



Universidade Federal de Pernambuco UFPE

24/09/2010

Prof. Hélio Magalhães de Oliveira, Docteur ENST

Contato hmo@ufpe.br <http://ufpe.academia.edu/hmdeoliveira>
visite/ <http://www2.ee.ufpe.br/codec/deOliveira.html>

Agradecimentos ao Prof. Alcione Alves, MSc

O primeiro dispositivo que permitiu a gravação do som e sua reprodução foi o **fonógrafo** de **Edison**, em 1877 (*Mary had a little lamb...*), no qual as ondas sonoras que chegavam a um diafragma faziam vibrar uma agulha, marcando uma ranhura de profundidade variada num **cilindro** que girava e que estava revestido com uma lâmina metálica.

O cilindro, que deveria ser girado a mão, funcionava com uma agulha unida a um diafragma.



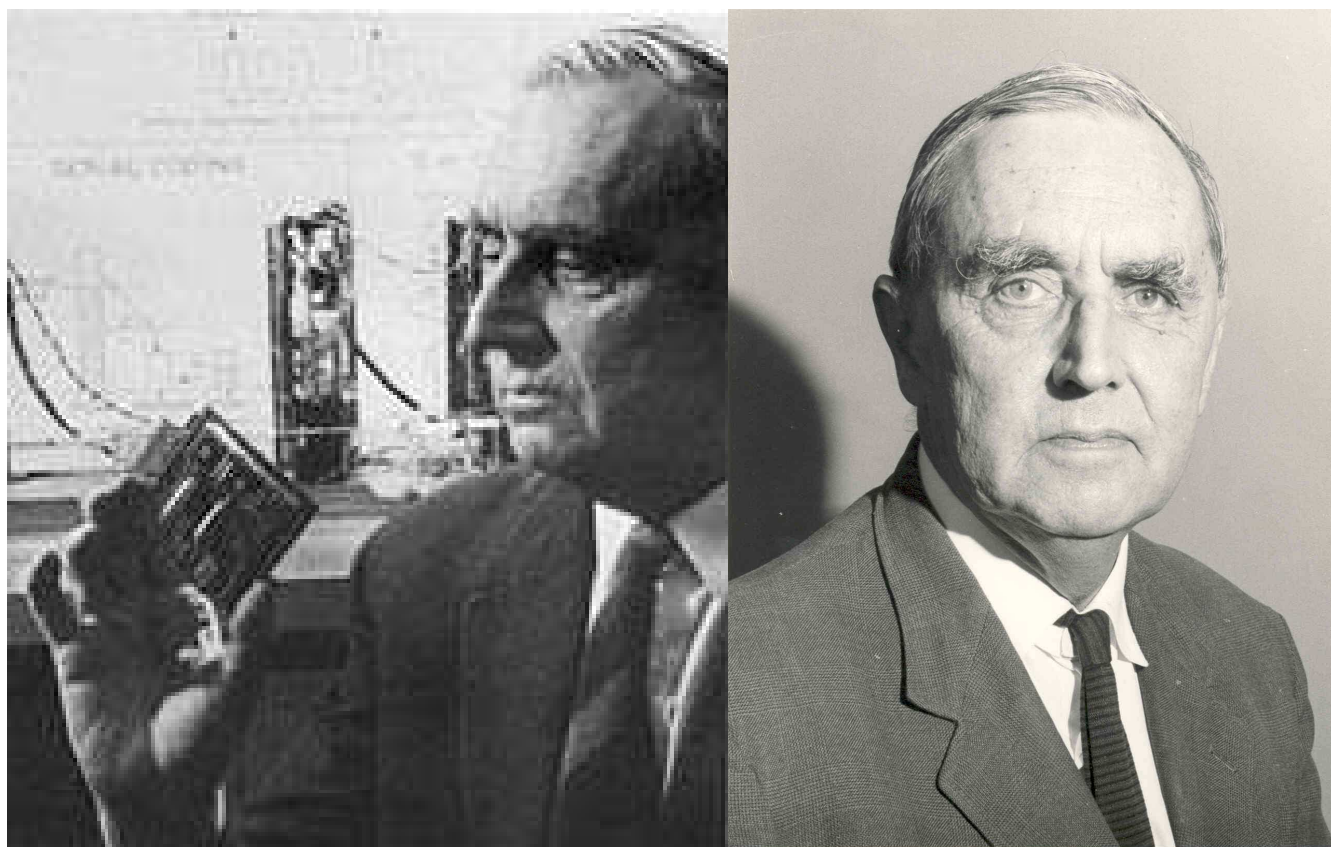
Em 1887, [Emile Berliner](#) desenvolveu o **gramofone**, que utilizava **discos** em vez de cilindros; a grande vantagem deste aparelho era que os discos podiam ser produzidos em massa a partir de um "modelo". 1908

LP

Em 1948, o disco "longa duração" (*long-play*, em inglês), melhorou a qualidade do som e estendeu o tempo de reprodução para mais de **20 minutos** (!) por lado.



33 rpm 45 rpm 78 rpm



Sir [Alec Reeves](#), patente 1937 PCM

Início da possibilidade de gravação digital:

o primeiro conversor A/D...

Tecnologia Digital: *Compact Disc* (Digitalização em CDs)

1983 "*CD Digital Audio Systems*" (Compact Disc)

- sinal de áudio com 44.100 amostras por segundo.(*)
- conversão A/D de 16 bits.

As amostras, cada com 32 bits (16 por canal, direito e esquerdo), são agrupadas em um quadro de 6 amostras.

* Nota histórica: *Karaoke*

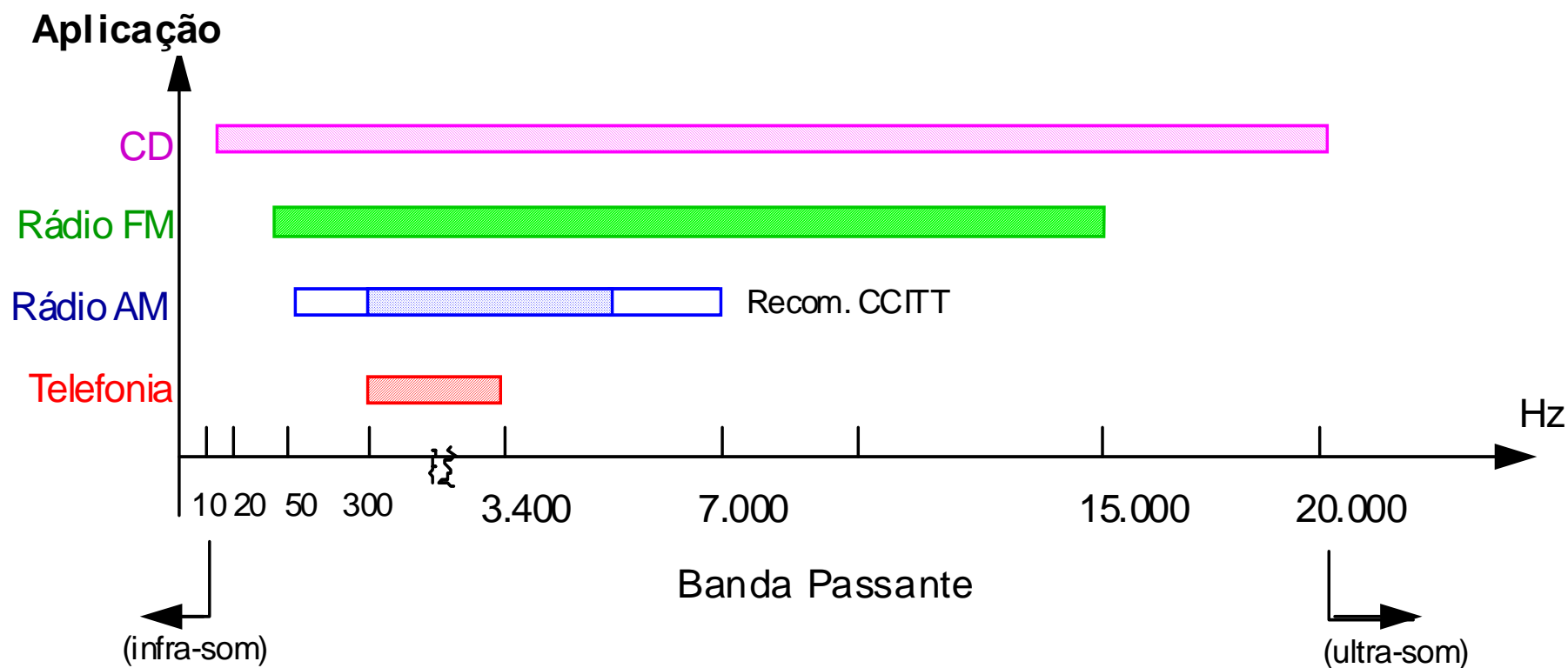


Figura - Requisitos de banda passante do sinal de áudio para diversas aplicações.

Uma disputa sugerira tempo de gravação (natural) de 60 minutos (1 hora de gravação) em um disco de diâmetro 100 mm (**Sony** 16-bit 44.100 amostras/s) ou 115 mm* (**Philips** 14-bit 44.000 amostras/s).

* escolhido com base em uma cartela de papelão de cerveja Heineken.

Este tamanho inicial, um pouco menor que o atual, foi substituído pelos engenheiros da Sony (na parceria que gerou os CDs), adotando um diâmetro capaz de armazenar integralmente a 9ª sinfonia de Beethoven tocada no *Bayreuth Festival* (**74 minutos**), por sugestão de **Von Karajan**.



Kornelis ("Kees") Antonie Schouhamer Immink 1979



Toshitada Doi (土井利忠), CD 1979

tb: Aibo

Asimo

Codificação concatenada de dois códigos de **Reed-Solomon**, reduzindo a probabilidade de erro por byte de 10^{-4} para 10^{-15} !

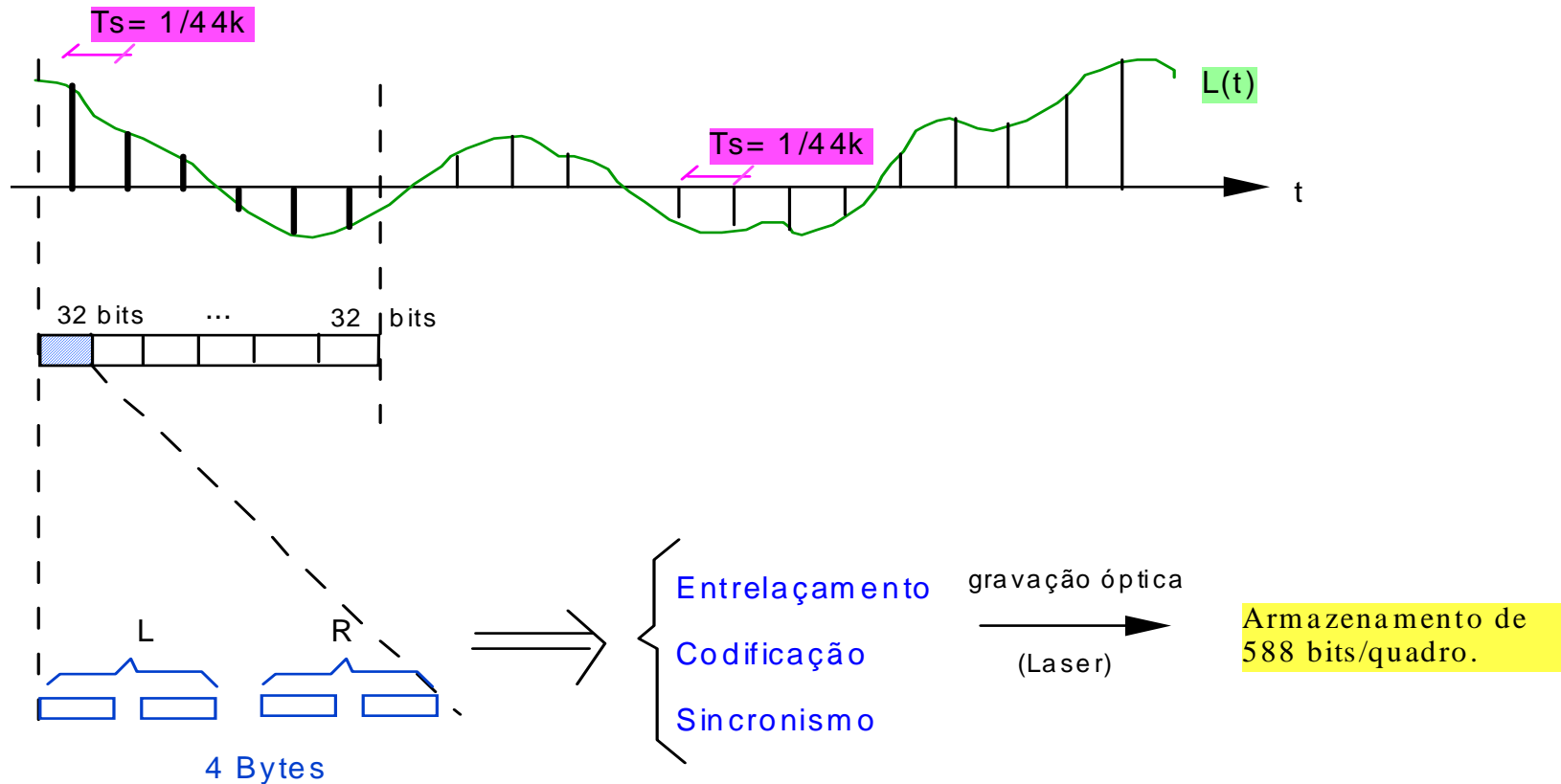


Figura- CD: Conversão A/D do áudio.

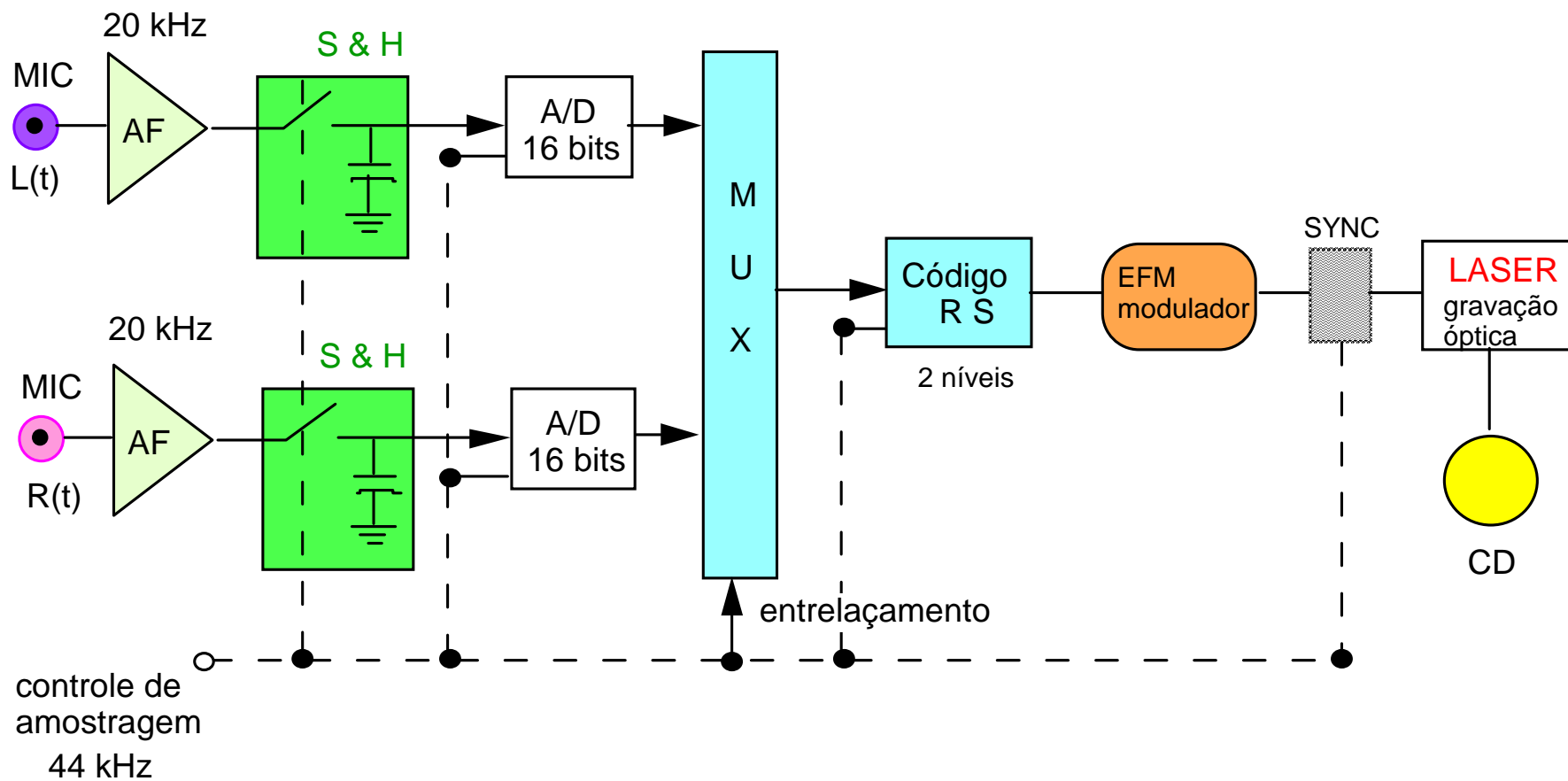


Figura - Esquema de áudio digital em CD (Gravação Digital).

SAMPLE & HOLD

O sistema colhe o valor das amostras de acordo com o teorema da amostragem e as retém por τ segundos. "circuito de amostragem e retenção" (*Sample & Hold*).

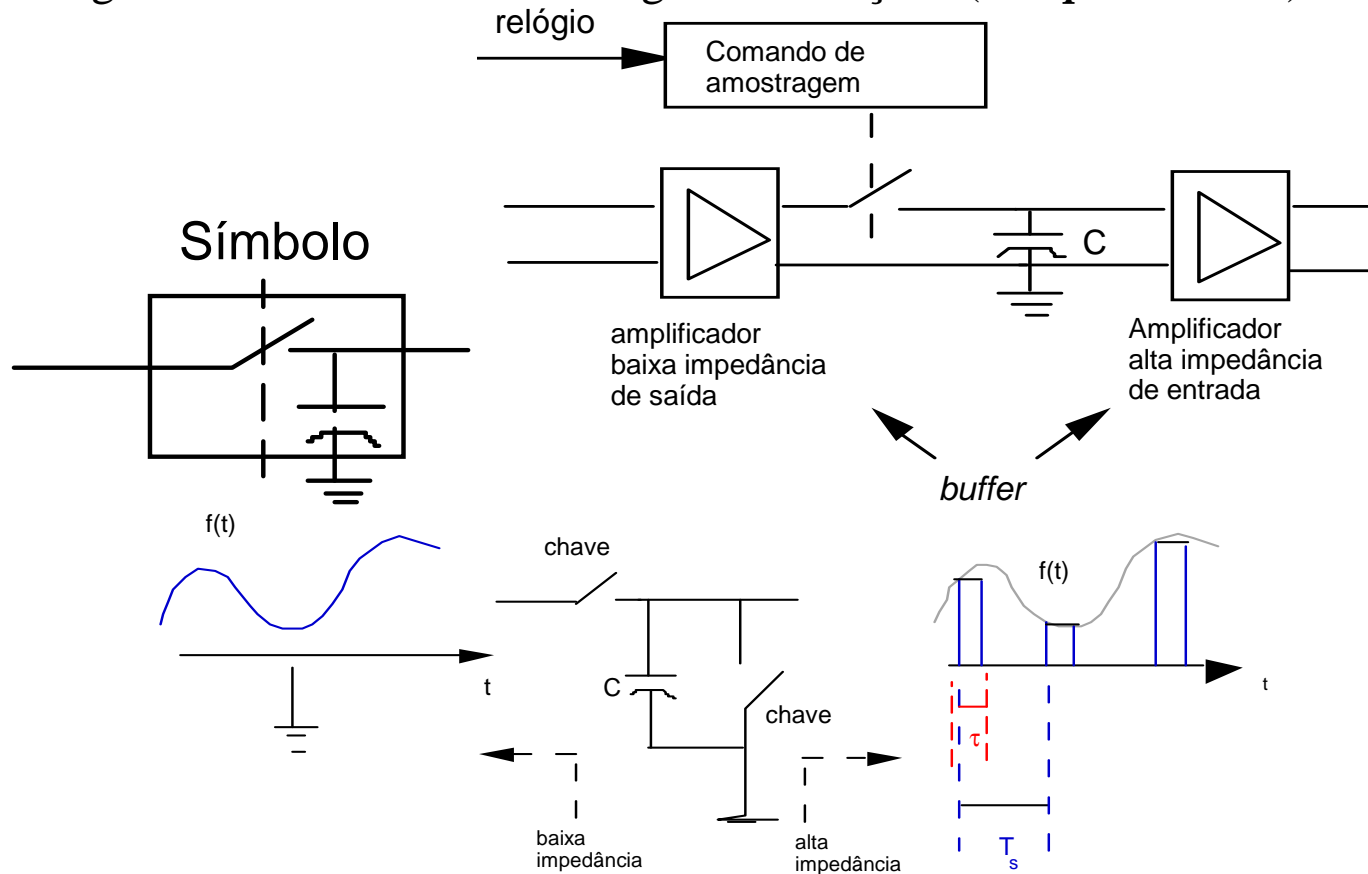


Figura - Sistema PAM Topo plano (circuito *sample & hold*).

Grande parte da superfície de leitura na qual ocorre a incidência do Laser é plana, sendo chamada de região de terra (*land*).

Certo número de baixos relevos na “paisagem”, chamados de poços (*pits*), também aparece no percurso da leitura.

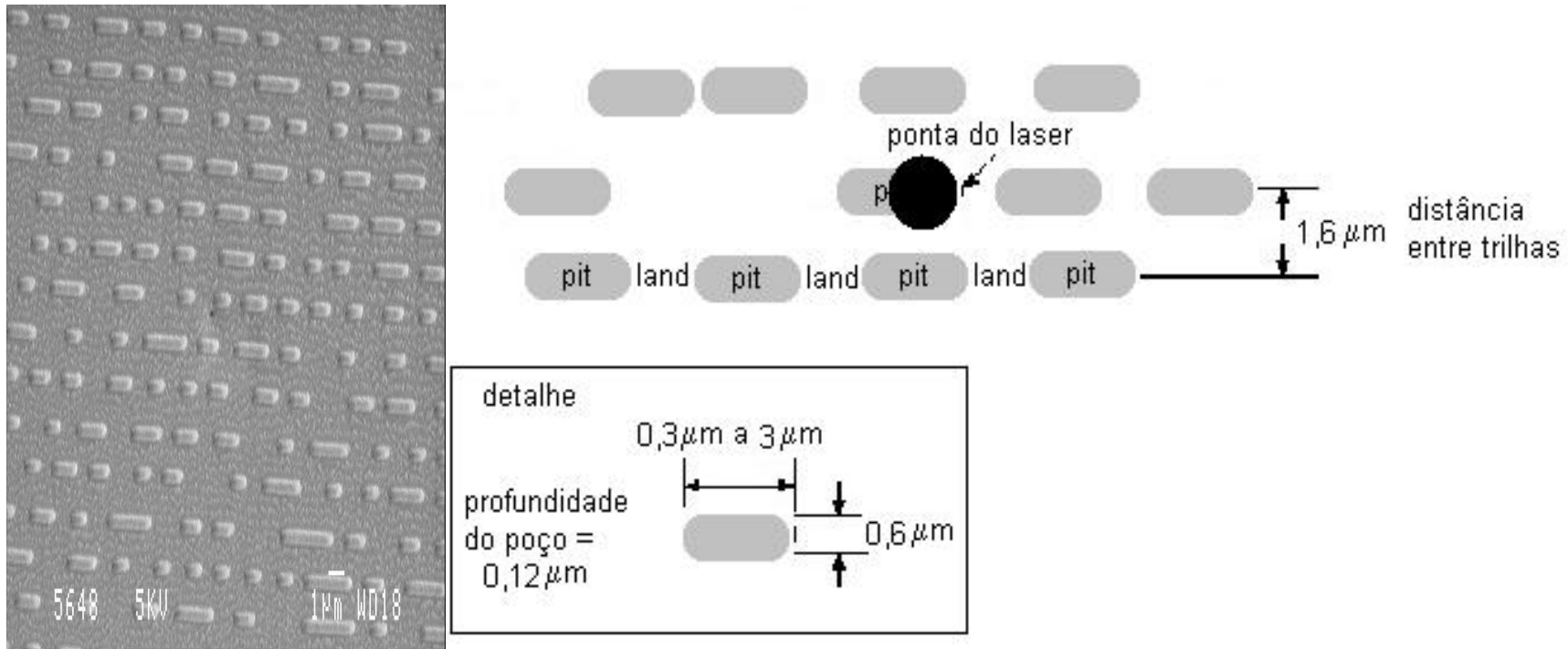
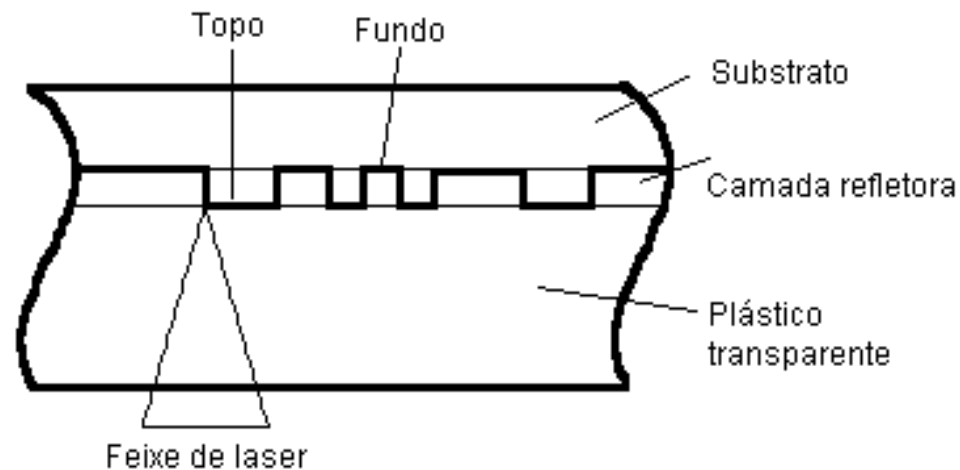


Figura – Detalhes da gravação em um CD: três trilhas são ilustradas. ponta do laser=circulo em preto.

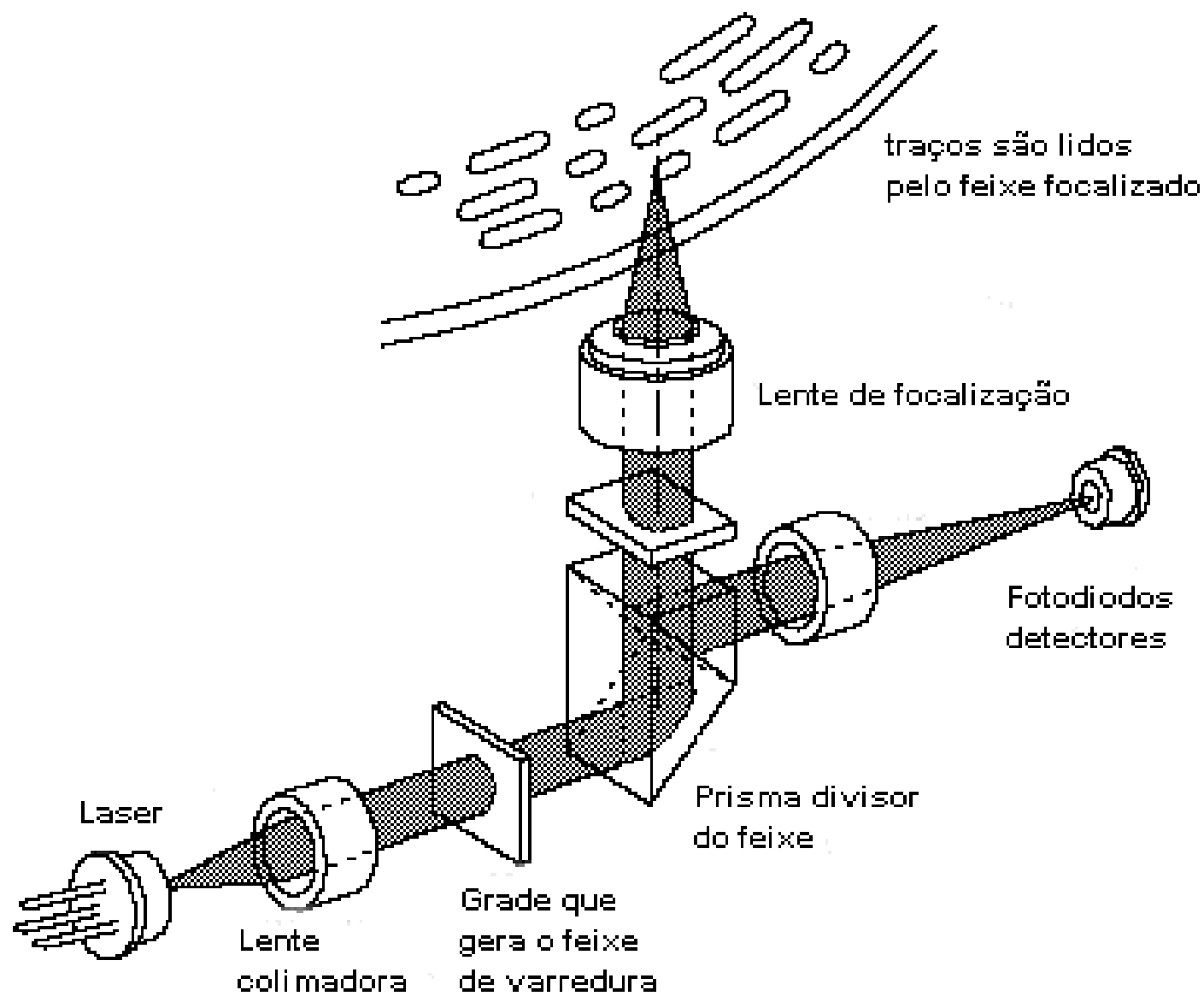
- feixe de laser incide “terra” => toda luz é refletida.
- feixe de laser ilumina “poço” => toda luz é eliminada.

A diferença de altura entre as duas partes é exatamente $\lambda/4$, e o feixe é eliminado por interferência entre o feixe refletido da superfície e o feixe refletido do poço.



Nos discos de vinil, uma agulha percorre sulcos, reproduzindo mecanicamente os sinais eletrônicos que os geraram.

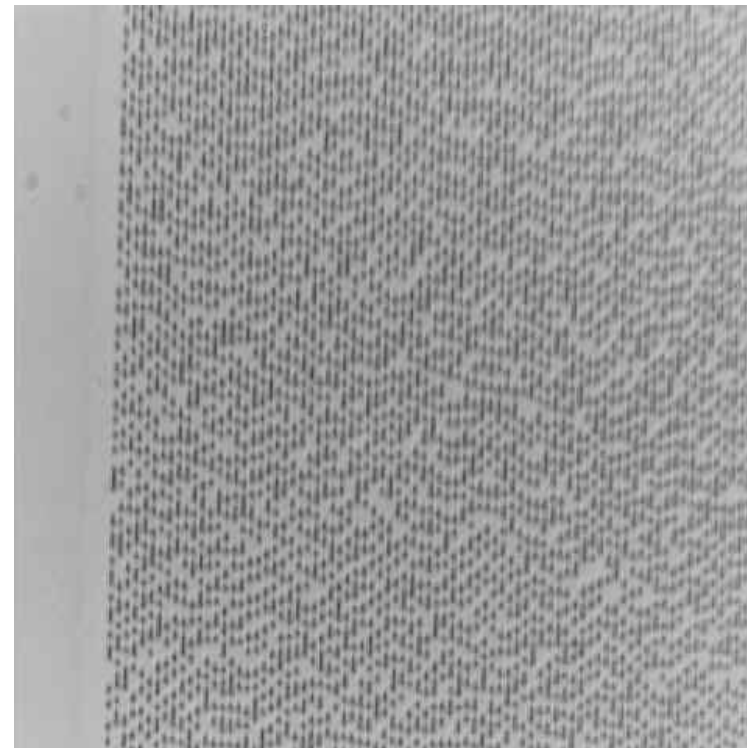
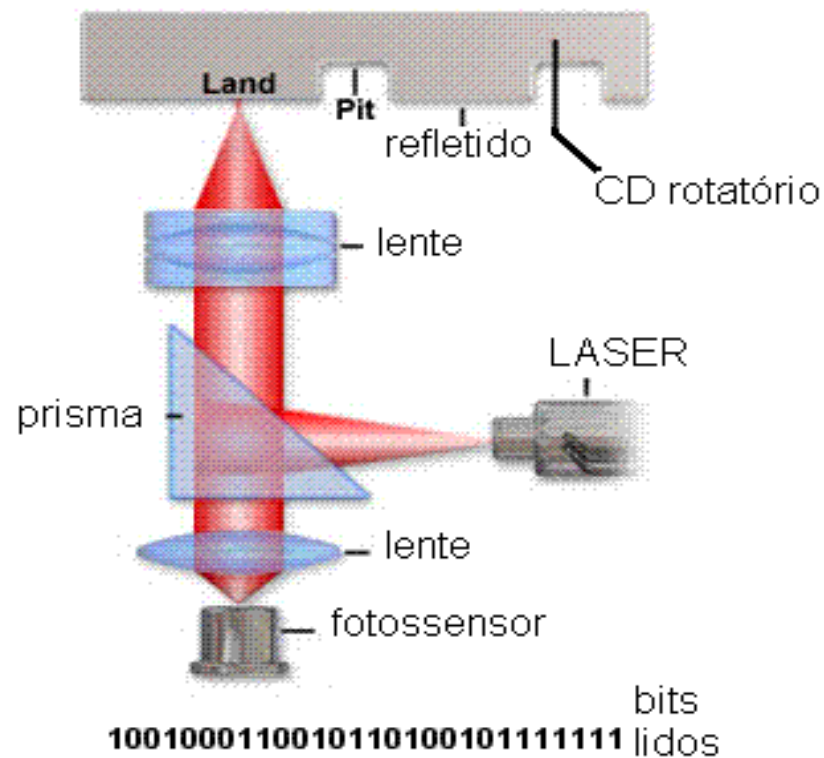
No CD, em vez de sulcos, existe uma seqüência de traços com um milésimo de largura e profundidade igual a um sexto dessa largura. **Não existe contato mecânico com esses traços:** a leitura é feita por um feixe de laser de 0,0009 mm. O feixe refletido (ou não) é dirigido a um conjunto de detectores.



Nas trilhas em espiral espaçadas de $1,6 \mu\text{m}$ no CD padrão 540 MB, o armazenamento dos dados é realizado através da seqüência de poço/terra (*pits/land*).

A velocidade de leitura padrão é constante e igual a $2,25 \text{ m/s}$ (CD×1, equivale a 150 kB/s).

Um CD×32 tem uma velocidade de leitura de 72 m/s (equivale a 4.800 kB/s).



Figura– Leitor de CD. Ao incidir no CD, o laser pode ser refletido (*land*) ou cancelado (*pit*). O sinal captado no fotossensor identifica se foi armazenado 0 ou 1.

<http://micro.magnet.fsu.edu/electromag/java/cd/> How CD works?

Codificação 8-14 para CDs

Kees A. Schouhamer_Immink

(Código EFM – *Eighth-to-Fourteen modulation*)

Para uma leitura adequada do laser (com diâmetro de 1 μm), não deve haver mais de dois 1's consecutivos em uma trilha gravada.

A solução encontrada foi adotar um código que realiza um mapeamento de palavras de um byte em palavras fixas de quatorze bits, as quais não contêm nenhuma “carreira de 1’s” sucessivos (chamado de EFM).

EFM pertence à classe dos códigos de comprimento de corrida *Run Length Limited* (RLL);

- O espectro (densidade espectral de potência) da seqüência codificada anula-se em baixa freqüência
- O valor mínimo e máximo do número de bits consecutivos de mesmo tipo está entre limites especificados.

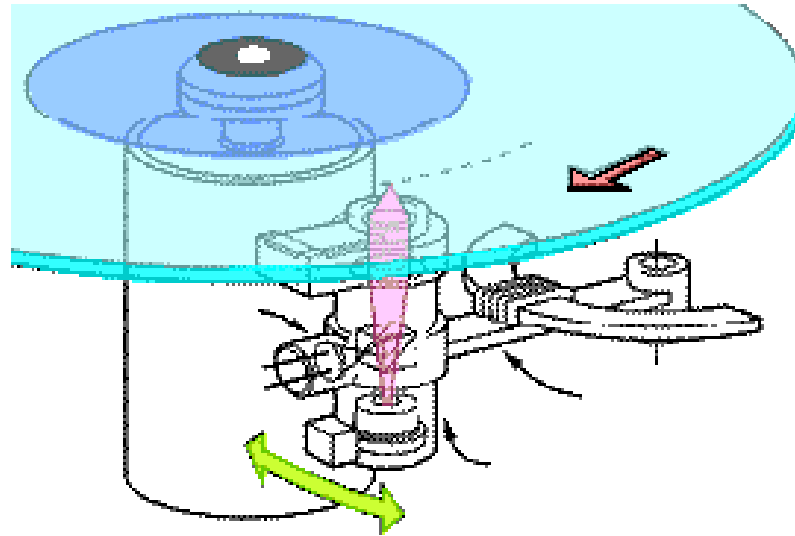
Devido a modulação (leia-se codificação de linha) EFM assegurar pelo menos dois zeros entre cada dois uns, ela garante que cada poço e terra acontece a cada três ciclos de relógio.

Esta propriedade é útil por reduzir a demanda do “driver” óptico usado no mecanismo de “reprodução”. O máximo de dez zeros consecutivos assegura o pior caso de recuperação do **sincronismo** do relógio.

Trecho da tabela (*look-up Table*) EMF

10	00001010	10010001000000
11	00001011	10001001000000
12	00001100	01000001000000
13	00001101	00000001000000
14	00001110	00010001000000
15	00001111	00100001000000
16	00010000	10000000100000
17	00010001	10000010000000
18	00010010	10010010000000
19	00010011	00100000100000
20	00010100	01000010000000
21	00010101	00000010000000
22	00010110	00010010000000
23	00010111	00100010000000
24	00011000	01001000010000
25	00011001	10000000010000
26	00011010	10010000010000
27	00011011	10001000010000
28	00011100	01000000010000
29	00011101	00001000010000
30	00011110	00010000010000

- 74 min 783,216,000



Exercício. Se a capacidade nominal de armazenamento de um CD de 4,8" (12 cm de diâmetro) é de 700 MB, quantos minutos de áudio estéreo podem ser gravados?

Solução. Cada amostra de áudio é convertida em 2 bytes
($2^{16}=65,536$ níveis de quantização).

44.100 amostras/canal/segundo \times 2 bytes/amostra=

88.200 bytes/canal/segundo,

Como a gravação é estereofônica, dois canais são usados.

$2 \times 5.292.000$ bytes/minuto.

Assim, dispõe-se de aproximadamente de $\left\lfloor \frac{783,216 \text{ M}}{2 \times 5,292 \text{ M}} \right\rfloor \approx \mathbf{74 \text{ minutos}}$

(CD). [A nona de Beethoven & Von Karajan atacam... listen](#)

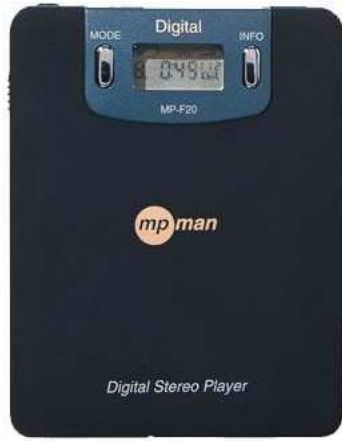
[Symphony No. 9 in D minor, Op. 125, 'Choral': Finale: Presto](#)

A inserção de um *overhead* para controle de erros reduz o espaço útil em um CD. O rendimento é aproximadamente 87%. Por exemplo, para o CD áudio de 74 minutos, tem-se aproximadamente 682 MB úteis.

São 75 blocos/segundo (4.500 blocos/minuto), totalizando 4.500 blocos/min 74 min=333 kblocos.

mp3 (áudio)

Estudos de percepção humana para sinais audíveis permitiram estabelecer os chamados “modelos psico-acústicos”. De fato, o sistema auditório não responde uniformemente na faixa audível (este fato pode ser explorado para descartar informações menos significantes ao ouvido).



1º mp3 32MB 8 músicas- Coreia do Sul 1998
SaeHan Information Systems



mp3 2GB 2004

Uma exploração dos limites de audibilidade mostra limiares nos quais o som pode ser ouvido e estes são dependentes da frequência.

A medição destes limites resulta em uma curva de limiares absolutos de escuta. O pico de sensibilidade ocorre entre 1 kHz e 5 kHz (drasticamente reduzido com a idade).

As curvas de contorno de sons de igual intensidade (*Equal-loudness contours*) (**Fletcher & Munson**).

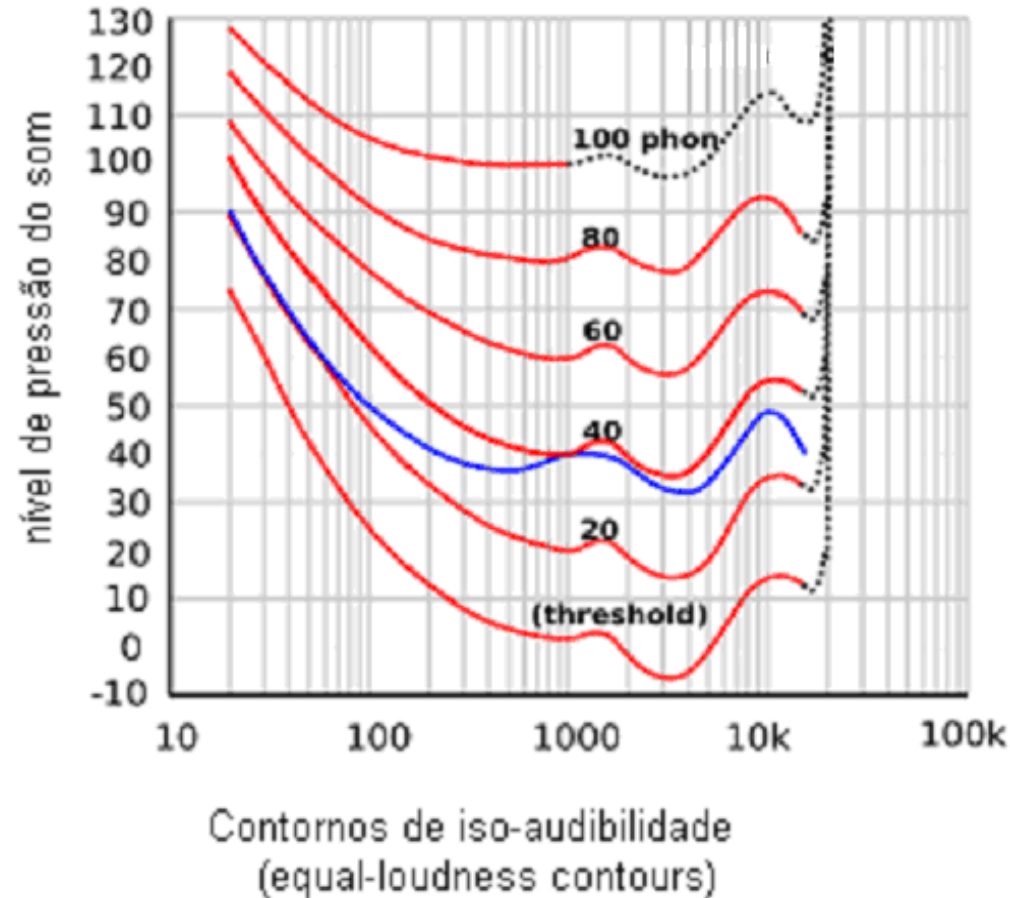


Figura - Contornos de iso-audibilidade: de 0 fon a 100 fon.

Os codificadores mp3 criados por engenheiros europeus:



Gesellschaft 58 Intitutos de Pesquisa !

- **Fraunhofer Society** (*Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung, Erlangen*)
- **Philips**
- **CCETT** (*Centre commun d'études de télévision et télécommunications*)

1. **PCM** procura reproduzir a forma de onda do áudio
2. **MP3** reproduz uma forma de onda que soa idêntica ao ouvido.

limiars de mascaramento: o ouvido humano não é capaz de perceber frequências fracas após frequências fortes. Em termos técnicos, isto é referido como “efeito de mascaramento”.

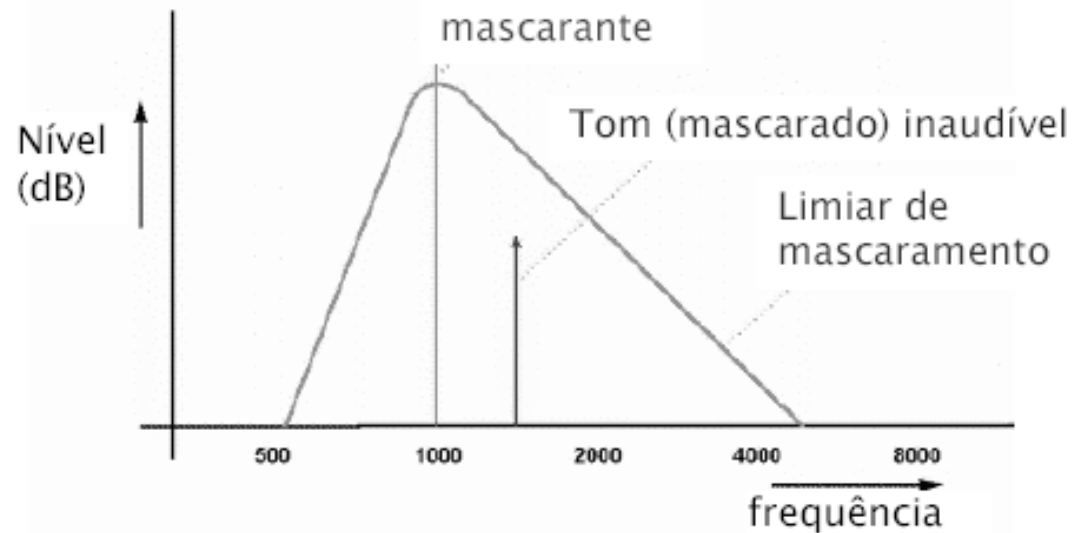
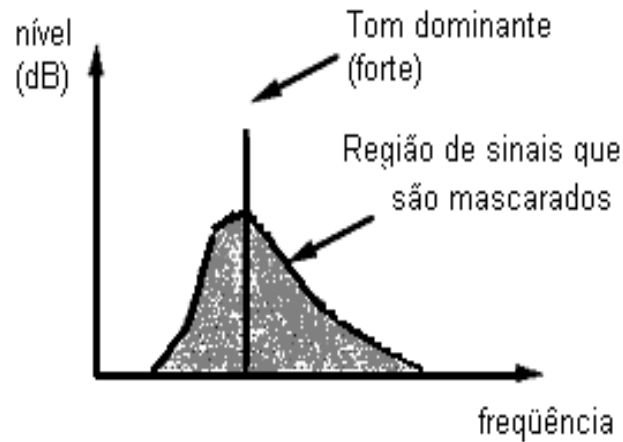


Figura - Mascaramento de áudio.

Dentro de cada sub-banda a presença de um tom dominante pode mascarar uma região de sinais mais fracos.

Uma família de padrões de compressão de áudio baseados em “codificação perceptual” é o **MPEG** (*Motion Picture Experts Group*, formado pela ISO em Janeiro de 1988 para criar **codecs de áudio e vídeo**).

Em particular, o sistema MPEG-1 é um padrão de compressão para áudio, estéreo ou mono, aprovado como padrão internacional em 1992.

Há três tipos de sistemas, com complexidade variável:

	Camada 1
MPEG-1	Camada 2
	Camada 3

O mais sofisticado deles, o sistema MPG-1 camada III é conhecido mundialmente com formato de compressão mp3 (nome simplificado).

Em MPEG (mantendo a qualidade de CD), tem-se redução de:

- **1:4** para **camada 1** (384 kbps para sinal estéreo)
- **1:6** para **camada 2** (256..192 kbps para sinal estéreo)
- **1:10** para **camada 3** (128..320 kbps para sinal estéreo)

Um arquivo MP3 criado em taxas 128 kbit/s resulta em um fator de compressão é cerca de 1:10 com relação aos arquivos de áudio comuns em CD.

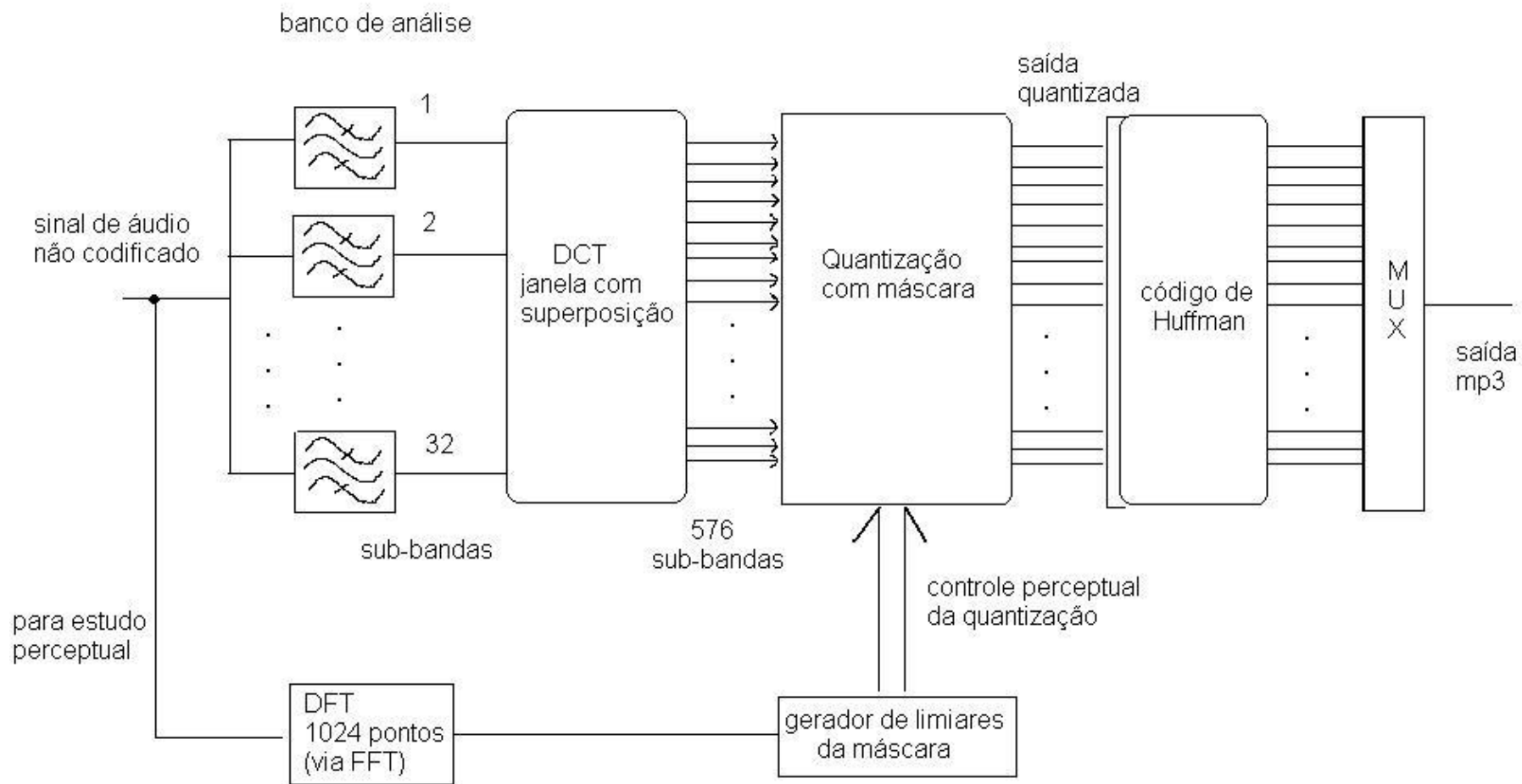


Figura – Codificador mp3.

A informação espectral para análise psico-acústica **via FFT** de 1.024 pontos.

Este espectro é usado para determinar os limiares das máscaras, dependendo da banda de frequência, intensidade e tonalidade.

O sinal é analisado em **32 sub-bandas** e usa-se uma **DCT** (transformada discreta do cosseno) para obter melhor compressão.

DCT com “janelamento” de blocos superpostos para cada sub-banda analisada, gerando 18 sub-bandas.

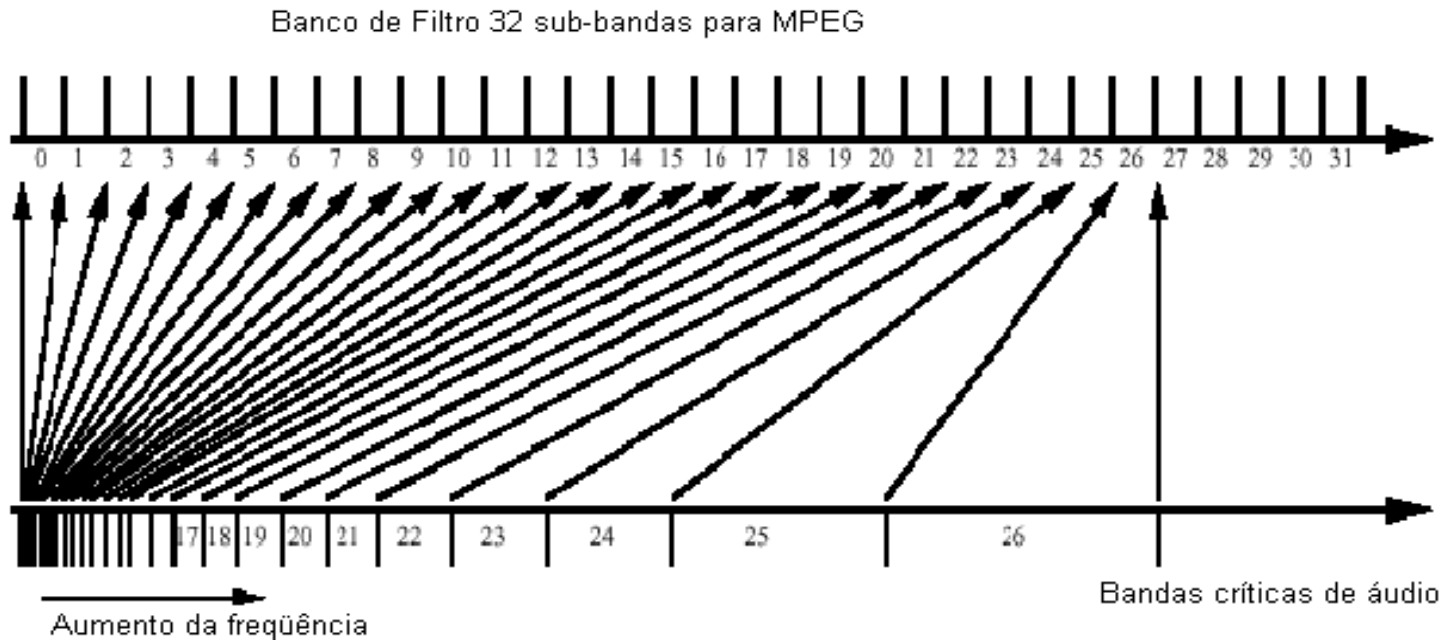


Figura - Banco de filtros para subdividir em 32 sub-bandas.

As bandas críticas vão de 100 Hz a mais de 4 kHz.

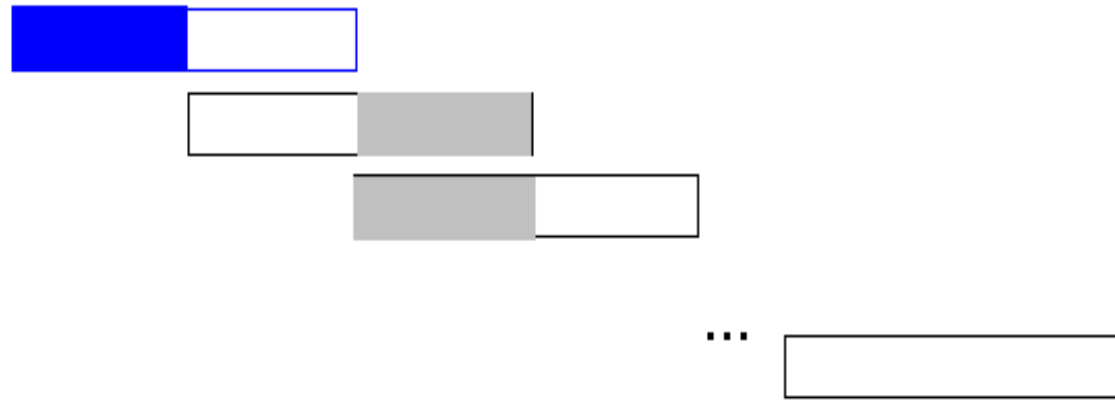


Figura - Blocos usados na DCT com superposição em janelas deslizantes. A primeira metade do bloco coincide com a segunda metade do bloco precedente.

Muitos dos coeficientes da DCT são desprezíveis e são “assassinados” (compressão com perdas controladas).

O espectro da DCT é quantizado com máscaras de áudio e o resultado é comprimido com codificador de [Huffman](#).

$$F[k] := \frac{2c[k]}{\sqrt{N}} \sum_{n=0}^{N-1} f[n] \cos\left(\frac{(2n+1)\pi kn}{2N}\right), \quad k=0,1,\dots, N-1,$$

em que $c[k] := \begin{cases} 1 & k = 0 \\ \sqrt{2} & k \neq 0. \end{cases}$ *Ainda transformadas!*

32 sub-bandas de áudio \times 18 sub-bandas DCT por sub-banda de áudio =

Divididos em 1152 amostras/bloco.

Cabeçalho (4 bytes iniciais de um quadro) –contém:
Sincronismo de quadro, camada MPEG, Frequência de amostragem, número de canais, CRC, etc.

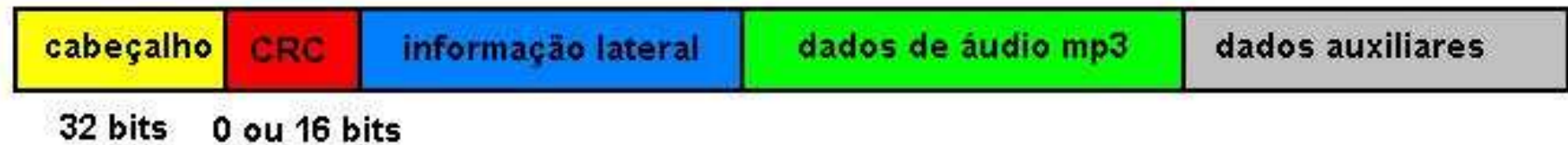


Figura – Quadro de áudio mp3.

O cabeçalho tem 32 bits. Contém palavra de sincronismo e um descritor do quadro.

- *Sync* (12 bits). A palavra de sincronismo é '1111 1111 1111'.
- *Id* (1 bit) especifica a versão do MPEG (1 se MPEG-1, 0 se MPEG-2).
- Especifica a *camada* (2 bits)
 - 00 reservado
 - 01 camada 3
 - 10 camada 2
 - 11 camada 1
- *Bit de Proteção* (1 bit)
 - (1 se usa CRC, 0 caso contrário)
- *Taxa de codificação* em kbps (4 bits)

	MPEG-1			MPEG-2		
Bits	camada I	camada II	camada III	camada I	camada II	camada III
0 0 0 0						
0 0 0 1	32	32	32	32	32	8
0 0 1 0	64	48	40	64	48	16
0 0 1 1	96	56	48	96	56	24
0 1 0 0	128	64	56	128	64	32
0 1 0 1	160	80	64	160	80	64
0 1 1 0	192	96	80	192	96	80
0 1 1 1	224	112	96	224	112	56
1 0 0 0	256	128	112	256	128	64
1 0 0 1	288	160	128	288	160	128
1 0 1 0	320	192	160	320	192	160
1 0 1 1	352	224	192	352	224	112
1 1 0 0	384	256	224	384	256	128
1 1 0 1	416	320	256	416	320	256
1 1 1 0	448	384	320	448	384	320
1 1 1 1						

- Frequência de amostragem (2 bits)

bits	MPEG1	MPEG2	MPEG2.5
00	44.100	22.050	11.025
01	48.000	24.000	12.000
10	32.000	16.000	8.000
11			

- *Padding* bit (1 bit)

Ajuste: no caso de taxa 128 kbps e frequência de amostragem 44,1 kHz, os quadros terão 417 bytes os invés de 418 bytes.

- bit privado (1 bit)

- Modo de gravação (2 bits)

00 estéreo
01 estéreo conjunto
10 dual estéreo
11 mono

- Extensão (2 bits)

Usados somente quando selecionado estéreo conjunto

- Bit de *copyright* (1 bit): 1 indica presença de *copyright*, 0 significa áudio livre

- *Home* (1 bit)
- Ênfase Dolby (2 bits)
 - 00 nenhum
 - 01 50/15 ms
 - 10 reservado
 - 11 CCITT J.16

O **CRC** possui 0 bytes ou 16 bytes, dependendo do bit de proteção.

A difusão dos avanços tecnológicos é comparável ao alastramento de um incêndio: Em primeiro, surge uma faísca, logo brota uma chama tremula; segue-se então um fogaréu, avançando com rapidez e pujança cada vez maiores. Nikola Tesla.

Agradeço a atenção de todos. Espero ter fornecido uma visão panorâmica dos mecanismos de gravação de áudio!

Disponível @ http://www2.ee.ufpe.br/codec/Nassau24_09_10.pdf