Application of wavelet filtering to image compression

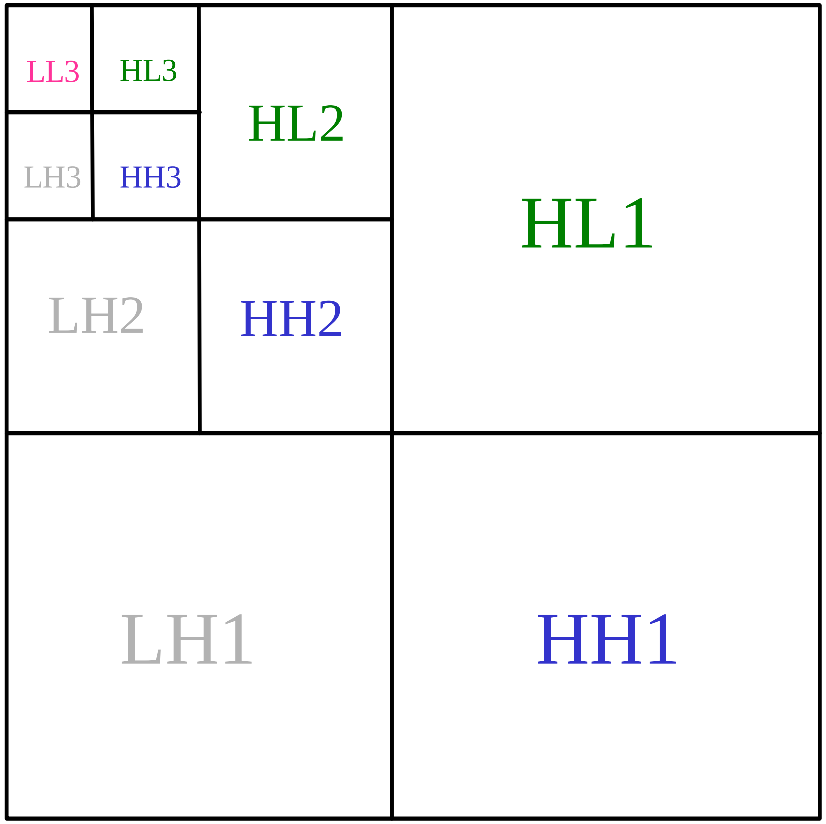
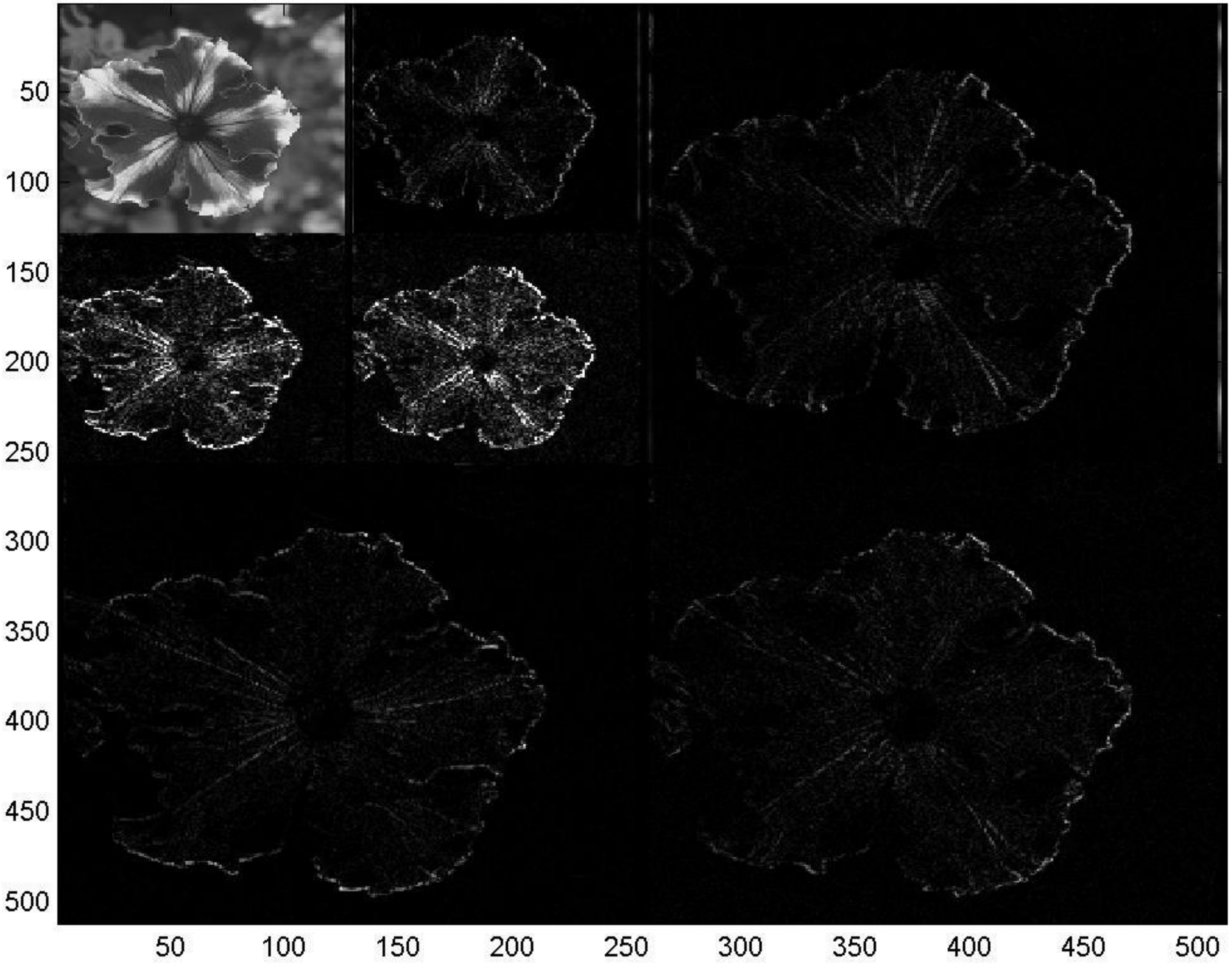


Fig. 9.1 Wavelet decomposition of image.

Application to image compression



Application to image compression

The HHX matrices correspond to high horizontal- high-vertical

filtering,

the HLX matrices correspond to high horizontal-low vertical filtering,

the LHX matrices correspond to low horizontal-high vertical filtering,

the LLX matrices correspond to low horizontal-low vertical filtering.

For decomposition with r levels we obtain $3r+1$ matrices of reducing size. Most of the energy is in the low-lowpass subband LL3, other matrices add details.

Application to image compression

The quantized highpass subbands usually contain many zeros. They can be efficiently compressed using zero run lengths coding followed by the Huffman coding of pairs (run length, amplitude) or arithmetic coding.

The lowpass subbands usually do not contain zeros at all or contain small number of zeros and can be encoded by the Huffman code or by the arithmetic code.

More advanced coding procedures used, for example, in MPEG-4 standard try to take into account dependencies between subbands. One of such methods is called zero-tree coding.

Zero-tree coding

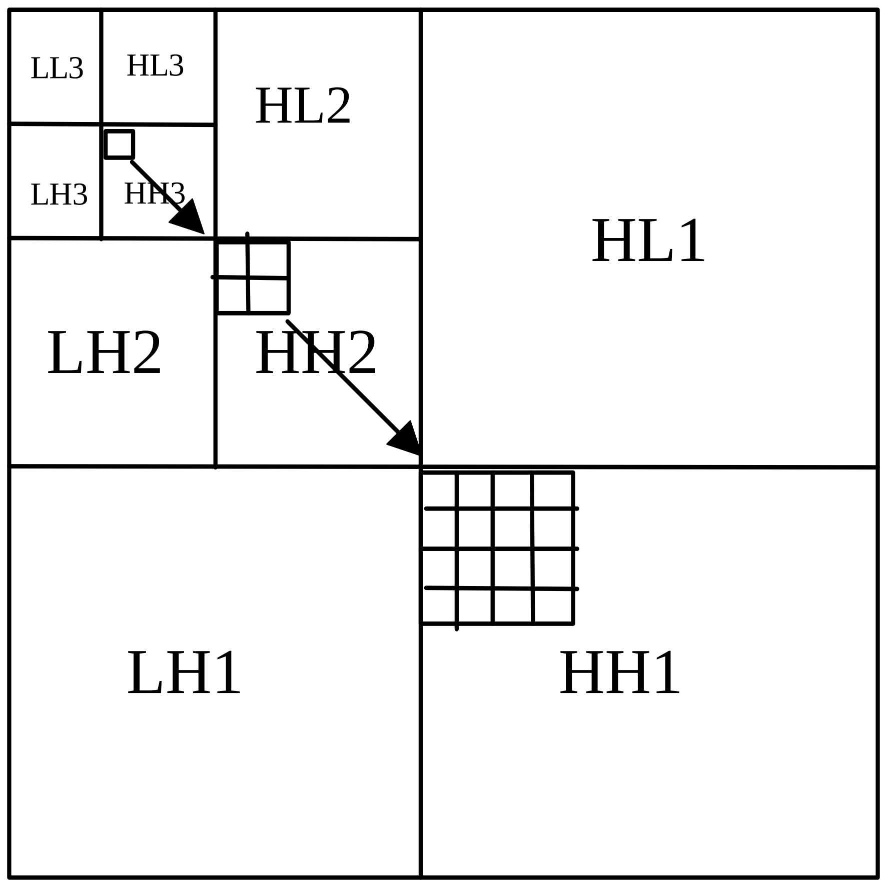
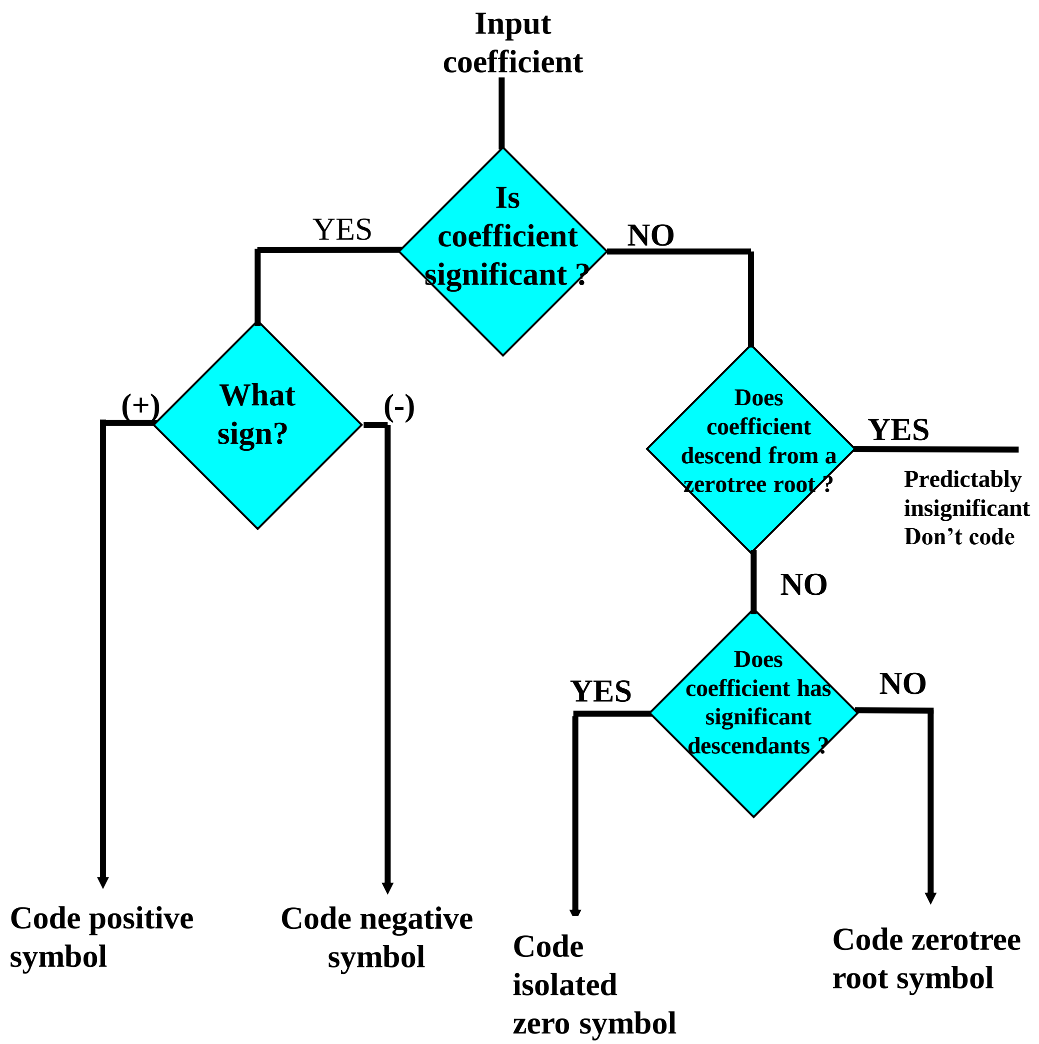


Fig.9.2 Parent-child dependencies of subbands


Input coefficient
Is coefficient significant ?
YES
NO
What sign?
(+)
(-)
Does coefficient descend from a zerotree root ?
YES
Predictably insignificant Don’t code
NO
Does coefficient has significant descendants ?
YES
NO
Code positive symbol
Code negative symbol
Code isolated zero symbol
Code zerotree root symbol

EZW

Zerotree coding combined with bit-plane coding is called **E**mbedded **Z**erotree **W**avelet (EZW) coding.

Bit plane is formed by bits of different coefficients with the same significance. It is used for lossless and scalable lossy image coding.

Two passes are used: sorting and refinement

Sorting: coefficients larger than 2^n are classified as significant (their signs and positions are transmitted)

Refinement: the n-*th* MSB of coefficients larger than 2^{n+1} is transmitted and $n \leftarrow n-1$

SPIHT

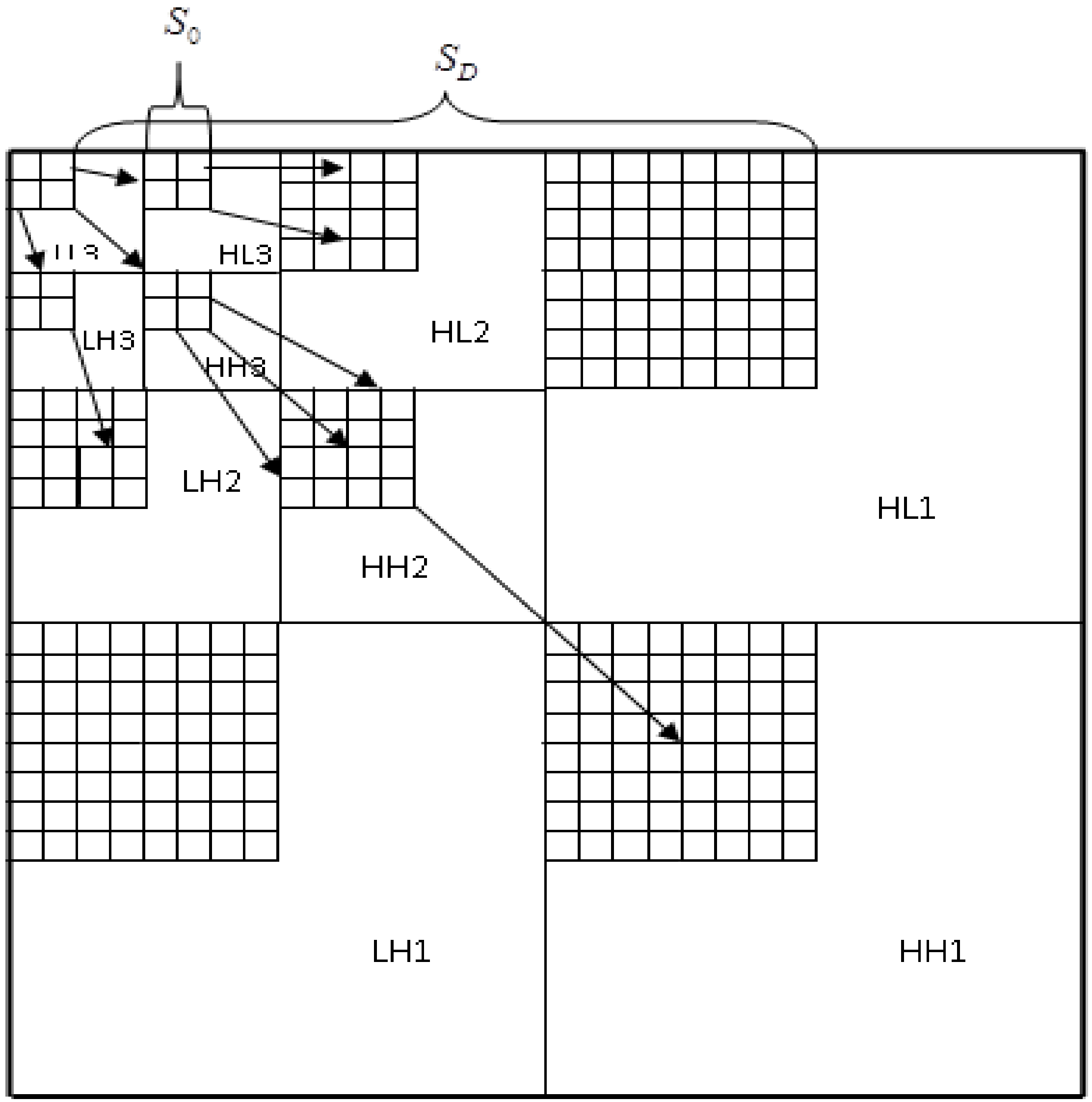
Set **P**artitioning **I**n **H**ierarchical **T**rees exploits wavelet subbands dependences and bit-plane coding.

The main idea is to group coefficients into subsets corresponding to the spatial orientation trees and then apply bit-plane coding to the obtained subsets.

The algorithm classifies not each coefficient but subsets of coefficients as significant or insignificant. If a subset is encoded as insignificant then **0** is transmitted otherwise **1** is transmitted. The significant subsets are further partitioned into new subsets and the significance test is then applied to them.

A subset is classified as significant is the maximum absolute value of its coefficient exceeds a threshold.

SPIHT(Set Partitioning in Hierarchical Trees)



SPIHT

The indices of all coefficients are put into three lists: the list of insignificant coefficients (LIC), the list of significant coefficients (LSC) and the list of insignificant sets of coefficients (LIS).

The following partition is used :

1. If the significant set is $S_D(i, j)$ it is split into $S_L(i, j)$ and the four coefficients which are offsprings of the coefficient with coordinates (i,*j*).

2. If the significant set is $S_L(i, j)$ it is split into four $S_D(k, l)$ sets which are sets of descendants of the offsprings (k,l) of the coefficient with coordinates (i,*j*).

Speech coding for multimedia applications

The most important attributes of speech coding are:

- **Bit rate.** From 2.4 kb/s for secure telephony to 64 kb/s for network applications.
- **Delay.** For network applications delay below 150ms is required. For multimedia storage applications the coder can have unlimited delay.
- **Complexity.** Coders can be implemented on PC or on DSP chips. Measures of complexity for a DSP or a CPU are different. Also complexity depends on DSP architecture. It is usually expressed in MIPS.
- **Quality.** For secure telephony quality is synonymous with intelligibility. For network applications the goal is to preserve naturalness and subjective quality.

Standardization

- International Telecommunication Union (ITU)
- International Standards Organization (ISO)
- Telecommunication Industry Association(TIA),NA
- R&D Center for Radio systems (RCR),Japan

Standard	Bit rate	Frame size / look-ahead	Complexity
ITU standards			
G.711 PCM	64 kb/s	0/0	0.01 MIPS
G.726 , G.727 ADPCM	16, 24, 32, 40 kb/s	0.125 ms/0	2 MIPS
G.722 Wideband coder	48, 56, 64 kb/s	0.125/1.5 ms	5 MIPS
G.728 LD-CELP	16 kb/s	0.625 ms/0	30 MIPS
G.729 CS-ACELP	8 kb/s	10/5 ms	20 MIPS
G.723.1 MPC-MLQ	5.3 & 6.4 kb/s	30/7.5 ms	16 MIPS
G.729 CS-ACELP annex A	8 kb/s	10/5 ms	11 MIPS
Cellular standards			
RPE-LTP(GSM)	13 kb/s	20 ms/0	5 MIPS
IS-54 VSELP (TIA)	7.95 kb/s	20/5 ms	15 MIPS
PDC VSELP(RCR)	6.7 kb/s	20/5 ms	15 MIPS
IS-96 QCELP (TIA)	8.5/4/2/0.8 kb/s	20/5 ms	15 MIPS
PDC PSI-CELP (RCR)	3.45 kb/s	40/10 ms	40 MIPS
U.S. secure telephony standards			
FS-1015 LPC-10E	2.4 kb/s	22.5/90 ms	20 MIPS
FS-1016 CELP	4.8 kb/s	30/30 ms	20 MIPS
MELP	2.4 kb/s	22.5/20 ms	40 MIPS

History of speech coding

•40s – PCM was invented and intensively developed in 60s

•50s – delta modulation and differential PCM were proposed

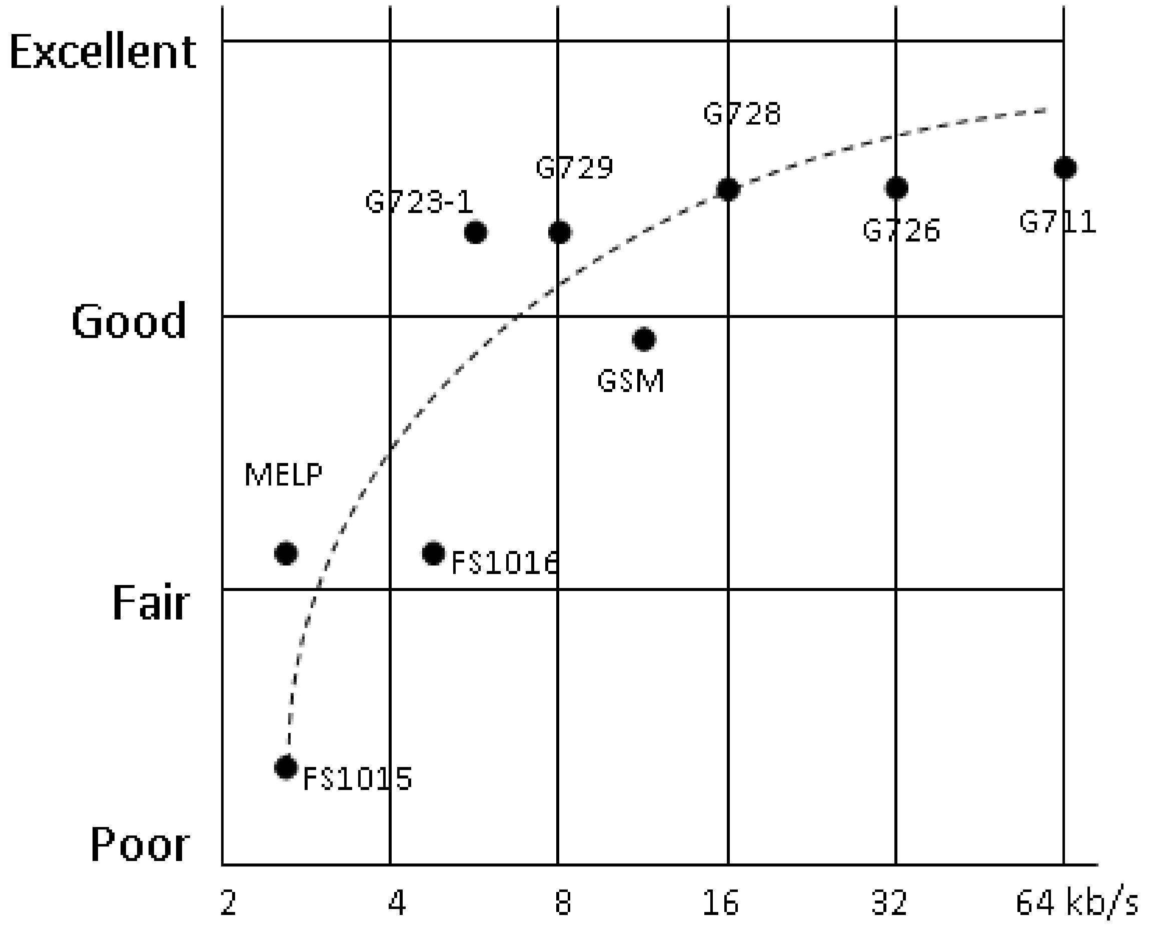
•1957 – $\mu -$law encoding was proposed (standardised for telephone networks in 1972 (G.711))

•1974 – ADPCM was developed (1976 – standards G.726 and G.727)

•1984 – CELP coder was proposed (majority of coding standards for speech coding today use a variation of CELP)

•1995 – MELP coder was invented

Subjective quality of speech coders



Speech coders. Demo.

Original file 8kHz, 16 bits, file size is 73684 bytes

A-law 8 kHz, 8 bits, file size is 36842 bytes

ADPCM, 8 kHz, 4 bits, file size is 18688 bytes

GSM , file size is 7540 bytes

CELP-like coder, file size is 4896 bytes

MELP-like coder, file size 920 bytes

MELP-like coder, file size 690 bytes

MELP-like coder, file size 460 bytes

Direct sample-by-sample quantization: Standard G.711

The G.711 PCM coder is designed for telephone bandwidth speech signals. The input signal is speech signal sampled with rate 8 kHz. The coder does direct sample-by-sample quantization. Instead of uniform quantization one form of nonuniform quantization known as companding is used.

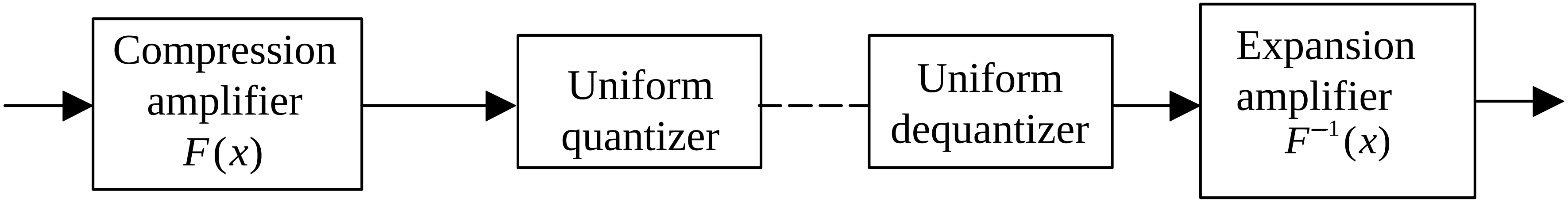
The name of the method is derived from the words "compressing-expanding".

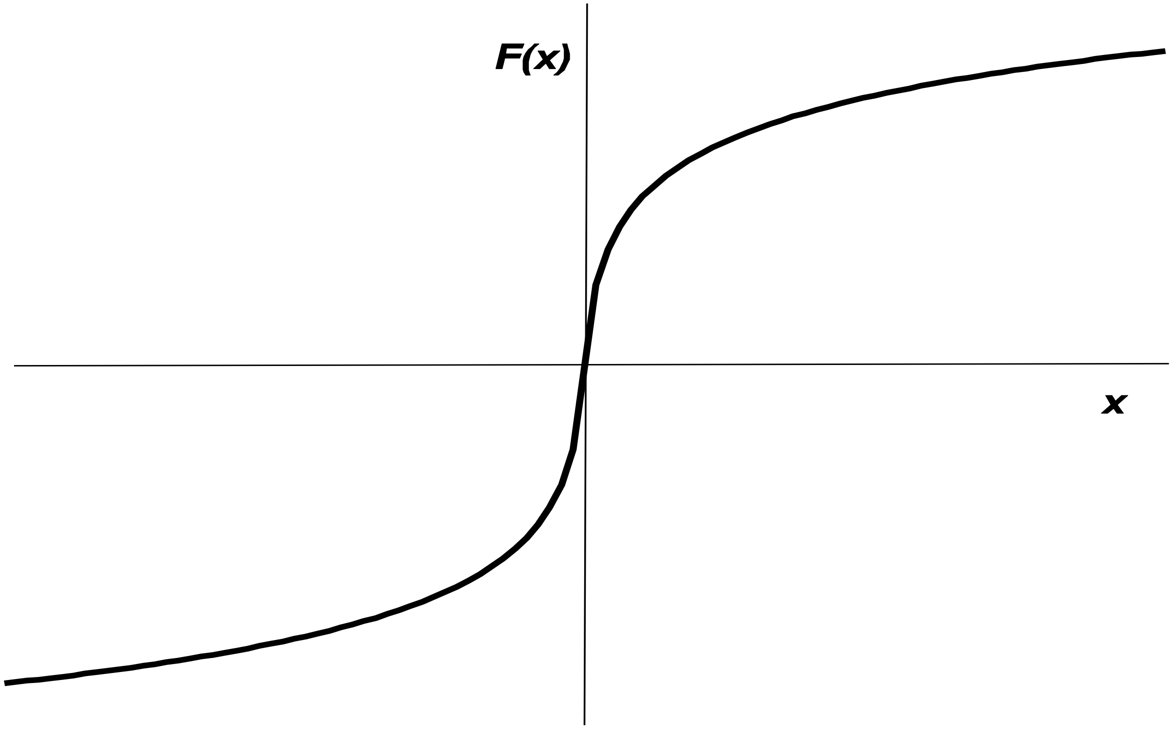
First the original signal is compressed using a memoryless nonlinear device.

The compressed signal is then uniformly quantized.

The decoded waveform must be expanded using a nonlinear function which is the inverse of that used in compression.

Standard G.711





Standard G.711

Companding is equivalent to quantizing with steps that start out small and get larger for higher signal levels.

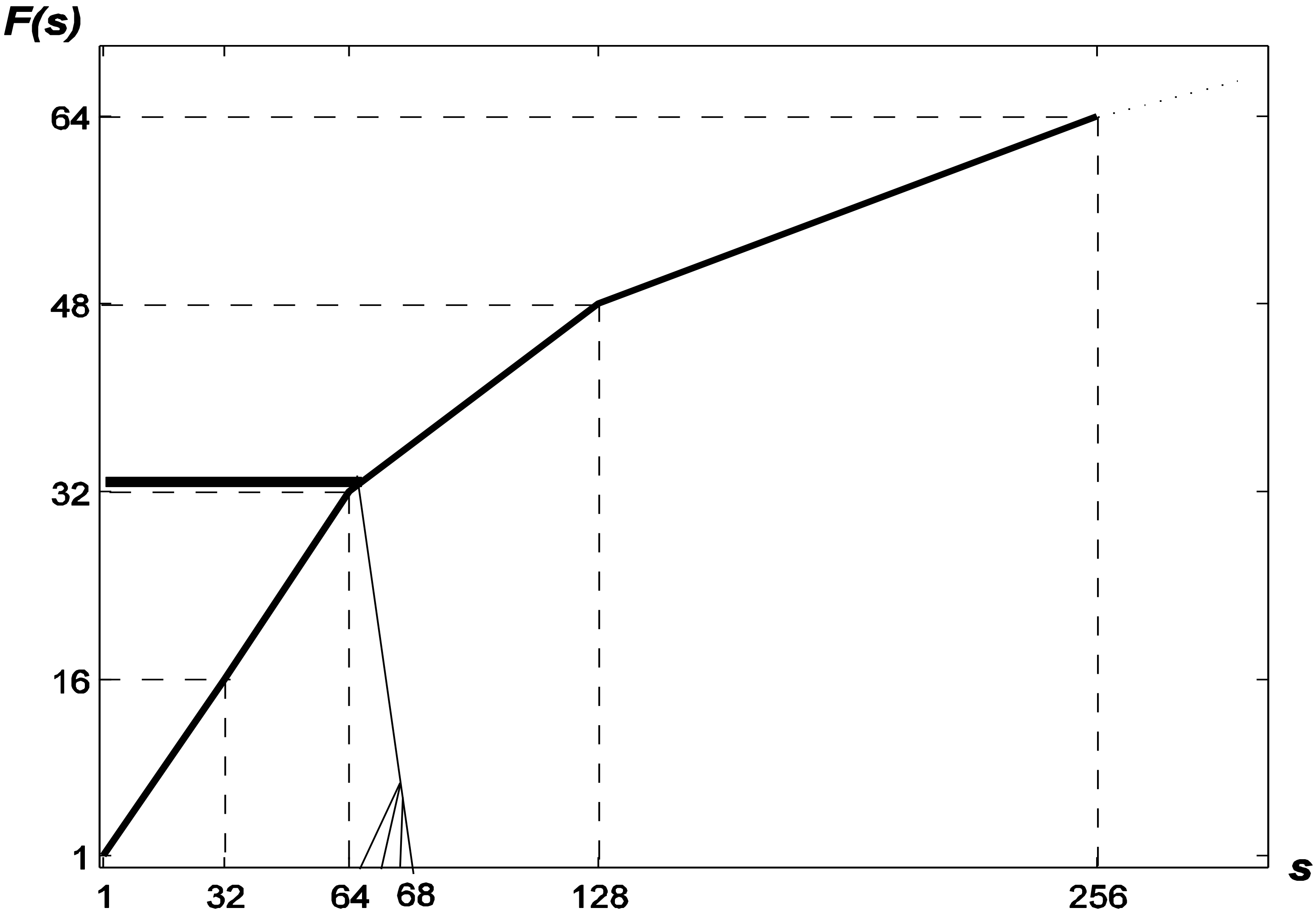
There are different standards for companding.

North America and Japan have adopted a standard compression curve known as a μ-law companding. Europe has adopted a different, but similar, standard known as A-law companding. The μ -law compression formula is

$$F(s) = F_{\max} \operatorname{sgn}(s) \frac{\ln\left(1 + \mu \left| s / s_{\max} \right|\right)}{\ln(1 + \mu)},$$

where s is the input sample value. G.711 uses $\mu = 255.$

Implementation of G.711



Implementation of G.711

The $\mu = 255$ curve is approximated by a piecewise linear curve. When using this law in networks the suppression of the all zero character signal is required.

The positive portion of the curve is approximated by 8 straight line elements. We divide the positive output region into 8 equal segments. The input region is divided into 8 corresponding nonequal segments.

To identify which segment the sample lies we spend 3 bits.

The value of sample in each segment is quantized to 4 bits number.

1 bit shows the polarity of the sample.

In total we spend 4+3+1 bits for each sample.

Example

In order to quantize the value 66 we spend 1 bit for sign,

The number of the segment is 010,

The quantization level is 0000 (values 65,66,67,68 are quantized to 33).

At the decoder using the codeword 10100000 we reconstruct the approximating value (64+68)/2=66.

The same approximating value will be reconstructed instead of values 65,66,67,68.

Each segment of the input axis is twice as wide as the segment to its left. The values 129,130,131,132,133,134,135,136 are quantized to the value 49. The resolution of each next segment is twice times as bad as of the previous one.

ADPCM coders: Standards G.726, G.727

Coders of this type are based on linear prediction method.

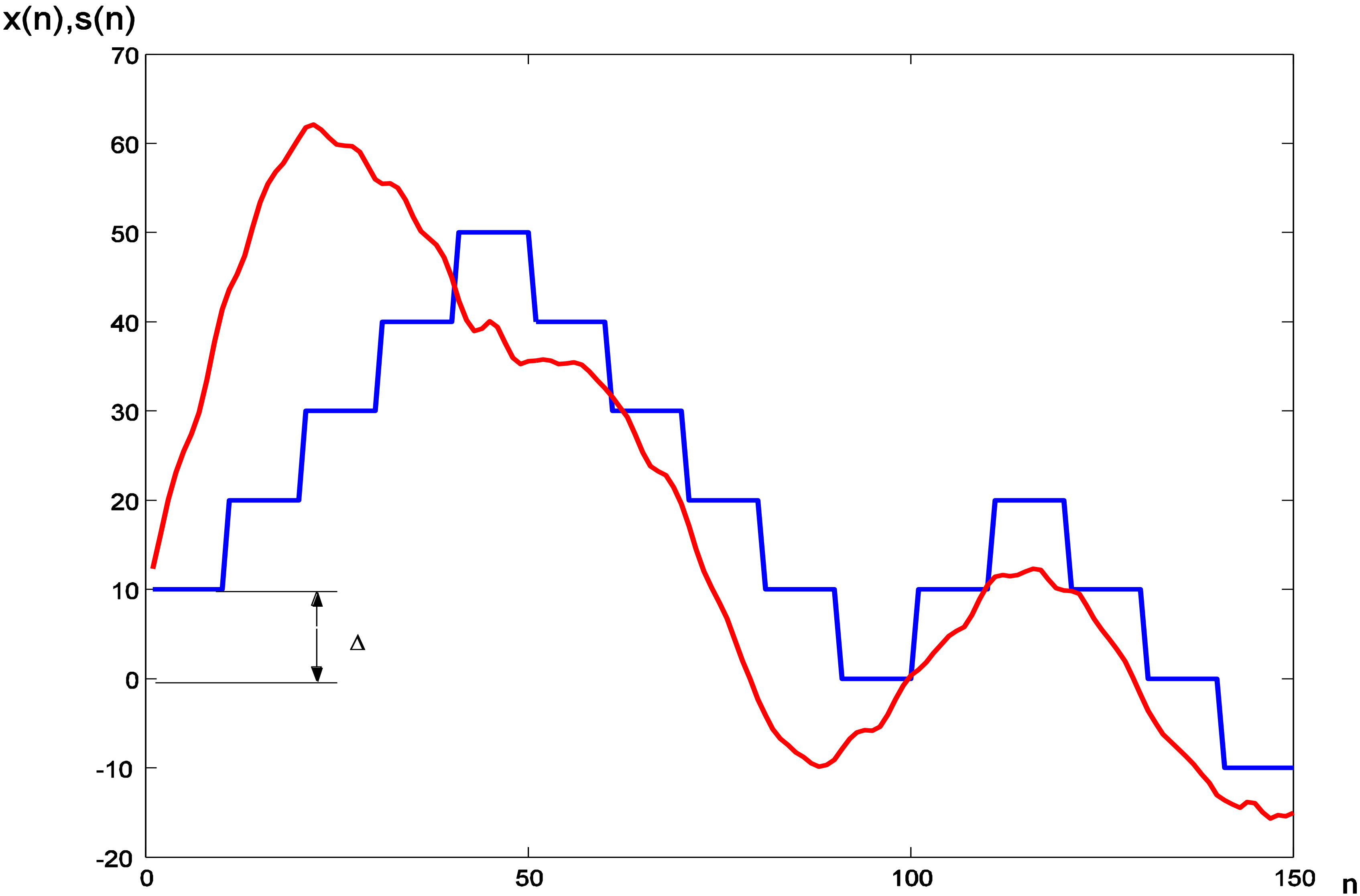
The main feature of these coders is that they use nonoptimal prediction coefficients and prediction is based on past reconstructed samples. These coders provides rather low delay and have low computational complexity as well.

Delta-modulation is a simple technique for reducing the dynamic range of the numbers to be coded.

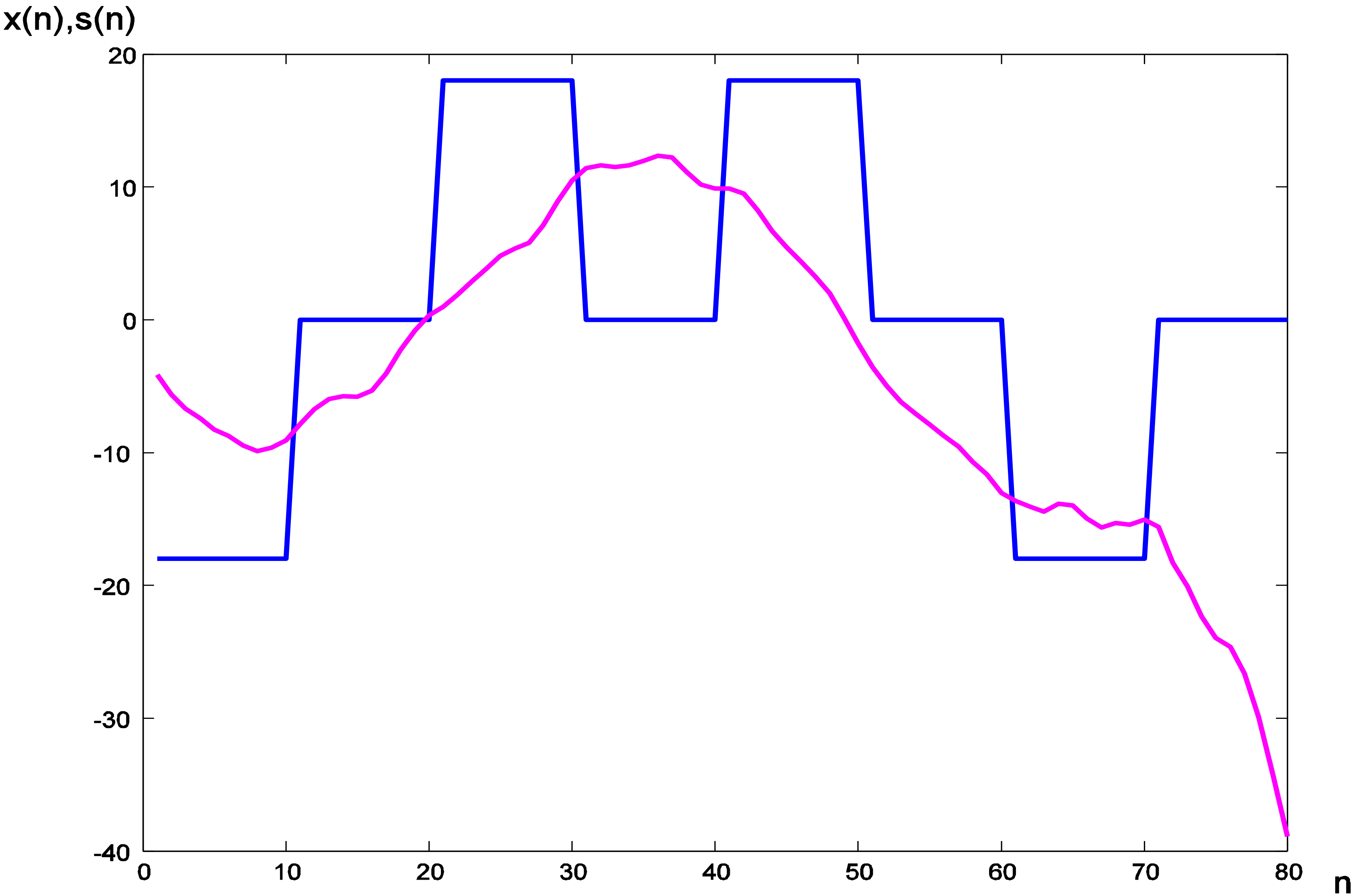
Instead of sending each sample value, we send the difference between sample and a value of a staircase approximation function.

The staircase approximation can only either increase or decrease by step Δ at each sample point.

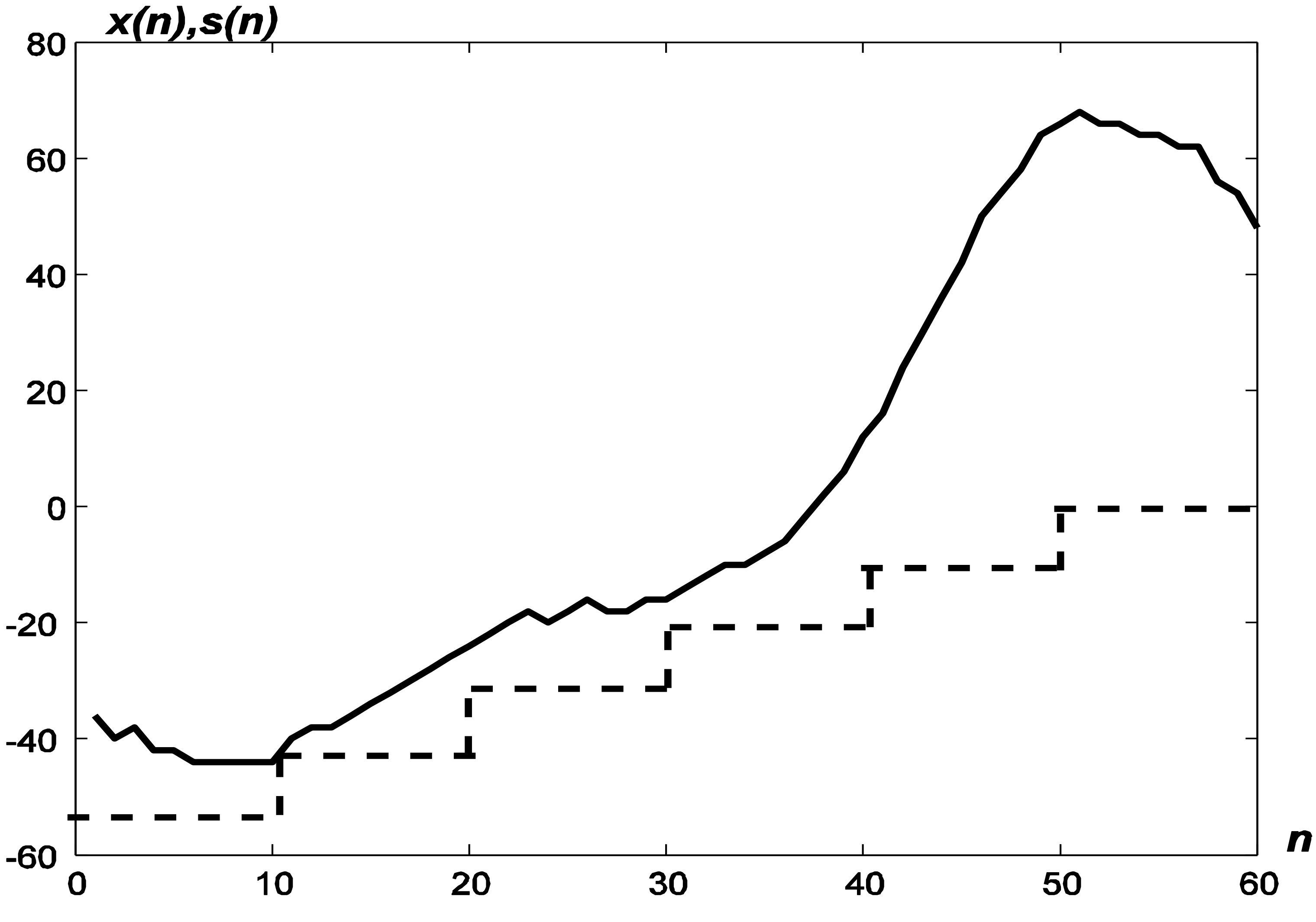
Delta-modulation



Granular noise



Slope overload



Delta-modulation

The choice of step size Δ and sampling rate is important.

If steps are too small we obtain a slope overload condition where the staircase cannot trace fast changes in the input signal.

If the steps are too large, considerable overshoot will occur during periods when the signal is changing slowly. In that case we have significant quantization noise, known as granular noise.

Adaptive delta-modulation is a scheme which permits adjustment of the step size depending upon the characteristics of the input signal.

Adaptive delta-modulation

The idea of step size adaptation is the following:

If output bit stream contains almost equal number of 1's and 0's we assume that the staircase is oscillating about slowly varying analog signal and reduce the step size.

An excess of either 1's or 0's within an output bit stream indicates that staircase is trying to catch up with the function. In such cases we increase the step size.

Usually delta-modulators require sampling rate greater than the Nyquist rate.

They provide compression ratios 2-3 times.

G.726, G.727

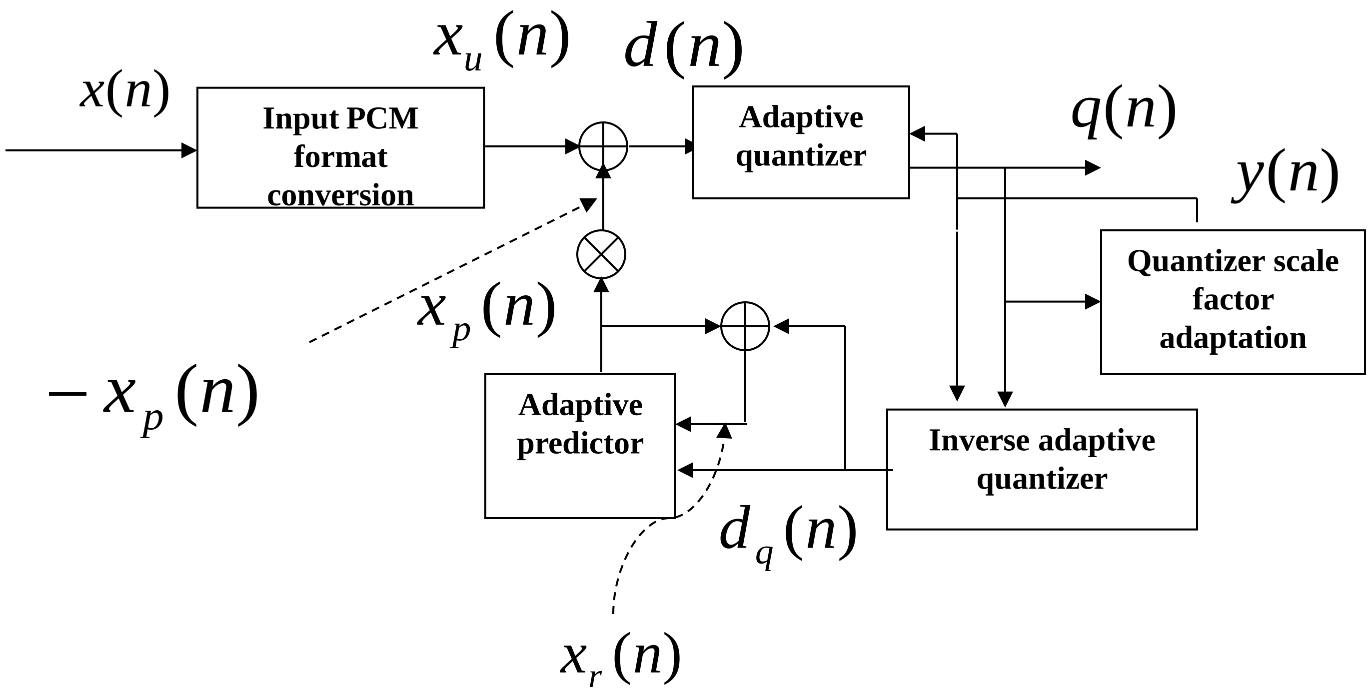
The speech coders G.726 and G.727 are adaptive differential pulse-coded modulation (ADPCM) coders for telephone bandwidth speech. The input signal is 64 kb/s PCM speech signal (sampling rate 8 kHz and each sample is 8 bit integer number).

The format conversion block converts the input signal $x(n)$ from A-law or μ-law PCM to a uniform PCM signal $x_u(n)$.

The difference $d(n) = x_u(n) - x_p(n)$, here $x_p(n)$ is the predicted signal, is quantized. A 32-, 16-, 8- or 4 level non-uniform quantizer is used for operating at 40, 32, 24 or 16 kb/s, respectively. Prior to quantization $d(n)$ is converted and scaled

$$l(n) = \log_2 |d(n)| - y(n),$$ where $y(n)$ is computed by the scale factor adaptation block.

G726,G727



$x(n)$
Input PCM format conversion
$x_u(n)$
$d(n)$
Adaptive quantizer
$q(n)$
$y(n)$
Quantizer scale factor adaptation
$-x_p(n)$
$x_p(n)$
Adaptive predictor
Inverse adaptive quantizer
$d_q(n)$
$x_r(n)$

G.726, G727

The value $l(n)$ is then scalar quantized with a given quantization rate.

For bit rate 16 kb/s $|l(n)|$ is quantized with $R = 1$ bit/sample and one more bit is used to specify the sign. Two quantization intervals are: $(-\infty, 2.04)$ and $(2.04, \infty)$. They contain the approximating values 0.91 and 2.85, respectively.

A linear prediction is based on the two previous samples of the reconstructed signal $x_r(n) = x_p(n) + d_q(n)$ and the six previous samples of the reconstructed difference $d_q(n)$:

$$x_p(n) = \sum_{i=1}^{2} a_i(n-1)x_r(n-i) + e(n),$$

$$e(n) = \sum_{i=1}^{6} b_i(n-1)d_q(n-i).$$

G726,G727

The transfer function of the prediction filter is

$$H_e(z) = \frac{E(z)}{X(z)} = \frac{1 - \sum_{i=1}^{2} a_i z^{-i}}{1 + \sum_{i=1}^{6} b_i z^{-i}}$$

G.726, G.727

The predictor coefficients as well as the scale factor are updated on sample-by-sample basis in a backward adaptive fashion. For the second order predictor :

$$a_1(n) = (1-2^{-8})a_1(n-1) + (3\cdot 2^{-8})\,\mathrm{sgn}(p(n))\,\mathrm{sgn}(p(n-1))$$

$$a_2(n) = (1-2^{-7})a_2(n-1) + 2^{-7}\left\{\mathrm{sgn}(p(n))\,\mathrm{sgn}(p(n-2))\right\}$$
$$-2^{-7} f\left\{a_1(n-1)\right\}\mathrm{sgn}(p(n))\,\mathrm{sgn}(p(n-1)),$$

$$p(n) = d_q(n) + e(n)$$

For the sixth order predictor :

$$b_i(n) = (1-2^{-8})b_i(n-1) + 2^{-7}\,\mathrm{sgn}(d_q(n))\,\mathrm{sgn}(d_q(n-i)),$$

$$i = 1,2,\ldots,6.$$