Applications in Compression

"That which shrinks must first expand."

Lao-Tzu, *Tao Te Ching*

Compression systems based on linear transforms

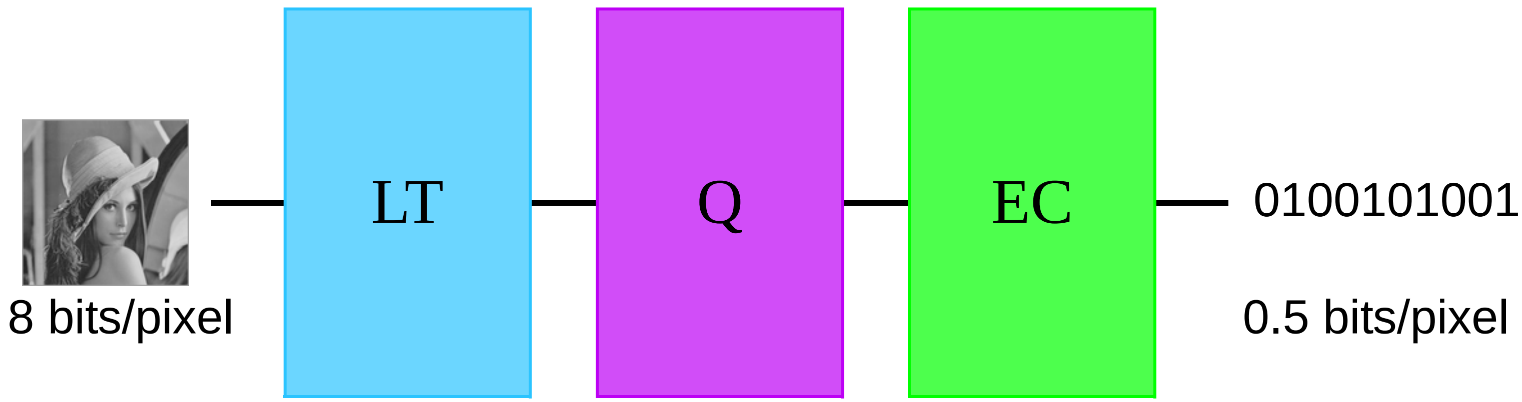
Speech and audio compression

Image compression

Video compression

Compression systems based on linear transforms

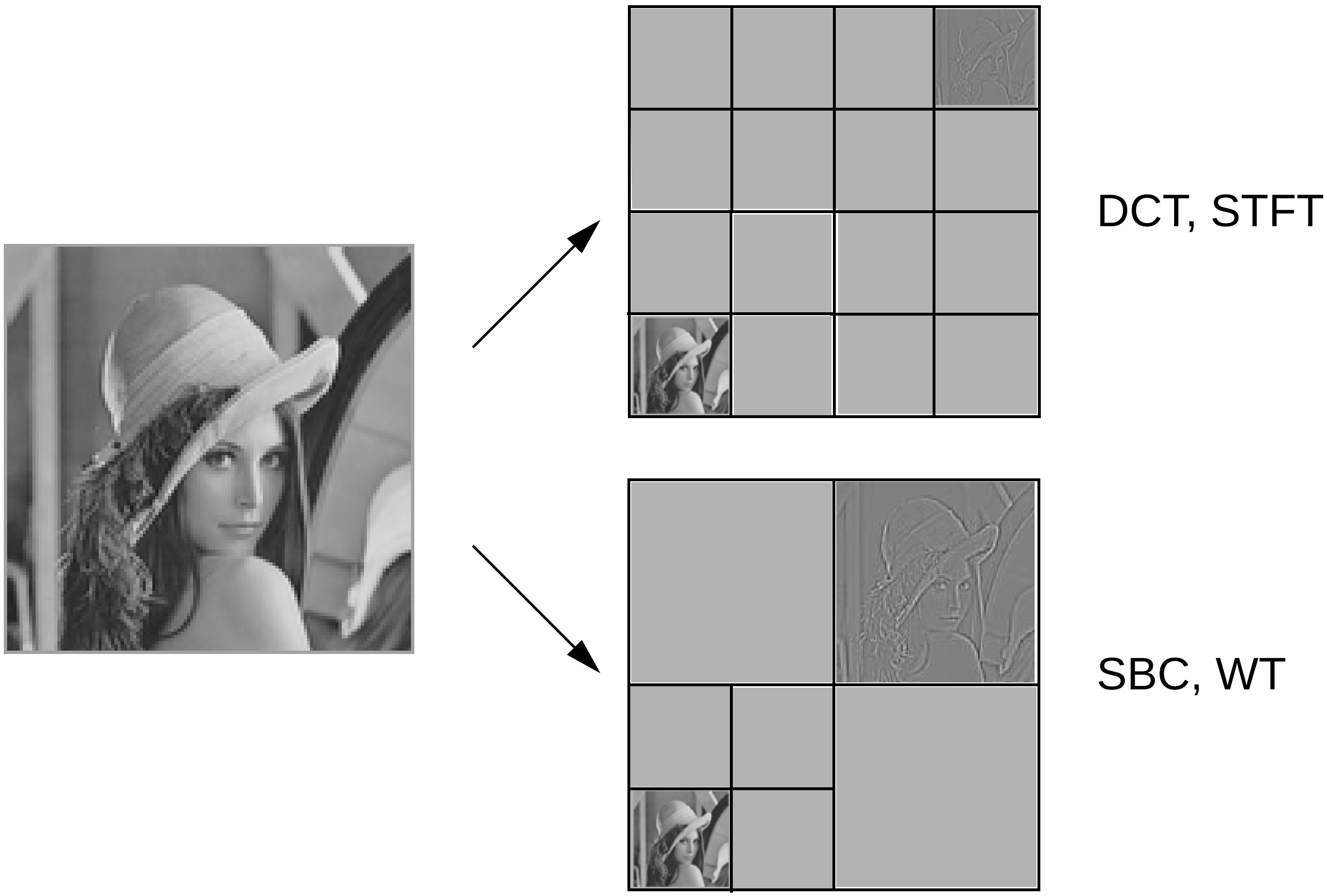
Goal: remove built-in redundancy, send only necessary info



- LT: linear transform (KLT, WT, SBC, DCT, STFT)
- Q: quantization
- EC: entropy coding

Linear transforms

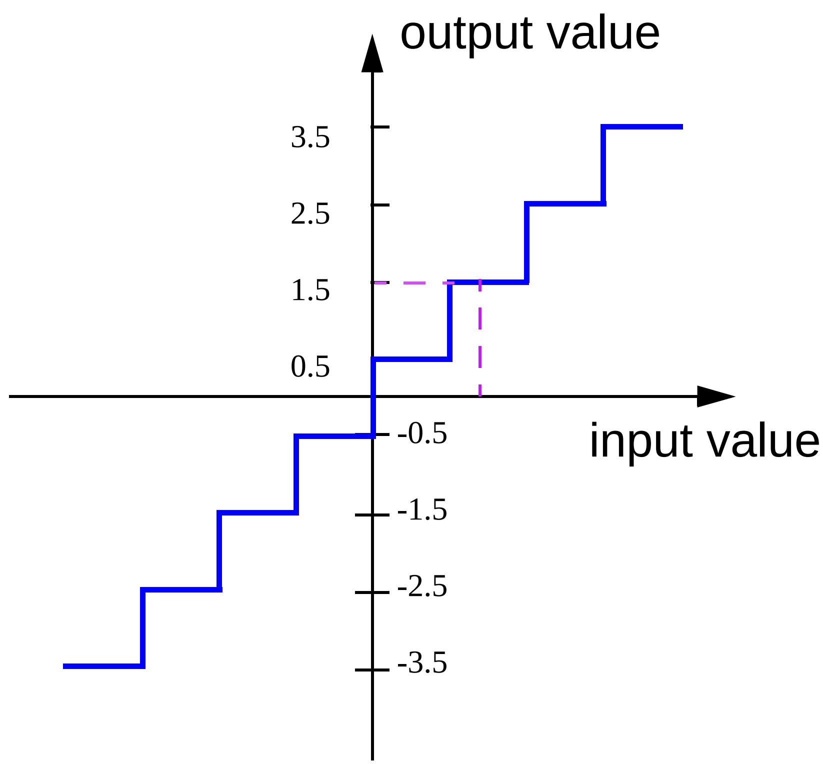
Variation on a theme of KLT:
decorrelates data and efficiently packs energy into a small number of coefficients



Quantization

Continuous amplitude set becomes discrete: digital signal

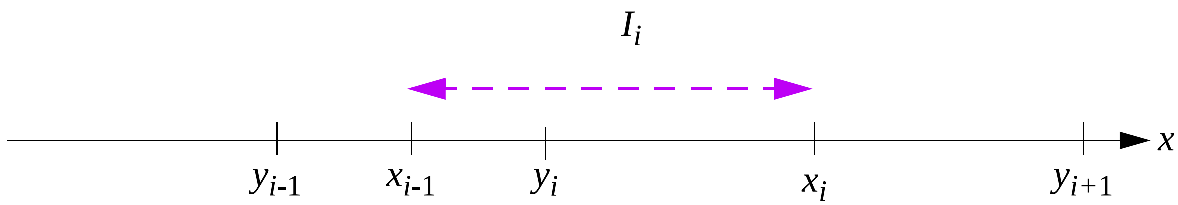
Usually, loss of information occurs only here



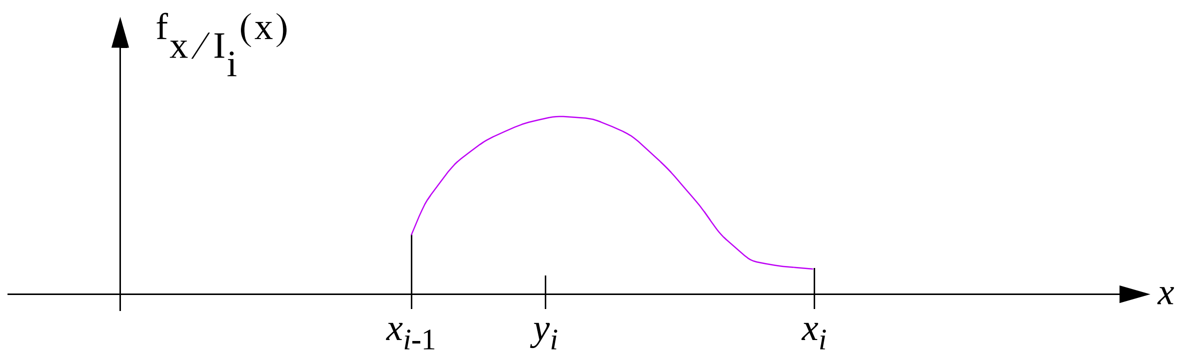
Behavior: $D(R) = c \cdot \sigma^2 2^{-2R}$ **where** D **is MSE,** R **is number of bits and** c **is a constant dependent on distribution**

Quantization ...
... optimality conditions for scalar quantizers

Nearest neighbor condition



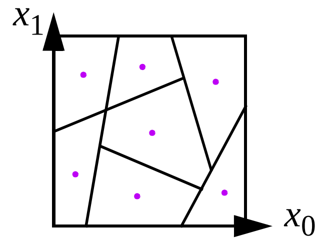
Centroid condition



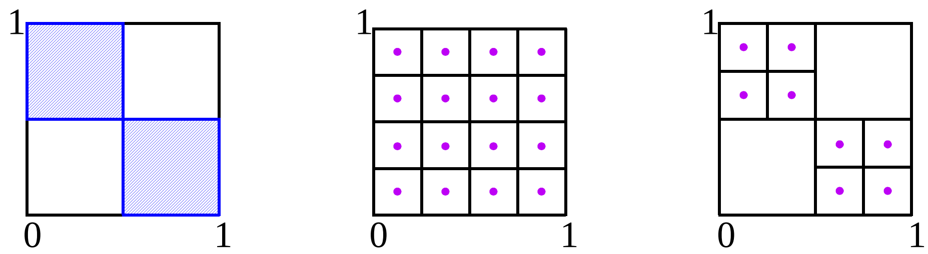
Algorithm: Lloyd-Max

Quantization ...
... vector quantization

Regular vector quantizer in two dimensions



Comparison of scalar and vector quantizers



two-dimensional probability density function

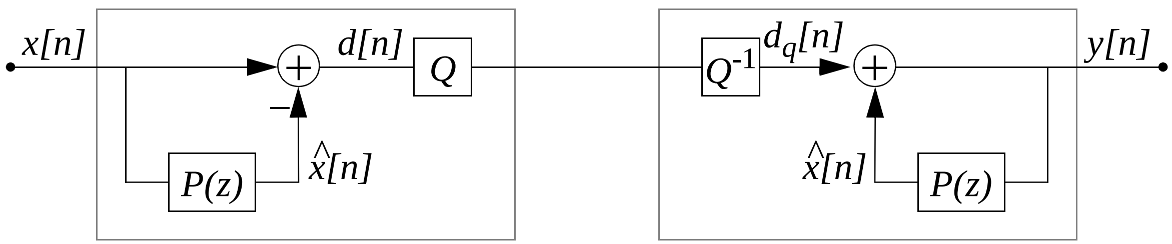
scalar quantizer 2.0 bits/sample

vector quantizer 1.5 bits/sample

Quantization...
... predictive quantization

Quantize the differences instead of the samples themselves

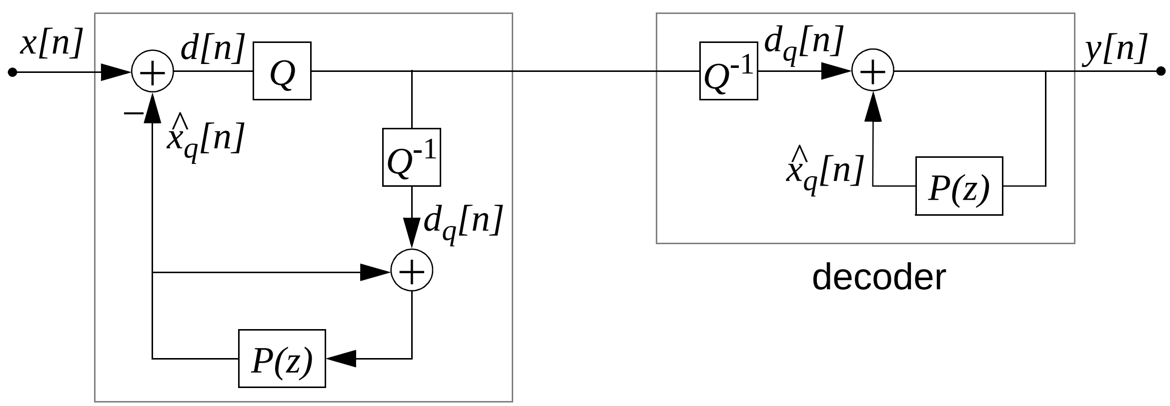
- open-loop predictive quantization



encoder

decoder

- closed-loop predictive quantization: DPCM



decoder

encoder

Quantization ...
... bit allocation

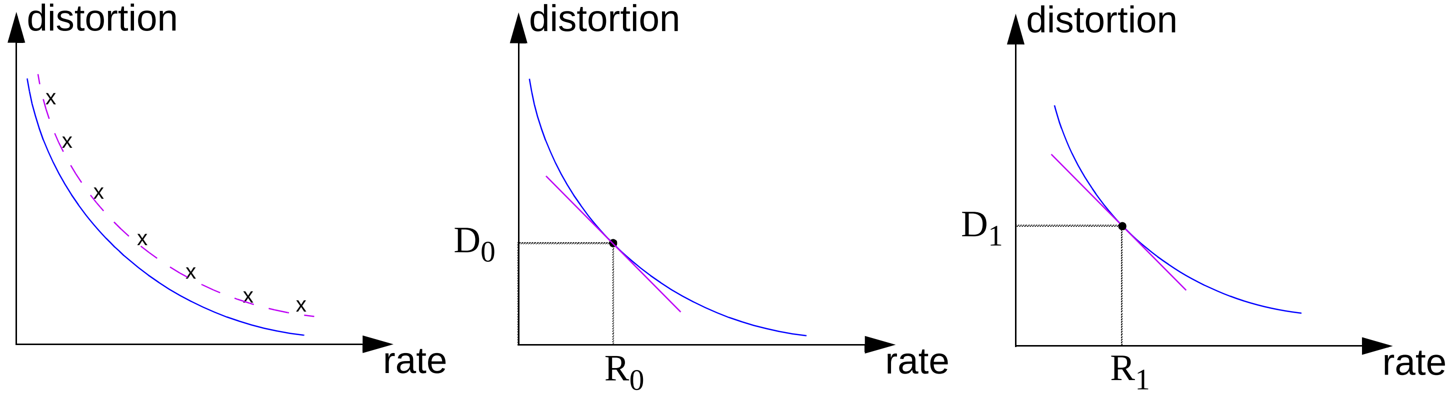
How to choose quantizers for various transform coefficients?

Fundamental trade-off: between rate R **and distortion** D

- rate-distortion theory

Minimize D **given a budget** R

- assume rate and distortion are additive
- use Lagrange multiplier
- we get a constant-slope solution



- $R = R_0 + R_1$

Entropy coding

Represents a sequence of symbols as a bit-stream

Example: Huffman coding

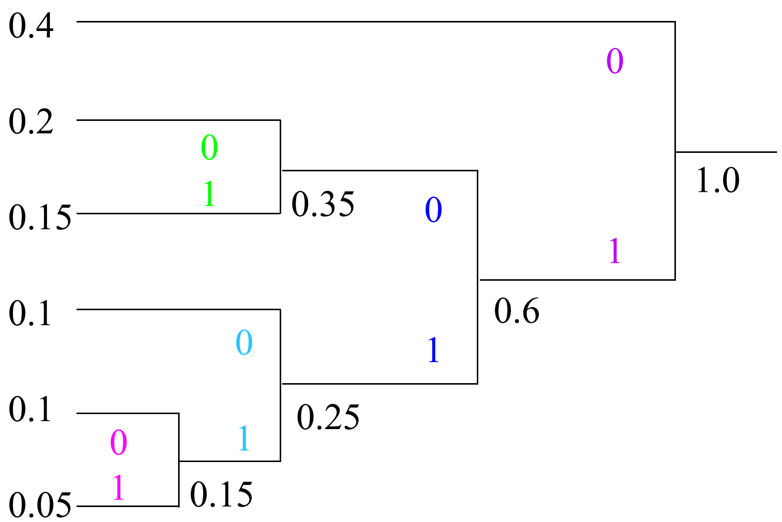
1 a_i	$p(a_i)$	b_i
0	0.40	0
1	0.20	100
2	0.15	101
3	0.10	110
4	0.10	1110
5	0.05	1111

2 a_i	$p(a_i)$	b_i
0	0.40	0
1	0.20	100
2	0.15	101
4+5	0.15	111
3	0.10	110

3 a_i	$p(a_i)$	b_i
0	0.40	0
3+4+5	0.25	11
1	0.20	100
2	0.15	101

4 a_i	$p(a_i)$	b_i
0	0.40	0
1+2	0.35	10
3+4+5	0.25	11

5 a_i	$p(a_i)$	b_i
1+...+5	0.60	1
0	0.40	0



Transform coding and the KLT

Best linear approximation

Result: Given an orthonormal basis for a space S, $\{g_n\}$

$$f = \sum_n \langle f, g_n \rangle \cdot g_n$$

the best linear approximation is given by the projection onto a fixed subspace of size M

$$\hat{f}_M = \sum_{n=0}^{M-1} \langle f, g_n \rangle \cdot g_n$$

The error is thus (MSE)

$$\hat{\varepsilon}_M = \|f - \hat{f}\|^2 = \sum_{n=M}^{\infty} |\langle f, g_n \rangle|^2$$

Transform coding and the KLT

Vector processes

$$X = [X_0, X_1, \ldots, X_{N-1}]^T \qquad E[X_i] = 0$$

$$E[X \cdot X^T] = R_X$$

Consider linear approximation in a basis

$$\hat{X}_M = \sum_{n=0}^{M-1} \langle X, g_n \rangle \cdot g_n \qquad M < N$$

Then:

$$E[\hat{\varepsilon}_M] = \sum_{n=M}^{N-1} \langle R_X, g_n \rangle \cdot g_n$$

Result: For $0 < M < N-1$, the expected error is minimized for the basis $\{g_n\}$ where g_m are the eigenvectors of R_X ordered in order of decreasing eigenvalues.
This is the Karhunen-Loeve transform (KLT).

Proof: Eigenvector argument inductively

Transform coding and the KLT

Jointly Gaussian vector process

- X is jointly Gaussian
- fine scalar quantization of transform coefficients

$$D(R) = c \cdot \sigma^2 \cdot 2^{-2R}$$

- optimal bit allocation between transform coefficients

Result: Among all orthogonal fixed transforms, the KLT achieves lowest overall distortion in MSE sense.

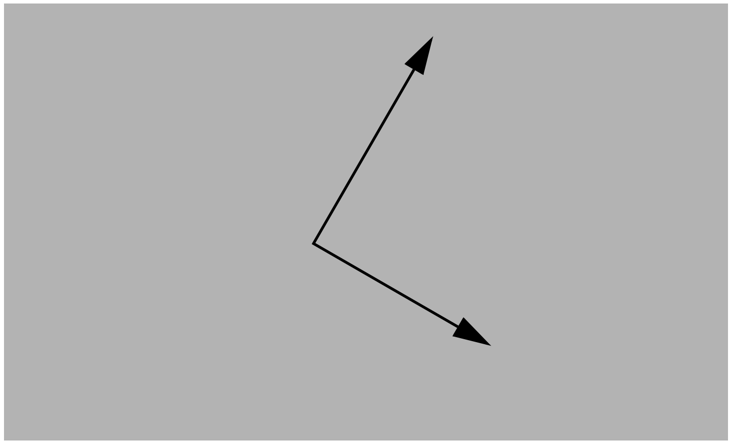
Proof: Determinant inequality (see [GG92])

Coding gain: Given variances of transform coefficients

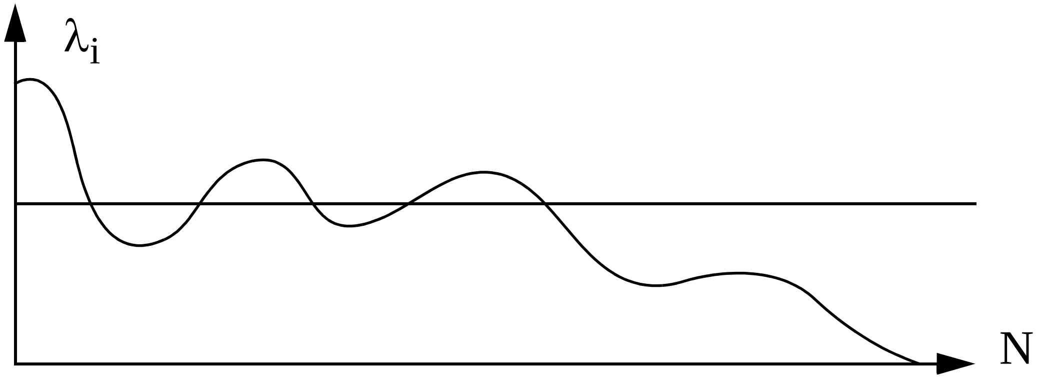
$$\frac{D_{PCM}}{D_{KLT}} = \frac{1/N \cdot \sum \sigma_i^2}{(\prod \sigma_i^2)^{1/N}} = \frac{\text{arith\_mean}}{\text{geom\_mean}}$$

Transform coding and the KLT

Geometric intuition: principal axes of distribution



Theorem: [Waterpouring] As first approximation, this is quantizing uniformly all coefficients above threshold



This can be used to show $D(R)$**, assuming vector coding**

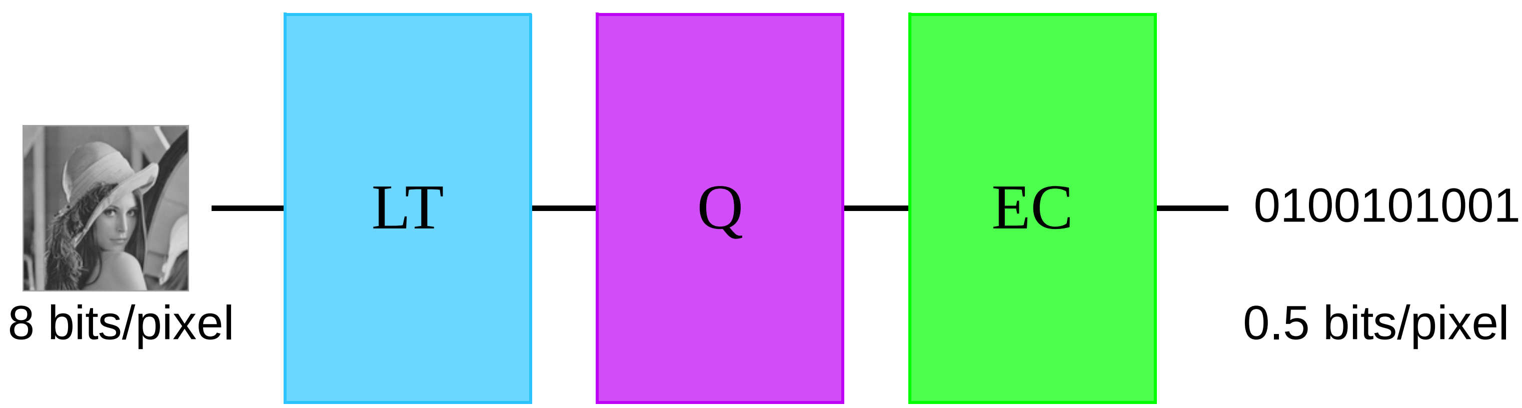
Transform coding and the KLT

The optimality of KLT is usually invoked as the basis for practical transform coding:

Assume process is first-order Markov with correlation ρ

$$R_X = \sigma^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \dots \\ \rho & 1 & \rho & \dots \\ \rho^2 & \rho & 1 & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

then the DCT is a close approximation to the KLT for ρ close to 1 and/or large block sizes



Speech compression

Good production model is available

- vocal cords produce excitation: pulse/noise-like
- vocal tract, mouth and lips act as a filter

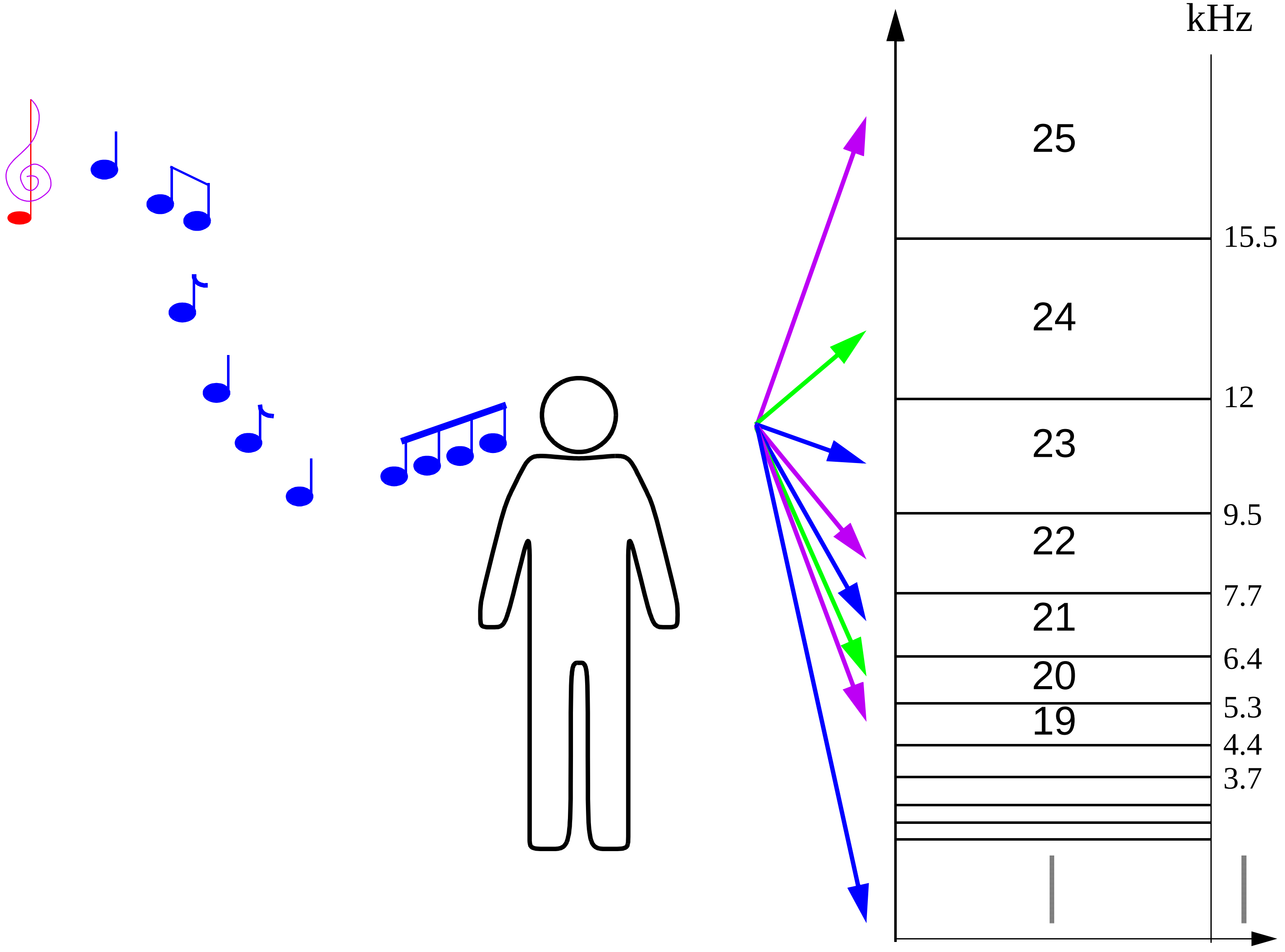
LPC: find an inverse filter, then examine residual

- for voiced/unvoiced Í small number of coefficients

Good results

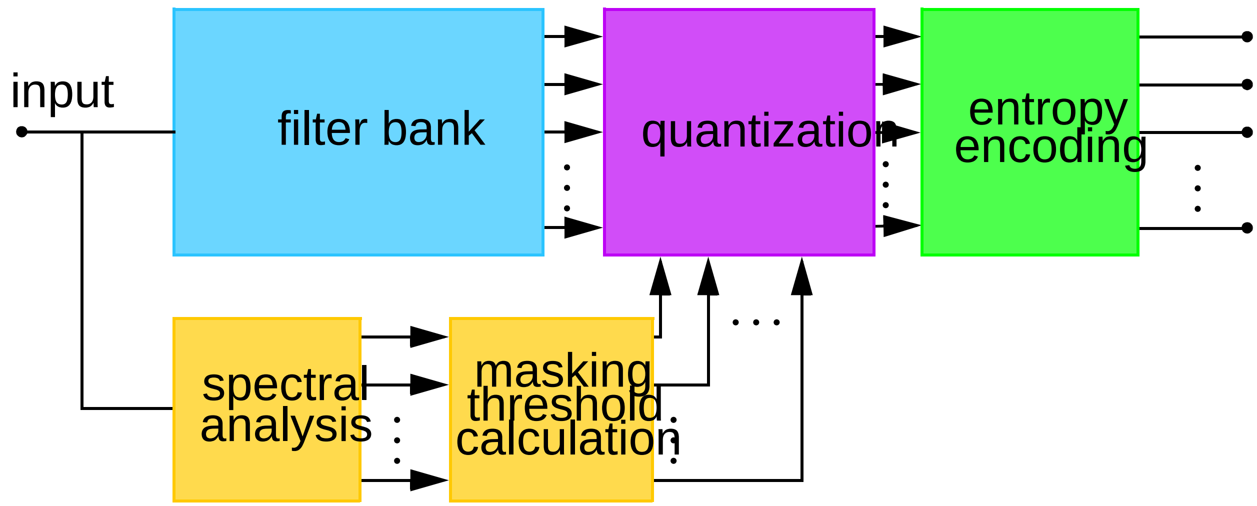
- speech at 64 kbits/sec Í 2.4 kbits/sec
  (sampled at 8 kHz)
- high-quality speech: sampled at 14 kHz
  (similar to audio, more emphasis on perception)

Audio compression



Audio compression

Generic perceptual coder

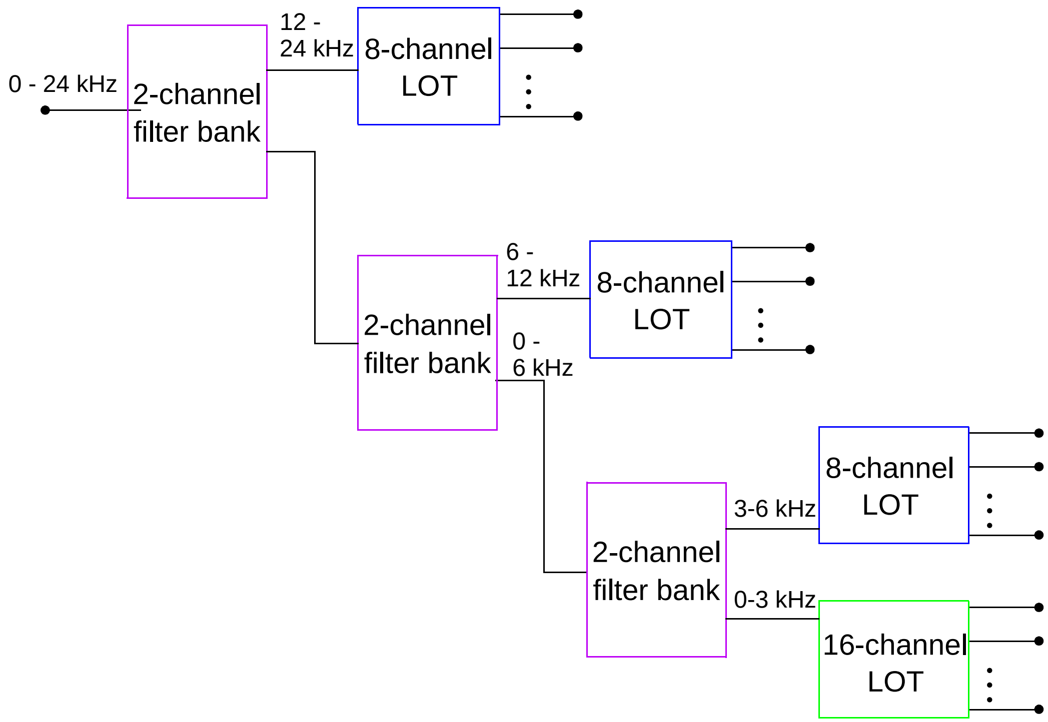


Examples

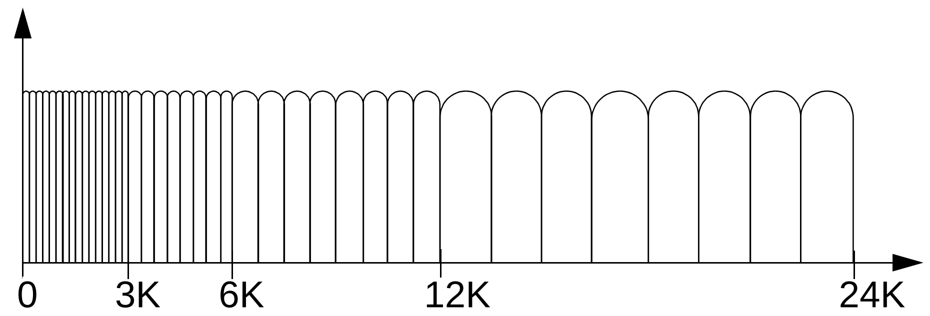
- MUSICAM: known as MPEG-I, 32-band uniform FB
- PAC: intended for DAB, cosine-modulated FB
- Dolby: in NTSC multichannel, cosine-modulated FB

Audio compression

Architecture

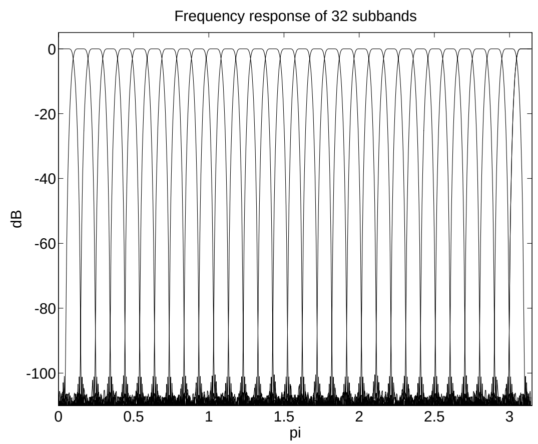


Frequency resolution



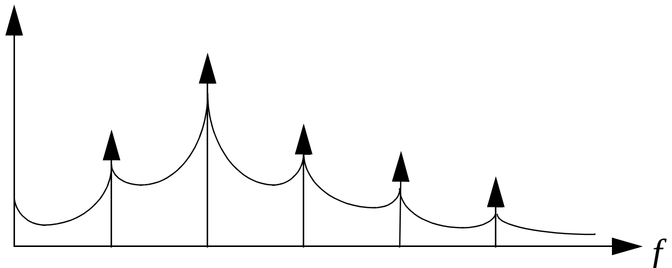
Audio compression ...
... MUSICAM

32-channel modulated filter bank

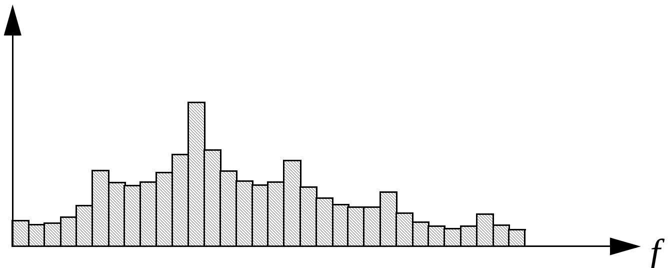


magnitude response

Quantization based on psychoacoustics

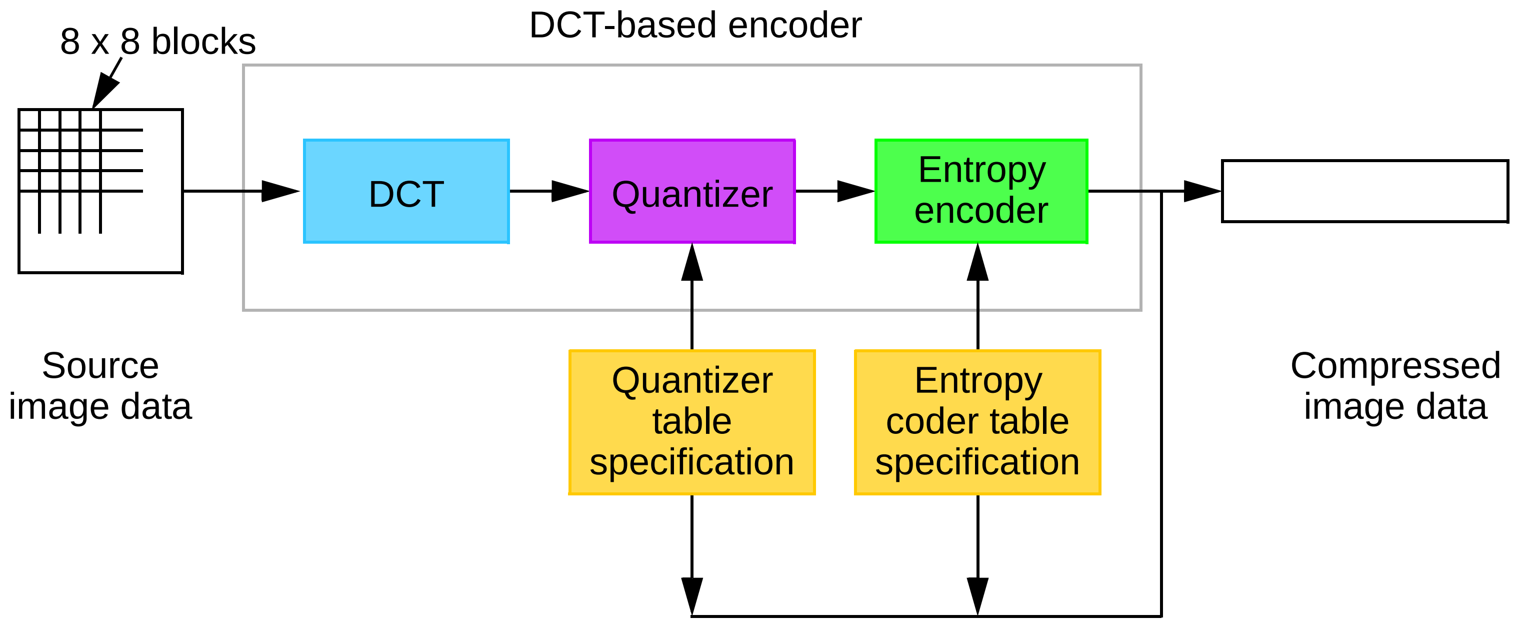


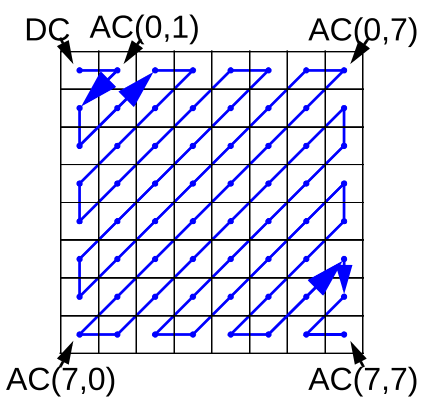
line spectrum and masking

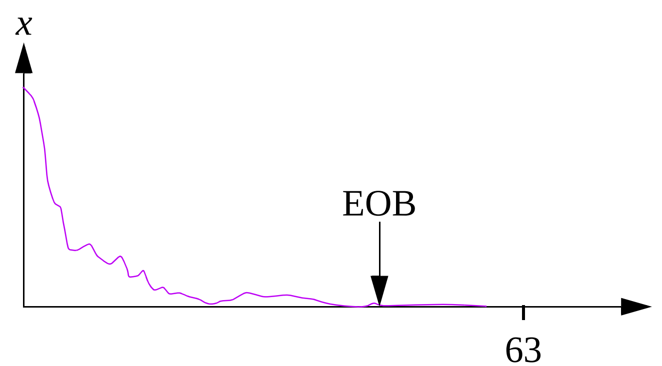


quantization noise

Image compression ...
.. JPEG







zig-zag scanning of 8x8 DCT coefficients

Next image coding standard... JPEG 2000

All the best coders based on wavelets

- 24 full proposals and a few partial ones
- 18 used wavelets, 4 used DCT and 5 used others
- top 75% are wavelet-based
- top 5 use advanced wavelet oriented quantization
- systems requirements ask for multiresolution

Final JPEG 2000 is wavelet based

Image compression ...
... pyramid scheme

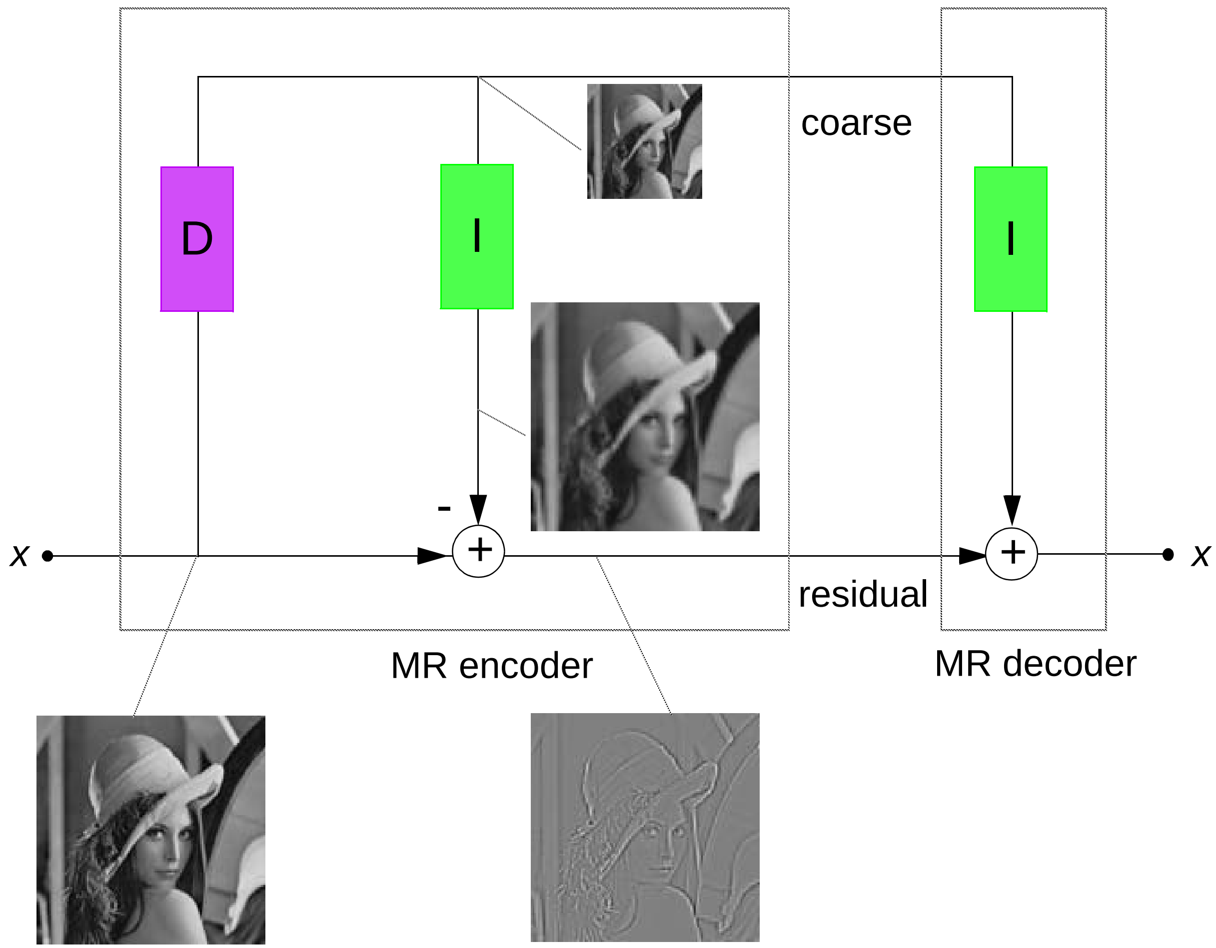
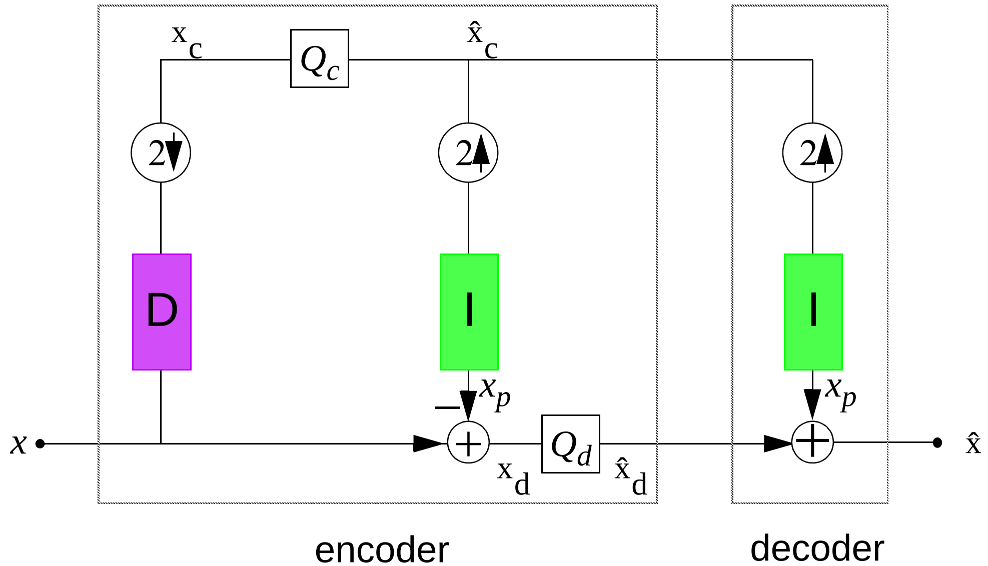


Image compression ...
... pyramid scheme

One-step pyramid coding



Advantages

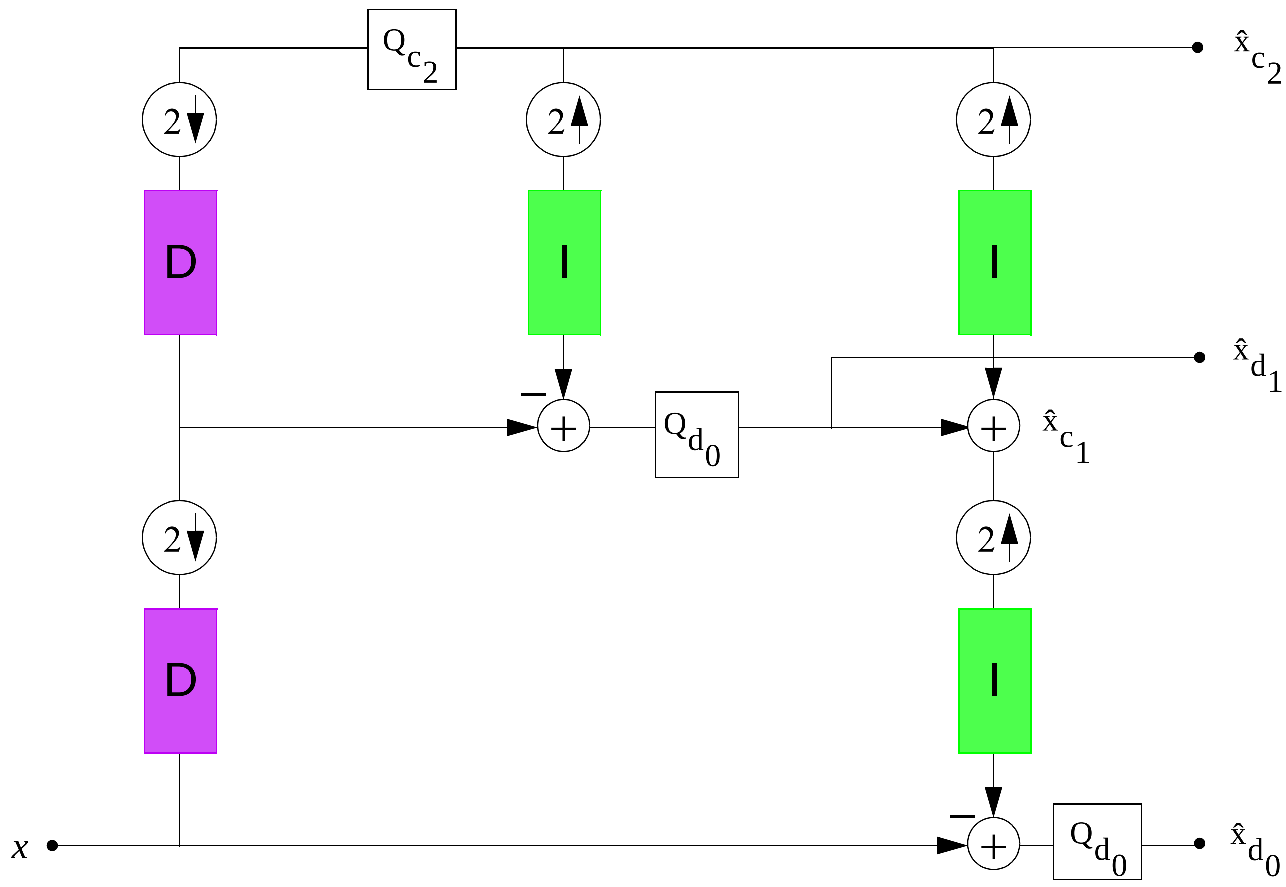
- robustness
- decimation and interpolation operators: arbitrary

Drawbacks

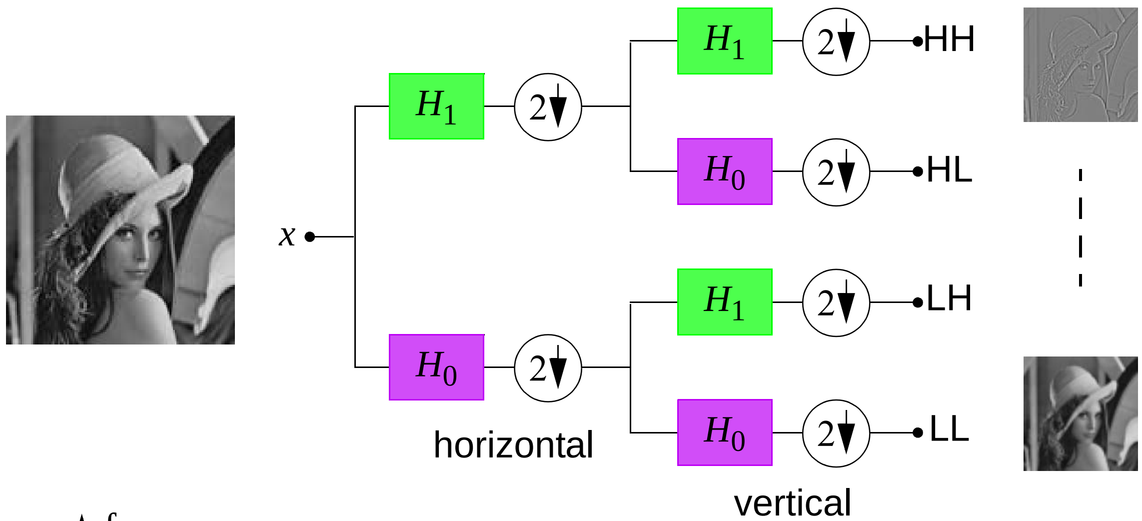
- not critically sampled

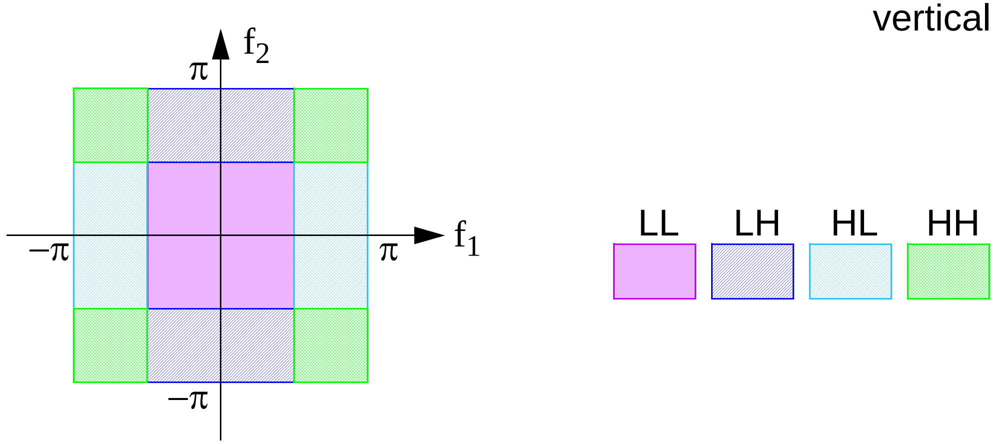
Pyramid scheme ...
... quantization noise analysis

Source of quantization noise limited to the last quantizer



Subband coding of images

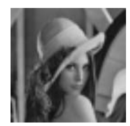




One can introduce perceptual criteria

Subband coding of images...
... typical subband outputs

162	162	161	159	162	161	158	159
162	162	161	159	162	161	158	159
162	162	160	159	162	160	158	159
162	160	159	159	158	159	156	154



LL

162	160	161.5	158.5
161.5	159.25	159.75	156.75

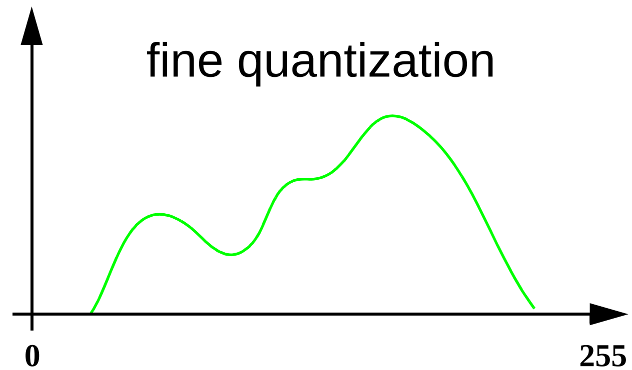
HH

0	0	0	0
-0.5	0.25	0.75	-0.75

Subband coding of images...
... how to compress?

LL

162	160	161.5	158.5
161.5	159.25	159.75	156.75



HH

0	0	0	0
-0.5	0.25	0.75	-0.75

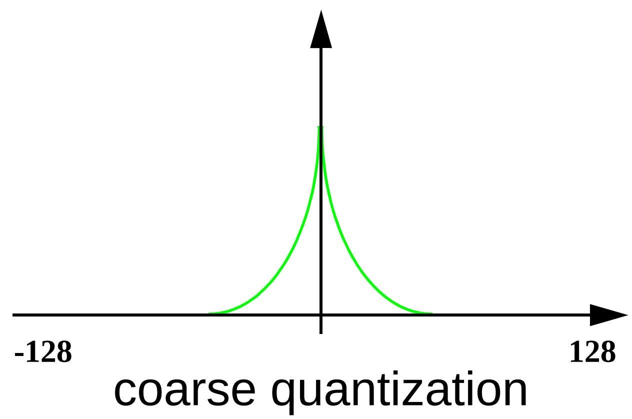
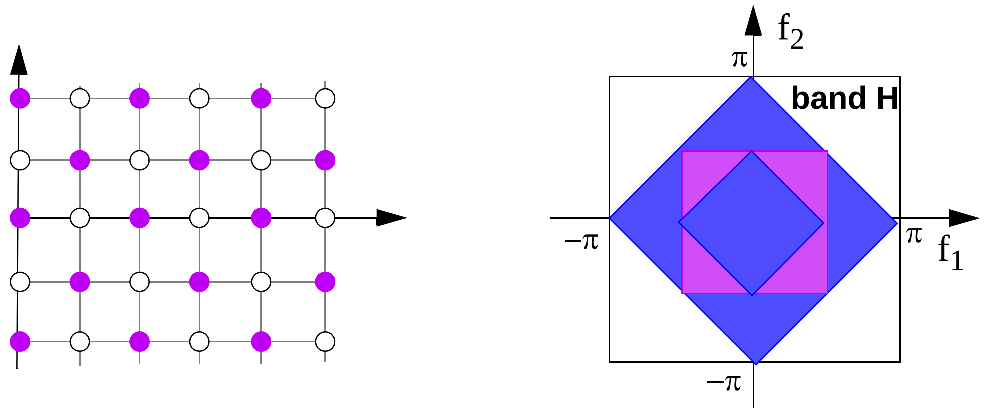
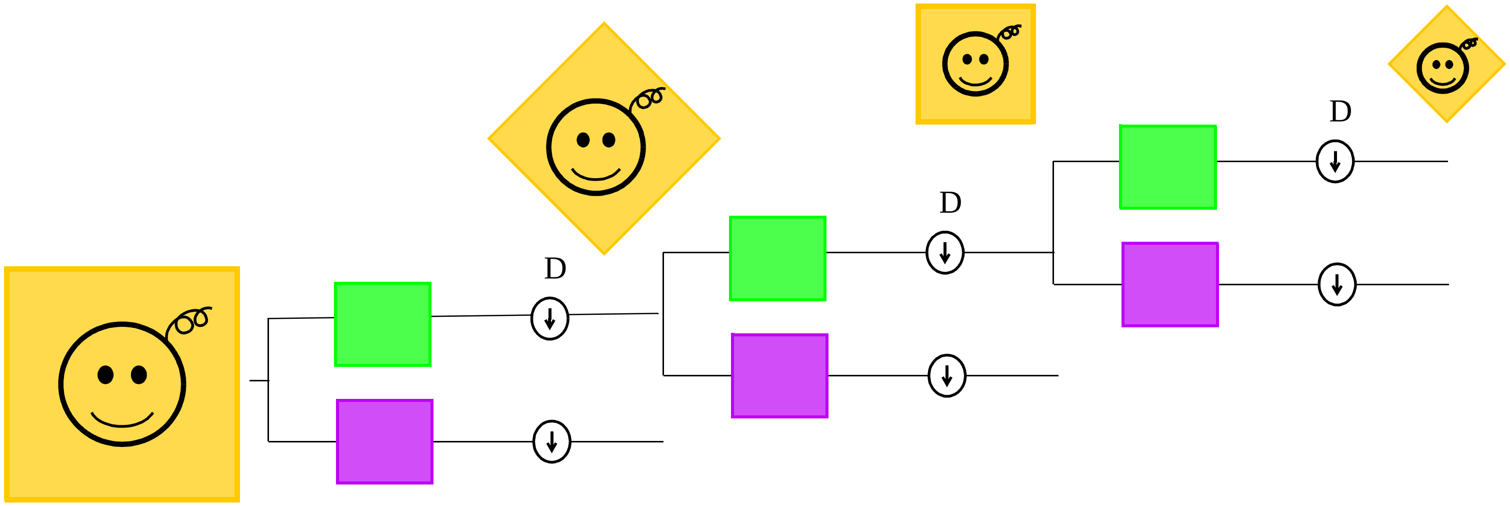


Image coding using quincunx sampling

Nonseparable, data reduction by 2





Successive approximation source coding

Theoretical foundation [Equitz & Cover]

- certain sources can be refined without loss (but not all...)

Practice of source coding: many schemes can be seen as successive approximation

- pyramid coding
- wavelet and subband coding
- tree-structured vector quantization

Standard coding schemes are not successively refinable (at least not efficiently)

- JPEG
- MPEG

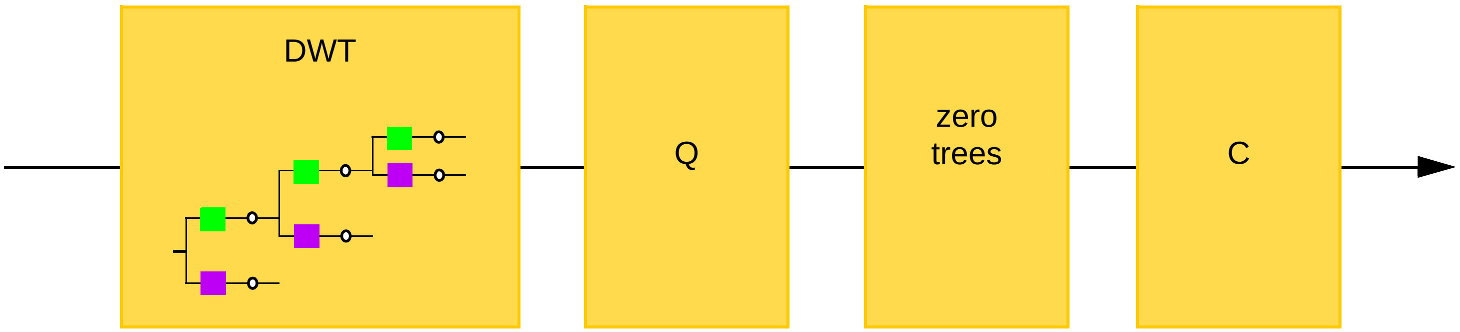
The most successful current image compression schemes are refinable (lucky coincidence?)

- for example, wavelet compression using EZW

Embedded wavelet hierarchical image coder (EZW)

[Shapiro]

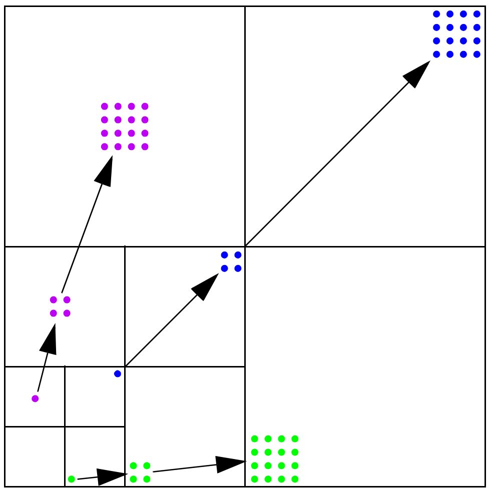
Most state-of-the-art image coders based on this



Features

- bits are ordered in importance
- decoder can cease decoding at any point
- DWT: compact MR representation of significant data
- zerotrees predict insignificance across scales
  if a parent is insignificant, so are all his children
- bit-plane encoding: looks at most significant bits

Embedded wavelet hierarchical image coder ...
... zerotrees



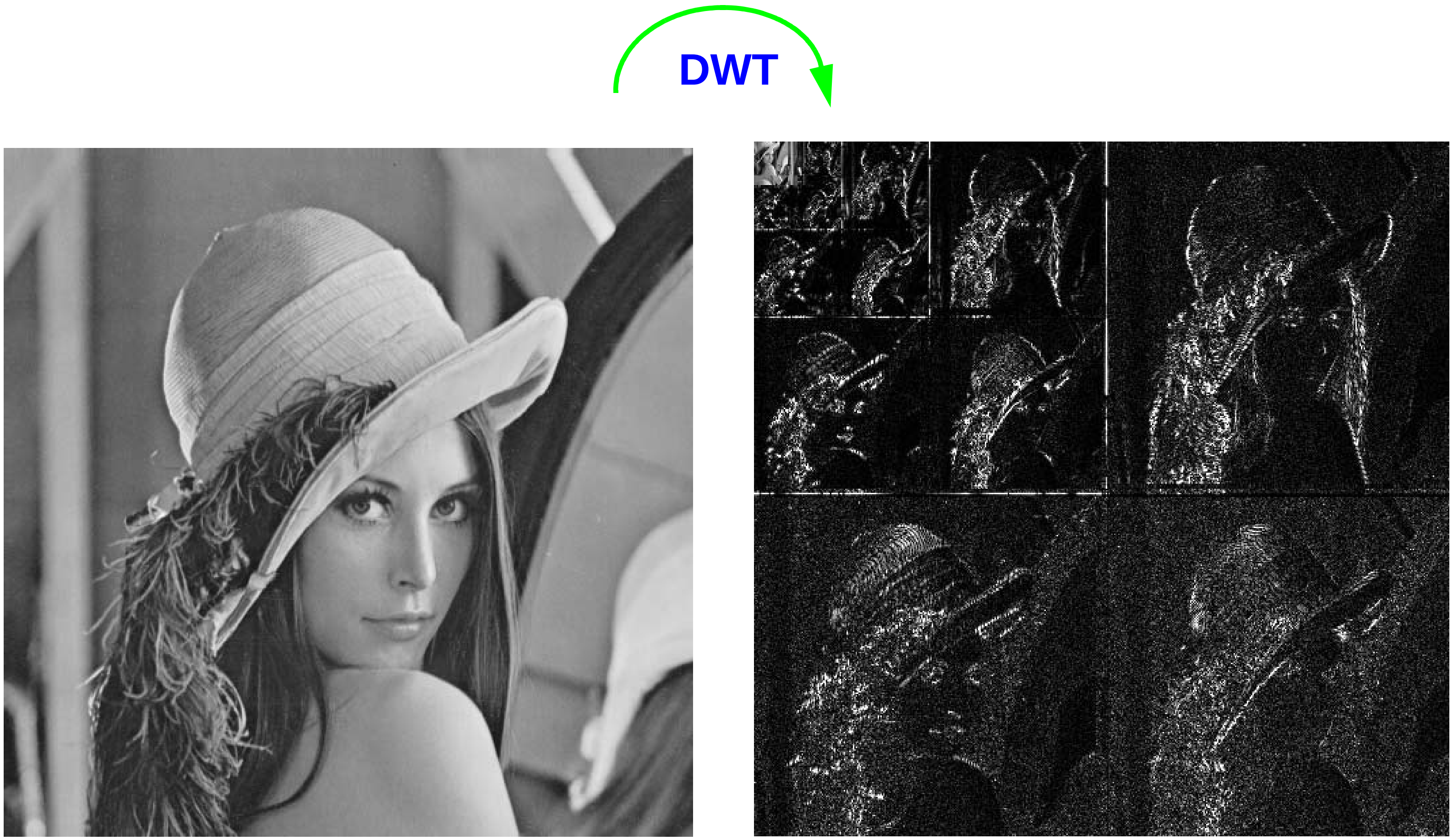
Data structure

- analogous to zig-zag scanning and EOB in DCT
- define a tree of zero symbols starting at the root
- significant high-frequency energy unlikely if there is little low-frequency energy at the same spatial location

- if there is an edge,
  there is probably an edge in all higher bands
  (slow decay)
- if there is no edge $\Rightarrow$ zerotree

Embedded wavelet hierarchical image coder...
... dependence across scales

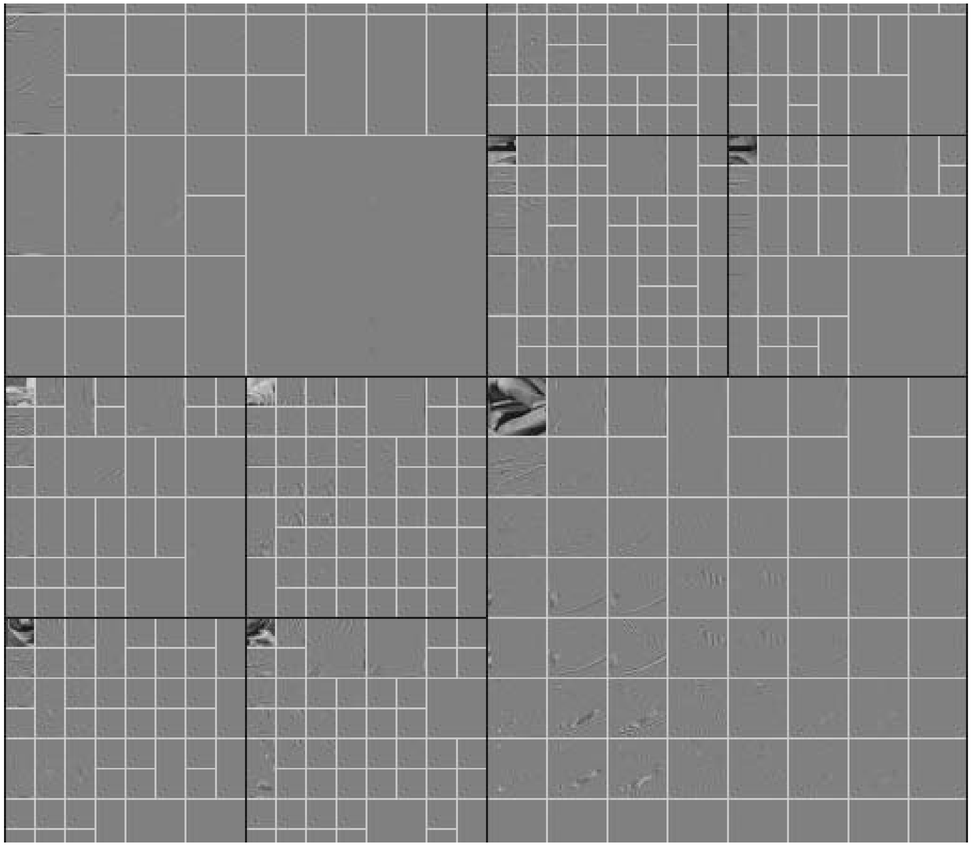
Key insight - dependence across scales



Embedded wavelet hierarchical image coder ...
... extension

Consider splitting in time, not only frequency

- jointly split in frequency and time
  using double-tree algorithm



Embedded wavelet hierarchical image coder ...
... lessons to be learned

The world is not Gaussian!

- decorrelation does not mean independence
- position coding is key
- self-similarity across scales
- most gain from zerotrees or structure across scales
- prediction of position of features

Strong interactions

- transform/quantization/entropy coding interact!
- in particular, matching quantization and entropy coding gains a lot

Embedded bitstreams make great systems

- progressive transmission
- unequal error protection
- browsing

Universal compression

Problem

- compression of an unknown source (from a class)
- compression of a time-varying source

Goal

- perform as well as if the source statistics
  or the variation of the statistics were known

Application

- deep space photographic compression
- nonstationary sources (e.g. mixed documents)
- flexible and robust compression

Method

- learn the statistics on the fly from the data
- backward and/or forward adaptation

Examples

- lossless compression (Lempel-Ziv, arithmetic coding ...)
- lossy compression (universal transform coding, VQ)

Universal transform coding

[Goyal, Zhuang, Vetterli]

Goal: adaptive KLT

- will learn statistics on the fly
- will track time-varying sources
- can be backward adaptive or forward adaptive

Problems

- can it be done in a backward manner? (yes!)
- performance (speed, accuracy)
- convergence

Results

- LMS type algorithm for backward adaptation
- proof of convergence in certain cases (Gaussian)
- speed of tracking unsatisfactory
  for ‘‘real’’ nonstationary images

What is a ‘‘good’’ basis?

Quality?

- norms L_1, L_2, L_{inf}
- approximation in mean, minimax
- ‘‘perceptual sense’’
- some derived ‘‘quality’’ (e.g. detection)

Harmonic analysis

- convergence properties, function classes

Approximation theory

- speed of convergence, properties of approx. fct, ...

Information theory

- rate-distortion
- limit behavior

Signal processing

- computational complexity
- operational $R(D)$
- ‘‘real’’ signals

Best basis for Lena?

Easy

- first vector is ============$\Rightarrow$
- other vectors: orthogonal

Expansion

- single nonzero component

What is wrong with this picture?

- complexity of the model
- problem has just been shifted to describing the transform...

Yet, there is something to be learned

- adapting the transform can help

Goal: universal compression schemes

Adaptive best bases

Question

given a signal you have never seen before, what is the ''best'' transform to code it?

Fourier versus wavelet bases

- linear versus octave-band frequency scale
- different trade-offs

Wavelet packets

- arbitrary but fixed time/frequency resolution
- algorithm to find the best wavelet packet
- ''nonstationary'' case
- adapt tiling with time:
  binary time-segmentation or arbitrary segmentation
- computational procedures:
  tree pruning and dynamic programming
- adaptive transform
  not linear: $T[a+b]$ not equal to $T[a]+T[b]$
- operator is unitary $T^*T = I$: energy conservation

Adaptive best bases

What is a "good" transform?

- packing property
  thresholding: keep first or largest m coefficients
  rate of decay
- classes of signals
  statistical description: classic framework
  smoothness classes [Donoho, DeVore-Jawerth-Lucie]
- quantization of inner products, entropy coding, $R(D)$

Classes of transforms

- T: unitary $N\mathrm{x}N$, $O[N^2]$ parameters...
- side information
- fast computation of expansion

Cost function which allows finding the best basis is

- additive over disjoint sets (R,D), independence
- basic step: T_a is better than T_b if

$$\mathrm{Cost}(T_a[x]) < \mathrm{Cost}(T_b[x])$$

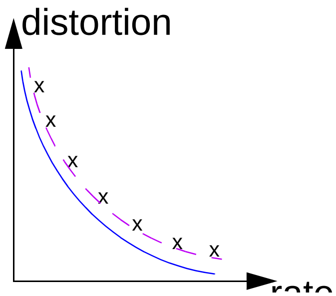
Adaptive best bases

What is an ‘‘optimal’’ tiling?

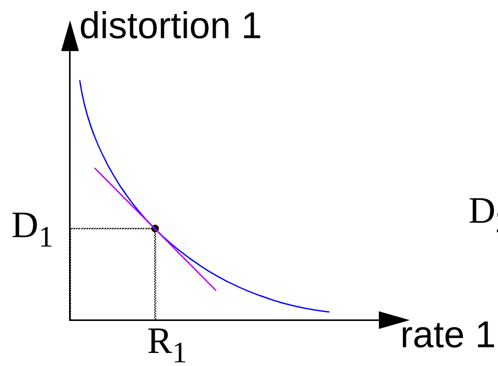
- choose criterion:
  best approximation? (minimum distortion)
  least information? (minimum rate)
- best in a rate-distortion sense:
  minimize $J = D + \lambda R$: two-sided cost
- search over λ such that budget is met

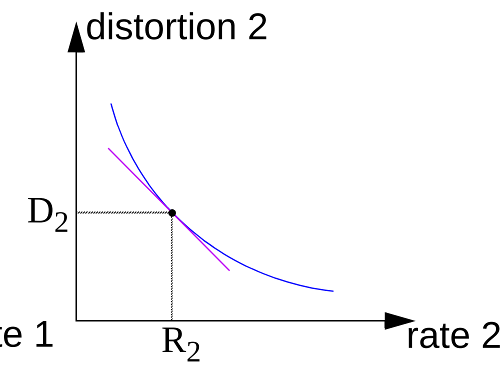
Operational $R(d)$

- easily computed
- ‘‘close’’ to real performance



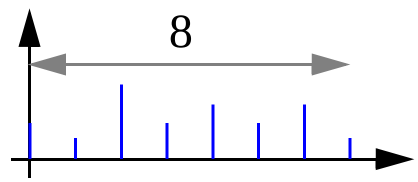
Resource allocation among competing units



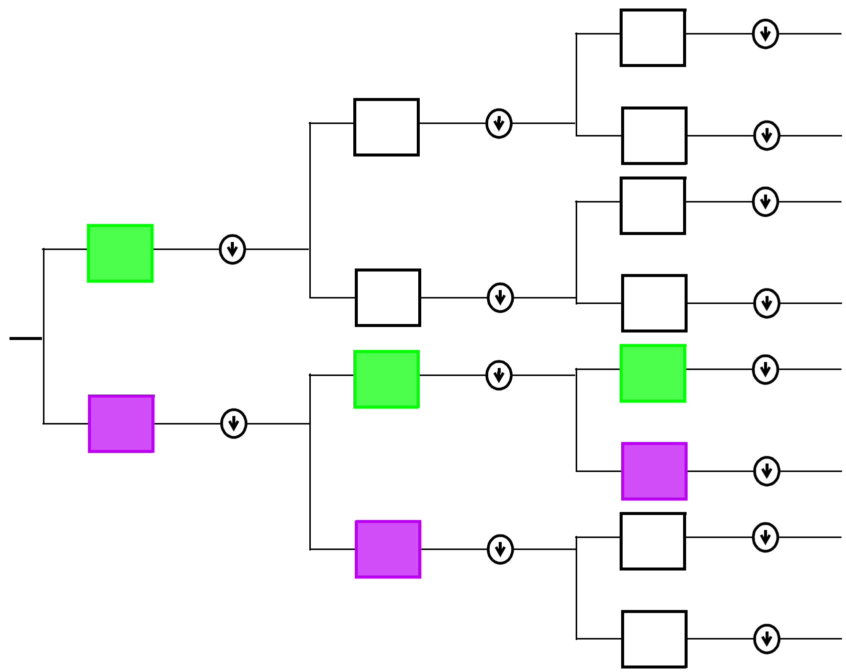


Wavelet packets

[Coifman&Wickerhauser]



Among all possible trees, find the best one



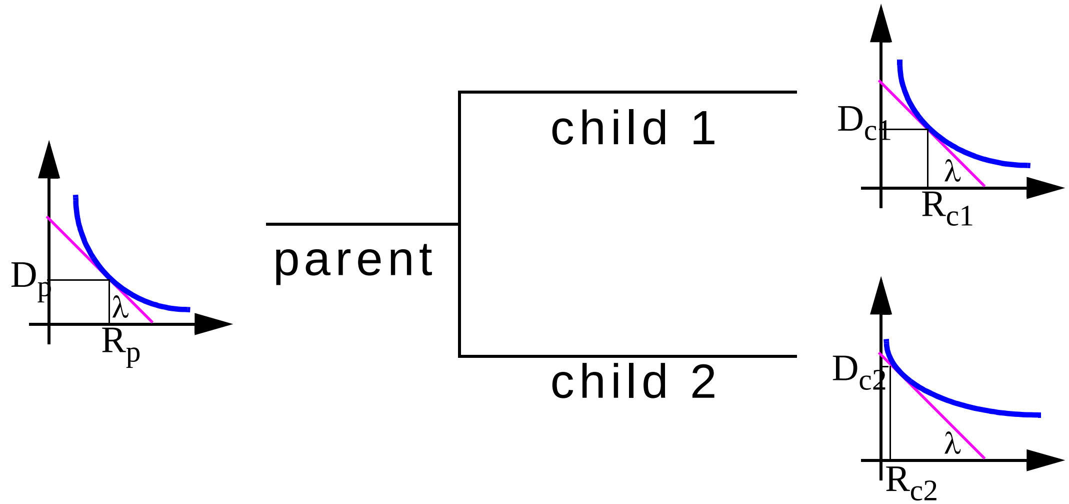
Wavelet packets ...
... adding rate-distortion

[Ramchandran & Vetterli]

Previous scheme: uses entropy as the "goodness" criterion

Extension: for a given bit rate budget,
choose the best basis together with the optimal quantizer

Best basis search: using rate-distortion criteria



For quality λ, prune if

$$(D_{c1} + D_{c2}) + \lambda(R_{c1} + R_{c2}) > (D_p + \lambda R_p)$$

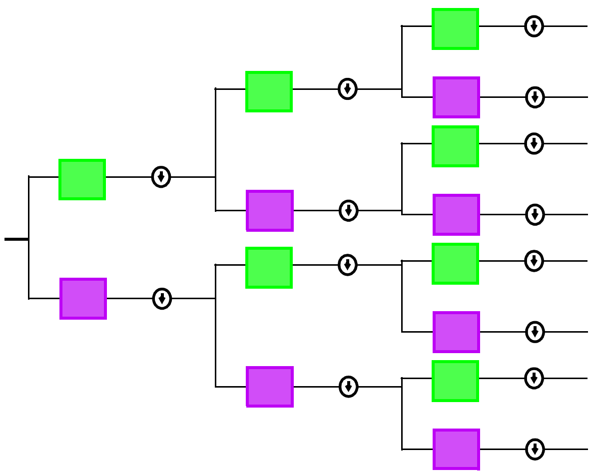
Wavelet packets

Criterion for search: entropy

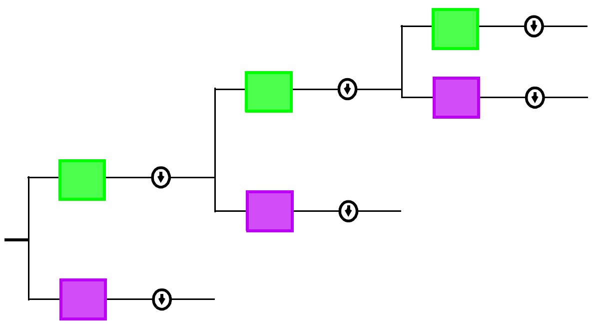
Cut signal into pieces and find best basis for each one

Contains STFT- and wavelet-like schemes as special cases

STFT

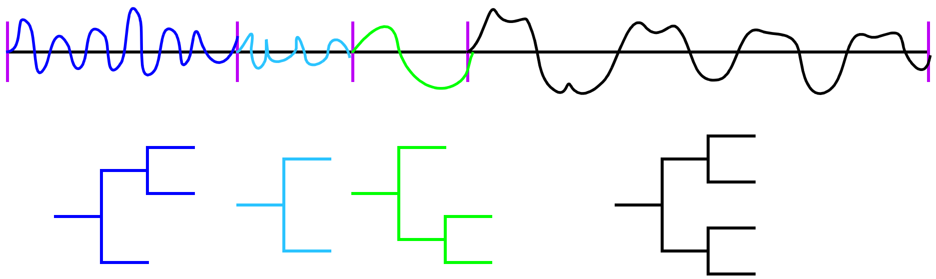


wavelet



Wavelet packets...
... adoptively

Splitting the signal in frequency works, how about time?



Question:
What is the ‘‘best’’ segmentation of a signal TOGETHER with the best wavelet packet in each segment?

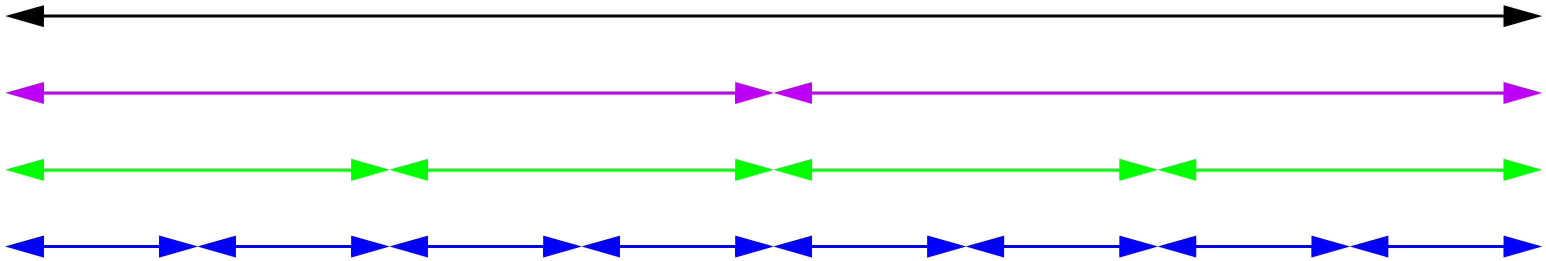
This has to be done jointly!
(not segmentation followed by WP)

Two problems to solve

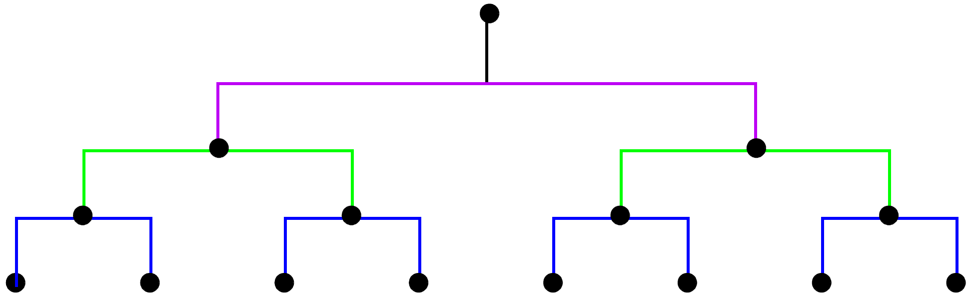
- find time-varying orthogonal filter banks
  boundary filters
  wavelets on the interval
- find algorithms to search for the ‘‘best’’ basis
  double-tree algorithm

Wavelet packets ...
... double-tree algorithm

- start with binary segmentation in time



- for a given λ, find best WP over each segment
- populate a binary tree, prune tree

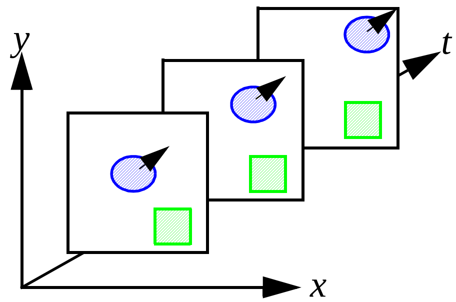


- iterate over λ using bisection
- complexity: $O[N(\log N)^2]$

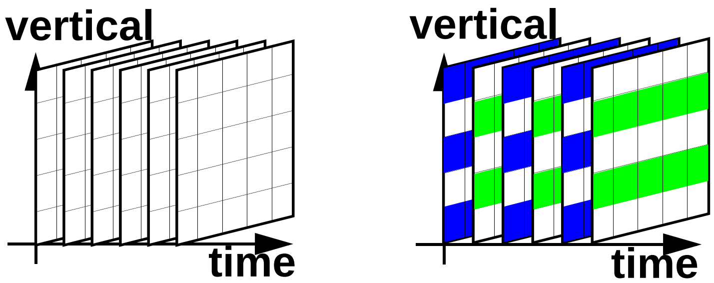
Video compression

Key problems

- motion models



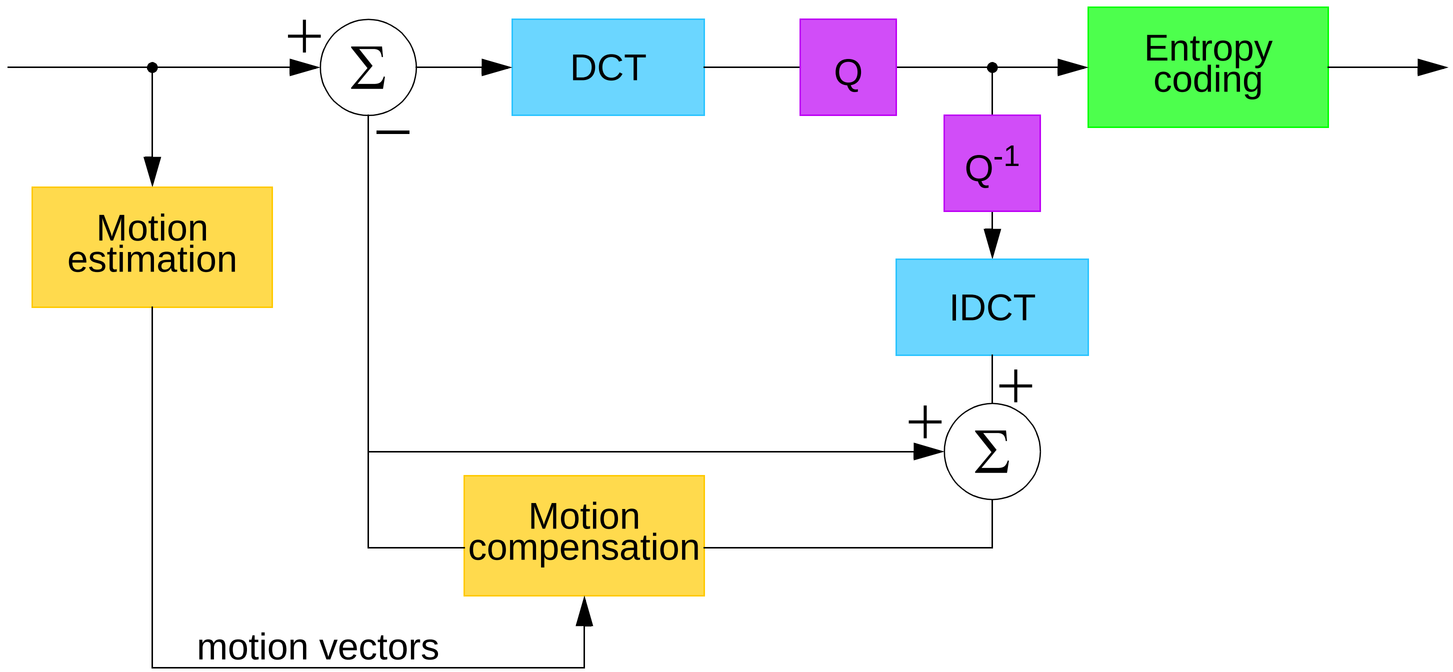
- transform-domain view
- perceptual point of view
- progressive and interlaced scanning



- compatibility

Motion-compensated video coding

Hybrid motion-compensated predictive DCT coder



Part of several standard coding algorithms (e.g. H.263)

Multiresolution coding of HDTV

[Uz,Vetterli,LeGall]

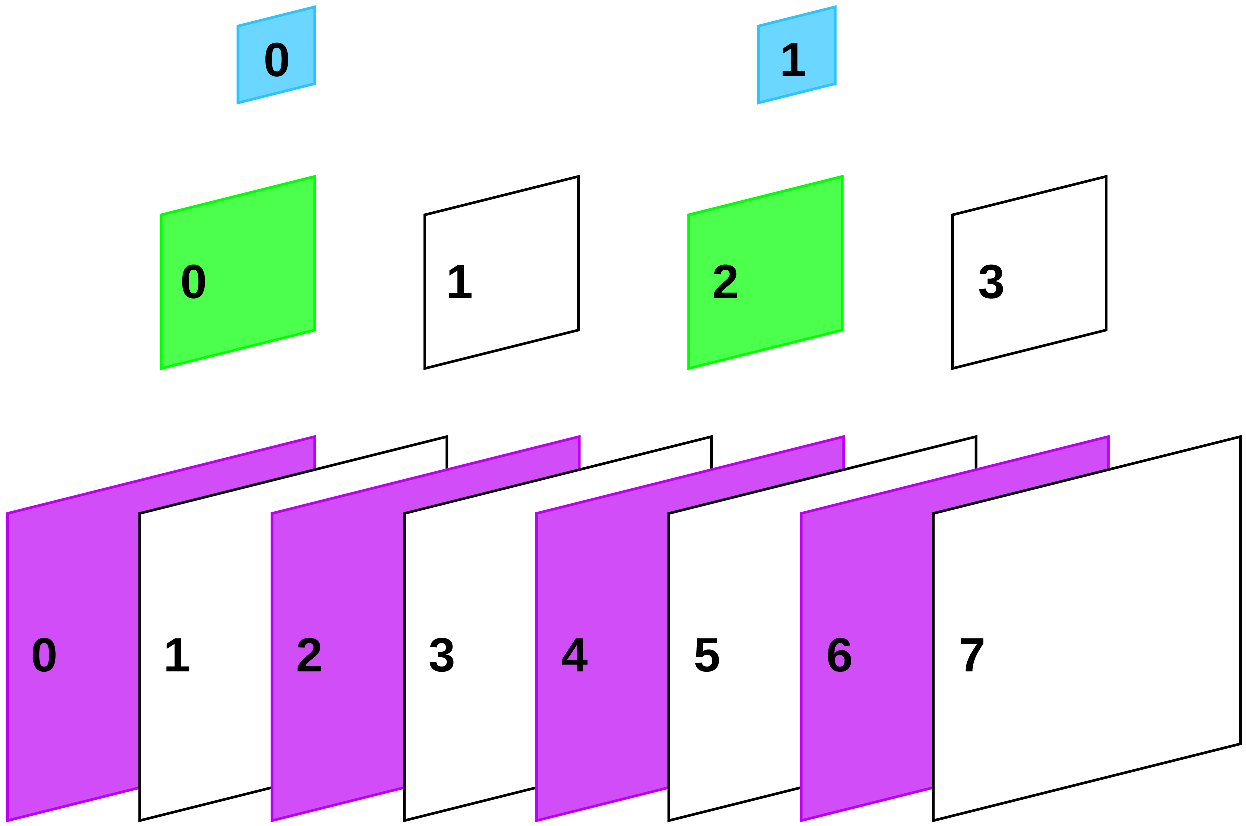
Goal: develop a high-quality coding method with

- signal decomposition for compression
- compatible subchannels
- tight control over coding error
- easy joint source/channel coding
- robustness to channel errors
- easy random access for digital storage

These point toward

- multiresolution scheme
- pyramid coding
  allow for better filters (compatible subchannels)
  allow for inclusion of motion-based techniques
  quantization error easily controlled
  oversampling overhead: 14%

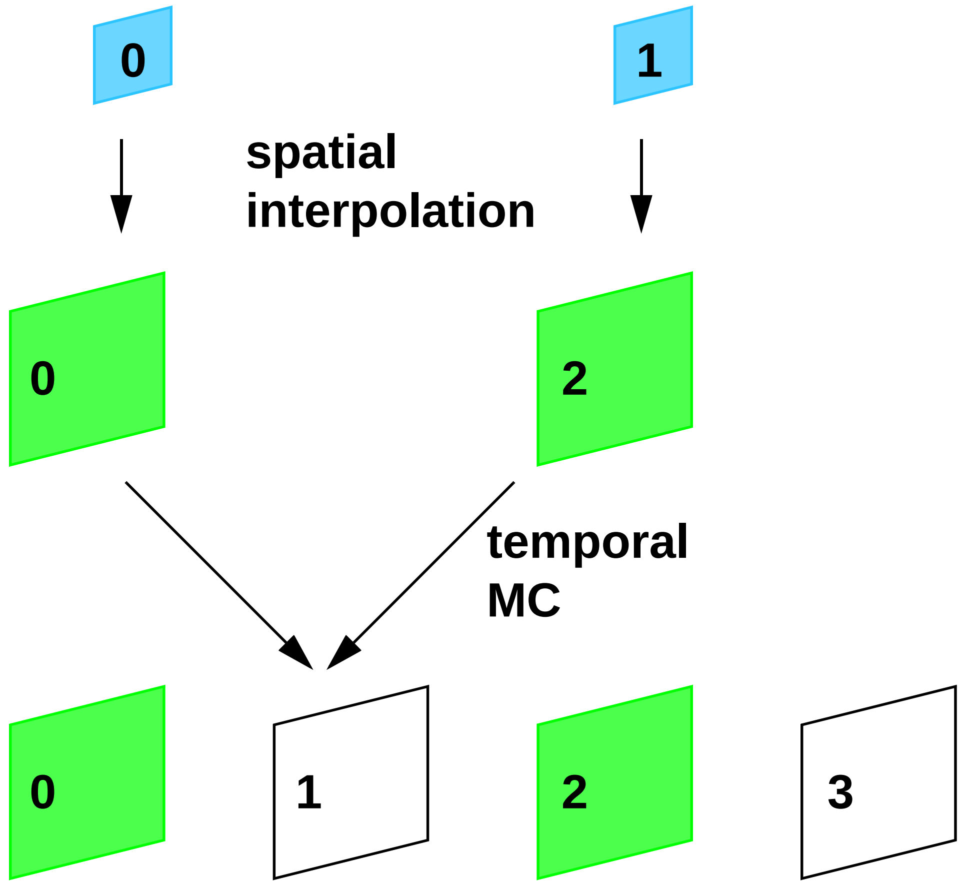
Multiresolution coding of HDTV
... spatio-temporal pyramid



Transmitted:

lowest resolution (1/64) + difference signals + MV

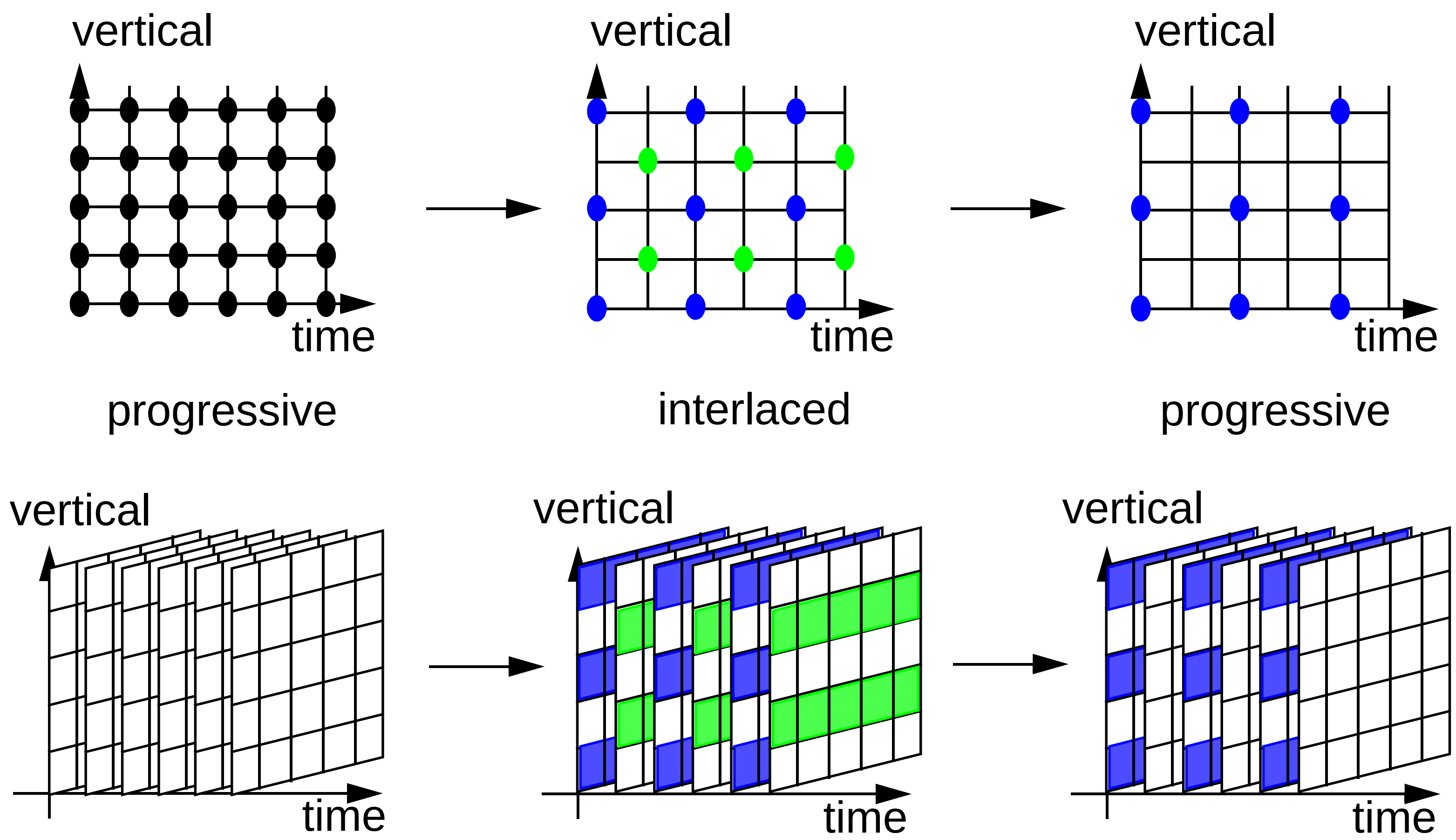
Multiresolution coding of HDTV
... interpolation step



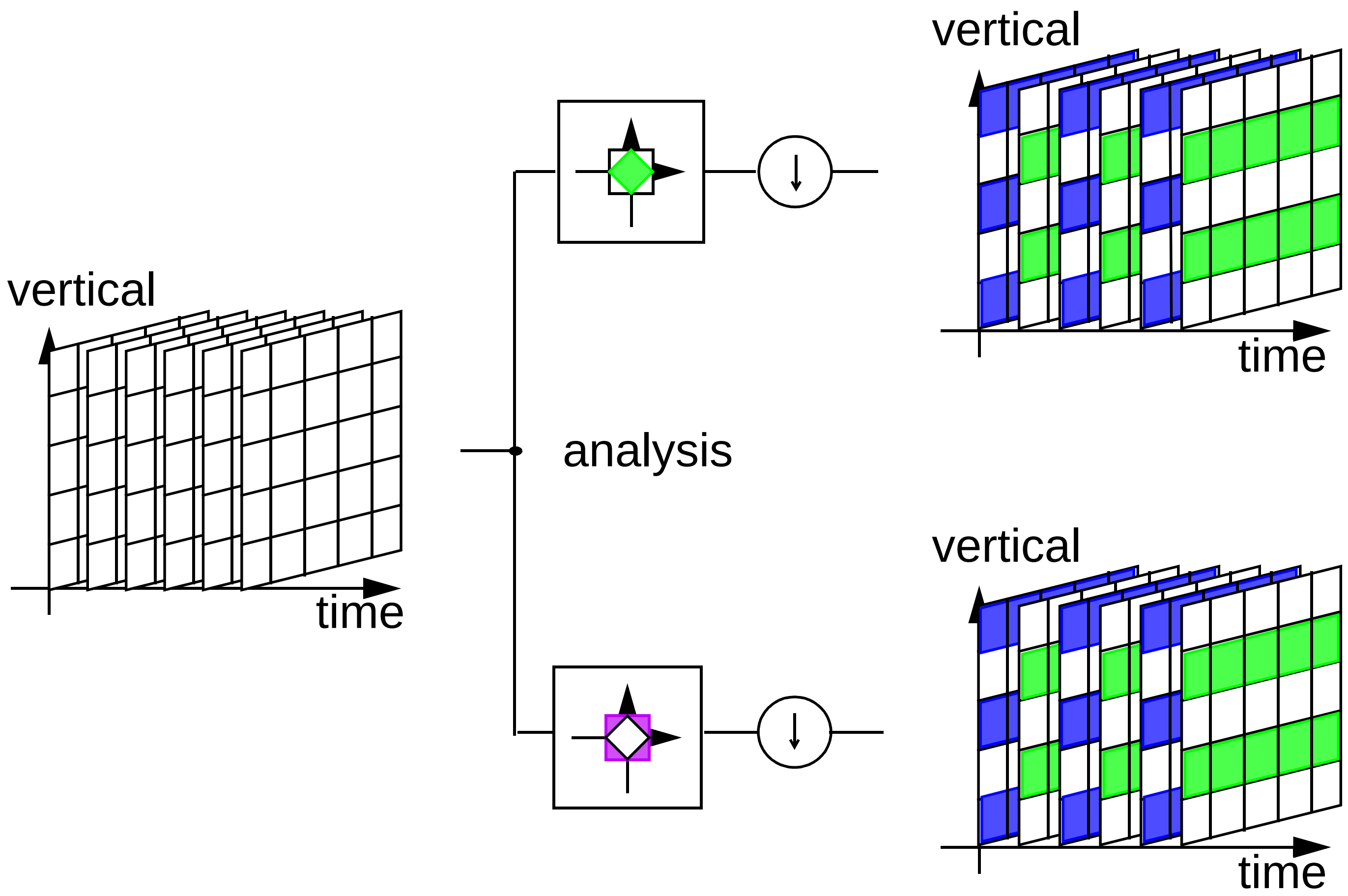
Quincunx sampling for HDTV

[Vetterli,Kovacevic,LeGall]

Nonseparable filtering/sampling

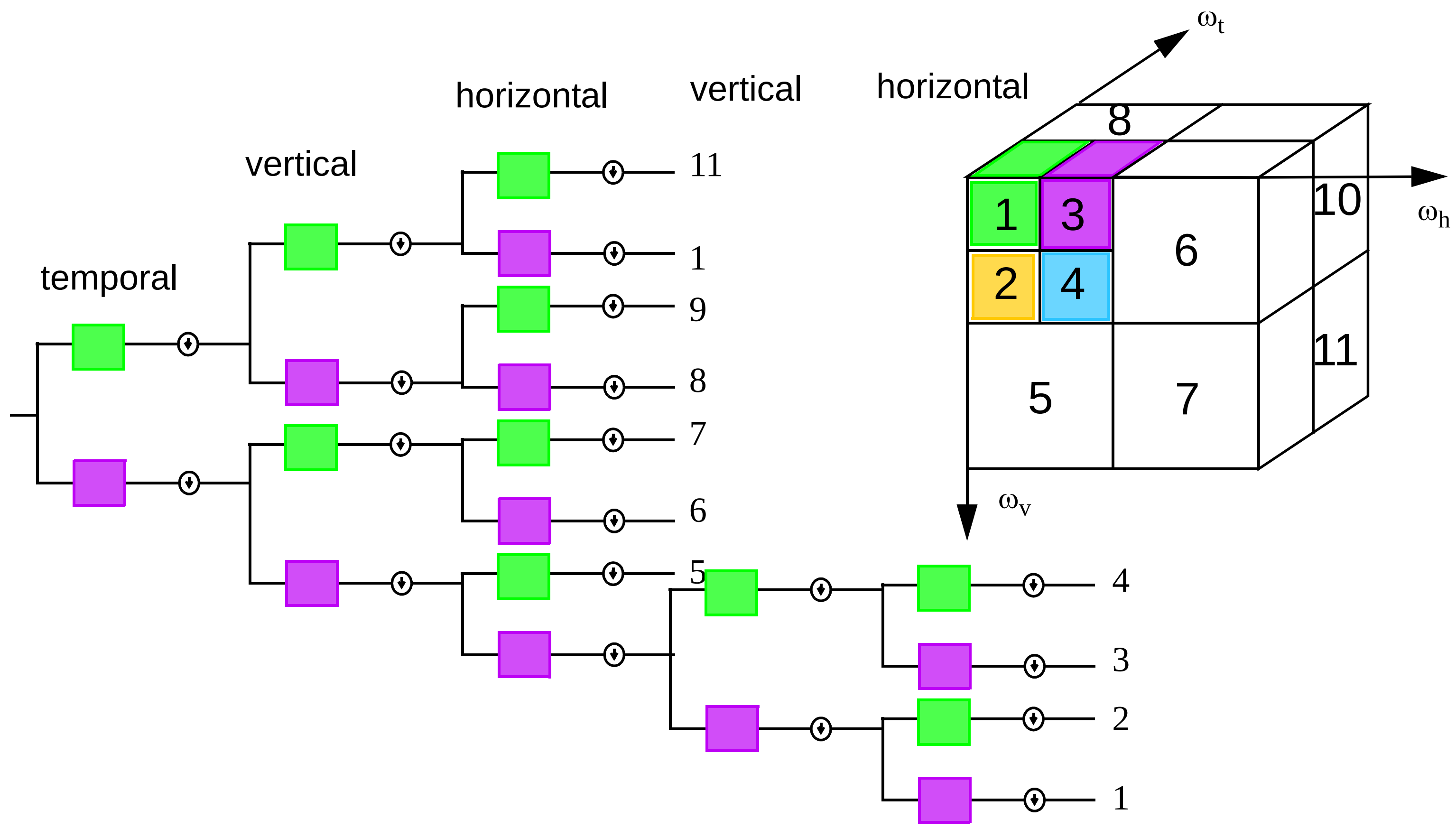


Quincunx sampling for HDTV
... progressive to interlaced



3D subband coding of packet video

[Karlsson & Vetterli]



Separable filtering/sampling

- channel 1: DPCM, highest priority
- channels 2-11: PCM, lower priority

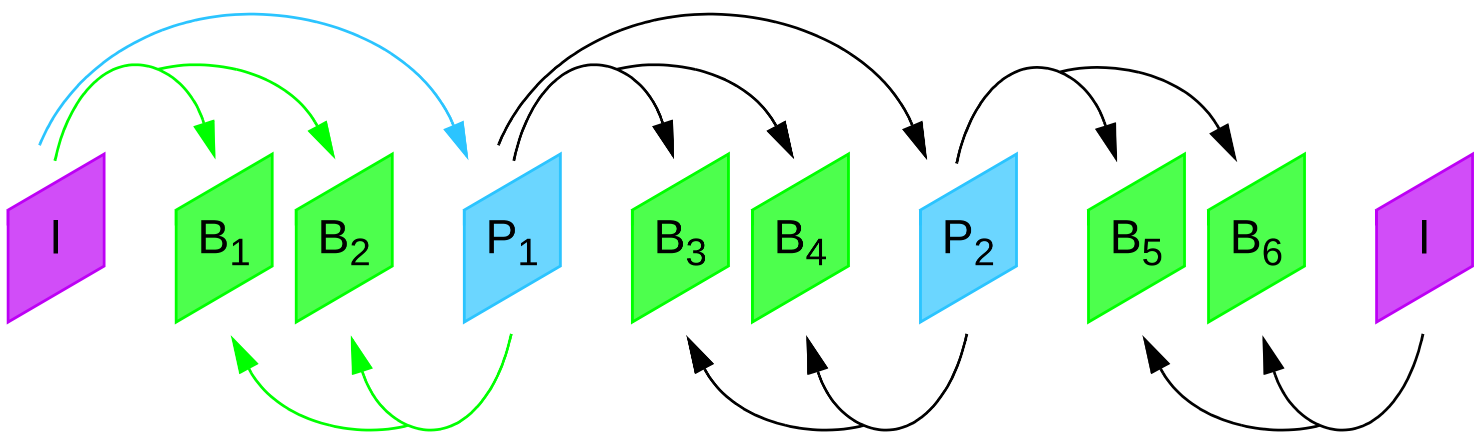
MPEG

Versions

- MPEG-I: bit rate 1 Mbit/sec
- MPEG-II: bit rate 5-10 Mbit/sec
- MPEG-IV: currently being developed

Uses hybrid motion-compensated predictive DCT

- segments the sequence into group of blocks (GOP)



Conclusion on compression

Audio coding

- standards include wavelets/filter banks
- more sophisticated coders use state-of-the-art time-frequency tiling

Image compression

- JPEG 2000 includes wavelets
- interesting new nonlinear approximation schemes

Video compression

- motion residual coding does include transforms as well as more sophisticated, wavelet like ideas

Thus, compression is probably the area where filter banks and wavelets have had the most visible impact

The ideas developed in compression (e.g. EZW) have lead to results in approximation theory