Brice Lecture, Rice Sept. 19 2002,

Signal Representations: from Fourier to Wavelets and Beyond

Martin Vetterli
EPFL & UC Berkeley

1. The Problem and its History

2. Mathematical Representation of Signals

3. Information Theory, Signal Processing and Wavelets

4. Wavelets and Approximation Theory

5. Approximation and Applications in Denoising and Compression

6. Going to Two Dimensions: Nonseparable Bases

7. Conclusions

Acknowledgements

Swiss National Science Foundation

Collaborators

- T. Blu (EPFL)
- M. Do (UIUC)
- P.L. Dragotti (Imperial College)
- P. Marzilliano (Genimedia)
- P. Roud (EPFL)

1. The Problem and its History

Henry the 8th looks for a new spouse

Anne de Clève, Holbein, 1539

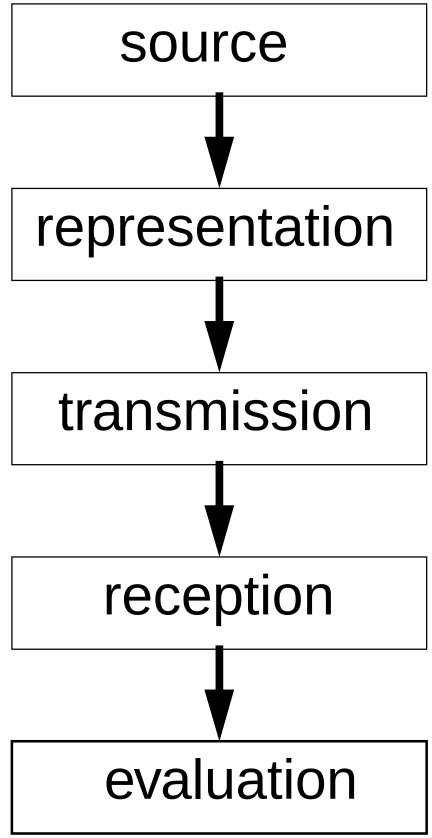
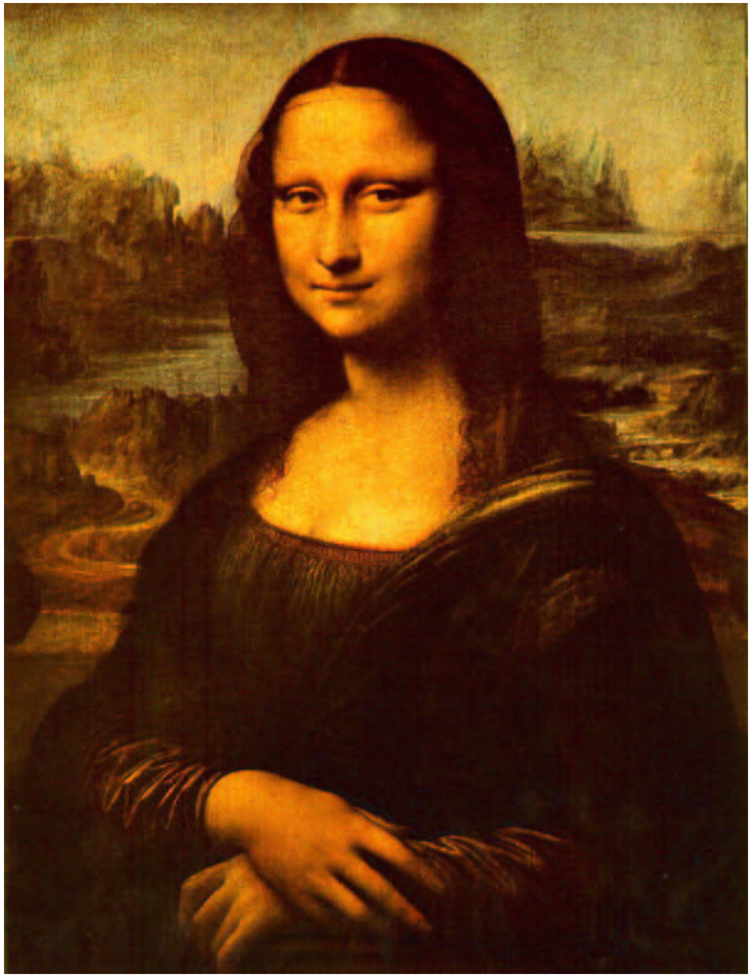


Image communication is an old problem...

How many bits for Mona Lisa?



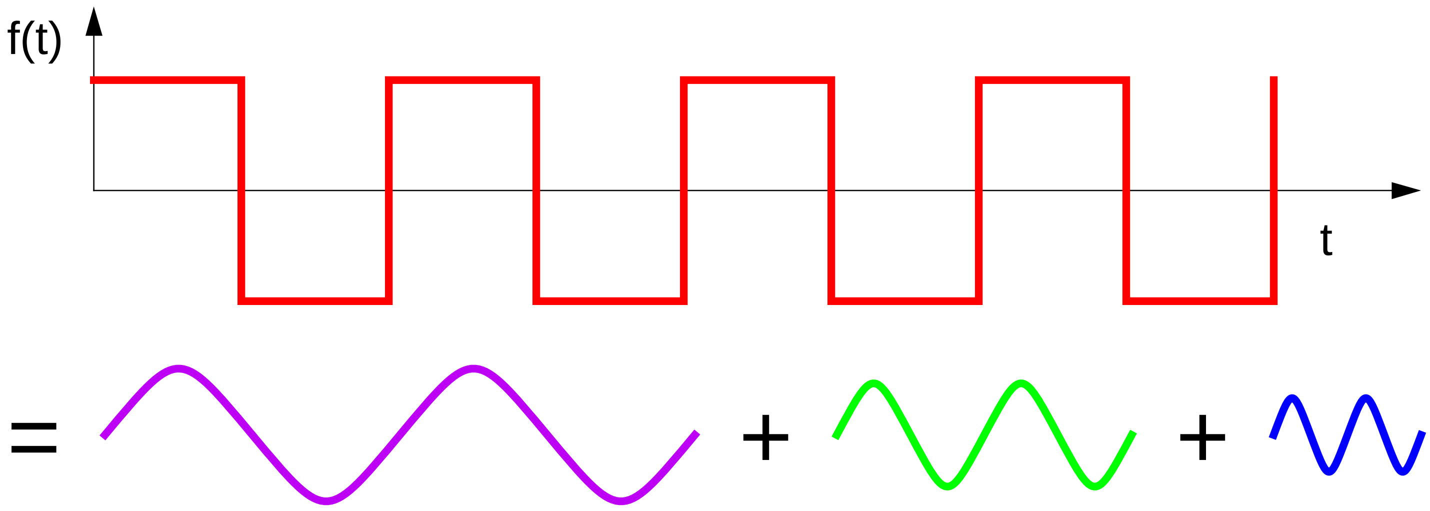
<=> {0,1}

2. Mathematical Representation of Signals

Joseph Fourier (1768-1830)

Studies the heat equation (in Egypt...)

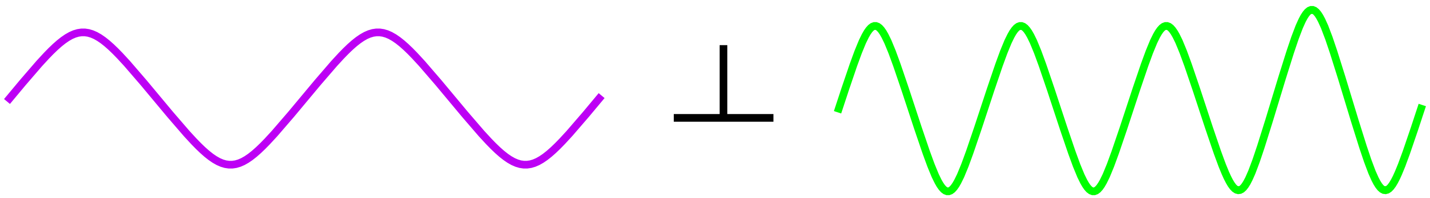
1807: Fourier upsets the French Academy....



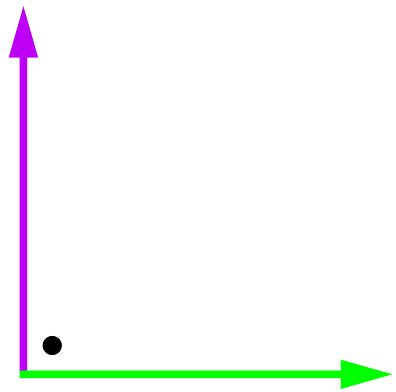
Fourier Series:

- Harmonic series
- Frequency changes, f_0, $2f_0$, $3f_0$, ...

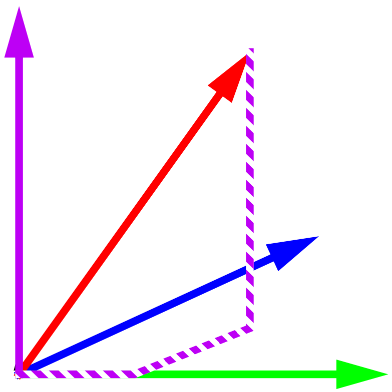
“What is the magic trick?”



or, with Euclid:



orthogonality

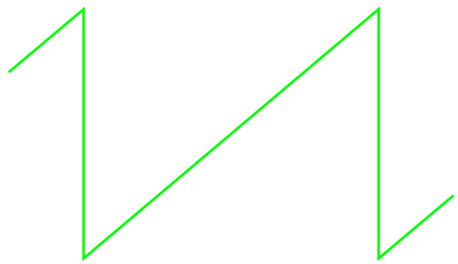


coordinate system

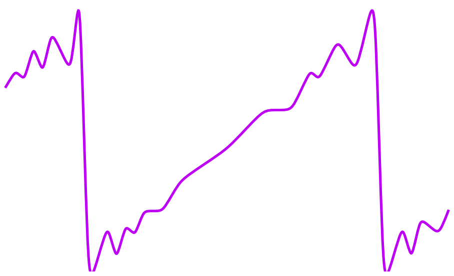
and: successive approximation

But

1898: Gibbs' paper

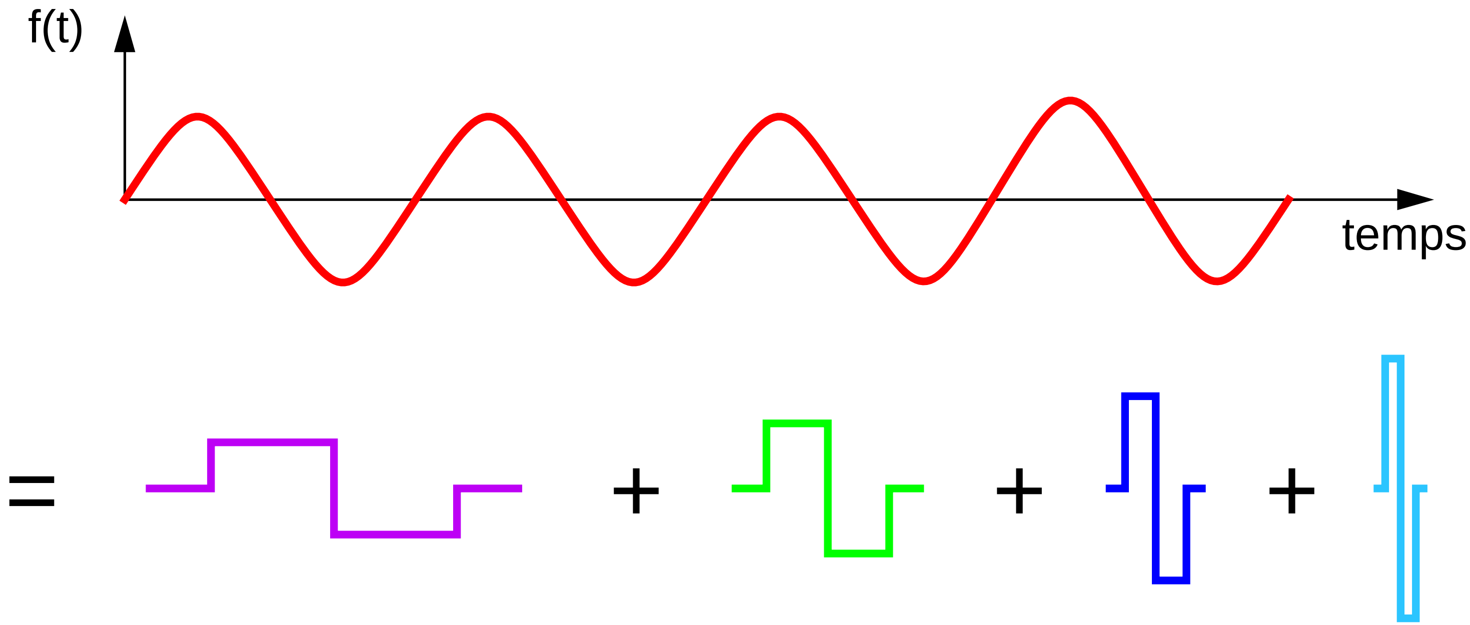


1899: Gibbs' correction



and it will take almost another 60 years to settle the convergence question (Carleson 1964).

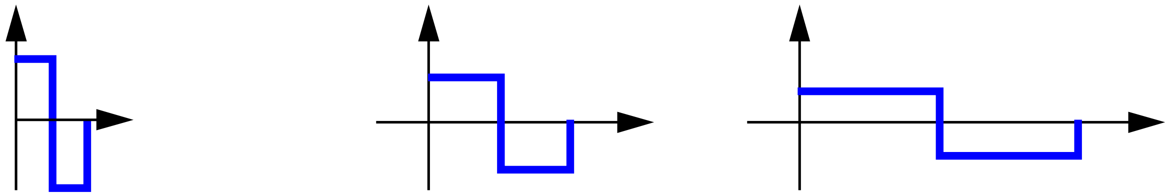
1910: Alfred Haar discovers the Haar wavelet dual to the Fourier construction



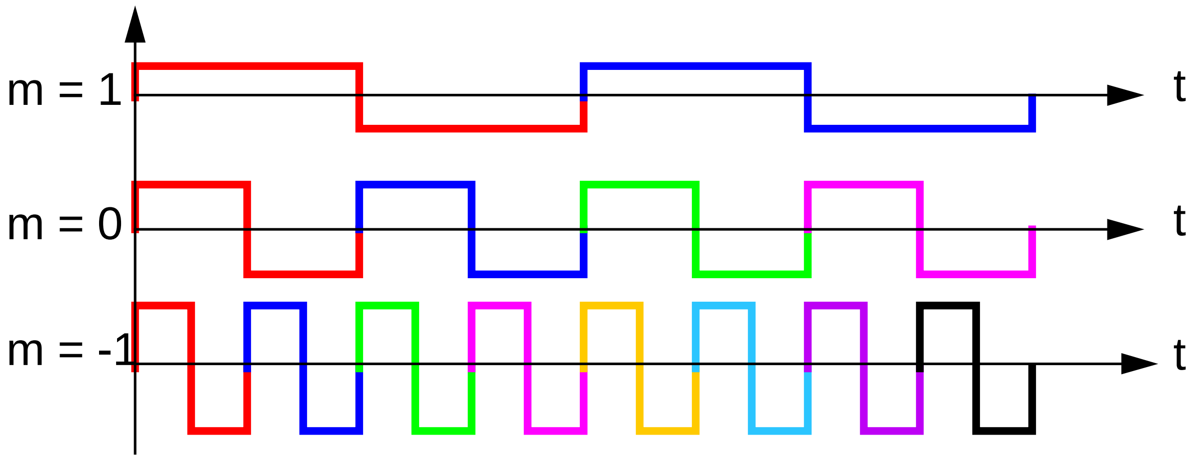
Haar series:

- Scale change, scales S_0, $2S_0$, $4S_0$, $8S_0$
- Time shift

The Haar system



Again a set of orthonormal vectors!

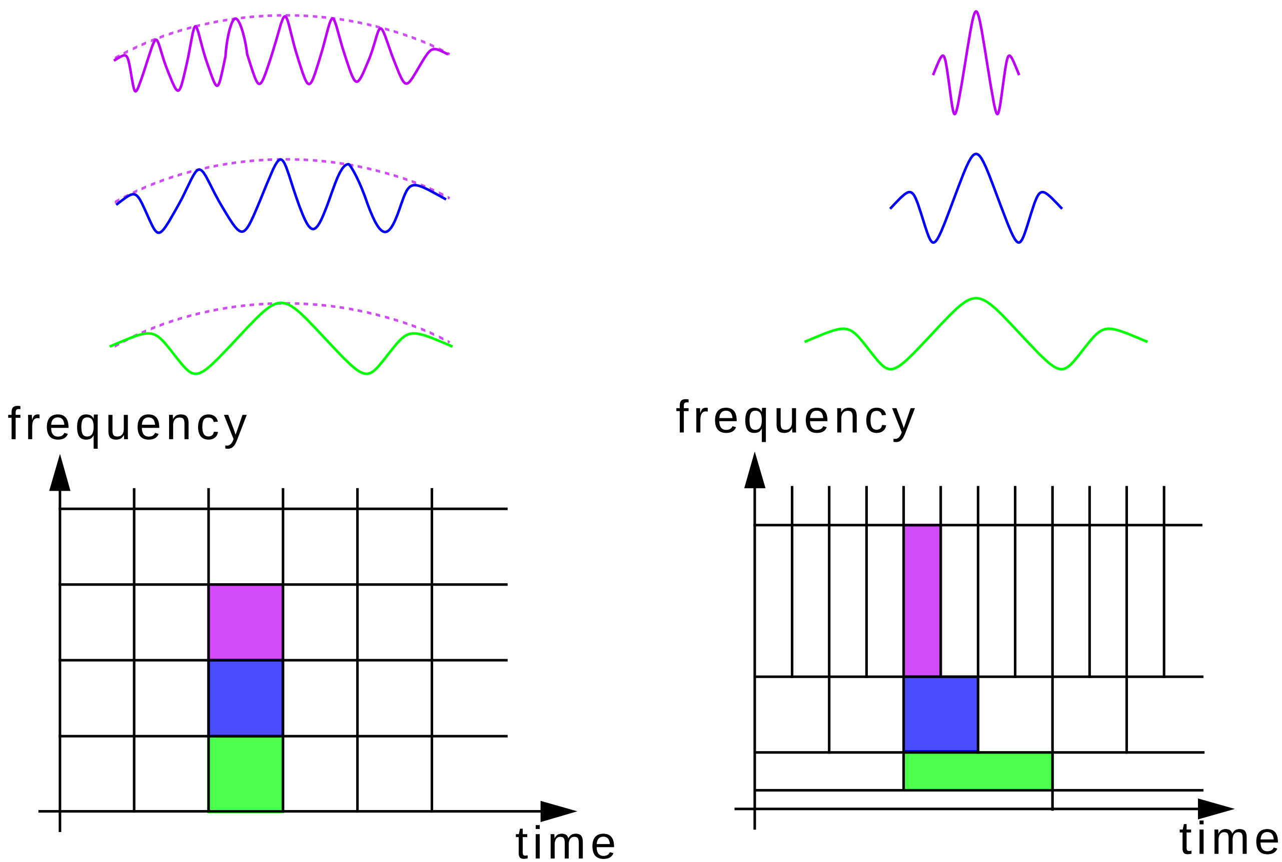


Size: length proportional to 2^m

''frequency'' : f_0, $2f_0$, $4f_0$, $8f_0$, ... octaves!

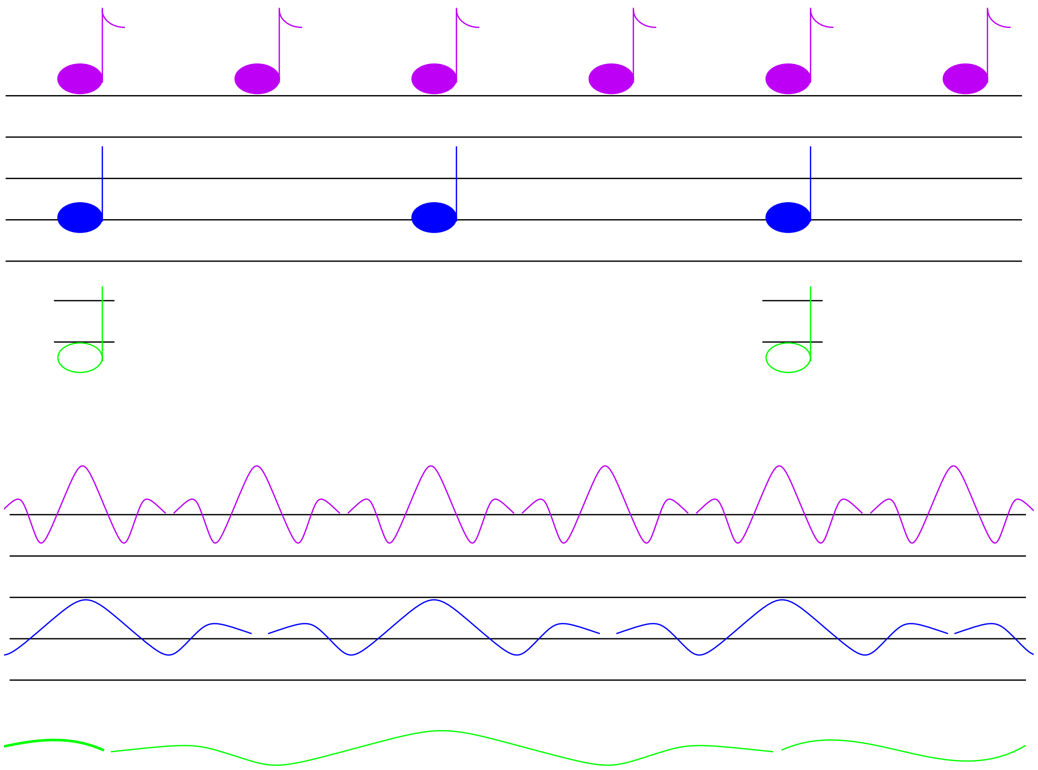
1945: Gabor localizes the Fourier transform $\Rightarrow$ STFT

1980: Morlet proposes the continuous wavelet transform

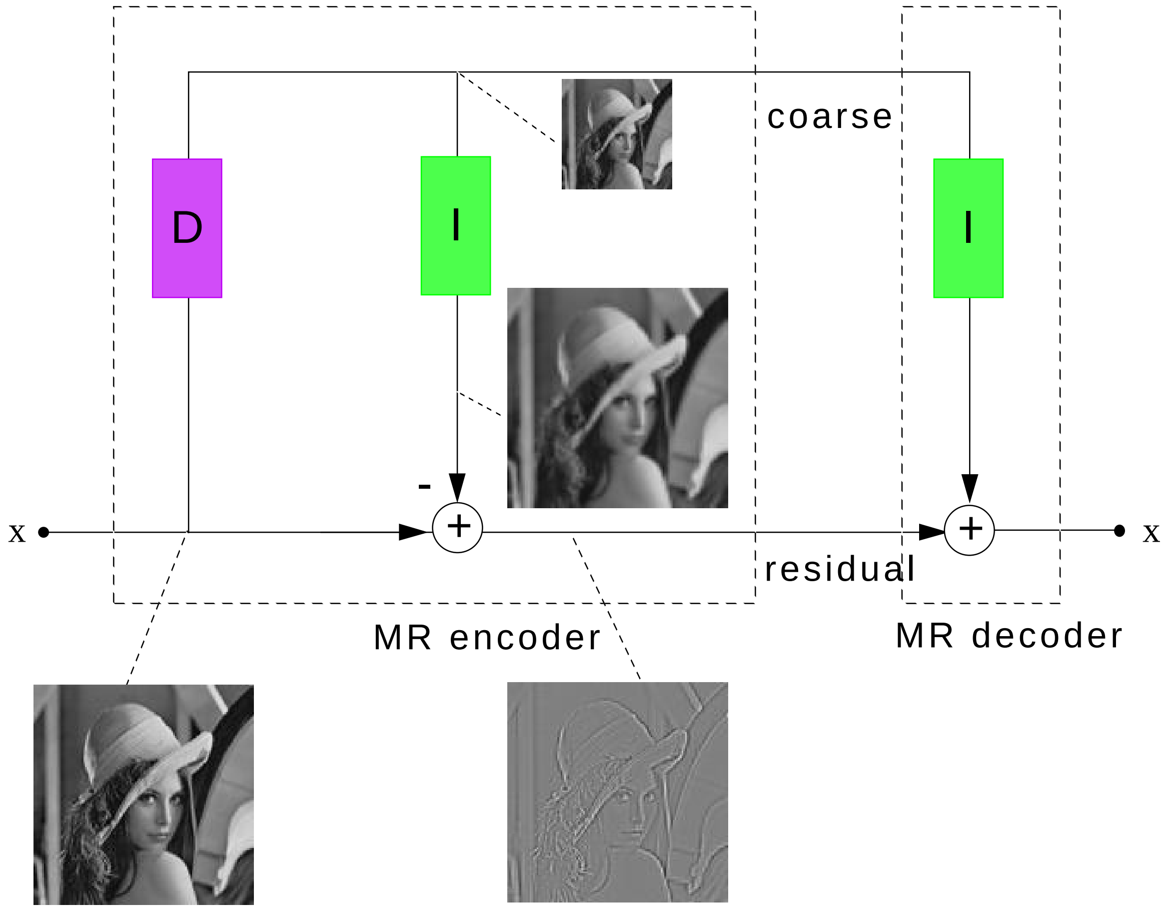


short-time Fourier transform wavelet transform

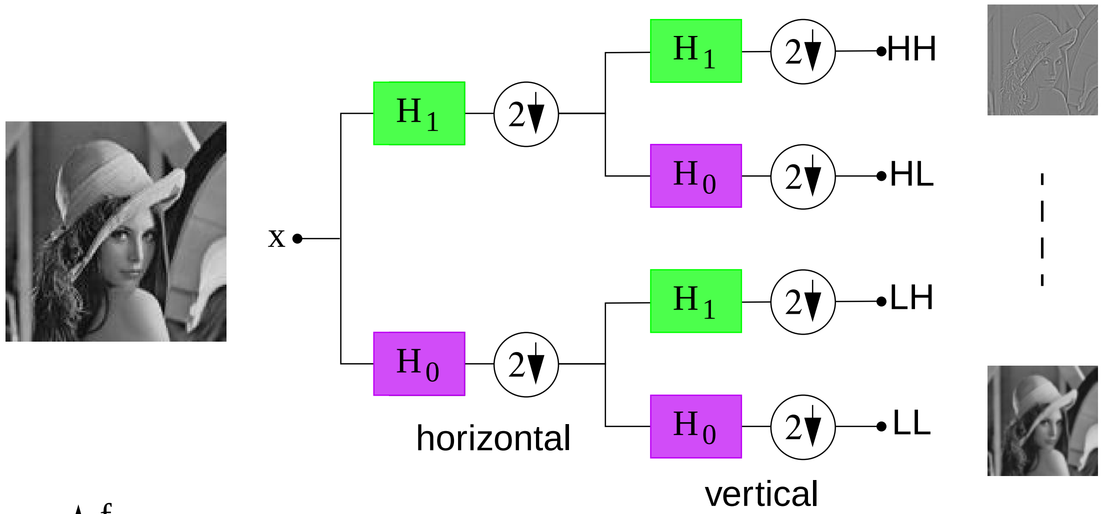
Analogy with the musical score
Bach knew about wavelets!

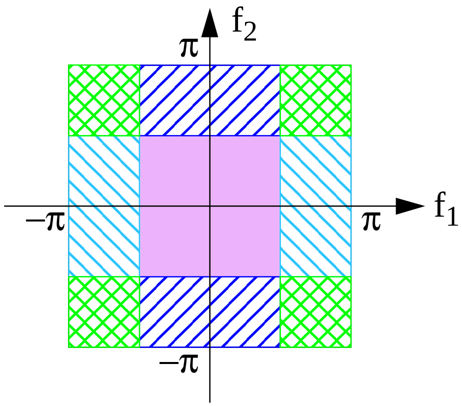


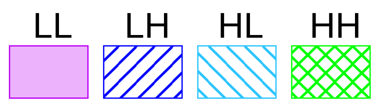
1983: Lena discovers pyramids (actually, Burt and Adelson)



1984: Lena gets critical (subband coding)





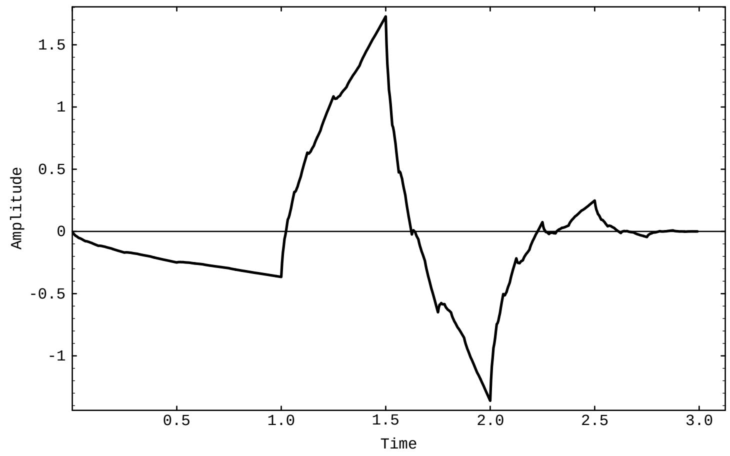


1986: Lena gets formal...
(multiresolution theory by Mallat, Meyer...)

=

+

1988: Ingrid discovers Daubechies' wavelets!



- New families of orthonormal bases, (generalizing Haar)
- Biorthogonal families, frames
- many new applications

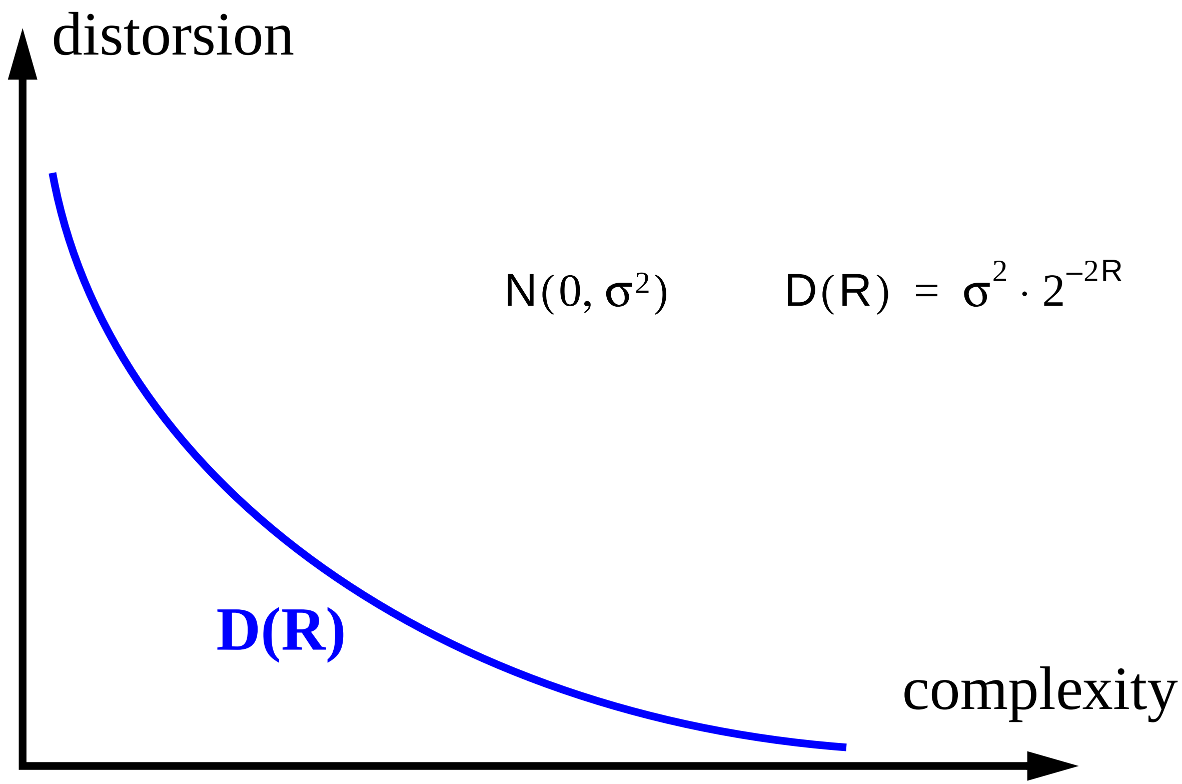
3. Information Theory, Signal Processing and Wavelets

Claude Shannon: The founding genius

1. Source coding
2. Channel coding
3. Separation of source and channel coding

Source Coding

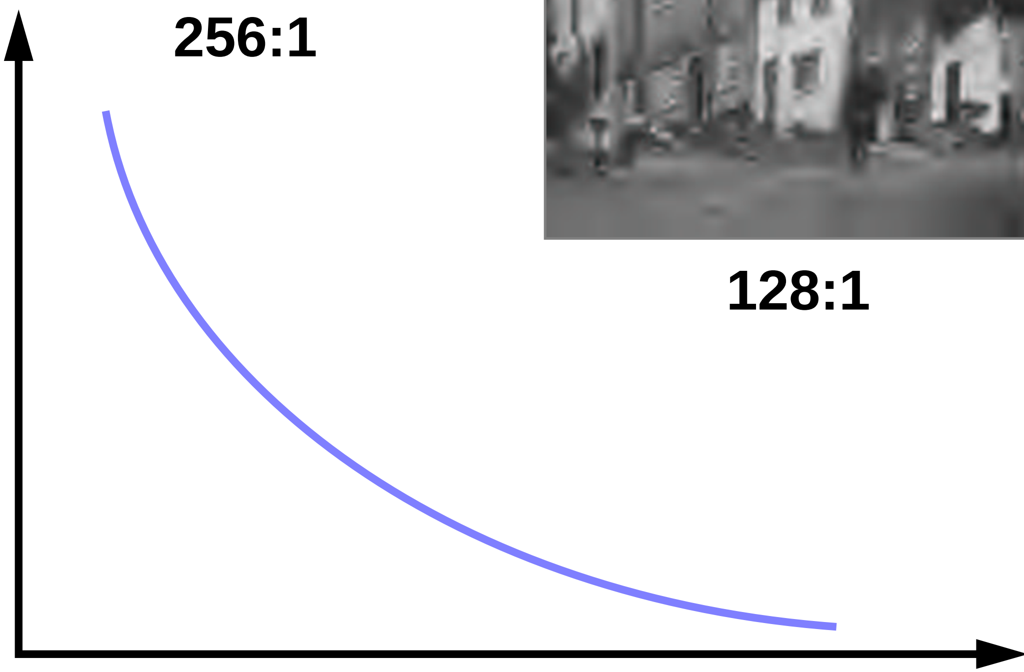
exchanging description complexity for quality



Again, successive approximation is key

256:1

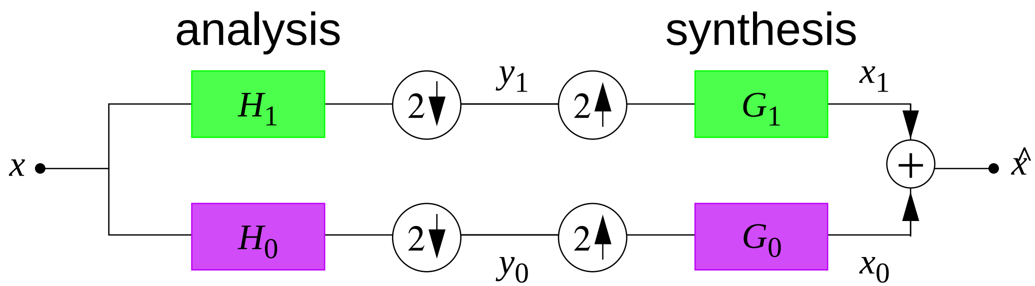
128:1

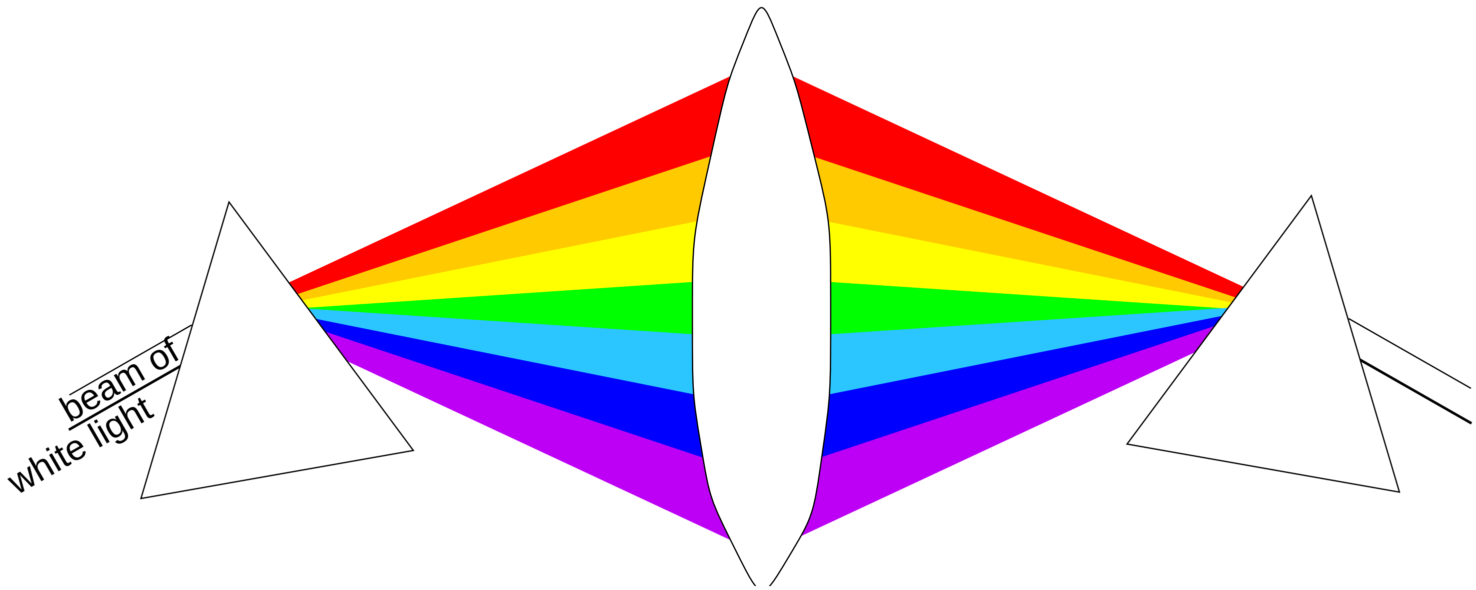


32:1

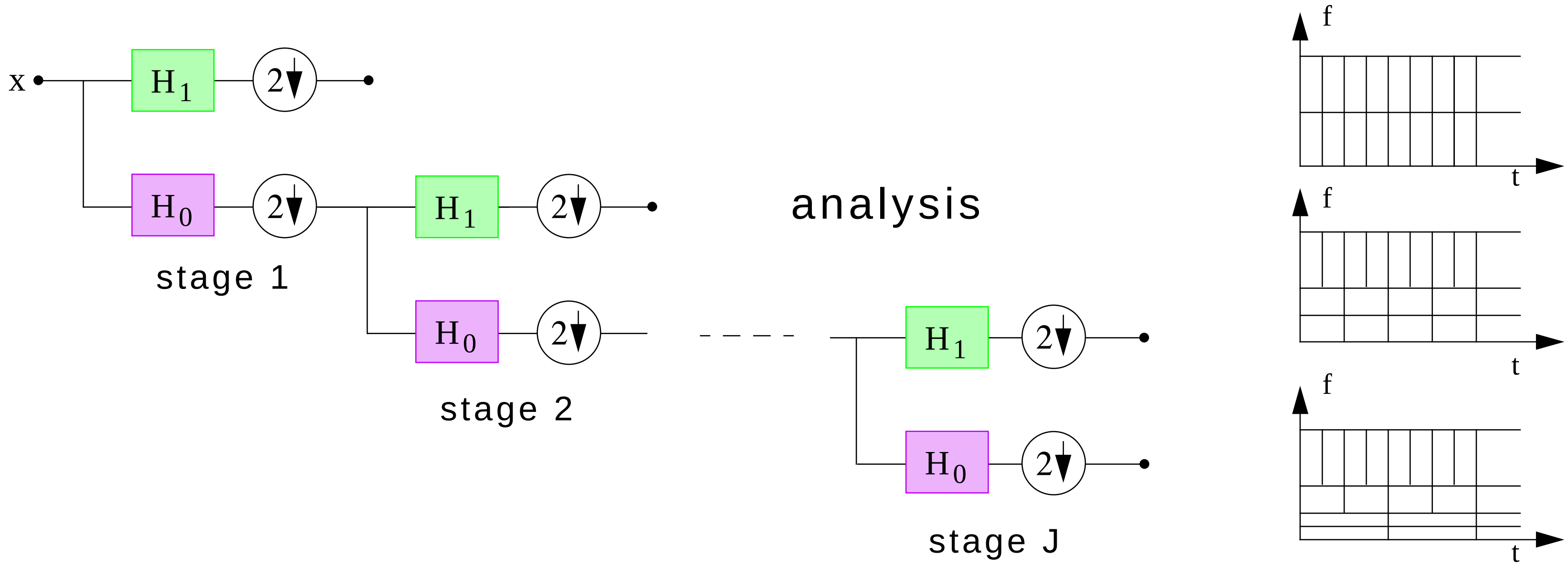
Signal Processing

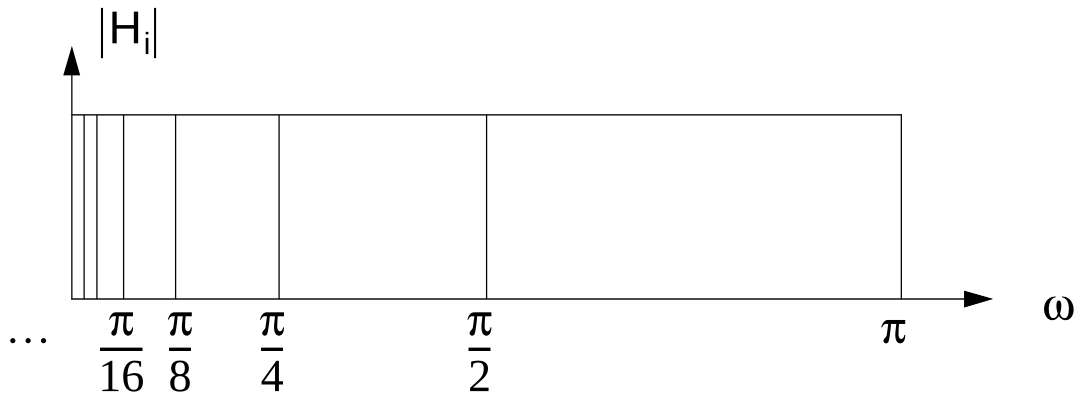
Subband coding





Iterated filter banks





Frequency division

Separable application in 2D

An image and its wavelet decomposition

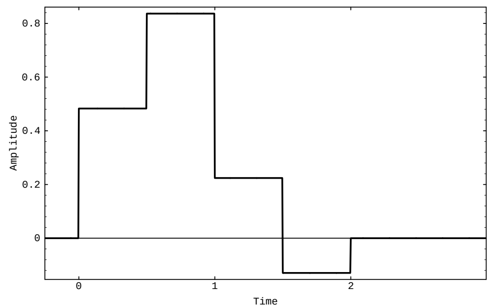
Important:

- auditory system works in octaves
- visual system works in frequency bands

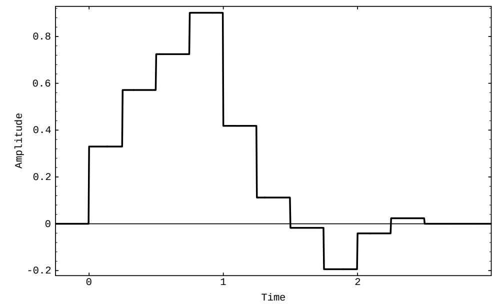
The iterated filter bank leads to wavelets

The Daubechies iterative wavelet construction

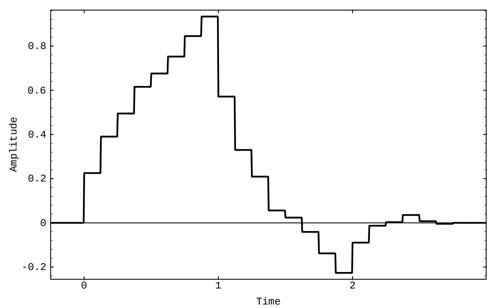
i = 1



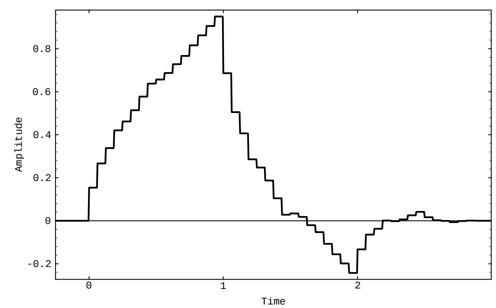
i = 2



i = 3



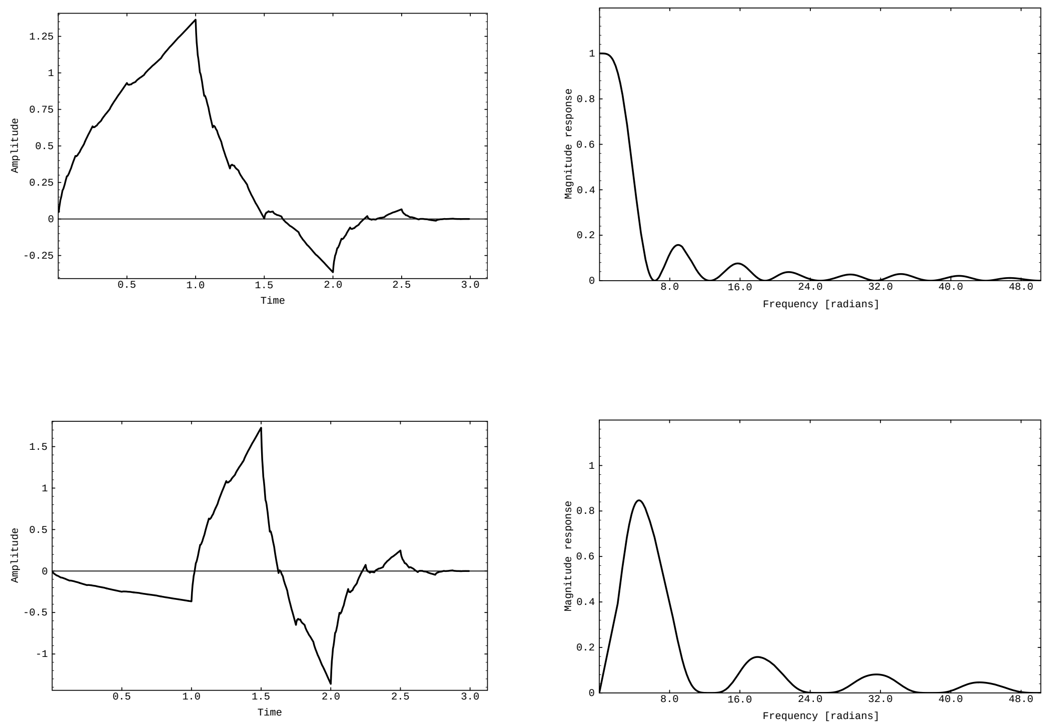
i = 4



Scaling function and Wavelet

scaling function

wavelet



lowpass

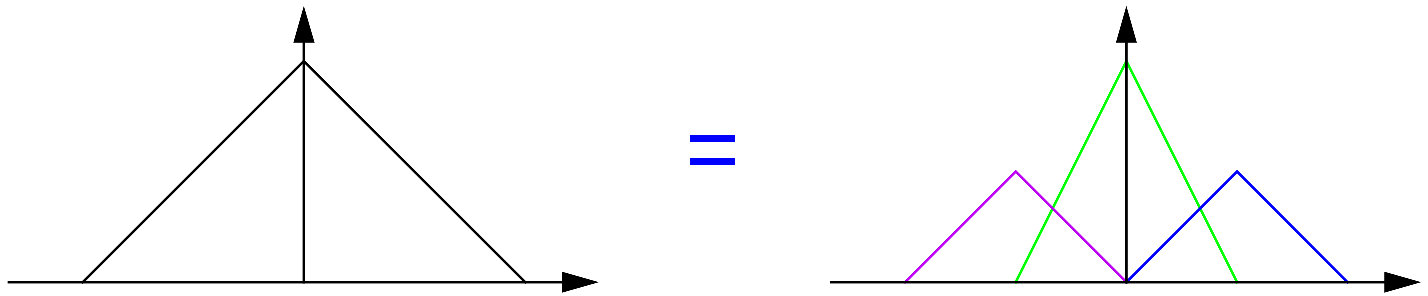
bandpass

Finite length, continuous $\varphi(t)$ and $\psi(t)$, based on L=4 iterated filter

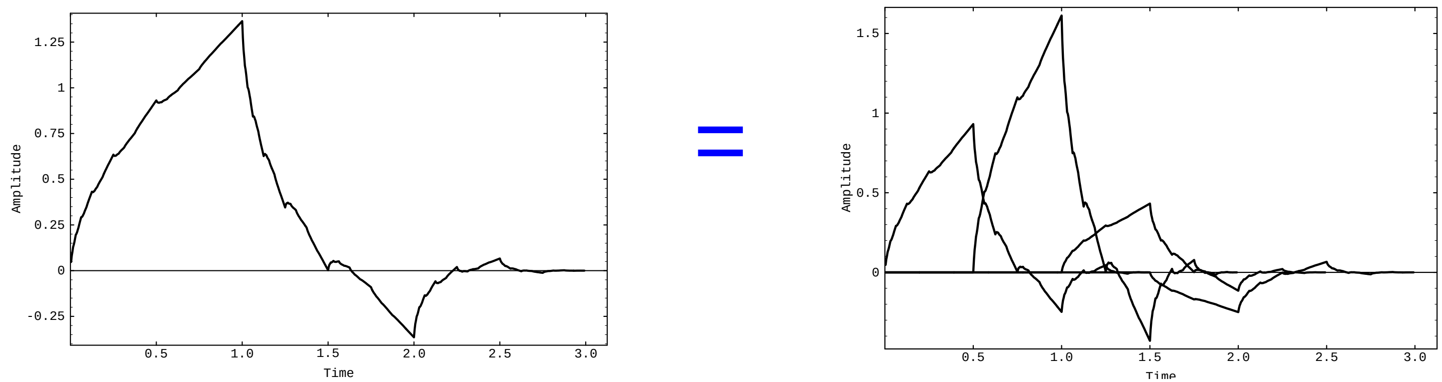
Iterated filter banks lead to two-scale equations

$$\varphi(t) = \sum_n c_n \varphi(2t - n)$$

Hat function



Daubechies' scaling function

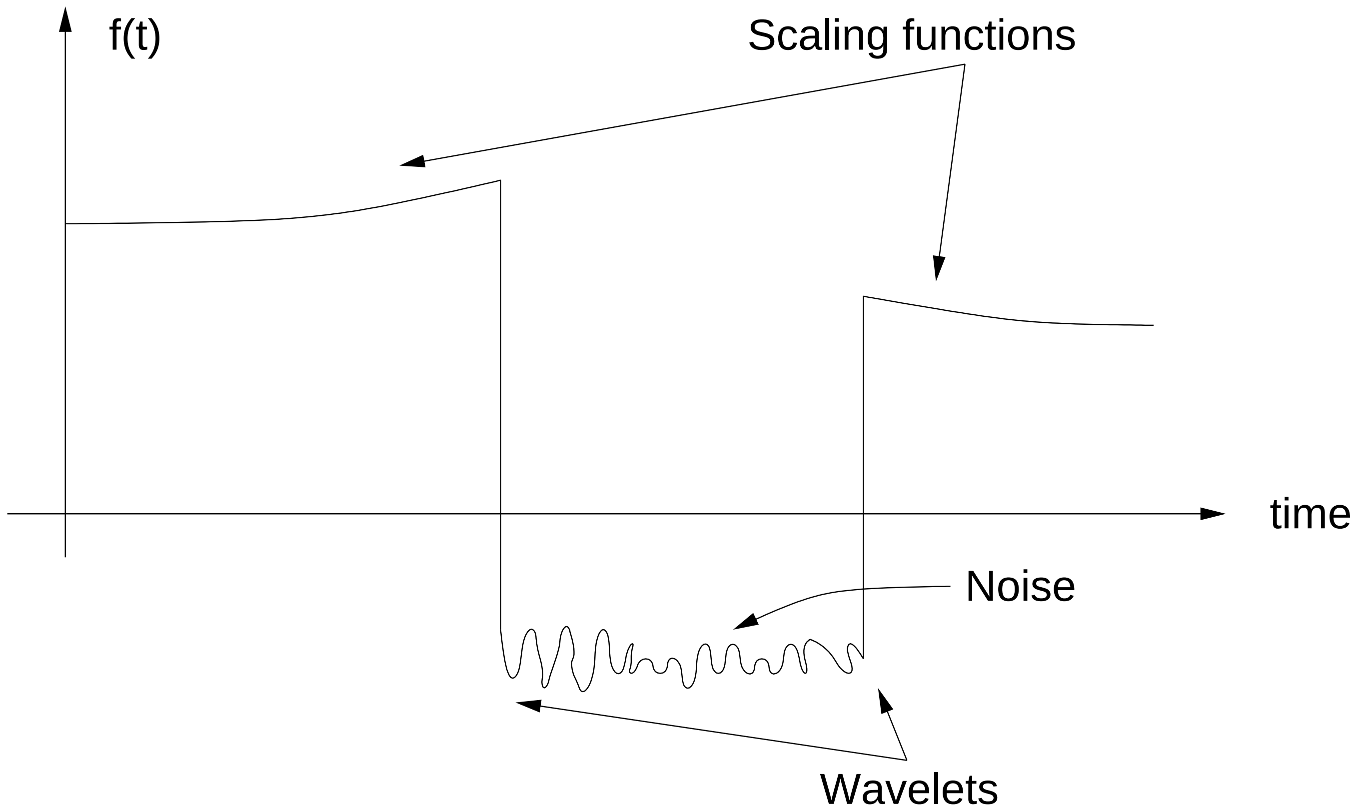


Relation to self-similarity useful for analysis and charaterization of fractal processes

4. Wavelets and Approximation Theory

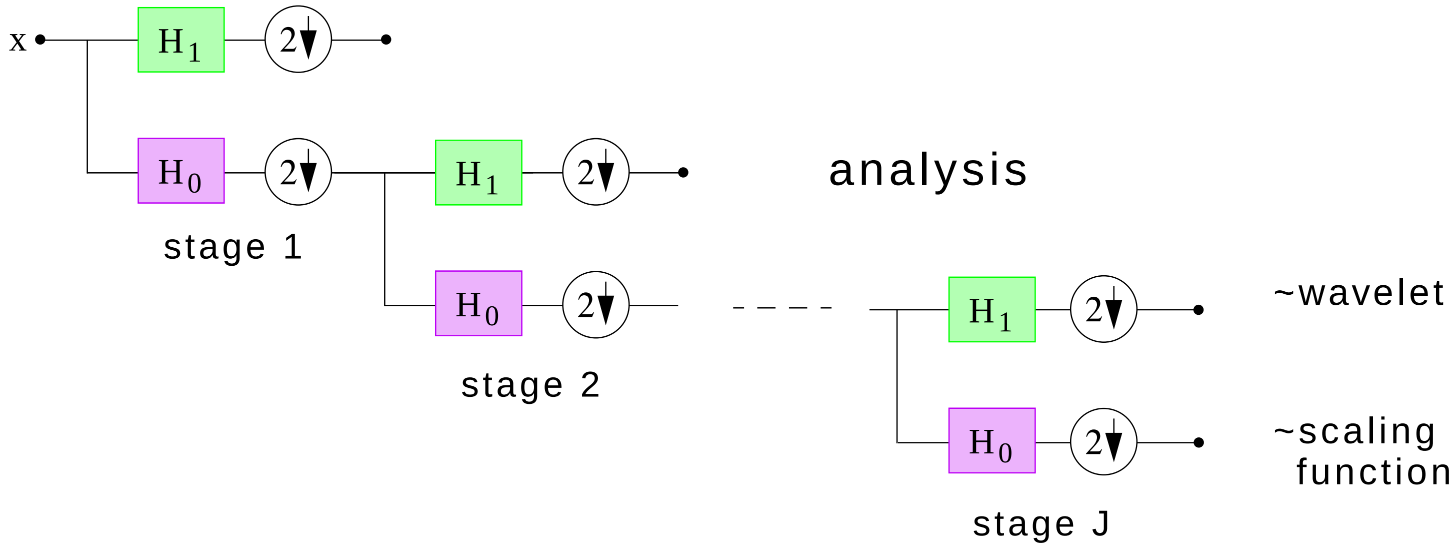
Consider piecewise smooth signals

- Wavelet act as singularity detectors
- Scaling functions catch smooth parts
- “Noise” is circularly symetric



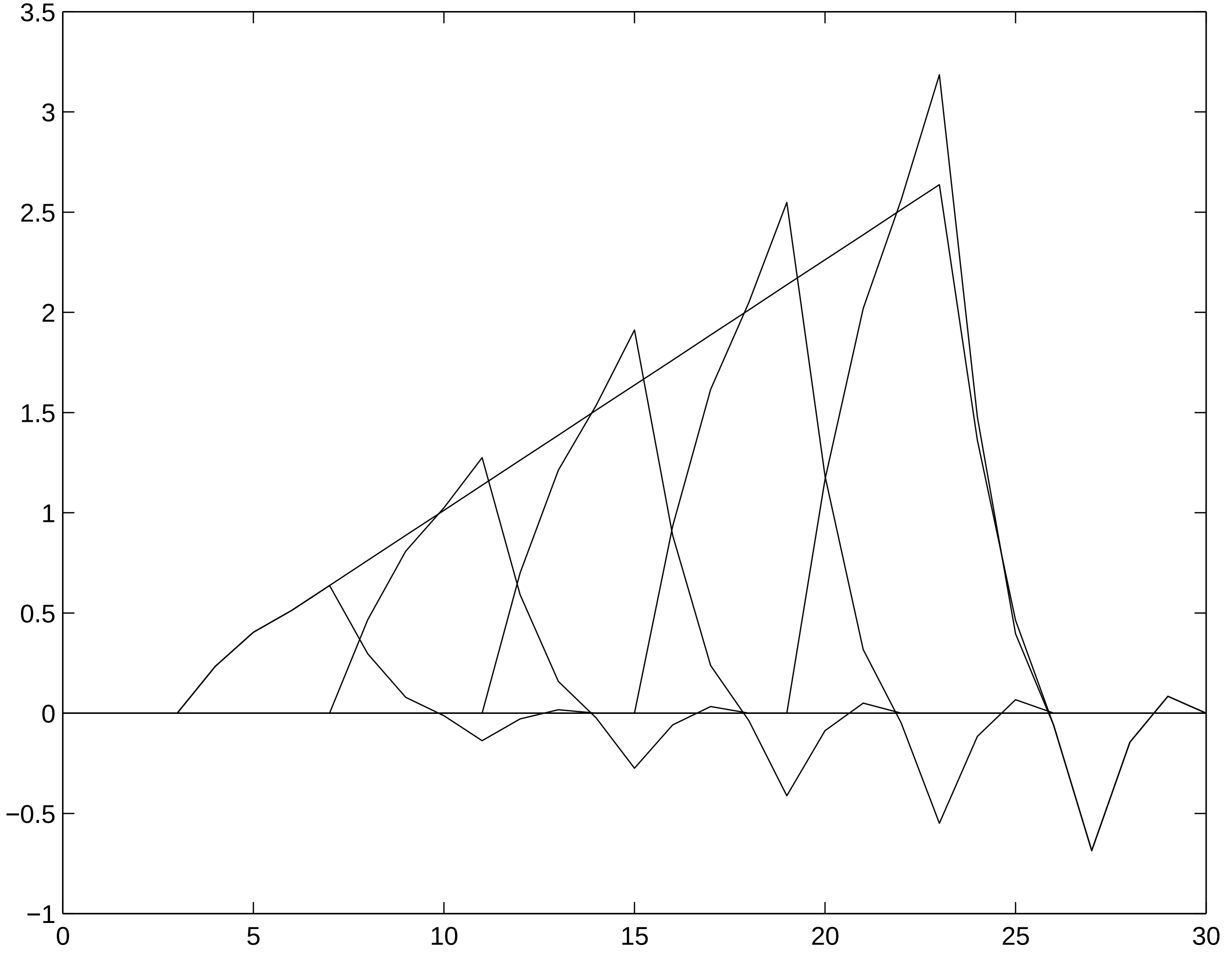
How does this work? Proper choice of filters!

Iterated filter bank ($H_j(z) = G_j(z^{-1})$)



- polynomials are ‘‘eaten’’ in the highpass
- polynomials are reproduced by the lowpass channel
- discontinuities are detected by the wavelets

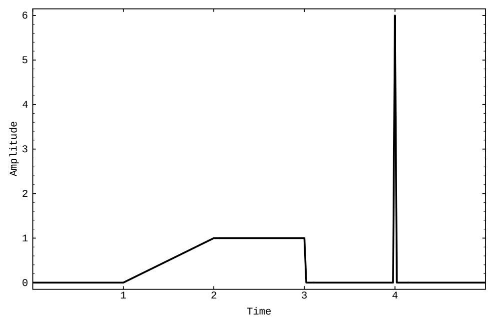
Example: S_4 reproduces linear fcts

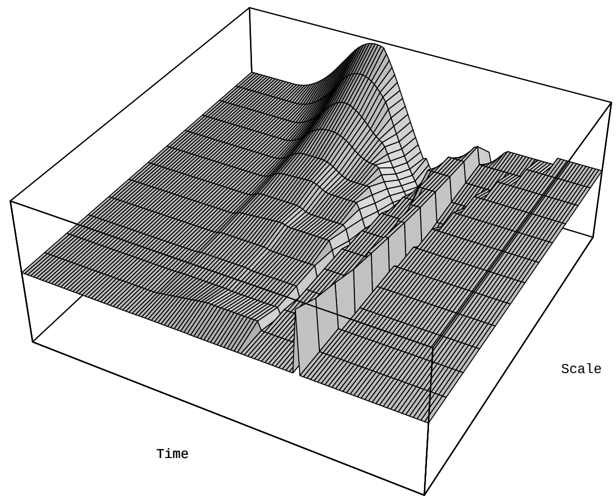


How about singularities?

If we have a singularity of order n at the origin (-1 Dirac, 0: Heaviside,...), the CWT transform behaves as

$$X(a, 0) = c_n \cdot a^{n/2}$$

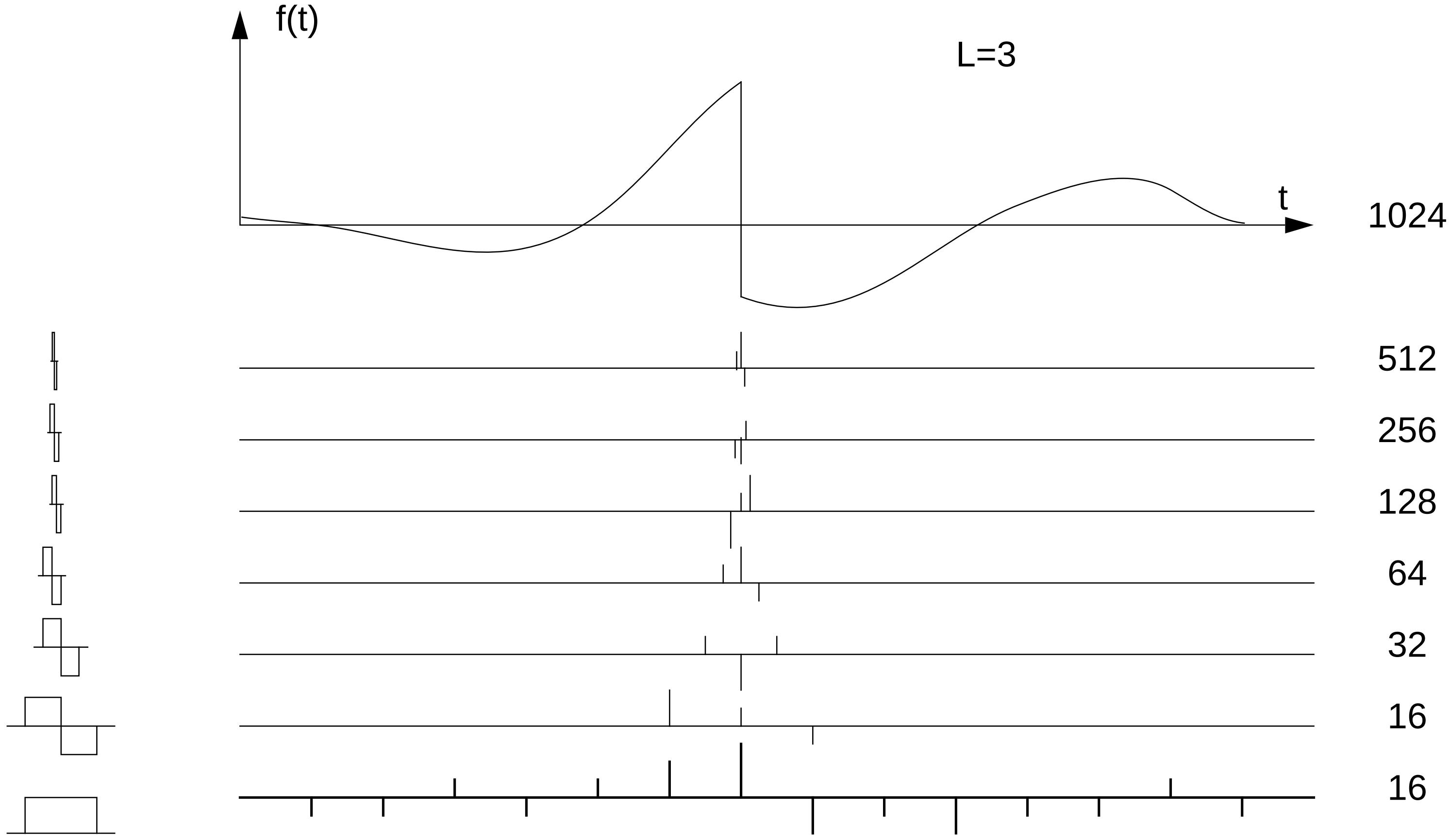




In the orthogonal wavelet series: same behavior, but only L=2N-1 coefficients influenced at each scale!

- e.g. Dirac/Heaviside: behavior as $2^{-m/2}$ and $2^{m/2}$

Example:



- phase changes randomize signs, but not decay
- a singularity influences only L wavelets at each scale ($L=2N-1=3$)

Approximation: linear versus non-linear

Given an orthonormal basis $\{g_n\}$ for a space S and a signal

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

- the best linear approximation is given by the projection onto a fixed subspace of size M (independent of f!)

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

- the best nonlinear approximation is given by the projection onto an adapted subspace of size M (dependent on f!)

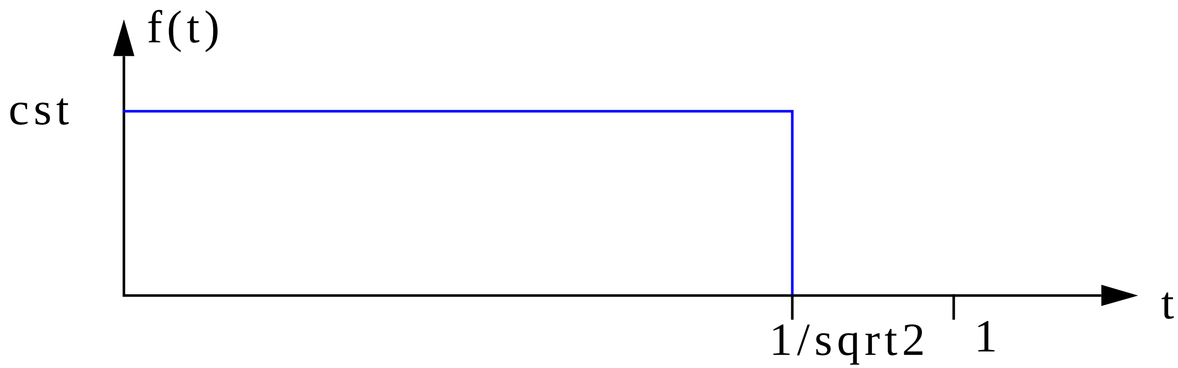
$$\tilde{f}_M = \sum_{n \in I_M} \langle f, g_n \rangle \cdot g_n \longrightarrow I_M\text{: set of largest } M \text{ coeffs}$$

or: take the first M coeffs (linear) or take the largest M coeffs (non-linear)

Nonlinear approximation

Nonlinear approximation power depends on basis

Example:



Two different bases for $[0,1]$:

- Fourier series $\{e^{j2\pi kt}\}_{k \in \mathfrak{I}}$
- Wavelet series: Haar wavelets

Linear approximation in Fourier or wavelet bases

$$\hat{\varepsilon}_M \sim 1/M$$

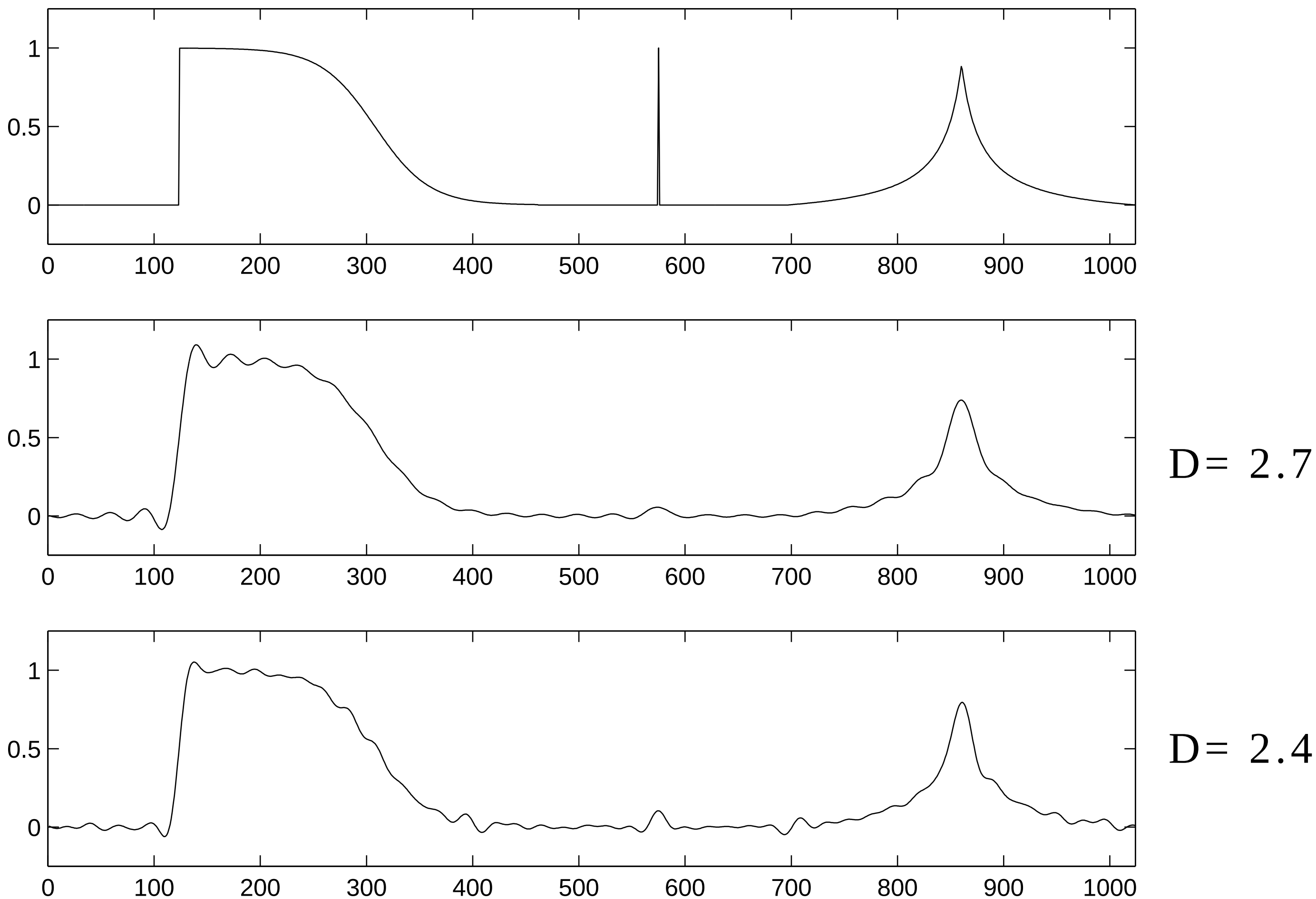
Nonlinear approximation in a Fourier basis

$$\tilde{\varepsilon}_M \sim 1/M$$

Nonlinear approximation in a wavelet basis

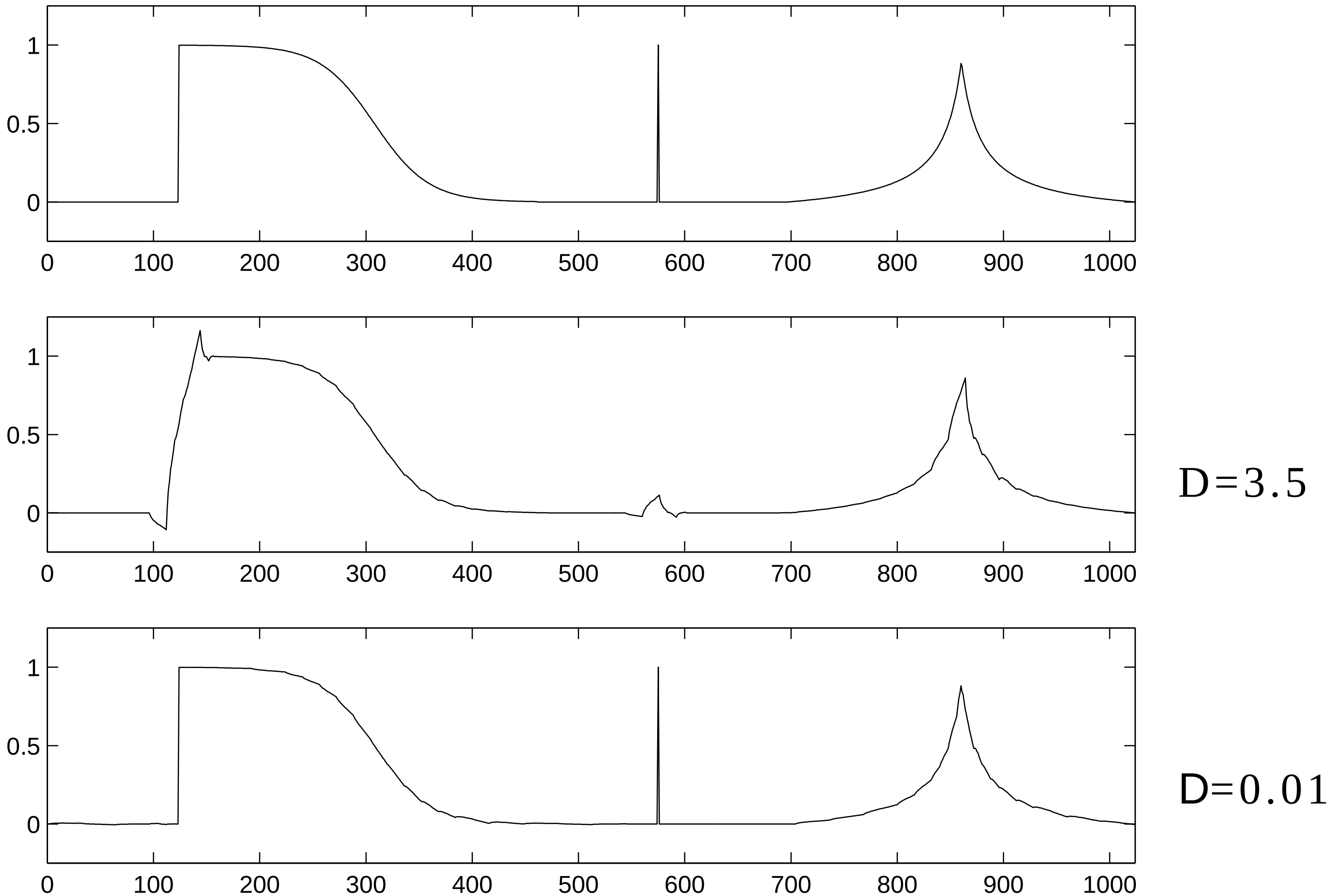
$$\tilde{\varepsilon}_M \sim 1/2^M$$

Fourier Basis: N=1024, M= 64, linear versus nonlinear



- nonlinear approximation is not necessarily much better!

Wavelet basis: N=1024, M= 64, J=6, linear versus non-linear



- nonlinear approximation is vastly superior!

5. Approximation and Applications in Denoising and Compression

Wavelets approximate piecewise smooth signals with few non-zero coefficients

This is good for

- Compression
- Denoising
- Classification
- Inverse problems

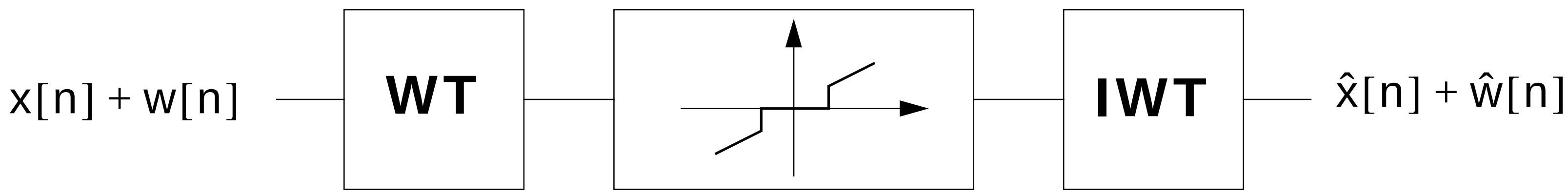
Thus: sparsity is good!

Denoising

Idea:

- Dominant features are caught by large wavelet coefficients
- Noise is spread uniformly over all coefficients
- Thresholding small coefficients to 0 keeps the signal but removes the noise

Schematically:



Now:

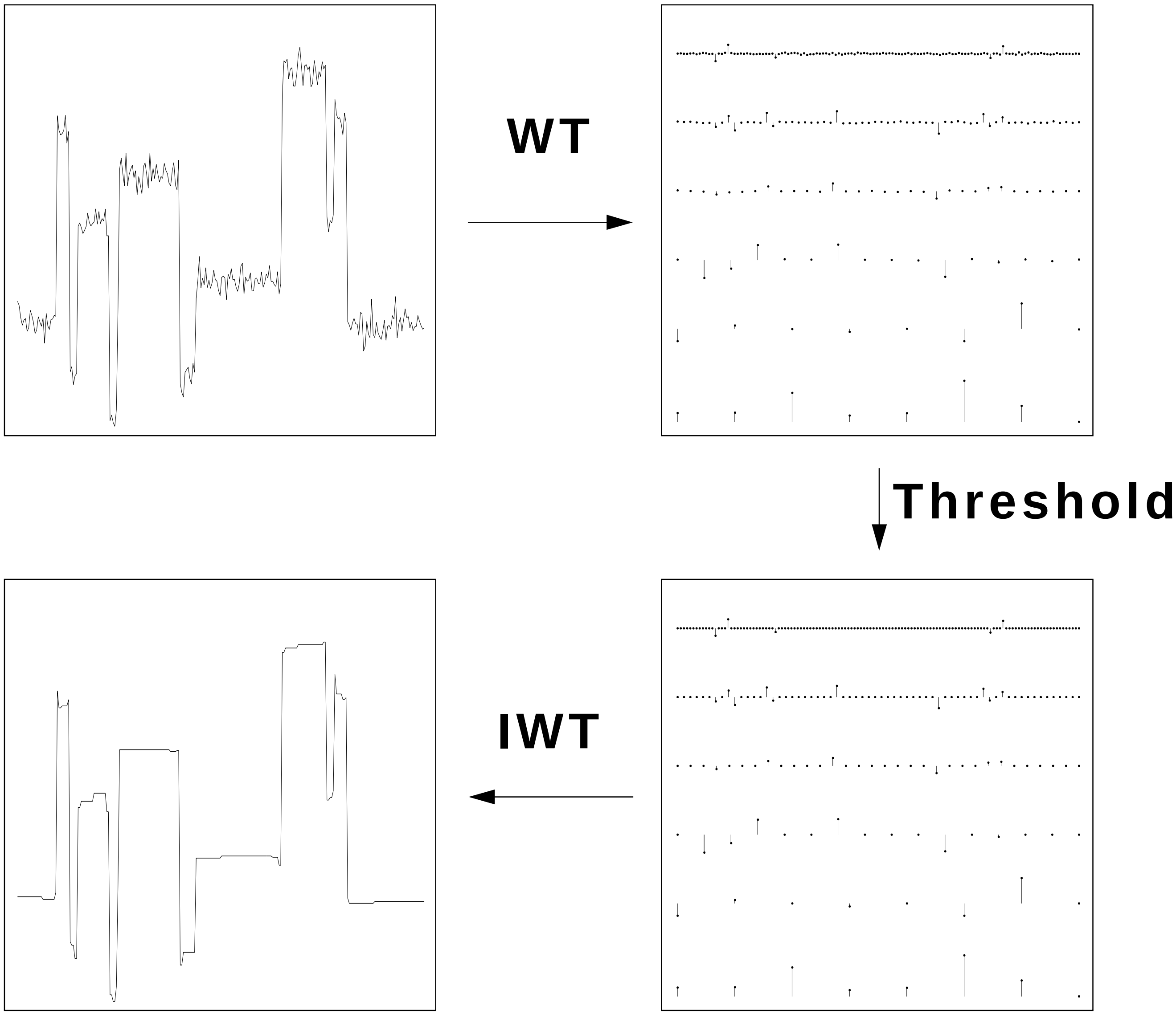
$$\hat{x}[n] \sim x[n]$$

$$|\hat{w}[n]| \ll |w[n]|$$

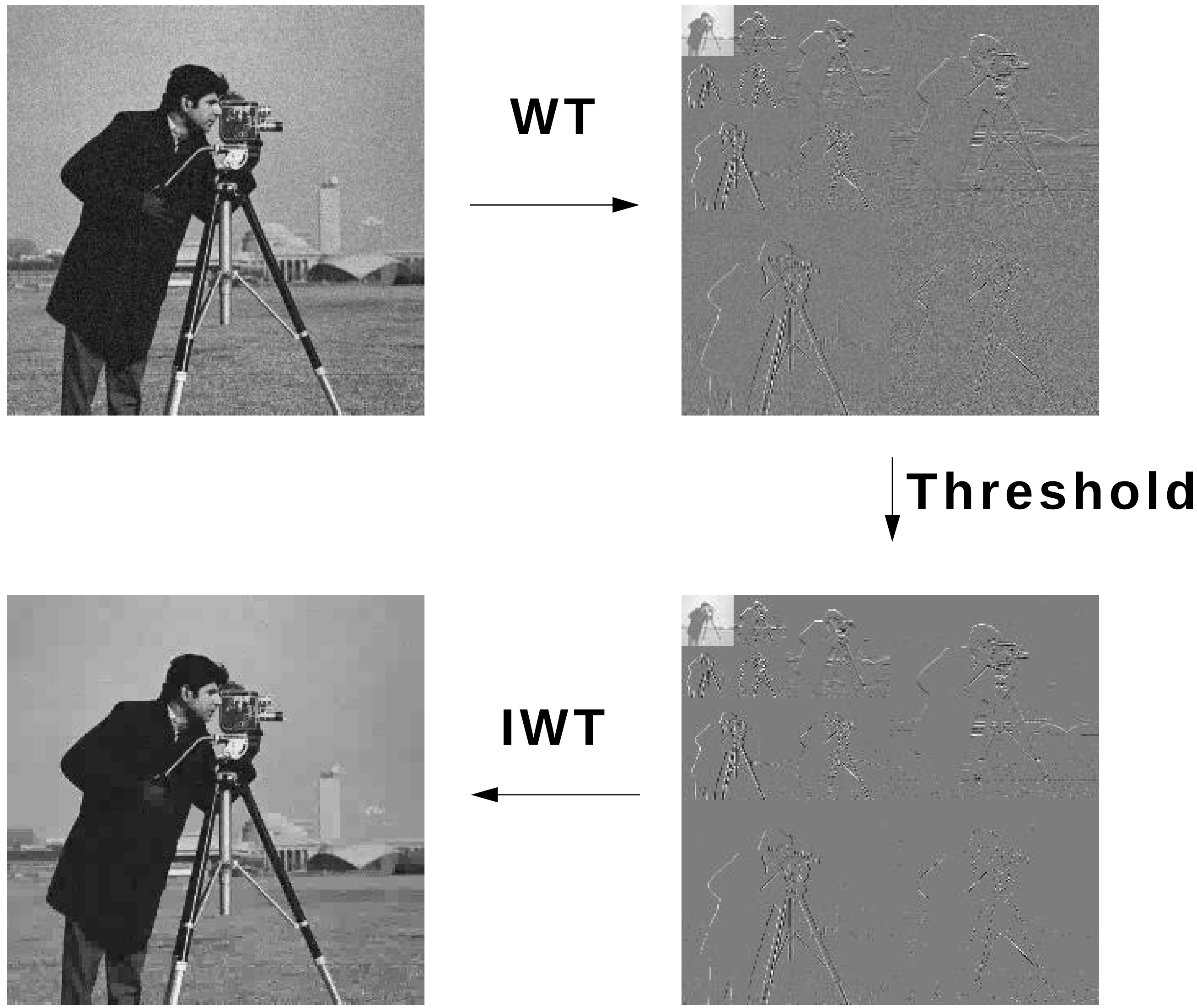
Note:

- very simple
- works well for piecewise smooth signals
- for jointly gaussian, standard linear methods (Wiener filter) are fine

Example: 1D Signal



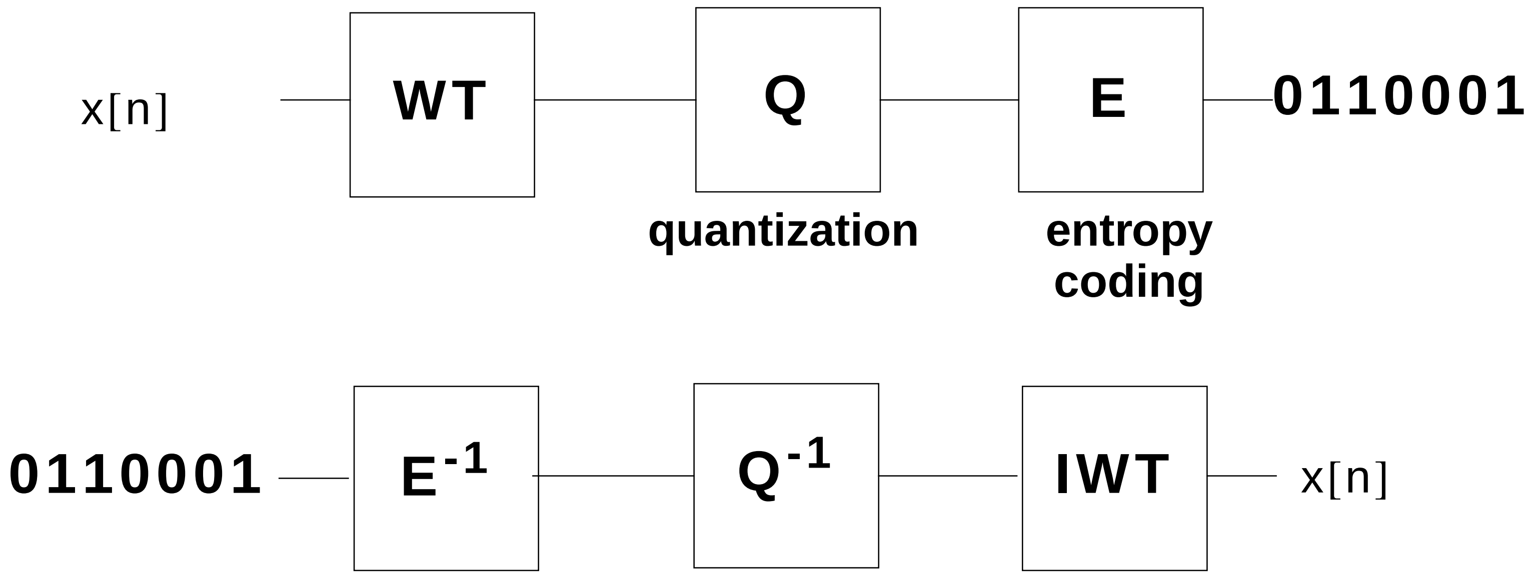
Example: 2D signal



Compression

Idea

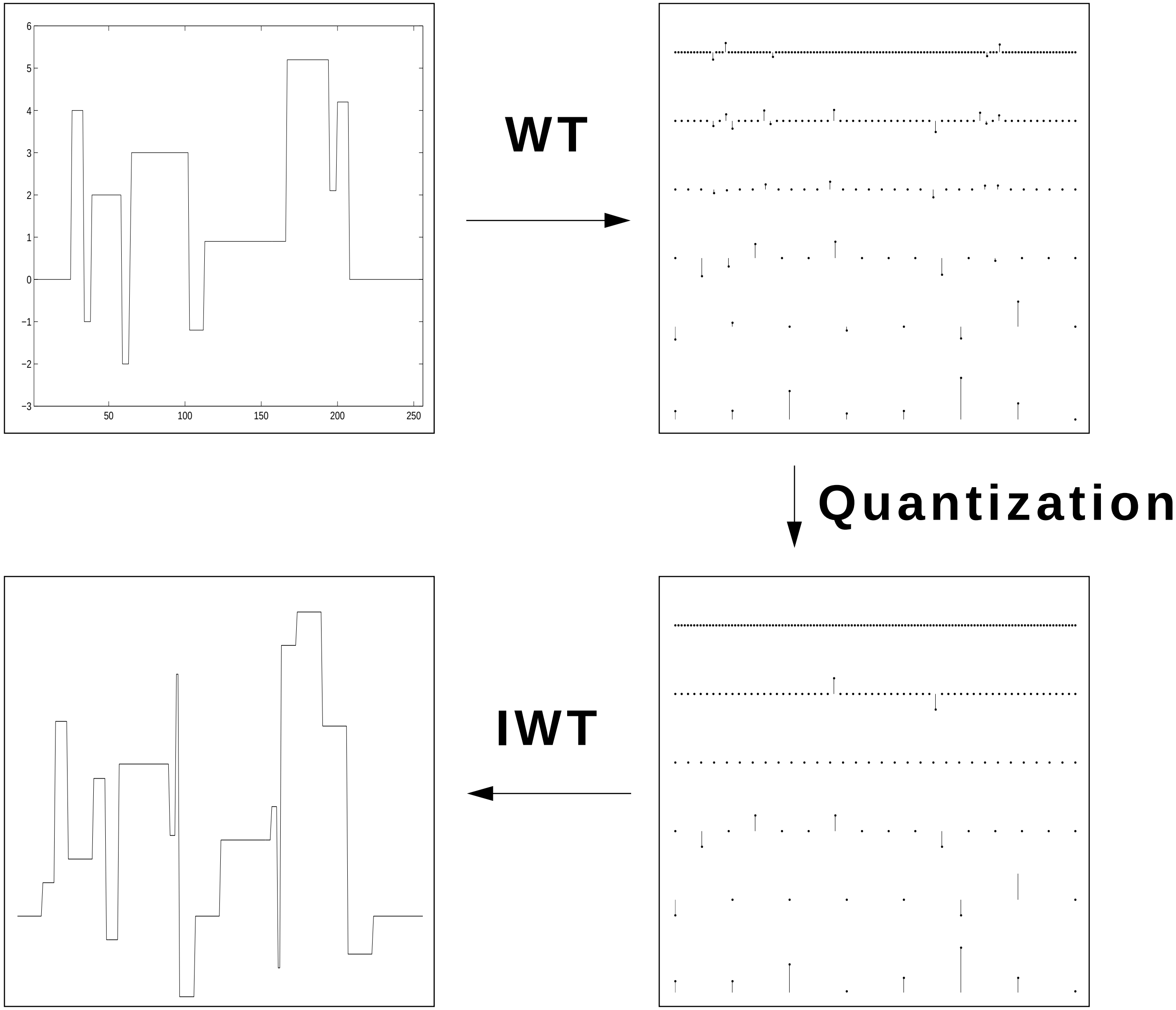
- sparse representation should be good for compression
- transform, keep large coefficients through quantization
- reconstruction gives good quality



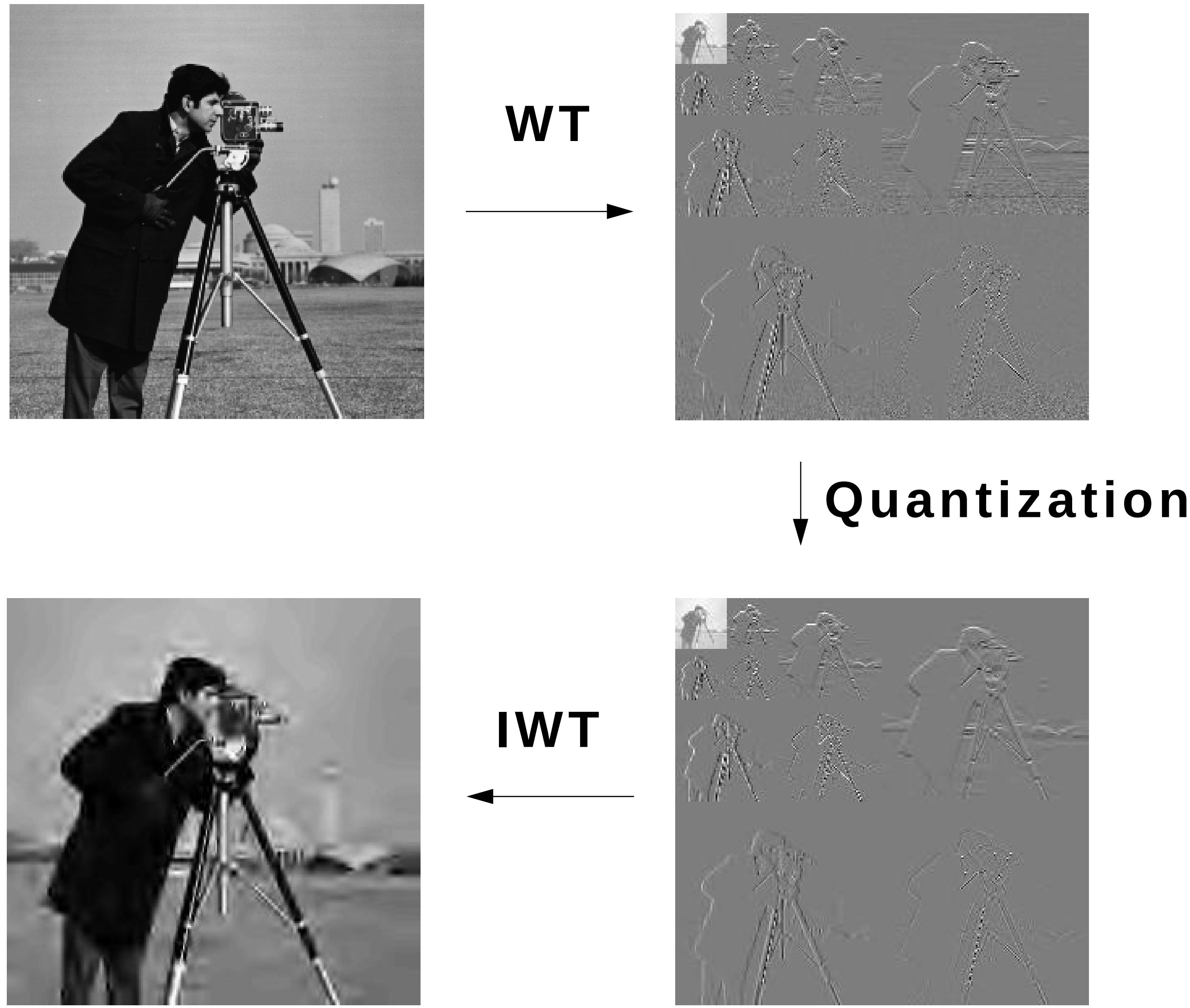
Note

- simple
- at the heart of JPEG 2000
- for jointly Gaussian, standard linear approach (KLT) is optimal

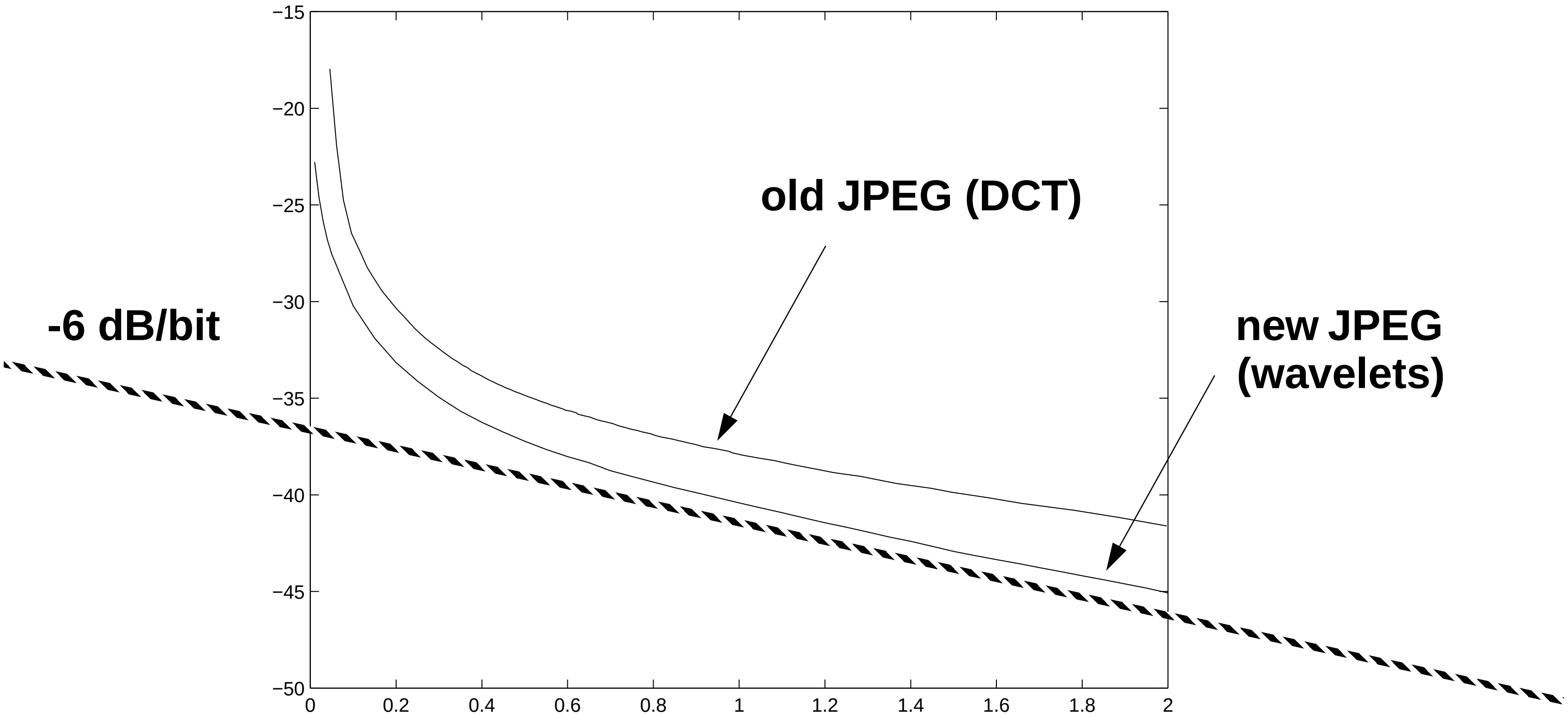
Example: 1D



Example: 2D



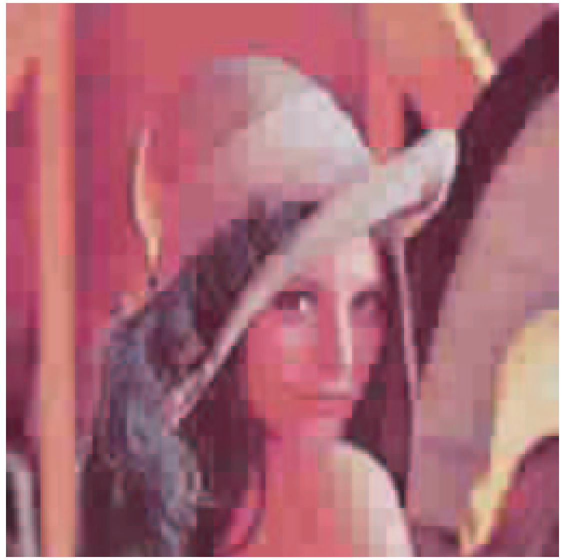
Old Versus New JPEG: D(R) on log scale



Notes

- improvement by a few dB's
- lot more functionalities (e.g. progressive download on internet)
- low rate behavior
- is this the limit?

Original Lena Image (256 x 256 Pixels, 24-Bit RGB)



JPEG Compressed (Compression Ratio 43:1)

JPEG2000 Compressed (Compression Ratio 43:1)

From the comparison, JPEG fails above 40:1 compression while JPEG2000 survives

Images courtesy of www.dspworx.com

So, are wavelets closing the "How many bits for Mona Lisa" question?

(un) fortunately: No!

Reason:

Shannon tells us

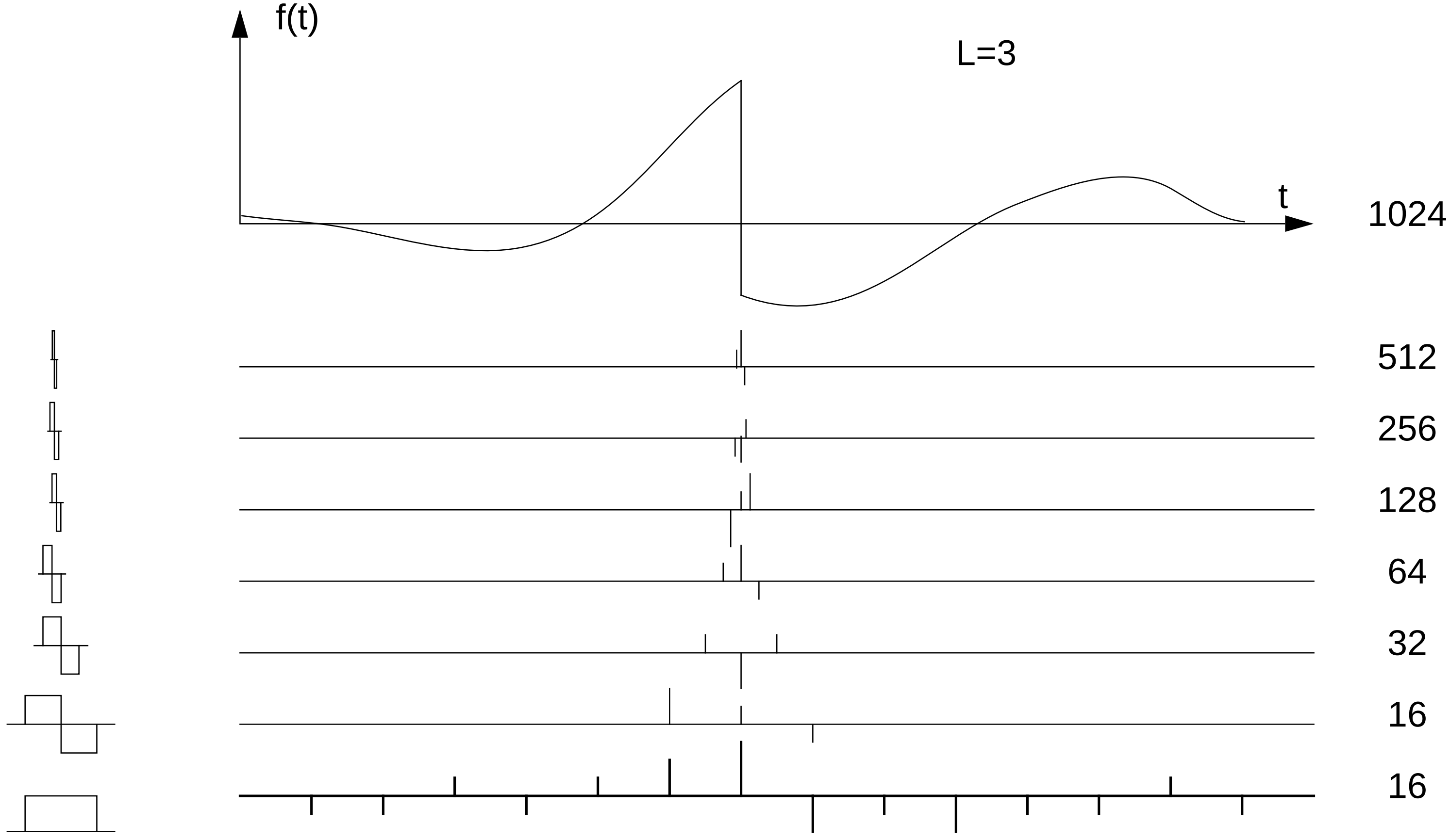
$$D(R) \sim \alpha_1 2^{-\beta_1 R}$$

but wavelets give

$$D_W(R) \sim \alpha_2 \sqrt{R}\, 2^{-\beta_2 \sqrt{R}}$$

for certain classes of simple signals

Reason: independent coding of dependent information

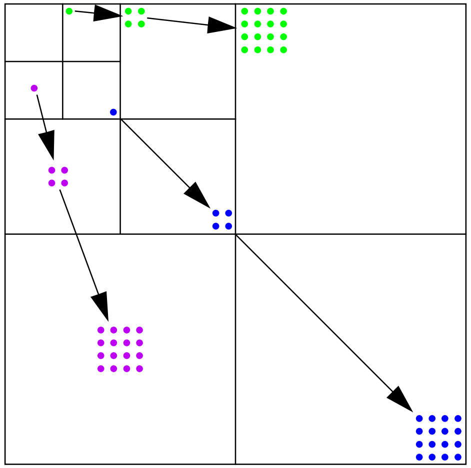


All these wavelets coefficients correspond to a single degree of freedom!

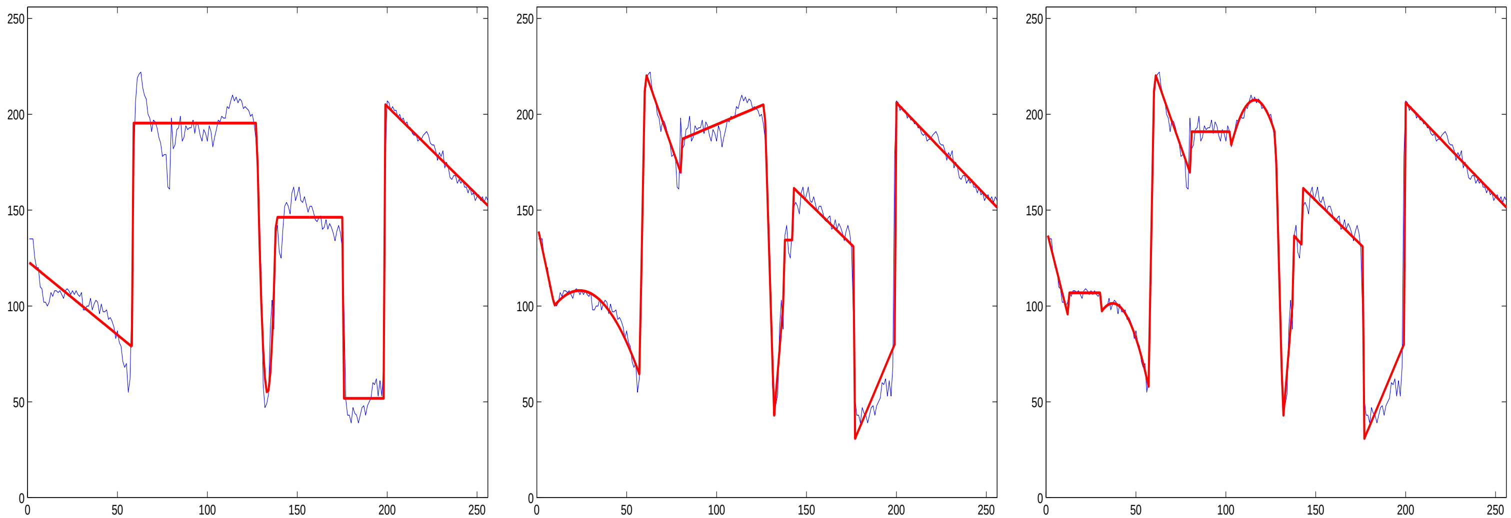
Solution: model dependencies between wavelets coefficients

Various proposals

- Markov models (Baraniuk)
- Zero trees
- Footprints



Example: An optimal algorithm



146 bits 297 bits 351 bits

This uses dynamic programming [Prandoni:00]

Wavelet Footprints [Dragotti:01]

Can we ‘’fix’’ the wavelet scenario?

That is, achieve the same rate-distortion performance as an oracle or a dynamic programming method but with the simplicity of wavelet methods?

The structure of wavelet representation of singularities is simple:

- location: random
- structure accross scales: deterministic!

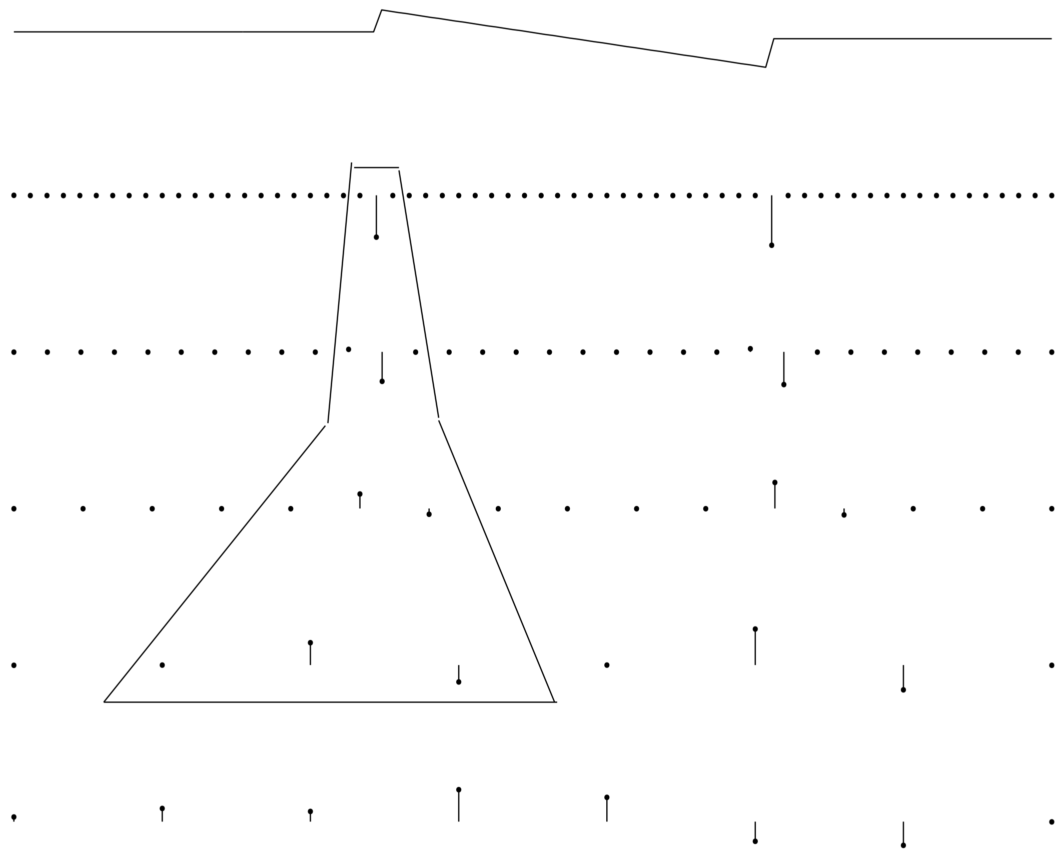
Data structure to capture discontinuities in wavelet domain

- in orthogonal expansion
- in frame

This leads to a simple and intuitive data structure

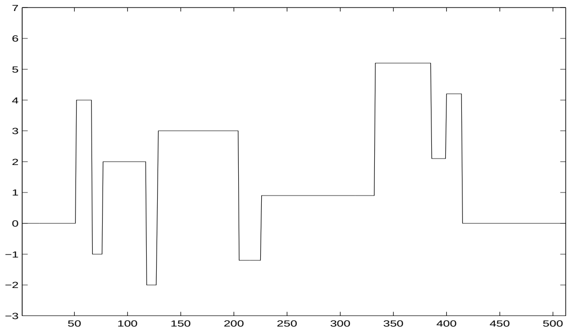
Wavelet Footprint

The wavelet footprint

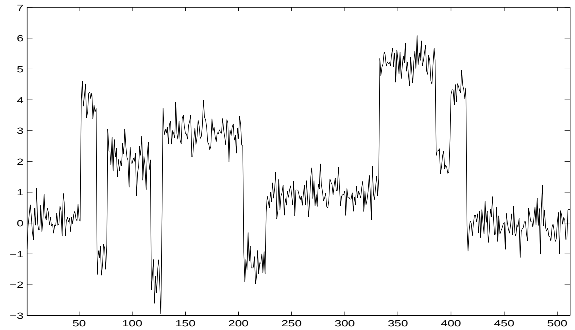


- this is the signature of the discontinuity
- behaviour well understood (classic wavelet analysis)

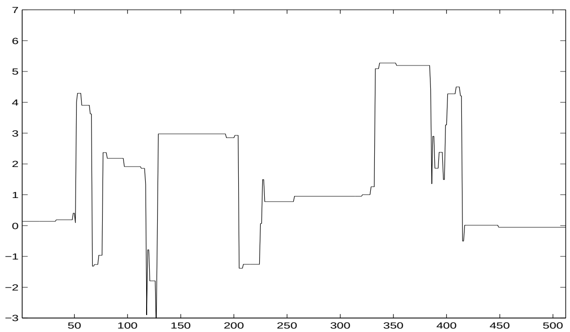
Denoising



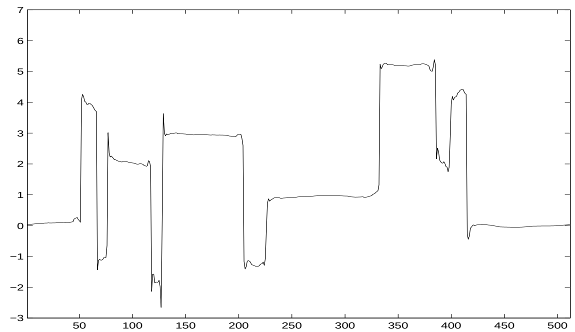
Original signal



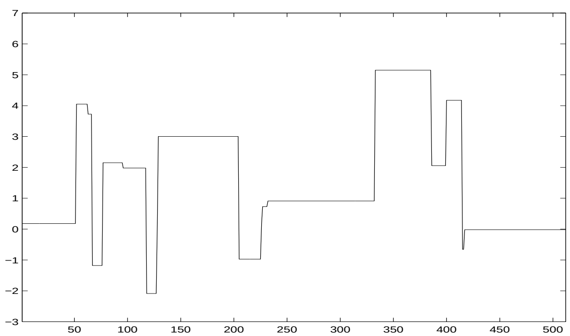
Noisy Signal (SNR=15.62dB)



Hard-Thresholding (SNR=21.3dB)

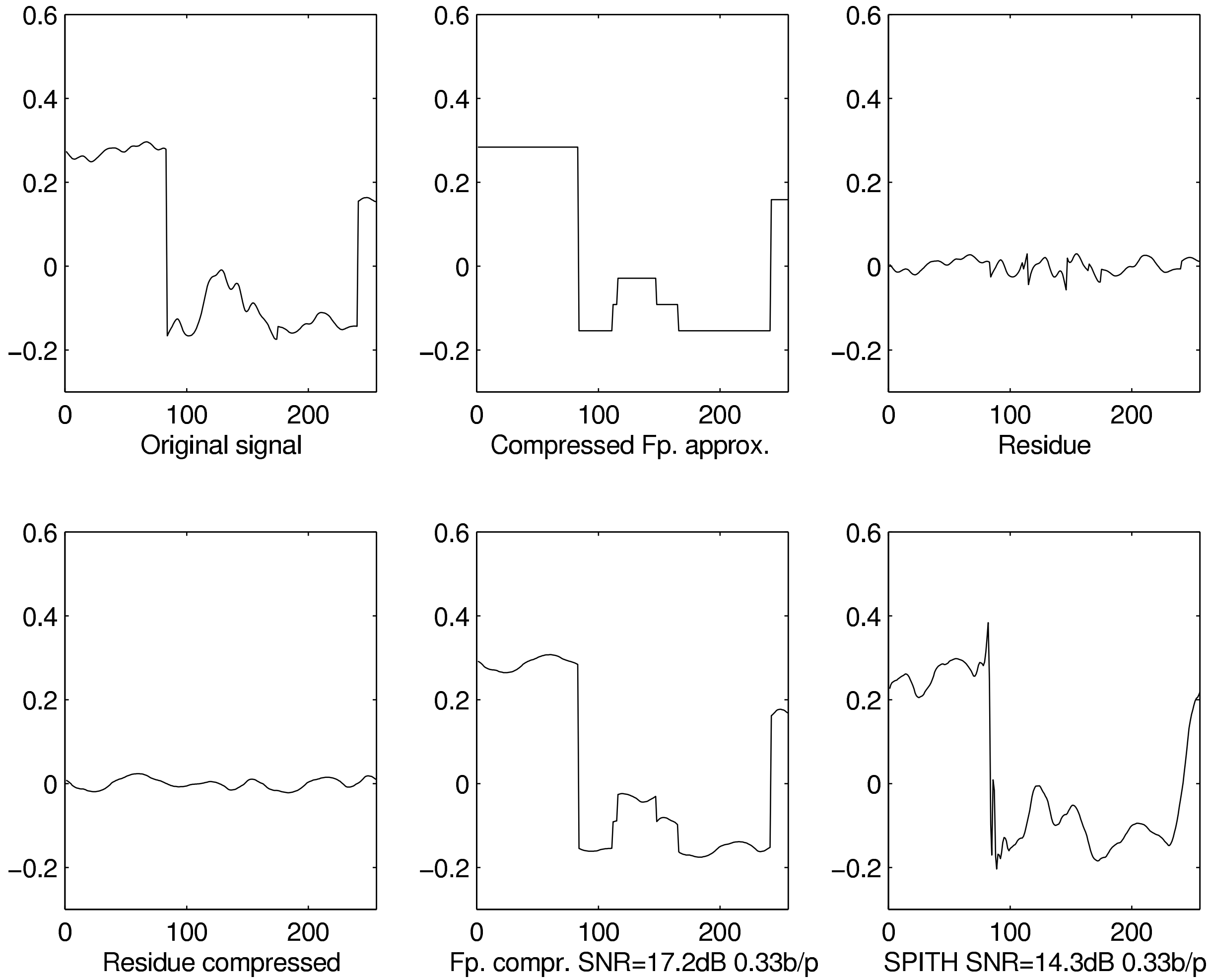


Cycle-Spinning (SNR=25.4dB)



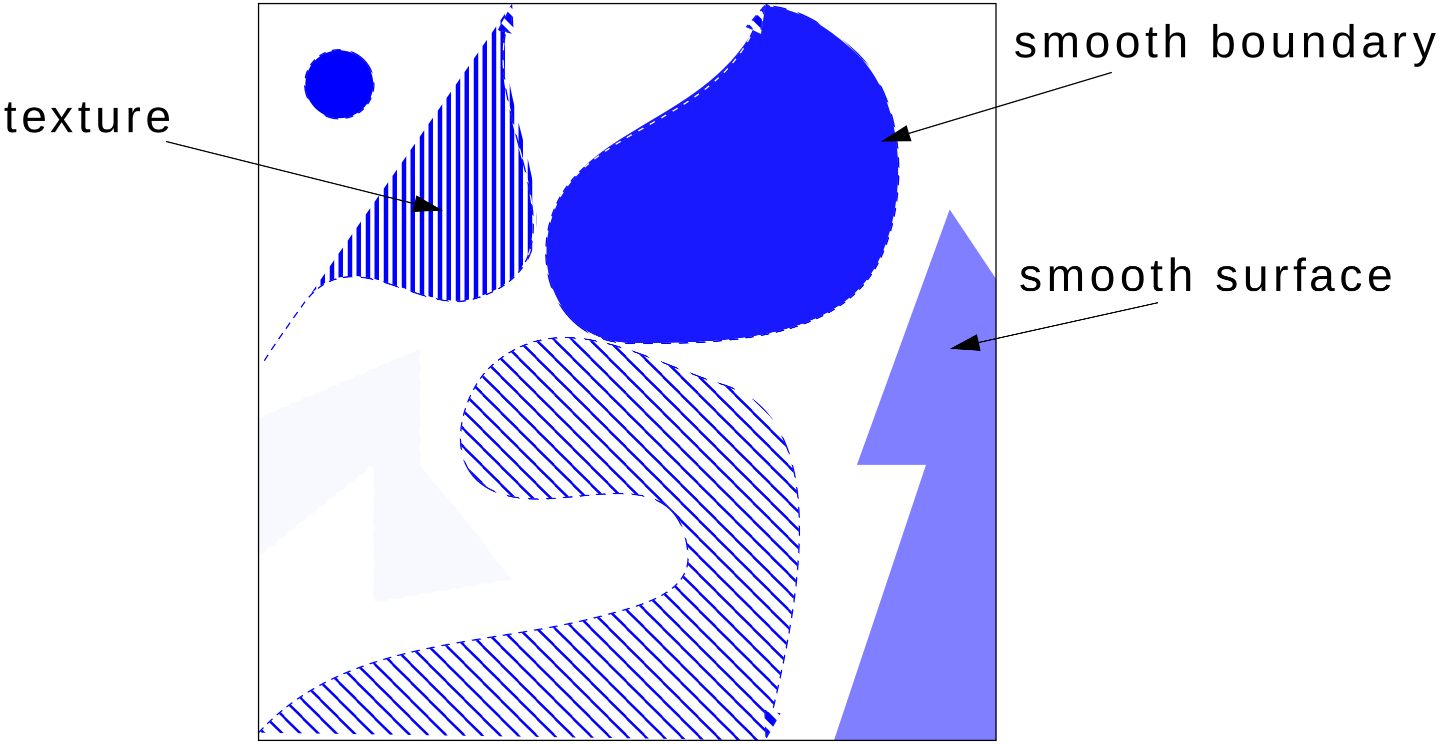
Denoising with Footprints (SNR=27.2dB)

Compression



6. Going to Two Dimensions: Nonseparable Bases

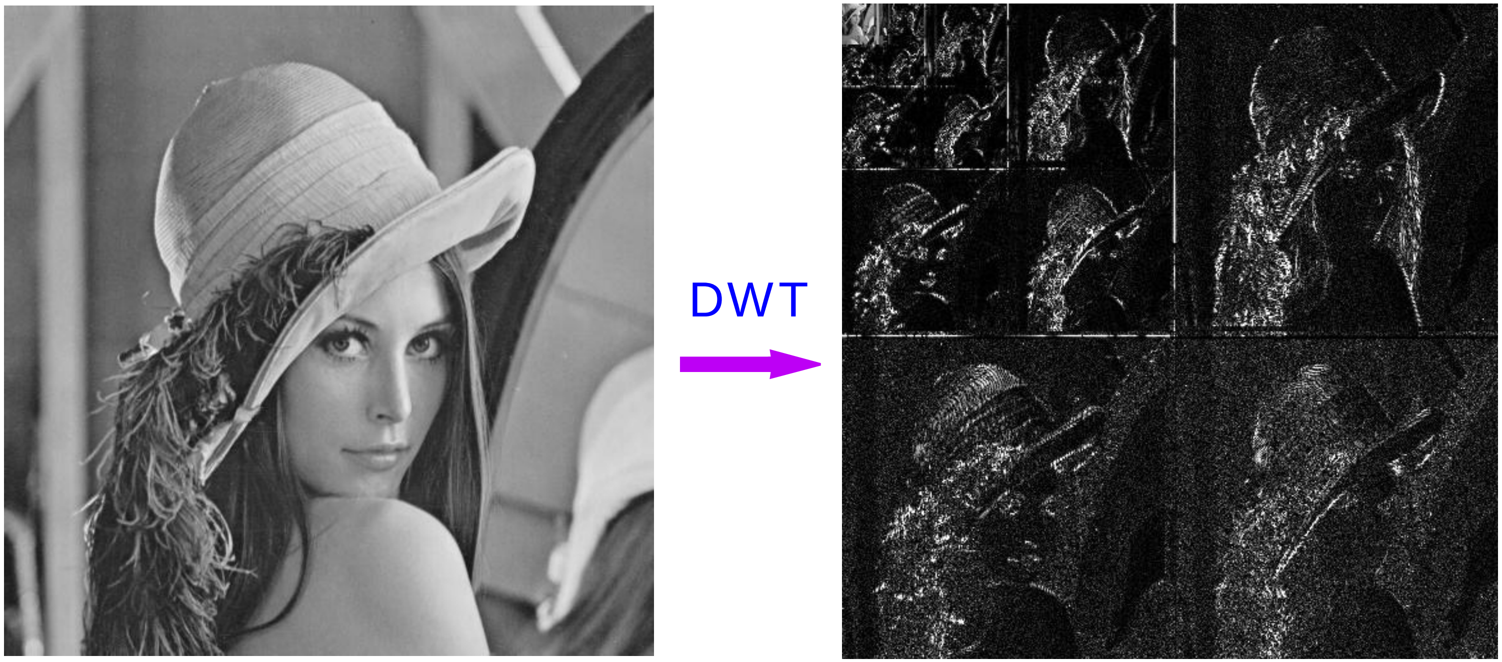
Objects in two dimensions we are interested in



- textures: $D(R) = C_0 \cdot 2^{-2R}$ per pixel
- smooth surfaces: $D(R) = C_1 \cdot 2^{-2R}$ per object!

Current approaches to two dimensions....

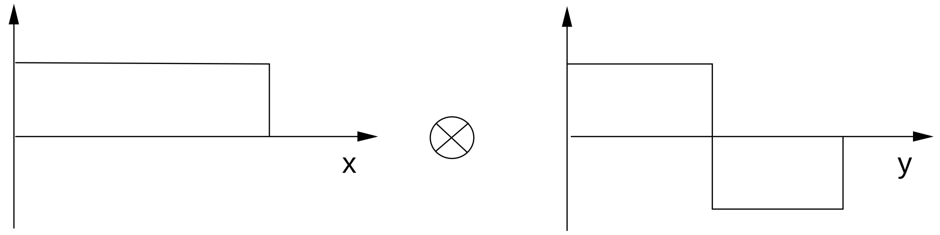
Mostly separable, direct products



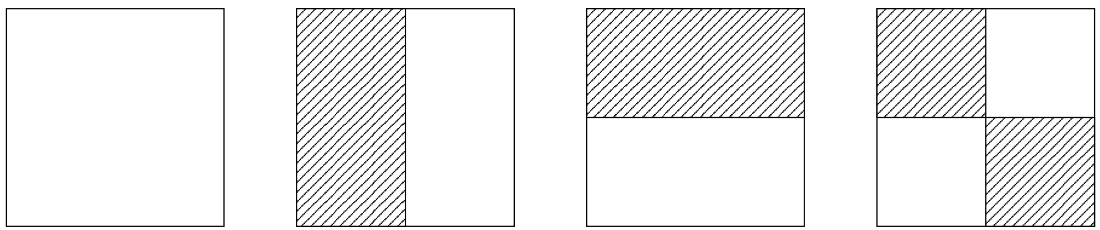
Wavelets: good for point singularities
but what is needed are sparse coding of edge singularities!

Two dimensonal wavelet bases

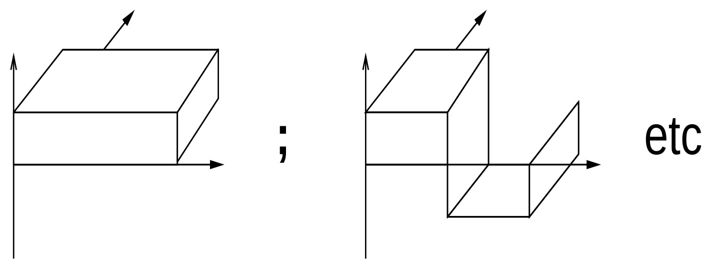
Ex: Tensor products of Haar functions



That is



or in 3D



That is very little directionality!

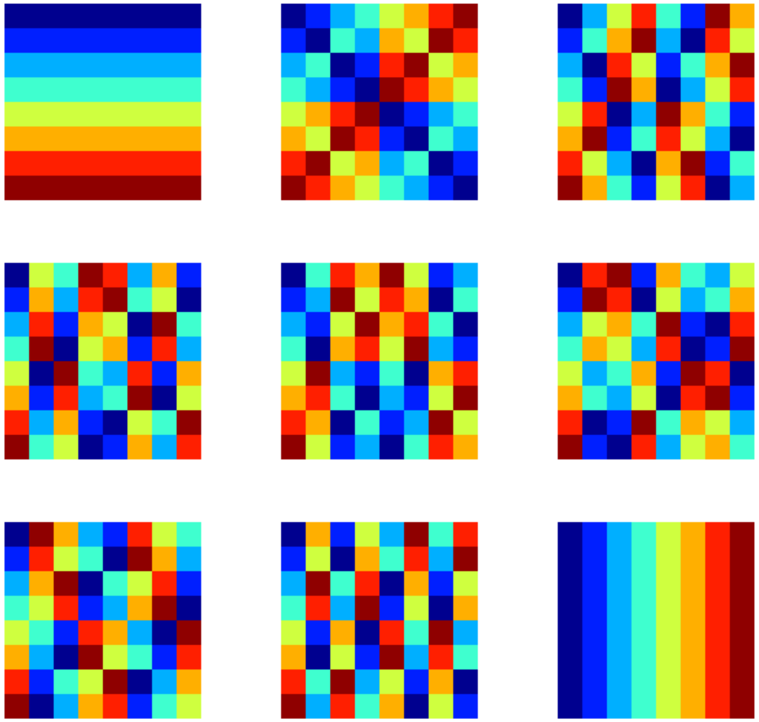
What is needed are directional bases

- Local Radon transform
- Ridgelets
- Curvelets
- Contourlets
- etc

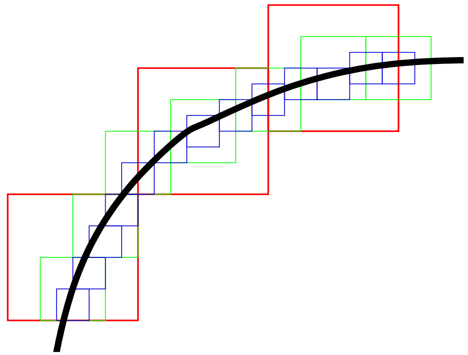
That is:

a zoo of true two dimensional animals

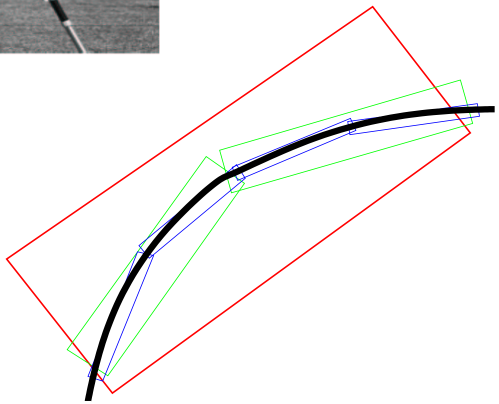
Example: a directional block transform [Do:01]



Multiresolution Contour Approximation



Wavelet



Xlet

Consider object c^2 boundary between two cst

- # of wavelet coeffs: 2^j
- # of curvelet coeffs: $2^{j/2}$

Rate fo approximation, M-term NLA

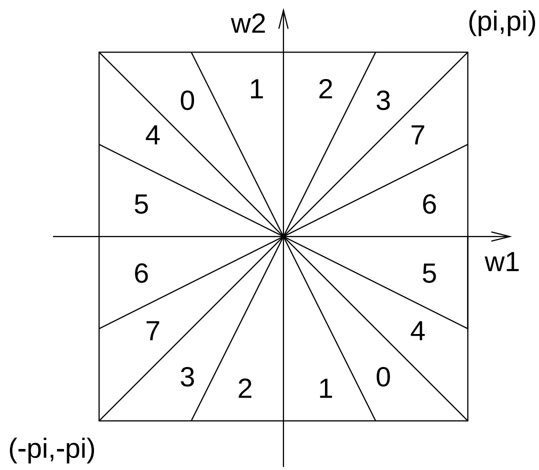
- Fourier: $O(1/\sqrt{M})$
- Wavelets: $O(1/M)$
- Curvlets: $O(1/M^2)$

Operational Solution

Directional Analysis (as in Radon transform)
+
Multiresolution as in wavelets

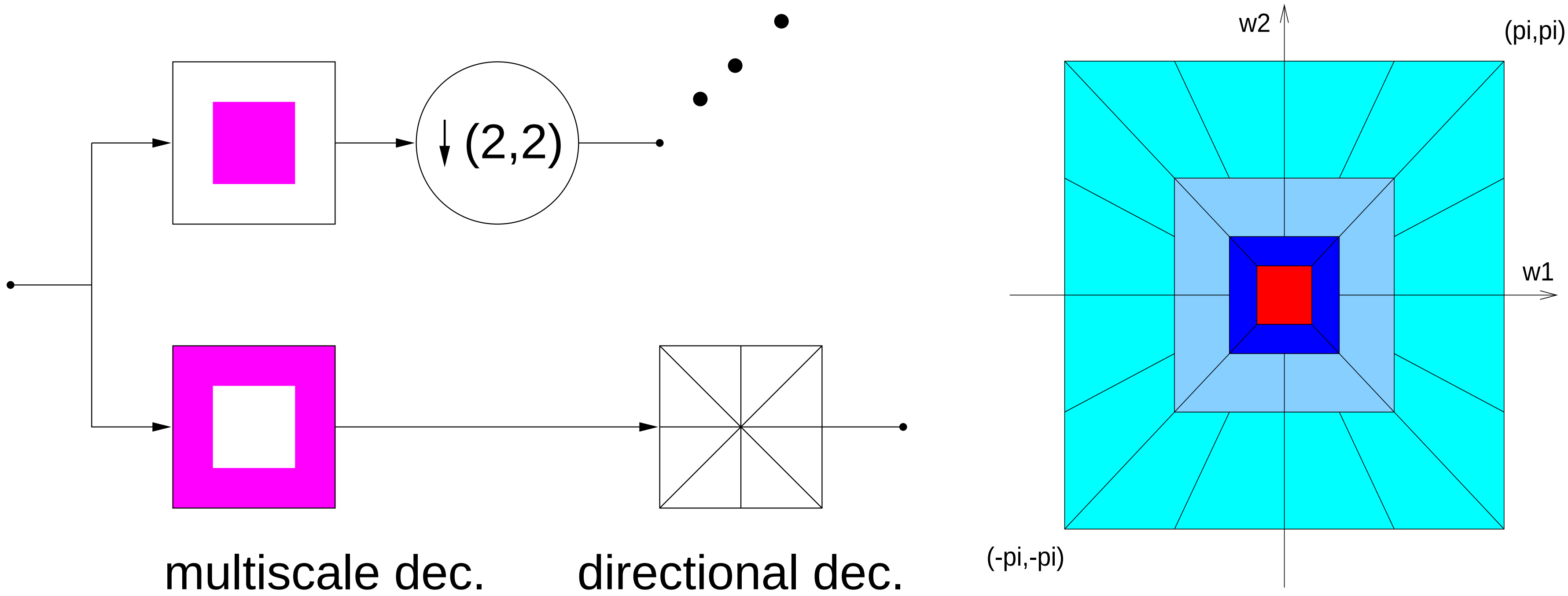
Directional Filter Banks

- division of 2-D spectrum into fine slices using iterated tree structured filter banks



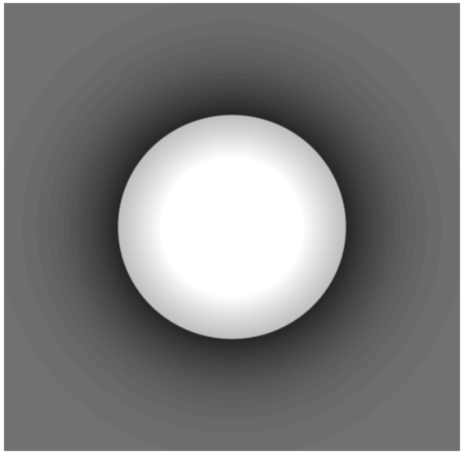
Pyramidal Directional Filter Banks (PDFB)

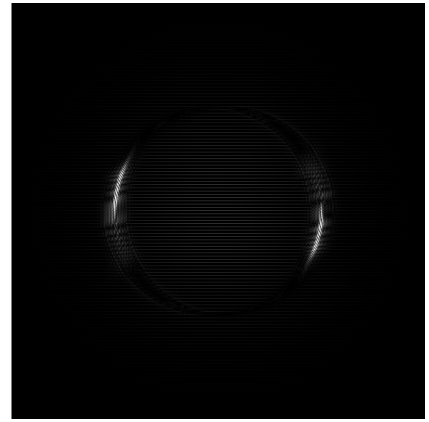
Motivation: + add multiscale into the directional filter bank
+ improve its non-linear approximation power

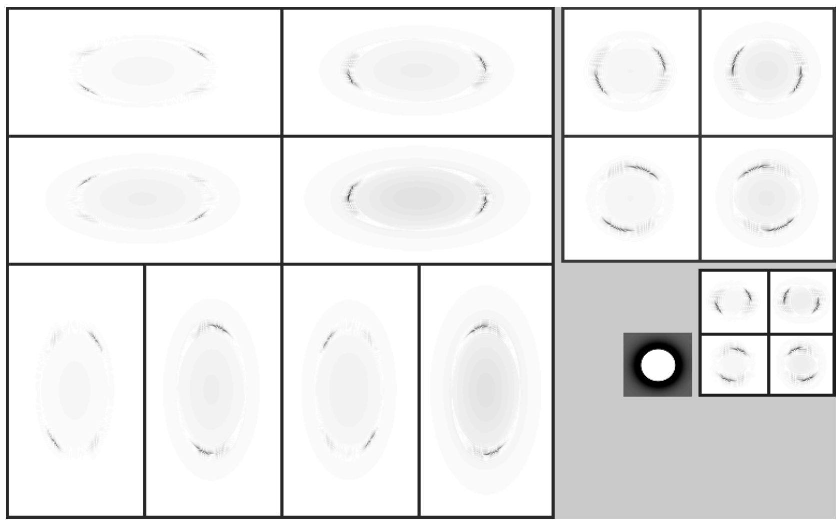


Properties: + Flexible multiscale and directional representation for images (can have different number of direction at each scale!)

Example: A pyramidal directional filter bank







Compression, denoising, inverse problems: mostly open!

7. Conclusions

Multiresolution is good for you!

- Perception and mathematics (mostly) agree...

Non-linear can buy a lot...

- in approximation, the difference can be huge!

Compression is hard but generic

- understanding complexity is fundamental

Multiple dimension is (infinitely) harder than one...

The search for the ultimate basis is a fascinating and timeless topic

References

For a tutorial:

M. Vetterli, Wawelets, Approximation and Compression, Signal Processing, May 2001

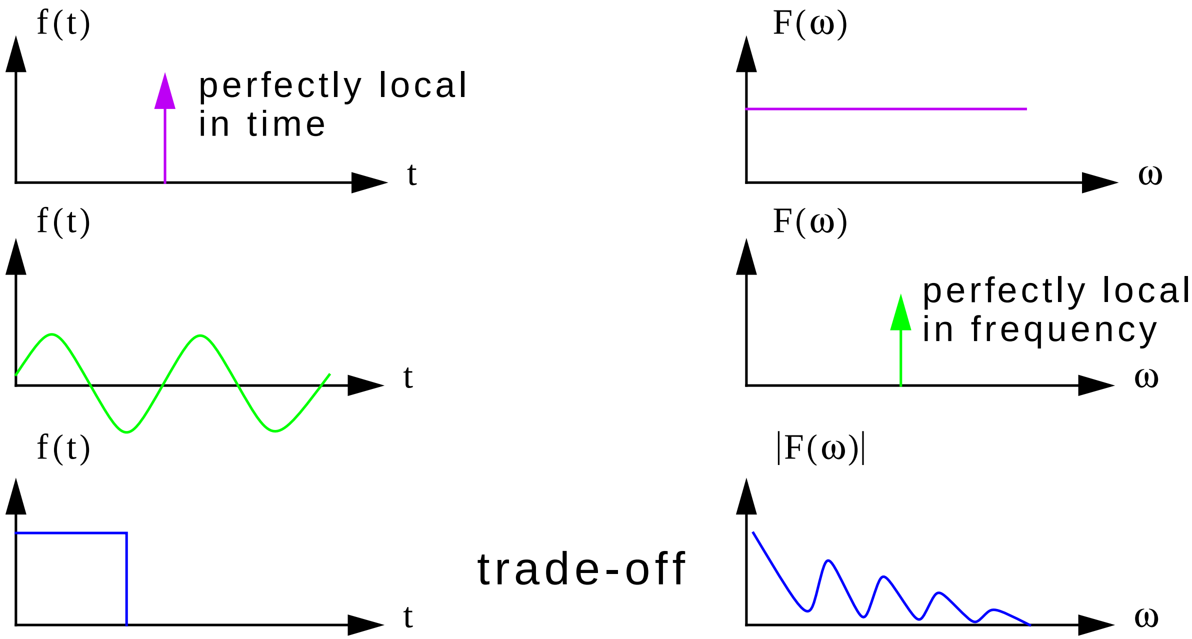
For more details

P. L. Dragotti, M. Vetterli, Wavelet Footprints, IEEE Transactions on Signal Processing, submitted

M. Do, M. Vetterli, Pyramidal Directional Filter Banks, to be submitted, 2002

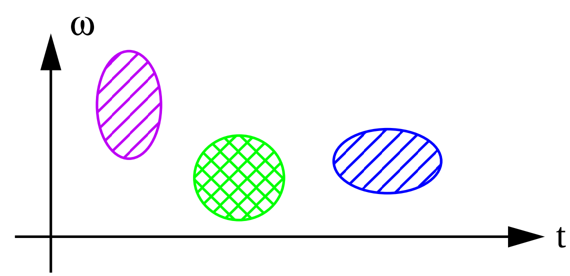
Appendix:

1930: Heisenberg discovers that you cannot have your cake and eat it too!



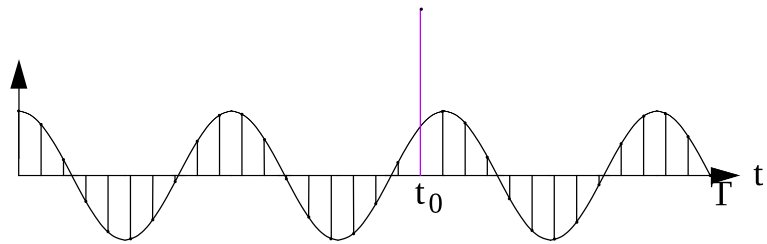
Uncertainty principle

- lower bound on TF product

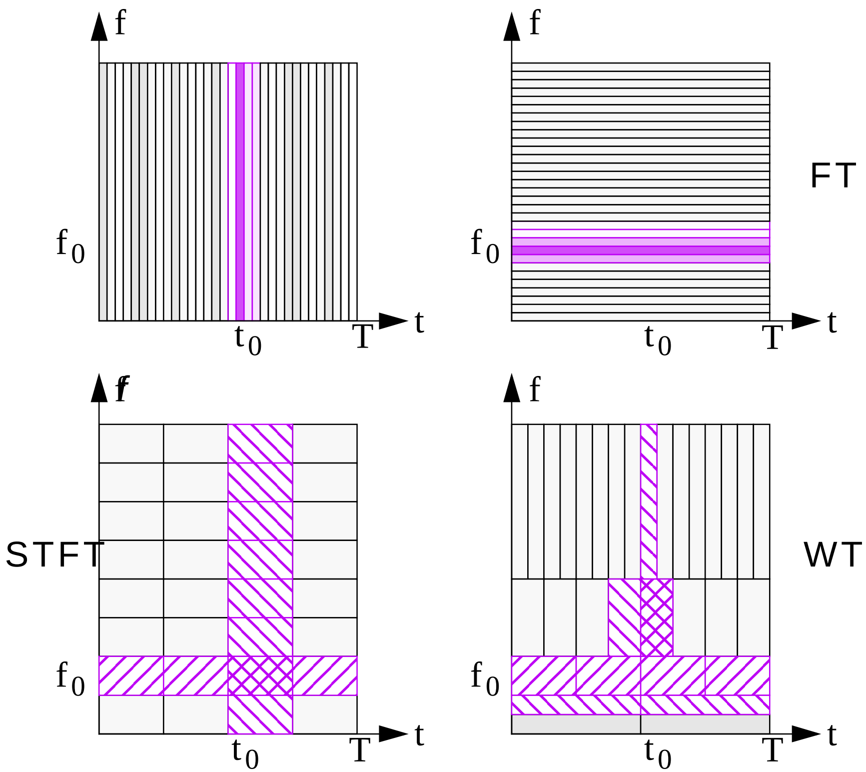


Time-frequency tiling for a

sine + Delta



$$x[n] = \cos 2\pi f_0 + A\delta[n - t_0]$$



so....

what is a good basis?