$$

em que

$$P_{e0} = P(n > V) = \int_V^\infty p_N(n) dn = Q\left(\frac{V}{\sigma_n}\right) \quad (4.25)$$

$$\begin{aligned} P_{e1} &= P(n < V - A) = \int_{-\infty}^{V-A} p_N(n) dn \\ &= \int_{-\infty}^V p_N(n - A) dn = Q\left(\frac{A - V}{\sigma_n}\right) \end{aligned} \quad (4.26)$$

e $Q(k)$ é a probabilidade da cauda gaussiana¹⁰. O ábaco da tabela 4.2, no fim deste capítulo, permite determinar alguns valores para $Q(k)$. A probabilidade de erro pode ser escrita sob a forma

$$P_e = \frac{1}{2} \left[Q\left(\frac{V}{\sigma_n}\right) + Q\left(\frac{A - V}{\sigma_n}\right) \right] \quad (4.27)$$

Se os símbolos são originalmente equiprováveis e se o ruído, em média, afecta igualmente os símbolos transmitidos independentemente da amplitude destes, então não há razão para que o regenerador altere esta

⁹como se justificará mais adiante na Teoria da Informação

¹⁰ $Q(k)$ é a probabilidade de uma variável aleatória gaussiana de média m e variância σ^2 tomar um valor maior do que $m + k\sigma$, ié,

$$Q(k) \doteq \frac{1}{\sqrt{2\pi}} \int_k^\infty e^{-x^2/2} dx$$

também chamada área sob a cauda gaussiana

regra produzindo uma sequência $\tilde{x}(t)$ em que os símbolos não sejam equiprováveis. Isto significa que o valor óptimo da amplitude de limiar, V_{opt} , que é também o valor que minimiza P_e , se deve situar no ponto onde as duas *funções de densidade de probabilidade* se intersectam — o ponto médio do intervalo $[0, A]$ — ao qual correspondem áreas de cauda iguais e portanto probabilidades de erro iguais. Assim sendo, $V_{opt} = A/2$.

A figura 4.7 ilustra esta discussão mostrando uma situação em que $V < V_{opt} = A/2$. Neste caso, a área sombreada mais escura é menor do que a área sombreada mais clara, significando que a probabilidade de erro quando $a_k = A$ será menor do que a probabilidade de erro quando $a_k = 0$, decidindo o regenerador mais favoravelmente pelo valor 1 do que pelo valor 0, em média. Com $V = V_{opt} = A/2$ tem-se $Q\left(\frac{V}{\sigma_n}\right) = Q\left(\frac{A-V}{\sigma_n}\right) =$

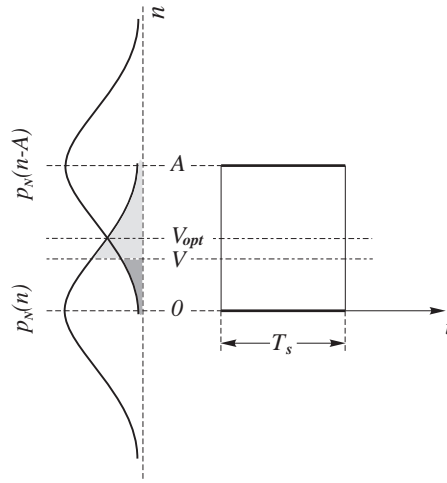


Figura 4.7: $Fdps$, amplitude de limiar de decisão e probabilidades de erro

$Q\left(\frac{A}{2\sigma_n}\right)$ pelo que da equação 4.27 resulta finalmente a probabilidade de erro para a transmissão binária unipolar

$$P_e = Q\left(\frac{A}{2\sigma_n}\right) \quad (4.28)$$

No caso geral em que os símbolos gerados pelo emissor não são equiprováveis, o valor óptimo da amplitude de limiar, que minimiza P_e , é a solução da equação $\frac{dP_e}{dV} = 0$ com P_e definido pela equação 4.22. A resolução desta equação que envolve a diferenciação dos integrais 4.25 e 4.26, conduz à

seguinte relação

$$P_0 \cdot P(n > V_{opt}) = P_1 \cdot P(n < V_{opt} - A) \quad (4.29)$$

Os valores σ_n e A podem expressar-se em função das potências médias do ruído e do sinal respectivamente, dado que $N = \sigma_n^2$ e, para o sinal unipolar, $S = A^2/2$

$$P_e = Q \left(\sqrt{\frac{1}{2} \frac{S}{N}} \right) \quad (4.30)$$

Por outro lado o filtro passa-baixo à entrada do regenerador não pode ter largura de banda inferior a metade do ritmo de símbolos, isto é, $B_T \approx B_N \geq r_s/2$ pelo que a potência do ruído é sempre

$$N = \eta B_T \geq \eta \frac{r_s}{2} \quad (4.31)$$

podendo-se obter o seguinte limite inferior para a probabilidade de erro

$$P_e \geq Q \left(\sqrt{\frac{S}{\eta r_s}} \right) \quad (4.32)$$

donde se conclui que, para a mesma potência média do sinal e densidade de ruído, a probabilidade de erro aumenta quando o ritmo de símbolos aumenta (e vice-versa).

É usual, ainda, representar P_e em função da *energia média por símbolo*, E_s . Para tal basta notar que $E_s = S \cdot T_s = S/r_s$ Joules e vem

$$P_e \geq Q \left(\sqrt{\frac{E_s}{\eta}} \right) \quad (4.33)$$

A quantidade

$$\gamma_s = \frac{E_s}{\eta} \quad (4.34)$$

é um parâmetro importante na medida de desempenho de sistemas de transmissão digitais e é comparável ao parâmetro *razão sinal-ruído*, $\frac{S}{N}$ que é mais utilizado para a medida de desempenho de sistemas de transmissão analógicos.

Os casos de outros códigos de linha podem ser tratados de forma semelhante. Na codificação binária *polar*, por exemplo, os símbolos terão

amplitudes $a_k = \pm A/2$. A amplitude de limiar mantém o seu valor relativo a estas amplitudes, isto é, $V_{opt} = 0$ e a potência média do sinal será $S = A^2/4$.

A tabela 4.1 resume as várias formas de expressar a probabilidade de erro para os sinais unipolares e polares.

Tabela 4.1: Probabilidades de erro

P_e	
Unipolar	Polar
$Q\left(\frac{A}{2\sigma_n}\right)$	$Q\left(\frac{A}{2\sigma_n}\right)$
$Q\left(\sqrt{\frac{1}{2}\frac{S}{N}}\right)$	$Q\left(\sqrt{\frac{S}{N}}\right)$
$\geq Q\left(\sqrt{\frac{E_s}{\eta}}\right)$	$\geq Q\left(\sqrt{\frac{2E_s}{\eta}}\right)$
$\geq Q\left(\sqrt{\gamma_s}\right)$	$\geq Q\left(\sqrt{2\gamma_s}\right)$

Exêmplo 4.2 *Um computador transmite por uma porta de comunicações, pulsos unipolares ao ritmo de 10^6 bps = 1 Mbps para transmissão por um sistema com ruído de densidade espectral de potência 4×10^{-20} W/Hz. Pretende-se determinar o valor para a potência média do sinal de modo a que a taxa de erros não exceda um bit por hora.*

Tem-se portanto $r_s = 10^6$ e $\eta = 4 \times 10^{-20}$. A taxa de erros máxima define o valor máximo da probabilidade de erro que é

$$\begin{aligned}
 P_e &= \frac{n^o \text{ de dígitos errados em 1 hora}}{n^o \text{ total de dígitos transmitidos em 1 hora}} \\
 &\leq \frac{1}{3600 \times r_s} = \frac{1}{3600 \times 10^6} \approx 2.8 \times 10^{-10}
 \end{aligned}$$

consultando o ábaco de $Q(k)$ verifica-se que $Q(6.2) \approx 2.8 \times 10^{-10}$ e, como o sinal é unipolar, $\sqrt{S/\eta r_s} = 6.2$, donde resulta

$$\begin{aligned}
 S &\geq 6.2^2 \times \eta r_s = 6.2^2 \times 4 \times 10^{-20} \times 10^6 \approx 1.5 \times 10^{-12} = 1.5 \text{ pW} \\
 S_{dBm} &\geq -88 \text{ dBm}
 \end{aligned}$$

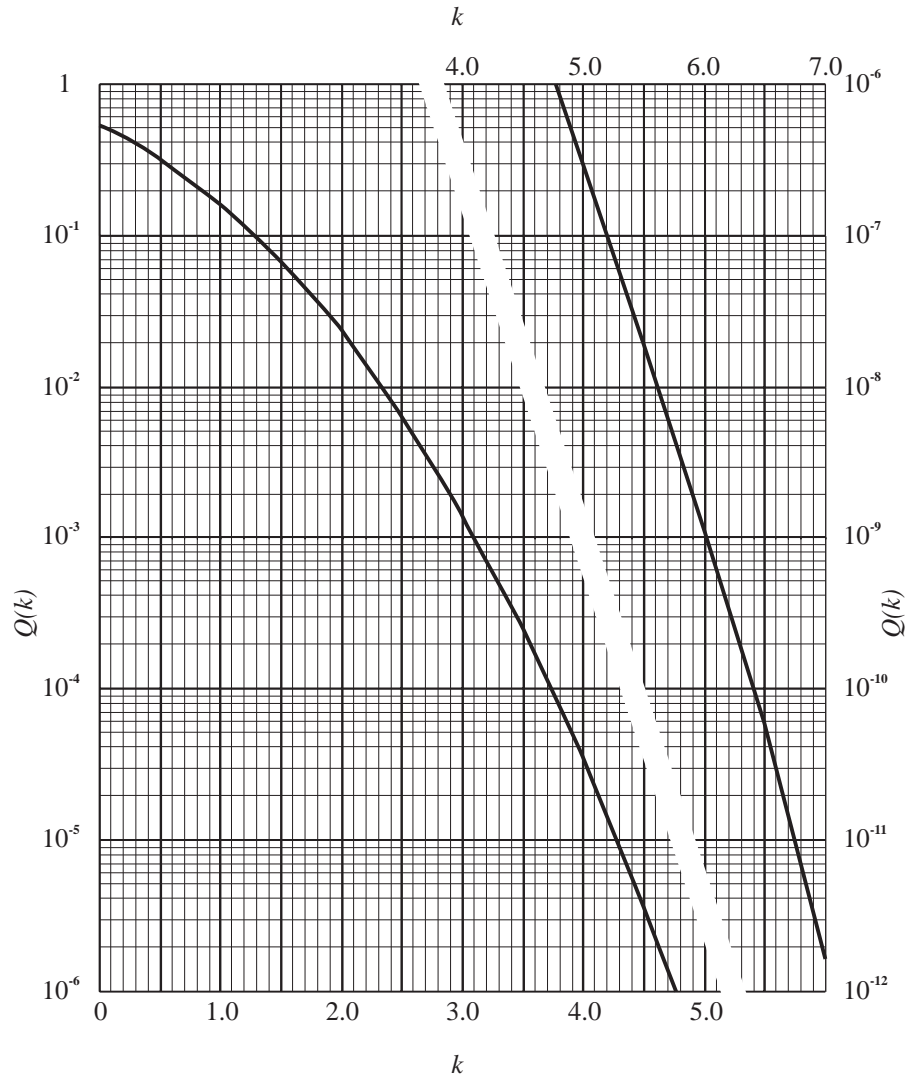
Trata-se de um valor bastante pequeno. No entanto representa a potência mínima do sinal que deve chegar ao receptor. Se se pretendesse determinar a potência mínima do sinal no transmissor dever-se-ia adicionar a este valor a atenuação introduzida pelo sistema de transmissão.

Tabela 4.2: Probabilidades da Cauda Gaussiana

A probabilidade de uma variável aleatória gaussiana de média m e variância σ^2 tomar um valor maior do que $m + k\sigma$ é dada por

$$Q(k) \doteq \frac{1}{\sqrt{2\pi}} \int_k^{\infty} e^{-x^2/2} dx$$

$$P(X > m + k\sigma) = P(X \leq m - k\sigma) = Q(k)$$



4.4 Problemas

- 4.1 – A partir da equação 4.11 calcule o valor eficaz (raiz do valor quadrático médio) do ruído térmico aos terminais de uma resistência de $1\ \Omega$ à temperatura de 29 Kelvin. Utilize a equação 4.12 para determinar qual a percentagem desse valor que é devida às componentes de frequência do ruído na banda $|f| < 1\ \text{GHz}$.
- 4.2 – O ruído térmico produzido por uma resistência de $10\ \text{K}\Omega$ à temperatura ambiente é aplicado a um filtro passa-baixo ideal com $B_T = 2.5\ \text{MHz}$ e ganho unitário. A tensão de ruído assim filtrada é aplicada a um retificador de onda completa, produzindo uma tensão $z(t) = |n(t)|$. Calcule o valor médio (componente contínua) e o valor quadrático médio (potência média) de $z(t)$.
- 4.3 – Considere o sistema de Butterworth de ordem- n definido (no capítulo 3) por

$$|H(f)| = \frac{1}{\sqrt{1 + \left(\frac{f}{B_T}\right)^{2n}}}$$

Mostre que a largura de banda equivalente de ruído está relacionada com a largura de banda a 3 dB por

$$B_N = \frac{\pi B_T}{2n \sin\left(\frac{\pi}{2n}\right)}$$

e que portanto $B_N \rightarrow B_T$ quando $n \rightarrow \infty$.

- 4.4 – Considere um sistema de transmissão para sinais binários unipolares com símbolos equiprováveis e $S/N = 50$ no receptor. Calcule as probabilidades de erro para cada um dos dois símbolos P_{e0} e P_{e1} e a probabilidade total de erro P_e quando a amplitude de limiar tem um valor não-ótimo $V = 0.4 A$. Compare o valor de P_e com o valor mínimo dado pela equação 4.28.
- 4.5 – Determine a razão S/N no receptor de um sistema binário unipolar com ruído branco e gaussiano de modo a que $P_e = 0.001$. Qual seria a probabilidade de erro de um sistema polar com a mesma S/N ?
- 4.6 – Um sistema de transmissão binário é afectado de ruído branco e gaussiano com $\eta = 10^{-8}/r_s$. Determine a potência média do sinal no receptor, para sinalização tanto polar como unipolar por forma a que $P_e \leq 10^{-6}$.

- 4.7 – Utilize a equação 4.29 para obter uma expressão para a amplitude de limiar ótima num sistema binário polar com ruído branco e gaussiano quando $P_0 \neq P_1$ (isto é, quando $P(a_k = -A/2) \neq P(a_k = +A/2)$).
- 4.8 – Um sistema de transmissão binário polar possui 3 repetidores cada um com $S/N = 20$ dB. Determine P_e quando os repetidores são regenerativos e não-regenerativos.

fim do capítulo 4

Capítulo 5

Digitalização

A digitalização é o processo pelo qual se transformam os sinais analógicos, contínuos no tempo, em sequências de números, com um número limitado de dígitos, que representam a amplitude do sinal em instantes do tempo regularmente espaçados.

Assim, a digitalização de um sinal analógico pressupõe não só que se consideram apenas alguns dos valores do sinal contínuo, ignorando-se os valores intermédios (discretização no tempo), como também que a amplitude analógica original deixa de poder tomar quaisquer valores dentro do intervalo contínuo de variação da amplitude do sinal, para tomar apenas alguns (discretização na amplitude), única forma de poder ser representada por um número limitado (e normalmente pequeno) de dígitos.

No processo de digitalização para transmissão distinguem-se quatro fases:

- (i) A *amostragem*, que consiste na recolha periódica de valores do sinal (*amostras*) substituindo-se o sinal completo apenas por esses valores;
- (ii) A *quantização*, pela qual as amplitudes das amostras são aproximadas a um número limitado de níveis quânticos;
- (iii) A *conversão analógico-a-digital*, que consiste na representação do valor das amplitudes quantizadas por números, normalmente na base dois e
- (iv) A *codificação de linha*, através da qual os dígitos binários dos números que representam as amostras são transformados em formas

de onda com as características mais apropriadas à sua transmissão.

Neste capítulo analisam-se as três primeiras fases deste processo bem como o processo inverso que permite recuperar o sinal analógico a partir dos valores numéricos das amostras. A codificação de linha é um processamento comum quer aos sinais analógicos que foram digitalizados, quer aos sinais de fontes originalmente digitais e que se destina prepará-los para a transmissão no canal e será abordada em capítulo próprio.

5.1 Teoria da Amostragem

Suponhamos um sinal de informação analógico qualquer $x(t)$, representado na figura 5.1 a), possuindo um espectro com característica de amplitude $|X(f)|$ e largura de banda B Hz. Suponhamos ainda que o espectro do sinal é limitado à banda $[-B, +B]$, isto é, $|X(f)| = 0$ para $|f| \geq B$. A figura 5.1 b) ilustra esta situação.

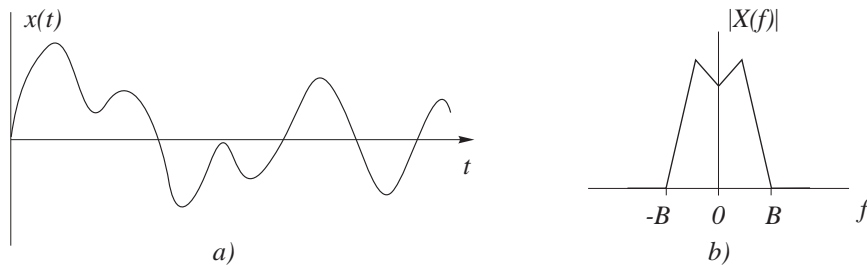


Figura 5.1: Sinal analógico de espectro limitado à banda $[-B, +B]$

5.1.1 A operação de amostragem

A operação de amostragem do sinal $x(t)$ consiste em tomar os valores do sinal em instantes regularmente espaçados no tempo e considerar que nos intervalos de tempo entre esses instantes o sinal tem amplitude nula. De uma maneira simplista, pode dizer-se que o circuito da figura 5.2 c) executa esta operação. À entrada do circuito introduz-se o sinal $x(t)$. Quando o comutador estiver na posição 2 a saída é zero e quando estiver na posição 1 a saída é igual à entrada. Se o comutador estiver na posição 2 e periodicamente "tocar" na posição 1 obtem-se à saída o sinal amostrado

$x_a(t)$ que consiste numa sequência de amplitudes do sinal $x(t)$ tomadas a intervalos regulares de tempo. Esses intervalos de tempo constituem o *período de amostragem* que designaremos por T_a e cujo inverso é a *frequência de amostragem* em Hertz, $f_a = 1/T_a$ Hz.

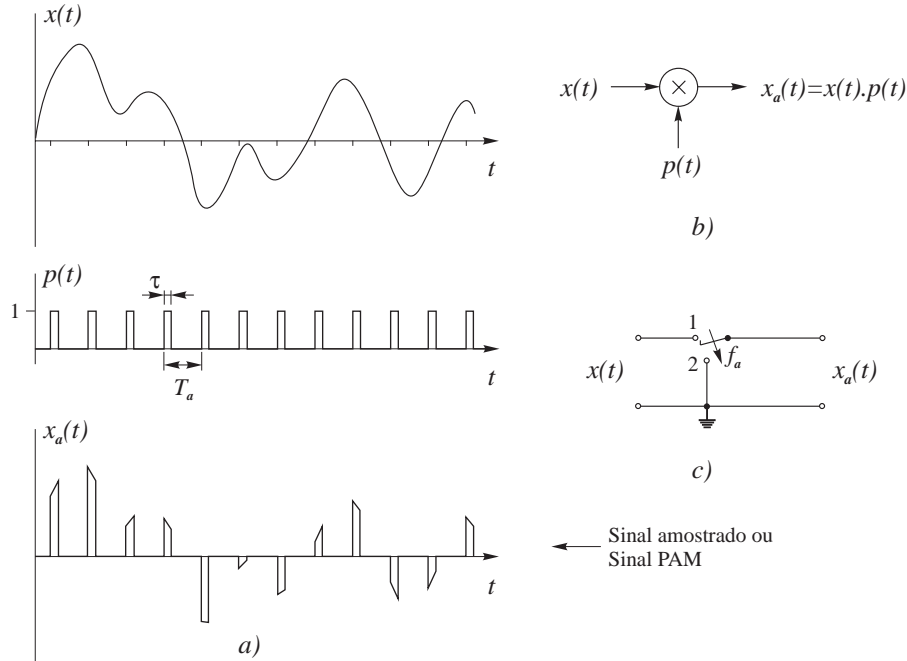


Figura 5.2: A operação de amostragem

Matematicamente, a operação de amostragem é descrita pelo produto de $x(t)$ por uma função $p(t)$ que é uma sequência de pulsos rectangulares com período de repetição igual a T_a e duração muito pequena. A figura 5.2 b) representa esta operação de multiplicação. Na realidade, a duração dos pulsos, τ , corresponde ao tempo em que o comutador está na posição 1 quando nela "toca". Assim, a operação de amostragem acaba por ser também uma operação de *modulação de pulso em amplitude* tendo por *portadora* a sequência de pulsos $p(t)$. Por esta razão o sinal amostrado $x_a(t)$ também se designa por sinal PAM (Pulse Amplitude Modulation).

A operação de amostragem, portanto, substitui os valores de amplitude que um sinal analógico toma ao longo de todo o tempo por apenas alguns desses valores, recolhidos em instantes determinados pela frequência de

amostragem, como a figura 5.1 a) mostra.

Coloca-se agora a seguinte questão fundamental: Serão as *amostras* do sinal suficientes para descrever completamente o sinal original e, se assim fôr, como é que se pode recuperar o sinal original $x(t)$ a partir apenas das amostras, ou seja, do sinal amostrado, $x_a(t)$?

A resposta a esta questão é dada pelo Teorema da Amostragem que a analisa no domínio da frequência de uma forma muito simples. Basicamente o teorema determina qual o espectro do sinal depois de amostrado, $X_a(f)$ e a seguir verifica se é possível *isolar* desse espectro o espectro do sinal original, $X(f)$. Ser-se capaz de isolar este espectro significa, na realidade, ser-se capaz de recuperar o sinal analógico original.

Este Teorema é de uma importância capital no processo de digitalização dos sinais de informação produzidos originalmente sob a forma analógica ao permitir, como se referiu no início do capítulo, lidar apenas com um conjunto de valores do sinal em número finito e, realce-se, sem qualquer perda da informação que o sinal original contenha.

5.1.2 Teorema da Amostragem

Seja $x_a(t)$ o sinal obtido por amostragem do sinal analógico $x(t)$ cujo espectro $X(f)$ é limitado à banda de frequências $[-B, +B]$. O sinal amostrado é definido pela operação

$$x_a(t) = x(t) \cdot p(t) \quad (5.1)$$

em que $p(t)$ é uma sequência de pulsos rectangulares de amplitude unitária, de período T_a e duração dos rectangulos muito pequena, $\tau \ll 1$.

Como $p(t)$ é periódica pode ser representada pela sua série de Fourier

$$p(t) = \sum_{n=-\infty}^{+\infty} C(n f_a) e^{j2\pi n f_a t} \quad (5.2)$$

onde $f_a = 1/T_a$ é a frequência de amostragem e os coeficientes da série podem ser obtidos a partir da definição dando

$$\begin{aligned} C(n f_a) &= \frac{1}{T_a} \int_{-T_a/2}^{+T_a/2} p(t) e^{-j2\pi n f_a t} dt \\ C(n f_a) &= f_a \tau \operatorname{sinc}(n f_a \tau) \end{aligned} \quad (5.3)$$

Calculemos agora o espectro do sinal amostrado $X_a(f)$

$$\begin{aligned}
 X_a(f) &= \mathcal{F}[x(t) \cdot p(t)] \\
 X_a(f) &= \int_{-\infty}^{+\infty} \left[x(t) \sum_{n=-\infty}^{+\infty} C(n f_a) e^{j2\pi n f_a t} \right] e^{-j2\pi f t} dt \\
 X_a(f) &= \sum_{n=-\infty}^{+\infty} C(n f_a) \int_{-\infty}^{+\infty} [x(t) e^{j2\pi n f_a t}] e^{-j2\pi f t} dt \\
 X_a(f) &= \sum_{n=-\infty}^{+\infty} C(n f_a) \int_{-\infty}^{+\infty} x(t) e^{-j2\pi(f - n f_a)t} dt \\
 X_a(f) &= \sum_{n=-\infty}^{+\infty} C(n f_a) X(f - n f_a)
 \end{aligned} \tag{5.4}$$

A última destas relações (5.4) diz que o espectro do sinal amostrado, $X_a(f)$ é a soma de infinitas réplicas do espectro $X(f)$ do sinal original cada uma delas deslocada na frequência de um múltiplo inteiro $n = 0, \pm 1, \pm 2, \dots$ da frequência de amostragem, incluindo $n = 0$, ou seja, o espectro do sinal original. Cada réplica vem multiplicada por um factor de escala em amplitude de valor igual a $C(n f_a)$ mas como se está a supôr que τ é muito pequeno, a equação 5.3 dá $C(n f_a)$ praticamente constante.

Se a frequência de amostragem f_a for igual ou superior ao dobro da largura de banda B do sinal analógico $x(t)$ então as referidas réplicas do espectro de $x(t)$ encontram-se separadas umas das outras e o espectro de amplitude do sinal amostrado definido pela equação 5.4 tem a representação gráfica que se mostra na figura 5.3 Nestas condições, o sinal original pode efecti-

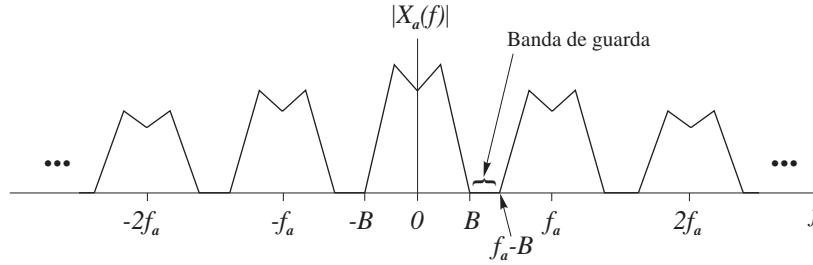


Figura 5.3: Espectro de amplitude do sinal amostrado $x_a(t)$, com $f_a > B$

vamente ser recuperado do sinal amostrado $x_a(t)$ bastando para isso que este seja filtrado por um filtro passa-baixo de largura de banda $B_T = B$.

Deste modo, todas as réplicas serão eliminadas excepto a réplica central a que corresponde $n = 0$ que é exactamente o espectro do sinal analógico original $x(t)$. A figura 5.4 ilustra a operação de filtragem passa-baixo para a recuperação do sinal original. Matematicamente, esta operação de

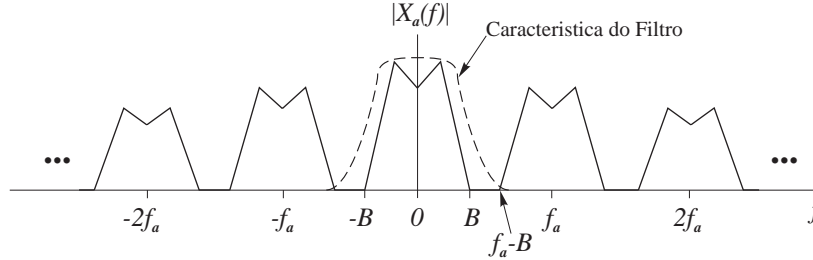


Figura 5.4: Recuperação do sinal original por filtragem do sinal amostrado

filtragem pode ser descrita pelo produto

$$X(f) = X_a(f) \cdot \Pi\left(\frac{f}{2B}\right) \quad (5.5)$$

isto é, pela multiplicação do espectro $X_a(f)$ pela função de transferência $\Pi(f/2B)$ do filtro passa-baixo ideal com largura de banda $B_T = B$. Conclui-se então que

Teorema 5.1 (Teorema da Amostragem) *Um sinal de espectro limitado à banda de frequências $[0, B]$ fica completamente definido pelas suas amostras desde que recolhidas a uma frequência igual ou superior a $2B$,*

$$f_a \geq 2B \quad (5.6)$$

podendo o sinal ser recuperado a partir das amostras por filtragem passa-baixo com largura de banda do filtro B_T igual a B Hz.

Se a frequência de amostragem for inferior àquele valor (não se verificando a desigualdade (5.6)) então, regra geral, as réplicas do espectro do sinal original $x(t)$ ficarão sobrepostas não sendo possível recuperar o sinal original. Esta situação está ilustrada na figura 5.5 que representa o espectro do sinal *sub*-amostrado não havendo nenhum filtro que consiga isolar nenhuma das réplicas. Caso se tente recuperar o sinal com o filtro anterior acontece um fenómeno designado por *aliasing*¹ em que parte da

¹lê-se *eliæsing* em inglês, do latim *alias*, f.p. de *alium* que significa "outra vez"

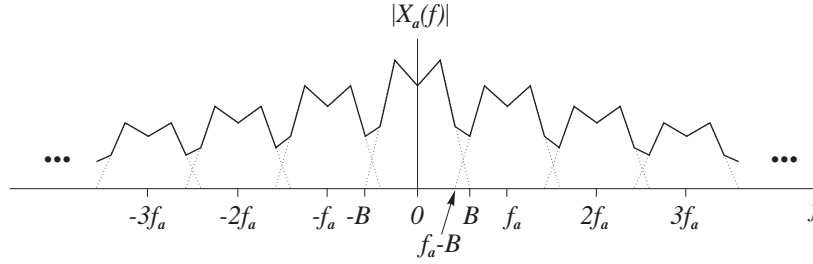


Figura 5.5: Fenómeno de *aliasing* espectral quando $f_a < 2B$

banda inferior aparece duas vezes dando origem a um sinal com maiores componentes de alta frequência e portanto diferente do original.

5.1.3 Amostragem na prática

Embora o Teorema da Amostragem garanta, em teoria, que um sinal analógico possa ser totalmente recuperado a partir das suas amostras se estas tiverem sido obtidas a uma frequência $f_a = 2B$, bastando para tal que o sinal possua um espectro estritamente limitado a uma banda de largura B Hz, na prática isto não se consegue por duas razões:

- (i) Não se conseguem realizar filtros ideais e
- (ii) Os sinais na prática não possuem espectros de banda limitada. Por serem limitados no tempo, são ilimitados na frequência.

pelo que a frequência de amostragem utilizada é normalmente

$$f_a > 2B \quad (5.7)$$

A primeira destas situações está ilustrada na figura 5.4 em que a característica do filtro é visivelmente não-ideal havendo a necessidade de deixar livre uma *banda de guarda* como se indica na figura 5.3.

A segunda mostra-se na figura 5.6 a) na qual, embora o sinal tenha uma largura de banda de B Hz, o seu espectro estende-se para além desta banda com componentes não nulas, isto é, $|X(f_1)| \neq 0$ para $|f_1| > B$.

Depois de amostrado mesmo a uma frequência $f_a > 2B$, ocorre o fenómeno de *aliasing* como se pode ver na figura 5.6 b). Nestas condições, a

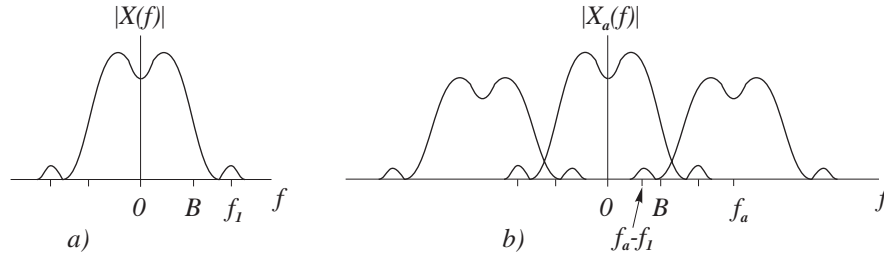


Figura 5.6: *Aliasing* espectral dos sinais da prática mesmo com $f_a > 2B$

única forma de evitar o *aliasing* é filtrar o sinal antes da amostragem limitando-se assim, forçadamente, a sua banda espectral ao intervalo $[-B, +B]$ como mostra a figura 5.7 a). A figura 5.7 b) mostra o espectro do sinal amostrado após a filtragem prévia, sem *aliasing*.

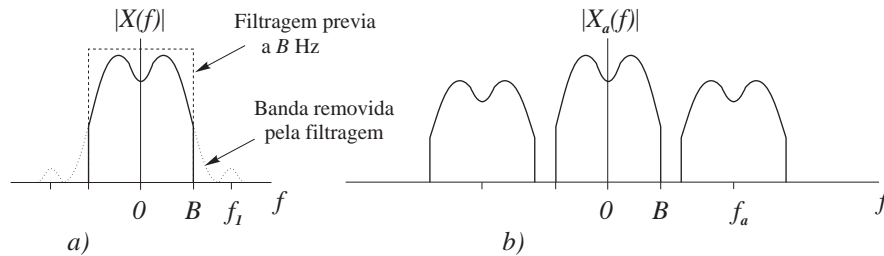


Figura 5.7: Filtragem prévia do sinal evitando o *aliasing* na amostragem

Na prática o sinal é amostrado por pulsos de duração praticamente nula $p_\delta(t)$. A amplitude instantânea da amostra resultante é depois *mantida* constante durante um certo tempo, o necessário para se completarem as operações subsequentes de digitalização, antes de voltar a zero.

O circuito da figura 5.8 b) modela este comportamento designado de *amostragem e manutenção*, ou *sample and hold*. O comutador I_2 está normalmente na posição 3. Quando o comutador I_1 "toca" na posição 1 o condensador é carregado *mantendo* aos seus terminais uma tensão igual à da amplitude da amostra. Ao fim de um certo tempo (antes da amostra seguinte) o comutador I_2 "toca" na posição 4 descarregando o condensador altura em que a tensão passa a zero. Os pulsos PAM resultantes possuem agora amplitude constante como se mostra na figura 5.8 a).

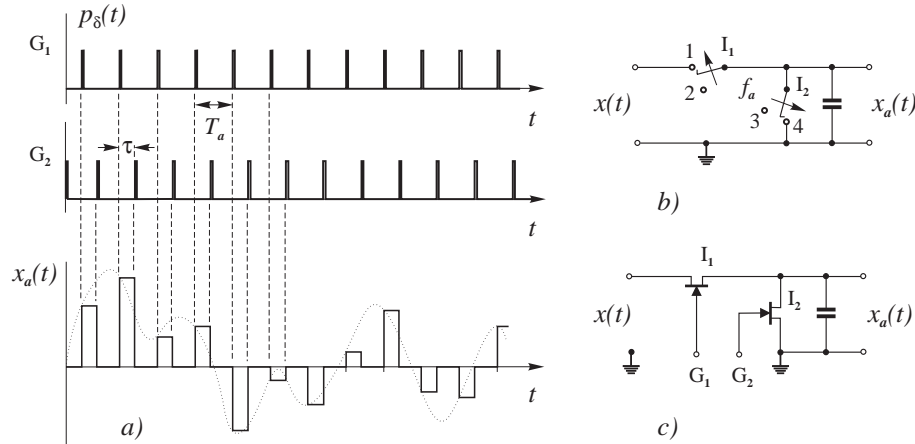


Figura 5.8: Amostragem e manutenção da amplitude (*sample and hold*)

Na prática não são utilizados comutadores mecânicos como as figuras 5.2 c) 5.8 b) sugerem. A comutação é feita por circuitos electrónicos, normalmente transistores, controlados por circuitos lógicos de temporização. A figura 5.8 c) mostra uma realização de um circuito *sample and hold* electrónico equivalente ao da figura 5.8 b). Os comutadores são transistores FET² cujas *gates* G_1 e G_2 são controladas por circuitos cuja temporização é a indicada na figura 5.8 a), basicamente dois relógios à mesma frequência com G_2 atrasado em relação a G_1 de τ s.

5.2 Quantização

Por forma a codificar as amplitudes das amostras em números, aquelas só podem tomar um número finito de valores dentro do intervalo contínuo de variação do sinal. Este processo de discretização das amplitudes designa-se por *quantização*. Ao contrário da operação de amostragem, neste processo perde-se alguma informação sobre o sinal – que não é possível recuperar aquando da sua reconstituição – sob a forma de um *erro de quantização* o qual, para efeitos práticos, é equivalente a um ruído pois pode ser considerado uma amplitude que se adiciona ou se subtrai ao sinal original a fim de o levar a um dos níveis quânticos.

²Field Effect Transistor

5.2.1 Quantização uniforme

Um *quantizador* divide o intervalo de variação das amplitudes das amostras, $x_a(t)$, em q intervalos, ou níveis quânticos, e aproxima-as ao nível mais próximo $x_q(t)$. Se os níveis quânticos estão igualmente espaçados entre si a quantização diz-se *uniforme*.

Por forma a visualizar a relação entre $x_a(t)$ e $x_q(t)$ considere-se que o sinal analógico foi normalizado a $|x(t)| \leq 1$. O quantizador uniforme sub-divide o intervalo $[-1, +1]$ das amplitudes de pico em q degraus de amplitude $2/q$ como mostra a figura 5.9 a). Os níveis quânticos são então $\pm 1/q, \pm 3/q, \dots, \pm (q-1)/q$ no caso usual em que q é um inteiro par. Uma amostra quantizada de amplitude $x_q(t_1) = 5/q$ resulta de qualquer amostra cuja amplitude esteja situada no intervalo $4/q < x_a(t_1) < 6/q$. O número de níveis quânticos q a utilizar em cada situação depende da

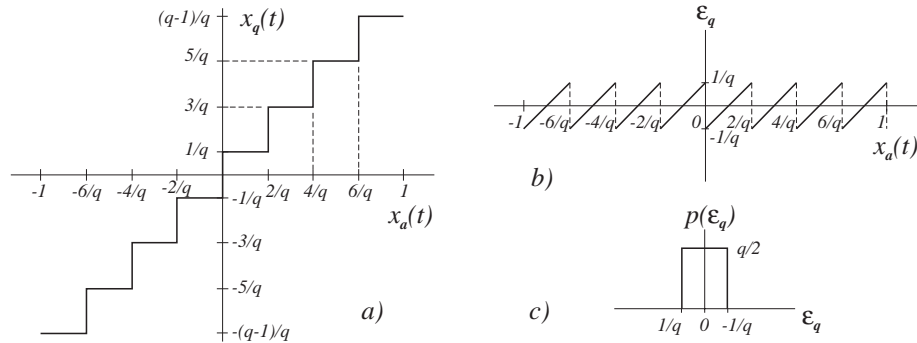


Figura 5.9: Funções característica e de erro da quantização uniforme

precisão com que se deseja representar digitalmente o sinal, ou seja, do número de dígitos, k , a utilizar. Regra geral a codificação é binária pelo que o valor de q mais apropriado será o de uma potência inteira de 2.

$$q = 2^k \quad k = \log_2 q \quad (5.8)$$

A escolha de um valor para q determina também o erro ou ruído de quantização que será introduzido.

5.2.2 Ruído de quantização

O erro de quantização numa amostra, ϵ_q , será a diferença

$$\epsilon_q = x_a(t) - x_q(t) \quad (5.9)$$

Como se pode verificar na figura 5.9 b) o erro da quantização uniforme varia no intervalo $-1/q < \epsilon_q < +1/q$ e dado tratar-se de um intervalo normalmente muito pequeno, no qual se pode considerar que as amostras do sinal $x_a(t)$ se distribuem uniformemente, pode supor-se também que possui média nula e se distribui uniformemente nesse intervalo, tal como se mostra na figura 5.9 c) que representa a sua função de distribuição de probabilidade. Supondo que os valores do erro ϵ_q são independentes entre si, o seu valor quadrático médio $\overline{\epsilon_q^2}$ representa a potência do ruído de quantização, N_q . Assim, a potência do ruído de quantização é

$$N_q = \sigma_q^2 = \overline{\epsilon_q^2} = \int_{-1/q}^{1/q} \epsilon_q^2 p(\epsilon_q) d\epsilon_q = \frac{q}{2} \int_{-1/q}^{1/q} \epsilon_q^2 d\epsilon_q = \frac{1}{3q^2} \quad (5.10)$$

que mostra que o ruído de quantização decresce quando o número de níveis de quantização aumenta. A razão entre a potência do sinal e a potência do ruído de quantização é

$$\begin{aligned} \frac{S}{N_q} &= 3q^2 S \\ \frac{S}{N_q} &\leq 3q^2 \end{aligned} \quad (5.11)$$

pois como estamos a considerar $|x(t)| \leq 1$ também a sua potência média será $S \leq 1$. O resultado (5.11) é usualmente expresso em decibéis e em função do número de dígitos binários utilizados nos números que representam as amplitudes das amostras quantizadas, ou seja,

$$\begin{aligned} \left(\frac{S}{N_q} \right)_{dB} &\leq 10 \log_{10}(3 \times 2^{2k}) \text{ dB} \\ \left(\frac{S}{N_q} \right)_{dB} &\leq 4.8 + 6.0 k \text{ dB} \end{aligned} \quad (5.12)$$

No telefone digital utilizam-se oito dígitos para codificar cada amostra do sinal de voz, pelo que $k = 8$ e portanto $(S/N_q)_{dB} \leq 52.8 \text{ dB}$ o que significa que a potência do ruído devido à quantização é cerca de 200 mil vezes inferior à do sinal o que representa uma fidelidade muito boa quando o objectivo é manter a inteligibilidade e a identificação do interlocutor.

5.2.3 Quantização não-uniforme

Verifica-se que os sinais analógicos de informação possuem elevados valores de crista, isto é, ao longo do tempo a sua amplitude situa-se mais frequentemente na zona das amplitudes baixas do que na zona das amplitudes altas, facto que pode ser descrito pela relação $x_{max} \gg x_{ef} = \sqrt{x^2}$. Isto significa que a densidade de probabilidade das amplitudes $p(x)$ é maior na zona das amplitudes perto do valor médio do sinal (normalmente o zero) decrescendo até aos valores de pico como a figura 5.10 a) mostra. Uma análise semelhante à que foi feita na secção anterior

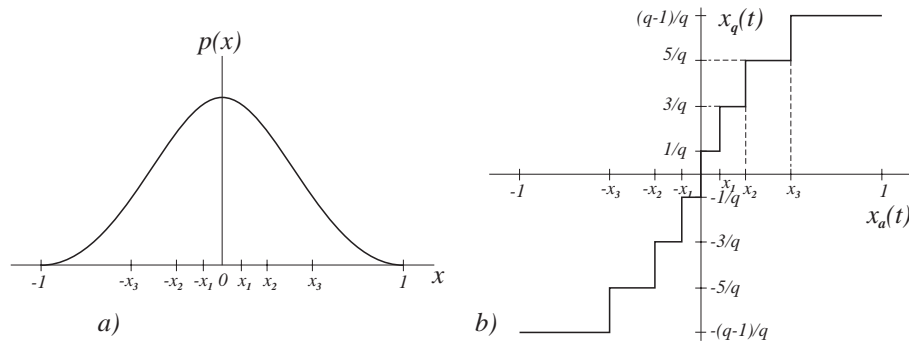


Figura 5.10: Fdp dos sinais e característica do quantizador não-uniforme

mostra que o valor médio do erro de quantização e portanto também a potência do ruído de quantização é mínima se os limiares de transição $\pm x_1, \pm x_2, \dots, \pm x_{q/2-1}$ dos níveis quânticos estiverem menos espaçados para as amplitudes mais baixas e mais espaçados nas amplitudes mais altas o que significa uma quantização *não-uniforme* ou *não-linear* tal como se exemplifica na figura 5.10 b).

Porém, a realização de um quantizador não-uniforme é mais complexa e dispendiosa do que a dum quantizador uniforme. O que se faz então na prática é utilizar um quantizador uniforme após uma *compressão* não-linear do sinal, em que as características do compressor são determinadas a partir de estudos experimentais com sinais representativos.

A partir de considerandos teórico-práticos chegou-se à conclusão que a característica do *compressor* que melhor *uniformiza* a densidade de probabilidade das amplitudes dos sinais que aparecem na prática (em especial os sinais de audio) é linear a partir da amplitude zero e até um certo valor

($1/A$) das amplitudes e depois logaritmica até ao seu valor máximo de acordo com a lei

$$y = \begin{cases} \frac{Ax}{1 + \ln A} & \text{para } |x| \leq \frac{1}{A} \\ \frac{1 + \ln Ax}{1 + \ln A} & \text{para } \frac{1}{A} < |x| \leq 1 \end{cases} \quad (5.13)$$

designada por *lei-A* e representada graficamente na figura 5.11 a). Assim, na prática, os sinais $x(t)$ são primeiramente comprimidos segundo esta lei, dando um sinal $y(t)$, o qual é depois amostrado e quantizado uniformemente dando o sinal $y_q(t)$. O sinal comprimido $y(t)$ é recuperado das suas amostras quantizadas $y_q(t)$ por filtragem (a menos do erro de quantização) o qual é depois *expandido* por multiplicação (analgica) pela função inversa de (5.13).

O conjunto dos procedimentos de *compressão* e *expansão* é designado de *comparing*³.

Nos casos práticos a lei-A de *comparing* é aproximada por segmentos lineares. O declive de cada segmento representa a razão de compressão no correspondente intervalo de variação do sinal. Para a quantização a $q = 256$ níveis (8 bits) foi adoptada como norma uma aproximação poligonal de 13 segmentos em que o segmento central possui uma razão de compressão de 16:1 e as dos restantes estão em progressão geométrica de razão $1/2$ como mostra a figura 5.11 b). O sinal comprimido é depois

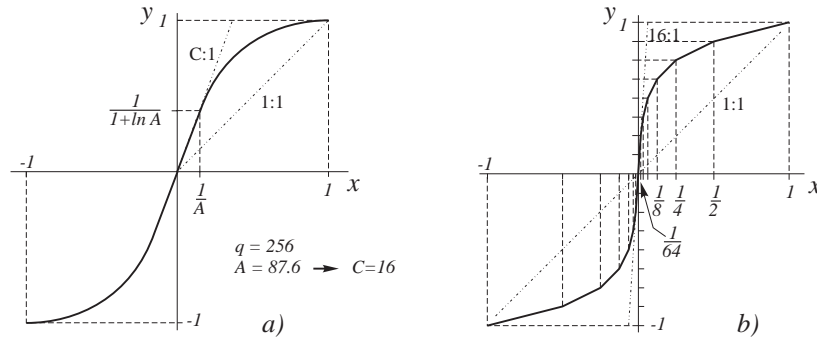


Figura 5.11: Compressor de lei-A e sua aproximação com 13 segmentos quantizado uniformemente a 256 níveis.

³conjunção das designações em língua inglesa compressing e expanding

Se um sinal analógico possuir uma função densidade de probabilidade de amplitude não-uniforme (do tipo da da figura 5.10 a)) e for quantizado uniformemente, os códigos binários das amostras irão conter mais zeros do que uns, visto que a amplitude do sinal será mais frequentemente pequena do que grande, uma situação a evitar na transmissão, visto que depois de serializada, a sequência binária fica composta por dígitos binários de igual valor seguidos, podendo dificultar a sincronização. Por outro lado, como se verá quando se tratar a teoria da informação, uma ocorrência mais frequente de certos valores do que de outros, significa que os dígitos do código contêm uma quantidade de informação por dígito menor do que a que seria teoricamente possível com o número de dígitos utilizados para codificar as amostras. Ver-se-á também que uma alternativa à compressão antes da quantização é a compressão depois da quantização, como parte integrante da codificação.

Nos Estados Unidos da América a lei de quantização logarítmica adoptada difere da lei-A e é designada de *lei- μ* expressa na seguinte forma

$$y = \frac{\ln(1 + \mu x)}{\ln(1 + \mu)} \quad (5.14)$$

A característica de compressão da *lei- μ* possui um comportamento assintótico sendo

- linear para $x \ll 1/\mu$ e
- logaritmico para $1/\mu \ll x \ll 1$

Ambas as leis são objecto de normalização na Recomendação G.711 da ITU⁴, a *lei-A* com $A = 87.6$ e a *lei- μ* com $\mu = 255$, e elas são de facto muito semelhantes. Contudo, a diferença das suas taxas de compressão na origem, $C_A = 16$ e $C_\mu \approx 46$ fazem com que a razão sinal-ruído de quantização seja melhor na *lei- μ* do que na *lei-A* para sinais fracos. E, também, a diferença entre as duas leis é suficientemente significativa de modo a tornar um compressor-A incompatível com um expansor- μ e vice-versa.

⁴International Telecommunications Union, organismo de normalização internacional em telecomunicações com sede em Geneve

5.3 Conversão analógico a digital

Uma vez quantizadas, as amostras constituem ainda uma amplitude analógica mas dado que se encontram discretizadas a q valores podem desde já ser consideradas valores digitais, isto é, representam números com apenas um dígito da base q . Na prática, porém, utilizam-se representações na base 2. É a conversão analógico-digital que executa esta conversão da base de numeração. Existem várias técnicas de conversão e aqui abordar-se-ão apenas duas, a Modulação de Impulso Codificada (MIC), ou PCM, e a Modulação Delta (MD), as mais usadas em comunicações.

5.3.1 Modulação de Impulso Codificada, MIC ou PCM

Os números binários resultantes da codificação das amostras quantizadas, com os dígitos serializados no tempo constituem uma sequência designada por *Modulação de Impulso Codificada* (MIC) ou, na designação inglesa, *Pulse Code Modulation* (PCM).

A figura 5.12 a) representa um conversor analógico-digital para $q = 8$ níveis de quantização uniforme. O quantizador está normalmente integrado no conversor e consiste no divisor de tensão constituído pelas resistências de $R \Omega$. Uma tensão de referência de 1 V é dividida em quatro intervalos iguais com limites em $1/4$, $1/2$ e $3/4$ V os quais são comparados com a amplitude das amostras rectificadas. Os comparadores produzem uma saída de valor lógico 1 sempre que a tensão da amostra (+) for superior à tensão de referência correspondente (–) e o valor lógico 0 em caso contrário. Assim, consoante o valor absoluto da amplitude duma amostra o conjunto das saídas dos comparadores mostrará os valores (000), (001), (011) ou (111) representativos do nível do sinal e que são depois convertidos em números de dois dígitos binários, a_0 e a_1 pelo bloco designado de *codificador binário* que realiza as funções de lógica combinatória apropriadas. Neste conversor A/D as amostras são rectificadas pelo facto dos níveis quânticos serem simétricos. Basta um comparador com zero da amostra $x_a(t)$ para determinar o sinal algébrico da amplitude, representado pelo valor do dígito a_2 . Se assim não fosse, o conversor deveria ter um divisor de tensão de oito intervalos e sete comparadores. O número binário resultante

$$D = (a_2 a_1 a_0) = (-1)^{a_2} (2a_1 + a_0) \quad (5.15)$$

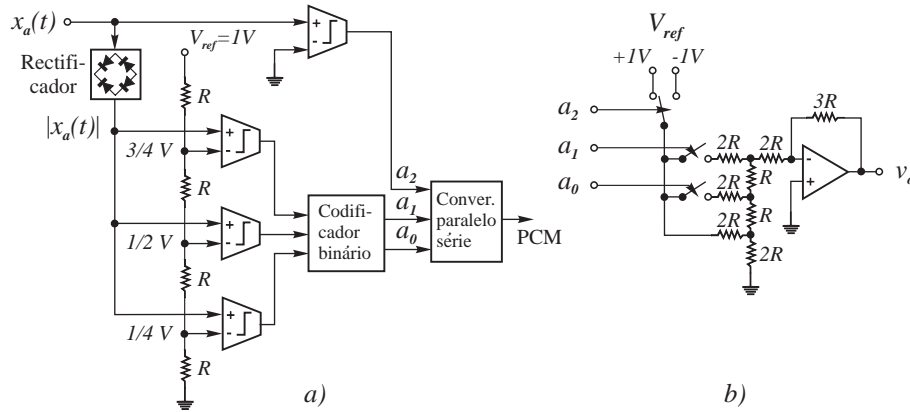


Figura 5.12: Conversão analógico-digital e digital-analógico

é depois serializado por um *conversor paralelo-série*, normalmente constituído por um registo de deslocamento, produzindo assim a sequência PCM. O código binário produzido por este conversor A/D é designado de código *dobrado* pois os números são simétricos sendo representados sob a forma de *sinal algébrico* e *mantissa inteira*. A tabela 5.1 mostra a correspondência entre os valores das amostras e o respectivo código PCM. A figura 5.13 exemplifica a saída PCM. Por cada amostra i é produzido

Tabela 5.1: Código binário PCM do conversor do exemplo

$x_a(t)$	$x_q(t)$	a_2	a_1	a_0
1	7/8	1	1	1
3/4	5/8	1	1	0
1/2	3/8	1	0	1
1/4	1/8	1	0	0
0	1/8	1	0	0
-1/4	-1/8	0	0	0
-1/2	-3/8	0	0	1
-3/4	-5/8	0	1	0
-1	-7/8	0	1	1

um número D_i de três bits. Donde se conclui que o ritmo binário de um

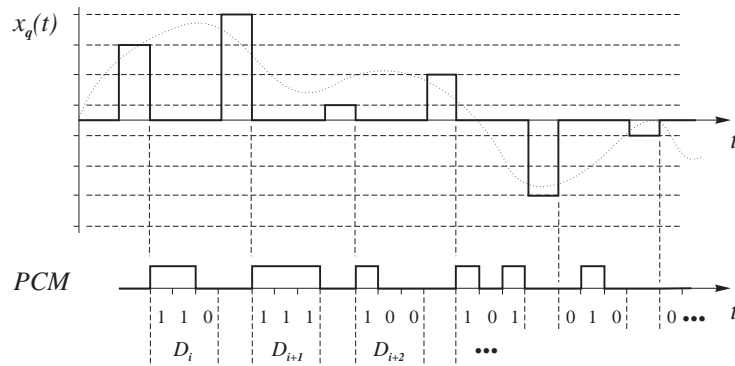


Figura 5.13: Sinal PCM resultante da codificação das amostras

canal PCM codificado a k bits por amostra é

$$r_c = k f_a \geq 2kB \quad (5.16)$$

e que a largura de banda exigida dum canal para transmissão digital é k vezes maior do que se a transmissão fosse feita analogicamente

$$B_T \geq \frac{r_c}{2} \geq kB \quad (5.17)$$

Na figura 5.12 b) representa-se o conversor digital-analógico, inverso deste, que recebe como entrada o número digital depois de paralelizado e produz uma tensão analógica de valor

$$v_0 = \frac{(-1)^{a_2}}{8} (4a_1 + 2a_0 + 1) \quad (5.18)$$

que reproduz a amplitude da amostra quantizada, isto é, $x_q(t) = v_0(t)$. Este conversor é constituído por uma malha de resistências em *escada* arranjadas de modo a que quando o valor de um dos dígitos binários é igual a 1 é introduzida no ponto da malha à entrada do amplificador operacional, uma corrente proporcional à posição desse dígito no número binário. A malha adiciona algebricamente as contribuições de cada um dos dígitos somando analogicamente as respectivas correntes efectuando exactamente as operações indicadas na equação 5.18. A malha está concebida por forma a exigir apenas resistências de 3 valores distintos (R , $2R$, e $3R$) qualquer que seja o número de dígitos do conversor.

O conversor da figura 5.12 a) que serviu de exemplo a esta discussão utiliza o método da comparação simultânea. É o método mais rápido de conversão dado que todos os dígitos são determinados ao mesmo tempo. Exige 2^{k-1} comparadores, isto é, por cada dígito por amostra o número de comparadores duplica o que, mesmo para valores moderados de k , representa alguma complexidade e um custo significativo. Existem porém outros métodos de conversão como por exemplo os que recorrem a *aproximações sucessivas* em que os dígitos do código vão sendo determinados sucessivamente ao ritmo de um relógio, ou os métodos *diferenciais* em que o que é codificado é a diferença de amplitude entre amostras sucessivas e não as amplitudes das amostras em si. Nos métodos *preditivos* o conversor *estima*, por extrapolação, o valor da amostra seguinte e quando ela chega ao conversor este determina o erro da *estimativa* e é este valor que ele codifica e transmite.

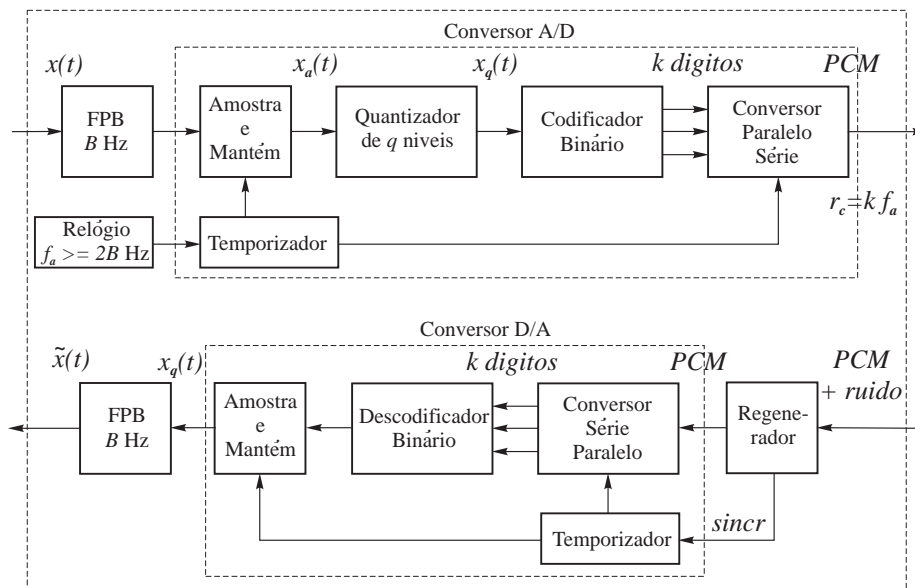


Figura 5.14: Esquema simplificado de um Codec PCM

Interessa notar que tanto o conversor A/D como o D/A normalmente integram num único bloco funcional as diferentes fases do processo de digitalização até aqui referidas (amostragem, quantização e codificação) bem como a temporização que as regulam, já que na maior parte das realizações físicas não é possível nem prático separá-las. Os conversores A/D

e D/A são normalmente fabricados aos pares e o conjunto é denominado um *Codec*⁵. O transmissor do *codec* é o conversor A/D e o receptor é o conversor D/A cada um deles ligado a um canal distinto do sistema de transmissão digital. A figura 5.14 ilustra estes conceitos.

A representação digital de alguns sinais analógicos mais utilizados em telecomunicações tais como a telefonia (voz), a música e o vídeo, têm sido normalizadas pela ITU permitindo a compatibilidade entre os diversos sistemas que são realizados industrialmente. A tabela 5.2 indica os parâmetros normalizados para alguns sinais.

Tabela 5.2: Valores de digitalização PCM de alguns sinais típicos

Sinais telefônicos (ITU, Recomendação G.711)	
Frequência de amostragem:	$f_a = 8 \text{ KHz}$
Quantização:	não-uniforme a $q = 256$ níveis
Palavra PCM:	$k = 8 \text{ bits}$
Ritmo binário (um canal):	$r_b = k f_a = 64 \text{ Kbps}$
Lei de quantização Europeia:	compressão digital segundo a <i>lei-A</i> , com 13 segmentos.
Código:	binário dobrado com inversão dos bits de ordem par.
Lei de quantização Americana:	compressão digital segundo a <i>lei-μ</i> , com 15 segmentos.
Código:	binário dobrado com inversão de todos os bits excepto o primeiro (bit do sinal)
Sinais de Vídeo (Televisão)	
Frequência de amostragem:	$f_a = 13.3 \text{ MHz}$
Quantização:	uniforme com $k = 8$ ou 9 bits
Gravação de Música	
Frequência de amostragem:	$f_a = 44.1 \text{ KHz}$
Quantização:	uniforme com $k = 16 \text{ bits}$
Ritmo binário:	$r_b \approx 0.7 \text{ Mbps}$

⁵conjunção das designações em língua inglesa Coder e decoder

Tabela 5.2 (cont.)

Transmissão de Música	
Frequência de amostragem:	$f_a = 32$ KHz
Quantização:	uniforme com $k = 14$ bits
Ritmo binário:	$r_b = 448$ Kbps ou
Quantização:	não-uniforme com $k = 12$ bits
Ritmo binário:	$r_b = 384$ Kbps
Lei de quantização:	lei-A com 5 segmentos ou
Quantização:	não-uniforme com $k = 10$ bits
Ritmo binário:	$r_b = 320$ Kbps
Lei de quantização:	lei-A com 13 segmentos

5.3.2 Ruído em PCM

Devido ao ruído no canal de transmissão os bits que constituem o código binário de cada amostra podem ser corrompidos e mudar de valor. Um erro no bit da posição m tem como consequência que o nível quântico em que a correspondente amostra irá ser decodificada estará desviado de uma quantidade $\epsilon_m = \pm(2/q) \cdot 2^m$ em relação ao que seria o correcto. ϵ_m é uma variável aleatória que representa a amplitude do ruído que se manifesta na posição m . A potência média deste ruído é dada pelo valor quadrático médio (para todas as k possíveis posições):

$$\overline{\epsilon_m^2} = \frac{1}{k} \sum_{m=0}^{k-1} \left(\frac{2}{q} 2^m \right)^2 = \frac{4}{kq^2} \sum_{m=0}^{k-1} 4^m = \frac{4}{kq^2} \frac{4^k - 1}{3} = \frac{4}{3k} \frac{q^2 - 1}{q^2} \approx \frac{4}{3k} \quad (5.19)$$

potência⁶ que está associada sómente às amostras que sofreram erro e que constituem apenas uma fracção da totalidade das amostras transmitidas. Uma amostra estará errada se na totalidade dos k bits que a constituem ocorrerem um ou mais bits em erro. A probabilidade disto acontecer dá então a referida fracção das amostras erradas. Esta probabilidade é dada pela lei (distribuição) binomial

$$\text{Prob(amostra errada)} = \text{Prob(bits errados} \geq 1)$$

⁶a soma dos k primeiros termos de uma série geométrica de razão z é $S_k = \frac{1-z^k}{1-z}$ e considerou-se que $q^2 \gg 1$ e portanto $\frac{q^2-1}{q^2} \approx 1$

$$= \sum_{i=1}^k C_i^k P_e^i (1 - P_e)^{k-i} \approx k P_e \quad (5.20)$$

onde P_e é a probabilidade de erro por bit, referida no capítulo 4, que depende da razão de potências sinal-ruído (S/N) no canal de transmissão e do código de linha utilizado (polar, unipolar, etc). Para o resultado obtido em (5.20) considerou-se a aproximação $P_e \ll 1$, o que é verdade nos sistemas de transmissão da prática, pelo que a soma é praticamente igual ao valor da primeira parcela o que por sua vez equivale a dizer que na prática as amostras erradas contêm apenas um erro⁷.

Portanto a potência do ruído (do erro) de descodificação é

$$N_d = k P_e \overline{\epsilon_m^2} = k P_e \frac{4}{3k} = \frac{4 P_e}{3} \quad (5.21)$$

e a potência total do ruído no destino será a soma deste ruído com o de quantização $N_q = 1/(3q^2)$ (equação 5.10) visto que estes dois ruídos são produzidos por processos essencialmente independentes:

$$N_D = N_d + N_q = \frac{4 P_e}{3} + \frac{1}{3q^2} = \frac{1 + 4q^2 P_e}{3q^2} \quad (5.22)$$

e

$$\left(\frac{S}{N}\right)_D \leq \frac{3q^2}{1 + 4q^2 P_e} \quad (5.23)$$

ou seja

$$\left(\frac{S}{N}\right)_D \leq \begin{cases} 3q^2 & \text{se } P_e \ll \frac{1}{4q^2} \\ \frac{3}{4P_e} & \text{se } P_e \gg \frac{1}{4q^2} \end{cases} \quad (5.24)$$

o que significa que o ruído de quantização é a componente dominante da qualidade da digitalização quando a P_e na transmissão (ou gravação) é pequena, mas o ruído de descodificação devido aos erros na transmissão (ou gravação) passa a ser o mais significativo quando a P_e é grande comparada com $1/(4q^2)$.

⁷ é fácil de verificar que a probabilidade de ocorrerem dois ou mais erros num bloco de k dígitos é muito inferior à probabilidade de ocorrer apenas um erro se a probabilidade de erro for muito inferior a 1

5.3.3 Modulação Delta

Consideremos um outro tipo de conversor muito utilizado na digitalização de sinais de informação, especialmente na gravação digital, designado de *Modulador Delta*. Para sinais de áudio, por exemplo, verifica-se que a qualidade obtida com o modulador delta é semelhante à obtida com PCM exigindo aquele um menor número de bits por amostra e o modulador delta é tão ou mais simples de realizar do que o codificador PCM.

O princípio de funcionamento do conversor por modulação delta é semelhante ao dos conversores por aproximações sucessivas que é o seguinte: Faz-se uma conversão A/D com um intervalo de quantização grande e toma-se o número binário resultante como primeira aproximação ao sinal digitalizado. Depois converte-se a analógico esta aproximação e compara-se o valor obtido com o valor do sinal, convertendo-se a digital a sua diferença. O processo repete-se até que a diferença seja inferior ao erro de quantização desejado. Cada uma das aproximações analógicas quantizadas ao sinal original constitui uma *estimativa* $\tilde{x}_q(t)$ do sinal analógico.

Modulação Delta Linear

A figura 5.15 representa duas versões simplificadas do modulador delta e a figura 5.16 as formas de onda em vários pontos da versão a) do circuito.

O sinal analógico $x(t)$ e a sua estimativa quantizada $\tilde{x}_q(t)$ são continuamente apresentados a um comparador cuja saída será o valor lógico 1 (representado por uma tensão V_o volts) ou o valor lógico 0 (representado por uma tensão de 0 volts) consoante o valor à entrada $+$ fôr, respectivamente, maior ou menor que o da entrada $-$ que é depois amostrado ao ritmo do relógio do conversor e a amplitude mantida durante todo o período de amostragem sendo dividida em duas tensões eléctricas simétricas que constituem os valores do erro quantizado $\epsilon_q(t)$:

$$\epsilon_q(t) = \begin{cases} +\Delta = +\frac{V_0}{2} & \text{se } x(t) \geq \tilde{x}_q(t) \\ -\Delta = -\frac{V_0}{2} & \text{se } x(t) < \tilde{x}_q(t) \end{cases} \quad (5.25)$$

É o sinal de erro quantizado $\epsilon_q(t)$, que constitui a digitalização do sinal analógico, que é transmitido.

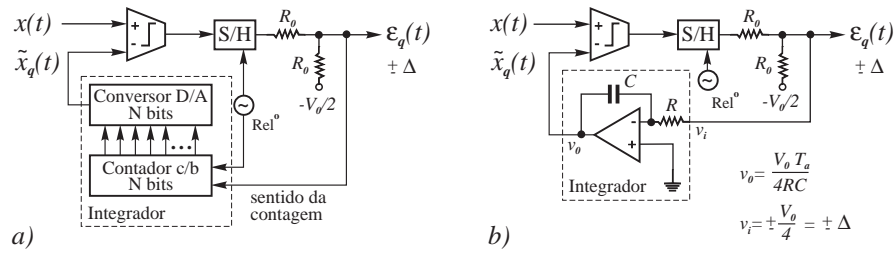


Figura 5.15: Circuitos para digitalização por modulação delta linear

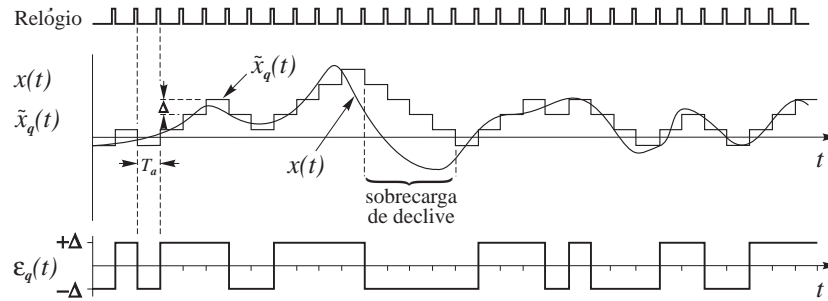


Figura 5.16: Formas de onda na modulação delta linear

Seja $t_i = i T_a$ um instante de amostragem. A estimativa $\tilde{x}_q(t_{i+1})$ é calculada da seguinte forma: Se $\epsilon(t_i) = +\Delta$, o que significa que $x(t_i) \geq \tilde{x}_q(t_i)$, então aumentar $\tilde{x}_q(t_i)$ de um *quantum* Δ ; se $\epsilon(t_i) = -\Delta$, o que significa que $x(t_i) < \tilde{x}_q(t_i)$, então diminuir $\tilde{x}_q(t_i)$ de um *quantum* Δ , ou seja, pela equação recursiva de diferenças:

$$\tilde{x}_q(t_{i+1}) = \tilde{x}_q(t_i) + \epsilon_q(t_i) \quad (5.26)$$

que representa uma integração no tempo. Esta integração pode ser realizada quer sob a forma digital através de um contador *para-cima-para-baixo*⁸ com comando de sentido de contagem cuja saída é convertida a analógico, conforme se ilustra na figura 5.15 a), quer directamante sob a forma analógica através de um integrador analógico constituído por um amplificador operacional devidamente realimentado como se ilustra na figura 5.15 b). Com um integrador analógico, como será fácil de verificar, $\tilde{x}_q(t)$ não tem a forma de degraus mas sim de rampas de declive

⁸up-down counter

$\pm V_0/4RC$ visto que a tensão de saída do amplificador, em função de T_a , é $v_0 = V_0 T_a / 4RC$.

É fácil de verificar que o receptor de um modulador delta, ou seja o correspondente conversor digital a analógico, terá de receber o erro quantizado $\epsilon_q(t)$ e produzir a estimativa $\tilde{x}_q(t)$ do sinal original que posteriormente será filtrada de modo a reproduzir o sinal analógico original. Portanto o receptor será constituído simplesmente por um circuito idêntico ao circuito de realimentação do modulador delta que se mostra na figura 5.17 para o caso digital.

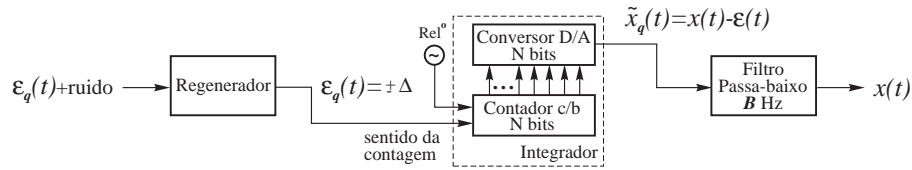


Figura 5.17: Receptor de Modulação Delta linear

A designação Modulador Delta (MD⁹) reflecte o facto de que cada amostra é codificada num único pulso de amplitude $+\Delta$ ou $-\Delta$. Mas também se pode considerar que $\epsilon_q(t)$ é uma *forma de onda binária* com um ritmo de sinalização de $r_b = f_a$, ou seja, *um bit por amostra*. Por esta razão a MD também é por vezes designada de *PCM de um-bit*. Estando-se a transmitir no canal ao ritmo $r_c = r_b$ então a largura de banda exigida pela MD é

$$B_T \geq \frac{r_c}{2} = \frac{f_a}{2} \geq B \quad (5.27)$$

consistentemente com as equações 5.16 e 5.17.

Consegue-se um bit por amostra porque se está a transmitir apenas o erro de estimativa e não o valor da amostra. Esta concepção é, no entanto, um tanto falaciosa porque na realidade é necessária uma frequência de amostragem superior de modo a resolver as situações de *sobrecarga de declive* que a figura 5.16 a) também ilustra. Uma sobrecarga de declive surge sempre que a variação da amplitude do sinal excede a velocidade com que a estimativa quantizada a pode acompanhar dando origem a erros superiores ao do *ruído granular* produzido pela própria quantização. De modo a se evitarem estas situações há que impôr a condição de que o

⁹DM, *Delta Modulator*

declive dos degraus (Δ/T_a) que constituem a estimativa do sinal quantizado nunca seja inferior ao declive máximo do sinal (obtido da derivada deste)

$$f_a \Delta \geq \left| \frac{dx(t)}{dt} \right|_{\max} \quad (5.28)$$

donde, para manter esta relação verdadeira e um Δ fixo, se terá que aumentar a frequência de amostragem f_a .

Exêmplo 5.1 *Considere-se que o sinal é sinusoidal $x(t) = A \cos(2\pi f_1 t)$. Tem-se $dx(t)/dt = -2\pi f_1 A \sin(2\pi f_1 t)$ e $|dx(t)/dt|_{\max} = 2\pi f_1 A \leq 2\pi B$, dado que $f_1 \leq B$ e que se tem estado a considerar que $A \leq 1$. A equação 5.28 exige então que*

$$f_a \geq 2\pi B/\Delta \gg \pi B \quad (5.29)$$

pois dever-se-á ter $\Delta \ll 2$ para que os níveis quânticos de $\tilde{x}_q(t)$ sejam pequenos comparados com a amplitude pico-a-pico do sinal $-1 \leq x(t) \leq +1$.

A determinação de uma expressão para a potência do ruído de quantização e consequentemente para a razão sinal-ruído na MD, tem de ter em consideração a condição imposta pela desigualdade 5.28 tornando-se um pouco mais complexa do que para o caso do PCM. Apenas se referirá que para atingir a mesma razão entre aquelas potências, isto é, para uma digitalização com a mesma qualidade, a MD exige uma maior largura de banda do que o PCM podendo ser entre duas a três vezes maior. O exêmplo seguinte dá uma ideia destes valores.

Exêmplo 5.2 *Um sinal de voz pré-filtrado a $B = 4$ KHz digitalizado em PCM binário com $k = 7$ bits por amostra produz um ritmo binário de $r_b = k f_a \geq 56$ kbps e exige uma banda de transmissão $B_T \geq 28$ KHz. O mesmo sinal digitalizado por Modulação Delta com a mesma qualidade (mesma razão sinal-ruído de granularidade) produz um ritmo binário $r_b = f_a \approx 2 \times 16 \times B = 128$ kbps exigindo portanto uma largura de banda de transmissão $B_T \geq 64$ KHz.*

Modulação Delta Adaptativa

Uma alternativa ao aumento da frequência de amostragem para reduzir a sobrecarga de declive, sem aumentar o ruído granular, é efectuar a quantização com *degraus de altura variável*. Pode verificar-se que quando

ocorre uma sobrecarga de declive os pulsos $\epsilon_q(t)$ têm a mesma polaridade, correspondendo a degraus sucessivos também com a mesma polaridade. Se sempre que se detectar que o degrau seguinte vai ter a mesma polaridade, se aumentar a amplitude do degrau, força-se $\tilde{x}_q(t)$ a aproximar-se mais rapidamente de $x(t)$.

Uma forma de realizar esta *adaptação* da altura do degrau a partir, por exemplo, do esquema do MD da figura 5.15 a), é actuar sobre a tensão de referência do conversor D/A através de um factor $1 < K < 2$ sempre que a polaridade se mantenha positiva e de um factor $1/K$ sempre que ela se mantenha negativa. Um circuito *controlador do degrau* fornece esta tensão de referência em função do sinal do degrau anterior conforme se representa no esquema da figura 5.18 que representa um *Modulador Delta Adaptativo* e que se rege agora pela equação de diferenças

$$\tilde{x}_q(t_{i+1}) = \tilde{x}_q(t_i) + g(t_i) \epsilon_q(t_i) \quad (5.30)$$

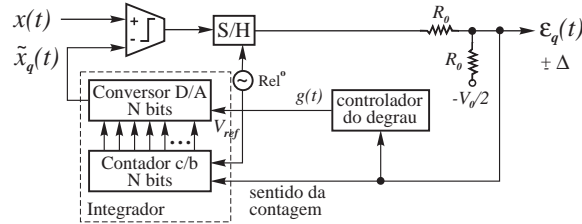


Figura 5.18: Modulador Delta Adaptativo

em que $g()$ representa um *ganho* definido por

$$g(t_i) = \begin{cases} g(t_{i-1}) \times K & \text{se } \epsilon_q(t_i) = \epsilon_q(t_{i-1}) \\ g(t_{i-1})/K & \text{se } \epsilon_q(t_i) \neq \epsilon_q(t_{i-1}) \end{cases} \quad (5.31)$$

Na figura 5.19 representa-se a digitalização por modulação delta adaptativa, com $K \approx 1.5$, da zona do sinal da figura 5.16 onde se verificava uma acentuada sobrecarga de declive. Pode observar-se que sempre que a polaridade se mantém igual à do degrau anterior os degraus vão aumentando de amplitude, a aproximação ao sinal analógico é mais rápida e consequentemente a sobrecarga de declive é menos acentuada e dura muito menos tempo.

Neste capítulo não se pretende estudar em detalhe a conversão A/D e D/A mas interessa referir que existem diversas outras técnicas que têm

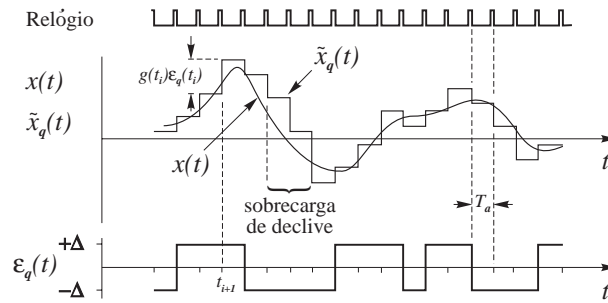


Figura 5.19: Formas de onda na modulação delta adaptativa

vindo a ser concebidos na procura de soluções que consigam reduzir ao mínimo o número de bits por amostra necessários à representação o mais fiel possível dos sinais analógicos sob a forma digital e, consequentemente, reduzir ou o ritmo binário na linha quando da sua transmissão, ou o espaço total de *memória* se o objectivo é a sua gravação digitalizada em suportes magnéticos ou ópticos como é o caso dos discos compactos (CD) ou dos ficheiros de computador em disco magnético. A tabela 5.3 compara, nestes aspectos, as características de diversos métodos utilizados na codificação de áudio com qualidade telefónica.

Tabela 5.3: Comparação entre vários métodos de codificação

Método de codificação	Sigla	f_a KHz	Bits por amostra	Ritmo Kbps
Modulação de Impulso Codificado	PCM	8	7–8	56–64
Mod de Imp Codif Diferencial	DPCM	8	4–6	32–48
Mod de Imp Codif Dif Adaptativa	ADPCM	8	3–4	24–32
Modulação Delta	DM	64–128	1	64–128
Modulação Delta Adaptativa	ADM	48–64	1	48–64
Codificação Preditiva Linear	LPC	0.04–0.1	≈ 80	3–8

5.4 Problemas

5.1 – Um sinal de banda de base, $x(t)$, tem por espectro

$$X(f) = \begin{cases} 1 - \frac{|f|}{200} & \text{para } |f| < 200 \\ 0 & \text{para } |f| \geq 200 \end{cases}$$

- a) Suponha que $x(t)$ é amostrado idealmente a $f_a = 300$ Hz. Esquematize o espectro de $x_a(t)$ para $|f| < 200$ Hz.
b) Repetir a) com $f_a = 400$ Hz.

5.2 – Um sinal $x(t) = 2 \cos(400\pi t) + 6 \cos(640\pi t)$ é amostrado idealmente a $f_a = 500$ Hz. Se o sinal amostrado passar através de um filtro passa-baixo ideal com frequência de corte de 400 Hz, que componentes de frequência aparecem à saída do filtro?

5.3 – Um sinal de banda passante possuindo o espectro de amplitude da figura 5.20 é amostrado idealmente. Esquematize o espectro do sinal amostrado quando $f_a = 20, 30$ e 40 Hz. Para cada um dos casos,

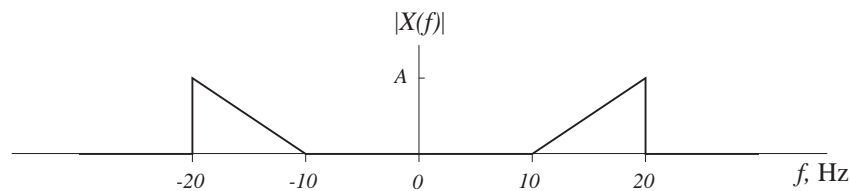


Figura 5.20:

verifique se o sinal analógico original pode ser recuperado a partir das amostras e, caso afirmativo, mostre como.

5.4 – A figura 5.21 mostra a função de densidade de probabilidade $p(x)$ das amplitudes das amostras de um sinal analógico $x(t)$. Defina um quantizador uniforme de quatro níveis para este sinal e calcule o valor da razão entre as potências do sinal e do ruído de quantização.

5.5 – a) Defina um quantizador não-uniforme de quatro níveis de erro quadrático médio mínimo para a função de densidade de probabilidade da figura 5.21.

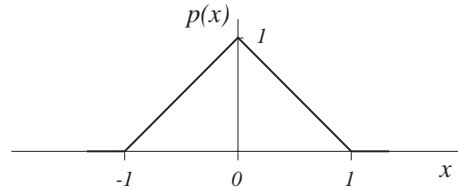


Figura 5.21:

- b) Calcule a razão de potências sinal-ruído de quantização para esse quantizador.
 - c) Defina um compressor e um expensor tais que a quantização não-uniforme possa ser efectuada utilizando esse compressor e um quantizador uniforme.
- 5.6 – Os q níveis x_i ($i = 1, 2, \dots, q$) de quantização são por vezes escolhidos de modo a que as amplitudes quantizadas do sinal ocorram com igual probabilidade, isto é, $p(x_i) = 1/q$.
- a) Defina um quantizador de quatro níveis de igual probabilidade para a função de densidade de probabilidade $p(x)$ do problema 5.4
 - b) Compare a razão de potências sinal-ruído de quantização do quantizador de igual probabilidade com a do quantizador de erro quadrático médio mínimo (Prob 5.5).
- 5.7 – Um sinal analógico com $B = 15$ KHz deve ser quantizado a $q \geq 200$ níveis e transmitido em PCM M -ário com $M = 2^n$. Determinar os valores máximos permissíveis para k e f_a e o correspondente valor de n se a largura de banda de transmissão disponível for $B_T = 50$ KHz.
- 5.8 – Repetir o problema 5.7 com $B_T = 80$ KHz
- 5.9 – A *hiperquantização* é o processo pelo qual N amostras quantizadas sucessivas são representadas por um único pulso de q^N amplitudes possíveis. Descreva como é que através do PCM hiperquantizado se pode obter uma *compressão de banda*, de modo a que $B_T < B$.
- 5.10 – Suponha que o erro de quantização ϵ_q não pode ser superior a P % da amplitude pico-a-pico do sinal. Determinar uma expressão para k em função da base de numeração M e de P .

- 5.11 – Conceba um circuito lógico para o codificador binário do conversor A/D da figura 5.12.
- 5.12 – Esquematize circuitos equivalentes aos da figura 5.12 para conversores de 4 bits.
- 5.13 – Um sinal de voz com $B = 3$ KHz e $S = 1/4$ deve ser transmitido em PCM M -ário. Determinar valores para a base de numeração M , número de dígitos k , e frequência de amostragem f_a , de modo a que $S/N \geq 40$ dB se $B_T = 16$ KHz.
- 5.14 – Repetir o problema 5.13 com $S/N \geq 36$ dB e $B_T = 20$ KHz.
- 5.15 – Um sinal de audio com potência média normalizada $S = 0.3$ deve ser transmitido por um sistema PCM cujos parâmetros devem satisfazer o padrão de qualidade de radiodifusão audio.
- a) Se a base de numeração for $M = 2$ quais os valores exigidos para o número de dígitos k e para B_T ?
- b) Se $B_T = 4B$ então qual o valor mínimo para M ?
- banda de audio para radiodifusão: 100 Hz a 5 KHz,
padrão de qualidade: $S/N = 40$ a 50 dB
- 5.16 – Repetir o problema 5.15 para o padrão de qualidade de audio de alta fidelidade.
- banda de audio Hi-Fi: 20 Hz a 20 KHz,
padrão de qualidade: $S/N = 55$ a 65 dB
- 5.17 – No Modulador Delta da figura 5.15 a), $x(t)$ é o sinal sinusoidal $x(t) = 0.1 \sin 2\pi \times 10^3 t$. A frequência do relógio é de 2×10^4 Hz e a altura do degrau é de $\Delta = 20$ mV. Esquematize as formas de onda $x(t)$, $\tilde{x}_q(t)$ e $\epsilon_q(t)$. Suponha que $x - \tilde{x}_q = \Delta$ quando $t = 0$.
- 5.18 – Repetir o problema 5.17 para $\Delta = 4$ mV e para $\Delta = 60$ mV. Notar que $\Delta = 20$ mV fornece a melhor estimativa $\tilde{x}_q(t)$. Mostre que a melhor altura do degrau é $\Delta_{opt} \approx 4A/(f_a/f_m)$ onde A e f_m são a amplitude de pico e a frequência do sinal sinusoidal.

fim do capítulo 5

Capítulo 6

Multiplexagem

Quando o canal de transmissão possui uma capacidade muito superior ao débito da fonte, ou seja, quando $B_T \gg B$ ou $B_T \gg \frac{r_s}{2}$, pode utilizar-se esse canal para transportar os sinais de mais do que uma fonte. Quando assim se procede diz-se que o canal está a ser partilhado ou *multiplexado*.

Essencialmente, pode partilhar-se o *tempo* ou a *frequência*.

No primeiro caso, designado por TDM ou Multiplexagem por Divisão do Tempo (Time Division Multiplexing), cada fonte ocupa toda a largura de banda disponível do sistema durante parte do tempo. No segundo caso, designado por FDM ou Multiplexagem por Divisão da Frequência (Frequency Division Multiplexing), cada fonte ocupa uma fracção da largura de banda disponível do sistema durante todo o tempo.

O FDM é o método mais antigo e mais conhecido visto que se lida com ele diariamente quando, por exemplo, se sintoniza uma estação de rádio ou de televisão. O espaço livre constitui o meio comum de transmissão que é multiplexado em frequência pelas várias estações de radiodifusão.

O FDM é o mais apropriado quando a transmissão é analógica. Actualmente, com a progressiva digitalização das comunicações tem perdido importância a favor do TDM que constitui a base dos modernos sistemas de telecomunicações e certamente da maioria dos futuros.

Existem também as formas de multiplexagem por divisão de espaço, SDM (Space Division Multiplexing), e por divisão de código, CDMA (Code Division Multiple Access), mais apropriados a outros contextos que não o da transmissão e que se abordarão na altura própria. Começaremos

por estudar o TDM e faremos apenas, na parte final deste capítulo, uma referência ao FDM.

6.1 Multiplexagem por divisão do tempo, TDM

Um sinal amostrado é nulo nos intervalos de tempo entre as amostras, ou seja, durante grande parte do tempo. Se se pretender transmitir o sinal sob esta forma (PAM) esses intervalos podem ser utilizados, em particular, para transmitir as amostras de sinais de outras fontes, intercalando-as entre si.

A figura 6.1 a) representa as características essenciais de um sistema TDM onde várias fontes de sinal distintas partilham um único canal de transmissão. Cada um dos sinais é previamente filtrado a fim de lhe limitar a banda espectral. Um *comutador rotativo* de amostragem do lado do transmissor extrai uma amostra de cada entrada por cada rotação. A saída deste comutador, fig 6.1 b), é um sinal PAM contendo as amostras de cada fonte periodicamente intercaladas no tempo.

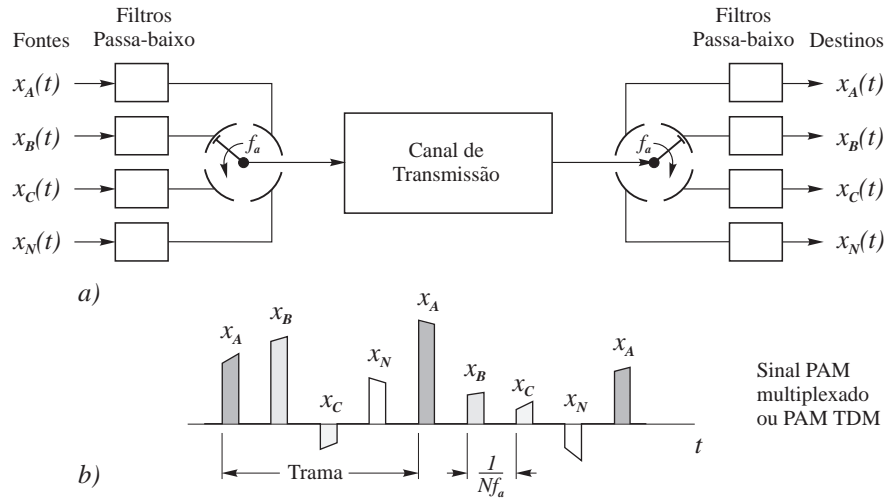


Figura 6.1: Multiplexagem por divisão do tempo. Sistema TDM

Do lado do receptor existe um comutador rotativo semelhante, ou *distribuidor*, que separa as amostras e as distribui por filtros passa-baixo, um

por cada saída, que reconstituem os sinais analógicos originais. Convém relembrar que, embora ambos os conjuntos de filtros possuam características passa-baixo, eles destinam-se a fins completamente diferentes. O primeiro conjunto de filtros *limita* a banda dos sinais para evitar o *aliasing* e o segundo conjunto *reconstrói* os sinais a partir das suas amostras.

Se todas as fontes produzirem sinais com a mesma largura de banda B , o comutador deverá rodar ao ritmo $f_a \geq 2B$ ciclos/s por forma a que amostras sucessivas de cada entrada estejam espaçadas no tempo de $T_a = 1/f_a$ segundos. A um conjunto, ordenado no tempo, constituído por uma amostra de cada entrada chama-se uma *trama*. A figura 6.1 b) mostra duas tramas TDM que transitam no canal de transmissão TDM. Se houver N pares de fonte/destino o espaçamento temporal entre amostras consecutivas numa *trama* será então de $T_a/N = 1/Nf_a$. Assim, o ritmo de pulsos PAM no canal de transmissão será de

$$r_c = Nf_a \geq 2NB \quad (6.1)$$

que é o *ritmo de transmissão* do *sinal TDM*.

Na prática, tal como se referiu no capítulo anterior, a comutação é feita electronicamente. A figura 6.2 a) mostra uma realização de um comutador electrónico para TDM em que as *gates* dos transistores que constituem as portas de modulação de pulso são controladas por um registo de deslocamento realimentado¹ constituído por flip-flops tipo-D que mantêm um único 1 lógico a *circular* ao ritmo de um relógio de frequência Nf_a Hz. A circulação daquele dígito provoca o *fecho* sequencial das portas (comutadores) emulando assim o comutador mecânico rotativo. As saídas dos comutadores são adicionadas num somador analógico que produz finalmente o sinal TDM para transmissão. Na figura 6.2 b) estão representados os diagramas de temporização do registo de deslocamento, das *gates* dos comutadores e do circuito de *manutenção (hold)*.

6.1.1 Sincronização

Uma questão essencial nestes sistemas de multiplexagem é a da *sincronização* entre o comutador e o distribuidor. Trata-se de uma questão crítica pelo facto de cada amostra ter de ser entregue ao destino correcto no instante devido. Isto significa que o distribuidor deve estar posicionado

¹pode verificar-se que se trata de um *ring counter* auto-corrector

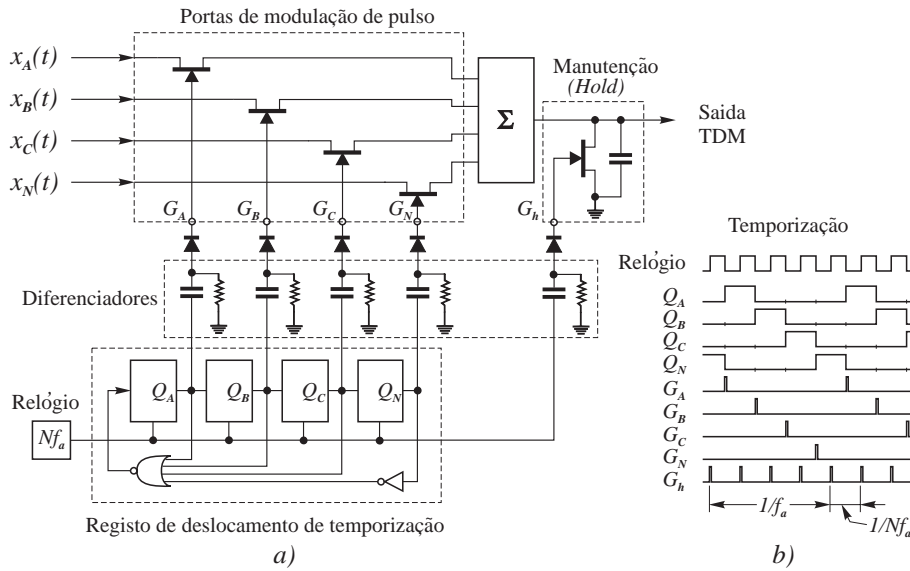


Figura 6.2: Realização de um comutador electrónico para TDM

na saída do destino i sempre que chega uma amostra originária da fonte i . Quando isto acontece diz-se que a trama está *alinhada*.

Uma técnica de sincronização simples e elementar é aquela que dedica uma *célula* temporal por *trama* à transmissão de um pulso-marca que é caracterizado por possuir uma amplitude superior ao valor máximo da amplitude que alguma vez possa ocorrer numa amostra qualquer, como mostra a figura 6.3. A *marca* é o último pulso transmitido em cada trama.

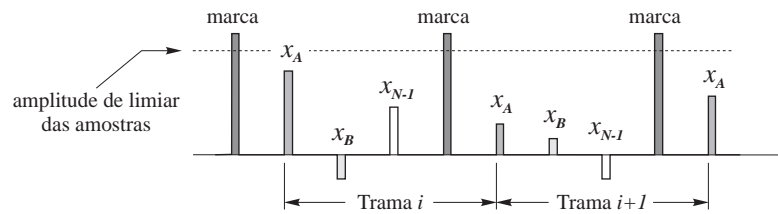


Figura 6.3: Sinal PAM TDM com *marcas* para sincronização

Assim, um circuito electrónico detector de limiar (ou de nível) colocado no

receptor², antes do distribuidor, ao detectar a *marca* provoca um *reset* no circuito lógico de controlo do distribuidor de modo a que este se posicione inicialmente na primeira saída. Estas *marcas* definem a frequência de trama f_a no receptor mas o número de pares de fonte/destino de sinal é reduzida a $N - 1$. Existem, naturalmente, outras técnicas de sincronização mais sofisticadas que se abordarão mais adiante.

6.1.2 Conceito de canal virtual

Do ponto de vista de cada par individual fonte/destino tudo se passa como se existisse entre eles um *caminho dedicado* pelo qual transitam as amostras do respectivo sinal. Considerando que podem existir $N > 1$ pares fonte/destino, o sistema de transmissão está a fornecer N caminhos distintos sobre um único canal *físico*, neste caso, o canal TDM. Cada um destes caminhos distintos é um canal de comunicação fim-a-fim ou *canal virtual* fonte-destino. Através deste conceito tudo se passa pois como se cada fonte estivesse ligada ao destino correspondente através de um circuito físico dedicado distinto.

O conceito de canal virtual (ou canal lógico) aparece frequentemente em outros contextos das telecomunicações, em especial na comunicação de dados e nos protocolos de comunicação entre computadores. Embora esses contextos sejam diferentes, como se verá, o conceito é essencialmente o mesmo.

A figura 6.4 ilustra o conceito de canal virtual ao representar o sistema TDM da figura 6.1 a) do ponto de vista da comunicação fim-a-fim entre cada fonte e o correspondente destino.

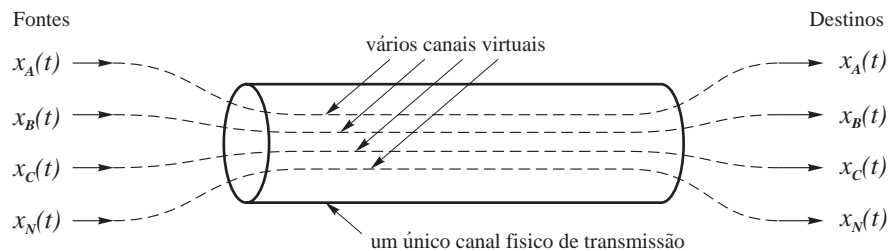


Figura 6.4: Representação do conceito de canal virtual

²ter-se-ia de conceber a lógica para lidar com estas *marcas*

6.1.3 Técnicas TDM

Na técnica que se acabou de descrever, os símbolos sucedem-se regularmente no tempo. As tramas são também contíguas no tempo, ou seja, ocorrem sem interrupção.

Se uma determinada fonte deixar de transmitir, os intervalos de tempo que lhe estão atribuídos em cada trama têm de decorrer embora não estejam a ser utilizados pois é a ordenação temporal desses intervalos que identifica o canal a que cada um pertence. Quando assim é, a técnica designa-se por *TDM síncrono*.

Quando se organiza a multiplexagem de modo a que não se exige nem a referida ordenação temporal nem a contiguidade das tramas, pode-se aproveitar o tempo que de outro modo seria desperdiçado quando determinado canal não produz *tráfego*, e está-se perante um *TDM assíncrono*, técnica também designada por *multiplexagem estatística* ou *Modo de Transferência Assíncrono* (ATM³).

6.2 TDM síncrono

A primeira forma de TDM síncrono apareceu com a digitalização PCM do sistema de telecomunicações eléctrico mais antigo — o sistema telefónico. A *normalização* da estrutura do TDM síncrono começou pois por estar fortemente influenciada pela preocupação de transmitir canais de voz de qualidade telefónica, donde resultou o designado sistema de multiplexagem PCM. A sua estrutura, porém, mostrou-se menos adequada a outros serviços de telecomunicações que foram surgindo posteriormente, como sejam, a televisão digitalizada, a comunicação de dados de alta velocidade entre computadores, etc, que se destinam a transmitir fontes de informação com características muito diferentes da da voz. Surgiram assim outras estruturas de multiplexagem, como sejam o SDH e o SONET.

Estudaremos os fundamentos do TDM síncrono tomando como referência o sistema PCM e abordaremos depois alguns aspectos doutros sistemas.

³o termo ATM (*Asynchronous Transfer Mode*) tem sido reservado para o contexto da comutação

6.2.1 Organização das tramas

Os sistemas TDM digitais multiplexam os canais sob a forma digital binária. Se as fontes são originalmente analógicas então estas são previamente convertidas a digital. A cada amostra PAM, depois de quantizada e codificada em binário, passa assim a corresponder um grupo de k bits que é designado por *palavra*. O ritmo *unitário* adoptado, designado por *ritmo básico*, é o resultante da digitalização de um canal telefónico ($B = 4$ KHz) quantizado a 8 bits por amostra, isto é, a unidade de velocidade digital é de

$$r_{\text{básico}} = 2 \times B \times k = 2 \times 4 \times 8 = 64 \quad \text{Kbps}$$

Uma *trama*, que passa a ser constituída pela multiplexagem no tempo de N canais de k bits, tem uma estrutura repetitiva que pode ser organizada de duas maneiras diferentes tal como mostra a figura 6.5:

- ou por *canais entrelaçados*, em que a trama é sub-dividida em N ou mais *células temporais* cada uma das quais contendo um grupo de k dígitos todos pertencentes ao mesmo canal digital;
- ou por *dígitos entrelaçados*, em que a trama é sub-dividida em k células temporais cada uma contendo os N ou mais dígitos da mesma ordem de cada um dos canais.

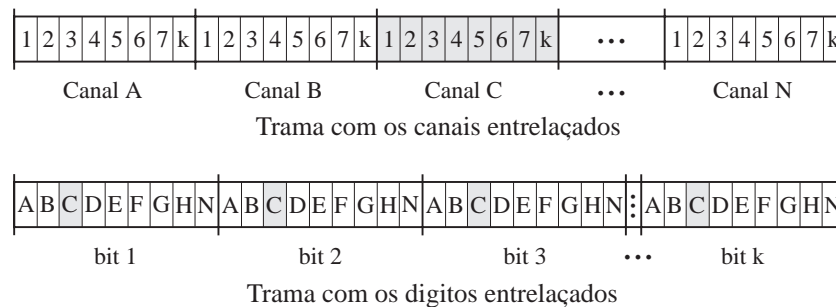


Figura 6.5: Formas de organização de tramas

A multiplexagem primária PCM é feita por canais entrelaçados e a multiplexagem de ordem superior é feita por dígitos entrelaçados.

6.2.2 Alinhamento das tramas

A operação de *alinhamento* de trama, típica da multiplexagem síncrona, consiste na sincronização do equipamento terminal de recepção (distribuidor), tanto em frequência como em fase, à sequência de símbolos que está a receber, isto é, na detecção do início da trama. É evidente que esta operação é necessária, pelo menos, cada vez que o receptor entra em operação (ao ser inicialmente ligado ou após uma interrupção, qualquer que seja a natureza desta). Uma vez *alinhado*, o receptor necessita, mesmo assim, de uma referência temporal periódica de modo a verificar o seu *isocronismo* e detectar eventuais desvios de fase. A referência temporal necessária consiste num determinado *padrão* de vários bits transportados periodicamente pela trama de acordo com uma das seguintes organizações:

- padrão de alinhamento *agrupado*: constituído por ν bits consecutivos no início de cada trama;
- padrão de alinhamento *distribuído*: os ν bits do padrão estão espalhados por determinadas posições dentro de uma trama ou ao longo de várias tramas.

Quando o receptor perde o alinhamento de trama, passa a procurar por este padrão de modo a se realinhar no mais curto intervalo de tempo possível (poucas tramas). Diz-se que o receptor está em *modo caça*⁴. Existe porém o *perigo de simulação* do padrão de alinhamento devido à ocorrência fortuita de uma combinação dos bits de informação igual à do padrão. Pode proteger-se o sistema de multiplexagem desta situação de várias maneiras, nomeadamente:

- escolhendo-se um padrão de alinhamento com baixa autocorrelação, impossível de reproduzir por deslocação e associação com dígitos vizinhos aleatórios, como por exemplo
 $\nu = 3$: 110
 $\nu = 7$: 0011011
- bloqueando-se todos os canais da trama, isto é, substituindo-os pela emissão de sequências determinísticas, quando o alinhamento é perdido na recepção.

⁴*hunt mode*

- confirmando o correcto alinhamento através de um critério diferente, por exemplo, pela ausência do padrão de alinhamento em tramas alternadas.

Contudo, quando existe alinhamento, a reacção do desmultiplexador a uma recepção incorrecta do padrão de alinhamento deve ser retardada (histerese) por forma a evitar reacções àquilo que pode apenas ser um erro de transmissão em que o alinhamento não foi efectivamente perdido.

6.2.3 Sinalização

Designa-se por *sinalização* a transmissão de informação auxiliar entre os equipamentos de multiplexagem de modo a estes se poderem controlar entre si. Por exemplo, pode desejar-se que o sinal da entrada número 3 seja entregue à saída número 7, situação típica de uma função de *comutação*.

A informação de sinalização trocada entre os equipamentos deverá, portanto, possuir uma semântica própria, o que não acontece com a informação transportada entre as fontes e destinos ligados aos equipamentos a qual é transferida de forma *transparente* (não interpretada), e é digital por natureza constituindo comandos, confirmações, etc.

Podem conceber-se várias soluções para a transferência de informação de sinalização:

- *sinalização dentro-do-octeto*, também chamada por *roubo de bit*⁵, pela qual o bit (dígito) menos significativo do octeto da amostra codificada é periodicamente utilizado para a sinalização (por exemplo, de 6 em 6 tramas). O resultado é uma degradação imperceptível da correspondente transmissão quando esta é analógica mas uma grave restrição à utilização desse canal digital para transmissão de dados.
- *sinalização fora-do-octeto, por canal*, pela qual são acrescentados um ou mais dígitos de sinalização aos k dígitos de informação de cada canal. Estes dígitos de sinalização de cada um dos N canais poderão estar *distribuídos* ao longo da trama, isto é, seguirem-se imediatamente aos dígitos de informação do canal correspondente, ou estar *agrupados* numa ranhura temporal (por exemplo, num canal) especificamente reservado para esse efeito, o qual é atribuído ciclicamente a cada um dos N canais de informação da trama.

⁵ *bit stealing*

Em qualquer dos casos a atribuição dos dígitos de sinalização é feita de uma forma fixa, segundo uma regra pré-estabelecida.

No primeiro caso o ritmo de sinalização é de pelo menos 8 Kbps por canal e, no segundo caso, de até $64/N$ Kbps por canal.

- *sinalização em canal comum*, pela qual é reservado um canal por trama para sinalização, o qual é atribuído *ocasionalmente* e *de acordo com as necessidades*, a um ou a outro canal. A sinalização é então efectuada através de *mensagens etiquetadas*, isto é, numa determinada trama o valor dos k dígitos do canal de sinalização constituem a *etiqueta*, que identifica o canal a que a sinalização (*mensagem*) das tramas subsequentes se refere.

O que deve estar pré-estabelecido é apenas em que tramas é que o valor do canal de sinalização deve ser interpretado como *etiqueta*.

O ritmo de sinalização instantâneo disponível, neste caso para um canal de cada vez, é de 64 Kbps.

A *sinalização em canal comum* é a técnica mais utilizada actualmente, que serve não apenas para a sinalização dos canais mas também para a troca de informação auxiliar, ou de control, entre os próprios multiplexadores.

6.2.4 Sistemas de multiplexagem PCM

Sistemas Europeu e Americano

Sob pena de proliferarem sistemas de multiplexagem incompatíveis, os diversos parâmetros destes sistemas — ritmos de transmissão, número de canais por trama, método de sinalização, etc — foram sujeitos a normalização que se iniciou mais cedo nos Estados Unidos da America e no Japão do que na Europa não se tendo aqui optado exactamente pelas normas Americanas por entretanto a experiência ter mostrado possuírem algumas deficiências ou já não se adaptarem tão bem à tecnologia mais recente. Aparecem assim os chamados sistema Americano (ou sistema AT&T) e o sistema Europeu ambos especificados em Recomendações da ITU⁶.

⁶International Telecommunications Union

Estrutura da trama PCM primária de 2 Mbps

O sistema de multiplexagem primário Europeu é caracterizado por um ritmo total de 2 048 Kbps, abreviadamente referido por ritmo de 2 Mbps, resultante da composição da trama de $125 \mu\text{s}$ de duração por 32 canais básicos. A figura 6.6 representa a estrutura desta trama. Os canais 1 a 15 e 17 a 31 constituem os 30 canais de informação utilizáveis quer para transmissão telefónica PCM quer para transmissão digital directa (transmissão de dados, telefax, etc) e sob o ponto de vista do utilizador estão numerados de 1 a 30, razão pela qual este sistema primário se designa por MIC30⁷.

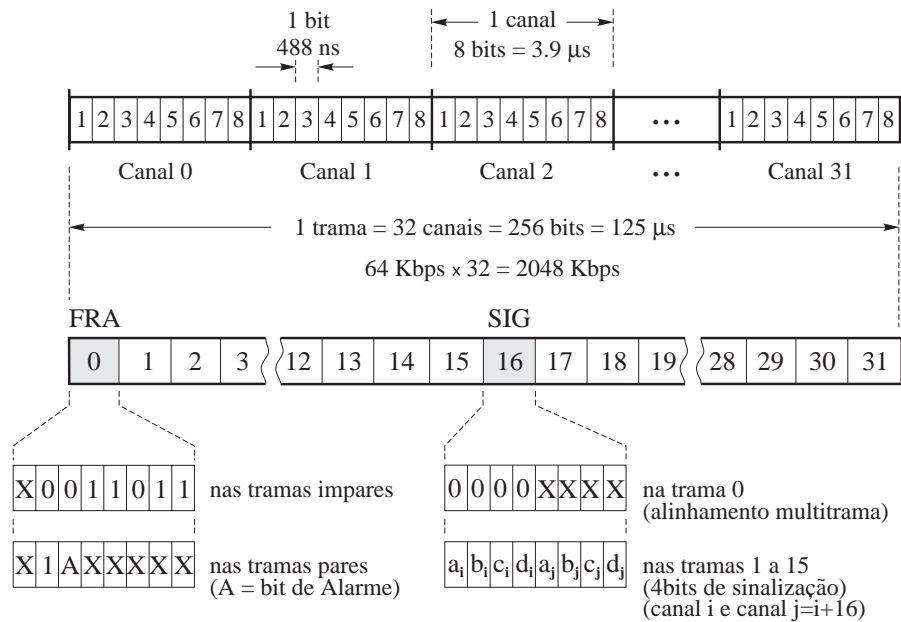


Figura 6.6: Estrutura da trama de multiplexagem PCM de 2 Mbps

O canal 0 destina-se à transmissão, em tramas alternadas, do padrão de alinhamento de trama, FRA⁸. Nas restantes tramas este canal transporta um bit de alarme *A* que anuncia a perda de alinhamento de trama na direcção oposta.

⁷Modulação de Impulso Codificada com 30 canais

⁸FRame Alignment

A sinalização é do tipo *fora-do-octeto*, canal a canal, estando atribuídos 4 dígitos por canal, agrupados nas duas metades do canal 16 designado canal SIG⁹. São pois necessárias 15 tramas para transportar a sinalização dos 30 canais úteis, as quais, juntamente com uma 16^a trama (a trama número 0), constituem uma *multitrama* como se mostra na figura 6.7. O canal 16 da trama número 0 contém um padrão de *alinhamento de multitrama* que permite numerar, sem ambiguidade, as tramas que constituem uma multitrama. Cada canal dispõe portanto de 4 bits de sinalização em cada 16 tramas, isto é, a intervalos de 2 ms, o que dá um ritmo de sinalização médio por canal de 2 Kbps.

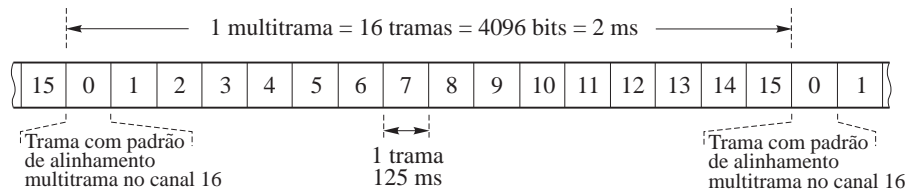


Figura 6.7: Multitrama PCM de 2 Mbps

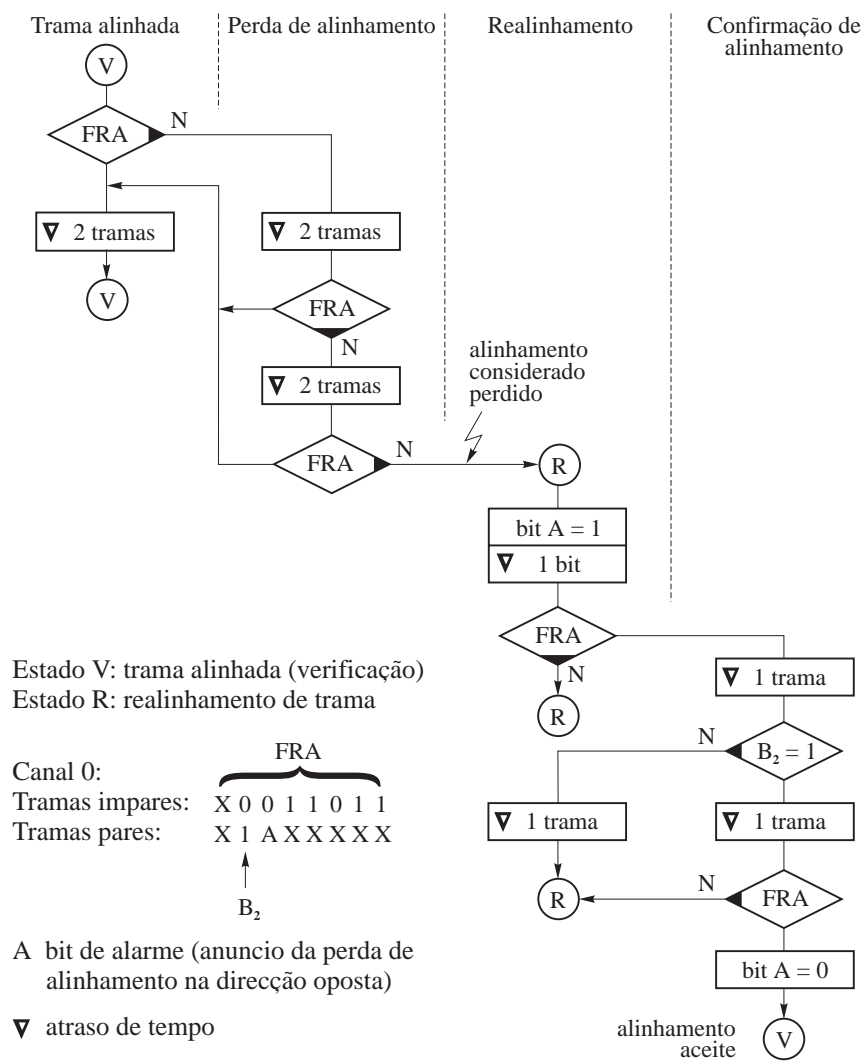
Alinhamento e interpretação da trama

A figura 6.8 representa o algoritmo completo do procedimento de sincronização, verificação de alinhamento e re-alinhamento, tal como se encontra especificado na Recomendação G.732 da ITU. São de salientar os seguintes pontos:

- a histerese do processo de monitorização: só se considera ter ocorrido perda de alinhamento após três ausências consecutivas do padrão de alinhamento, FRA;
- a confirmação de alinhamento pela presença de um valor diferente do bit da segunda posição (B_2) do canal 0 da trama seguinte àquela em que o padrão FRA foi reconhecido;
- o anúncio da perda de alinhamento através do bit de alarme A , emitido na direcção oposta.

Dependendo da situação, o realinhamento demora entre $250 \mu s$ e $375 \mu s$.

⁹SIGnaling channel



Estrutura da trama PCM primária de 1.5 Mbps

No sistema Americano de multiplexagem primária, a trama tem a mesma duração de $125 \mu s$ mas é composta por apenas 24 canais básicos também com 8 bits por canal. O alinhamento de trama neste sistema consiste na utilização de um único bit inicial de sincronismo por trama. Assim, uma trama é constituída por $24 \times 8 + 1 = 193$ bits sendo o ritmo primário então de $193 \text{ bits} / 125 \mu s = 1544 \text{ Kbps}$, abreviadamente designado de 1.5 Mbps. A figura 6.9 representa a estrutura da trama do sistema primário Americano. As tramas são consideradas alinhadas quando na

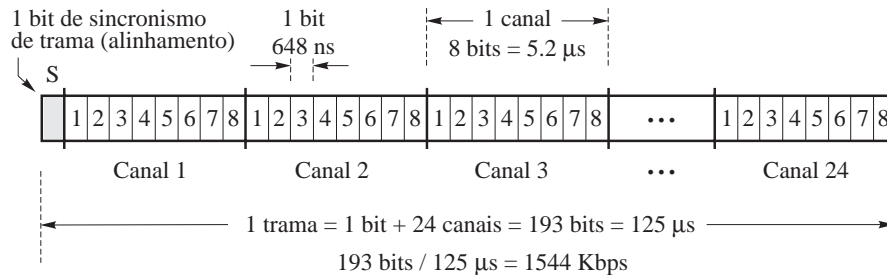


Figura 6.9: Estrutura da trama de multiplexagem PCM de 1.5 Mbps

sequência de valores do bit de sincronismo se observar o seguinte padrão cíclico: 100011011100. Isto significa que o sincronismo é efectuado com base numa multitrama de 12 tramas. Numa multitrama, as tramas são identificadas de 1 a 12 de acordo com o valor, ordenado, do respectivo bit de sincronismo.

Para a sinalização é roubado o último dígito (o menos significativo) de cada canal nas tramas 6 e 12 deixando livres, nestas tramas, apenas sete bits por canal para a informação útil.

Trata-se, portanto, de um sistema de sinalização *dentro-do-octeto* sendo roubados 1 bit por canal de 6 em 6 tramas. Conclui-se então, conforme se referiu na secção 6.2.3 a propósito deste tipo de sinalização, que as tramas 6 e 12 não podem ser utilizadas para transmissão de dados sob pena de se introduzir uma elevadíssima taxa de erros. Assim, sempre que um canal esteja a ser utilizado para comunicação de dados, o símbolo (grupo de 8 bits) seguinte ao que vai nesse canal da trama 5, segue na trama 7 seguinte; e o seguinte ao que vai nesse canal da trama 11 segue na trama 1 seguinte.

Um exemplo de utilização da sinalização é precisamente o tratamento deste caso, em que o multiplexador tem de fazer saber ao distribuidor que determinado canal vai ser usado para comunicação de dados de modo a que o sistema se comporte da maneira descrita.

Neste sistema a sinalização está organizada em dois canais de sinalização: o *canal A*, constituído pelos bits de sinalização transportados nas tramas 6, e o *canal B*, constituído pelos bits de sinalização transportados nas tramas 12. A figura 6.10 ilustra a localização destes canais na multitrama.

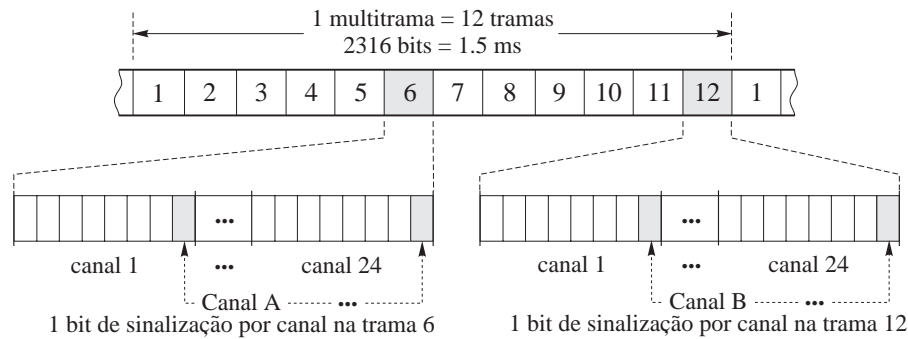


Figura 6.10: Multitrama PCM de 1.5 Mbps e canais de sinalização

6.2.5 Hierarquia de multiplexagem PDH

Multiplexagem de ordem superior

Para multiplexar um maior número de canais básicos do que aquele que o sistema primário admite (30 no sistema Europeu e 24 no sistema Americano) recorre-se à hierarquização por estágios sucessivos de multiplexagem. Assim, as saídas de multiplexadores primários são por sua vez multiplexadas num multiplexador de segunda ordem apresentando, necessariamente, um ritmo agregado de saída mais elevado. A figura 6.11 ilustra a hierarquia do sistema Europeu de multiplexagem PDH¹⁰ em que a ordem dos multiplexadores é identificada pelas designações M01, M12, M23, etc., e os ritmos normalizados são designados sucessivamente pelos identificadores E1, E2, E3, etc. Na versão Americana desta hierarquia,

¹⁰ *Plesiochronous Digital Hierarchy* ou Hierarquia Digital Plesiócrona

muitas vezes designada hierarquia AT&T¹¹, os ritmos normalizados são identificados por T1, T2, T3, etc.

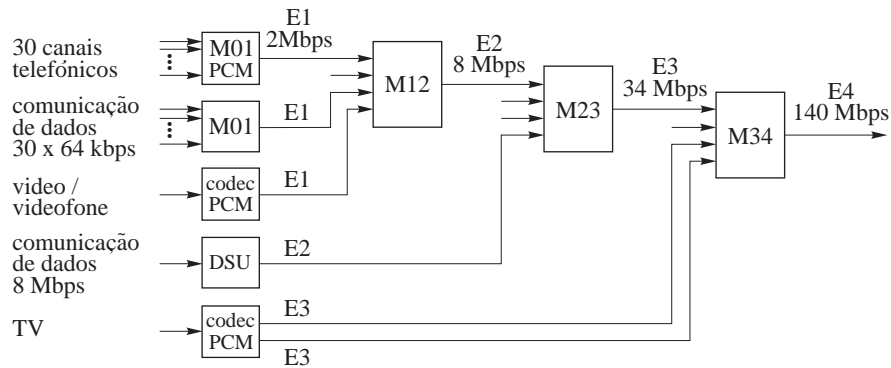


Figura 6.11: Exêmplo de uma multiplexação PDH Europeia

A tabela 6.1 indica os ritmos normalizados para cada um desses sistemas. Cada nível de multiplexação necessita também de dígitos de alinhamento

Tabela 6.1: Hierarquias de Multiplexação PDH

Ordem	Sistema Europeu ITU-T G.732			Sistema Americano ITU-T G.733		
	Entradas	Ritmo de saída (Kbps)		Entradas	Ritmo de saída (Kbps)	
1	30	2 048	E1	24	1 544	T1
2	4	8 448	E2	4	6 312	T2
3	4	34 368	E3	7	44 736	T3
4	4	139 264	E4	6	274 176	T4
5	4	564 992	E5			

da trama correspondente a esse nível, bem como de dígitos para a sinalização. Por essa razão a trama E2, por exêmplo, não é transmitida ao ritmo de $4 \times 2048 = 8192$ Kbps mas sim ao ritmo ligeiramente superior de 8448 Kbps.

¹¹Os sistemas e normas de telecomunicações Americanos são muitas vezes associados à AT&T, a American Telegraph & Telephone, a maior Companhia e operadora de telecomunicações dos EUA que os concebeu e adoptou

Interessa notar que a partir da segunda ordem, os multiplexadores apenas lidam com sinais digitais, tanto à entrada como à saída. Aliás, actualmente, os codecs vêm já incluídos nos equipamentos terminais das fontes analógicas de modo que os próprios equipamentos multiplexadores de primeiro nível já só vêm preparados para entradas digitais e os esquemas das figuras 6.1 e 6.2 para estes multiplexadores já não são muito usados na prática, isto é, raramente, a não ser para aplicações muito especiais, o multiplexador inclui o codec.

Donde resulta que a forma mais prática de organizar as tramas passa a ser por dígitos entrelaçados.

Se se pretender transmitir uma fonte com um débito nominal superior ao ritmo básico, como por exemplo um sinal de vídeo digitalizado, este teria de entrar directamente no nível M12.

Exemplo 6.1 *Um sinal de vídeo representando imagens em movimento de 300×400 pixels¹² por imagem com 8 tonalidades de cinzento (3 bits por pixel) a 16 imagens por segundo, constitui uma fonte com um débito binário de*

$$r_b = 300 \times 400 \times 3 \times 16 = 5\,760 \text{ Kbps}$$

É comum conseguir-se uma codificação binária comprimida de aproximadamente 30% pelo que o ritmo binário de transmissão no canal seria de

$$r_c = 0.3 \times 5\,760 = 1\,728 \text{ Kbps}$$

exigindo $N = 1\,728/64 = 27$ canais básicos para a sua transmissão. O codec de vídeo, neste caso, teria de organizar as amostras para as distribuir por esse número de canais e teria de possuir 27 saídas básicas.

Alternativamente, o sinal digitalizado poderia entrar directamente no nível primário de 2 Mbps, M12. Neste caso o codec teria de possuir um formatador de trama de acordo com a estrutura primária normalizada bastando possuir apenas uma saída primária.

As infraestruturas de transmissão digital das operadoras de telecomunicações estão estruturadas segundo esta hierarquia pelo que o utente (assinante) deverá possuir o equipamento terminal apropriado ao serviço que pretende utilizar. Se o equipamento que constitui a fonte de informação que se pretende transmitir não formata a saída de acordo com

¹²pixel é uma abreviatura de *picture element*, ou seja, elemento de imagem

o serviço que pretende utilizar (E1, E2, etc), então terá de se intercalar entre esse equipamento e o serviço um equipamento que o faça. Esse equipamento é designado genericamente por DSU (Digital Service Unit). Na figura 6.11 exemplificam-se alguns tipos de fontes de sinal e o nível da hierarquia que normalmente utilizam.

Anisocronismo das tributárias

Os multiplexadores do nível imediatamente inferior a um determinado nível de multiplexagem são designados de *tributárias* deste nível.

Na construção de um multiplex de ordem n a partir das tramas de ordem $n - 1$ é-se confrontado com o problema do *anisocronismo das tributárias*. Com efeito, as tramas que vão ser agrupadas no nível n podem ter origem em equipamentos terminais geograficamente distantes e cujos relógios possuam frequências não exactamente iguais (*tributárias plesiócronas*¹³) ou, na melhor das hipóteses, possuam frequências iguais (*tributárias isócronas*) mas estejam desfasados entre eles. O desfaseamento, que é um atraso relativo, pode mesmo variar ligeiramente no tempo. Esta variação do atraso tem o nome de *jitter*.

No entanto, como é fácil de concluir, a multiplexagem síncrona a qualquer nível n exige um perfeito *sincronismo* dos bits, isto é, que cada bit se apresente no instante ou intervalo de tempo que lhe está destinado. Face à tecnologia que é hoje utilizada na realização física dos multiplexadores, baseada em microprocessadores que incluem memórias *buffer* de entrada que se comportam como *memórias elásticas*¹⁴, esse sincronismo é relativamente simples de obter pois os bits são prévia e temporariamente memorizados nos buffers e só são de lá retirados no instante exacto da respectiva inclusão na trama de saída, o que é feito sob control do software (ou *firmware*) do multiplexador.

Para compensar a tolerância de variação de ritmo admissível nas tributárias, o ritmo nominal de saída dos multiplexadores é ligeiramente superior ao da soma dos ritmos nominais das entradas.

Isto significa que, em média, de tempos a tempos, o multiplexador não tem bits disponíveis nalguns dos seus buffers de entrada nos instantes em que esses bits deviam passar para a trama de saída. Nestas ocasiões, o

¹³do grego, *plēsio* (quase ou próximo) + *kronos* (tempo)

¹⁴são filas de espera

multiplexador insere nesse intervalo um *bit de recheio*¹⁵, isto é, um valor qualquer, mas tem de *sinalizar* que nesse intervalo de bit o valor não corresponde ao da entrada e deve ser ignorado.

Assim, quando o ritmo de uma tributária é inferior ao nominal, mas ainda dentro da tolerância, a frequência de *bits de recheio* na trama de saída correspondentes a essa tributária, é elevado, e quando aquele ritmo é superior ao nominal, a frequência de *bits de recheio* é pequena ou nula. O limite superior da tolerância de ritmo de uma tributária é aquele para o qual o multiplexador deixa totalmente de introduzir *bits de recheio* e o limite inferior é aquele para o qual o multiplexador introduz o número máximo de *bits de recheio* que é de um bit por trama por tributária.

É pelo facto de ocorrer plesiocronismo nas tributárias a cada nível desta hierarquia que ela toma o nome de Hierarquia Digital Plesiócrona ou PDH.

Estrutura da trama PDH de segunda ordem a 8 Mbps

A figura 6.12 representa a estrutura das tramas E2, ou de 8 Mbps, que constituem a segunda ordem da hierarquia Europeia, definida na Recomendação G.742 da ITU-T, e constitui um exemplo de multiplexagem com recheio. Estas tramas designam-se, em geral, de tramas de recheio (ou *stuffing frames*). As especificações deste nível são as seguintes:

- ritmo binário do multiplex, $r_M = 8448$ Kbps, toler: $\pm 30 \times 10^{-6}$.
- número de bits por trama: 848 bits, e portanto $T_{trama} \approx 100.4 \mu s$.
- número de tributárias a 2048 Kbps: 4.
- estrutura da trama: bits entrelaçados, um por tributária.
- alinhamento da trama: padrão de 10 bits contíguos no início.
- bits por tributária por trama: 206 (205 caso bit de recheio presente).
- sinalização de recheio: 3 bits por tributária, distribuídos, com decisão tomada por maioria.

A utilização de 3 bits de sinalização de recheio por tributária destina-se a proteger esta sinalização dos erros de transmissão. A decisão por maioria

¹⁵*stuffing bit*

consiste em considerar que existe bit de recheio na trama corrente, para uma determinada tributária, se pelo menos dois dos bits de sinalização correspondentes a essa tributária tiverem o valor 1. Os bits de recheio, quando ocorrem, ocorrem logo a seguir ao terceiro bit de sinalização. Na trama exemplificada na figura 6.12 foi introduzido um bit de recheio no fluxo da tributária II, o que está indicado pelo valor 1 dos 3 bits de sinalização correspondentes.

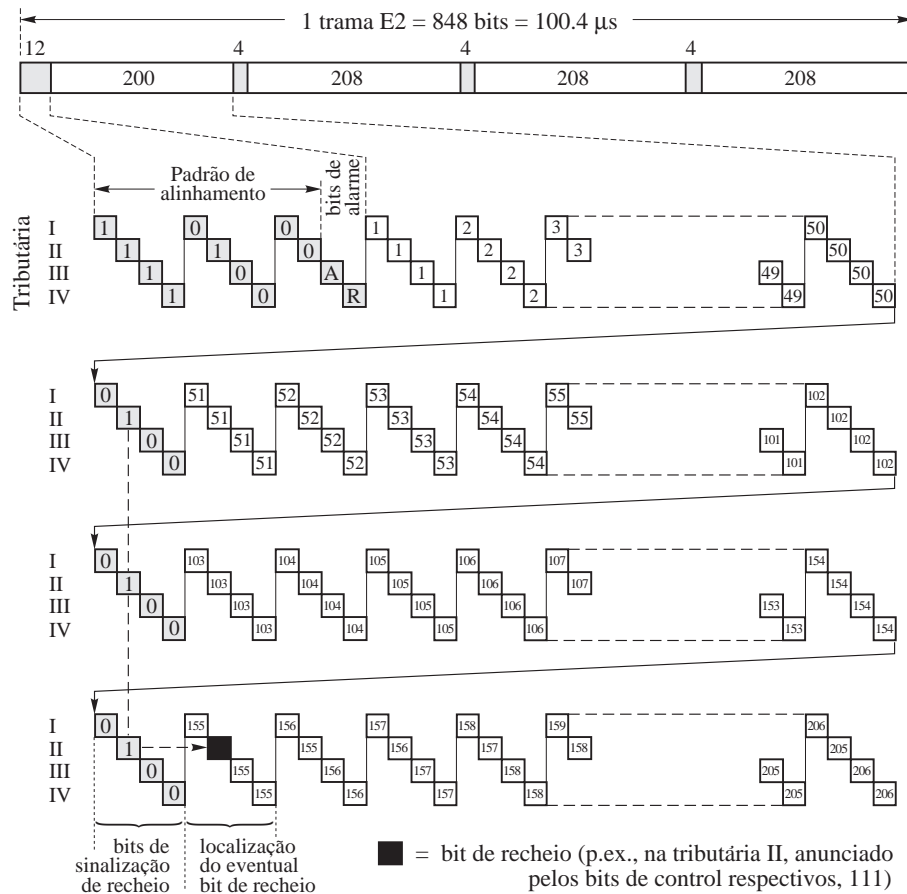


Figura 6.12: Formato da trama de multiplexagem PDH de 8 Mbps

Destas características podem retirar-se as seguintes conclusões:

- a) a trama E2, de 100.4 μ s, não tem qualquer relação com as tramas de amostragem das tributárias E1, de 125 μ s. O multiplex secundário

ignora a estrutura das tramas primárias e considera apenas o fluxo dos bits individualmente.

- b) a capacidade máxima de compensação através do mecanismo de recheio é, para cada tributária independentemente, de 1 bit em cada 212 bits, o que permite compensar uma diferença relativa de 0.47 % entre o ritmo da tributária, r_T , e o ritmo que lhe está atribuído, que é de $r_M/4$.
- c) para o ritmo primário nominal de $r_T = 2048$ Kbps, é necessário emitir, em média, $r_T \cdot T_{trama} = r_T \cdot (848/r_M) = 205\frac{19}{33}$ bits significativos por trama, por tributária, o que significa que, em média, 14 tramas em 33 contêm um bit de recheio.

Arquitectura dos multiplexadores

Os equipamentos multiplexadores são basicamente constituídos por uma configuração de microprocessador digital programável e portas série de entrada/saída utilizando USARTs¹⁶ às quais estão associadas *memórias buffer*, como se ilustra na figura 6.13. Cada USART, independentemente

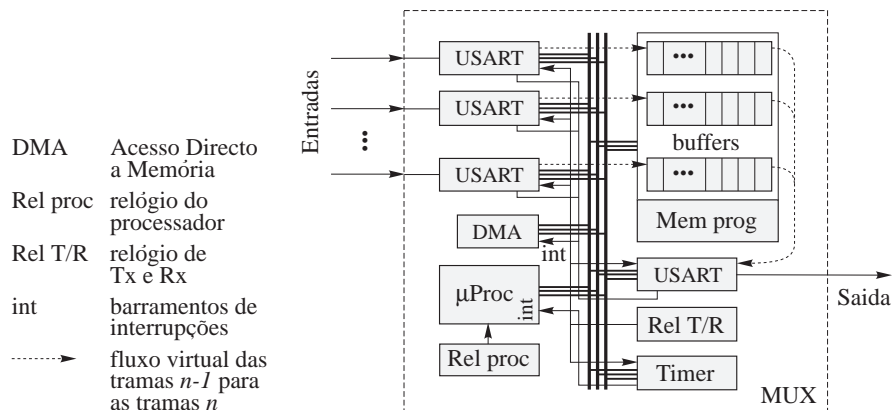


Figura 6.13: Multiplexador baseado em microprocessador

das restantes, recebe os bits das tramas $n - 1$ da tributária que está ligada à respectiva porta de entrada e, através de interrupções a uma unidade

¹⁶Universal Synchronous Asynchronous Receiver Transmitter

de *acesso directo a memória*, DMA¹⁷, coloca-os no *buffer* correspondente. Os *buffers* são zonas pré-definidas da memória, um por cada porta de entrada devendo cada um deles possuir uma dimensão tal que lhe permita armazenar alguns bits do nível $n - 1$.

O processador percorre ciclicamente todos os *buffers* lendo um bit de cada um deles e compõe a trama do nível n à medida que a vai transmitindo pela respectiva USART de saída.

Estas arquitecturas e, de uma maneira geral, quaisquer arquitecturas de equipamentos de comunicações, incluem um sub-sistema de transmissão/recepção de bits constituído por USART, *buffers*, processador DMA, processador de interrupções e relógio do processador e um sub-sistema de temporização das tramas constituído por um relógio (T/R) e por *timers* destinados a marcar os ritmos da transmissão/recepção.

Comunicação duplex

Por forma a permitir uma comunicação duplex em cada canal, torna-se necessário prover, a cada nível de multiplexagem e por cada multiplexador, um desmultiplexador que interligue os mesmos canais em sentido inverso. O desmultiplexador é normalmente designado *multiplexador inverso*. A figura 6.14 ilustra esta situação e em que ambas estas

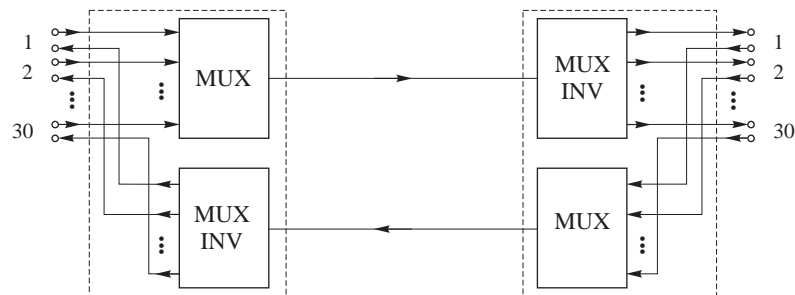


Figura 6.14: Duplexidade dos multiplexadores

funções estão integradas num mesmo equipamento (indicado a tracejado na figura). Muitas vezes, mas nem sempre, os equipamentos multiplexadores vêm já preparados para funcionar em duplex, mais propriamente, em *full-duplex*. Assim, cada uma das 30 portas indicadas na figura é

¹⁷Direct Memory Access

uma porta bidireccional. A interligação entre os multiplexadores faz-se, por exemplo, através ou de um par de fios para cada sentido (4 fios), ou de um par de fios para ambos os sentidos (2 fios) onde os dois sentidos são multiplexados em frequência (como se verá mais adiante), ou de uma fibra óptica para cada sentido. Na figura 6.14 cada linha com uma seta representa, obviamente, ou um par de fios condutores ou uma fibra óptica.

6.2.6 Hierarquia de multiplexagem SDH e SONET

As duas hierarquias PDH referidas nas secções anteriores são um exemplo da multiplicidade de sistemas TDM distintos que realmente existem na prática, o que dificulta a sua interoperação. Por outro lado, a evolução da tecnologia óptica e a sua crescente disponibilidade, permitindo a utilização de ritmos binários muito superiores, criou a oportunidade de repensar a normalização e de se conceber um sistema *padrão* de multiplexagem TDM. A iniciativa partiu das operadoras de telecomunicações norte-americanas, que propuseram a norma SONET (*Synchronous Optical NETwork*), a que se juntou a ITU que produziu a série de Recomendações G.707, G.708 e G.709 normalizando o sistema a que chamou SDH (*Synchronous Digital Hierarchy*). Na altura em que se iniciou a definição SONET/SDH, os níveis E4 e T4 não eram praticamente utilizados devido a dificuldades de os realizar com tecnologia eléctrica.

O desejo de uniformidade absoluta não foi, porém, completamente conseguido (como é tradicional!) mas estas duas normas apenas diferem em aspectos não essenciais, pelo que, numa abordagem necessariamente introdutória, as diferenças não serão praticamente notórias.

Os principais objectivos da concepção dos sistemas SDH e SONET foram:

- (i) a adopção de uma hierarquia única para TDM síncrono.
- (ii) a continuação da hierarquia até e para além do gigabit/seg (Gbps).
- (iii) permitir a multiplexagem directa de canais muito mais lentos nos níveis superiores.
- (iv) o enriquecimento da estrutura de sinalização de modo a suportar melhor serviços de administração, operação e manutenção do sistema.

- (v) permitir o transporte de tráfego assíncrono (estatístico) e que tal se efectue de forma rentável (eficiente).

Definições

Um sistema SDH/SONET é constituído, em geral, por multiplexadores, repetidores e comutadores. A figura 6.15, que representa o modelo de referência funcional e terminológico, mostra um *caminho* entre dois equipamentos terminais ligados a um sistema SDH/SONET. Cada equipamento terminal está ligado a um multiplexador que por essa razão se designa por multiplexador terminal. No trajecto entre os mux terminais existem repetidores-regeneradores de sinal e multiplexadores ou comutadores intermédios.

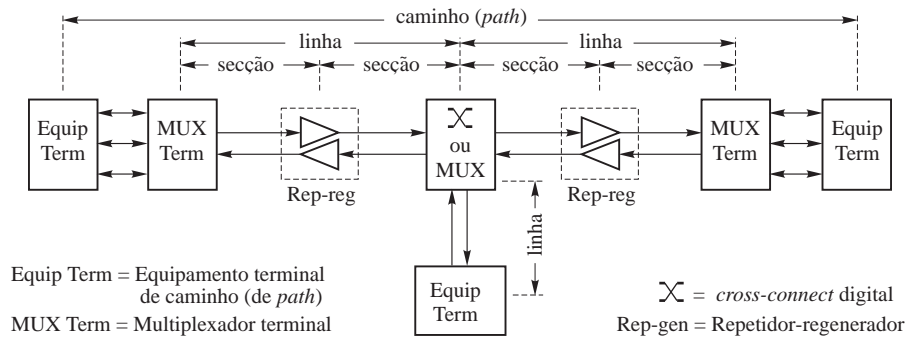


Figura 6.15: Definições de secção, linha e caminho no SDH e SONET

Nestes sistemas, a fibra óptica ininterrupta que liga dois quaisquer destes dispositivos é designada de *secção*. O trajecto entre dois mux ou comutadores, tenha ou não repetidores intercalados, é uma *linha*. E o trajecto entre equipamentos terminais, eventualmente com um ou mais mux e repetidores, é chamado um *caminho*, ou *path* que é o termo original em língua inglesa.

Formato das tramas SDH e SONET

As tramas básicas SDH e SONET são constituídas por blocos de 810 bytes de 125 μ s de duração. Esta duração da trama coincide com o período de amostragem dos canais PCM e portanto também são produzidas 8000

tramas por segundo. Trantando-se de sistemas síncronos, as tramas são sempre emitidas, continuamente, quer haja informação útil a ser transmitida quer não.

As tramas são descritas nas normas através de uma matriz de bytes com 90 colunas e 9 linhas. A figura 6.16 mostra duas dessas tramas. Os bytes são transmitidos da esquerda para a direita, por linhas, começando na primeira linha.

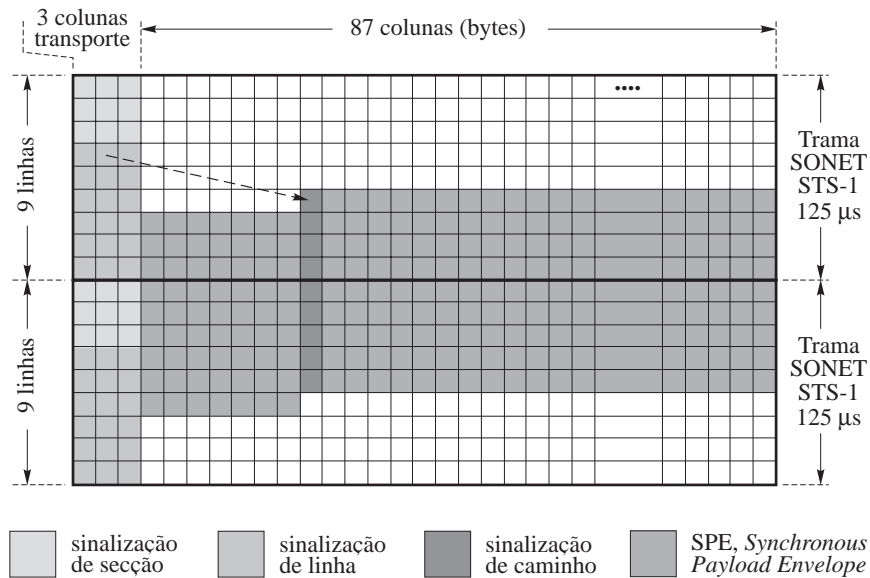


Figura 6.16: Duas tramas SONET STS-1 contíguas

Como são transmitidos $810 \times 8 = 6480$ bits em 125μ s o ritmo binário bruto é de 51.84 Mbps que é o ritmo do canal básico SONET e a trama correspondente é designada STS-1 (Synchronous Transport Signal-1). Todos os restantes ritmos são múltiplos do STS-1.

Em cada trama, os bytes das três primeiras colunas estão reservados para a gestão do sistema. Destes, os das três primeiras linhas, isto é, os bytes de ordem 1,2,3, 91, 92, 93, 181, 182 e 182, são destinados a sinalização de secção e os das seis linhas seguintes, isto é, os de ordem 271, 272, 273, 361, ..., 723, estão reservados para sinalização de linha.

Estes bytes constituem o que se designa por *overhead*, termo utilizado para referir tudo aquilo que não constitui informação útil dos utilizadores

originada nas entradas ou destinada às saídas dos equipamentos terminais. Trata-se de redundância, necessária para a gestão e operação do sistema, mas que contribui para baixar o seu rendimento (*utilização*), sob o ponto de vista desses equipamentos terminais.

Os bytes de *overhead* de secção são gerados e verificados no início e no final de cada secção e os bytes de *overhead* de linha são verificados no início e no fim de cada linha.

As restantes 87 colunas transportam $(87 \times 9 \times 8) / 125 \times 10^{-6} = 50.112$ Mbps de dados de utilizador — as origens/destinos ligados aos equipamentos terminais. Contudo, estes dados, designados SPE (Synchronous Payload Envelope)¹⁸, não têm necessariamente de começar ou ser inseridos na coluna 4 da linha 1. O SPE pode começar em qualquer posição dentro da trama. Para isso, a primeira linha do *overhead* de linha contém um apontador para o primeiro byte do SPE como se exemplifica na figura 6.16. A primeira coluna ocupada pelo SPE contém *overhead* de *path*. São bits para sinalização entre os equipamentos terminais desse caminho.

O facto do SPE poder começar em qualquer posição de uma trama e de se estender por duas tramas como mostra a figura 6.16 torna o sistema mais flexível. Se os dados (uma *payload*) chegam a um mux quando está a ser formada uma trama, eles podem ser imediatamente inseridos nessa trama em lugar de terem de esperar pelo começo da seguinte. Esta característica também se torna útil quando a *payload* não cabe exactamente numa trama como é o caso, por exemplo, duma fonte ATM cujos dados vêm organizados em células contíguas de 53 bytes. A primeira linha do *overhead* de linha pode então apontar para o início da primeira célula completa.

Hierarquias SDH e SONET

Os sistemas SDH e SONET estão organizados hierarquicamente como mostra a tabela 6.2. A multiplexagem em cada nível é feita por bytes entrelaçados. O *sinál* STS-3 resultante da multiplexagem de três entradas STS-1 é uma trama de 270 colunas por 9 linhas a qual, sendo transmitida em 125 μ s, define para aquele sinal o ritmo de 155.52 Mbps.

¹⁸os dados úteis em sistema de telecomunicações são geralmente chamados, na terminologia inglesa, de *payload*, querendo-se com isso significar *carga-taxada*, a carga da responsabilidade do utilizador

O sistema SDH só começa nos 155.52 Mbps e o respectivo nível é designado STM-1 (Synchronous Transfer Mode-1).

Tabela 6.2: Hierarquias de Multiplexagem SDH e SONET

SONET		SDH	Ritmo binário (Mbps)		
Eléctrico	Óptico	Óptico	Bruto	SPE	Útil
STS-1	OC-1		51.84	50.112	49.536
STS-3	OC-3	STM-1	155.52	150.336	148.608
STS-9	OC-9	STM-3	466.56	451.008	445.824
STS-12	OC-12	STM-4	622.08	601.344	594.432
STS-18	OC-18	STM-6	933.12	902.016	891.648
STS-24	OC-24	STM-8	1244.16	1202.688	1188.864
STS-36	OC-36	STM-12	1866.24	1804.032	1783.296
STS-48	OC-48	STM-16	2488.32	2405.376	2377.728

Ao sinal STS- i transmitido sob a forma óptica o sistema SONET designa por OC- i (Optical Carrier- i). O sistema SDH só define sinais portadores ópticos.

A tabela também mostra, para cada nível, os ritmos brutos (na linha de transmissão), os ritmos de SPEs e os ritmos úteis que se obtêm à saída dos caminhos. A figura 6.17 ilustra uma situação de multiplexagem hierárquica SDH.

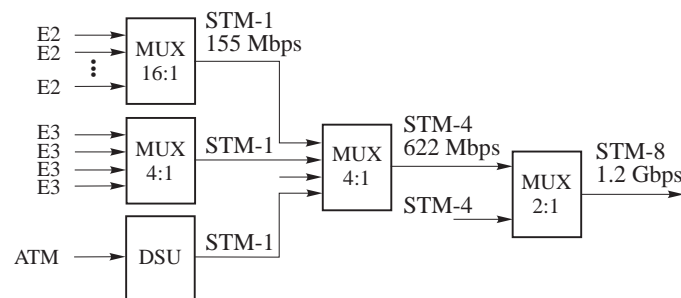


Figura 6.17: Multiplexagem hierárquica SDH

Quando um sinal (ou *carrier*), como por exemplo o STS-3 ou o OC-3, não resulta de uma multiplexagem mas transporta dados de uma única fonte, identifica-se esse facto adicionando a letra *c*, significando *concatenação*,

à sua designação. Assim, STS-3 e OC-3 indicam sinais de 466.56 Mbps e 155.52 Mbps resultantes da multiplexagem de três sinais distintos, respectivamente, STS-1 e OC-1, mas STS-3c e OC-3c indicam sinais de uma única fonte a 466.56 Mbps e 155.52 Mbps respectivamente. A quantidade de dados úteis num sinal concatenado STS- i c é ligeiramente superior ao de um STS- i visto que só existe uma coluna de *overhead* de *path* no SPE em lugar de três. Por exemplo, um OC-3c oferece 260 colunas para dados do utilizador, das 270 disponíveis na trama, enquanto que um OC-3 apenas oferece 258.

6.3 TDM estatístico

A multiplexagem síncrona é apropriada para a transmissão digitalizada de fontes que produzem informação a um ritmo contínuo ou regular, isto é, sem interrupções. Existem situações, porém, em que as fontes (equipamentos terminais) não se comportam dessa maneira. É o caso, por exemplo, dos computadores, dos terminais de dados, das estações de trabalho de computação, etc., em que o tráfego de dados que produzem tem origem nas aplicações de comunicações que os utilizadores invocam, como sejam, a transferência de ficheiros, o correio electrónico, o acesso por terminal remoto, etc. Este tráfego é de natureza aleatória e o multiplexador pode tirar partido desse facto alocando dinamicamente as ranhuras temporais das tramas de saída de acordo com a existência ou não de dados em cada instante nos *buffers* de entrada e com o seu estado de ocupação. A multiplexagem efectuada segundo este princípio designa-se por multiplexagem estatística ou assíncrona.

6.3.1 Características do tráfego gerado por computadores

Verifica-se que o tráfego de dados gerado pelas comunicações entre computadores é intermitente não-regular, de carácter aleatório, com um elevado *factor de crista*, isto é, a razão entre o ritmo binário nominal e o ritmo binário médio é muito maior que 1. A figura 6.18 ilustra esta característica. Na realidade, este tipo de tráfego de dados é caracterizado por sequências de bits contíguos intercaladas por longos períodos de ausência de transmissão, isto é, os dados aparecem aos mólhos (*bursts*): ou há transmissão — intervalos de tempo τ_i — e ela efectua-se ao ritmo nominal, r_b , ou não há transmissão — restantes intervalos de tempo. A

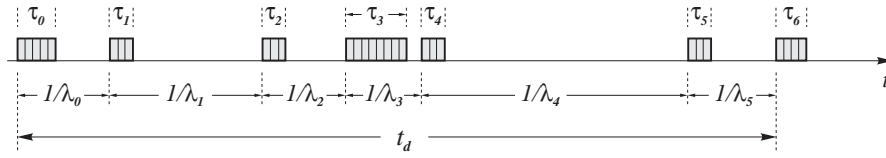


Figura 6.18: Característica intermitente das transmissões por computador

duração das intermitências, $1/\lambda_i - \tau_i$, é aleatória e o ritmo médio, considerando um intervalo de tempo t_d suficientemente grande, é definido por

$$\bar{r}_b = \frac{\langle \text{número de bits transmitidos durante } t_d \rangle}{t_d} \quad (6.2)$$

e o factor de crista, fc ,

$$fc = \frac{r_b}{\bar{r}_b} = \frac{t_d}{\sum_i \tau_i} \gg 1 \quad (6.3)$$

ou seja, os equipamentos terminais apenas transmitem durante uma fracção do tempo cujo valor médio é

$$\alpha = \frac{1}{fc} \quad \text{donde} \quad 0 < \alpha < 1 \quad (6.4)$$

Uma outra característica importante diz respeito à *unidade de dados* objecto de transmissão. Quando um equipamento transmite, nunca emite apenas um bit. Ou emite oito bits seguidos (*byte*), ou um múltiplo de bytes seguidos ou vários bits seguidos em número não necessariamente múltiplo de oito. Uma unidade de dados (ou *Data Unit*, DU) é um conjunto de bits emitidos consecutivamente que deve ser tratado e interpretado sempre como um todo. As DU poderão ter um número fixo ou um número variável de bits consoante as regras protocolares adoptadas em cada situação, dizendo-se de comprimento fixo ou de comprimento variável, respectivamente. As DU, especialmente as de comprimento variável, costumam-se por vezes designar genericamente por *mensagens*. Na figura 6.18 os intervalos de tempo entre emissões de DUs foi designado por $1/\lambda_i$ segundos pelo que o valor médio de λ_i é o número médio de DUs produzidas por segundo.

6.3.2 Multiplexagem estatística

O multiplexador estatístico (ou assíncrono) explora esta característica do tráfego não regular, atribuindo ranhuras temporais da trama de saída às linhas de entrada que tenham tráfego nesse instante. Tal como o multiplexador síncrono, o multiplexador estatístico possui várias linhas de entrada e uma linha de saída multiplexada de ritmo mais elevado. A cada linha de entrada está associado um *buffer* destinado a armazenar temporariamente as unidades de dados que vão chegando até estas serem recolhidas para envio na trama de saída. No TDM estatístico, o facto de nem todos os equipamentos de entrada estarem a transmitir ao mesmo tempo, permite um ritmo nominal de saída inferior à soma dos ritmos nominais das entradas. Assim, um multiplexador estatístico pode utilizar um ritmo de saída inferior ao do multiplexador síncrono para o mesmo número e ritmos de entrada. Ou, alternativamente, para os mesmos ritmos de saída, o estatístico pode suportar um maior número de entradas. A figura 6.19 contrasta a multiplexagem (TDM) estatística com a síncrona mostrando quatro fontes e os dados por elas emitidos em quatro épocas

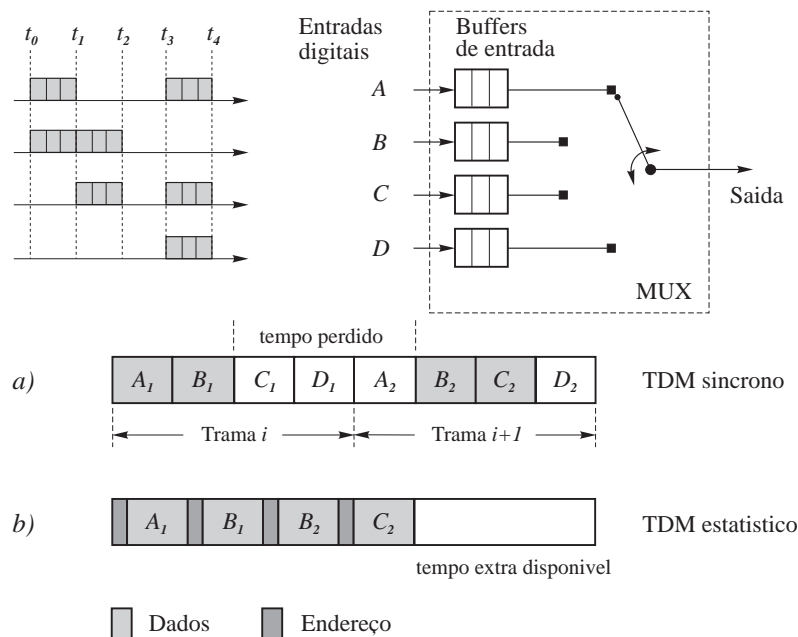


Figura 6.19: Contraste entre TDM síncrono e TDM estatístico

t_0 , t_1 , t_2 e t_3 regularmente espaçadas no tempo.

Se o multiplexador é síncrono, apresentará um ritmo de saída quatro vezes superior ao de cada entrada (supondo o destas iguais). Em cada época são preenchidas quatro ranhuras temporais da trama de saída. Na época t_1 duas dessas ranhuras ficam vazias, que é o caso da trama i na situação *a*), dado que os terminais C e D nada produziram na época t_0 .

Se o multiplexador é estatístico, não havendo mais dados de uma determinada época, passa aos da época seguinte não enviando ranhuras vazias. Na trama de saída são introduzidos dos dados de A e de B na época t_1 seguidos dos dados de B e de C na época t_2 e, não havendo dados na época t_3 , o multiplexador pode encerrar a trama e transmiti-la começando a formar uma nova trama a partir desse instante. Contudo, neste multiplexador perde-se o significado posicional das ranhuras, isto é, os canais não possuem uma posição pré-definida na trama TDM visto que tanto a chegada dos seus dados como a sua distribuição pelas tramas são imprevisíveis. Torna-se necessário fazê-los acompanhar do endereço, ou identificador, do canal a que pertencem. Esta informação extra constitui uma redundância (*overhead*) necessária mas mesmo assim poupa-se um certo tempo como se mostra na figura 6.19 *b*).

A estrutura da trama na multiplexagem estatística influencia o desempenho desta. É desejável minimizar o peso relativo dos dígitos de *overhead* podendo-se para isso transmitir o maior número possível de dados por cada entrada de cada vez. Como, no caso geral, o comprimento da trama não é fixo esta poderá necessitar tanto de um padrão de alinhamento inicial como de um padrão final para se poder determinar onde é que a trama acaba. A figura 6.20 mostra várias maneiras de construir uma trama de TDM estatístico. Na figura está-se a designar por *sub-trama TDM estatístico* ao conteúdo da trama entre padrões de alinhamento. Pode-se formar a sub-trama apenas com os dados de uma única entrada de cada vez, caso *a*), quando se pretende transmitir muitos dados por entrada de cada vez. Pode-se formar a sub-trama com dados de várias entradas, caso *b*), mas agora o multiplexador terá de retirar de cada *buffer* um número fixo de unidades de dados de cada vez de modo a se saber onde começam e onde acabam os dados (e os endereços) respeitantes a cada canal. Se isso não acontecer, então a estrutura dos dados de cada canal, para além de conter o endereço do canal terá de incluir um *campo de comprimento* que indique quantos, dos bits que se seguem, pertencem a esse canal e, por conseguinte, onde começa a ranhura do canal seguinte, caso *c*).

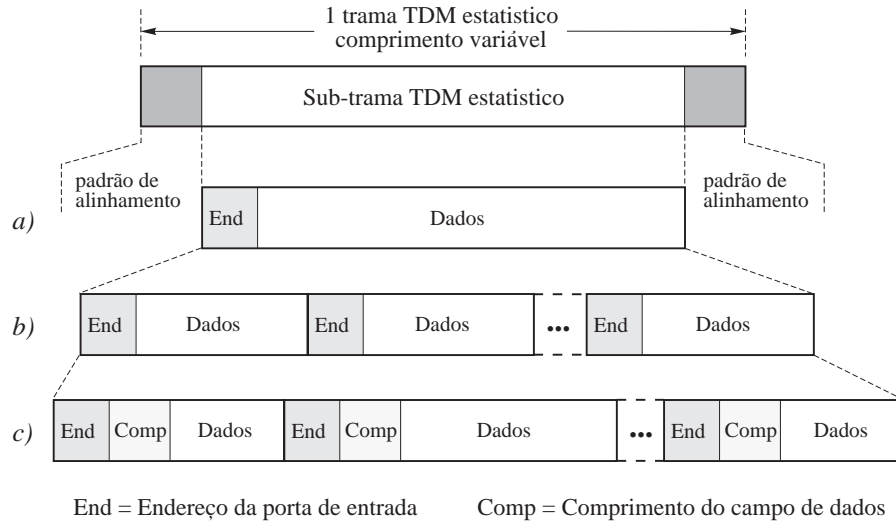


Figura 6.20: Formatos das tramas do TDM estatístico

Por sua vez, tanto o campo do endereço como o do comprimento, se este existir, podem ter comprimento fixo ou variável (ver problema 6.12). No campo do comprimento, por exemplo, pode convencionar-se que se os dois primeiros bits forem 11, esse campo tem um determinado número de bits que se lhe seguirão; e, se os dois primeiros bits forem 00, 01 ou 10, que apenas há esses dois bits para definir o comprimento, a que correspondem, respectivamente, um, dois e três unidades de dados.

6.3.3 Desempenho

Referiu-se atrás que o ritmo nominal de saída do mux estatístico pode ser inferior à soma dos ritmos nominais das entradas. Isto é possível porque se prevê que o ritmo médio agregado das entradas seja inferior ao seu ritmo nominal agregado. Uma dificuldade, porém, reside no facto de que, devido à aleatoriedade, poderão ocorrer períodos de pico de tráfego durante os quais o tráfego de entrada excede a capacidade da saída. Os *buffers* de entrada destinam-se precisamente a armazenar temporariamente esse excesso de tráfego comportando-se como autênticas *filas de espera*.

Existe um compromisso entre o tamanho dos *buffers* a utilizar e o ritmo binário na linha de saída. A situação ideal seria ter *buffers* pequenos e

ritmos de saída baixos mas uma redução deste ritmo faz aumentar as filas de entrada. A preocupação em ter filas curtas não é tanto a poupança de memória *buffer* (a memória é barata) mas com o facto de que, quantos mais dados houver em *buffer*, maior é o atraso na sua entrega ao destino. De facto, o compromisso é entre o *tempo de resposta* do multiplexador, como sistema, e a velocidade de transmissão na linha multiplexada (ritmo nominal de saída).

Para analisar e quantificar este compromisso recorre-se à *teoria das filas de espera* que é a que melhor descreve o comportamento do multiplexador estatístico.

Modelo de filas de espera

Na teoria das filas de espera um sistema é constituído por uma ou mais *filas de espera* e por um ou mais pontos de *serviço*. Uma situação de fila de espera acontece quando, do exterior do sistema, chega, um *cliente* que pretende ser servido mas ao encontrar o servidor ocupado é forçado a esperar juntando-se a uma fila. Ao chegar a sua vez, cada cliente transita da fila para o ponto de serviço no qual é *servido*. Em geral o serviço consiste num certo número de *operações* elementares com uma determinada duração. Após ser servido o cliente sai do sistema.

Há várias configurações para um sistema de filas de espera: a de uma única fila para um único servidor; a de múltiplas filas para um único servidor; a de uma fila para vários servidores e a de várias filas para vários servidores. No entanto, a segunda configuração pode reduzir-se à primeira e a quarta à terceira ou à primeira. Assim, para efeitos de análise apenas duas são verdadeiramente distintas: a de fila única e servidor único e a de fila única e servidor múltiplo. Aquela que descreve o multiplexador estatístico é a de servidor único representada na figura 6.21.

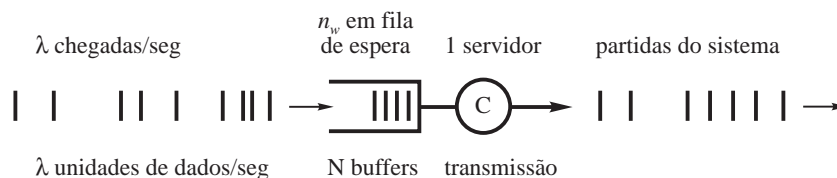


Figura 6.21: Modelo de fila de espera do MUX estatístico

Os clientes chegam ao sistema segundo uma determinada estatística a um ritmo médio de λ *chegadas por segundo*. O servidor possui uma capacidade de operação de C *operações por segundo* e o número de operações por cliente, que é imprevisível antes da sua chegada ao sistema, segue também uma determinada estatística com um valor médio de $1/\mu$ *operações*, isto é, cada cliente necessita, em média, de $1/\mu$ operações do servidor para ser considerado servido e sair do sistema. O tempo de serviço, S , é assim, em geral, uma variável aleatória cujo valor médio é

$$\bar{S} = \frac{1}{\mu C} \quad (6.5)$$

μC representa pois o número médio de clientes que o servidor serve por segundo. O servidor atende um cliente de cada vez, retirado da fila segundo uma determinada política, como por exemplo, a ordem de chegada¹⁹, ou por prioridade segundo um critério pré-definido como por exemplo, o mais rápido primeiro²⁰. É óbvio que quando o número médio de chegadas ao sistema excede o número médio de clientes que o servidor é capaz de despachar, a fila de espera tende a aumentar indefinidamente e diz-se que o sistema está em desequilíbrio. A condição para que um sistema de fila de espera esteja em equilíbrio é, portanto

$$\rho = \frac{\lambda}{\mu C} = \lambda \bar{S} < 1 \quad (6.6)$$

O parâmetro ρ mede a utilização do servidor, ou seja, representa a fracção do tempo total que o servidor está ocupado.

Um sistema de filas de espera é identificado pela notação A/B/m em que A designa a distribuição dos tempos entre chegadas, B a distribuição dos tempos de serviço e m o número de servidores. A e B tomam valores no conjunto de símbolos que representam a respectiva função de densidade de probabilidade (fdp) indicados na tabela 6.3 em que a função $\delta(x - 1/\mu)$ para a fdp determinística representa uma linha de altura unitária (impulso unitário) em $x = 1/\mu$.

O objectivo principal da teoria das filas de espera é o de se obter, para cada situação A/B/m, as distribuições dos tempos e do número de clientes tanto em espera em fila, t_w e n_w , como no sistema, t_q e n_q , donde se possam determinar as várias médias estatísticas. O *atraso* sofrido por

¹⁹FCFS, *First Come First Served* ou *FIFO, First In First Out*

²⁰*shortest job first*

Tabela 6.3: Alguns modelos probabilísticos para filas de espera

Símb	Distribuição	fdp
M	Markoviana (exponencial)	$p(x) = \mu e^{-\mu x}$
E_r	Erlang de r estágios	$p(x) = \frac{r\mu(r\mu x)^{r-1}e^{-r\mu x}}{(r-1)!}$
H_R	Hiperexponencial, R estágios	$p(x) = \sum_{i=1}^R \alpha_i \mu_i e^{-\mu_i x}$
D	Determinística	$p(x) = \delta\left(x - \frac{1}{\mu}\right)$
G	Geral	$p(x)$ é arbitrária

um cliente é o tempo gasto à espera na fila mais o tempo que demora a ser servido, $t_q = t_w + S$, e o número de clientes no sistema é o número de clientes à espera em fila mais o que está a ser servido.

Modelo para o multiplexador estatístico

Considere-se que as unidades de dados têm um comprimento fixo de k bits. O multiplexador estatístico é modelado por um sistema de fila de espera de servidor único, onde os clientes são as unidades de dados que chegam a cada uma das suas N entradas ao ritmo médio de $\alpha r_{be}/k$ unidades por segundo. As N entradas são equivalentes a uma entrada única com chegadas ao ritmo médio $N\alpha r_{be}/k$ unidades de dados por segundo.

O serviço consiste na transmissão das unidades de dados para a linha de saída a um ritmo fixo de r_{bs} bps. Trata-se de um serviço de capacidade $C = r_{bs}$ efectuando $1/\mu = k$ operações por cliente. S é o tempo que demora a transmitir uma unidade de dados de k bits e é constante, portanto, determinístico.

O parâmetro ρ , neste caso, é a utilização da linha ou a fracção da capacidade total da linha de saída que está a ser usada. Estas correspondências traduzem-se nas seguintes equações:

$$\lambda = N\alpha \frac{r_{be}}{k} \quad (6.7)$$

$$\overline{S} = \frac{1}{\mu C} = \frac{k}{r_{bs}} \quad (6.8)$$

$$\rho = \lambda \overline{S} = N \alpha \frac{r_{be}}{r_{bs}} \quad (6.9)$$

Considere-se as chegadas das unidades de dados aleatória (Poisson), ou seja, que os tempos entre chegadas — os valores $1/\lambda_i$ da figura 6.18 —, se distribuem segundo uma função exponencial negativa. Trata-se de um modelo aproximado para o comportamento das chegadas mas suficientemente realista na maior parte das situações. Considere-se também que as unidades de dados são transmitidas FCFS. O mux estatístico comporta-se então como um sistema M/D/1.

Um dos teoremas fundamentais da teoria das filas de espera é o Teorema de Pollaczek-Khinchin cujo resultado, que se apresenta sem demonstração, para o valor médio do tempo de espera em fila no sistema M/D/1 é

$$\overline{t}_w = \frac{\rho \overline{S}}{2(1 - \rho)} \quad (6.10)$$

O tempo médio total no sistema \overline{t}_q , ou tempo médio de atraso (*delay*), de uma unidade de dados será a soma do tempo médio de transmissão \overline{S} com o tempo médio de espera em fila \overline{t}_w

$$\boxed{\overline{t}_q = \overline{S} + \frac{\rho \overline{S}}{2(1 - \rho)}} \quad (6.11)$$

Por outro lado, pelo teorema de Little, o número médio de clientes no sistema \overline{n}_q , é igual ao produto do ritmo médio de chegadas λ , pelo tempo médio de atraso no sistema \overline{t}_q , ou seja, $\overline{n}_q = \lambda \cdot \overline{t}_q$, o que dá para o número médio de unidades de dados no multiplexador

$$\boxed{\overline{n}_q = \rho + \frac{\rho^2}{2(1 - \rho)}} \quad (6.12)$$

A figura 6.22 representa graficamente as equações (6.12), número médio de unidades de dados no multiplexador, e (6.11), o tempo médio de atraso no multiplexador, neste caso para vários ritmos nominais de saída e unidade de dados de 1000 bits de comprimento, ambos em função da utilização do sistema, ou seja, da utilização da linha de saída, ρ . Esta figura realça a natureza do compromisso entre o *tempo de atraso* no sistema e

a velocidade na linha multiplexada. A utilização da linha é expressa em fracção da sua capacidade (ritmo nominal). Verifica-se que o tempo de atraso cresce de forma aproximadamente linear com a utilização até cerca de 0.6. A partir deste valor o crescimento aumenta rapidamente e, por consequência, a fila de espera (ocupação dos buffers) cresce também mais rapidamente. Uma utilização acima de 0.8 conduz a uma situação de instabilidade que deve ser evitada. Por estas razões se costuma considerar o valor de $\rho = 0.6$ como ponto óptimo de funcionamento de um sistema de fila de espera.

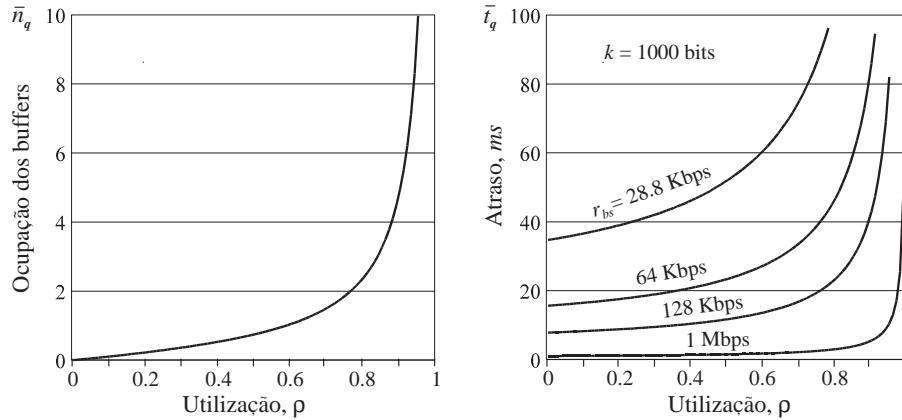


Figura 6.22: Ocupação e atraso médios no MUX estatístico

Para além dos valores médios das distribuições de n_q e de t_q , é importante, evidentemente, ter em consideração médias de ordem superior dado que tanto a ocupação dos buffers como o tempo de atraso, pela sua natureza estatística variam em cada instante podendo mesmo, ocasionalmente, atingir valores muito superiores ao da média. Quando isso acontece, no caso de n_q , é necessário ter espaço de reserva nos buffers para acomodar o excesso temporário de DUs, sob pena de sobrelotação (*overflow*) e da *perda* irremediável de DUs.

A partir da distribuição do número em fila, n_w , é possível determinar a probabilidade de n_w exceder um determinado valor, que é precisamente a probabilidade de *overflow* do buffer caso o seu comprimento seja exactamente esse valor. A expressão dessa distribuição para o sistema M/D/1 é relativamente complexa para ser aqui analisada, no entanto, representam-se na figura 6.23 gráficos para aquelas probabilidades em função do com-

primento do buffer, para diversos valores da utilização.

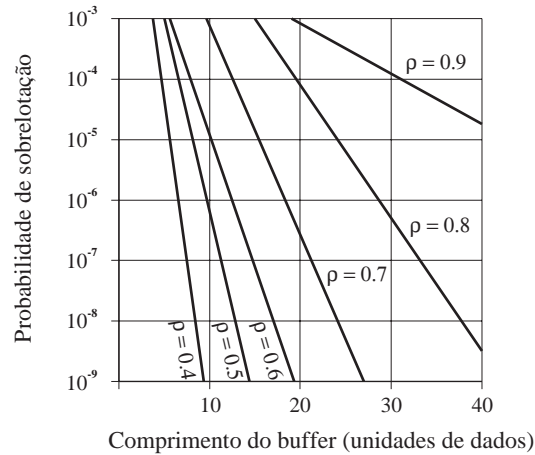


Figura 6.23: Probabilidade de sobrelotação (*overflow*) dos *buffers*

É fácil verificar que a variância do número em fila cresce com a utilização e por conseguinte o comprimento do buffer tem de crescer igualmente para cada determinado valor da probabilidade de sobrelotação. Pode-se, assim, determinar qual o tamanho que os buffers do multiplexador devem ter para se garantir uma taxa máxima de perda de unidades de dados ou, alternativamente, conhecido o comprimento dos buffers, determinar qual aquela taxa ou ainda, dimensionar o número de entradas, ou o ritmo de transmissão à saída de modo a se obter o desempenho desejado, etc.

Exemplo 6.2 *Considere-se um multiplexador estatístico com entradas de 64 Kbps e saída a 128 Kbps. Suponha-se que os equipamentos ligados à entrada transmitem DUs de comprimento fixo igual a 1000 bits durante 40% do tempo a um ritmo aleatório (Poisson), isto é, com tempos entre emissões de DUs distribuídos segundo uma função exponencial negativa. Esta situação traduz-se nos seguintes valores:*

$$\begin{aligned} k &= 1000 && \text{bits} \\ r_{be} &= 64000 && \text{bps} \\ r_{bs} &= 128000 && \text{bps} \\ \alpha &= 0.4 \end{aligned}$$

donde o tempo médio de serviço, que é o tempo médio de transmissão de

uma unidade de dados é constante e igual a

$$\overline{S} = \frac{k}{r_{bs}} = \frac{1000}{128000} \approx 7.8 \text{ ms} \quad (6.13)$$

Considere-se agora os dois casos seguintes.

Caso 1: Estão três equipamentos ligados ao multiplexador, $N = 3$.

O ritmo médio de chegadas de DUs será de $\lambda = N\alpha r_{be}/k = 76.8 \text{ DU/s}$ e o coeficiente de utilização da linha multiplexada é $\rho = \lambda \overline{S} = 0.6$.

O tempo médio de atraso, que é o tempo médio de trânsito de uma DU pelo multiplexador, calculado da equação (6.11), é de $\bar{t}_q \approx 13.7 \text{ ms}$ e o número médio de DUs dentro do multiplexador, calculado pela equação (6.12), é de $\bar{n}_q = 1.05 \text{ DUs}$, ou seja, em média, estão 1050 bits dentro do multiplexador.

Para se garantir que a perda de DUs não excede uma em mil, os buffers de entrada terão de ter espaço para pelo menos 6 DUs, ou seja, 6000 bits (750 bytes), resultado obtido por leitura no gráfico da figura 6.23 para $\rho = 0.6$ e $P_{ovf} = 10^{-3}$.

Se se pretender uma menor taxa de perdas de DUs, por exemplo não inferior a uma em cem mil, então os buffers deverão ser de pelo menos 10000 bits (1.22 Kbytes).

Caso 2: Estão quatro equipamentos ligados ao multiplexador, $N = 4$.

O ritmo médio de chegadas e o coeficiente de utilização são superiores aos do caso anterior, como era de esperar, $\lambda = 102.4 \text{ DU/s}$ e $\rho = 0.8$. O tempo médio de atraso é agora quase o dobro, $\bar{t}_q \approx 23.4 \text{ ms}$, e o número médio de DUs dentro do multiplexador excede o dobro do caso anterior, $\bar{n}_q = 2.4 \text{ DUs}$.

Para garantir que a perda de DUs não excede uma em mil, os buffers de entrada terão de ter espaço para pelo menos 15 DUs, ou seja, 15000 bits (1.83 Kbytes), e para garantir uma taxa de perdas não inferior a uma em cem mil, os buffers terão de acomodar pelo menos 25 DUs, ou 25000 bits (3.125 Kbytes).

Seria, talvez, de esperar, que os valores obtidos para os comprimentos dos buffers em cada caso, fossem divididos proporcionalmente pelas entradas, isto é, por 3 e por 4 respectivamente. Estatisticamente tal só seria válido se o número de entradas fosse muito grande. No caso do exemplo, o buffer de cada linha de entrada deve ter o comprimento que foi calculado.

6.4 Multiplexagem por divisão da frequência

A multiplexagem de um meio de transmissão pode fazer-se também por divisão de frequência, FDM, a qual é possível quando a largura de banda útil excede a largura de banda exigida pelos sinais a transmitir. Vários canais FDM podem coexistir simultaneamente no mesmo suporte físico se os sinais correspondentes forem modulados em *portadoras* de diferentes frequências e estas forem suficientemente separadas por forma a que as bandas dos sinais não se sobreponham senão cada sinal original não poderá ser recuperado por filtragem.

6.4.1 Método

A figura 6.24 mostra os vários aspectos de um sistema FDM. Em *a*), vários sinais de banda base $x_i(t)$, $i = 1..N$, analógicos ou digitais, destinam-se a ser multiplexados no mesmo meio de transmissão. Supondo que estes sinais possuem a mesma largura de banda, B , com espectros $|X_i(f)|$ como se mostra em *b*), eles são previamente filtrados por filtros passa-baixo (FPB) de modo a eliminar todas as suas componentes acima de B Hz resultando o espectro que se mostra em *c*). Cada sinal $x_i(t)$ é modulado em onda contínua e em banda lateral única (BLU²¹) numa portadora de frequência f_{spi} . A modulação em BLU consiste em retirar uma das bandas do espectro do sinal que foi gerada pela modulação. Recorde-se que a modulação adiciona, na banda das frequências positivas, a banda simétrica do sinal que se encontra no intervalo das frequências negativas. Embora simétricas essas duas bandas são iguais bastando transmitir uma delas para se replicar a outra no receptor. Poupa-se assim em largura de banda do canal. Tanto se pode filtrar a banda lateral superior como a inferior dizendo-se que a transmissão se faz, respectivamente, em USB ou LSB²².

Visto serem utilizadas várias portadoras estas são designadas de *sub-portadoras* para se distinguir da modulação quando é efectuada para a transmissão. Pode ser utilizado qualquer tipo de modulação (amplitude, frequência ou fase). Com esta operação, cada sinal resultante $x_{spi}(t)$ possui um espectro deslocado na frequência cujo limite superior é precisamente a frequência f_{spi} da portadora caso se tenha retirado a banda

²¹SSB, *Single SideBand*

²²Upper *SideBand* ou Lower *SideBand*

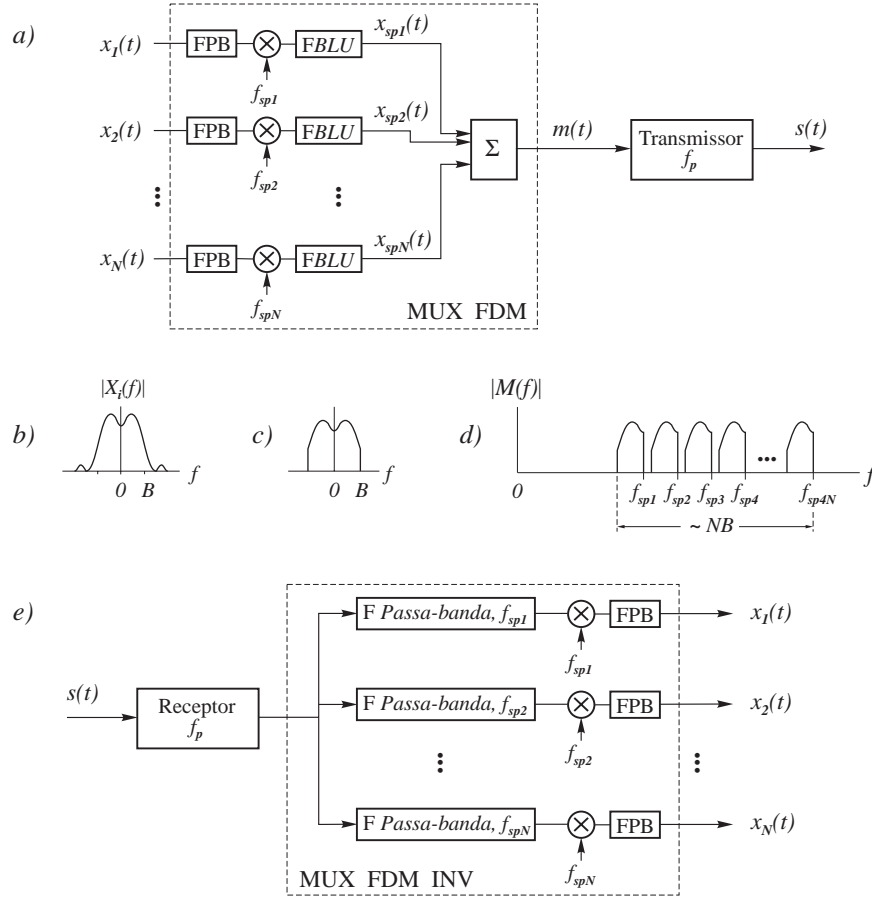


Figura 6.24: Multiplexagem por divisão de frequência, FDM

superior do sinal.

As frequências das subportadoras são escolhidas de modo a que as bandas de cada um dos sinais se situem umas imediatamente a seguir às outras, apenas com uma pequena banda de guarda entre elas. Os $x_{spi}(t)$ são depois somados produzindo um sinal composto $m(t)$ cujo espectro $|M(f)|$ é o representado em d) na figura 6.24.

Por sua vez, o sinal composto $m(t)$ poderá ser modulado quer por uma outra subportadora de um nível superior de multiplexagem, quer por uma portadora de transmissão no canal de acordo com a localização da banda de passagem deste. A largura de banda, B_m , de $m(t)$, será a

soma das larguras de banda de cada um dos sinais nas subportadoras, $B_m = \sum_{i=1}^N B_{spi}$ ou seja, neste caso, $B_m \approx N B$.

Na recepção, depois da desmodulação de transmissão, retirada a portadora f_p , o sinal composto $m(t)$ é apresentado a uma bateria de N filtros passa-banda, cada um com largura de banda B e frequência superior igual a f_{spi} . Desta forma o sinal composto é dividido nas suas partes componentes. Cada componente é então desmodulada por uma subportadora com a mesma frequência superior do filtro e o sinal original desse canal $x_i(t)$ é assim recuperado como se mostra no esquema do multiplexador inverso na figura 6.24 e).

6.4.2 Hierarquias FDM

A tabela 6.4 resume as especificações das hierarquias de multiplexagem FDM americana e a normalizada pela ITU que é adoptada na europa.

Tabela 6.4: Hierarquias de Multiplexagem FDM

	Canais de voz	Largura de Banda	Espectro	Designação	
				Americana	ITU
1	12	48 KHz	60-108 KHz	Group	Group
2	60	240 KHz	312-552 KHz	Supergroup	Supergroup
3	300	1232 KHz	812-2044 KHz	—	Mastergroup
3	600	2520 KHz	564-3084 KHz	Mastergroup	—
4	900	3872 KHz	8516-12388 KHz	—	Supermaster group
×	$N \times 600$			Mastergroup multiplex	—
5	3600	16984 KHz	564-17548 KHz	Jumbogroup	—
6	10800	57442 KHz	3124-60566 KHz	Jumbogroup multiplex	—

Tal como no TDM, o canal básico de referência para a multiplexagem por divisão de frequência é o canal de voz de qualidade telefónica, isto é, o sinal de banda base com $B = 4$ KHz depois de filtrado, incluindo uma banda de guarda. As normas, tanto europeias como americanas, para a multiplexagem de primeiro nível, designado um *grupo*, multiplexa 12 canais de 4 KHz em subportadoras de 64, 68, 72, ..., 108 KHz, em banda lateral inferior, resultando um sinal composto $m(t)$ com largura de banda $B_m = 48$ KHz e frequências limites da banda 60 e 108 KHz.

Formam-se multiplexes de nível superior tratando cada sinal $m_i(t)$ exatamente da mesma forma que se trataram os $x_i(t)$, isto é, modulando-os nas subportadoras definidas para esse nível. A norma para o segundo nível, um *supergrupo*, junta 5 entradas de banda [60, 108 KHz] para formar um sinal composto de largura de banda 240 KHz. As subportadoras que formam o *supergrupo* estão pois separadas de 48 KHz e são de 420, 468, ..., 612 KHz com a modulação também em banda lateral inferior.

Cada nível da hierarquia é identificado por um nome que é o indicado nas duas últimas colunas. Os sistemas diferem no nível de terceira ordem terminando as normas ITU no nível 4. O sistema americano admite multiplexes indefinidos desde que sejam múltiplos de um *Mastergroup*.

6.5 Problemas

- 6.1 – Oito sinais são amostrados e multiplexados em TDM/PAM. Seis dos sinais têm uma largura de banda de 4 KHz e os restantes de 12 KHz.
- a) Qual o ritmo r_c no canal de transmissão se todos os canais forem amostrados à mesma frequência?
 - b) Qual o ritmo r_c se os canais forem amostrados à frequência mínima teórica para cada um deles?
 - c) Esquematize um comutador TDM capaz de efectuar a operação indicada em b).
 - d) Compare os requisitos de largura de banda de transmissão para os casos a) e b)
- 6.2 – Considere que no comutador TDM da figura 6.2 os circuitos de controlo dos comutadores (registo de deslocamento e diferenciadores) são substituídos por uma configuração de microprocessador com memória de programa, porta paralela de I/O, um *timer* com interrupt vectorizado ao CPU e um relógio separado do do processamento para alimentar o *timer*. Apresente um algoritmo para um programa que efectue o controlo da comutação TDM segundo a situação da alínea b) do problema anterior.
- 6.3 – Repita o problema 6.1 considerando que se torna necessário manter uma banda de guarda de 50% para evitar o *aliasing*.
- 6.4 – Mostre que o realinhamento de tramas E1 demora entre 250 e 375 μs .
- 6.5 – Qual o tempo máximo para se obter o alinhamento de tramas T1?
- 6.6 – Qual o ritmo de sinalização médio por canal de sinalização ao nível primário no sistema PDH Americano?
- 6.7 – Qual o ritmo disponível nos canais de sinalização do nível primário do sistema de multiplexagem PDH Europeu? Justifique.
- 6.8 – Quais as percentagens de redundância (*overhead*) nas linhas E1 e T1 respectivamente, isto é, que percentagem do ritmo nominal não é efectivamente utilizada para transmitir informação das fontes?

- 6.9 – O ritmo útil para o utilizador em STM-1 indicado na tabela 6.2 é de 148.608 Mbps. Mostre como é que este valor pode ser calculado a partir dos parâmetros do SDH STM-1.
- 6.10 – Se num sistema SDH os relógios apresentarem um *jitter* constante de 1 parte em 10^9 , quando é que o desvio (atraso ou avanço) iguala o tempo de 1 bit? Quais as implicações deste resultado?
- 6.11 – Qual o ritmo útil disponível numa ligação STM-4c?
- 6.12 – Os campos do endereço nas sub-tramas de multiplexagem estatística poderão ser reduzidos se se utilizar endereçamento relativo passando, possivelmente, a ter comprimento variável. Mostre como se pode efectuar esse endereçamento e, através de um exemplo, quantifique a redução de *overhead* que daí pode resultar.
- 6.13 – Sessenta terminais estão ligados a um concentrador. Metade deles transmite, em média, 1 mensagem cada 15 segundos e a outra metade, 1 mensagem cada 30 segundos, em média. O comprimento das mensagens é de 10 bytes. Determinar o atraso médio das mensagens no concentrador se a linha de saída transmite ao ritmo nominal de 60 bytes/s. Qual o número médio de mensagens dentro do concentrador?
- 6.14 – Repetir o Problema 6.13 para mensagens com metade do comprimento mas com taxa de chegadas dupla. Compare os resultados. Que conclusões pode tirar?
- 6.15 – Repetir o Problema 6.13 com todos os terminais a transmitir, em média, 1 mensagem cada 10 segundos utilizando mensagens de 8 bytes.
- 6.16 – Trinta terminais encontram-se ligados a um multiplexador estatístico. Cada terminal gera em média uma mensagem de comprimento fixo de 1200 bits (pacote) por minuto.
- Determinar a capacidade exigida para a linha de saída (ritmo nominal) de modo que a carga média no sistema, ou seja, o tráfego médio normalizado de saída seja de $\rho = 0.6$.
 - Determinar o tamanho que os buffers devem ter para que a probabilidade de sobrelotação (*overflow*) não exceda 0.0001.
 - Qual o atraso médio que se verifica neste caso?

- 6.17 – Representar graficamente os espectros dos sinais compostos de um grupo e de um supergrupo FDM. Quais das bandas laterais do sinal básico original estão presentes em cada caso?
- 6.18 – Mostre, através de gráficos espectrais, como é que um mux inverso de grupo FDM recupera ambas as bandas laterais dos sinais $x_i(t)$ e portanto, os próprios sinais originais.

fim do capítulo 6

Capítulo 7

Códigos de Linha

Imediatamente antes da sua transmissão no canal, a informação sob a forma digital é convertida em sinais físicos (eléctricos ou ópticos) possuindo certas características nomeadamente, uma determinada localização e largura espectrais, facilidade de detecção no receptor, rendimento de transmissão optimizado e pequena vulnerabilidade aos erros. Esta operação designa-se por *formatação* ou *codificação de linha* e o conjunto dos elementos de sinal ou símbolos físicos utilizados para representar os dígitos de informação ou suas sequências constituem o *código de linha*.

Fundamentalmente os sinais podem ser transmitidos sob duas formas:

1. em banda base
2. em banda passante ou banda de canal

Na primeira forma o sinal é transmitido na banda de frequências em que ele próprio é produzido e se situa – a banda base. Na segunda forma o sinal é modulado em onda contínua com uma frequência de portadora adequada à banda do sistema de transmissão. Assim, em banda de base a transmissão é digital e em banda de canal a transmissão é analógica embora em ambos os casos os *dados* da fonte de *informação* sejam digitais.

7.1 Códigos de banda base

Um sinal digital é uma sequência discreta, descontínua, de pulsos de tensão eléctrica ou intensidade luminosa. Um pulso é normalmente uma

forma aproximadamente rectangular no tempo e constitui o que se designa também por símbolo de linha. Cada pulso é um *elemento de sinal*.

Ritmo de modulação

O *ritmo binário de dados*, ou apenas *ritmo binário*, de um sinal é a velocidade, em dígitos binários por segundo (bps), em que os dados de informação são transmitidos. A *duração*, *período* ou *tempo* de um dígito binário é a quantidade de tempo que o transmissor demora a emitir o dígito. Para um ritmo binário de r_b bps a duração é $T_b = 1/r_b$ s. Por outro lado, designa-se por *ritmo de modulação*¹ à velocidade, r_c , com que os elementos de sinal se sucedem no canal de transmissão o que se expressa em *baud*. Uma outra forma de caracterizar o ritmo binário é a forma *normalizada* ao ritmo de modulação, ou seja, definindo-o pelo número médio de dígitos de dados por transição do sinal² que, no caso da codificação digital se expressa em *bits/aud* e no caso da codificação em onda contínua, em *bits/Hz*.

Sincronismo e regeneração

A interpretação, ou seja a determinação do *valor* de um sinal digital, faz-se aquando da sua regeneração como se referiu na secção 4.3 do capítulo 4. Um aspecto importante, senão mesmo essencial, é o do *sincronismo* entre o instante em que se toma a amostra ao (integral³ do) elemento do sinal e o instante central deste. Este sincronismo tem de ser obtido através da reprodução, no receptor, do relógio do transmissor, tanto em frequência como principalmente em fase. Dado que não se utilizam canais paralelos para transmitir o relógio, de alguma forma este tem de estar implícito no próprio sinal, isto é, tem de ser codificado com o sinal. Um bom código de linha, portanto, é aquele que melhor consegue codificar o relógio juntamente com o sinal sem com isso prejudicar nem as propriedades do sinal, como seja não aumentar a sua largura de banda, nem a complexidade do equipamento com o algoritmo de codificação.

¹ou apenas *ritmo* do sinal digital.

²Por razões históricas oriundas da utilização do código de Morse, os elementos de um sinal binário, também se costumam designar por *mark* e *space* ou *traço* e *ponto*, correspondendo respectivamente ao 1 e ao 0

³o filtro passa-baixo é um integrador

Polaridade do código

Os dados sob a forma binária são transmitidos codificando os bits em elementos de sinal. No caso mais simples existe uma correspondência um-para-um entre bits e elementos de sinal. Se estes possuírem todos o mesmo sinal algébrico, isto é, todos positivos ou todos negativos, o código de linha designa-se *unipolar*. Se um dos valores lógicos binários fôr representado por um nível de tensão positivo e o outro por um nível de tensão negativo o código designa-se *polar*.

Critérios de avaliação de desempenho

Uns códigos de linha serão melhores que outros mas a sua avaliação objectiva é feita face às seguintes propriedades:

- Espectro. A ausência de componentes de alta frequência no espectro significa que o sinal (código de linha) exige do sistema menor largura de banda de transmissão. A ausência de componente contínua também é desejável visto, por um lado, permitir o acoplamento electromagnético (por transformador) entre o transmissor e o meio de transmissão isolando-os electricamente e, por outro lado, evitar a utilização de energia de sinal numa componente que não transporta informação sendo, neste sentido, inútil.
- Sincronismo. Como se disse, a capacidade do código de linha incluir o relógio do transmissor de modo a poder ser totalmente reproduzido, por separação, no receptor e assim possibilitar o sincronismo na regeneração do sinal.
- Deteccção de erros. Independentemente do control de erros (deteccção ou correcção) se fazer também a outro nível, isto é, antes da transmissão e depois da recepção utilizando sequências de dígitos redundantes, como se verá no capítulo sobre Teoria da Informação, o código de linha poderá possuir alguma capacidade de deteccção de erros ao nível do próprio sinal físico.
- Imunidade ao ruído e à interferência. Alguns códigos poderão exibir um melhor desempenho face ao ruído e à interferência que outros.

- Custo e complexidade. Embora a codificação de linha se faça, hoje em dia, toda em hardware e o custo do hardware continue a baixar, uma maior complexidade do processo de codificação encarece o equipamento. Por outro lado, o custo deste também aumenta com o ritmo de modulação. Quanto maior for este ritmo em relação ao ritmo binário, mais caro se torna o equipamento.

7.1.1 Não Retorno a Zero (NRZ)

O formato mais simples e usual para transmitir sinais digitais binários é o que utiliza dois níveis diferentes de tensão eléctrica para os dois dígitos binários. Nos códigos que utilizam este formato o nível de tensão é constante durante o período do bit. Não existe transição de nível nem um retorno à tensão zero durante aquele intervalo de tempo.

A tensão zero pode ser utilizada para representar o 0 lógico e uma tensão positiva o 1 lógico. A situação mais usual, no entanto, é a *polar*, em que um dos valores lógicos é representado por uma tensão negativa e o outro por uma tensão positiva, respectivamente o 1 e o 0. Trata-se do código polar NRZ-L⁴ representado na figura 7.1. O NRZ-L é o código é o normalmente utilizado para interpretar os dados produzidos por equipamentos digitais ou circuitos lógicos. Para a transmissão à distância utilizam-se outros códigos que normalmente tomam este como referência definidora do valor de sequências de dados.

Uma variante deste é o NRZI ou Não Retorno a Zero Inversão. Este código é um NRZ que inverte a polaridade no início de um valor lógico igual a 1. Os dados são pois codificados pela ausência ou presença de uma transição de nível de tensão no início do período do bit. Trata-se de um exemplo do que se designa por *codificação diferencial*. A descodificação processa-se por comparação da polaridade de elementos de sinal adjacentes.

A codificação diferencial tem normalmente mais vantagens do que inconvenientes relativamente a uma codificação absoluta. É mais fiável, por exemplo, detectar uma transição de nível face ao ruído, do que o valor absoluto da amplitude de um elemento do sinal pois uma simples troca de fios num circuito a dois fios inverteria os valores lógicos.

Os códigos NRZ utilizam eficazmente a banda disponível como se ilustra na figura 7.2 onde se comparam as densidades espectrais de potência de

⁴*Non-Return to Zero-Level*

vários códigos de banda base. Nesta figura, o eixo da frequência está normalizado ao ritmo binário. Como se pode verificar, a maior parte da potência dos códigos NRZ localiza-se entre 0 Hz e metade do ritmo de modulação, r_c .

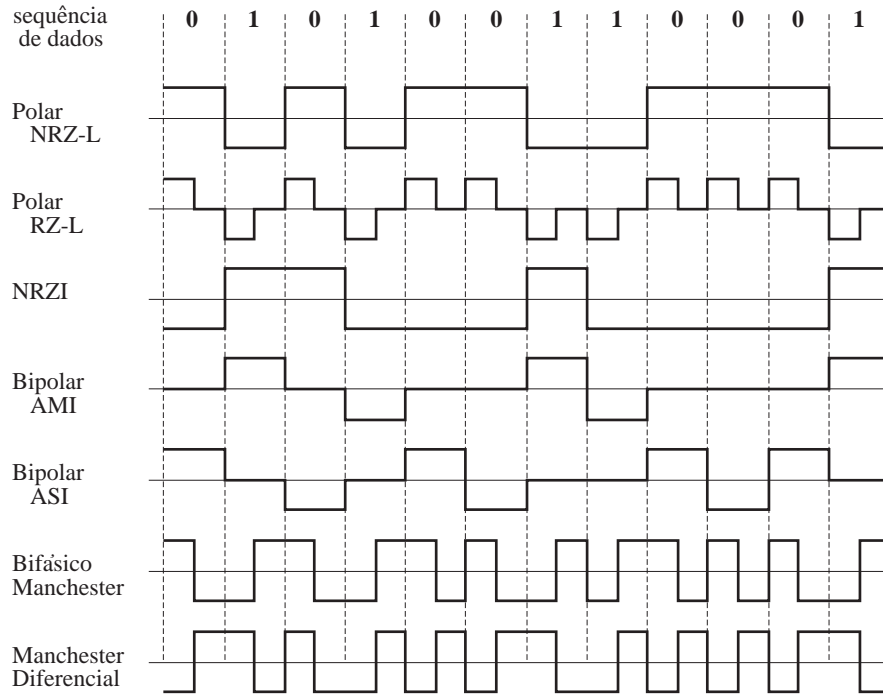


Figura 7.1: Códigos de linha de banda base

As maiores limitações dos códigos NRZ são a da presença de componente contínua e a da falta de capacidade de sincronismo. Uma sequência muito longa de 0 ou 1 seguidos para o NRZ-L ou de 0 para o NRZI resulta numa tensão contínua muito prolongada. Se houver um deslize na frequência do relógio do receptor relativamente à do transmissor poder-se-á perder o sincronismo de regeneração.

Dada a sua simplicidade e conteúdo de baixas frequências, os códigos NRZ são mais utilizados na gravação magnética digital. As limitações descritas tornam-nos pouco atractivos para a transmissão digital.

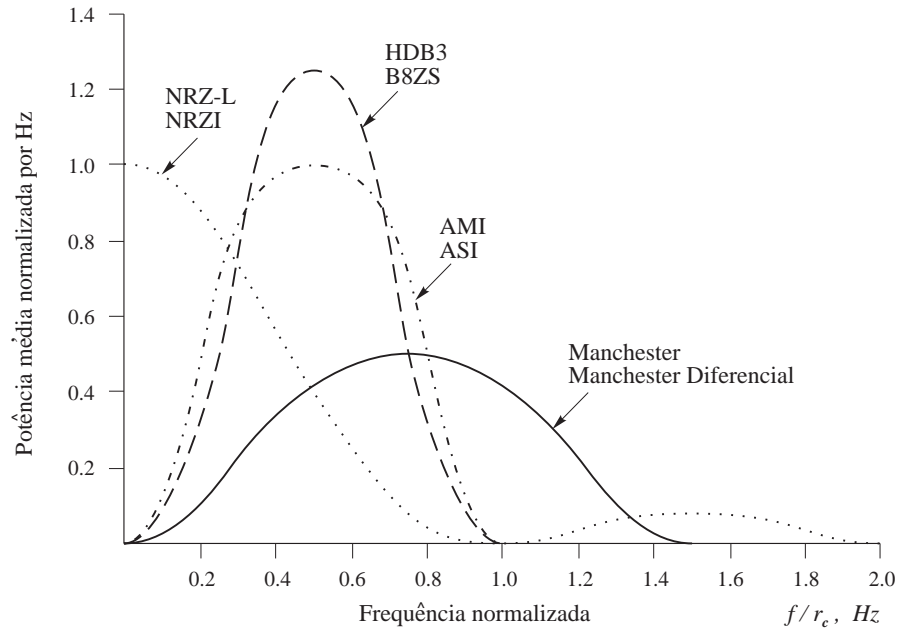


Figura 7.2: Densidade espectral de potência de códigos de banda base

7.1.2 Bipolar AMI

Os códigos bipolares, também designados de binários multinível, tentam resolver algumas das desvantagens dos NRZ. Um código de linha binário é *bipolar* quando um dos valores lógicos é representado alternadamente por dois níveis de tensão positiva e negativa, e o outro valor lógico pela tensão zero. Se é o lógico 1 que alterna de polaridade o código designa-se AMI (Alternate Mark Inversion), se é o lógico 0, designa-se ASI (Alternate Space Inversion).

Como estão em jogo três tensões distintas para representar valores binários os códigos bipolares são também designados de *pseudoternários*.

No AMI, uma sequência longa de 1 não dá origem a perda de sincronismo dado que há variação periódica de nível embora isso não aconteça em sequências prolongadas de 0.

Dado que os elementos de sinal de amplitude não-zero alternam sempre de polaridade, o espectro não possui componente contínua. Adicionalmente,

a largura de banda deste código é menor do que a do NRZ. Por outro lado, a alternância (obrigatória) de polaridade resulta num mecanismo de detecção de erros pois qualquer erro isolado *viola* esta regra.

Embora possuam alguma capacidade de sincronismo esta não é total, como se referiu. Para colmatar esta desvantagem existem técnicas que recorrem à inserção forçada de pulsos (códigos *scrambling*).

O código bipolar, por outro lado, não é tão rentável como o NRZ dado que necessita de três níveis para representar valores binários⁵.

Uma outra forma de abordar a questão da rentabilidade é a que se baseia na probabilidade de erro em função da energia média por símbolo (elemento de sinal), E_b/η . Atendendo a que o receptor tem de distinguir entre três níveis de tensão, $+A$, 0 e $-A$, uma análise deste código semelhante à que se fez no Capítulo 4 para o caso do código unipolar, revela que para a mesma probabilidade de erro por bit (P_e), o código bipolar exige mais cerca de 3 dB de potência do que os polares (NRZ). Ou seja, para a mesma razão sinal-ruído, o código NRZ possui um BER⁶ significativamente menor do que o bipolar. A figura 7.3 representa os BER destes códigos em função de E_b/η em decibéis que é a forma mais usual de representar a função da probabilidade de erro da tabela 4.2.

7.1.3 Bifásico Manchester

Um outro conjunto de códigos de linha designados *bifásicos* tentam também resolver as limitações do NRZ. Dois dos mais importantes são os códigos Manchester e Manchester Diferencial.

No código *Manchester*, o elemento de sinal possui sempre uma transição a meio do intervalo. Esta transição codifica simultaneamente o relógio e os dados. Se a transição é do nível de tensão negativo para o positivo ela representa o lógico 1 e se a transição é no sentido contrário representa o lógico 0.

No código *Manchester Diferencial* a transição a meio do período só é utilizada para recuperar a fase do relógio. A codificação do 0 é representada pela presença de uma transição no início do período e a do 1 pela ausência

⁵Cada elemento de sinal representa no máximo *um bit* de informação enquanto poderia representar $\log_2 3 = 1.58$ bits como se verá no capítulo sobre a Teoria da Informação.

⁶*Bit Error Rate* ou P_e

de transição no início do período. Trata-se portanto de uma codificação diferencial tal como a do NRZI.

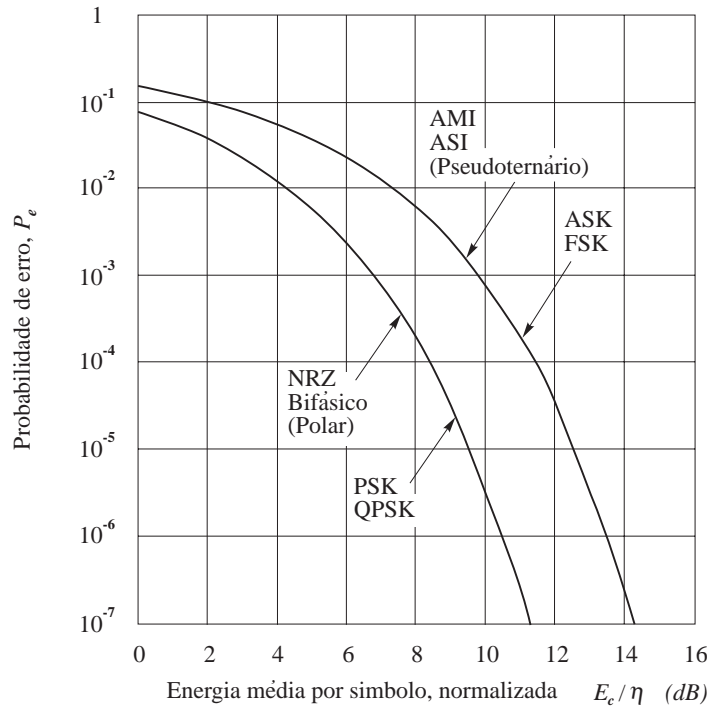


Figura 7.3: BER teórico de alguns códigos de linha

Todas as técnicas bifásicas possuem pelo menos uma transição por elemento de sinal (símbolo) podendo haver até duas transições. Portanto o ritmo máximo de modulação é o dobro do do NRZ o que significa que a largura de banda exigida pelo bifásico é maior. Esta desvantagem é compensada pelas seguintes vantagens:

- Sincronismo: Dado que existe sempre uma transição a meio do elemento de sinal, o receptor pode sincronizar-se nessa transição. Os códigos bifásicos são portanto auto-sincronizados.
- Ausência de componente contínua: vantagem já descrita anteriormente.
- Deteção de erros: A ausência da esperada transição pode ser utilizada para detectar um erro. O ruído na linha teria de inverter o

sinal tanto antes como depois da transição esperada de modo a que o erro passasse despercebido.

Como se pode observar na figura 7.2 a maior parte da potência nos códigos bifásicos está localizada entre 0.5 e 1 vezes o ritmo de modulação. A sua banda é, portanto, relativamente estreita e não contém componente contínua. Contudo, é mais larga do que a banda dos códigos bipolares.

Os códigos bifásicos são bastante utilizados na transmissão de dados especialmente em redes locais. O mais vulgar é o código Manchester que está adoptado nas normas IEEE 802.3 e *ethernet* para as redes locais CSMA/CD⁷ de banda base, tanto em cabo coaxial como em pares entrançados (UTP e STP⁸).

O código Manchester Diferencial é o estabelecido na norma IEEE 802.5 para as redes locais de banda base em anel com passagem de testemunho⁹ que utiliza pares entrançados com malha (STP) como meio de transmissão.

7.1.4 Códigos Scrambling

Embora as técnicas bifásicas sejam amplamente utilizadas nas redes locais de relativo alto ritmo (até 10 Mbps), não o são na transmissão a longas distâncias. A razão principal advém do facto de possuírem um ritmo de modulação superior ao ritmo binário. Dado que os custos dos circuitos de longa distância constituem uma parcela muito significativa dos custos totais dos sistemas, existe a preocupação em maximizar a sua utilização designadamente procurando-se códigos de linha que consigam transmitir o maior número de *bits/Hz* da banda disponível o que na prática se traduz na procura de ritmos de modulação igual ou inferior a 1.

Para tal recorre-se a técnicas de *scrambling*¹⁰ ou de redução de ritmo. As primeiras consistem basicamente na substituição das sequências binárias que resultariam em níveis de tensão constantes, por sequências de *enchimento* que forneçam um número de transições suficientes não só para

⁷ *Carrier-Sense Multiple Access/Collision Detection*

⁸ UTP, *Unshielded Twisted Pair*; STP, *Shielded Twisted Pair*, pares entrançados, sem e com malha respectivamente

⁹ *token ring*

¹⁰ os *scramblers* são *misturadores* pseudo-aleatórios destinados a tornar os dados ininteligíveis a quem não possua a *chave* para os decifrar. No caso da codificação de linha, porém, os *scramblers* são utilizados para *aleatorizar* os dígitos.

que o relógio do receptor se mantenha sincronizado como também para reduzir ao mínimo a componente contínua do sinal.

Essas sequências de *enchimento* têm de ser reconhecíveis pelo receptor que as substituirá pelas originais. Possuem o mesmo número de dígitos que as originais e portanto não introduzem aumento de ritmo de modulação.

As características de um código *scrambling* são portanto as seguintes:

- Ausência de componente contínua.
- Ausência de longas sequências de elementos de sinal de nível zero.
- Manutenção do ritmo binário.
- Capacidade de detecção de erros ao nível do sinal.

HDB3

O código HDB3, abreviatura de *High-Density Bipolar 3 zeros*, baseia-se no código AMI. As sequências de quatro zeros seguidos são substituídas por sequências possuindo um ou dois pulsos. Em cada um dos casos o quarto zero é substituído por uma violação do código AMI. A tabela 7.1 resume as regras de formação deste código de linha. Estas regras asseguram que duas violações AMI seguidas são de polaridade simétrica de modo a evitar a componente contínua.

Com efeito, a regra verifica se o número de pulsos desde a última violação do código AMI é par ou ímpar, bem como a polaridade do último pulso imediatamente anterior à ocorrência dos quatro zeros sucessivos, escolhendo depois a sequência de substituição apropriada. Este código de

Tabela 7.1: Regras de substituição do código HDB3

Polaridade do pulso precedente	Número de pulsos bipolares desde a última substituição	
	Ímpar	Par
–	000–	+00+
+	000+	–00–

linha é utilizado nos sistemas de transmissão europeus e japoneses.

B8ZS

O código B8ZS, abreviatura de *Bipolar 8 Zeros Substituição*, também se baseia no AMI que é modificado de acordo com as seguintes regras:

- i) Quando ocorre um octeto de zeros e o pulso precedente é positivo, o octeto é substituído pela sequência 000+−0−+
- ii) Quando ocorre um octeto de zeros e o pulso precedente é negativo, o octeto é substituído pela sequência 000−+0+−

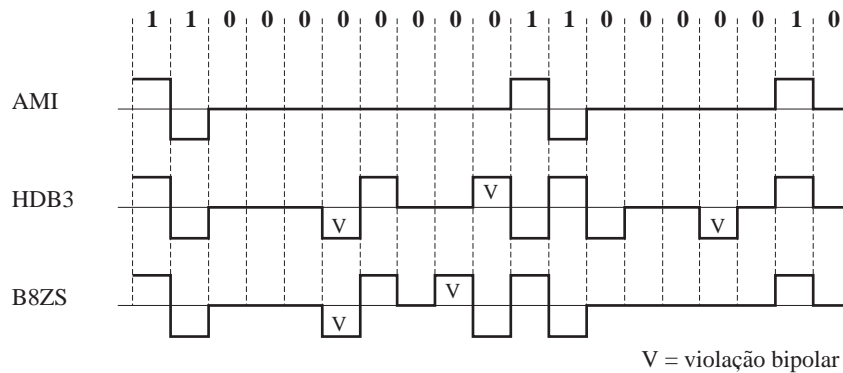


Figura 7.4: Códigos *scrambling* HDB3 e S8ZS

Estas regras forçam duas violações do código AMI situação pouco provável de poder ser atribuída aos efeitos do ruído. O receptor reconhece estes padrões e substitui-os por octetos de zeros. A figura 7.4 exemplifica a formação destes dois códigos. O código B8ZS é o utilizado nos sistemas de transmissão americanos.

As propriedades espectrais dos códigos HDB3 e B8ZS estão patentes na figura 7.2. Como se pode verificar, nenhum deles possui componente contínua e a maior parte da energia está concentrada numa banda relativamente estreita em torno de uma frequência que é igual a metade do ritmo binário.

Os códigos *scrambling* são utilizados nas transmissões de alto ritmo a longas distâncias. Os circuitos de acesso à RDIS de banda larga¹¹ utilizam estes códigos de linha.

¹¹Broadband ISDN, (*Integrated Services Digital Network* ou Rede Digital com Integração de Serviços)

7.1.5 Códigos de redução de ritmo, mBnL

Os códigos de redução de ritmo destinam-se a transmitir mais de um bit por elemento de sinal, ou seja, mais de um bit/s por baud. Uma das vantagens destes códigos é a de reduzirem o *crosstalk* porque as variações de amplitude entre pulsos consecutivos são menores. São designados por códigos *mBnL* o que quer dizer que uma sequência de m Bits de dados é representada por n pulsos de L níveis cada, com $n < m$ e $L > 2$.

4B3T

No código 4B3T, cada grupo de 4 dígitos Binários de dados são codificados em 3 pulsos que poderão ter três níveis de amplitude distintos, ou seja, em 3 dígitos Ternários. Obtém-se assim uma redução de ritmo de pulsos na linha de 1/4 relativamente ao ritmo binário dos dados. Os três pulsos do código de linha transmitidos por cada quatro bits da sequência de dados são seleccionados de uma das quatro colunas da tabela 7.2. Associado a

Tabela 7.2: Padrões de codificação de linha 4B3T

Sequência binária (dados)	1		2		3		4	
	Código	Col seg	Código	Col seg	Código	Col seg	Código	Col seg
0001	0 - +	1	0 - +	2	0 - +	3	0 - +	4
0111	- 0 +	1	- 0 +	2	- 0 +	3	- 0 +	4
0100	- + 0	1	- + 0	2	- + 0	3	- + 0	4
0010	+ - 0	1	+ - 0	2	+ - 0	3	+ - 0	4
1011	+ 0 -	1	+ 0 -	2	+ 0 -	3	+ 0 -	4
1110	0 + -	1	0 + -	2	0 + -	3	0 + -	4
1001	+ - +	2	+ - +	3	+ - +	4	- - -	1
0011	0 0 +	2	0 0 +	3	0 0 +	4	- - 0	2
1101	0 + 0	2	0 + 0	3	0 + 0	4	- 0 -	2
1000	+ 0 0	2	+ 0 0	3	+ 0 0	4	0 - -	2
0110	- + +	2	- + +	3	- + +	4	- - +	3
1010	+ + -	2	+ + -	3	+ - -	2	+ - -	3
1111	+ + 0	3	0 0 -	1	0 0 -	2	0 0 -	3
0000	+ 0 +	3	0 - 0	1	0 - 0	2	0 - 0	3
0101	0 + +	4	- 0 0	1	- 0 0	2	- 0 0	3
1100	+ + +	4	- + -	1	- + -	2	- + -	3

Nota: o código 000 é descodificado como 0000

cada código numa coluna existe um número que indica a coluna de onde o código seguinte deve ser escolhido. No exemplo da figura 7.5 o código

para a sequência 0111 é escolhido da coluna 1 ($-0+$) e a coluna seguinte é a 1. O código para a sequência seguinte, 1000, é escolhido da coluna 1 ($+00$) e a coluna seguinte é a 2. O código para a sequência seguinte, 1110, é escolhido da coluna 2 ($0+-$) e a coluna seguinte é a 2, e assim sucessivamente.

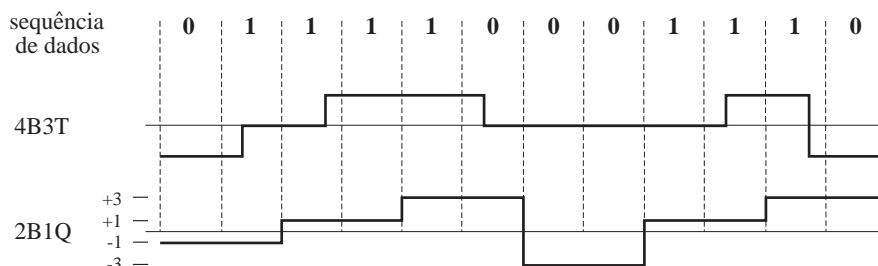


Figura 7.5: Códigos de redução de ritmo, $mBnL$

Existem $3^3 = 27$ possíveis ternos de pulsos ternários para codificar as $2^4 = 16$ diferentes quadras de bits de dados pelo que $27-16=11$ ternos não são usados. Se no receptor aparecer algum destes ternos pode concluir-se que houve erro na transmissão. A redundância do código permitiu torná-lo pelo menos detector de erros. Na escolha dos ternos do código procura-se que a uma sequência binária aleatória de dados corresponda um sinal com uma largura de banda e valor médio o mais reduzidos possível.

2B1Q

No código 2B1Q, cada par de bits dos dados é codificado num único pulso de quatro níveis, ou pulso Quaternário. Na figura 7.5 os quatro níveis estão representados pelas amplitudes $+3$, $+1$, -1 e -3 indicando que os níveis são simétricos relativamente a zero e estão igualmente espaçados entre si. O primeiro bit em cada par indica a polaridade e o segundo a amplitude.

Este código não é redundante mas apresenta uma redução de ritmo de $1/2$. É um dos códigos de linha mais utilizados.

Opera a ritmos binários de 160 Kbps em linhas de pares de fios entrançados de vários quilómetros de comprimento e é utilizado nos circuitos de acesso básico RDIS (RDIS de banda estreita).

7.2 Códigos de banda de canal

Os códigos de banda de canal, também chamados de banda passante, são obtidos por modulação digital de onda contínua. São os códigos adotados para possibilitar a transmissão digital em circuitos originalmente concebidos para transmissão analógica pelo que são largamente utilizados em *modems* destinados a transmitir sobre linhas telefônicas analógicas, quer comutadas quer dedicadas (ponto-a-ponto). Com a digitalização dos circuitos e das redes de comunicações, estes códigos tenderão a ser cada vez menos utilizados.

7.2.1 Técnicas fundamentais de codificação

Nesta modulação o sinal digital modulante faz variar uma das três variáveis de uma onda portadora sinusoidal: a amplitude, a frequência ou a fase dando origem, respectivamente, às três técnicas de modulação digital:

- ASK: Amplitude Shift Keying
- FSK: Frequency Shift Keying
- PSK: Phase Shift Keying

que se encontram ilustradas na figura 7.6. O termo *keying* permanece desde o tempo em que a modulação digital foi utilizada pela primeira vez na transmissão dos sinais binários do código Morse os quais eram gerados por uma *chave* (*key*), ou interruptor, que permitia a emissão de uma portadora, ou não, consoante era premida ou não de acordo com a sequência binária dos dados a transmitir. Este modo de transmissão em que ao valor 1 correspondia a existência de portadora de amplitude e frequência constantes e ao valor 0 a ausência de portadora, ou seja, amplitude zero, é designado por On-Off Keying (OOK).

Em qualquer das técnicas de modulação o espectro do sinal resultante ocupa uma banda centrada na frequência da portadora.

Em ASK os valores binários são representados por duas amplitudes diferentes da portadora de frequência constante f_p . Normalmente uma das amplitudes é zero. O sinal resultante é

$$\text{ASK} \quad s(t) = \begin{cases} A \cos(2\pi f_p t + \theta_p) & \text{binário 1} \\ 0 & \text{binário 0} \end{cases}$$

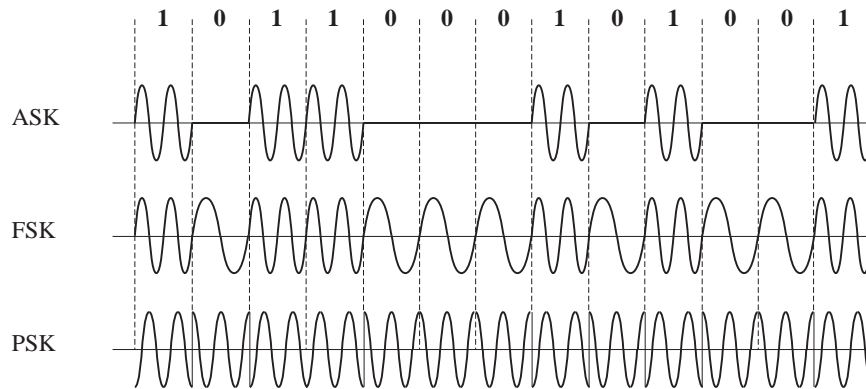


Figura 7.6: Códigos de linha em banda de canal (modulação digital)

Em FSK os dois valores binários são representados por dois valores da frequência, f_1 e f_2 , acima e abaixo da frequência da portadora, e igualmente espaçadas desta, $f_2 - f_p = f_p - f_1$. O sinal resultante é

$$\text{FSK} \quad s(t) = \begin{cases} A \cos(2\pi f_1 t + \theta_p) & \text{binário 1} \\ A \cos(2\pi f_2 t + \theta_p) & \text{binário 0} \end{cases}$$

A modulação FSK é menos susceptível a erros do que a ASK pois o ruído nos sistemas de transmissão é essencialmente aditivo e portanto manifesta-se essencialmente na amplitude do sinal que no caso FSK não transporta informação sobre o valor do dígito binário. Esta informação está contida no valor da frequência da portadora ao longo do tempo.

Em modulação digital de fase a amplitude e a frequência da portadora são mantidas constantes e a fase é deslocada de um certo angulo de acordo com a sequência binária dos dados. Uma forma de modulação de fase utiliza duas portadoras de igual frequência fixa mas diferindo 180° em fase, isto é, duas sinusoides inversas aditivas uma da outra. O sinal resultante é

$$\text{PSK} \quad s(t) = \begin{cases} A \cos(2\pi f_p t + 180^\circ) & \text{binário 1} \\ A \cos(2\pi f_p t) & \text{binário 0} \end{cases}$$

e trata-se de uma codificação, ou modulação, PSK. Dado que com esta técnica é necessário reproduzir e manter no receptor uma fase de referência de modo a compará-la com a do sinal, este método é designado

de PSK coerente. Na prática este requisito traduz-se numa grande complexidade do receptor/desmodulador agravado ainda pelo facto de a transmissão estar frequentemente sujeita a grandes variações aleatórias de fase. Por estas razões é normalmente utilizada uma variante desta técnica pela qual o valor binário dos dados não está contido no valor absoluto da fase da portadora mas sim na variação de fase relativamente à do dígito anterior.

Assim uma mudança de fase de 90° relativamente à fase corrente representa o binário 0 e uma mudança de 270° representa o binário 1. Desta forma o receptor só necessita determinar o valor da mudança de fase. Esta técnica é designada de PSK diferencial, DPSK, e é a mais utilizada na prática.

$$\text{DPSK} \quad s(t) = \begin{cases} A \cos(2\pi f_p t + \theta_{i-1} + 270^\circ) & \text{binário 1} \\ A \cos(2\pi f_p t + \theta_{i-1} + 90^\circ) & \text{binário 0} \end{cases}$$

A figura 7.7 ilustra a ocupação espectral de cada uma destas técnicas. Considerando que o espectro do sinal digital é o representado em a), cor-

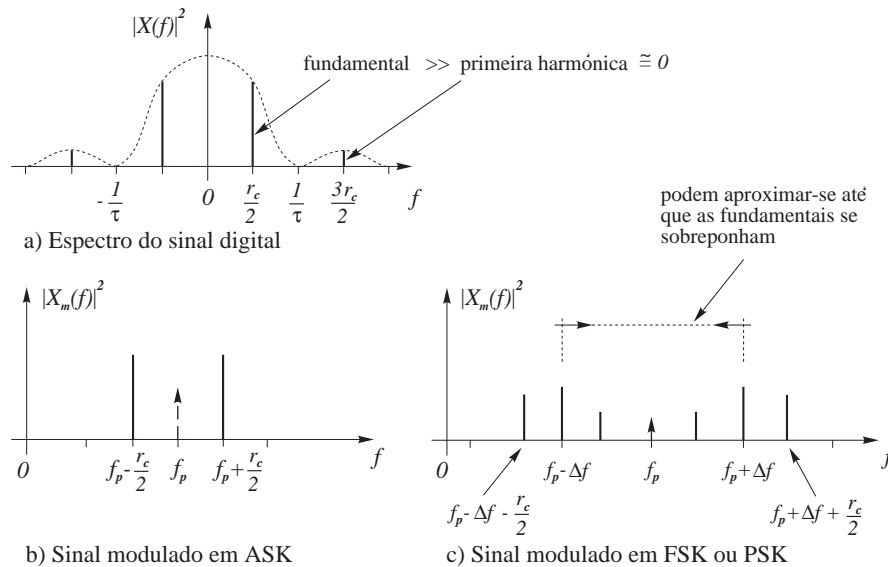


Figura 7.7: Espectros dos sinais ASK, FSK e PSK

respondente a uma sequência periódica de pulsos rectangulares de duração $\tau = T_c$ de tal forma que o ritmo de símbolos é $r_c = 1/T_c$ simb/s, então

a sua codificação em ASK com uma portadora de frequência f_p terá o espectro representado em *b*) e a sua codificação quer FSK quer PSK ou DPSK possuirá uma ocupação espectral como se representa em *c*). É de notar que no caso ASK a largura da banda ocupada pelo sinal codificado é igual ao ritmo de símbolos, dado que neste caso as harmónicas de ordem par são nulas, e nos casos FSK e PSK a largura da banda é de pelo menos 2 vezes superior àquele ritmo. Na realidade a variação Δf mínima da frequência da portadora, f_p , tem de ser superior a r_c para que as componentes fundamentais acima de $f_p - \Delta f$ e abaixo de $f_p + \Delta f$ não se sobreponham.

7.2.2 Codificação *M*-ária

Quando não é possível aumentar o ritmo binário devido a limitação de largura de banda do canal (condição de Nyquist) ou quando se deseje uma ocupação mínima da banda disponível, por questões de rentabilidade, recorre-se a codificação multinível, uma forma de codificação por redução de ritmo ($r_c < r_b$), desde que a consequente diminuição da qualidade da comunicação (devido ao aumento da probabilidade de erro) seja ainda aceitável e, evidentemente, não se pretenda ultrapassar o limite teórico estabelecido pelo Teorema de Shannon da capacidade do canal¹².

Assim, com pulsos multinível *M*-ários, tem-se $\log_2 M$ bits por pulso (símbolo) e portanto $r_b = r_c \cdot \log_2 M$ bits/s. Os casos mais frequentes são $M = 4$ ou $M = 8$ em que os elementos de sinal se designam respectivamente por *dibits* e *tribits* pelo facto de representarem respectivamente dois e três dígitos binários.

Na codificação *M*-ária em ASK utilizam-se *M* amplitudes diferentes para a portadora e em FSK a portadora varia entre *M* frequências diferentes igualmente espaçadas.

Mas as realizações mais usuais empregam a modulação digital de fase *M*-ária, também designada por modulação em *quadratura* ou QPSK (Quadrature Phase Shift Keying). Assim, por exemplo, o sinal quaternário ($M = 4$) QPSK possuirá quatro fases distintas em que os valores binários da informação estão contidos ou no valor absoluto da fase (QPSK coerente) ou na variação de fase relativamente ao símbolo anterior (modulação QPSK Diferencial ou DQPSK).

¹²discutido no capítulo sobre Teoria da Informação

Em lugar de serem representados analiticamente por expressões como

$$\text{QPSK} \quad s(t) = \begin{cases} A \cos(2\pi f_p t + 45^\circ) & \text{binário 11} \\ A \cos(2\pi f_p t + 135^\circ) & \text{binário 10} \\ A \cos(2\pi f_p t + 225^\circ) & \text{binário 00} \\ A \cos(2\pi f_p t + 315^\circ) & \text{binário 01} \end{cases} \quad (7.1)$$

são representados pelo que se designa por *constelação* do sinal que é o conjunto dos valores do sinal sob a forma de afixos no plano complexo. Os vectores desses afixos representam a amplitude e a fase de cada elemento do sinal QPSK. A figura 7.8 a) representa a constelação do sinal QPSK da equação 7.1 etiquetada com os correspondentes valores binários.

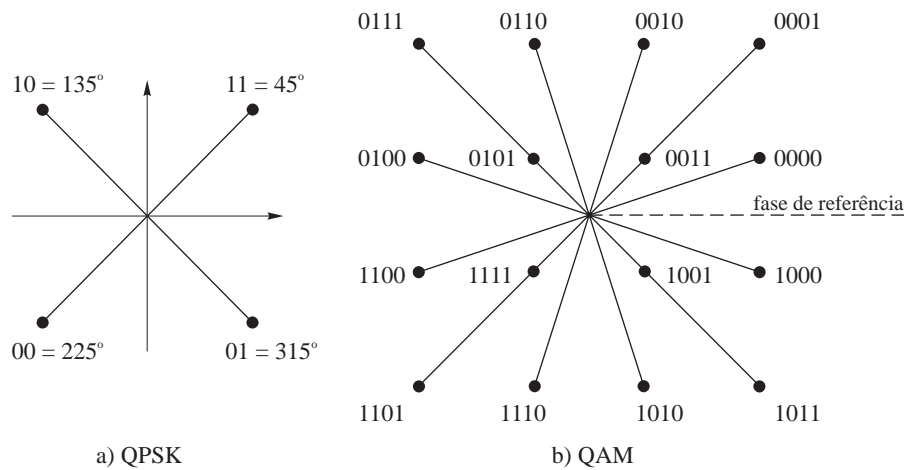


Figura 7.8: Constelações de sinais modulados em quadratura

7.2.3 Técnicas híbridas

Uma das formas mais eficazes de ultrapassar as limitações da condição de Nyquist é a utilização de técnicas híbridas que consistem na combinação das técnicas fundamentais numa única operação de codificação. Uma das técnicas híbridas mais usadas resulta da combinação da modulação em amplitude com a modulação de fase, técnica que se designa por QAM (Quadrature Amplitude Modulation).

Desta forma pode obter-se um sinal QAM representando, por exemplo, quatro bits por elemento (símbolo), combinando 12 valores de fase e 3

valores de amplitude para produzir 16 elementos (valores) distintos do sinal. A figura 7.8 b) representa a constelação deste sinal. Esta foi, na realidade, a primeira implementação de QAM, que permitia um ritmo binário de 9600 bps (bits por segundo) com o sinal operando a 2400 baud (símbolos por segundo) numa linha telefónica analógica cuja largura de banda é, como se sabe, de cerca de 3400 Hz.

A figura 7.9 ilustra a forma de onda (contínua) de uma sequência binária resultante de uma hipotética modulação em quadratura em que se combinaram 2 valores de fase e 2 valores de amplitude para codificar 2 dígitos binários por elemento de sinal.

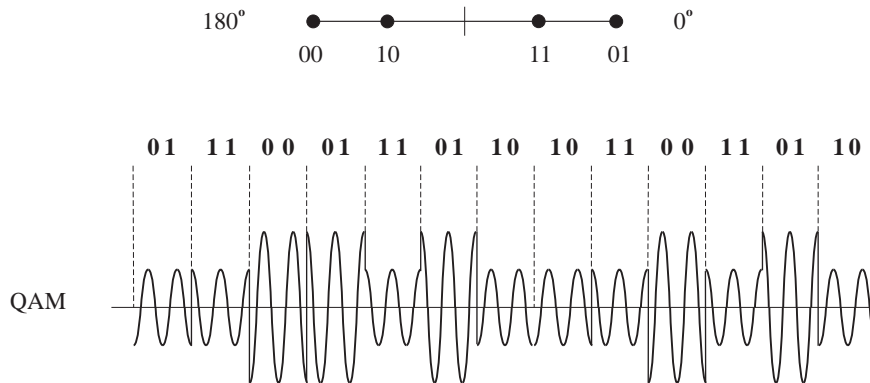


Figura 7.9: Forma de onda de uma hipotética modulação em quadratura

Existem outras técnicas de codificação mais sofisticadas permitindo transmitir mais dígitos binários por elemento de sinal (símbolo) mantendo uma probabilidade de erro controlada mas que recorrem a uma codificação prévia dos dígitos de informação antes de efectuarem propriamente a codificação de linha, ou seja antes da modulação propriamente dita. Trata-se na realidade de codificação para controlo de erros¹³ que utiliza dígitos redundantes para validação dos dados recebidos. Uma dessas técnicas é chamada de TCM (Trellis Coded Modulation) que utiliza codificação binária convolucional e que é utilizada em certos equipamentos de modulação e desmodulação — os modems ECM (Error Correcting Modem).

¹³este assunto será tratado em pormenor mais adiante

7.3 Problemas

- 7.1 – Considere a seguinte técnica de codificação de linha. Os dígitos binários, d_m ($m=1,2,3,\dots$), são apresentados ao codificador que os processa em dois passos. No primeiro, é produzido um novo conjunto de dígitos binários $b_m = (d_m + b_{m-1}) \bmod 2$ e no segundo, estes são codificados da seguinte forma $c_m = b_m - b_{m-1}$. No receptor, os dígitos originais são recuperados fazendo $d_m = c_m \bmod 2$.
- Verificar que os valores recebidos são iguais aos transmitidos.
 - De que tipo de codificação de linha se trata?
- 7.2 – Desenvolva um diagrama de estados do autômato finito gerador de um código pseudoternário.
- 7.3 – Para que o código B8ZS seja eficaz, a probabilidade de ocorrência de mais do que uma violação de código devido a um erro deve ser bastante pequena. Qual a probabilidade de ocorrência de mais do que uma violação de código num octeto quando a probabilidade de erro por bit é $P_e = 10^{-6}$? E quando $P_e = 10^{-3}$?
- 7.4 – Esquematize as formas de onda resultantes da codificação de linha em banda base da sequência 01000011101110100001, para cada um dos códigos representados nas figuras 7.1, 7.4 e 7.5.
- 7.5 – A forma de onda da figura representa um sinal binário codificado em Manchester. Determinar o início e o fim dos períodos dos bits e a sequência binária que foi codificada.



- 7.6 – Considere a figura 7.8. Oito das fases utilizam apenas um nível de amplitude e a constelação do sinal apenas codifica quatro bits. Quantos bits poderiam ser codificados se essas fases utilizassem também dois níveis de amplitude?

fim do capítulo 7

Capítulo 8

Teoria da Informação

A teoria dos sinais e dos sistemas de comunicação, constituindo embora uma ferramenta válida de análise, não aborda o processo fundamental das comunicações que é o da produção e transferência de *informação*. Reconhecendo a necessidade de uma teoria própria para este processo, Claude Shannon, matemático dos Laboratórios Bell que aí trabalhara em sistemas de comunicação durante a segunda grande guerra, desenvolveu uma "Teoria Matemática das Comunicações" que publicou pela primeira vez em 1948.

Segundo Shannon, a questão principal da engenharia de comunicações era a seguinte: Dada uma fonte de informação qualquer, como é que devem ser representadas as mensagens por ela emitidas, de modo a poderem ser transmitidas fiavelmente através de um canal de comunicação, dadas as inerentes limitações físicas deste?

Ao atacar esta questão Shannon concentrou-se na informação da mensagem em si, e não nos sinais utilizados para a transmitir. Esta abordagem deu origem ao que é hoje conhecido por *Teoria da Informação*, que se tem desenvolvido como disciplina híbrida da matemática e da engenharia.

A Teoria da Informação estuda os três problemas básicos seguintes:

1. A *medida* da informação produzida por uma fonte;
2. A *capacidade* do canal de comunicação para transmitir informação;
3. A *codificação*, quer da fonte quer do canal, como mecanismo para melhor utilização da capacidade do canal.

Nesta teoria o termo *canal* é utilizado em sentido lato, designando o sistema (ou a *via*) de comunicação completo entre a fonte e o destino da informação. O termo *codificação* é utilizado igualmente no seu sentido mais genérico de representação de mensagens quer por formas de onda contínuas quer por formas de onda discretas ou outros símbolos. Os três problemas atrás mencionados conjugam-se naquele que constitui o teorema fundamental da Teoria da Informação, a saber:

Dado um canal de comunicação e uma fonte cujo débito de informação não excede a capacidade do canal, existe um código tal que a informação pode ser transmitida através do canal com uma frequência de erros arbitrariamente pequena, apesar da presença do ruído.

O aspecto surpreendente deste teorema é a promessa que faz de transmissão de informação *sem erros* através de um canal ruidoso, o que se torna possível através de operações de *codificação*.

O processo de codificação envolve, em geral, duas operações distintas de codificação/descodificação, representadas em diagrama na figura 8.1. O *codificador* e o *descodificador do canal* executam a tarefa de *control de erros*, detectando-os e eventualmente corrigindo-os através de mecanismos que se estudarão no próximo capítulo e pelos quais os efeitos do ruído podem ser reduzidos ou eliminados. Assim, podemos desde já considerar que o conjunto dos blocos da figura 8.1 contidos no bloco tracejado, equivale a um *canal sem ruído*. A teoria da informação vai um passo mais além afirmando que a codificação óptima do canal dá origem a um canal equivalente sem ruído com uma capacidade de transmissão de informação bem definida. O *codificador* e o *descodificador* da fonte têm o papel de adaptar a fonte ao canal equivalente sem ruído desde que o débito de informação da fonte não ultrapasse a capacidade do canal. Começaremos por considerar o caso da informação representada sob a forma digital ou *discreta*. Trataremos a medida da informação, a codificação da fonte, a transmissão da informação e a capacidade do canal discreto. Estes conceitos serão depois extendidos ao caso mais realista da transmissão de informação em canais *contínuos* onde as mensagens tomam a forma de sinais variáveis no tempo, concluindo com a *lei de Hartley-Shannon* que define a capacidade do canal em função da largura de banda e da razão sinal-ruído, servindo de termo de comparação para a avaliação do desempenho de sistemas de comunicação.

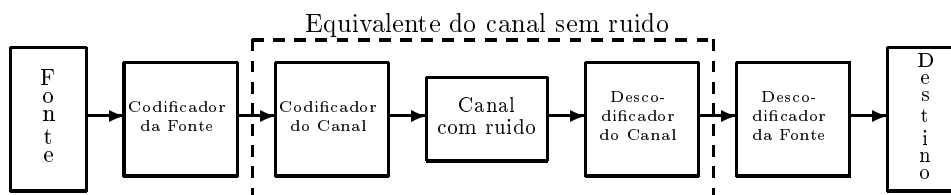


Figura 8.1: Sistema de comunicação com codificação da fonte e do canal

Convém recordar o que se disse no início relativamente ao conceito de *canal* em Teoria da Informação: é a *via* de comunicação simplex completa a partir da saída do codificador da fonte e até à entrada do respectivo decodificador no destino, isto é, o bloco a tracejado na figura 8.1.

8.1 Medida da Informação

O ponto de partida da teoria da informação é a medida da *informação*, termo técnico que não deve ser confundido com *dados*, nem com *conhecimento* ou *significado*, conceitos cuja definição e medida são ainda objecto de debate.

No contexto das comunicações, informação não é mais do que o produto, o bem ou o objecto imaterial útil produzido por uma fonte que tem de ser transferido para um utilizador num destino. Se a informação está previamente disponível no destino a transferência será zero.

Suponhamos, por exemplo, uma pessoa que planeia as suas férias de verão para o Algarve e ouve uma das seguintes previsões do tempo para essa altura:

No Algarve o sol vai nascer

No Algarve vai chover

No Algarve vai nevar

A primeira mensagem não contém nenhuma informação pois é certo que o sol nasce todos os dias em toda a parte. A previsão de chuva porém contém informação não disponível anteriormente. A terceira previsão fornece ainda mais informação dado que nevar no Algarve é um acontecimento raro e porventura inesperado.

É senso comum que quanto menos provável for uma determinada mensagem maior a quantidade de informação nela contida. Pode pois concluir-se que a medida da informação deve relacionar-se com o grau de *incerteza* do destinatário quanto à mensagem que vai receber. Posto de outra forma, a informação mede a *liberdade de escolha* exercida pela fonte ao seleccionar uma mensagem dentro do conjunto universo das possíveis mensagens.

8.1.1 Informação própria

Torna-se evidente, portanto, que a medida da informação deve ser uma função da *probabilidade* de ocorrência da mensagem.

Seja $f()$ essa função e x_i uma mensagem arbitrária tal que a probabilidade do acontecimento x_i ser seleccionado para transmissão é $P(x_i) = P_i$. A quantidade de informação, I_i , associada à ocorrência de x_i , designada de *informação própria* ou *auto-informação* de x_i , será então $I_i = f(P_i)$.

A função $f()$ deve possuir as seguintes propriedades:

- (i) $f(P_i) \geq 0$ para $0 \leq P_i \leq 1$
- (ii) $\lim_{P_i \rightarrow 1} f(P_i) = 0$
- (iii) $f(P_i) > f(P_j)$ para $P_i < P_j$

Estas propriedades fundamentam-se, ainda, no senso-comum. A primeira indica que a informação nunca é negativa, a segunda que ela é nula se o acontecimento for certo e a terceira indica que a informação aumenta com a incerteza.

Existem muitas funções que obedecem a estas três propriedades. Contudo, considere-se o caso em que a fonte produz duas mensagens sucessivas e independentes, x_i e x_j , com probabilidade conjunta $P(x_i x_j) = P_{ij} = P_i P_j$. Pretende-se que a informação total seja igual à soma da informação das mensagens individuais o que exige a seguinte propriedade adicional:

- (iv) $f(P_i P_j) = f(P_i) + f(P_j)$

A *única* função que satisfaz estas quatro propriedades é a função logarítmica negativa, $-\log_b(\cdot)$. A base b que se utilizar para o logaritmo define

a *unidade* de medida da informação. A convenção adoptada na teoria da informação é tomar $b = 2$ e designar a unidade correspondente por *bit*.

Definição 8.1 — Bit como unidade de medida de informação

O bit é a quantidade de informação necessária para escolher uma entre duas alternativas igualmente prováveis ou, a quantidade de informação contida numa mensagem emitida por uma fonte capaz de emitir apenas duas mensagens distintas e equiprováveis.

Portanto, e por definição, a quantidade de informação, ou informação própria, I_i numa mensagem x_i é dada por:

$$I_i \stackrel{\text{def}}{=} \log_2 \frac{1}{P_i} \quad \text{bits} \quad (8.1)$$

Se a fonte possuir um alfabeto de apenas duas mensagens, x_1 e x_2 , e se $P(x_1) = P(x_2) = 1/2$, tem-se $I_1 = I_2 = \log_2 2 = 1$ *bit*.

Deve ter-se cuidado em distinguir *bits* de informação dos *dígitos binários* utilizados para a representar ou codificar — especialmente porque um dígito binário pode transportar mais ou menos do que um bit de informação.

De modo a evitar a confusão, o dígito binário é muitas vezes designado por *binit* e, no contexto da Teoria da Informação, o termo *bit* designa a unidade de medida de informação e não o dígito binário.

Por simplicidade utiliza-se também o termo *símbolo* em alternativa ao termo *mensagem*.

É frequente ter de se converter logaritmos de uma base b qualquer, por exemplo a natural ou a decimal, para a base 2. É fácil verificar que

$$\log_2 \alpha = \frac{\log_b \alpha}{\log_b 2} = \frac{\ln \alpha}{\ln 2} = \frac{\log_{10} \alpha}{\log_{10} 2} \quad (8.2)$$

8.1.2 Entropia e Débito de informação

Consideremos agora uma fonte de informação que emite uma sequência de símbolos (mensagens) seleccionados de um alfabeto de m símbolos distintos. Seja $X = \{x_1, x_2, \dots, x_m\}$ esse conjunto alfabeto. Podemos tratar cada símbolo x_i como uma mensagem que ocorre com probabilidade P_i

e transporta a auto-informação I_i . Os valores das probabilidades devem obviamente satisfazer a igualdade $\sum_{i=1}^m P_i = 1$.

Suponhamos que os sucessivos símbolos são estatisticamente independentes e são produzidos pela fonte a um débito médio de r_s símbolos por segundo. Suponhamos ainda que a fonte é estacionária, o que significa que as probabilidades não variam com o tempo. Estas propriedades definem o que se designa por *fonte discreta sem memória*.

A quantidade de informação produzida pela fonte durante um intervalo de símbolo arbitrário é uma variável aleatória discreta que toma valores I_1, I_2, \dots, I_m . A informação média por símbolo é então dada pela média estatística,

$$\mathcal{H}(X) \stackrel{\text{def}}{=} \sum_{i=1}^m P_i I_i = \sum_{i=1}^m P_i \log_2 \frac{1}{P_i} \text{ bits/símbolo} \quad (8.3)$$

grandeza que é chamada *entropia* da fonte. A equação 8.3 é interpretada da seguinte maneira: quando uma fonte emite uma sequência de $n \gg 1$ símbolos, a informação total a produzida é aproximadamente igual a $n \mathcal{H}(X)$ bits. Dado que a fonte produz r_s símbolos por segundo, a duração desta sequência é de n/r_s seg. A informação deve pois ser transferida a um débito médio $n \mathcal{H}(X)/(n/r_s) = r_s \mathcal{H}(X)$ bits por segundo. O *débito médio de informação* de uma fonte é então definido por

$$\mathcal{R} \stackrel{\text{def}}{=} r_s \mathcal{H}(X) \text{ bits/seg} \quad (8.4)$$

uma grandeza crítica em sistemas de transmissão. O teorema fundamental da teoria da informação (devido a Shannon) diz que a informação produzida por qualquer fonte discreta sem memória pode ser codificada em dígitos binários e transmitida através de um canal sem ruído a um ritmo binário r_b , com $r_b \geq \mathcal{R}$. O valor de $\mathcal{H}(X)$ para uma dada fonte depende das probabilidades dos símbolos, P_i , e da cardinalidade, m , do alfabeto estando limitado por

$$0 \leq \mathcal{H}(X) \leq \log_2 m \quad (8.5)$$

O limite inferior corresponde à inexistência de incerteza o que acontece quando um dos símbolos tem probabilidade $P_j = 1$ e todos os restantes $P_i = 0$ ($i \neq j$) e portanto a fonte emite sempre o mesmo símbolo. O limite superior corresponde à máxima incerteza que ocorre quando $P_i = 1/m \forall i$, isto é, quando todos os símbolos são igualmente prováveis.

Exêmplo 8.1 – Fonte binária

Para ilustrar a variação de $\mathcal{H}(X)$ entre aqueles extremos consideremos o caso especial mas importante da fonte binária ($m = 2$) com

$$P_1 = p \quad P_2 = 1 - p$$

Substituindo estes valores das probabilidades na equação 8.3 obtém-se a entropia da fonte binária

$$\mathcal{H}(X) = \Omega(p) \stackrel{\text{def}}{=} p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{1 - p} \quad (8.6)$$

Designamos esta entropia pela função $\Omega()$ pois vamos utilizá-la mais adiante. O gráfico de $\Omega(p)$ na figura 8.2 mostra um máximo centrado em $p = 1 - p = 1/2$ onde se tem $\mathcal{H}(X) = \log_2 2 = 1$ bit/símb. $\mathcal{H}(X)$ decresce monótonamente para zero quando $p \rightarrow 1$ ou $p \rightarrow 0$. A prova

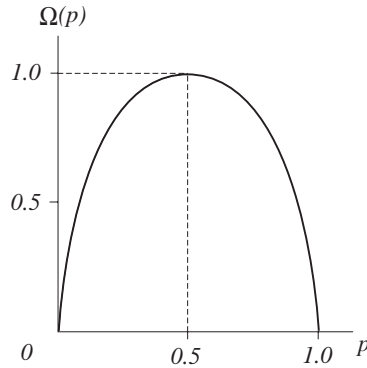


Figura 8.2: Função de entropia binária

do limite inferior na relação 8.5 obtém-se facilmente bastando notar que $\alpha \log_2(1/\alpha) \rightarrow 0$ quando $\alpha \rightarrow 0$. A prova do limite superior $\mathcal{H}(X) \leq \log_2 m$ exige mais atenção (resolver o problema 8.6 para demonstrar este resultado).

Exêmplo 8.2 – Fonte quaternária

Suponhamos uma fonte que emite $r_s = 2000$ simb/seg seleccionados de um alfabeto de quatro símbolos ($m = 4$) com probabilidade dos símbolos e auto-informação dados na tabela 8.1. A equação 8.3 fornece o valor da

Tabela 8.1:

x_i	P_i	I_i
A	$1/2$	1
B	$1/4$	2
C	$1/8$	3
D	$1/8$	3

entropia desta fonte

$$\mathcal{H}(X) = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 = 1.75 \text{ bits/símb}$$

que é ligeiramente inferior ao valor máximo $\log_2 m = 2$. O débito de informação é

$$\mathcal{R} = 2000 \times 1.75 = 3500 \text{ bits/seg}$$

e através de codificação apropriada, deve ser possível transmitir a informação da fonte a um ritmo binário $r_b \geq 3500 \text{ dig bin/seg}$ (resolver o problema 8.8).

8.2 Codificação da fonte discreta sem memória

Quando uma fonte discreta sem memória produz m símbolos¹ equiprováveis de tal forma que $\mathcal{R} = r_s \log_2 m$, todos os símbolos transportam a mesma quantidade de informação e pode obter-se uma transmissão com bom rendimento utilizando codificação de linha multi-nível com m níveis distintos (codificação M-ária) e com um ritmo na linha² r_c igual ao débito de símbolos r_s . Mas quando os símbolos possuem probabilidades de ocorrência diferentes e se tem

$$\mathcal{R} = r_s \mathcal{H}(X) < r_s \log_2 m \quad (8.7)$$

torna-se necessário *codificar a fonte* em função da quantidade de informação variável por símbolo.

¹tenha-se em atenção que, neste contexto, símbolo é sinónimo de mensagem

² r_c designa o ritmo de símbolos do código no canal de transmissão. Quando esses símbolos são dígitos binários o respectivo ritmo costuma designar-se por r_b .

Nesta secção aborda-se um método para efectuar esta codificação considerando-se apenas a codificação binária. Também se podem obter resultados equivalentes para o caso da codificação não-binária desde que se adopte uma base apropriada para os logaritmos.

O codificador binário da figura 8.3 converte os símbolos da fonte em *palavras de código* compostas por dígitos binários produzidos a um ritmo de r_b *dígitos binários/seg.* A saída do codificador pode ser encarada como uma fonte binária com entropia

$$\mathcal{H}(X) = \Omega(p) = p \log_2 \frac{1}{p} + (1-p) \log_2 \frac{1}{1-p} \quad \text{bits/dig-bin} \quad (8.8)$$

e com débito de informação

$$\mathcal{R} = r_b \Omega(p) \leq r_b \log_2 2 = r_b \quad \text{bits/s} \quad (8.9)$$

em que p é a probabilidade de um dos dois símbolos binários. É óbvio que a operação de codificação não pode gerar informação adicional àquela que é produzida pela fonte, nem mesmo perder informação desde que o código seja unívocamente decifrável. Assim, igualando os débitos de

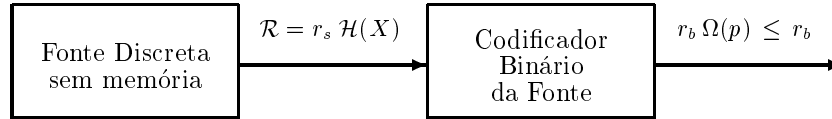


Figura 8.3: Fonte, Codificador da Fonte e débitos de informação

informação à entrada e à saída do codificador tem-se $r_s \mathcal{H}(X) = r_b \Omega(p) \leq r_b$, ou seja, $r_b/r_s \geq \mathcal{H}(X)$, em que a quantidade $\overline{N} = \frac{r_b}{r_s}$ é um parâmetro importante designado por *comprimento médio do código*. Fisicamente, \overline{N} corresponde ao número médio de dígitos binários por símbolo da fonte. Se for N_i o comprimento da palavra de código correspondente ao símbolo i , \overline{N} é a média estatística

$$\overline{N} = \sum_{i=1}^m P_i N_i \quad (8.10)$$

Podemos agora enunciar um dos teoremas fundamentais da Teoria da Informação de que faremos uso, embora não procedamos à sua demonstração:

Teorema 8.1 – Teorema de Shannon da Codificação da Fonte

Toda a fonte de informação caracterizada por um valor da entropia $\mathcal{H}(X)$ bits/símbolo, pode ser codificada em binário de tal forma que o comprimento médio do código, \overline{N} , é limitado por

$$\mathcal{H}(X) \leq \overline{N} \leq \mathcal{H}(X) + \epsilon \quad (8.11)$$

em que ϵ é uma quantidade positiva tão pequena quanto se desejar.

O código óptimo é aquele em que $\epsilon = 0$, ou seja, $\overline{N} = \mathcal{H}(X)$. Na prática, nem sempre se consegue obter um código nestas condições sendo satisfatório um código sub-óptimo com $\overline{N} > \mathcal{H}(X)$ desde que este possua um bom rendimento o que é dado por

$$\rho = \frac{\mathcal{H}(X)}{\overline{N}} = \frac{\mathcal{R}}{r_b} \leq 1 \quad (8.12)$$

O teorema da codificação da fonte pressupõe um código unívocamente decifrável de modo a garantir que não há perda de informação. Esta exigência impõe uma condição adicional, embora indirecta, aos valores que N_i pode tomar. Uma condição necessária e suficiente para que um código binário seja unívocamente decifrável é que os comprimentos das palavras de código N_i sejam tais que obedecem à *desigualdade de Kraft*

$$K = \sum_{i=1}^m 2^{-N_i} \leq 1 \quad (8.13)$$

O processo de codificação mais simples gera *códigos de comprimento fixo* com todas as palavras de código possuindo o mesmo comprimento $N_i = \overline{N}$. É a codificação binária pura. A desigualdade de Kraft dá $K = m2^{-\overline{N}} \leq 1$ e portanto, para ser decifrável, o código terá de ter comprimento médio $\overline{N} \geq \log_2 m$ e o seu rendimento será dado por $\frac{\mathcal{H}(X)}{\overline{N}} \leq \frac{\mathcal{H}(X)}{\log_2 m}$. Quando $\mathcal{H}(X) < \log_2 m$ só se poderão obter melhores rendimentos quando se utilizarem *códigos de comprimento variável*.

Exemplo 8.3 – Codificação da fonte quaternária

Na fonte do exemplo 8.2 os quatro símbolos A, B, C e D correspondem às formas de onda obtidas por amostragem e quantização a quatro níveis de um sinal analógico (PAM quantizado). Determinar um código binário para a fonte e comparar o ritmo binário obtido com o débito de informação da fonte.

Os resultados obtidos no exemplo 8.2 foram $\mathcal{H}(X) = 1.75$ bits/símbolo e $\mathcal{R} = 3500$ bits/s. Consideremos, na tabela 8.2, vários potenciais códigos binários para esta fonte: O código I tem comprimento fixo com $\overline{N} =$

Tabela 8.2: Códigos binários para a fonte quaternária

x_i	P_i	Código I	Código II	Código III	Código IV
A	1/2	00	0	0	0
B	1/4	01	1	01	10
C	1/8	10	10	011	110
D	1/8	11	11	0111	111
\overline{N}		2.0	1.25	1.875	1.75
K		1.0	1.5	0.9375	1.0

$\log_2 m = 2$ dig bin/simb o que, comparado com $\mathcal{H}(X) = 1.75$ bits/simb, dá um rendimento $\rho = \mathcal{H}(X)/\overline{N} = \frac{1.75}{2} \approx 88\%$ que não é muito baixo.

Aplicando a média estatística ao código II obtem-se

$$\overline{N} = \frac{1}{2} + \frac{1}{4} + 2 \times \frac{1}{8} + 2 \times \frac{1}{8} = 1.25 \leq \mathcal{H}(X)$$

e o número de Kraft dá

$$K = 2^{-1} + 2^{-1} + 2^{-2} + 2^{-2} = 1.5 > 1$$

mas o resultado $\overline{N} < \mathcal{H}(X)$ não tem significado porque $K > 1$ o que quer dizer que o código II não é unívocamente decifrável. Por exemplo, a sequência de código 10011 poderia ser decodificada como BAABB, CABB, CAD, etc. Este código destrói realmente a informação da fonte e portanto não pode ser adoptado.

O código III, conhecido por código de vírgula, tem $K < 1$ assegurando a sua decifragem. O início de cada palavra de código é marcada pelo dígito 0, chamado a vírgula. As vírgulas acabam por ser dígitos redundantes e conduzem a $\overline{N} = 1.875 > \mathcal{H}(X)$ o que no entanto já é melhor do que o código binário puro (código I).

O código IV é um código em árvore que possui a propriedade de nenhuma palavra de código aparecer como prefixo de nenhuma outra. Assim, por exemplo, a sequência 110010111 representa sem ambiguidade a sequência

de mensagens CABD. Este código é ótimo relativamente à fonte em questão dado que $\overline{N} = 1.75 = \mathcal{H}(X)$ com $K = 1$.

Construção do código

Existem vários algoritmos para calcular o código para uma dada fonte. Mencionaremos um dos mais conhecidos e simples de construir e que conduz a códigos decifráveis com bom rendimento. São os códigos de Huffman também designados de códigos de Shannon-Fano.

Códigos de Huffman

São códigos de comprimento variável cujo comprimento da palavra aumenta à medida que a probabilidade do símbolo diminui. O algoritmo conduz a um código em árvore e é o seguinte:

- (1) Ordenar os símbolos por ordem decrescente de probabilidade;
- (2) Dividir o conjunto assim ordenado em dois subconjuntos tais que a soma das probabilidades em cada um deles seja o mais aproximadamente possível igual a metade da soma das probabilidades no conjunto anterior. Manter a ordenação.
- (3) O dígito seguinte do código binário dos símbolos do primeiro dos sub-conjuntos é o **0** e o dos do outro é o **1**;
- (4) Se os sub-conjuntos contêm um só elemento, a codificação terminou para esses sub-conjuntos;
- (5) Repetir para cada um dos restantes sub-conjuntos (passo 2.)

Exemplo 8.4 – Codificação de Huffman

Aplicar o algoritmo de Huffman para codificar a fonte com $m = 8$ cuja estatística é a indicada na tabela 8.3.

Os símbolos são ordenados por ordem decrescente de probabilidade e são efectuados os passos de codificação indicados na tabela 8.4.

O código resultante possui comprimento médio $\overline{N} = 2.18$ dig bin e um rendimento $\rho = \frac{2.15}{2.18} \approx 99\%$ (resolver o problema 8.11).

Tabela 8.3: Estatística da fonte

x_i	A	B	C	D	E	F	G	H
P_i	0.50	0.15	0.15	0.08	0.08	0.02	0.01	0.01

Tabela 8.4: Codificação da fonte

x_i	P_i	Passos de codificação						Código
		1	2	3	4	5	6	
A	0.50	0						0
B	0.15	1	0	0				100
C	0.15	1	0	1				101
D	0.08	1	1	0				110
E	0.08	1	1	1	0			1110
F	0.02	1	1	1	1	0		11110
G	0.01	1	1	1	1	1	0	111110
H	0.01	1	1	1	1	1	1	111111
$\mathcal{H}(X) = 2.15$								$\bar{N} = 2.18$

8.3 Fontes com memória

Até aqui considerámos fontes discretas sem memória cujos símbolos sucessivos são estatisticamente independentes. No entanto a maior parte das fontes de informação possui *memória*, na medida em que as probabilidades dos símbolos dependem do símbolo ou dos símbolos anteriores produzidos pela fonte. A linguagem escrita que se rege pelas regras da gramática constitui um exemplo de uma fonte com memória. Verifica-se que a letra U , por exemplo, que poderá ocorrer com uma probabilidade $P(U) \approx 0.2$ numa base de frequência relativa, ocorre com probabilidade igual a 1 sempre que a letra anterior fôr Q . A probabilidade condicional desta situação será indicada por $P(U|Q) \approx 1$, isto é, praticamente não existe na forma escrita da língua portuguesa nenhum caso em que a letra Q não venha seguida da letra U .

Esta influência pode estender-se a vários símbolos anteriores. É quase

certo, por exemplo, que a seguir à frase *PODE MOSTRAR-SE* aparece a palavra *QUE*, ou seja, $P(\text{QUE}|\text{PODE MOSTRAR-SE}) \approx 1$.

O efeito da *memória* é o de reduzir a incerteza dando origem a um valor para a entropia mais baixo do que aquele que seria obtido se se utilizasse a estatística absoluta da fonte em lugar da condicional.

Suponhamos que uma fonte possui *memória de primeira ordem*, isto é, a fonte só se "lembra" do símbolo precedente. Para formular uma expressão da entropia, consideremos a probabilidade condicional $P(x_i|x_j)$ do símbolo x_i ser escolhido depois do símbolo x_j . Substituindo $P_i = P(x_i)$ por $P(x_i|x_j)$ na equação 8.3 obtem-se a *entropia condicional* relativamente ao símbolo x_j

$$\mathcal{H}(X|x_j) \stackrel{\text{def}}{=} \sum_{i=1}^m P(x_i|x_j) \log_2 \frac{1}{P(x_i|x_j)} \quad (8.14)$$

que representa a informação média por símbolo dado que o símbolo anterior foi um determinado x_j . Fazendo a média estatística para todos os possíveis valores do símbolo imediatamente anterior ter-se-á

$$\mathcal{H}_c(X) = \sum_{j=1}^m P(x_j) \mathcal{H}(X|x_j) \quad (8.15)$$

expressão que fornece a *entropia condicional da fonte com memória de primeira ordem* e que pode ser escrita sob a forma desenvolvida

$$\mathcal{H}_c(X) = \sum_{i=1}^m \sum_{j=1}^m P(x_j) P(x_i|x_j) \log_2 \frac{1}{P(x_i|x_j)} \quad (8.16)$$

$$\mathcal{H}_c(X) = \sum_{i=1}^m \sum_{j=1}^m P(x_i x_j) \log_2 \frac{1}{P(x_i|x_j)} \quad (8.17)$$

Pode-se escrever uma expressão equivalente para o caso geral de uma fonte com memória de ordem n . A notação torna-se bastante complexa pois x_j tem de ser substituído pelo *estado* da fonte em termos dos n símbolos anteriormente emitidos e haverá m^n possíveis estados a considerar. A fonte pode então ser modelada por um autómato de estados finitos.

Uma fonte com memória é dita *redundante* quando as probabilidades condicionais reduzem significativamente $\mathcal{H}_c(X)$ em relação ao limite superior $\log_2 m$. A redundância num texto em língua Portuguesa pode ser

tão elevada como 50%, o que significa que metade das letras ou palavras num texto longo não são essenciais à sua compreensão e portanto à entrega da informação completa ao destinatário. Vários factores contribuem para que tal seja possível como seja a dedução por contexto, isto é, deve ser capaz de ler o mesmo fazendo algumas letras.

Se a incerteza é reduzida por efeito de memória então a *previsibilidade* aumenta. A codificação de fontes com memória pode portanto ser baseada em métodos de previsão. O estudo destes métodos está fora do âmbito deste curso pelo que sómente fazemos menção à sua existência.

Concluimos esta secção notando que a codificação da fonte é um processo que *retira a redundância* produzida pela fonte, processo que também se designa por *compressão da fonte*. A redundância, especialmente na linguagem escrita, constitui um mecanismo natural de correcção de erros. Uma vez retirada essa redundância, resultando uma frase *comprimida*, como a que se exemplificou atrás, a perda ou alteração de uma letra poderá torná-la parcial ou completamente incompreensível.

A compressão obtida numa determinada codificação da fonte mede-se normalmente pela *redução* percentual do número médio de dígitos binários por símbolo relativamente à codificação binária pura, isto é

$$\text{compressão (\%)} = \frac{N_{bin} - \bar{N}}{N_{bin}} \times 100 \%$$

8.4 Transmissão de informação: o canal

Iremos estudar de seguida a transmissão de informação em canais de comunicação sob o ponto de vista da teoria da informação. Começaremos por considerar o caso em que tanto a fonte como o canal de transmissão são discretos, isto é, as mensagens são representadas e transmitidas sob a forma de símbolos de um determinado alfabeto finito. Passaremos depois ao caso mais realista da fonte e do canal contínuos onde as mensagens são produzidas e transmitidas sob a forma de funções contínuas do tempo.

8.4.1 Canais discretos

Informação mútua

Considere-se o sistema de transmissão de informação representado na figura 8.4. A fonte discreta selecciona símbolos do seu alfabeto X para transmissão através do canal. Por seu lado, o ruído e outras formas de perturbação da transmissão podem transformar os símbolos transmitidos noutros do alfabeto Y do destino. Pretende-se medir a informação transmitida. Os vários tipos de probabilidades de símbolos que estarão em

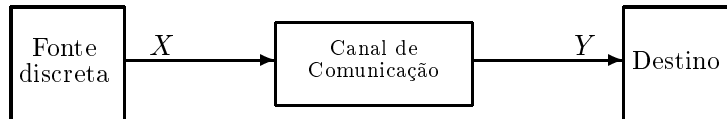


Figura 8.4: Sistema discreto de transmissão de informação

causa são os seguintes:

- $P(x_i)$ probabilidade da fonte seleccionar o símbolo x_i para transmissão;
- $P(y_j)$ probabilidade de ser recebido no destino o símbolo y_j ;
- $P(x_i y_j)$ probabilidade conjunta de ser transmitido o símbolo x_i e recebido o símbolo y_j ;
- $P(x_i|y_j)$ probabilidade condicional de ter sido transmitido o símbolo x_i dado ter-se recebido o símbolo y_j ;
- $P(y_j|x_i)$ probabilidade condicional de se receber o símbolo y_j dado ter-se transmitido o símbolo x_i .

Consideramos o canal *invariante no tempo* e *sem memória*, portanto as probabilidades condicionais são independentes do tempo e dos símbolos precedentes. Em especial, os $P(y_j|x_i)$ representam as *probabilidades de transição directas* do canal, indicadas na figura 8.5 a qual ilustra o modelo de um canal m -ário com ruído.

Se o sistema é suposto entregar $Y = y_k$ quando $X = x_k$, então as *probabilidades de erro* dos símbolos são dadas por $P(y_j|x_i)$ para $j \neq i$. Por outro lado, ao receber-se um determinado símbolo y_j , a incerteza quanto àquele que lhe deu origem ter sido o símbolo x_i , depende da probabilidade condicional $P(x_i|y_j)$, a que corresponde uma *perda* de informação que designaremos por $I(x_i|y_j)$ medida em *bits*.

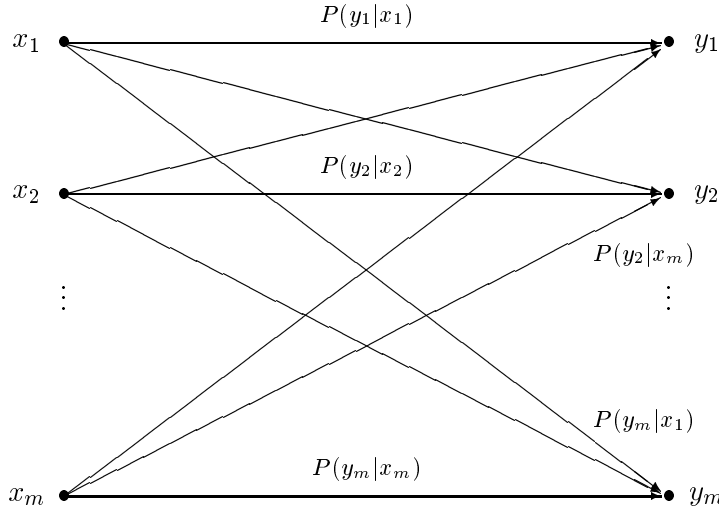


Figura 8.5: Probabilidades de transição para um canal discreto com ruído

A quantidade de informação transferida (recebida no destino) quando é transmitido o símbolo x_i e recebido o símbolo y_j é chamada de *informação mútua* e dada então por

$$I(x_i, y_j) = I(x_i) - I(x_i|y_j) \quad (8.18)$$

em que $I(x_i) = I_i$ é a auto-informação do símbolo x_i produzido na fonte. Em função das probabilidades tem-se

$$I(x_i, y_j) = \log_2 \frac{1}{P(x_i)} - \log_2 \frac{1}{P(x_i|y_j)} \quad (8.19)$$

Determinemos agora os valores médios da informação que, por consistência com a notação que se adoptou, vamos designar por *entropia*, \mathcal{H} . Estarão em causa várias entropias.

Assim, o valor médio da informação recebida num símbolo y_j em relação a todos os possíveis símbolos produzidos pela fonte será dado pela média estatística

$$\begin{aligned} I(X, y_j) &= \sum_{i=1}^m P(x_i) \log_2 \frac{1}{P(x_i)} - \sum_{i=1}^m P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} \\ &= \mathcal{H}(X) - \mathcal{H}(X|y_j) \end{aligned} \quad (8.20)$$

A informação média recebida por símbolo, considerados todos os possíveis símbolos do alfabeto do destino é dada novamente pela média estatística

$$\begin{aligned}
 I(X, Y) &= \mathcal{H}(X) - \sum_{j=1}^m P(y_j) \mathcal{H}(X|y_j) \\
 &= \mathcal{H}(X) - \sum_{j=1}^m \sum_{i=1}^m P(y_j) P(x_i|y_j) \log_2 \frac{1}{P(x_i|y_j)} \\
 &= \mathcal{H}(X) - \underbrace{\sum_{j=1}^m \sum_{i=1}^m P(x_i y_j) \log_2 \frac{1}{P(x_i|y_j)}}_{\mathcal{H}(X|Y)} \quad (8.21)
 \end{aligned}$$

A informação média recebida não é mais do que a informação mútua média, ou seja a *entropia mútua* $\mathcal{H}(X, Y) \equiv I(X, Y)$

$$\mathcal{H}(X, Y) = \mathcal{H}(X) - \mathcal{H}(X|Y) \quad (8.22)$$

Se agora se tomar em consideração que $P(x_i) P(y_j|x_i) = P(y_j) P(x_i|y_j)$ e que portanto $I(X, Y) = I(Y, X)$, pode trocar-se X e Y na equação 8.22 obtendo-se

$$\mathcal{H}(X, Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X) \quad (8.23)$$

tendo-se introduzido três novas quantidades com significado físico bem definido

$$\mathcal{H}(X|Y) = \sum_{j=1}^m \sum_{i=1}^m P(x_i y_j) \log_2 \frac{1}{P(x_i|y_j)} \quad (8.24)$$

$$\mathcal{H}(Y|X) = \sum_{i=1}^m \sum_{j=1}^m P(x_i y_j) \log_2 \frac{1}{P(y_j|x_i)} \quad (8.25)$$

$$\mathcal{H}(Y) = \sum_{j=1}^m P(y_j) \log_2 \frac{1}{P(y_j)} \quad (8.26)$$

em que

- $\mathcal{H}(X|Y)$ é a equivocação;
- $\mathcal{H}(Y|X)$ é a entropia do ruído;
- $\mathcal{H}(Y)$ é a entropia do destino.

A equação 8.22 diz que a transferência de informação média por símbolo é igual à entropia da fonte menos a equivocação. A equivocação representa a informação perdida no canal com ruído. A equação 8.23 diz que a

informação transferida é também igual à *entropia no destino* $\mathcal{H}(Y)$ menos a *entropia do ruído* $\mathcal{H}(Y|X)$ introduzida pelo canal. A interpretação de $\mathcal{H}(Y|X)$ como entropia do ruído segue-se à observação anterior segundo a qual o conjunto das probabilidades de transição directas $P(y_j|x_i)$ inclui as probabilidades de erro dos símbolos.

A entropia mútua pode então obter-se de duas maneiras: ou pela equação 8.22 ou pela equação 8.23.

Exêmplo 8.5 – Canal binário simétrico

Consideremos o canal binário simétrico (BSC^8) cujo modelo está representado na figura 8.6. A fonte produz dois símbolos com probabilidades

$$P(x_1) = p \quad P(x_2) = 1 - p$$

e no destino há dois possíveis símbolos com probabilidades de transição directa

$$\begin{aligned} P(y_1|x_2) &= P(y_2|x_1) = \alpha \\ P(y_1|x_1) &= P(y_2|x_2) = 1 - \alpha \end{aligned}$$

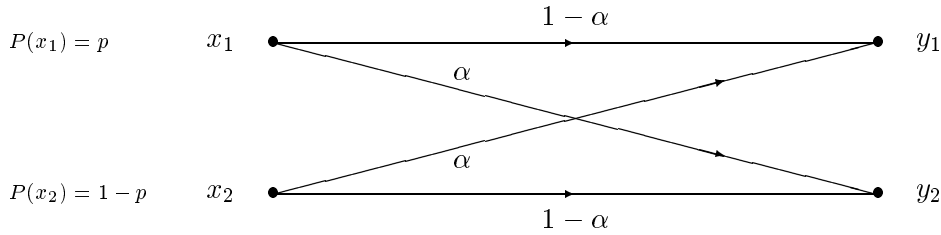


Figura 8.6: Modelo do canal binário simétrico

Este modelo representa qualquer sistema de transmissão binário no qual os erros são estatisticamente independentes e as probabilidades de erro são as mesmas para ambos os símbolos e portanto a probabilidade média de erro por símbolo é dada por

$$P_e = P(x_1) P(y_2|x_1) + P(x_2) P(y_1|x_2) = p\alpha + (1 - p)\alpha = \alpha$$

Dado conhecerem-se as probabilidades de transição directas, pode utilizar-se a equação $\mathcal{H}(X, Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X)$ para calcular a informação

⁸BSC – sigla para a denominação inglesa de *Binary Symmetric Channel*

média mútua, ou entropia mútua em termos de p e α . Esta será a informação média transferida para o destino.

A entropia no destino $\mathcal{H}(Y)$ pode ser obtida tratando a saída do canal como uma fonte binária com probabilidades de símbolos $P(y_1)$ e $P(y_2) = 1 - P(y_1)$, isto é

$$\mathcal{H}(Y) = \Omega[P(y_1)]$$

em que $\Omega()$ é a função de entropia binária definida pela equação 8.8 e onde

$$P(y_1) = P(x_1)P(y_1|x_1) + P(x_2)P(y_1|x_2) = \alpha + p - 2\alpha p$$

O cálculo da entropia do ruído $\mathcal{H}(Y|X)$ faz-se a partir da equação 8.25 que dá

$$\mathcal{H}(Y|X) = \Omega(\alpha)$$

resultado que é independente de p devido à simetria do canal.

Pode-se agora escrever uma expressão para a entropia mútua em função das expressões obtidas

$$\mathcal{H}(X, Y) = \Omega(\alpha + p - 2\alpha p) - \Omega(\alpha) \quad (8.27)$$

o que mostra que a transferência de informação através de um canal binário simétrico depende tanto da probabilidade de erro α como da probabilidade da fonte p . Se o ruído for pequeno, então $\alpha \ll 1$ e $\mathcal{H}(X, Y) \approx \Omega(p) = \mathcal{H}(X)$. Se o ruído for muito grande, então $\alpha = \frac{1}{2}$ e $\mathcal{H}(X, Y) = 0$ (resolver os problemas 8.14 e 8.15).

Capacidade do canal discreto

Cada canal possui normalmente alfabetos de origem e destino fixos e probabilidades de transição directas também fixas dado o sistema de transmissão completo, quando já em exploração, ser constituído por equipamento bem determinado resultante do respectivo projecto e funcionar com parâmetros fixos tais como o tipo de modulação, a potência do sinal, o meio físico de transmissão utilizado, etc. Deste modo, as únicas quantidades variáveis em $\mathcal{H}(X, Y)$ são as probabilidades da fonte $P(x_i)$. Consequentemente, para que a transferência de informação seja máxima é necessária uma determinada estatística para a fonte que poderá ser

obtida, por exemplo, através da codificação da fonte. Seja C_c o valor máximo de $\mathcal{H}(X, Y)$ assim obtido, isto é,

$$C_c \stackrel{\text{def}}{=} \max_{P(x_i)} \mathcal{H}(X, Y) \quad \text{bits/símbolo} \quad (8.28)$$

C_c representa a máxima quantidade de informação transferida por símbolo do canal e que é designada por *capacidade do canal*. É mais frequente, porém, medir a capacidade de um canal em termos do débito de informação. Assim, se for r_c o ritmo máximo de símbolos permitido no canal, então a sua capacidade é dada por

$$C = r_c C_c \quad \text{bits/seg} \quad (8.29)$$

que representa o ritmo máximo de transferência de informação. A importância do conceito de capacidade do canal torna-se mais evidente no contexto do *teorema fundamental da codificação do canal*, devido a Shannon, que se referiu no início deste capítulo e cujo enunciado é o seguinte:

Teorema 8.2 – Teorema de Shannon da Codificação do Canal

Se um canal possui capacidade C e uma fonte produz informação a um débito $R \leq C$, então existe um método de codificação para a transmissão pelo canal tal que a saída da fonte pode ser transmitida por esse canal com uma frequência de erros arbitrariamente pequena. Pelo contrário, se $R > C$, então não é possível transmitir a informação sem erros.

De notar a extrema importância deste teorema dado que ele promete a possibilidade de transmissão isenta de erros bastando para isso que $R \leq C$.

A sua demonstração geral está fora do âmbito deste curso pelo que terminaremos esta secção discutindo o princípio geral da codificação do canal. Este assunto será abordado em mais detalhe no capítulo que se segue.

Codificação para o canal binário simétrico

Consideremos o sistema de codificação representado na figura 8.7. O Canal Binário Simétrico (BSC) possui uma capacidade $C_c = 1 - \Omega(\alpha)$ com $\alpha < 1/2$ (ver Problema 8.16) e um ritmo de símbolos r_c simb do canal/s. A fonte, com entropia $\mathcal{H}(X)$, produz informação a um débito de \mathcal{R} bits/s.

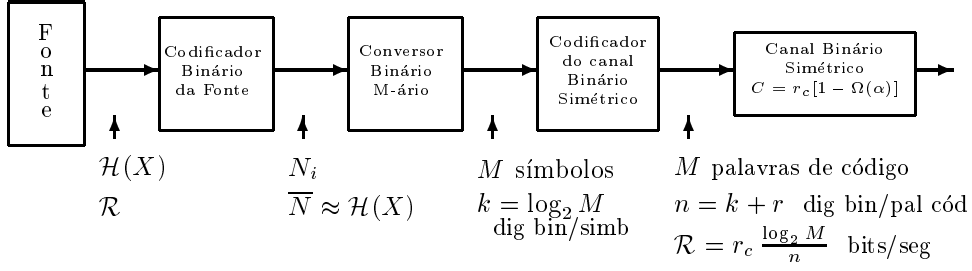


Figura 8.7: Codificação para um canal binário simétrico (BSC)

De modo a maximizar a informação mútua, um codificador binário da fonte óptimo opera sobre os símbolos da fonte e produz dígitos binários equiprováveis. A codificação do canal para control de erros é efectuada em dois estágios. No primeiro, os dígitos binários do código da fonte (os dígitos de informação) são agrupados em blocos de k dígitos produzindo assim $M = 2^k$ símbolos distintos (símbolos M -ários representados por $\log_2 M = k$ dígitos binários). No segundo, o codificador do canal binário simétrico, representa cada símbolo M -ário numa *palavra de código do canal*, com n dígitos binários por adição de r dígitos binários redundantes⁴ ($n = k + r$).

A informação média por dígito binário do canal é $\frac{\log_2 M}{n}$ bits/dig bin, sendo estes dígitos gerados ao ritmo de símbolos do canal, r_c . Podemos, pois, expressar o débito de informação da fonte por

$$\mathcal{R} = r_c \frac{\log_2 M}{n} \quad \text{bits/s} \quad (8.30)$$

em que M e n são os parâmetros do codificador do canal. O teorema de Shannon exige que $R \leq C = r_c C_c$ o que equivale a $\frac{\log_2 M}{n} \leq C_c$. Os parâmetros M e n devem portanto estar relacionados por

$$M = 2^{n(C_c - \epsilon)} \quad (8.31)$$

com $0 \leq \epsilon < C_c$. Pode mostrar-se que ϵ pode ser arbitrariamente pequeno e portanto $R \rightarrow C$ e que uma codificação do canal apropriada permite recuperar os símbolos M -ários no destino com uma probabilidade de erro arbitrariamente pequena, desde que o comprimento n da palavra de código seja *muito grande*. De facto, ter-se-á eventualmente que fazer

⁴redundantes no sentido de não transportarem informação da fonte

$n \rightarrow \infty$ para garantir uma transmissão isenta de erro. A codificação ideal do canal binário simétrico resulta assim num *tempo de atraso infinito*, pelo que qualquer sistema prático de codificação, devendo ter necessariamente um tempo de atraso limitado e um comprimento de palavra finito, ficará sempre aquém do desempenho ideal.

Por outro lado, Shannon não propôs qualquer algoritmo explícito para a determinação das palavras de código.

8.4.2 Canais contínuos

Começaremos esta secção pela definição da medida de informação produzida por uma fonte que emite mensagens sob a forma de sinais contínuos. Seguidamente, baseando-nos em hipóteses razoáveis àcerca da transmissão de sinais contínuos, chegaremos a uma expressão para a capacidade do canal contínuo em função da largura de banda e da razão de potências sinal-ruído, resultado que é conhecido por *lei de Hartley-Shannon*. Esta lei, que define o sistema de comunicação ideal, serve de padrão para a avaliação comparativa do desempenho de diferentes sistemas de comunicação e de termo de referência quando se projectam esses sistemas.

Informação contínua

Uma fonte de informação contínua produz um *signal* variável no tempo $x(t)$. Consideremos os possíveis sinais como um conjunto de formas de onda geradas por algum processo aleatório que supomos ser ergódico⁵. Suponhamos ainda que o processo tem largura de banda limitada, B_x , o que significa, pelo teorema da amostragem, que $x(t)$ fica completamente caracterizado pelos *valores das amostras* obtidas periodicamente a uma frequência de amostragem não inferior a $2B_x$ ($f_a \geq 2B_x$). Assim, em qualquer instante de amostragem, o conjunto dos possíveis valores de amplitude das amostras constitui uma *variável aleatória contínua* X_x descrita por uma função de densidade de probabilidade $p_X(x)$.

A quantidade média de informação por amostra de $x(t)$ é medida pela função de entropia

$$\mathcal{H}(X_x) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} p_X(x) \log_2 \frac{1}{p_X(x)} dx \quad (8.32)$$

⁵o que permite tomar as médias temporais pelas médias de conjunto

equação que é semelhante à que define a entropia de uma fonte discreta (equação 8.3), onde se substituíram o sinal de somatório pelo de integral e as probabilidades P_i pela *f.d.p.* $p_X(x)$. Contudo, a equação 8.32 representa uma medida *relativa* da informação e não uma medida absoluta. A entropia absoluta de uma fonte contínua é sempre infinita o que é razoável de admitir devido ao facto de X_x ser uma variável aleatória contínua e portanto poder tomar um número incontável de valores.

A entropia relativa, sendo finita, é uma medida útil de informação de fontes contínuas se se evitarem comparações com diferentes sinais de referência.

A questão que se coloca no caso das fontes contínuas é semelhante à que se colocou no caso das fontes discretas: Qual a função de densidade de probabilidade $p_X(x)$ que maximiza o valor da entropia $\mathcal{H}(X_x)$ para uma dada fonte?

O resultado que então se obteve (secção 8.1.2) — expresso na relação 8.5 e demonstrado no problema 6 — foi que $\mathcal{H}(X)$ era máxima quando os símbolos eram estatisticamente independentes, tendo-se concluído que esse máximo era obtido com $P_i = 1/m$.

Podemos então afirmar que se a variável aleatória contínua X_x toma um grande número de valores e estes são estatisticamente independentes e de valor quadrático médio limitado — os valores das amplitudes das amostras do sinal $x(t)$ — então, pelo *teorema do limite central*, a função de densidade de promplitudes das amostras do sinal $x(t)$ — então, pelo *teorema do limite central*, a função de densidade de probabilidade $p_X(x)$ dessa variável é Gaussiana de média nula e é a que maximiza $\mathcal{H}(X_x)$

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x^2}{2\sigma_x^2}} \quad (8.33)$$

O valor quadrático médio de X_x não é mais do que a potência média, $S = \sigma_x^2$, do sinal $x(t)$

$$S = \overline{x^2} = \int_{-\infty}^{+\infty} x^2 p_X(x) dx \quad (8.34)$$

que tem de ser finita.

O valor do máximo da entropia obtem-se de seguida substituindo o valor de $p_X(x)$ na equação 8.32, dando

$$\mathcal{H}(X_x)_{max} = \frac{1}{2} \log_2(2\pi e S) \quad (8.35)$$

Capacidade do canal contínuo

A transferência de informação num canal contínuo toma a forma de transmissão de um sinal. A fonte emite um sinal $x(t)$ o qual, depois de corrompido pelo ruído de transmissão, chega ao destino sob a forma de um outro sinal $y(t)$. A *informação mútua média*, ou *entropia mútua*, é definida por analogia com o caso discreto (ver equação 8.23)

$$\mathcal{H}(X_x, Y_y) = \mathcal{H}(Y_y) - \mathcal{H}(Y_y|X_x) \quad (8.36)$$

em que nos interessa tomar esta forma dado que, tal como no caso discreto, normalmente conhecemos explicitamente a *f.d.p* de transição directa $p_Y(y|x)$ e não a inversa $p_X(x|y)$.

Dado que as amplitudes do ruído, $n(t)$, são estatisticamente independentes e ele se manifesta aditivamente ao sinal, isto é,

$$y(t) = x(t) + n(t) \quad (8.37)$$

a entropia do ruído é definida por uma expressão equivalente à da equação 8.32

$$\mathcal{H}(Y_y|X_x) \stackrel{def}{=} \int_{-\infty}^{\infty} p_N(n) \log_2 \frac{1}{p_N(n)} dn \quad (8.38)$$

em que $p_N(n)$ é a função de densidade de probabilidade do ruído que será também Gaussiano

$$p_N(n) = \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{n^2}{2\sigma_n^2}} \quad (8.39)$$

O valor quadrático médio do sinal de ruído $n(t)$ representa a sua potência média, $N = \sigma_n^2$.

O valor da entropia do ruído obtem-se de igual forma substituindo o valor de $p_N(n)$ na equação 8.38

$$\mathcal{H}(Y_y|X_x) = \frac{1}{2} \log_2(2\pi e N) \quad (8.40)$$

Da equação 8.37 deduz-se que o sinal de destino $y(t)$ terá potência média $\overline{y^2} = S + N$ pelo que $\mathcal{H}(Y_y)$ se obtem da equação 8.35 substituindo S por $S + N$

$$\mathcal{H}(Y_y) = \frac{1}{2} \log_2[2\pi e(S + N)] \quad (8.41)$$

A máxima quantidade de informação transferida por amostra de $y(t)$ dá a *capacidade do canal contínuo*.

$$\begin{aligned} C_c &\stackrel{def}{=} \max_{p_X(x)} \mathcal{H}(X_x, Y_y) \\ &= \max_{p_X(x)} \{ \mathcal{H}(Y_y) - \mathcal{H}(Y_y|X_x) \} \\ &= \max_{p_X(x)} \left\{ \frac{1}{2} \log_2[2\pi e(S + N)] - \frac{1}{2} \log_2[2\pi eN] \right\} \quad (8.42) \end{aligned}$$

$$= \frac{1}{2} \log_2 \left(1 + \frac{S}{N} \right) \quad \text{bits/amostra} \quad (8.43)$$

Se o canal de transmissão tem uma *largura de banda* B_T *Hz* então o ritmo máximo de transmissão de amostras é, como se sabe,

$$r_c = 2 B_T \quad (8.44)$$

donde se obtem a capacidade do canal contínuo como sendo o máximo do débito médio de transferência de informação por amostra, tal como se definiu em 8.29 com $C = r_c C_c$:

$$C = B_T \log_2 \left(1 + \frac{S}{N} \right) \quad \text{bits/seg} \quad (8.45)$$

8.5 Problemas

- 8.1 – Uma fonte emite uma de quatro mensagens possíveis m_1 , m_2 , m_3 e m_4 com probabilidades $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$ e $\frac{1}{8}$ respectivamente. Calcular a quantidade de informação de cada mensagem e a informação média por mensagem.
- 8.2 – Uma carta é tirada de um baralho de cartas de jogo.
- É informado que a carta que tirou é uma espada. Quanta informação recebeu?
 - Quanta informação recebe se lhe fôr dito que a carta que tirou é um ás?
 - Quanta informação recebe se lhe fôr dito que a carta que tirou é um ás de espadas? Verifique a relação que existe entre este resultado e os obtidos em a) e b).
- 8.3 – Uma fonte emite uma sequência independente de símbolos de um alfabeto consistindo em 5 símbolos A, B, C, D e E com probabilidades de ocorrência dos símbolos $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{8}$, $\frac{3}{16}$ e $\frac{5}{16}$ respectivamente. Determinar a entropia desta fonte.
- 8.4 – Calcular o débito de entropia de uma fonte telegráfica que emite pontos e traços com probabilidades de ocorrência e tempos de duração do ponto e do traço respectivamente $P_p = \frac{2}{3}$, $\tau_p = 0.2$ s, $P_t = \frac{1}{3}$, $\tau_t = 0.4$ s. Qual o tempo médio de duração de um símbolo desta fonte?
- 8.5 – Uma fonte de dados tem oito símbolos que são produzidos em blocos de três a um débito de 1000 blocos por segundo. O primeiro símbolo de cada bloco é sempre o mesmo (presumivelmente para sincronismo) e os dois seguintes podem ser preenchidos por quaisquer dos oito símbolos com igual probabilidade. Qual é o débito de entropia, \mathcal{R} , da fonte?
- 8.6 – Mostrar que $\sum_{j=1}^m P_j \log_2(1/mP_j) = \mathcal{H}(X) - \log_2 m$ e utilizar este resultado e o facto de que $\ln(\alpha) \leq \alpha - 1$ para provar que $\mathcal{H}(X) \leq \log_2 m$.
- 8.7 – Uma fonte de informação tem um alfabeto de dimensão m . Um dos símbolos tem probabilidade ϵ enquanto os outros são igualmente prováveis. Determinar $\mathcal{H}(X)$ em termos de m e ϵ .

- 8.8 – Suponha que uma fonte tem $m = 3$ símbolos com probabilidades $P_1 = p$ e $P_2 = P_3$. Mostre que $\mathcal{H}(X) = \Omega(p) + 1 - p$. Esboce $\mathcal{H}(X)$ em função de p .
- 8.9 – Uma fonte de dados binária produz símbolos 0 e 1 com $p_0 = \frac{3}{8}$ e $p_1 = \frac{5}{8}$ e a influência entre símbolos em grupos de dois símbolos sucessivos é tal que $p_{1/0} = \frac{3}{4}$ e $p_{0/1} = \frac{1}{16}$. Calcular a entropia condicional $\mathcal{H}_c(X)$ e comparar com $\mathcal{H}(X)$.
- 8.10 – O código Morse internacional utiliza uma sequência de pontos e traços para transmitir letras do alfabeto alfanumérico. O traço é representado por um pulso de três unidades de tempo de duração e o ponto por um pulso de uma unidade de tempo de duração. A probabilidade de ocorrência de um traço é $\frac{1}{3}$ da probabilidade de ocorrência de um ponto.
- Calcular o conteúdo de informação de um ponto e de um traço.
 - Calcular a informação média do código.
 - Suponha que o traço dura 1 *ms* que é o mesmo intervalo de tempo que a pausa entre símbolos. Determinar o débito médio de transmissão de informação.
- 8.11 – Aplicar o algoritmo de Huffman à fonte do exemplo 8.3. O resultado será idêntico ao código IV. Confirmar que este código possui a propriedade ótima $N_i = I_i$ e que os dígitos binários 0 e 1 são equiprováveis.
- 8.12 – Uma fonte emite sequências independentes de símbolos de um alfabeto contendo cinco símbolos com probabilidades 0.4, 0.2, 0.2, 0.1 e 0.1.
- Calcule a entropia da fonte.
 - Defina um codificador para a fonte com dimensão de bloco $n = 2$.
- 8.13 – Determinar um código de comprimento variável para a fonte do problema 8.9
- Com rendimento não inferior a 60%.
 - Com rendimento não inferior a 80%.
 - Qual a compressão conseguida em cada caso?

- 8.14 – Confirmar que $\mathcal{H}(Y|X) = \Omega(\alpha)$ para o canal binário simétrico.
- 8.15 – Considere um canal com a propriedade de x_i e y_j serem estatisticamente independentes $\forall i, j$. Mostrar que $\mathcal{H}(X|Y) = \mathcal{H}(X)$ e que $\mathcal{H}(X, Y) = 0$. Que significa, fisicamente, x_i e y_j serem estatisticamente independentes?
- 8.16 – Considere o canal binário não-simétrico representado na figura 8.8 em que as probabilidades dos símbolos 0 e 1 chegarem ao destino errados são respectivamente α e β ($\alpha, \beta < 1$).

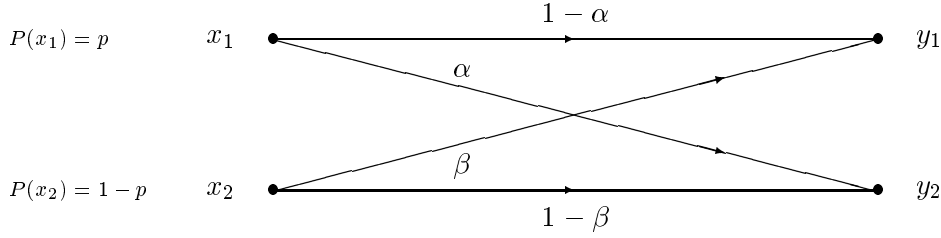


Figura 8.8: Canal binário não-simétrico

- a) Determine a entropia na fonte $\mathcal{H}(X)$, a entropia no destino $\mathcal{H}(Y)$, a equivocação $\mathcal{H}(X|Y)$ e a entropia do erro $\mathcal{H}(Y|X)$, quando $p(x = 0) = \frac{1}{4}$, $p(x = 1) = \frac{3}{4}$, $\alpha = 0.25$ e $\beta = 0.1$.
- b) Determine a capacidade do canal para $\alpha = 0.25$ e $\beta = 0.1$.
- c) Determine a capacidade do canal binário simétrico ($\alpha = \beta$).
- 8.17 – Mostre que a entropia mútua no canal binário não-simétrico do problema 8.16 é dada por

$$\mathcal{H}(X, Y) = \Omega[\beta + (1 - \alpha - \beta)p] - p\Omega(\alpha) - (1 - p)\Omega(\beta)$$

- 8.18 – Diz-se que o canal do problema 8.16 é *inútil* se $\beta = 1 - \alpha$. Justifique esta afirmação utilizando argumentos intuitivos e depois analíticos.

- 8.19 – Se a e b forem constantes, mostre que

$$\frac{d}{dp} \Omega(a + bp) = b \log_2 \frac{1 - a - bp}{a + bp}$$

e aplique esta relação para confirmar que $\mathcal{H}(X, Y)$ na equação 8.27 é máxima quando $p = 1/2$.

- 8.20 – O canal do problema 8.16 é chamado um canal Z quando $\beta = 0$. Utilize a relação do problema 8.19 para mostrar que se $\alpha = 1/2$, então $\mathcal{H}(X, Y)$ é máxima quando $p = 2/5$. Depois calcule C_c .
- 8.21 – Suponha que no canal do problema 8.16 se tem $\alpha = 1/4$ e $\beta = 0$. Utilize a relação do problema 8.19 para calcular C_c .
- 8.22 – Determine a capacidade do canal ternário simétrico no qual a probabilidade de erro por símbolo é de 2α ($2\alpha < 1$).

fim do capítulo 8

Capítulo 9

Códigos para Control de Erros

Este capítulo trata o problema da codificação da informação para a transmissão, designada de codificação do canal, cuja finalidade é a de permitir controlar os erros de transmissão em sistemas de telecomunicações não-fiaíveis ou ruidosos. Segundo o Teorema de Shannon da codificação do canal, existe um código para a informação que possibilita a sua transmissão desde a fonte até ao destino com um número de erros arbitrariamente pequeno.

Neste capítulo estabelecem-se as bases matemáticas que permitem a construção de alguns desses códigos e abordam-se os métodos mais utilizados quer na detecção quer na correcção de erros, conjuntamente designados por métodos de *control* de erros. Considera-se apenas o caso da transmissão digital binária.

9.1 Tipos de erros

Há fundamentalmente dois tipos de ruído que afectam as comunicações digitais: o ruído branco e *gaussiano* e o ruído *impulsivo*. E para enfrentar cada um destes tipos de ruído são utilizados códigos com estruturas distintas.

Os erros de transmissão causados pelo ruído branco e *gaussiano* são tais que a ocorrência de erro num determinado dígito (binário) não afecta os

dígitos subsequentes, isto é, as ocorrências de erros são estatisticamente independentes. Neste caso o canal discreto pode ser modelado pelo canal binário simétrico. Os erros de transmissão devidos ao ruído branco e gaussiano são designados de erros aleatórios.

A presença de ruído *impulsivo* é caracterizada por longos intervalos de tempo em que os dígitos não são corrompidos, intercalados por *mólhos*¹ de dígitos corrompidos. O ruído deste tipo, normalmente de grande energia, quando ocorre, afecta um conjunto (mólho) de dígitos seguidos, ou seja, as ocorrências de erros não são estatisticamente independentes.

As técnicas de control de erros destinadas a lidar com o primeiro tipo de erros utilizam os designados *códigos de correcção de erros aleatórios*² e com o segundo, *códigos de correcção de erros aos mólhos*³. Neste capítulo abordaremos apenas os códigos de correcção de erros aleatórios cuja base matemática é semelhante à dos destinados à correcção de erros aos mólhos visto que nos interessa apenas compreender os fundamentos desta teoria.

9.2 Tipos de códigos

Os códigos para control de erros são basicamente de dois tipos: *códigos de bloco* e *códigos convolucionais*.

Nos códigos de bloco, cada conjunto de k dígitos de informação é acompanhado de $n - k$ de dígitos redundantes, chamados dígitos de *verificação de paridade*, calculados a partir dos dígitos de informação e formando um *bloco* de tamanho fixo, de n dígitos, designado por *palavra de código*.

Nos códigos convolucionais não há dígitos de verificação distintos dos dígitos de informação. A sequência dos dígitos de informação dá origem a uma outra sequência de dígitos dependentes entre si de uma forma também calculada de certa maneira (por convolução).

Embora seja possível tratar matematicamente ambos os tipos de códigos numa teoria unificada, esta seria desnecessariamente complexa para este curso dificilmente permitindo uma visualização imediata das realizações práticas. Neste capítulo abordaremos apenas os códigos de bloco visto serem os mais utilizados nos protocolos de comunicação de dados.

¹*bursts* na terminologia inglesa.

²*random error correcting codes*

³*burst error correcting codes*

9.3 Códigos Lineares de Bloco

Os códigos de bloco mais usuais são os lineares. Cada bloco de k dígitos da mensagem (dígitos de informação) são codificados num bloco de $n > k$ dígitos pela adição de $n - k$ dígitos de verificação calculados a partir dos k dígitos da mensagem. A figura 9.1 ilustra a entrada e a saída do codificador do canal. Cada bloco de n dígitos à saída do codificador é

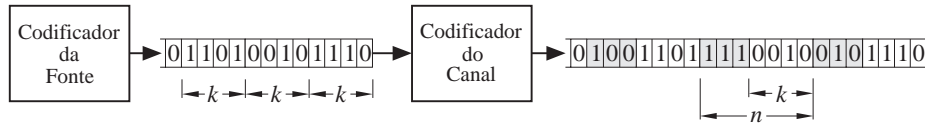


Figura 9.1: Codificador de bloco do canal

uma *palavra de código*. Os códigos nos quais os dígitos da mensagem aparecem destacados num sub-bloco conexo da palavra de código são chamados códigos *sistemáticos* ou *separáveis*. Se cada uma das 2^k diferentes palavras de código for obtida por combinação *linear* de k palavras de código independentes então o código é chamado de *código linear de bloco*.

Um bloco de mensagem (de dígitos de informação, a que chamaremos dígitos de *dados*) será um *tuplo* $D = (d_0 \ d_1 \ d_2 \ \dots \ d_{k-1})$ com $d_j \in \{0, 1\}$. Haverá portanto 2^k blocos de mensagem distintos. Cada bloco de mensagem é transformado numa palavra de código representada pelo *tuplo* $C = (c_0 \ c_1 \ c_2 \ \dots \ c_{n-1})$ com $c_j \in \{0, 1\}$ e haverá portanto apenas 2^k palavras de código distintas, uma para cada bloco de mensagem. As restantes $2^n - 2^k$ palavras que se podem formar com os n dígitos binários não fazem parte do dicionário do código e, se porventura forem recebidas no destino, denunciam a ocorrência de erro.

Estes códigos são designados de *códigos*-(n, k) ou, abreviadamente, $C(n, k)$. O seu *rendimento*, ρ , é igual a

$$\rho = \frac{k}{n}$$

pois aumentam o número de dígitos originalmente produzidos pelo codificador da fonte de um factor igual a $\frac{n}{k}$.

Num código sistemático linear de bloco os primeiros k dígitos do código são os dígitos de informação, isto é, $c_j = d_j$ para $j = 0, 2, \dots, k - 1$ e os últimos $n - k$ dígitos são os de verificação.

9.3.1 Distância de Hamming

Definição 9.1 Distância de Hamming entre duas palavras de um código de bloco, $d(C_i, C_j)$, é o número de posições em que as duas palavras, C_i e C_j , diferem.

A distância de Hamming entre as palavras (11100) e (10101), por exemplo, é igual a 2. Elas diferem apenas na segunda e na quinta posições. Duas palavras idênticas estão à distância zero uma da outra e duas palavras distintas estarão a uma distância igual ou superior a uma unidade. O conceito de distância de Hamming é passível de interpretação geométrica semelhante à da distância Euclideana (usual) entre dois pontos do espaço físico. Com efeito, pode estabelecer-se uma correspondência biunívoca entre as 2^n palavras distintas de n dígitos binários e os 2^n vértices de um hipercubo unitário no espaço de n dimensões como se mostra na figura 9.2 para vários valores de n . Cada palavra de n

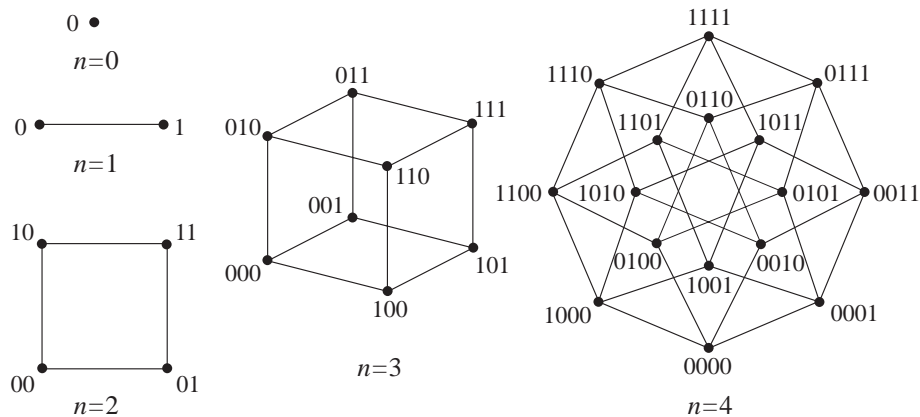


Figura 9.2: Palavras de código como coordenadas dos vértices do hipercubo unitário a n dimensões

dígitos define a coordenada de um vértice do cubo a n dimensões. Dois vértices adjacentes, isto é, situados nos extremos de uma mesma aresta de comprimento igual a 1 unidade, estão distanciados de uma unidade e

as suas coordenadas diferem apenas numa única posição dimensional. Se nos movermos de um vértice para outro, ao longo de d arestas, então a distância entre esses dois vértices será igual a d desde que não haja um número menor de arestas que liguem esses dois vértices.

Definição 9.2 *Distância mínima de um código de bloco, d_{\min} , é a menor das distâncias de Hamming entre quaisquer duas palavras desse código.*

Qualquer código com um número constante de dígitos será um sub-conjunto das 2^n palavras disponíveis e poderá ser representado mapeando-o no cubo unitário a n dimensões. As propriedades métricas das palavras de um código determinam a sua capacidade de control de erros.

9.3.2 Capacidade de control de erros dos códigos de bloco

Se se construir um código de tal forma que cada palavra esteja a uma distância igual ou superior a 2 de qualquer outra palavra de código então pode sempre detectar-se um único erro em qualquer palavra transmitida. Com efeito, todos os casos em que ocorra um único erro darão lugar a uma nova palavra à distância 1 da palavra correcta original e à distância *pelo menos* 1 de alguma outra palavra de código. Assim, a palavra corrompida será diferente (e poderá ser identificada) de qualquer palavra válida. Não será possível corrigir o erro dado que não é possível localizar a posição do dígito errado visto que a mesma palavra pode ter resultado da corrupção de várias palavras de código (válidas) em diferentes posições dos dígitos. Da mesma forma, se ocorrerem dois erros na mesma palavra o resultado pode dar origem a uma palavra válida e a corrupção não será detectada. Torna-se evidente que para detectar a ocorrência de até e erros numa palavra de código a distância mínima do código deve ser pelo menos igual a $e + 1$. De facto, e erros simultâneos darão origem a uma nova palavra à distância e da palavra original, mas também pelo menos à distância 1 de qualquer outra palavra de código.

Considere-se agora um código com distância mínima igual a 3. Um único erro em qualquer posição de uma palavra de código produzirá uma nova palavra à distância 1 da palavra original e pelo menos à distância 2 de qualquer outra palavra de código, portanto, mais perto daquela do que destas.

Se se adoptar o princípio da máxima verosimilhança⁴, é mais provável que a palavra original fosse a que está à distância 1, na qual terá ocorrido um único erro, do que qualquer outra à distância 2, em que teriam ocorrido dois erros.

Ao tomar esta decisão de correção diz-se que o decodificador do canal corrigiu um erro simples. Contudo, pode ter-se cometido um erro de descodificação caso tenham realmente ocorrido dois erros na palavra. Nesta situação a palavra corrompida estará à distância 1 de uma outra palavra de código e à distância 2 da palavra original.

Assim, a probabilidade de erro por dígito de informação entregue ao destino, P_E , será menor do que a probabilidade de erro por dígito no canal, P_e , porque se elimina, pelo menos, a possibilidade de erros simples por palavra. Na secção 9.5 abordar-se-á esta questão para o caso dos sistemas que utilizam o mecanismo da correção de erros progressiva.

Em conclusão, se se desejar corrigir até e erros por bloco (palavra de código) deve assegurar-se que o código utilizado possui uma distância mínima igual a $2e + 1$ pois, deste modo, e erros darão lugar a uma palavra à distância e da palavra correcta original e a pelo menos $e + 1$ de qualquer outra palavra válida (de código).

Seja d_{min} a distância mínima de um código,

Para detectar até e_d erros: $d_{min} = e_d + 1$ Para corrigir até e_c erros: $d_{min} = 2e_c + 1$

Existe sempre um compromisso entre a capacidade de correção de erros de um código e a sua capacidade de detecção de erros dado que um código que corrige e_c erros pode ser empregue, alternativamente, como um código detector de $e_d = 2e_c$ erros.

Definição 9.3 Peso de uma palavra C_i de um código de bloco, $p(C_i)$, é o número de dígitos 1 que a palavra C_i contém.

Definição 9.4 Peso mínimo de um código de bloco, $[p(C_i)]_{min}$ é o peso da palavra de menor peso desse código, exceptuando a palavra de peso zero.

⁴Num bloco de n dígitos, a probabilidade de ocorrerem apenas i erros é maior do que a probabilidade de ocorrerem $i + 1$ erros ($0 \leq i \leq n - 1$), i.é., $P(0 \text{ erros}) \gg P(1 \text{ erro}) \gg P(2 \text{ erros}) \gg \dots \gg P(n \text{ erros})$ quando $P(1 \text{ erro}) = P_e \ll 1$.

Como consequência da linearidade pode determinar-se a distância mínima de um código conhecido o seu peso mínimo:

Teorema 9.1 — Distância mínima

A distância mínima de um código de bloco é igual ao seu peso mínimo.

Demonstração

Seja $p(C_i)$ o peso da palavra C_i . A distância de Hamming entre quaisquer duas palavras do código, $C_i = (c_{0i} \ c_{1i} \ \dots \ c_{n-1i})$ e $C_j = (c_{0j} \ c_{1j} \ \dots \ c_{n-1j})$, será então

$$d(C_i, C_j) = p(C_i + C_j)$$

dado que $c_{ki} + c_{kj} = 1$ sempre que $c_{ki} \neq c_{kj}$. Portanto, a distância entre C_i e C_j é igual ao peso de uma outra palavra de código $C_z = C_i + C_j$. Mas se $C_j = (00 \dots 0)$ então $C_i + C_j = C_i$ e por conseguinte

$$d_{\min} = [p(C_i)]_{\min} \quad \text{com} \quad C_i \neq (00 \dots 0)$$

c.q.d.

9.3.3 Códigos de Hamming

Para cada valor de $n - k$ existe um código binário com $n - k$ dígitos de verificação e comprimento da palavra de código

$$n = 2^{n-k} - 1$$

A cardinalidade do dicionário é 2^k e o código é capaz de corrigir um erro simples em qualquer posição da palavra recebida. O rendimento ou taxa do código é

$$\rho = \frac{k}{n} = \frac{2^{n-k} - 1 - (n - k)}{2^{n-k} - 1} = 1 - \frac{n - k}{2^{n-k} - 1}$$

Um código que obedeça a estas condições é designado de código de Hamming.

Os códigos de Hamming são códigos correctores de erros simples ou detectores de erros duplos de elevado rendimento. A ocorrência de dois ou mais erros não pode, em geral ser corrigida e a ocorrência de três ou mais erros não pode, em geral ser detectada.

9.4 Códigos Cíclicos Binários

Os *códigos cíclicos* constituem uma sub-classe dos códigos lineares de bloco que é atractiva porque tanto a codificação, como a detecção ou a correcção dos erros no decodificador, podem realizar-se facilmente com registos de deslocamento realimentados lineares⁵ devido à simplicidade da estrutura matemática que possuem. Estes códigos são representados por *polinómios* no *corpo finito* binário.

O corpo binário é o corpo finito constituído por um conjunto de dois elementos, por exemplo, o conjunto $\mathcal{A} = \{0, 1\}$, onde estão definidas as operações de adição, $+$, e multiplicação, \cdot , obedecendo às seguintes regras:

$+$	0	1
0	0	1
1	1	0

\cdot	0	1
0	0	0
1	0	1

em que o elemento 0 é a identidade aditiva (o elemento zero do corpo) e o elemento 1 é a identidade multiplicativa (o elemento unidade do corpo) e onde se pode verificar que cada elemento é o seu próprio inverso aditivo (simétrico), isto é, $y = -y$, $\forall y \in \mathcal{A}$, e cada elemento diferente de zero é o seu próprio inverso multiplicativo, ou seja, $y = y^{-1}$, $\forall y \in \mathcal{A}_{-\{0\}}$. Por outras palavras, a multiplicação é a multiplicação usual e a adição é a adição usual tomada módulo-2, ou seja, é a operação *ou-exclusivo*.

9.4.1 Estrutura algébrica dos códigos cíclicos

Definição 9.5 Um código linear de bloco $C(n, k)$ é cíclico se possuir a seguinte propriedade:

Se o tuplo $C = (c_0, c_1, c_2, \dots, c_{n-1})$ for uma palavra de código então o tuplo $C^{(1)} = (c_{n-1}, c_0, c_1, \dots, c_{n-2})$ obtido por deslocação cíclica direita de uma posição de C também é uma palavra de código.

Desta definição resulta que

$$C^{(i)} = (c_{n-i}, c_{n-i+1}, \dots, c_{n-1}, c_0, c_1, \dots, c_{n-i-1})$$

⁵linear feedback shift registers ou lfsr

também é uma palavra de código. Esta propriedade dos códigos cíclicos permite considerar os elementos de cada palavra de código como coeficientes de um polinómio de grau $n - 1$ numa variável real x , isto é,

$$C \rightarrow C(x) = c_0 + c_1x + c_2x^2 + \dots + c_{n-1}x^{n-1}$$

em que os coeficientes $c_j \in \{0, 1\}$ do polinómio e as operações $+$ e \cdot , são, respectivamente, os elementos e as operações do corpo binário.

O polinómio de código $C^{(i)}(x)$ correspondente à palavra de código $C^{(i)}$ é

$$C^{(i)}(x) = c_{n-i} + c_{n-i+1}x + \dots + c_{n-1}x^{i-1} + c_0x^i + c_1x^{i+1} + \dots + c_{n-i-1}x^{n-1}$$

e pode mostrar-se que o polinómio $C^{(i)}(x)$ é o resto da divisão de $x^i C(x)$ por $x^n + 1$, isto é,

$$x^i C(x) = q(x)(x^n + 1) + C^{(i)}(x)$$

(Resolver o problema 9.1).

Enunciemos e demonstremos agora um importante teorema que nos permite gerar códigos cíclicos.

Teorema 9.2 — Geração de códigos cíclicos

Seja $g(x)$ um polinómio de grau $n - k$, que divide o polinómio $x^n + 1$. Então $g(x)$ gera um código cíclico (n, k) no qual o polinómio de código $C(x)$ correspondente ao polinómio de informação (de dados) $D(x) = d_0 + d_1x + d_2x^2 + \dots + d_{k-1}x^{k-1}$ é gerado por

$$C(x) = D(x) \cdot g(x) \tag{9.1}$$

Demonstração

Considerem-se os k polinómios $g(x), xg(x), x^2g(x), \dots, x^{k-1}g(x)$ de grau menor ou igual a $n - 1$. Qualquer combinação linear destes polinómios é um polinómio de grau menor ou igual a $n - 1$ e é múltiplo de $g(x)$. Em particular, a seguinte combinação linear está nessas condições

$$\begin{aligned} C(x) &= d_0g(x) + d_1xg(x) + d_2x^2g(x) + \dots + d_{k-1}x^{k-1}g(x) \\ &= D(x) \cdot g(x) \end{aligned}$$

Existe um total de 2^k polinómios $C(x)$ distintos correspondendo aos 2^k diferentes tuplos $D = (d_0, d_1, d_2, \dots, d_{k-1})$ e os polinómios de código $C(x)$

correspondentes a esses 2^k polinómios constituem um código linear (n, k) . Para mostrar que o código resultante é cíclico, considere-se que $C(x) = c_0 + c_1x + c_2x^2 + \dots + c_{n-1}x^{n-1}$ é um polinómio deste código. O produto de x por $C(x)$ dá

$$\begin{aligned} x \cdot C(x) &= c_0x + c_1x^2 + c_2x^3 + \dots + c_{n-2}x^{n-1} + c_{n-1}x^n \\ &= c_{n-1}(x^n + 1) + (c_{n-1} + c_0x + c_1x^2 + \dots + c_{n-2}x^{n-1}) \\ &= c_{n-1}(x^n + 1) + C^{(1)}(x) \end{aligned}$$

onde se fez $c_{n-1} + c_{n-1} = 0$. $C^{(1)}(x)$ obtem-se por deslocação cíclica direita de $C(x)$. Dado que $x \cdot C(x)$ e $x^n + 1$ são ambos divisíveis por $g(x)$, $C^{(1)}(x)$ também é divisível por $g(x)$. Assim, sendo $C^{(1)}(x)$ um múltiplo de $g(x)$, pode ser expresso como uma combinação linear dos polinómios $g(x)$, $xg(x)$, $x^2g(x)$, \dots , $x^{k-1}g(x)$, o que significa que $C^{(1)}(x)$ é também um polinómio de código. Em conclusão, pela definição de um código cíclico, o código linear gerado por $g(x)$, $xg(x)$, \dots , $x^{k-1}g(x)$ é um código cíclico (n, k) . c.q.d..

O polinómio $g(x)$ é chamado *polinómio gerador* do código cíclico. Pode mostrar-se que $g(x)$ é um polinómio irredutível no corpo binário, isto é, um polinómio primo (não factorizável) pelo que será sempre $g_{n-k-1} = g_0 = 1$.

9.4.2 Geração de códigos cíclicos sistemáticos

Dado o polinómio gerador $g(x)$ de um código cíclico, as respectivas palavras de código podem ser construídas de uma forma sistemática separando os dígitos de verificação de paridade dos dígitos da mensagem da seguinte maneira

$$C = \underbrace{(r_0, r_1, r_2, \dots, r_{n-k-1})}_{\substack{n-k \text{ dígitos} \\ \text{de verificação} \\ \text{de paridade}}}, \underbrace{(d_0, d_1, d_2, \dots, d_{k-1})}_{\substack{k \text{ dígitos} \\ \text{da mensagem}}} \quad (9.2)$$

onde $r(x) = r_0 + r_1x + r_2x^2 + \dots + r_{n-k-1}x^{n-k-1}$ é o *polinómio de verificação de paridade* correspondente ao polinómio da mensagem (a que chamaremos polinómio dos dados), $D(x) = d_0 + d_1x + d_2x^2 + \dots + d_{k-1}x^{k-1}$. É simples verificar que o polinómio $r(x)$ é o resto da divisão de $x^{n-k}D(x)$ por $g(x)$, ou seja

$$x^{n-k}D(x) = q(x) \cdot g(x) + r(x) \quad (9.3)$$

$$C(x) = r(x) + x^{n-k}D(x) \quad (9.4)$$

b) calculando os polinómios de código na forma sistemática

$$\begin{array}{r|l} x^5+x^4+x^3 & x^3+x+1 \\ x^5+ & x^2+x \\ \hline 0+x^4+0+x^2 & \\ & x^4+ & x^2+x \\ & \hline & 0+ & 0+x \end{array} = r(x)$$

donde resulta $q(x) = x^2 + x$ e $r(x) = x$ pelo que $C = (\underbrace{010}_{r(x)} \underbrace{1110}_{D(x)})$. De igual forma, podem obter-se outras palavras do código por rotação cíclica a partir desta. A terceira coluna da tabela 9.1 lista o código sistemático completo assim calculado e na quarta coluna indicam-se os pesos de cada palavra deste código.

Tabela 9.1: Código cíclico (7, 4) gerado por $g(x) = 1 + x + x^3$

Informação $D(x)$	Código criptográfico $C(x) = D(x) \cdot g(x)$	Código sistemático $C(x) = r(x) + x^{n-k}D(x)$	Peso $p(C_i)$
0 0 0 0	0 0 0 0 0 0 0	0 0 0 0 0 0 0	0
0 0 0 1	0 0 0 1 1 0 1	1 0 1 0 0 0 1	3
0 0 1 0	0 0 1 1 0 1 0	1 1 1 0 0 1 0	4
0 0 1 1	0 0 1 0 1 1 1	0 1 0 0 0 1 1	3
0 1 0 0	0 1 1 0 1 0 0	0 1 1 0 1 0 0	3
0 1 0 1	0 1 1 1 0 0 1	1 1 0 0 1 0 1	4
0 1 1 0	0 1 0 1 1 1 0	1 0 0 0 1 1 0	3
0 1 1 1	0 1 0 0 0 1 1	0 0 1 0 1 1 1	4
1 0 0 0	1 1 0 1 0 0 0	1 1 0 1 0 0 0	3
1 0 0 1	1 1 0 0 1 0 1	0 1 1 1 0 0 1	4
1 0 1 0	1 1 1 0 0 1 0	0 0 1 1 0 1 0	3
1 0 1 1	1 1 1 1 1 1 1	1 0 0 1 0 1 1	4
1 1 0 0	1 0 1 1 1 0 0	1 0 1 1 1 0 0	4
1 1 0 1	1 0 1 0 0 0 1	0 0 0 1 1 0 1	3
1 1 1 0	1 0 0 0 1 1 0	0 1 0 1 1 1 0	4
1 1 1 1	1 0 0 1 0 1 1	1 1 1 1 1 1 1	7

Quatro observações são devidas relativamente aos códigos deste exemplo e que também se aplicam à generalidade dos casos:

- (1) O conjunto das palavras de código é o mesmo em ambas as codificações embora possam não corresponder às mesmas palavras da mensagem.
- (2) O ciclo obtido pelas sucessivas deslocações cíclicas a partir de uma palavra de código pode não gerar *todas* as palavras de código.

Nestes casos haverá que repetir o procedimento a partir de uma palavra de mensagem cujo código não esteja contido num ciclo já gerado.

- (3) No código gerado através da equação 9.1 as posições dos dígitos de verificação e de informação não são localizáveis constituindo o que se designa por *código criptográfico*.
- (4) O peso mínimo dos códigos é 3 sendo a sua distância mínima também $d_{min} = 3$. Portanto, o código (7,4) é corrector de erros simples ($e_c = 1$) ou detector de erros duplos ($e_d = 2$). Trata-se de um código de Hamming pois verifica a relação $n = 2^{n-k} - 1$.

9.4.3 Codificação com registos de deslocamento

A maior vantagem que os códigos cíclicos oferecem é a simplicidade de realização prática dos codificadores e dos decodificadores bem como do control de erros.

As operações de codificação descritas pelas equações 9.1 e 9.4 envolvem respectivamente a multiplicação e a divisão de um polinómio pelo polinómio gerador $g(x)$. Em particular, no segundo caso, o polinómio de verificação de paridade, $r(x)$, é o resto daquela divisão. Para realizar a divisão utilizam-se *registos de deslocamento com realimentação*.

Considere-se a divisão de um polinómio *dividendo* qualquer $a(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$ de grau n pelo polinómio fixo *divisor* $b(x) = b_0 + b_1x + b_2x^2 + \dots + b_mx^m$ de grau m . O resultado será um polinómio *quociente*, $q(x)$, de grau $n - m$ e possivelmente um polinómio *resto*, $r(x)$, de grau menor que m

$$\frac{a(x)}{b(x)} = q(x) + \frac{r(x)}{b(x)}$$

O primeiro termo de $q(x)$ terá a forma $q_{n-m}x^{n-m}$ e pode deduzir-se q_{n-m} pelo facto de se ter de fazer $(q_{n-m}x^{n-m}) \cdot (b_mx^m)$ igual ao inverso aditivo do termo de maior ordem de $a(x)$, ou seja, a_nx^n . Assim, $q_{n-m} = a_nb_m^{-1}$. Na divisão executada manualmente continuar-se-ia, somando $a_nb_m^{-1}x^{n-m} \cdot b(x)$ a $a(x)$ eliminando-se o termo a_nx^n (e possivelmente vários outros). Repetir-se-ia o procedimento deduzindo o termo seguinte de $q(x)$ que seria $(a_{n-1} - a_nb_{m-1}b_m^{-1}) \cdot b_m^{-1}x^{n-m-1}$, ou seja,

$$\begin{array}{l}
a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \quad \left| \begin{array}{l} b_m x^m + b_{m-1} x^{m-1} + \dots + b_1 x + b_0 \\ a_n b_m^{-1} x^{n-m} + (a_{n-1} - a_n b_{m-1} b_m^{-1}) x^{n-m-1} + \dots \end{array} \right. \\
\\
\frac{a_n x^n + a_n b_{m-1} b_m^{-1} x^{n-1} + \dots + a_n b_1 b_m^{-1} x^{n-m+1} + a_n b_0 b_m^{-1} x^{n-m}}{(a_{n-1} - a_n b_{m-1} b_m^{-1}) x^{n-1} + \dots} \\
\frac{(a_{n-1} - a_n b_{m-1} b_m^{-1}) x^{n-1} + (a_{n-1} - a_n b_{m-1} b_m^{-1}) b_m^{-1} b_{m-1} x^{n-2} + \dots}{(a_{n-2} - (a_n b_{m-2} + a_{n-1} b_{m-1}) b_m^{-1} + a_n b_{m-1}^2 b_m^{-2}) + \dots} \\
\text{etc.}
\end{array}$$

para cada novo coeficiente do quociente q_i o polinómio $q_i b(x)$ é subtraído do dividendo e o procedimento repetido até não se encontrar novos termos para o quociente. Neste ponto restarão ainda alguns termos no dividendo que constituem o polinómio resto final, $r(x)$.

O circuito representado na figura 9.3 mecaniza esta operação de divisão polinomial. Nessa figura, as caixas quadradas representam circuitos multi-estáveis tipo-D interligados por adicionadores constituindo um registo

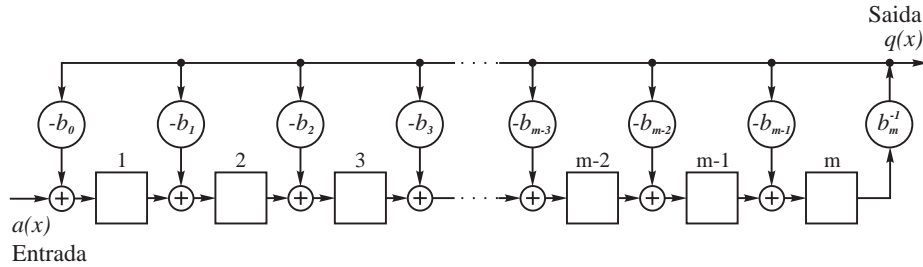


Figura 9.3: Circuito de divisão de polinómios

de deslocamento. Os circulos circunscrevendo os valores dos coeficientes do polinómio divisor são multiplicadores por escalar na álgebra considerada. A saída do registo de deslocamento, após multiplicação pela constante b_m^{-1} e, em cada estágio, pelo valor dum coeficiente do dividendo *realimenta* o registo nesse estágio adicionando-se ao valor da saída do estágio anterior.

No caso binário, que é o que interessa considerar, a adição é efectuada módulo-2, o multiplicador por escalar é uma ligação directa fechada se $b_i = 1$ ou uma ligação aberta se $b_i = 0$ e os multi-estáveis são bi-estáveis (*flip-flops*) tipo-D. A cada ocorrência do pulso de relógio de temporização (que não está representado) dos flip-flops, os valores à entrada destes transitam para as respectivas saídas. O registo é inicialmente colocado a zero. A saída $q(x)$ permanece a zero durante os primeiros m deslocamentos, ao fim dos quais o coeficiente da maior potência do dividendo a_n atinge o último estágio do registo. Aparece então o primeiro coeficiente da saída

não-zero, ou seja, $a_n b_m^{-1}$ (notar que no caso binário $b_i^{-1} \equiv b_i$). As ligações de realimentação asseguram que, para cada coeficiente do quociente, q_i , o polinómio $q_i b(x)$ seja subtraído do dividendo que é mantido no registo sob a forma de restos parciais. Após um total de $n + 1$ deslocamentos, todo o dividendo entrou no registo, todo o quociente já apareceu à saída e o polinómio resto final encontra-se no registo.

Deve notar-se que a divisão só começa quando o coeficiente da maior potência do dividendo chega ao último flip-flop, ou seja, apenas passados $n - k$ pulsos do relógio de temporização do circuito. Assim sendo e para que a divisão possa começar imediatamente, o polinómio dividendo deve entrar directamente somado à saída do último flip-flop cujo valor é 0 visto que o registo é inicializado a (00...0) antes do início da divisão.

Pode verificar-se que o mesmo circuito, alimentado à saída como se mostra na figura 9.4, efectua a divisão polinomial. Mas, para formar uma palavra de código sistemático, não é necessário o quociente mas sim o resto da divisão, portanto a saída útil deste circuito é o próprio dividendo seguido do resto final que ficou no registo.

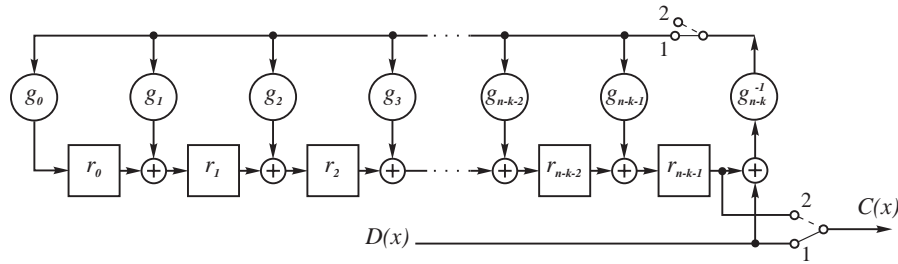


Figura 9.4: Codificador para um código sistemático (n, k)

Esta configuração é pois a mais apropriada para realizar a codificação sistemática pela qual são transmitidos em primeiro lugar os k dígitos da mensagem, seguidos dos $n - k$ dígitos do resto descarregados do registo, por deslocamento, após efectuada a divisão. Durante os k primeiros pulsos do relógio, os comutadores estão na posição 1, e durante os restantes $n - k$ pulsos na posição 2.

Exemplo 9.2 Definir um codificador para o código cíclico binário $(7, 4)$ gerado por $g(x) = 1 + x + x^3$ e verificar a sua operação utilizando a palavra de mensagem (0101).

Resolução

O codificador para este código está representado na figura 9.5 pois tem-se $n - k = 7 - 4 = 3$ e $g_0 = 1$, $g_1 = 1$, $g_2 = 0$ e $g_3 = 1$. Os passos da

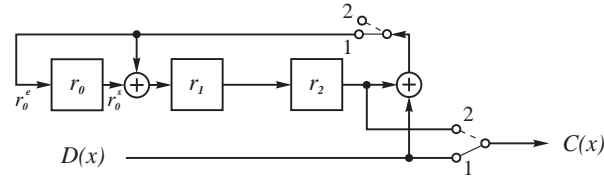


Figura 9.5: Codificador para o código cíclico sistemático $(7, 4)$ gerado por $g(x) = 1 + x + x^3$

operação de codificação estão discriminados na tabela 9.2 e portanto a palavra de código de $(0\ 1\ 0\ 1)$ é $(1\ 1\ 0\ 0\ 1\ 0\ 1)$.

Tabela 9.2: Operação de codificação da mensagem $(0\ 1\ 0\ 1)$

bit de entrada	entrada nos registos				saída dos registos			
$D(x)$	r_0^e	r_1^e	r_2^e		r_0^s	r_1^s	r_2^s	
—	0	0	0		0	0	0	
1	1	1	0	→	1	1	0	1 ^o deslocamento
0	0	1	1	→	0	1	1	2 ^o deslocamento
1	0	0	1	→	0	0	1	3 ^o deslocamento
0	1	1	0	→	1	1	0	4 ^o deslocamento

9.4.4 O síndrome: detecção e correcção de erros

Quando uma palavra de código C é transmitida através do canal, ela será eventualmente corrompida, chegando ao receptor uma palavra qualquer R . A partir desta o decodificador deverá determinar qual a palavra de código efectivamente transmitida. Para isso divide $R(x)$ pelo polinómio gerador $g(x)$ obtendo um resto $S(x)$ que é designado por *síndrome* de $R(x)$. Se o síndrome for zero a palavra recebida pertence ao código e portanto toma-a como correcta. Caso contrário decide que houve erro e

poderá ou não tentar corrigi-la. Essa divisão é representada pela equação

$$R(x) = P(x) \cdot g(x) + S(x) \quad (9.5)$$

em que $P(x)$ é o polinómio quociente. O polinómio síndrome $S(x)$ é de grau $n - k - 1$ ou inferior. Seja E a sequência de erro que ocorreu no canal. E pode ser representado por um polinómio $E(x)$, desconhecido, de grau $n - 1$ dando origem a $R(x)$ da seguinte forma:

$$R(x) = C(x) + E(x) \quad (9.6)$$

Substituindo este valor na equação 9.5 tem-se

$$C(x) + E(x) = P(x) \cdot g(x) + S(x) \quad (9.7)$$

e dado que $C(x) = D(x) \cdot g(x)$ tem-se

$$E(x) = [P(x) + D(x)] \cdot g(x) + S(x) \quad (9.8)$$

donde se verifica que o resto da divisão da sequência de erro pelo polinómio gerador é igual ao síndrome de $R(x)$. Por outras palavras, o síndrome do erro é igual ao síndrome da palavra recebida. Daqui se conclui que o síndrome contém informação acerca do padrão do erro e pode ser utilizado para corrigir R .

No código (7,4) dos exemplos anteriores o síndrome é um polinómio de grau $n - k = 2$, isto é, tem 3 dígitos, que é o número de dígitos suficiente para *apontar* para uma das sete possíveis posições do erro podendo portanto corrigi-lo. Caso tenham ocorrido dois ou mais erros o reduzido número de dígitos do síndrome neste código (7,4) não permite apontá-los a ambos. Apenas indicará que eles ocorreram.

Para a decodificação, é então efectuada uma divisão utilizando-se um circuito semelhante ao da figura 9.3 como se mostra na figura 9.6. Neste circuito existe um registo de deslocamento (*buffer*) de entrada com n flip-flops onde dá entrada a palavra recebida $R(x)$. Simultaneamente, é executada a divisão de $R(x)$ pelo polinómio gerador $g(x)$ no registo realimentado abaixo daquele, que conterà o resto logo que $R(x)$ esteja todo no *buffer*. A soma das posições em erro é $S(x) = r(x)$. Se ocorreu um erro simples então $S(x) = s_0 + s_1x + \dots + s_{n-k-1}x^{n-k+1}$ indica a posição desse erro.

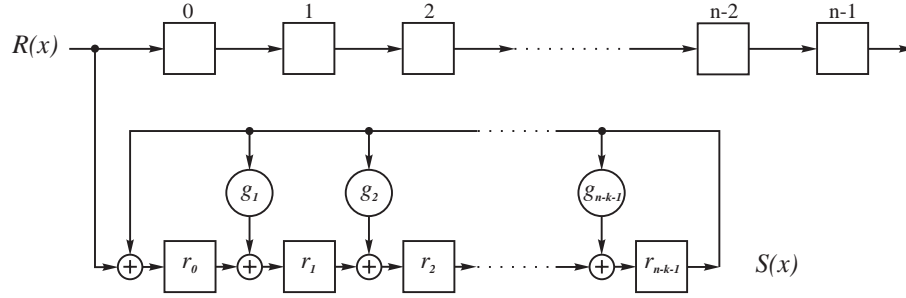


Figura 9.6: Divisão de $R(x)$ por $g(x)$ no decodificador

Caso o decodificador seja utilizado para detecção de erros, bastará calcular o ou-inclusivo dos valores dos registos do circuito divisor para se sinalizar a ocorrência de erro.

Caso seja utilizado para a correcção, numa solução elegante, adiar-se-á a correcção para a altura em que a palavra recebida saia do *buffer* de entrada para ser entregue ao destino. Assim, se o síndrome indica um erro na posição 2^i dever-se-á deixar sair do *buffer* $n - 1 - i$ dígitos de $R(x)$ e depois corrigir o dígito i quando ele aparecer à saída.

De modo a libertar o registo de divisão para a palavra R que venha logo a seguir, transfere-se $S(x)$ para um circuito divisor idêntico — o registo corrector —, temporizado sincronamente com os restantes, e leve-se o registo anterior a zero. A figura 9.7 ilustra a lógica de um circuito corrector de erros. Pode verificar-se que a função de correcção para o caso de erros simples é, como mostra a figura:

$$f(s_0, s_1, \dots, s_{n-k-1}) = \overline{s_0 \vee s_1 \vee s_2 \vee \dots \vee s_{n-k-1}} \quad (9.9)$$

9.4.5 Códigos BCH

A concepção óptima de códigos para controlo de erros envolve a procura de códigos com o menor comprimento possível de bloco, n , para um dado comprimento k do bloco de informação e para uma determinada distância mínima do código, d_{min} . Ou, para um dado comprimento do bloco n e valor do rendimento k/n , descobrir aqueles que possuem a maior distância mínima d_{min} . Isto é, normalmente deseja-se encontrar códigos com a maior capacidade de detecção ou de correcção de erros possível.

Os códigos BCH (Bose-Chaudhuri-Hocquenghem) constituem a sub-classe

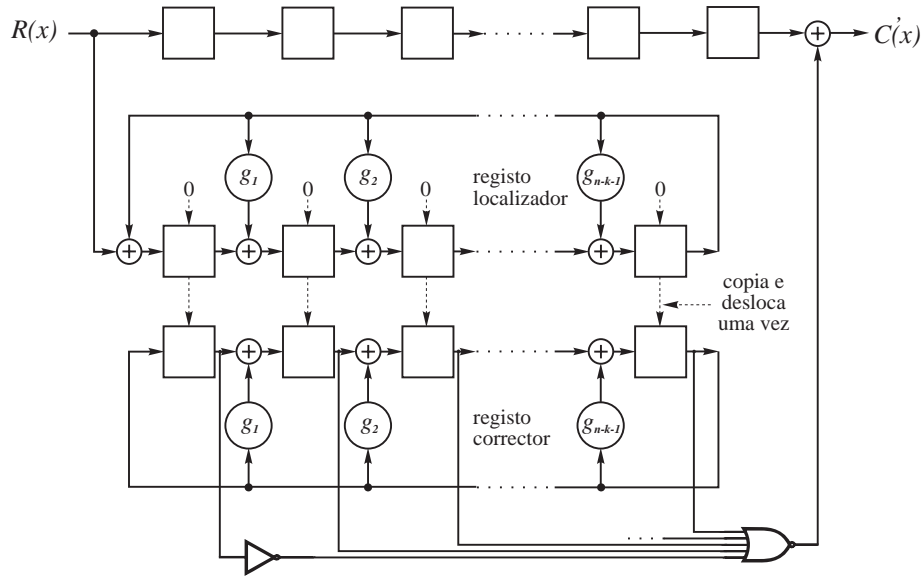


Figura 9.7: Circuito corrector de erros simples

mais poderosa dos códigos cíclicos correctores de erros. A sua análise detalhada, porém, exige tratamentos com álgebras em corpos finitos o que sai fora do âmbito deste curso. Enunciam-se, contudo, algumas propriedades dos códigos BCH que ilustram a sua potencialidade.

Assim, para quaisquer números inteiros positivos m e e_c (com $e_c < 2^{m-1}$) existe um código BCH com os seguintes parâmetros:

comprimento do bloco:	$n = 2^m - 1$
nº de dígitos de verificação:	$n - k \leq m e_c$
distância mínima:	$d_{min} \geq 2e_c + 1$

em que e_c é o número de erros que o código é capaz de corrigir.

Segundo o teorema 9.2, qualquer factor de $x^n + 1$ com grau $n - k$ pode servir de polinómio gerador de um código cíclico (n, k) mas não gera necessariamente um bom código. A tabela 9.3 lista os polinómios geradores de alguns códigos cíclicos que revelaram possuir as melhores características para serem usados como códigos correctores em sistemas de transmissão genéricos.

Tabela 9.3: Alguns bons códigos cíclicos

Tipo	n	k	ρ	d_{min}	$g(x)$
códigos de	7	4	0.57	3	$x^3 + x + 1$
Hamming	15	11	0.73	3	$x^4 + x + 1$
	31	26	0.84	3	$x^5 + x^2 + 1$
códigos	15	7	0.46	5	$x^8 + x^7 + x^6 + x^4 + 1$
BCH	31	21	0.68	5	$x^{10} + x^9 + x^8 + x^6 + x^5 + x^3 + 1$
	63	45	0.71	7	$x^{18} + x^{17} + x^{16} + x^{15} + x^9 + x^7 +$ $+x^6 + x^3 + x^2 + x + 1$
código	23	12	0.52	7	$x^{11} + x^9 + x^7 + x^6 + x^5 + x + 1$
Golay					

9.5 Técnicas de Correção de Erros

9.5.1 Correção Progressiva e ARQ

Quando um código para control de erros é utilizado directamente como corrector, o processo designa-se por *correção de erros progressiva* — *Forward Error Correction (FEC)*. É o único mecanismo utilizável quando só existe um canal num só sentido (comunicação simplex) e a retransmissão é impossível ou quando esta é impraticável. Como exemplos desta situação podem citar-se as transmissões por difusão nas quais existem múltiplos destinatários e as sondas espaciais que, pelo facto do tempo de propagação ser muito grande, utilizam essencialmente canais simplex.

Uma outra técnica, que é designada por *correção por pedido automático de repetição* — *Automatic Repeat reQuest (ARQ)*, exige um canal em sentido oposto — comunicação duplex —, que permita ao decodificador do canal pedir ao codificador a retransmissão de palavras (blocos) sempre que *detecte* a existência de erros. Nesta situação o código é utilizado apenas como detector de erros e a correção processa-se por repetição.

A técnica de correção progressiva é raramente utilizada nos sistemas de transmissão de dados visto funcionarem normalmente em duplex possibilitando a retransmissão e proporcionando reduções muito maiores da probabilidade global de erro com rendimentos substancialmente melhores. Ela é utilizada noutras aplicações, por exemplo, nas memórias de com-

putadores ou nas gravações digitais magnéticas ou ópticas.

Como termo de comparação com as técnicas ARQ, consideraremos neste capítulo apenas o caso da correcção progressiva.

As técnicas ARQ serão abordadas num próximo capítulo no contexto dos protocolos de comunicação de dados.

9.5.2 Probabilidade de Erro na Correcção Progressiva

Um código corrector é caracterizado por corrigir até e_c erros por palavra. Ocasionalmente corrigirá mais do que e_c erros. Portanto, uma palavra de n dígitos será incorrectamente decodificada, ou seja, estará normalmente errada quando ocorrerem $e_c + 1$ ou mais erros. A probabilidade, P_{pe} , de uma palavra errada será pois

$$P_{pe} \leq P(e_c + 1, n) + P(e_c + 2, n) + \dots + P(n, n) = \sum_{i=e_c+1}^n P(i, n)$$

em que $P(i, n)$ designa a probabilidade de existirem exactamente i erros numa palavra de n dígitos cujo valor, como se sabe, é dado pela distribuição binomial:

$$P(i, n) = C_i^n P_e^i (1 - P_e)^{n-i}$$

onde P_e é a probabilidade de erro por dígito no canal. Se se tiver $P_e \ll 1$, o que de facto se verifica nos sistemas reais, então $P(i + 1, n) \ll P(i, n)$, sendo aceitável tomar-se a aproximação

$$P_{pe} \approx P(e_c + 1, n) \approx C_{e_c+1}^n P_e^{e_c+1} \quad (9.10)$$

o que significa que uma palavra errada, tipicamente, possuirá $e_c + 1$ dígitos errados dos quais, em média,

$$\frac{k}{n}(e_c + 1)$$

são dígitos de informação e os restantes erros são nos dígitos de verificação.

Quando são transmitidas $N \gg 1$ palavras as quais contêm um total de kN dígitos de informação, espera-se receber NP_{pe} palavras erradas ou seja $\frac{k}{n}(e_c + 1)NP_{pe}$ dígitos de informação errados. Portanto, a probabilidade de se entregar ao destino um dígito de informação errado, isto é,

a probabilidade de erro por bit⁶, P_E , quando é utilizada a correcção de erros progressiva é

$$P_E = \frac{\frac{k}{n}(e_c + 1)NP_{pe}}{kN} = \frac{e_c + 1}{n} P_{pe}$$

ou seja,

$$\boxed{P_E = C_{e_c}^{n-1} P_e^{e_c+1}} \quad (9.11)$$

É importante notar que este resultado está de acordo com o Teorema de Shannon da codificação do canal, segundo o qual, uma melhoria (diminuição) da probabilidade de erro ($P_E < P_e$) e portanto menor equívocação e maior transferência de informação se faz à custa duma diminuição da velocidade de transmissão (débito) de informação, neste caso de um factor igual a $\rho = \frac{k}{n}$. O débito da fonte, r_s , e o débito no canal, r_c , estão pois relacionados por

$$\boxed{\rho = \frac{k}{n} = \frac{r_s}{r_c}} \quad (9.12)$$

9.6 Problemas

9.1 – Seja $C(x)$ um polinómio de grau $n - 1$ com coeficientes num corpo binário e $C^{(i)}(x)$ o polinómio resultante da rotação cíclica direita de grau $n - 1$ com coeficientes num corpo binário e $C^{(i)}(x)$ o polinómio resultante da rotação cíclica direita de i posições dos coeficientes de $C(x)$. Mostrar que o polinómio $C^{(i)}(x)$ é o resto da divisão de $x^i C(x)$ por $x^n + 1$, isto é, que

$$x^i C(x) = q(x) (x^n + 1) + C^{(i)}(x)$$

onde $q(x)$ é o polinómio quociente da divisão.

9.2 – Um código linear cíclico $(15, 5)$ tem como polinómio gerador:

$$g(x) = 1 + x + x^2 + x^4 + x^5 + x^8 + x^{10}$$

a) Esquematizar um codificador e um calculador de síndrome para este código.

⁶passaremos a designar o dígito binário por *bit* sempre que não haja confusão com a unidade de medida da informação

- b) Determinar o polinómio do código, quer sistemático quer criptográfico, correspondente ao polinómio de mensagem $D(x) = 1 + x^2 + x^4$.
- c) O polinómio $R(x) = 1 + x^4 + x^6 + x^8 + x^{14}$ pertence ao dicionário do código? Em caso negativo determine o respectivo síndrome.

9.3 – O polinómio gerador de um código cíclico $(15, 7)$ é

$$g(x) = 1 + x^4 + x^6 + x^7 + x^8$$

- a) Determinar a palavra do código sistemático para o polinómio de mensagem $D(x) = x^2 + x^3 + x^4$.
- b) Supôr que o primeiro e último dígitos do polinómio de código $C(x)$ para $D(x)$ sofreu um erro na transmissão. Determinar o síndrome de $R(x)$.

9.4 – Construir um código para correcção de erros simples com um comprimento de bloco da mensagem igual a 11 e mostrar, através de um exemplo, que ele pode efectivamente corrigir erros simples.

9.5 – Num código de repetição tripla, por cada dígito binário produzido pela fonte são enviados três. Por cada 0 é enviado 000 e por cada 1 é enviado 111.

- a) Mostre que este código é corrector de erros simples.
- b) Determine um polinómio gerador do código.
- c) Determine uma expressão para a probabilidade de erro por bit em função da probabilidade, P_e , de erro no canal.
- d) Compare este código com o de Hamming $(7,4)$.

9.6 – Mostre que um polinómio gerador de um código cíclico binário (n, k) tem sempre $g_{n-k} = g_0 = 1$ e que portanto só existem 2^{n-k-1} polinómios candidatos a geradores desse código.

9.7 – O código de Hamming $(7, 4)$ pode ser extendido de modo a formar um código $(8, 4)$ por adição de um quarto dígito de verificação escolhido de modo a tornar *par* a paridade de todas as palavras de código. Aplique este processo de extensão aos códigos da tabela 9.1 e mostre que $d_{min} = 4$. Determine o respectivo polinómio gerador.

fim do capítulo 9