Komprese zvuku

Motivace

Zvuk v CD kvalitě:

- 2 * 16 * 44100 = 1411,2 kb/s
- cca 10 MB/min

Proč komprese:

- Omezená propustnost datových linek (např. ADSL 128kb/s)
- Archivace záznamů
- …

Komprese/kódování zvuku

Požadavky:

- šířka pásma (např. 300 – 3400 Hz řeč, 20 – 20000 Hz hudba)
- kvalita
- možnost kombinování (multiplexování) s video daty (MPEG)
- malé zpoždění komprese/dekomrese

Komprese/kódování zvuku

Možné způsoby komprese:

- časová oblast (PCM→non-linear PCM [*mu-law* encoding, .au, SUN])
- frekvenční oblast (sub-band coding→MPEG, cca. 2 bity na vzorek)
- (modely založené na událostech - MIDI)

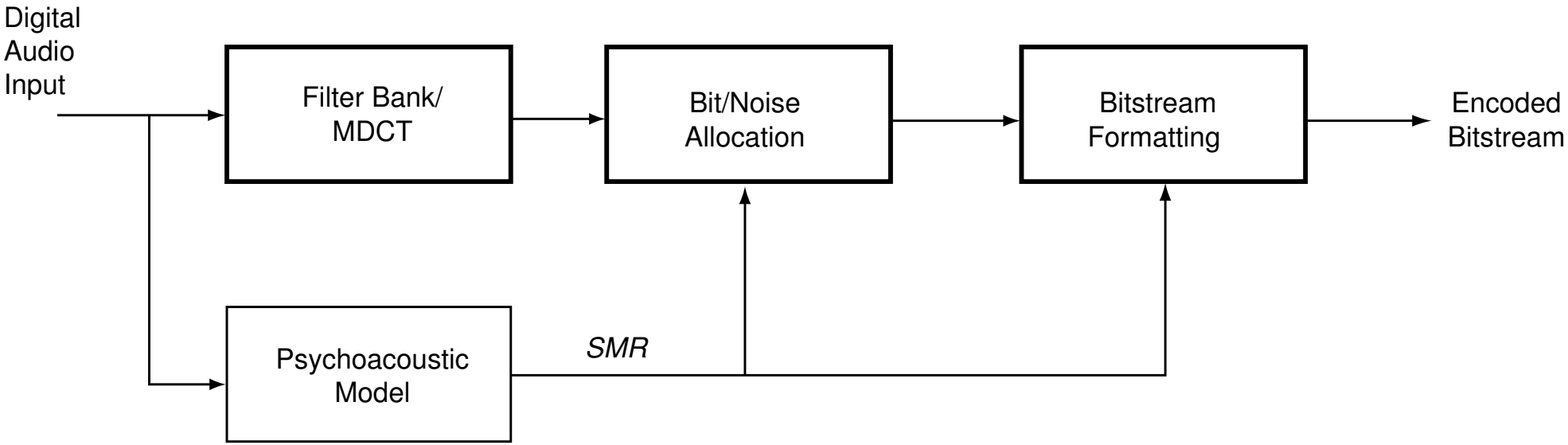
Získání vyšší komprese → psychoakustický model, maskování

Maskování:

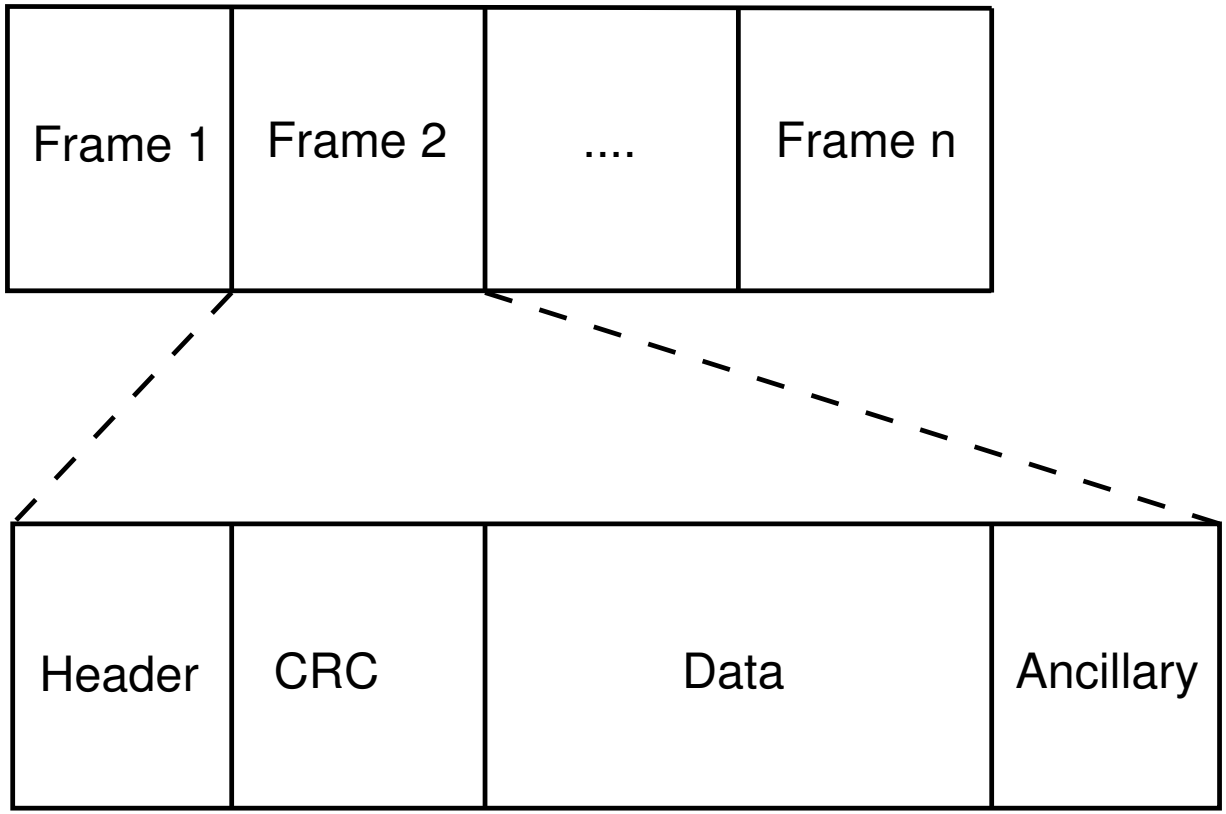
- frekvenční
- temporální

MPEG1 audio komprese

- Založena na psychoakustickém modelu
- Alokace/přiřazování bitů vzorkům na základě odstupu signálu od masky/prahu (SMR=signal-to-mask ratio)

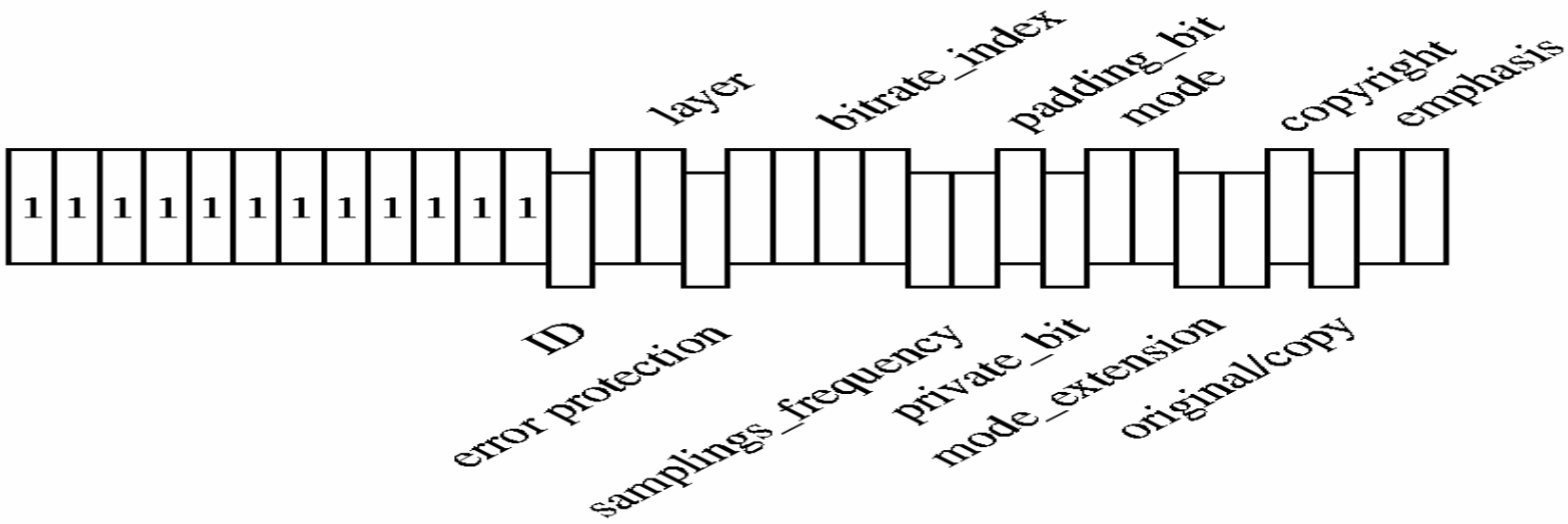


Struktura bitstreamu



Hlavička (header) obsahuje (mimo jiné):

- typ vrstvy (LayerI, II, III) [2bity]
- bit-rate index [4bity]
- vzorkovací frekvenci [2bity]
- vyplňovací bity (definuje počet „slotů“, na který je rámec zarovnán)
- mód [2bity]: stereo, joint-stereo, dual/single channel



Tři vrstvy:

- Layer I
  - základní
  - 32 frekvenčních pásem (neodpovídá vlastnostem lidského ucha, nižší frekvence – užší pásma, vyšší frekvence – širší)
  - každé pásmo má 12 vzorků (32*12=384 samples)
- Layer II
  - rámec obsahuje 3*12*32=1152 vzorků
  - definovány až 3 rozsahy/měřítka (scale factors) pro každé pásmo
- Layer III
  - frekvenční filtrace (sub-band filtering) + MDCT (MDCT přidána pro zvýšení frekvenčního rozlišení)
  - neuniformní kvantování
  - Huffmanovo kódování

Table 10.18 Bit rates available in MPEG-1 audio coding [487]

	Bit rate (Kbps)		
Index	Layer I	Layer II	Layer III
0000	free format	free format	free format
0001	32	32	32
0010	64	48	40
0011	96	56	48
0100	128	64	56
0101	160	80	64
0110	192	96	80
0111	224	112	96
1000	256	128	112
1001	288	160	128
1010	320	192	160
1011	352	224	192
1100	384	256	224
1101	416	320	256
1110	448	384	320

Source: © 1993 ISO/IEC.

Joint-stereo coding

- zvýšení komprese ≈ snížení redundance ve stereo signálu

• **intensity stereo coding**

- kódován je součtový signál společně s poměrem pro levý a pravý kanál
- frekv. spektrum zachováno, amplituda odlišná

• **MS stereo coding**

- kódováno M=L'+R, S=L-R

$$L = (M + S) / \sqrt{2}$$

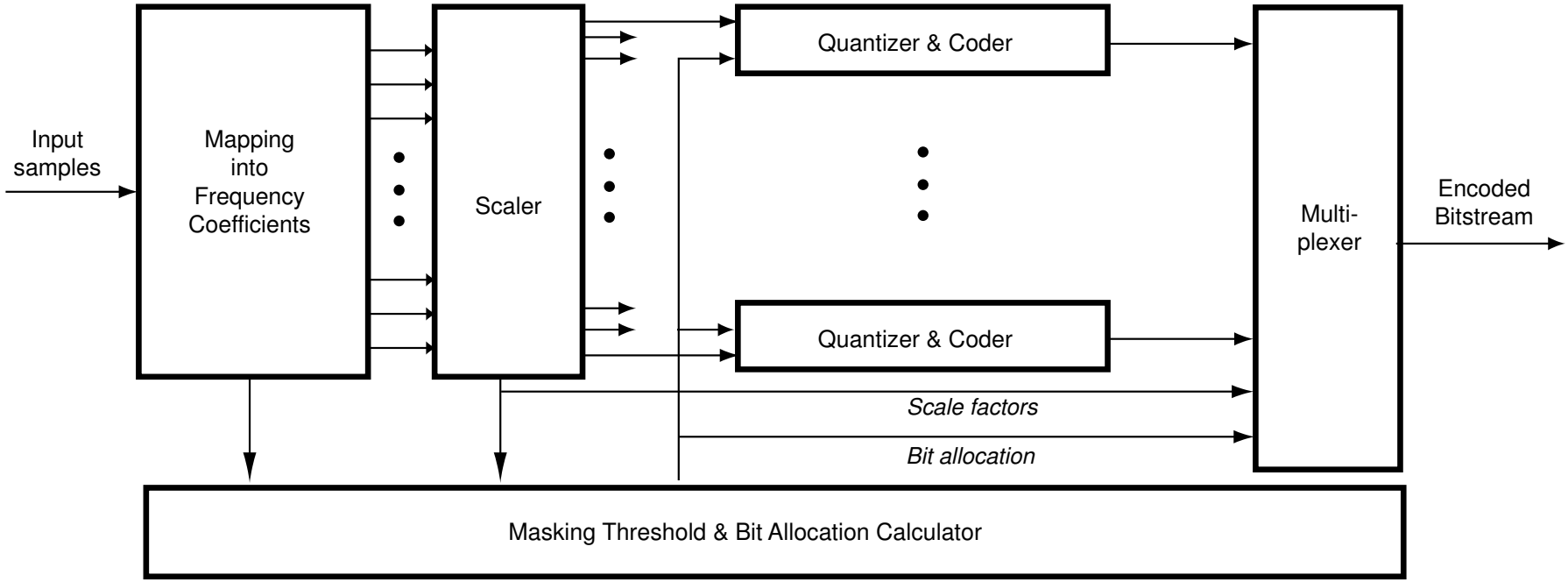
$$R = (M - S) / \sqrt{2}$$

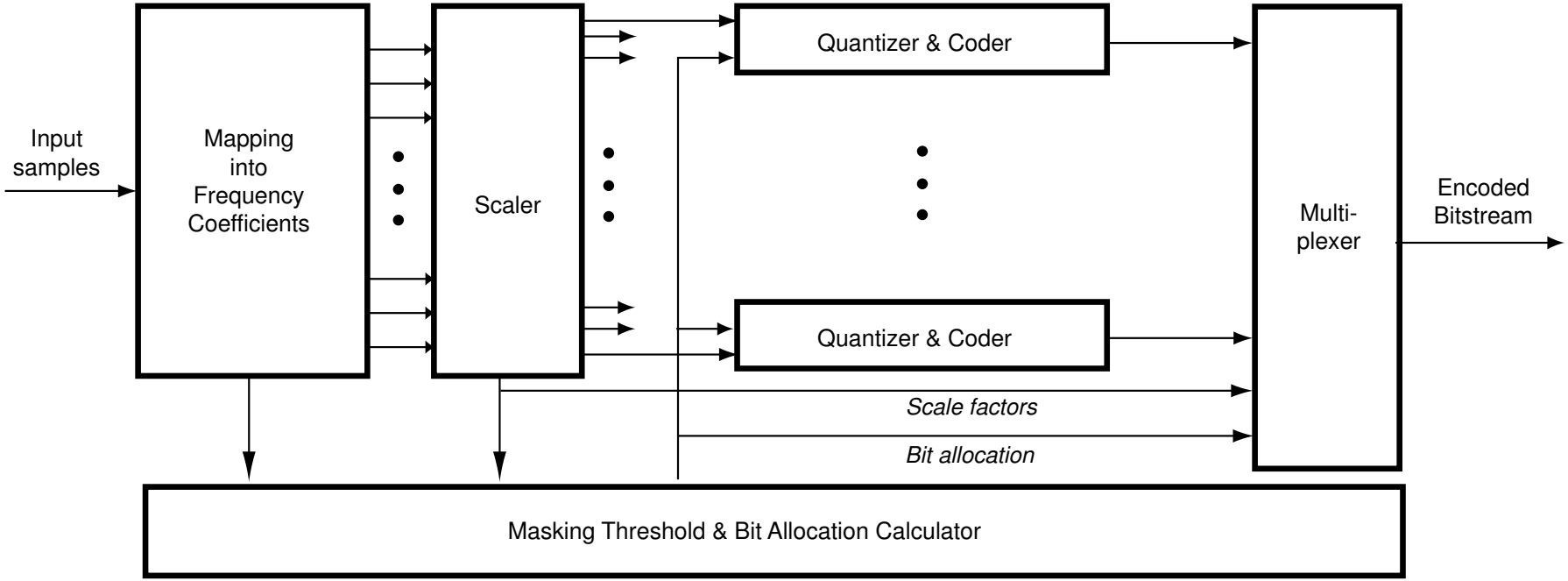
Algoritmy kódování

- **ASPEC** - použití MDCT pro převod do frekv. oblasti
- **MUSICAM** - sub-band filtering

Oba používají psychoakustický model pro řízení přiřazování bitů (*bit-allocation*) a kvantování.

Společná část pro ASPEC i MUSICAM





- převod do frekvenční oblasti
- normalizace (rozsah určen pomocí psychoakustického modelu)
- kvantování a kódování
- vytvoření rámce

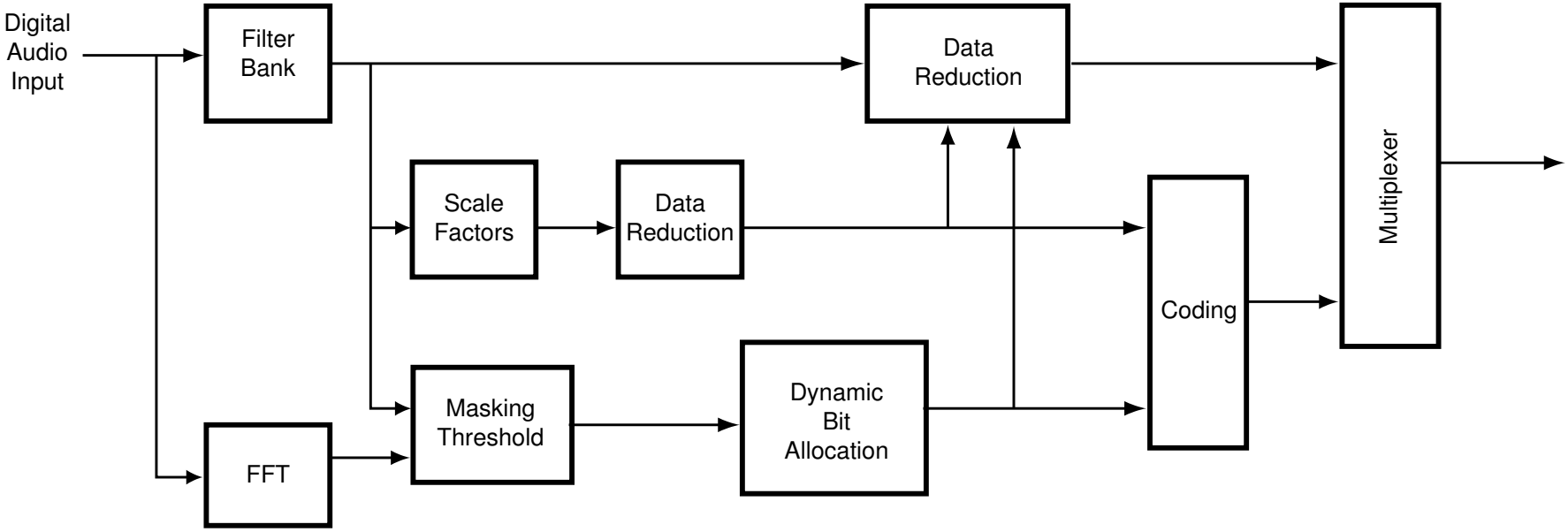
MUSICAM

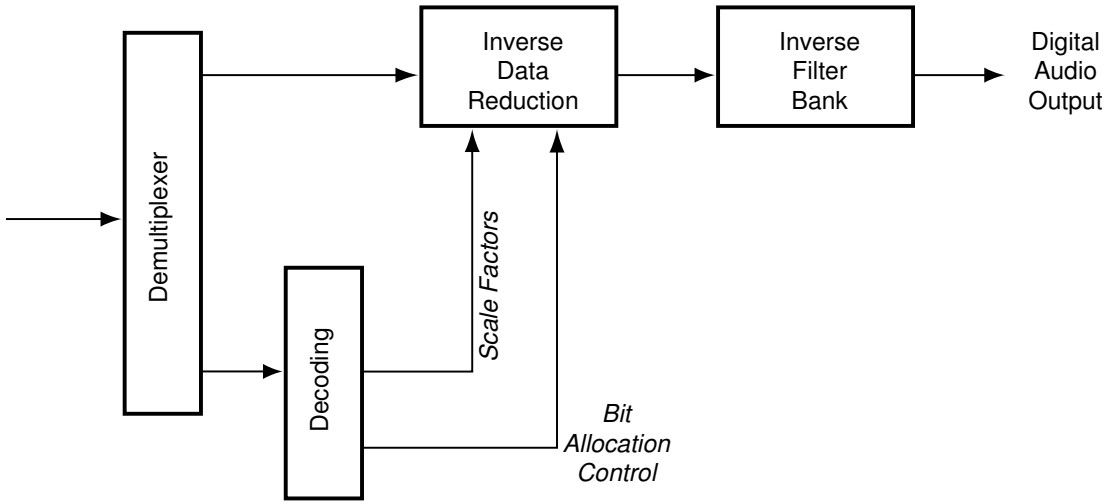
- signál rozdělen do 32 pásem
- každé pásmo 12 vzorků
- FFT k nalezení maskovacích frekvencí
- alokace bitů pro kvantování na základě minimalizace odstupu šumu od maskovacího prahu (NMR=noise-to-mask ratio)
- výpočet rozsahů (scale factors), adaptivní kvantování (vzorky v rozsahu [-1,1])
- informace o počtu přidělených bitů a rozsazích je kódována a přidána k rámci

ASPEC

- MDCT pro převod do frekvenční oblasti $2N_t \rightarrow N_f$
- definovány dvě délky bloků: 256 vzorků pro 96, 128kbps, 512 vzorků pro 32, 64kbps
- snižování aliasingu: TDAC (Windowing, Time Domain Aliasing Cancellation)

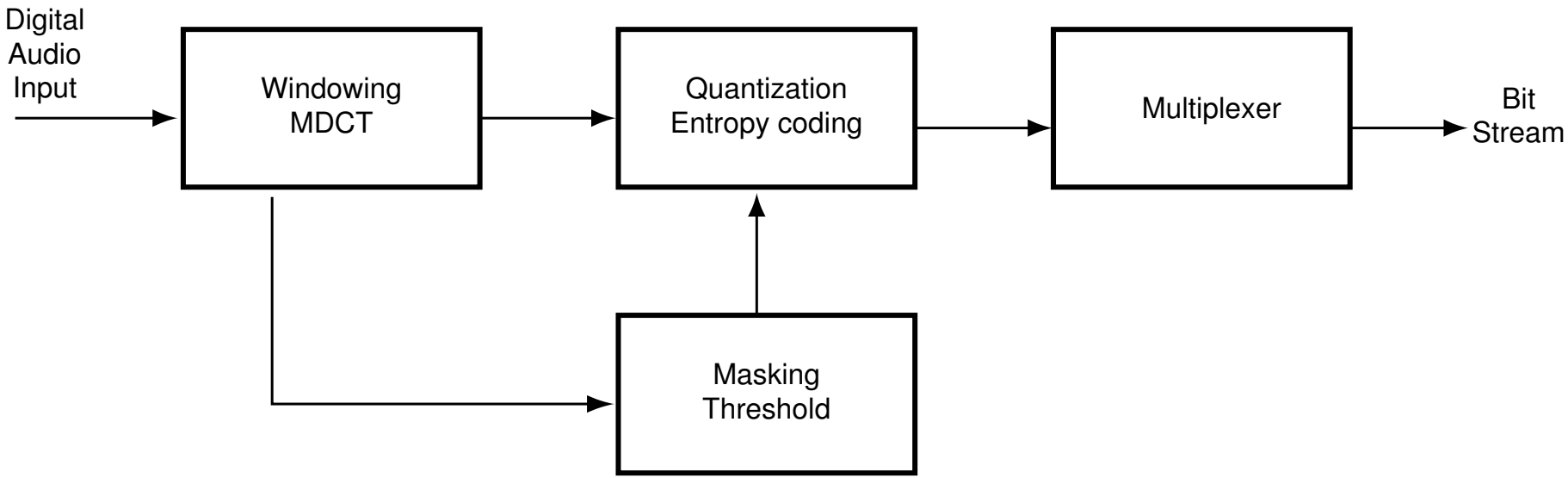
MUSICAM





Inverse
Data
Reduction

ASPEC


Digital
Audio
Input
Windowing
MDCT
Quantization
Entropy coding
Multiplexer
Bit
Stream
Masking
Threshold

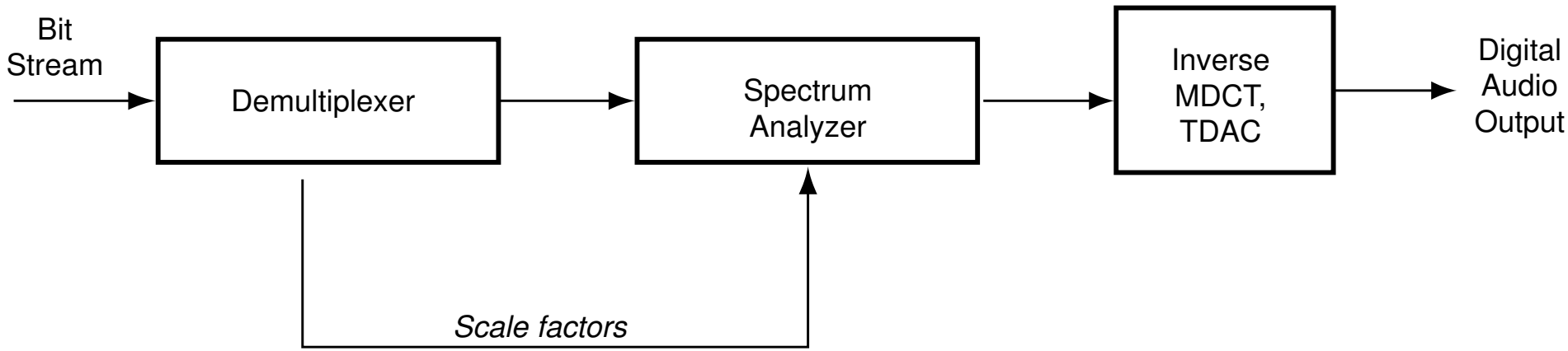

Bit
Stream
Demultiplexer
Spectrum
Analyzer
Inverse
MDCT,
TDAC
Digital
Audio
Output
Scale factors

Table 10.19 Technical features of the MUSICAM algorithm

Sampling frequency	32, 44.1, 48 KHz
Bit rate	64, 96, 128, 192 Kbps
Delay time	Encoder 24 ms Decoder 6 ms

Table 10.20 Technical features of the ASPEC algorithm

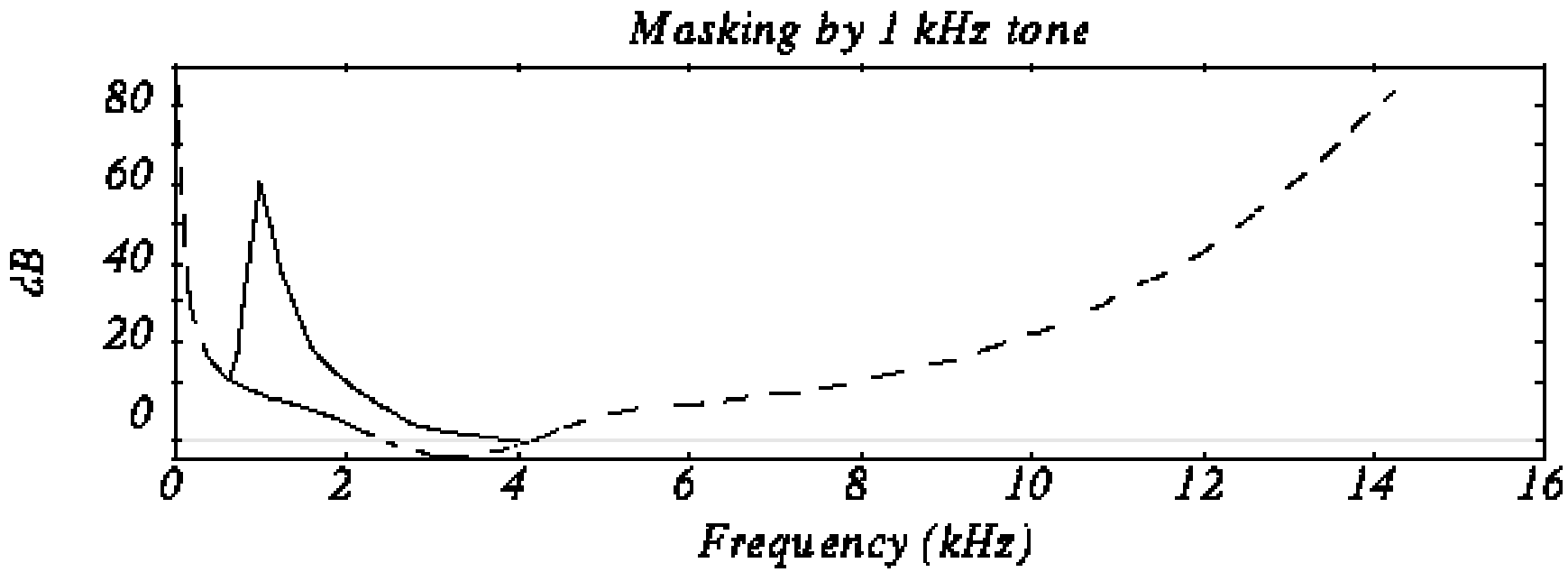
Sampling frequency	32, 44.1, 48 KHz
Bit rate	32, 64, 96, 128 Kbps
Block length	256, 512 samples

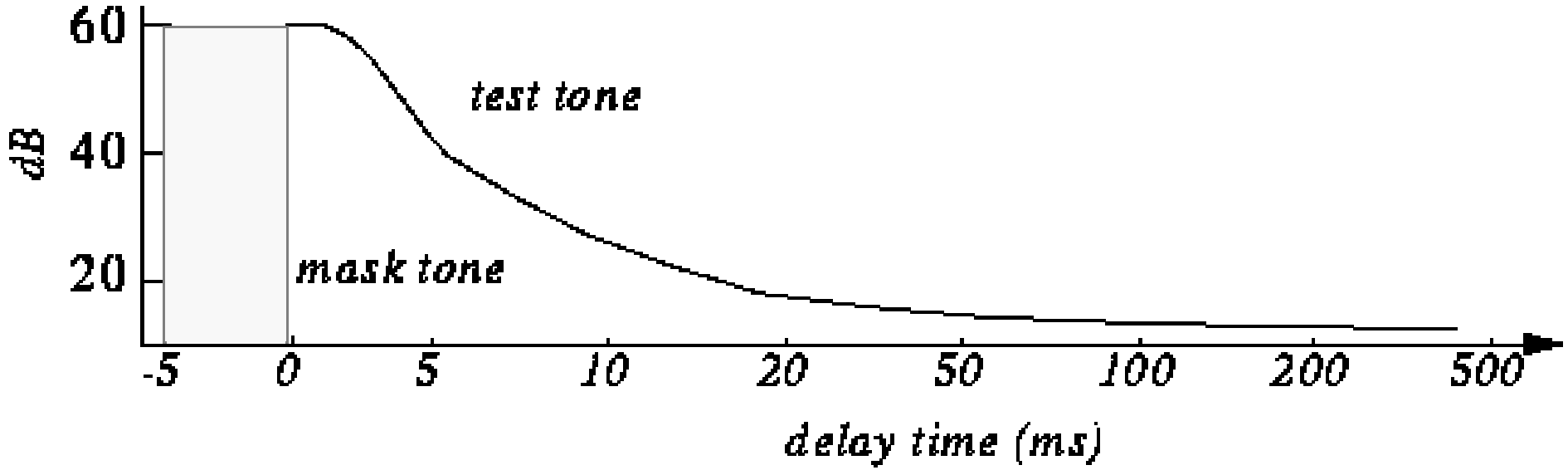
Psychoakustický model - maskování

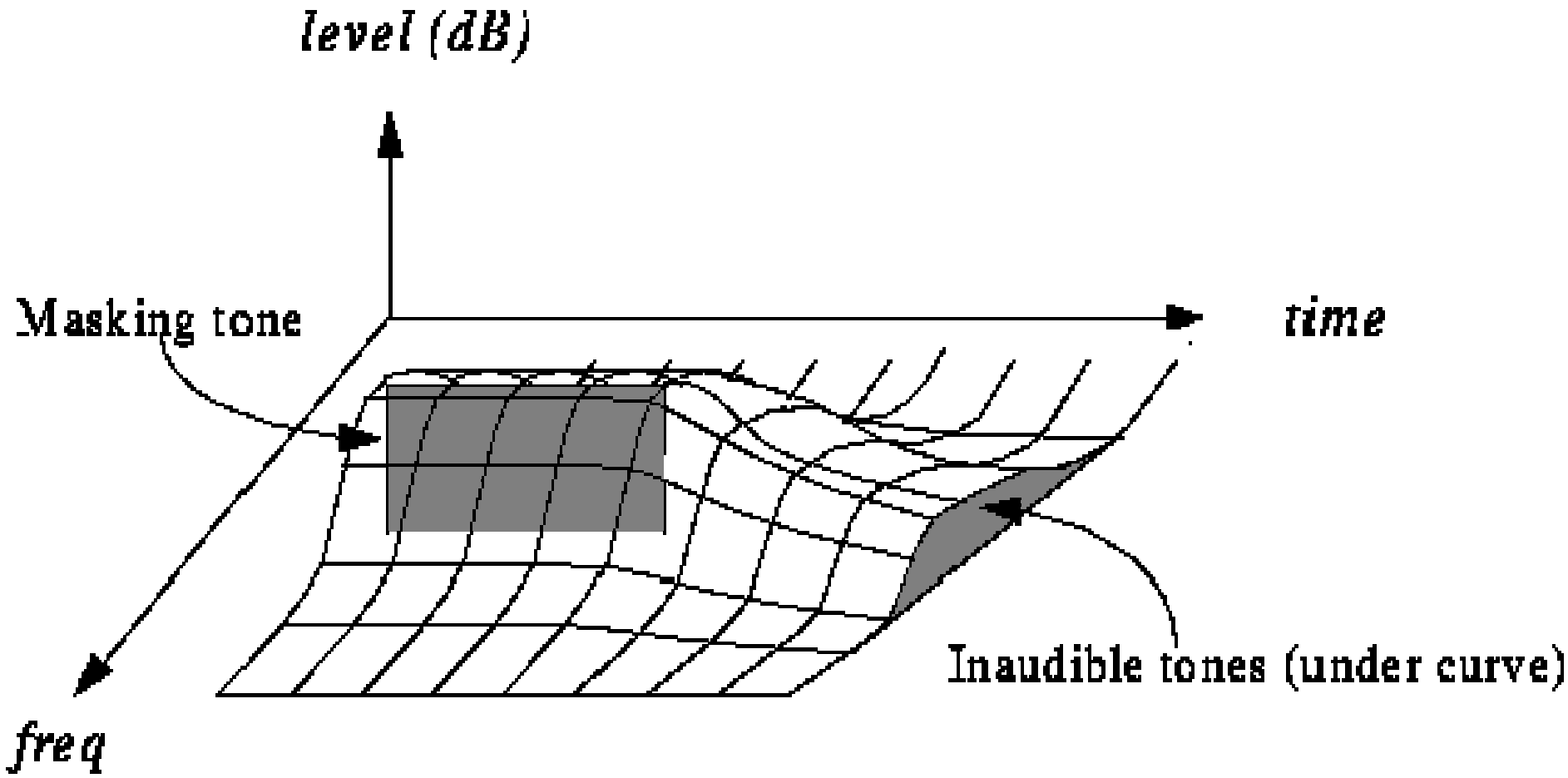
- **frekvenční** - frekvence o definované intenzitě maskují okolní frekvence s menší intenzitou

- **časové** - signál o určité frekvenci a intenzitě i po odeznění maskuje jiné frekvence o menší intenzitě

- kodér analyzuje data, určí množství frekv. vzorků, které mohou být zanedbány
- určí počet bitů nutných k zakódování vzorků pásma


Masking by 1 kHz tone
dB
80
60
40
20
0
0
2
4
6
8
10
12
14
16
Frequency (kHz)


dB
60
40
20
test tone
mask tone
-5
0
5
10
20
50
100
200
500
delay time (ms)


level (dB)
Masking tone
time
Inaudible tones (under curve)
freq

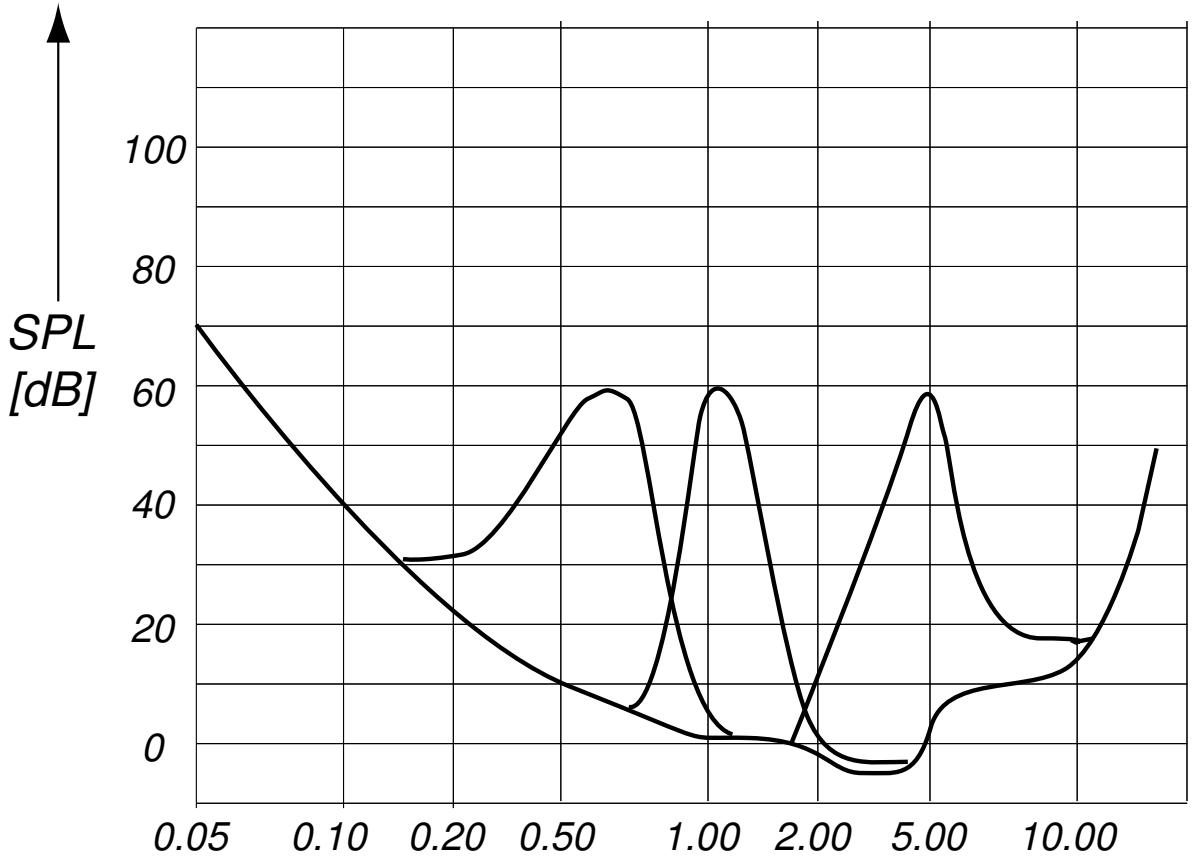
Psychoakustický model 1

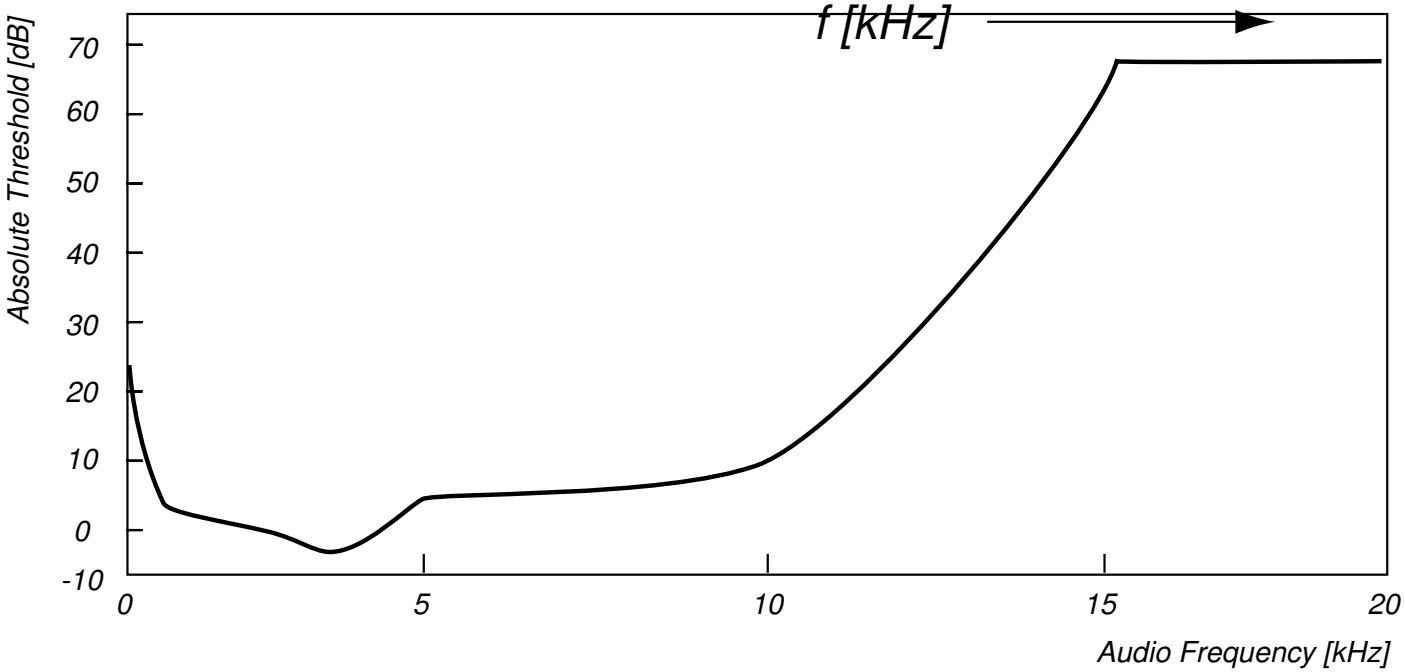
- FFT (512 vzorků - LayerI, 1024 - LayerII) + sub-band filtering
- v každém pásmu vypočten SPL (*sound pressure level*)
- určení prahu ticha (*absolute threshold*)
- určení znělých (harmonických) a neznělých (šum) komponent
- maskování vzhledem k absolutnímu prahu
- maskování vzhledem k nejsilnějším vzorkům do definované vzdálenosti
- pro každý vzorek vypočten maskovací práh
- výpočet odstupu signálu od maskovací frekvence

SMR, určeno k přiřazení počtu bitů při kódování (bit-allocation)

$$SMR_{sb}(n) = L_{sb}(n) - LT_{\min}(n)$$

$L_{sb}(n)$ … úroveň signálu v pásmu n
$LT_{\min}(n)$ … minimální celkový maskovací práh


SPL
[dB]
100
80
60
40
20
0
0.05
0.10
0.20
0.50
1.00
2.00
5.00
10.00
f [kHz]


Absolute Threshold [dB]
70
60
50
40
30
20
10
0
-10
0
5
10
15
20
Audio Frequency [kHz]

Psychoakustický model 2

- používá se v 3. vrstvě (Layer III)
- FFT vypočítána pro délku 192 a 576 vzorků
- uvažuje se rozšíření maskování na okolní pásma
- uvažuje se temporální maskování

$$SMR_n = 10\log_{10}(epart_n / npart_n)dB$$

epart … energie signáluv pásmu *n*
npart … úroveň šumu v pásmu *n*

Kódování vrstvy I (Layer I)

- Kódovaný zvuk se rozdělí na 32 frekvenčních pásem. Každé toto pásmo se kóduje zvlášť. Jednotlivá pásma jsou široká 625 Hz.
- Je použita kvantizace.
- Kvantizační koeficienty se vypočítávají dynamicky podle dosažené akustické hladiny zvuku tak, aby šum vzniklý použitím kvantizace byl pod rozlišovací schopností ucha.
- Vypočtený signál se kóduje entropickou metodou, např. Huffmanovým kódováním.

Kódování vrstvy I (Layer I)

- základní algoritmus
- Data (Layer I)

Bit Allocation (4 bits)	Scale Factors (6 bits)	Samples (2 ~ 15 bits)

- komprese obdobná jako MUSICAM
- sub-band filtering (pásmový filtr)
  - vstup 32 vzorků do posuvného registru (celkem 512 vzorků)
  - redukce aliasingu: vynásobení koeficienty (dáno normou, windowing)
  - sumace osmi vzorků (výběr každého 64.)
  - filtrace (konvoluční filtr)

$$M_{ik} = \cos\left((2i+1)(k-16)\frac{\pi}{64} \right)$$

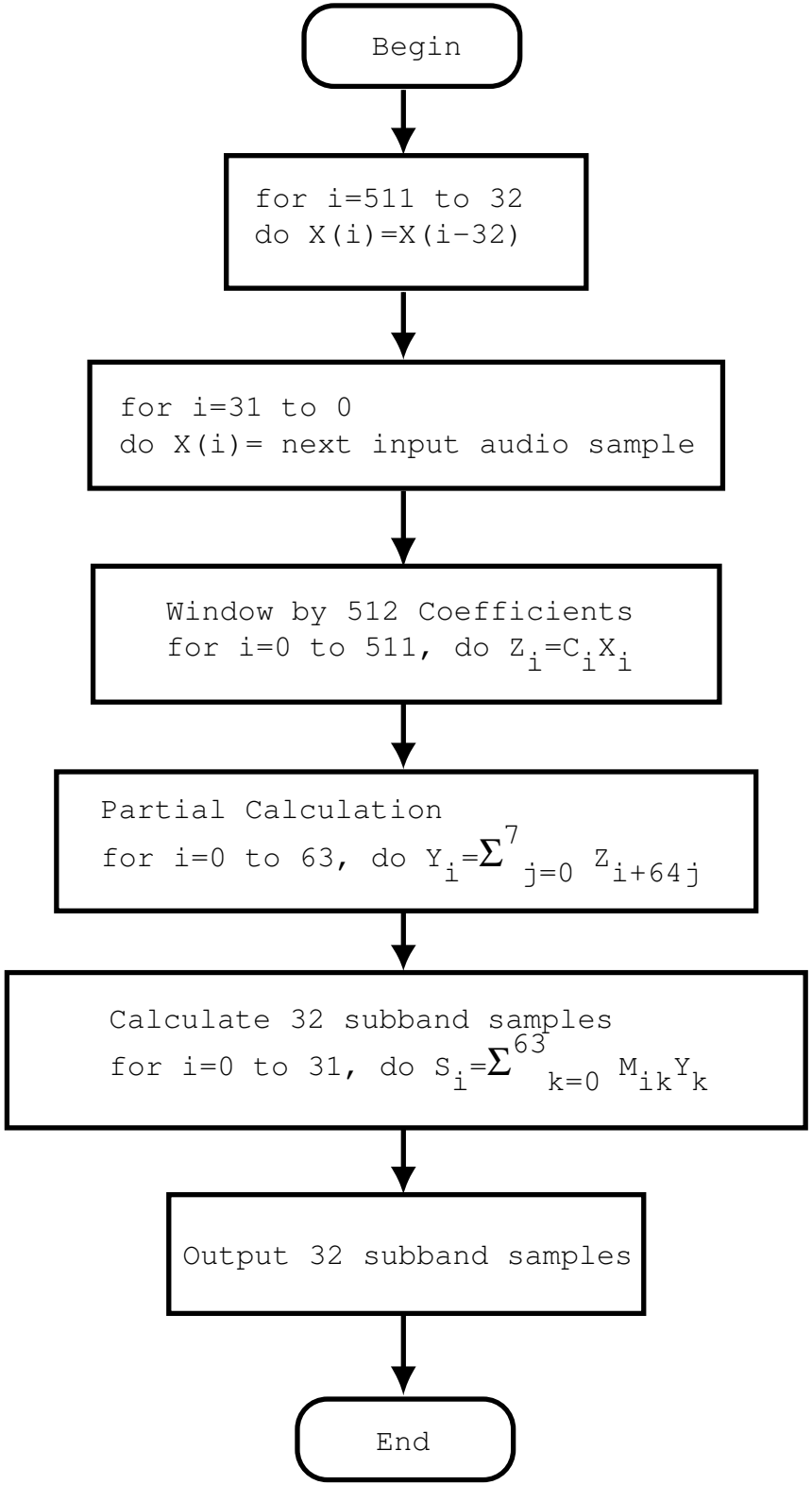
$$i = 0,1,\ldots 31$$

$$k = 0,1,\ldots 63$$

- aplikace psychoakustického modelu (model 1), určení SMR
- určení rozsahu (*scale factor*): nalezení maxima z 12 vzorků v pásmu, zakódování indexu do tabulky [6bitů]
- přiřazení bitů (*bit-allocation*): určení množství bitů potřebných k zakódování frekv. vzorků
- kvantování a zakódování: vzorky normalizovány a kvantovány

$$S_{qi} = \left\lfloor \left(A\left(\frac{S_i}{scf} \right) + B \right) \right\rfloor_N$$

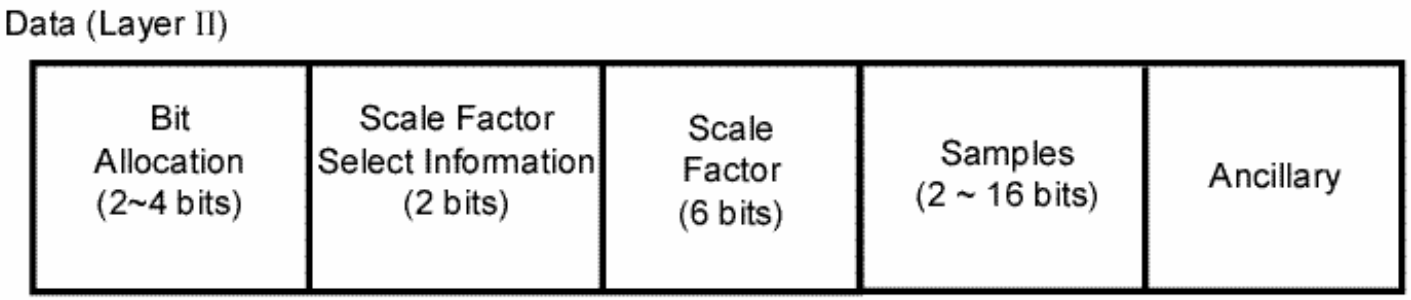
- skládání série bitů na výstup (*bit-stream formatting*)


Begin
for i=511 to 32
do X(i)=X(i-32)
for i=31 to 0
do X(i)= next input audio sample
Window by 512 Coefficients
for i=0 to 511, do $Z_i=C_iX_i$
Partial Calculation
for i=0 to 63, do $Y_i=\Sigma^{7}_{j=0} Z_{i+64j}$
Calculate 32 subband samples
for i=0 to 31, do $S_i=\Sigma^{63}_{k=0} M_{ik}Y_k$
Output 32 subband samples
End

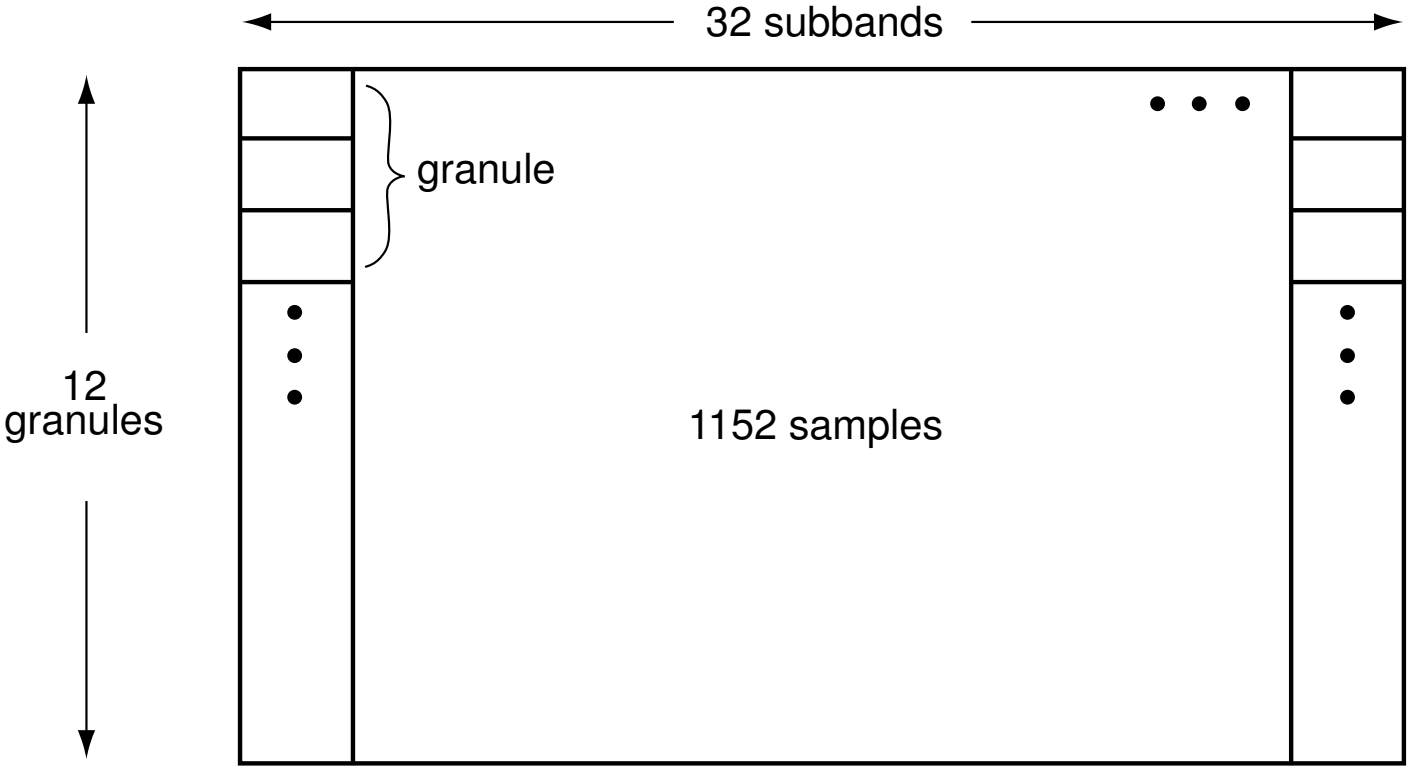
Kódování vrstvy II (Layer II)

- Komprimuje se najednou větší kvantum dat, délka frame je 1152.
- Algoritmy vycházejí z Layer-I, *scale factors* a *bit allocation* se navíc ještě kódují.
- V každém pásmu se kvantizovaná data kódují 0-15 bity, výsledek se násobí 6-bitovým faktorem.

Kódování vrstvy II (Layer II)

- 1152 vzorků
- 

- kódování koeficientů rozsahu (*scale factors*)
  - pásmo má 3*12 vzorků, obsahuje 3 koeficienty rozsahu



- výpočet diferencí rozsahů a jejich rozdělení do tříd

$$Dscf1 = scf2 - scf1$$

$$Dscf2 = scf3 - scf3$$

- třídy definují způsob zakódování *scale factors* (redukce redundance)

• přiřazení bitů (*bit-allocation*)

• kvantizace a kódování

- seskupování trojic vzorků (granulí)

```
for i=0 to 2
S(i)=(code)MOD(no.of steps)
code=(code)//(no.of steps)
                    … no.of steps = 3,5,9
```

- (dekódování: tři kódy jsou dekódovány a dekvantizovány použitím:

$$S'(i) = C\left(S''(i) + D\right)$$

• skládání série bitů na výstup (*bit-stream formatting*)

Table 10.23 Layer II scale factor transmission patterns [487]

Class 1	Class 2	Transmission pattern	Scale factor–select information
1	1	1 2 3	0
1	2	1 2 2	3
1	3	1 2 2	3
1	4	1 3 3	3
1	5	1 2 3	0
2	1	1 1 3	1
2	2	1 1 1	2
2	3	1 1 1	2
2	4	4 4 4	2
2	5	1 1 3	1
3	1	1 1 1	2
3	2	1 1 1	2
3	3	1 1 1	2
3	4	3 3 3	2
3	5	1 1 3	1
4	1	2 2 2	2
4	2	2 2 2	2
4	3	2 2 2	2
4	4	3 3 3	2
4	5	1 2 3	0
5	1	1 2 3	0
5	2	1 2 2	3
5	3	1 2 2	3
5	4	1 3 3	3
5	5	1 2 3	0

Source: © 1993 ISO/IEC.

Table 10.24 Coding of scale factor–select information

scfsi	Number of coded scale factors	Decoding scale factors
0 (00)	3	scf1, scf2, scf3
1 (01)	2	1st ⇒ scf1 and scf2 2nd ⇒ scf3
2 (10)	1	one for all
3 (11)	2	1st ⇒ scf1 2nd ⇒ scf2 and scf3

Kódování vrstvy III (Layer III)

- Kromě základního rozdělení zvuku do 32 frekvenčních pásem se ještě každé pásmo dělí do 6-18 komponent, tzv. *subbands*. Celkem se frekvenční spektrum rozdělí až do 576 nezávislých pásem.
- Kvantizátor je nestejnoměrný, segmentace je použita adaptivní.
- Kromě klasického psychoakustického modelu je využit *temporal masking*.

Kódování vrstvy III (Layer III)

- sub-band filtering (jako v I a II) + MDCT pro zpřesnění transformace
  - MDCT pro délku bloku 36 a 12 vzorků (krátký blok zlepšuje časové rozlišení náhlých změn signálu)
  - definováno „okno“ s koeficienty

$$h(k) = x(k)\sin\left(\frac{\pi}{N}\left(k+\frac{1}{2}\right)\right)$$

$$k = 0,1,\ldots 35$$

$$N = 36$$

  - je definována přechodová funkce - postupný přechod z dlouhého na krátký blok (a naopak)
- MDCT: získání *N/2* koeficientů z *N*

$$S_i = \sum_{k=0}^{N-1} x_k \cos\left(\frac{\pi}{2N}\left(2k+1+\frac{N}{2}\right)(2i+1)\right)$$

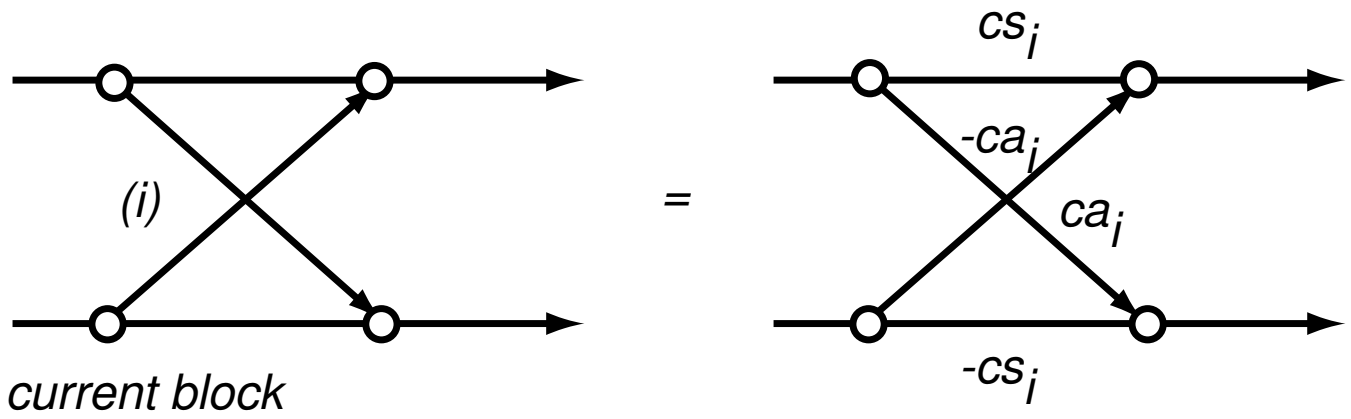
$$i = 0,1,\ldots\frac{N}{2}-1$$

$$N = 12,36$$

$$x_k = \sum_{i=0}^{\frac{N}{2}-1} S_i \cos\left(\frac{\pi}{2N}\left(2k+1+\frac{N}{2}\right)(2i+1)\right)$$

$$k = 0,1,\ldots N-1$$

- snížení aliasingu (i v dekodéru) (pouze pro dlouhé bloky)



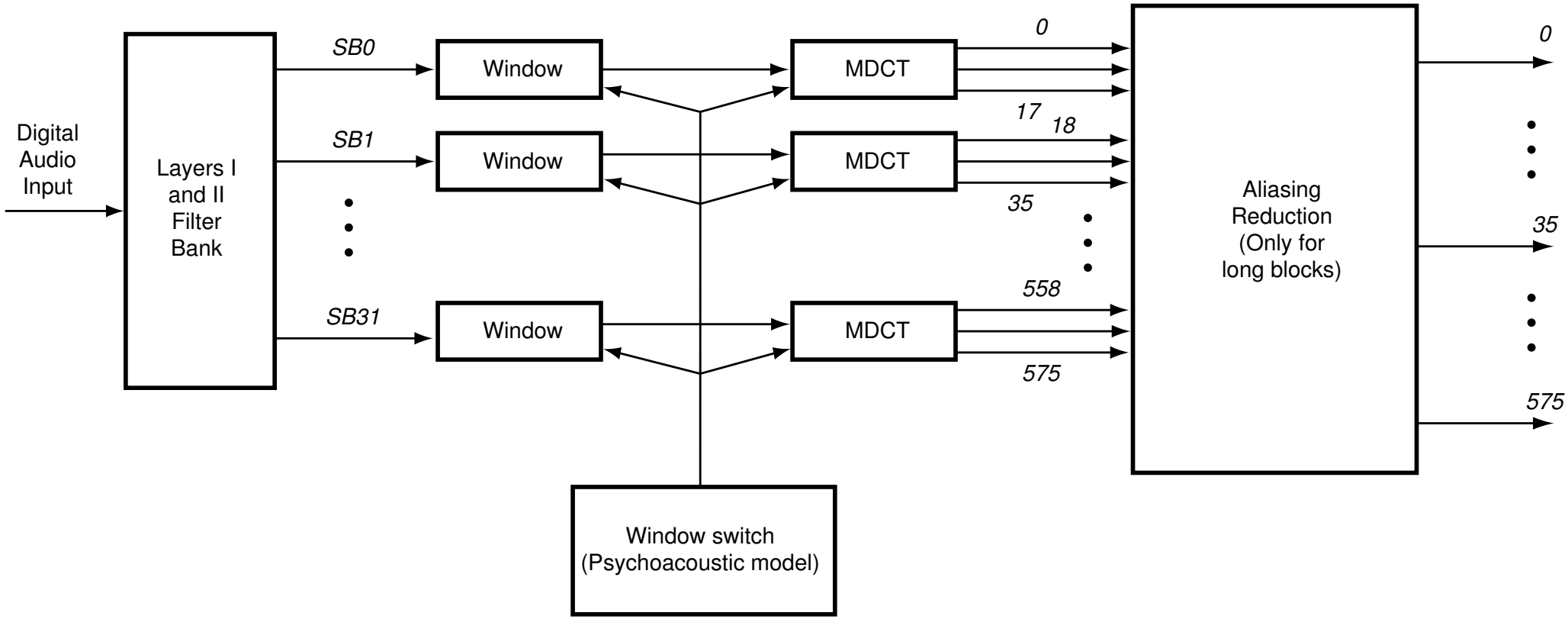
i=0, 1, ..., 7

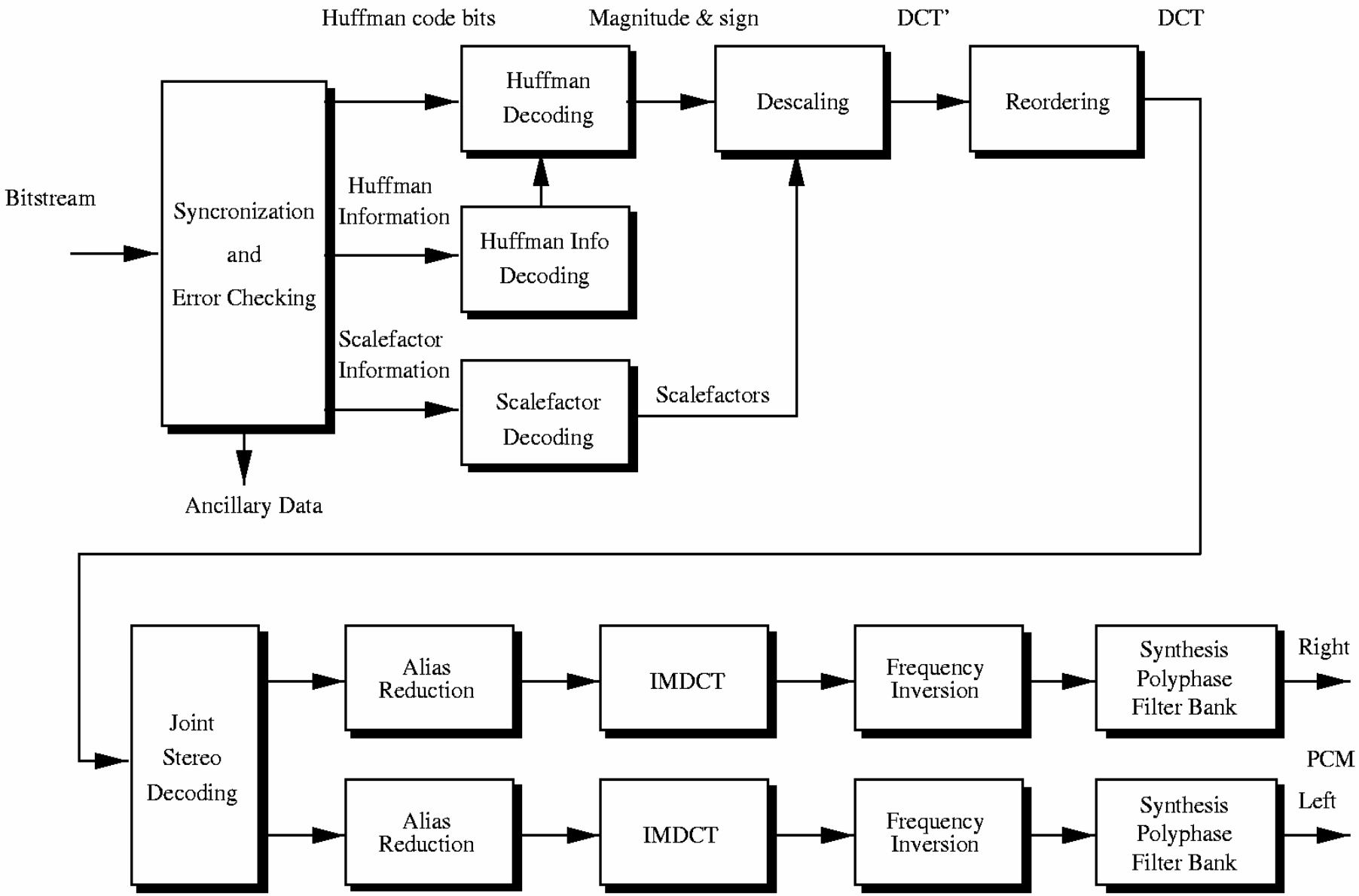
$$cs_i = \frac{1}{\sqrt{1+c_i^2}}$$

$$ca_i = \frac{c_i}{\sqrt{1-c_i^2}}$$

(c_i definováno normou)

- noise-allocation: rozšířené hledání vhodného počtu bitů (*bit-allocation*)
- Huffmanovo kódování
- bit-reservoir: „ušetřený“ počet bitů se přenáší k použití do dalšího rámce


Digital Audio Input
Layers I and II Filter Bank
SB0
SB1
SB31
Window
Window
Window
MDCT
MDCT
MDCT
0
17
18
35
558
575
Aliasing Reduction (Only for long blocks)
0
35
575
Window switch (Psychoacoustic model)


Huffman code bits
Magnitude & sign
DCT'
DCT
Bitstream
Syncronization and Error Checking
Huffman Information
Huffman Decoding
Huffman Info Decoding
Descaling
Reordering
Scalefactor Information
Scalefactor Decoding
Scalefactors
Ancillary Data
Joint Stereo Decoding
Alias Reduction
IMDCT
Frequency Inversion
Synthesis Polyphase Filter Bank
Right
PCM
Left
Alias Reduction
IMDCT
Frequency Inversion
Synthesis Polyphase Filter Bank