Filter Banks and Wavelets:
Relationships, New Results and Applications

Martin Vetterli, Columbia University

April 30 1990, NJIT

Overview

I Introduction

II Wavelets, Filter Banks and Multiresolution Analysis

- An early example
- STFT versus Wavelet transform
- Filter banks versus pyramids

III Theoretical results

- PR FB with unequal bandwidth
- Regular linear phase wavelets
- Multidimensional PR FB

IV Applications

- Progressive to interlaced conversion
- Multiresolution coding for HDTV

V Conclusions

Acknowledgements

C.Herley (CU)

J.Kovacevic (CU)

K.M.Uz (CU)

D.Anastassiou (CU)

O.Rioul (CNET)

D.LeGall (C^3)

I INTRODUCTION

- Applied Mathematics: theory of wavelets

- Computer Vision: multiresolution signal analysis

- Digital Signal Processing: subband coding

All the same!

Construction of orthonormal bases for signal expansion

- Particularity: different trade-off's between frequency and time resolution.

- Fundamental difference: modulation is replaced by scaling!

  scaling - SCALING

  modulation - MoDuLaTiOn

- Continuous case: more elegant!

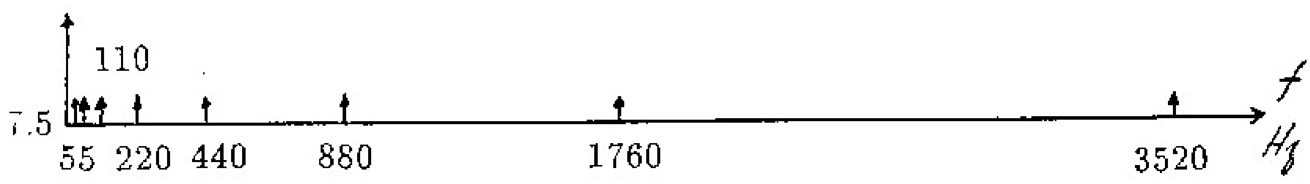
- Discrete case: more useful....

An early example:

Analysis of polyphonic music for music synthesis

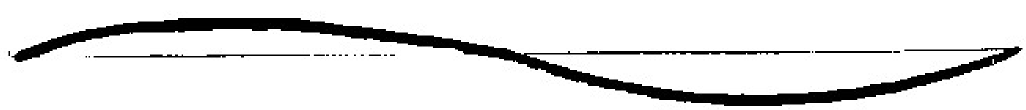
(CCRMA, Stanford, Tech. Rep. 1985)

Problem: piano has 8 octaves + harmonics

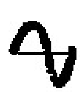


sampling at 44KHz

27.5 Hz = 1600 samples



440 Hz = 100 samples

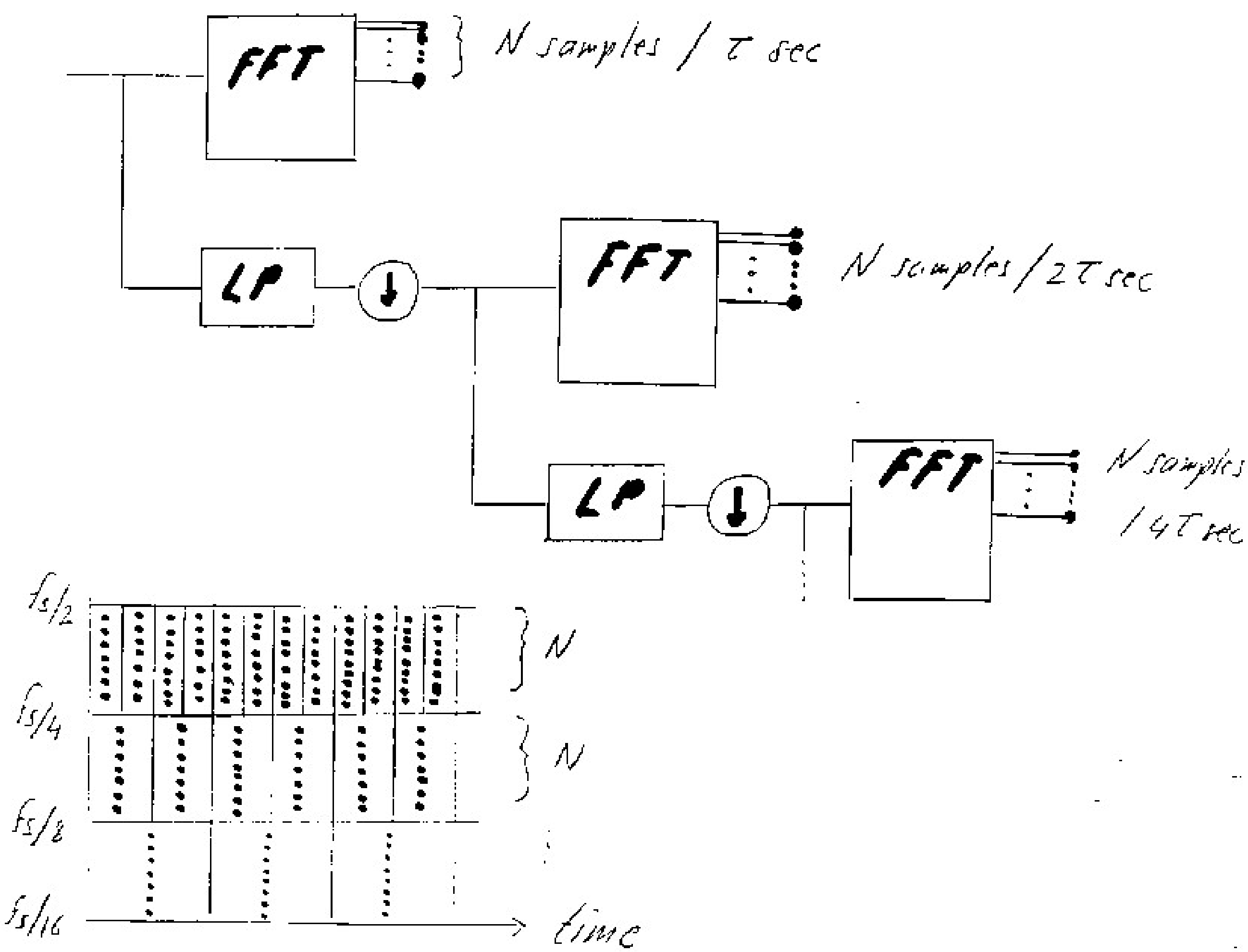


3520 = 12.5 samples

Solution: bounded Q transform

- Take running FFT of size 4N
- Keep upper half of the spectrum (N samples)
- Low-pass filter, subsample by 2, iterate

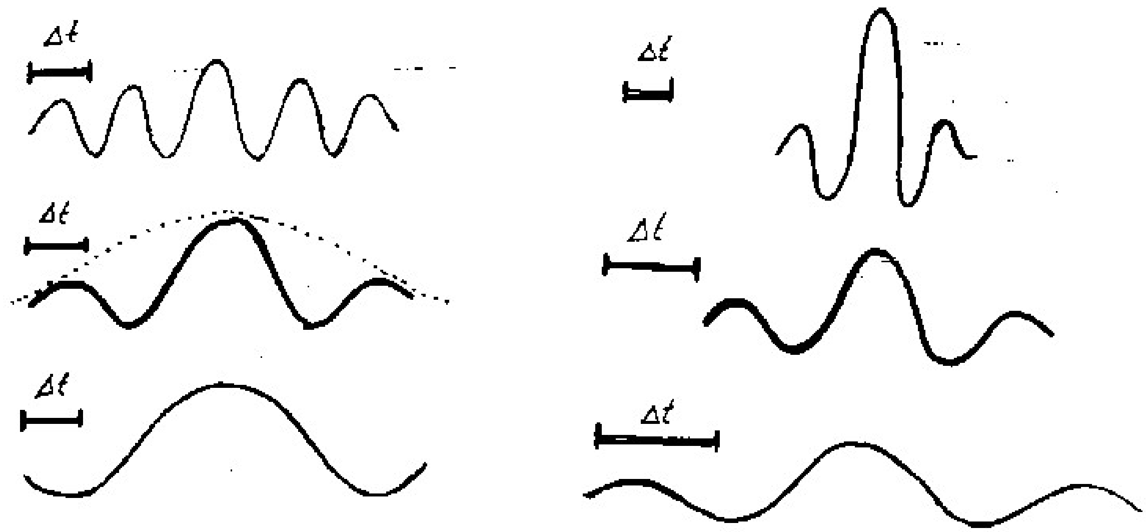
Result: octave by octave "spectrogram"



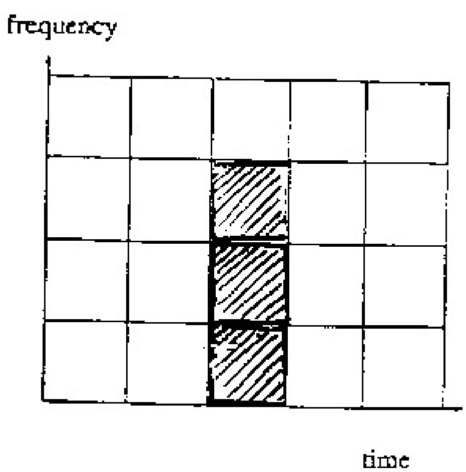
STFT versus Wavelet Transform

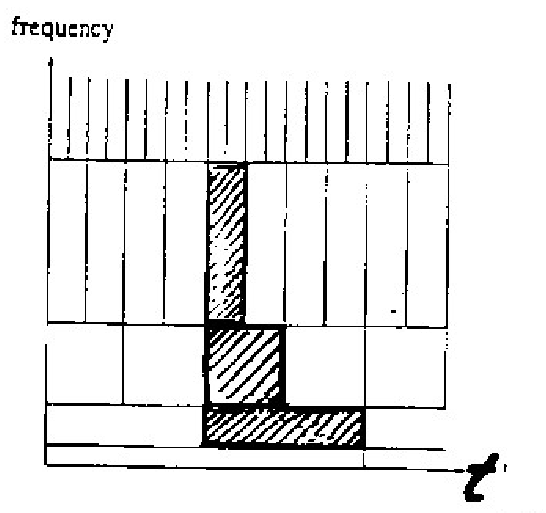
Basis Functions

$c_{mn}(t) = e^{-jm\omega_0 t} w(t - nt_0)$ $\qquad$ $h_{mn}(t) = a_0^{-m/2} h(t a_0^{-m} - n b_0)$



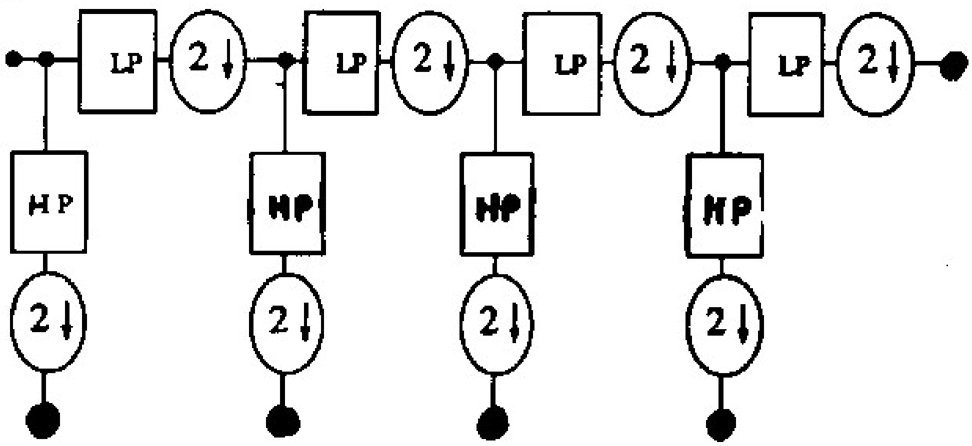
Sampling of time-frequency plane



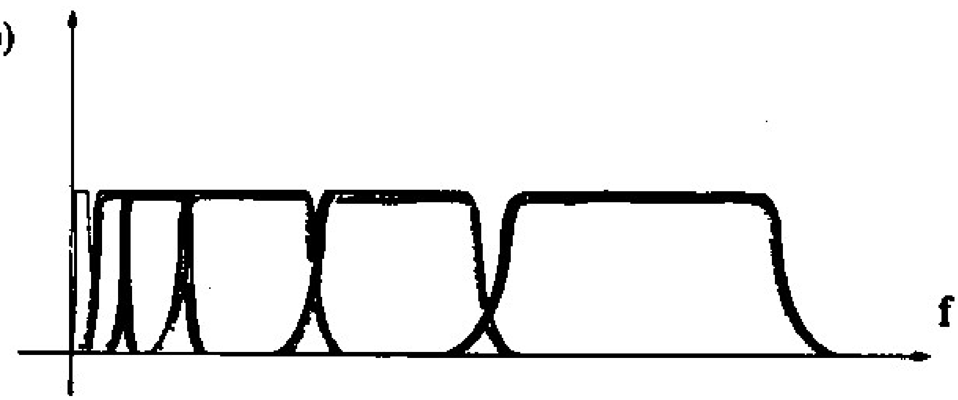


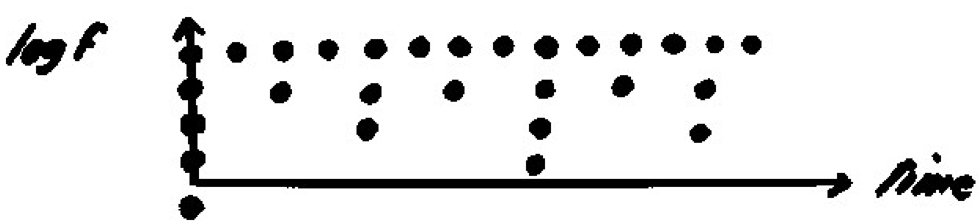
Discrete wavelet transform implemented with filter banks

a)



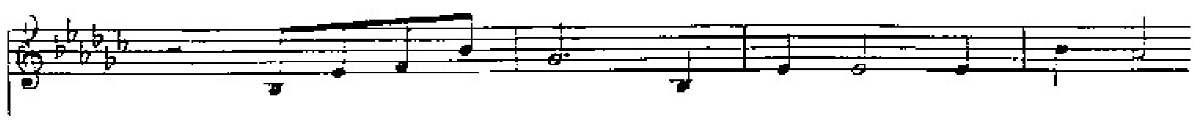
b)



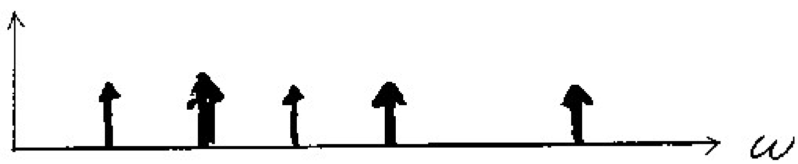


Scaling versus modulation

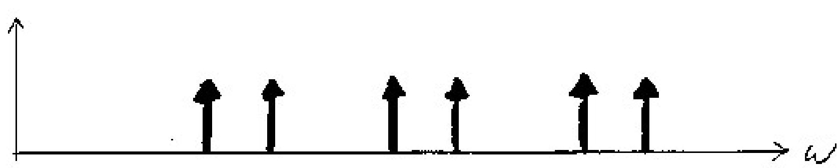
Musical score: logarithmic "scale"



Scaled version: x(2n)→ $X(\omega/2)$

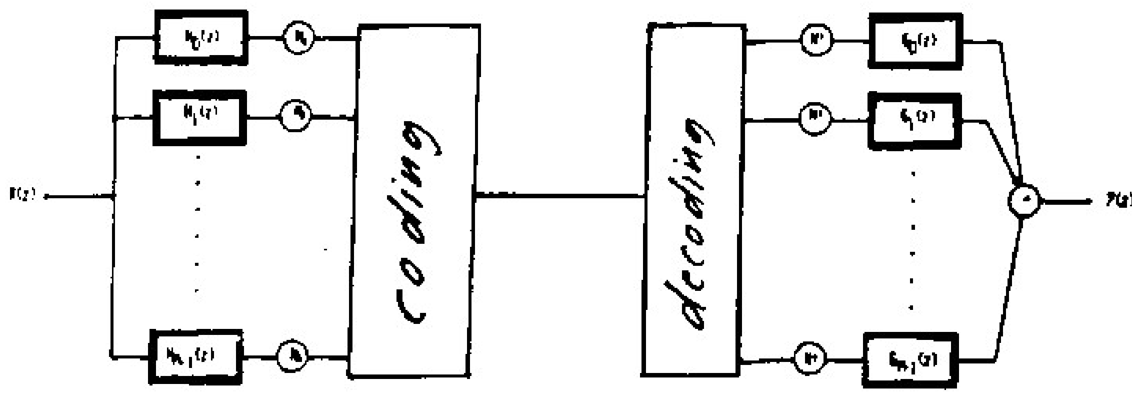


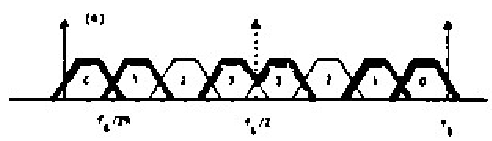
Modulated version: W^{kn} x(n)→ $X(\omega - k)$



Sub-Band Coding

- Divide a signal into sub-sampled sub-bands
- Apply adequate coding in the sub-bands





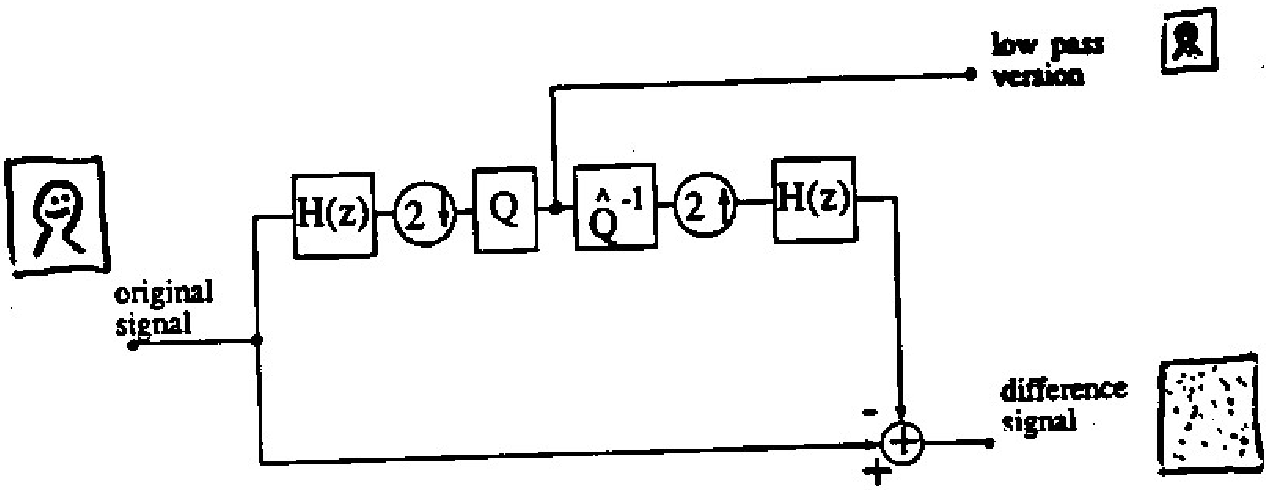
Speech: successful for medium compression

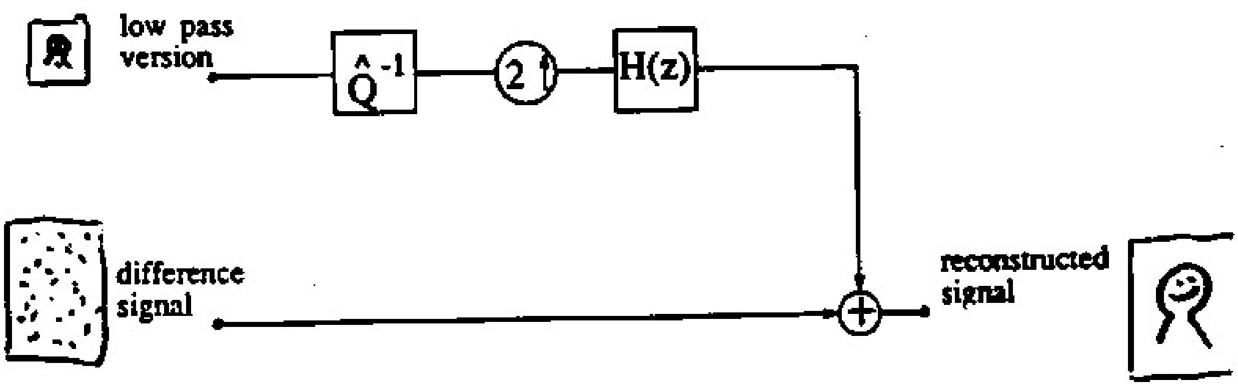
Images: seems promising

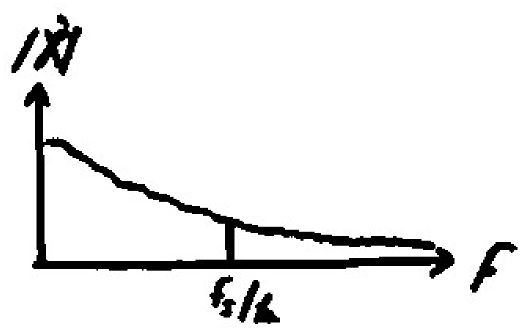
- improvement over DCT: no blocking
- improvement over pyramid: critical sampling

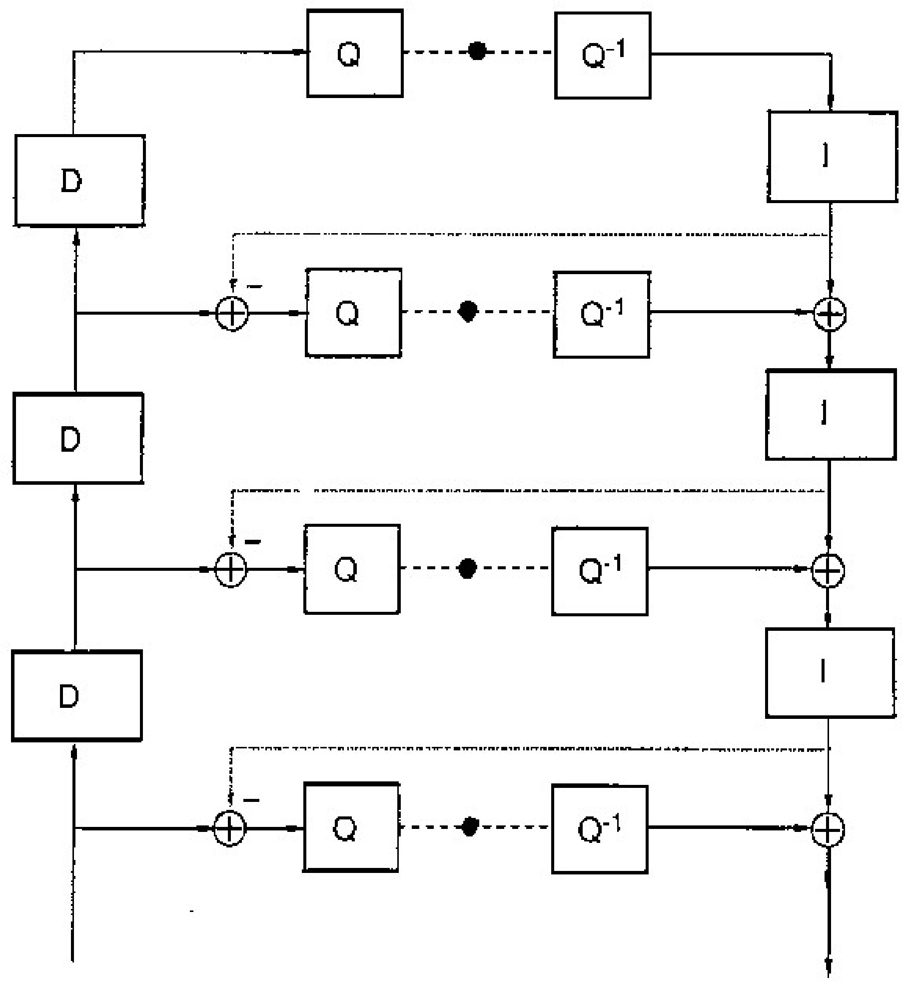
Video: under investigation

Pyramidal or Multiresolution Schemes








Q
Q-1
D
I
Q
Q-1
D
I
Q
Q-1
D
I
Q
Q-1

Pyramids versus critically sampled filter banks

Subband coders:

(1) critically sampled

(2) constrained filter design

(3) poor lowband....

(4) linear processing required

(5) sensitive to quantization

Pyramid coders:

(1) non critically sampled

(2) arbitrary filers

(3) good lowband

(4) non-linear processing possible

(5) only quantization of last difference matters...

(6) robust to errors in higher layers

- Oversampling in pyramids becomes negligible as dimensionality grows:

  1D: 1+1/2+1/4...< 2

  2D: 1+1/4+1/16...< 1.333

  3D: 1+1/8+1/64...< 1.1.142

- Bounds on quantization noise are weak in transform/subband schemes:

  Ex: max error of quantizer is δ = PCM error

  max error in pyramid is δ

  max error in DCT is $\sqrt{N}\delta$

  Ex:

$$\begin{pmatrix} \sqrt{2}\delta \\ 0 \end{pmatrix} = 1/\sqrt{2} \cdot \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} \delta \\ \delta \end{pmatrix}$$

- Inclusion of non-linear processing:

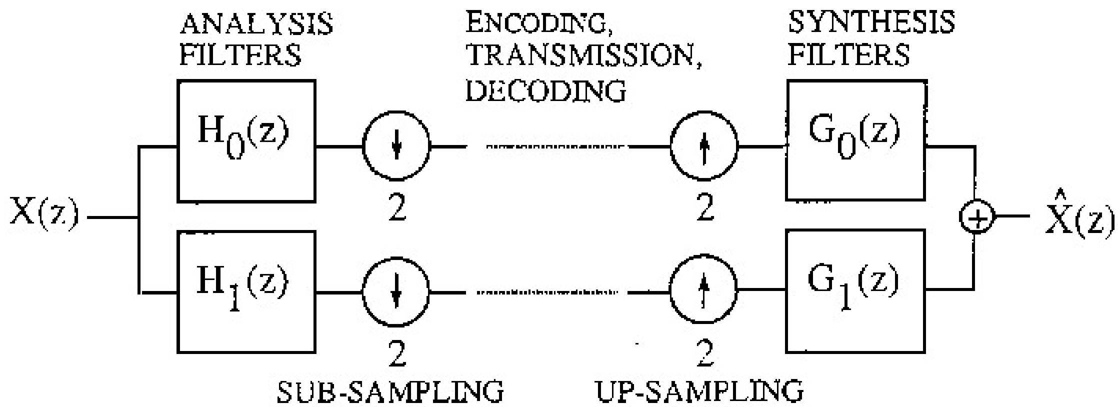
  difficult in subband coders

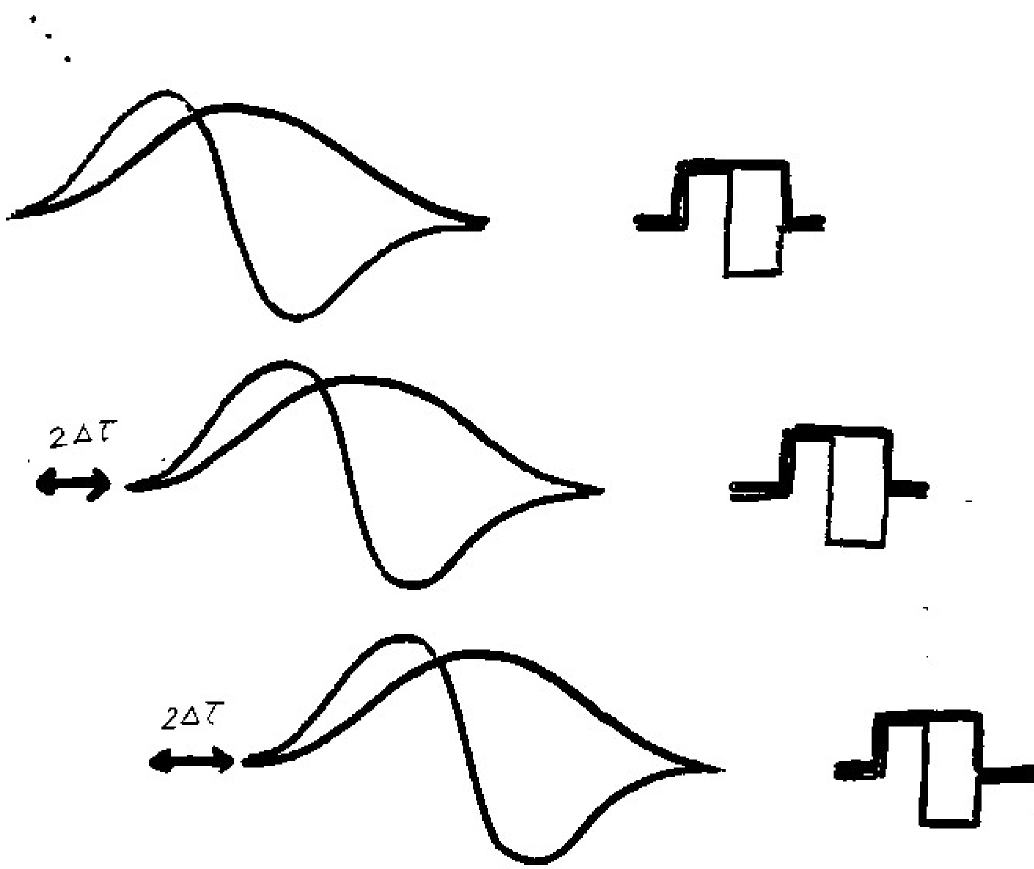
  easy in pyramids

- Multiresolution framework also useful in motion estimation:

  Hierarchical motion estimation

Two channel perfect reconstruction filter bank





Making an orthogonal base (FIR)

(i) Autocorrelation of $\{h_0(n)\}$:

single non-zero even coefficient at $n = 0$

$$H_0(z) \cdot H_0(z^{-1}) + H_0(-z) \cdot H_0(-z^{-1}) = 1$$

thus $H_0(z)$ has odd degree

(ii) Crosscorrelation $\{h_0(n), h_1(n)\}$: zero even terms

$$H_0(z) \cdot H_1(z^{-1}) + H_0(-z) \cdot H_1(-z^{-1}) = 0$$

Sufficient: $H_1(z) = z^{-2l+1} \cdot H_0(-z^{-1})$

then: $H_0(z) \cdot H_1(z^{-1}) = z^{-2l+1} \cdot H_0(z) \cdot H_0(-z)$ is odd

Necessity: $H_0(z) \cdot H_1(z^{-1}) = z^{2n+1} Q(z^2)$

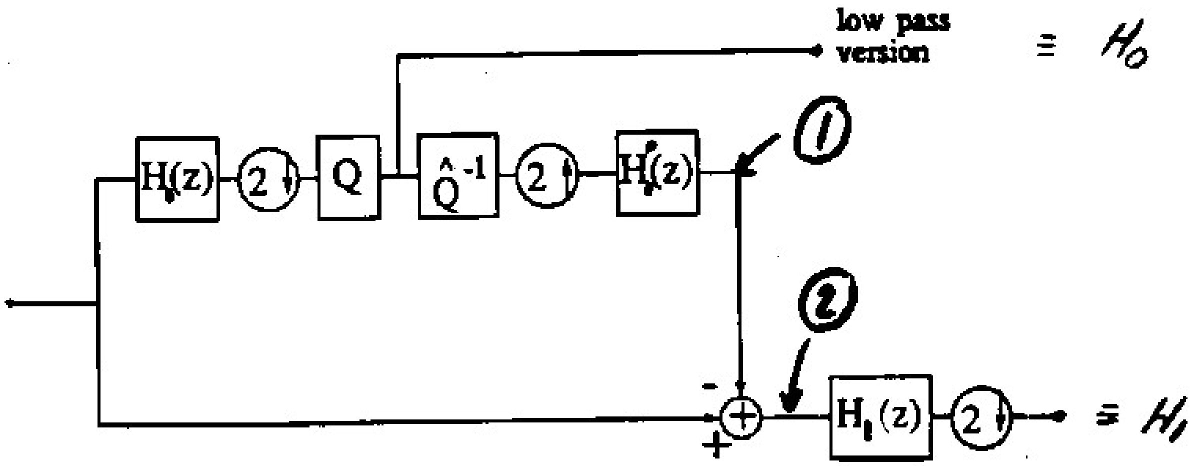
zero at $\alpha \Rightarrow$ zero at $-\alpha$

$H_0(z)$ cannot have a zero at α and at $-\alpha$

$H_0(z)$: zero at $\alpha \Rightarrow$ then $H_1(z)$: zero at $-1/\alpha$

Thus: the impulse responses $h_0(n), h_1(n)$ and their even shifted versions form an orthonormal basis on which the input can be projected and then reconstructed with the same basis vector

Back to multiresolution systems!



$$H_0 = \begin{bmatrix} \ddots & & & & & & & \\ 0 & h_{00} & h_{01} & \cdots & h_{0N-2} & h_{0N-1} & 0 & 0 \\ 0 & 0 & 0 & h_{00} & \cdots & & h_{0N-2} & h_{0N-1} \\ & & & & \ddots & & & \ddots \end{bmatrix}$$

① $H_0^* H_0$

② $I - H_0^* H_0$

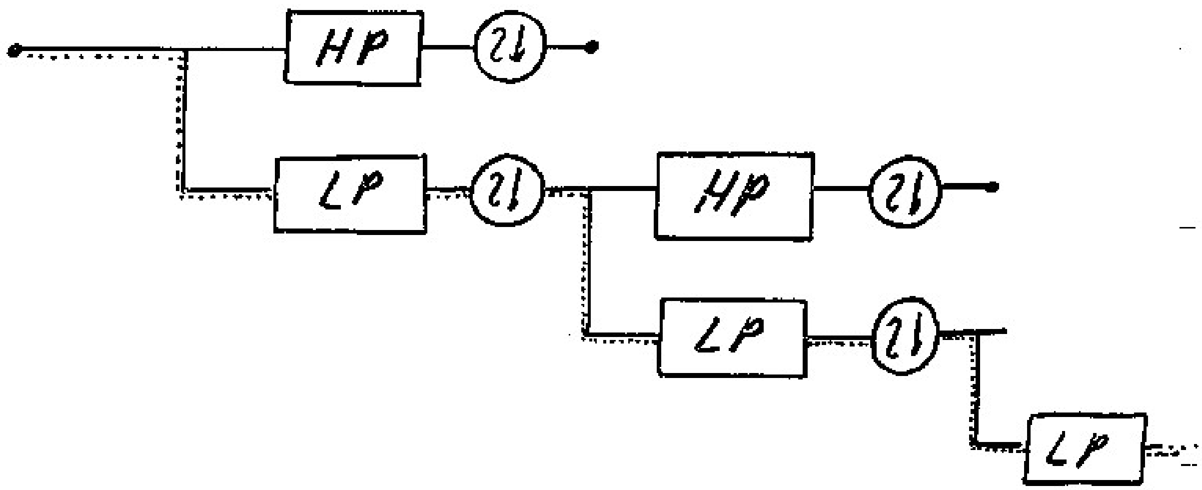
③ $H_0^* H_0 + H_1^* H_1 = I$ (PR System)

④ $I - H_0^* H_0 = H_1^* H_1$

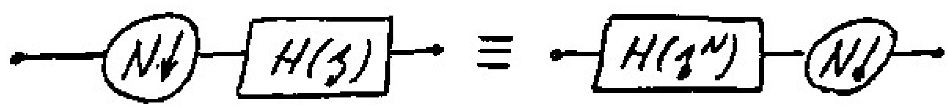
⑤ $H_1 H_1^* = I_{1/2} \Rightarrow H_1 H_1^* H_1 = H_1$

⑥ PR with H_0^*, H_1^*

The Regularity Question



Now:



Equivalent lowpass filter :

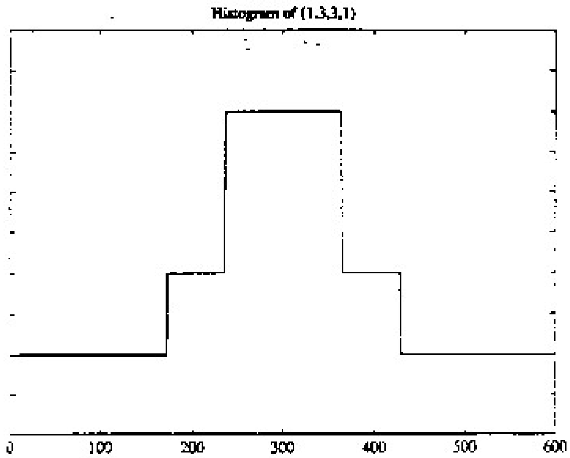
$$H_\infty(z) = \prod_{i=0}^{\infty} H(z^{2^i})$$

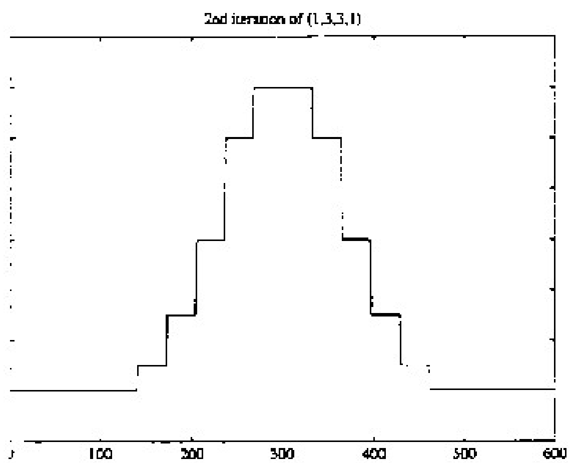
Question: does this converge to a continuous function ?

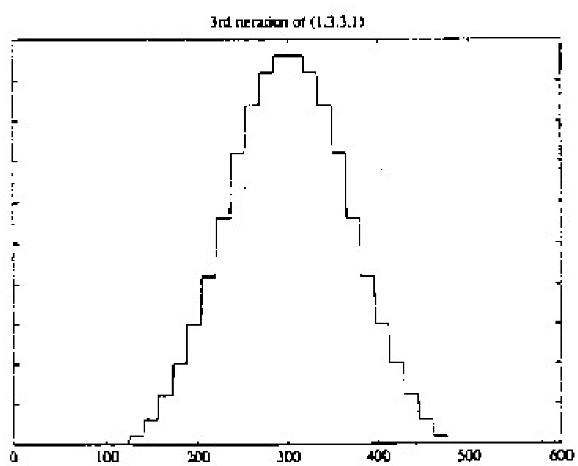
[Daubechies 1988]: $H_0(e^{j\omega}) = [1/2(1+e^{j\omega})]^N \, F(e^{j\omega})$

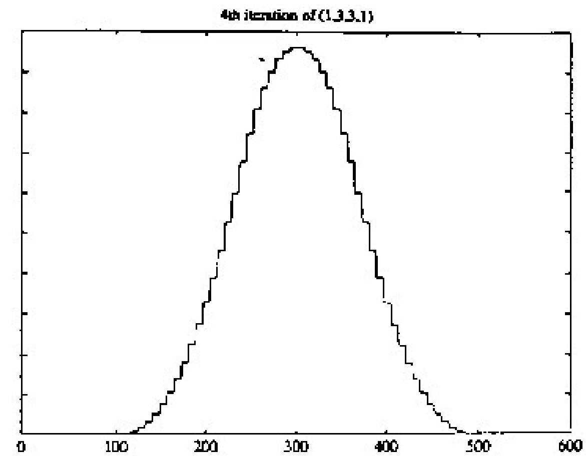
if $\sup_{\omega \in [-\pi,\pi]} | F(e^{j\omega}) | < 2^{N-1}$

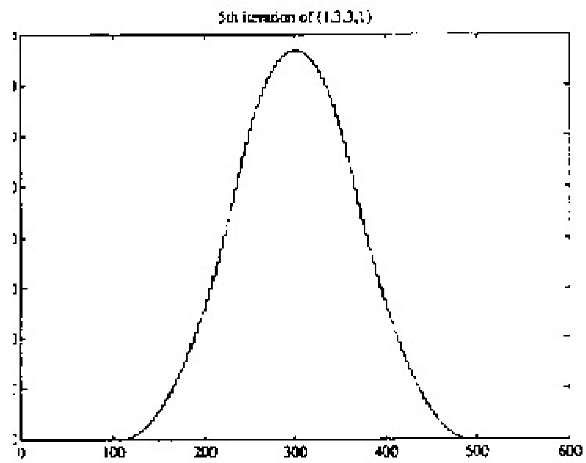
then $H_\infty(z)$ converges to a continuous function

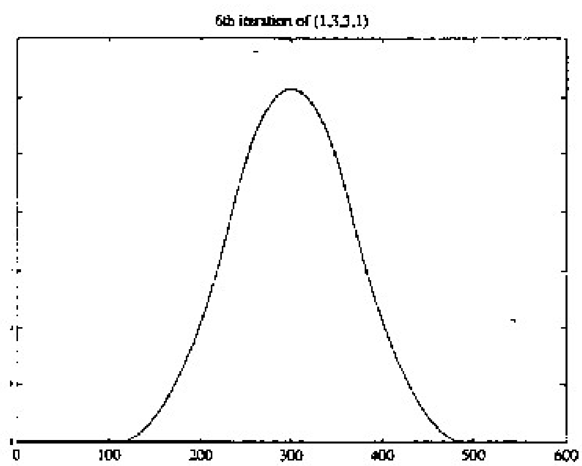

Histogram of (1,3,3,1)
0
100
200
300
400
500
600


2nd iteration of (1,3,3,1)
100
200
300
400
500
600

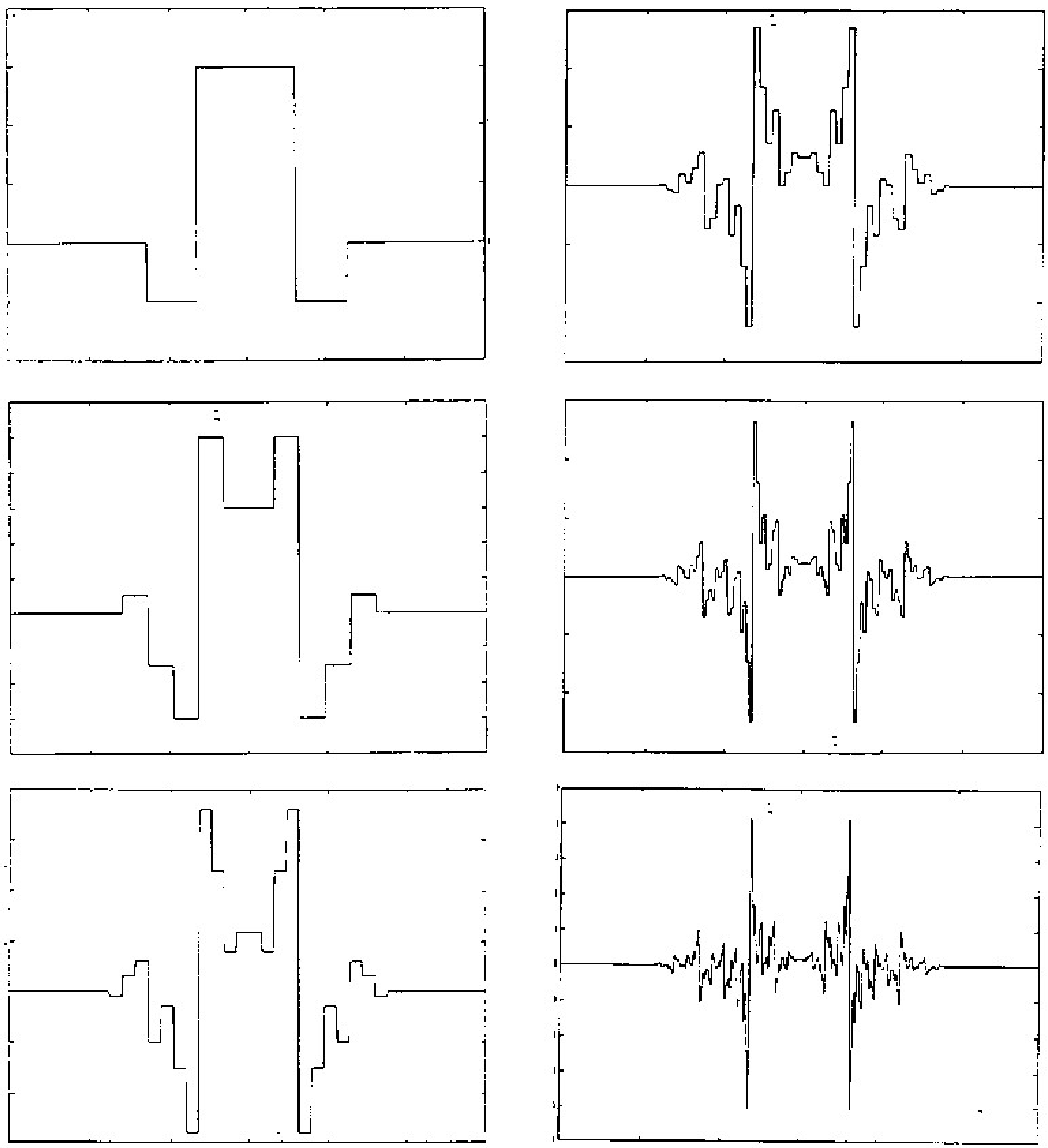

3rd iteration of (1,3,3,1)
0
100
200
300
400
500
600


4th iteration of (1,3,3,1)
0
100
200
300
400
500
600


5th iteration of (1,3,3,1)
0
100
200
300
400
500
600


6th iteration of (1,3,3,1)
0
100
200
300
400
500
600

-133-1



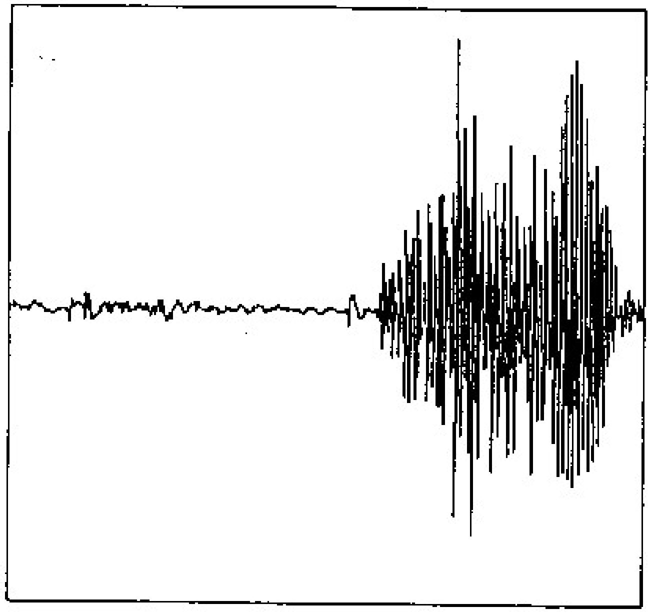


Fig 1: 0.82 s of speech filtered and subsampled 4 times by (1, -3, -3, 1)

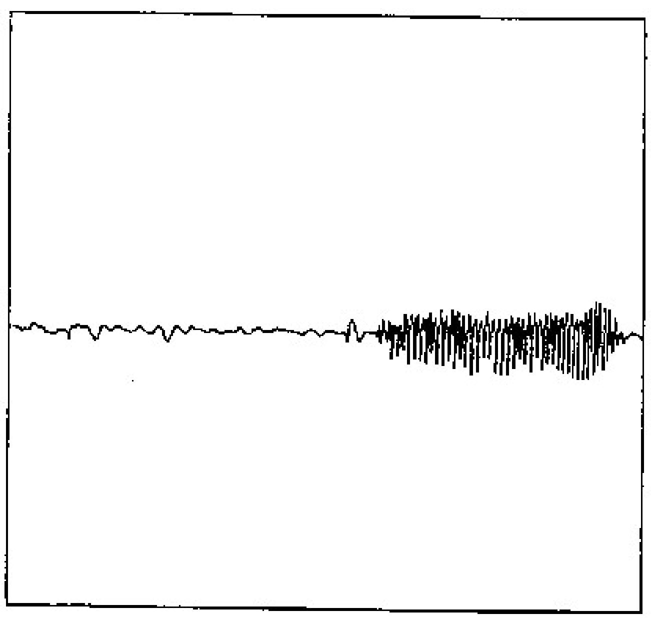
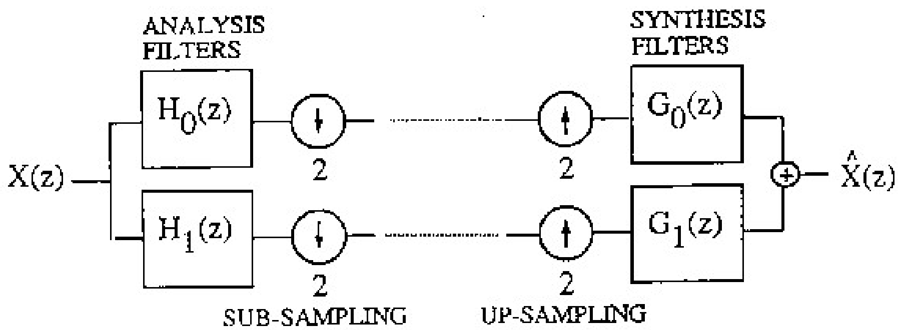


Fig 2: 0.82 s of speech filtered and subsampled 4 times by (1, 3, 3, 1)

Design of Regular Linear Phase Wavelets



Aliasing cancellation:

$G_0(z) = H_1(-z)$ and $G_1(z) = -H_0(-z)$

Perfect reconstruction:

$$\begin{aligned}\Delta(z) &= H_0(z)H_1(-z) - H_0(-z)H_1(z) \\ &= P(z) - P(-z) \\ &= cz^{-k-1}\end{aligned}$$

Paraunitary solution:

$$H_1(z) = z^{-L+1}H_0(-z^{-1}) \Rightarrow G_0(z) = z^{-L+1}H_0(z^{-1})$$

Regular analysis implies Regular synthesis

General case: Biorthogonal system

Check regularity of both $H_0(z)$ and $G_0(z)$!

Complementary filter method

- Choose $B(z) = (1+z)^N$
- Find a linear phase $R(z)$ such that:

$$B(z)R(z) - B(-z)R(-z) = z^{-k+1}$$

- Factorize:

$$P(z) = B(z)R(z) = H_0(z)H_1(-z)$$
$$= (1+z)^{N+1}\prod_i R_i(z)$$

- Gather factors into two linear phase regular filters:

$$H_0(z) = (1+z)^{N_1}\prod_{i\epsilon S_0} R_i(z)$$

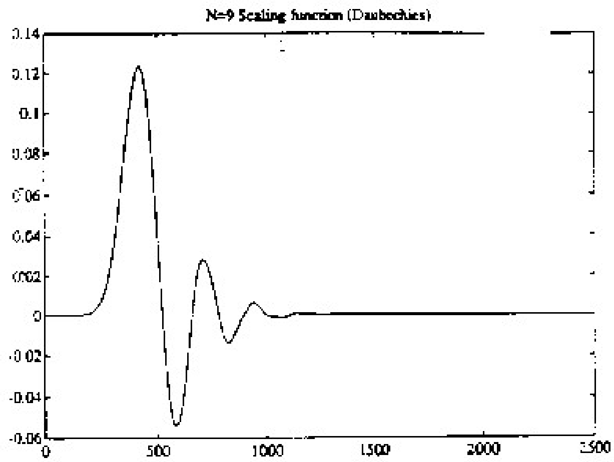
$$H_1(z) = (1+z)^{N_2}\prod_{i\epsilon S_1} R_i(z)$$

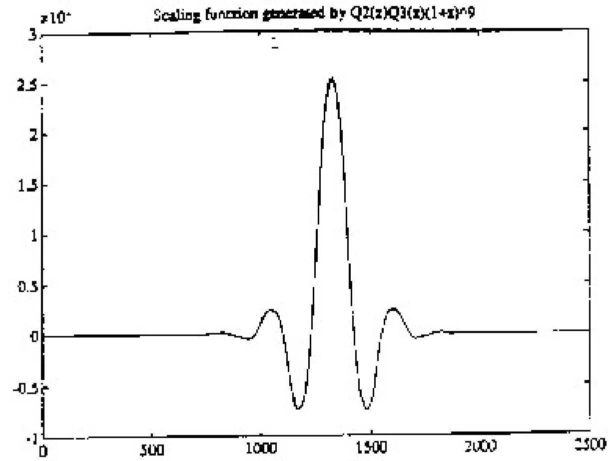
Alternative approach:

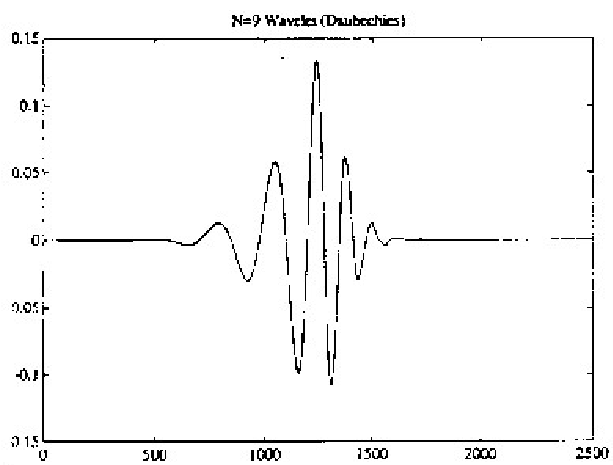
Solve the linear phase lattice so that $H_0(z) = (1+z)^N$

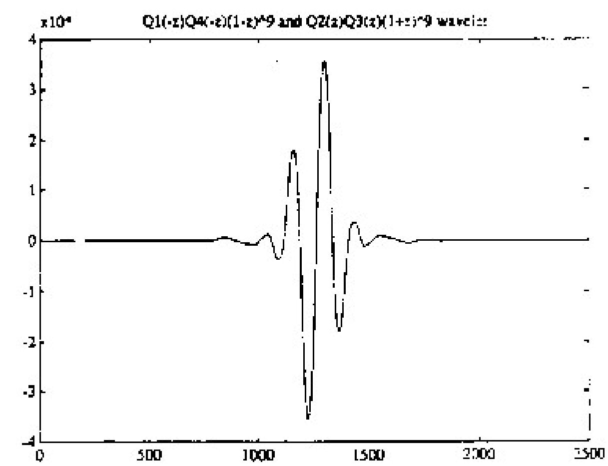
Then $H_1(z)$ is the complementary filter

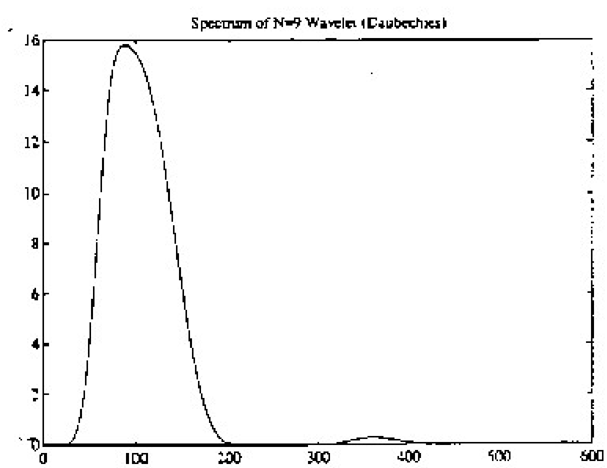
Note $R(z)$ has rational coefficients

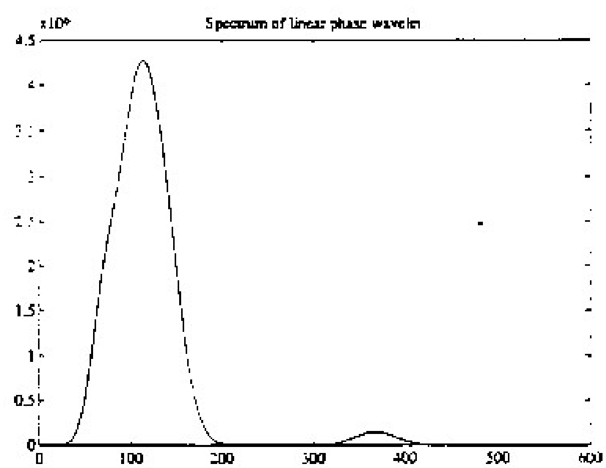

N=9 Scaling function (Daubechies)

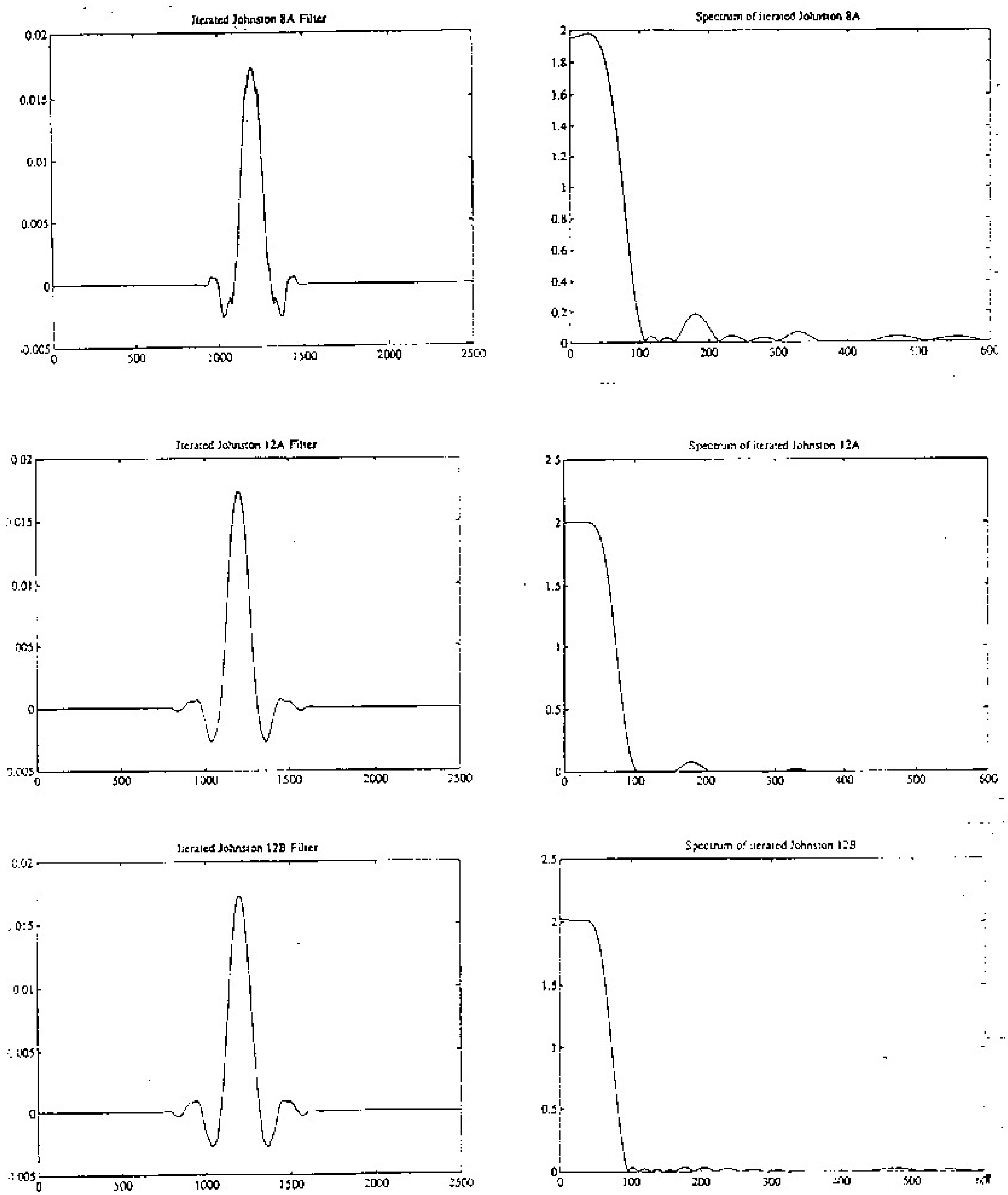

Scaling function generated by Q2(z)Q3(z)(1+z)^9


N=9 Wavelet (Daubechies)


Q1(-z)Q4(-z)(1-z)^9 and Q2(z)Q3(z)(1+z)^9 wavelet


Spectrum of N=9 Wavelet (Daubechies)

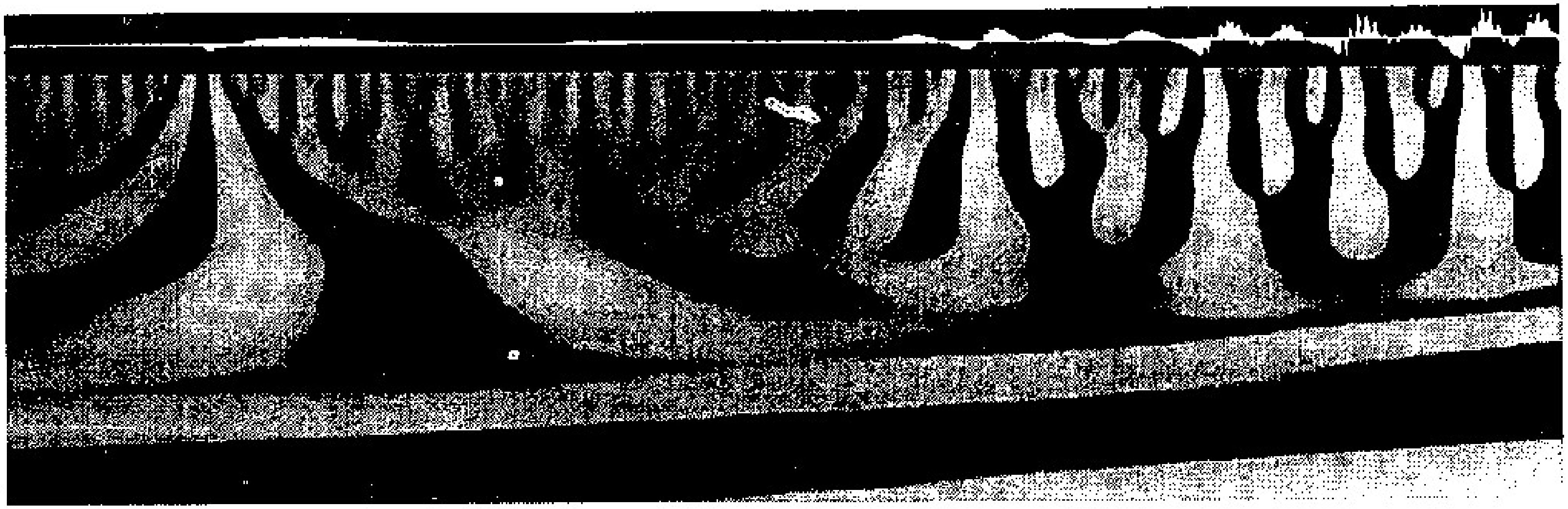

Spectrum of linear phase wavelet



Iterated impulse response

Magnitude of the FT

(a)



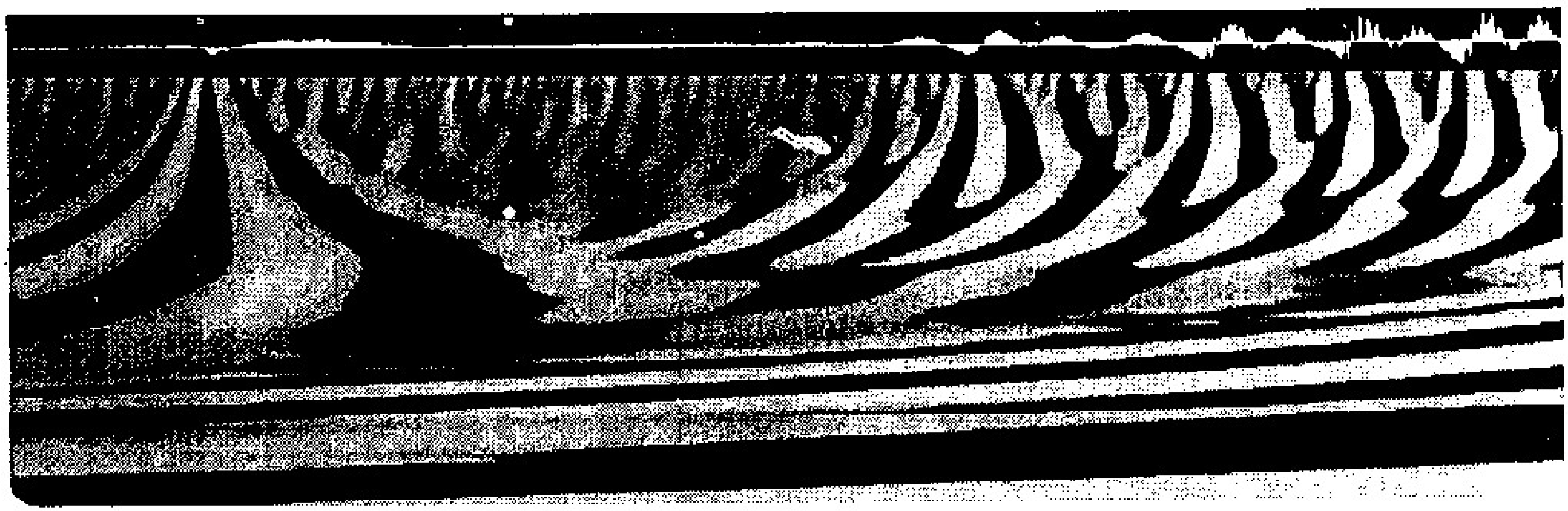
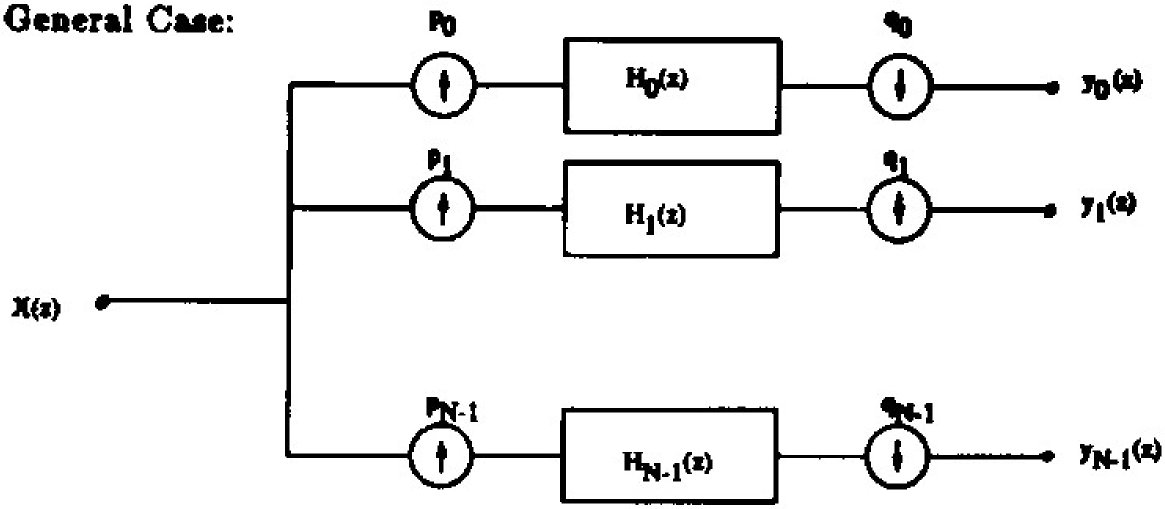


Figure 8: CWP plot of segment of speech showing onset of voicing (a) analysed using linear phase

Filter Banks with Rational Sampling Rates

General Case:



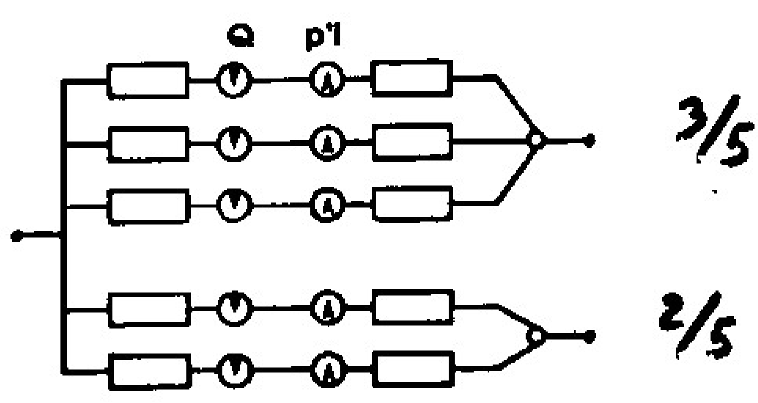
Critical Sampling

$$\sum_{i=0}^{N-1} \frac{p_i}{q_i} = 1$$

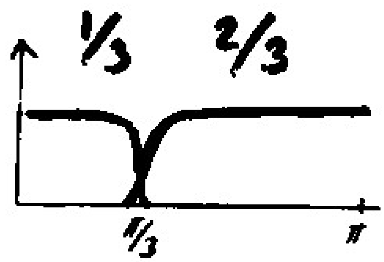
A trivial solution:

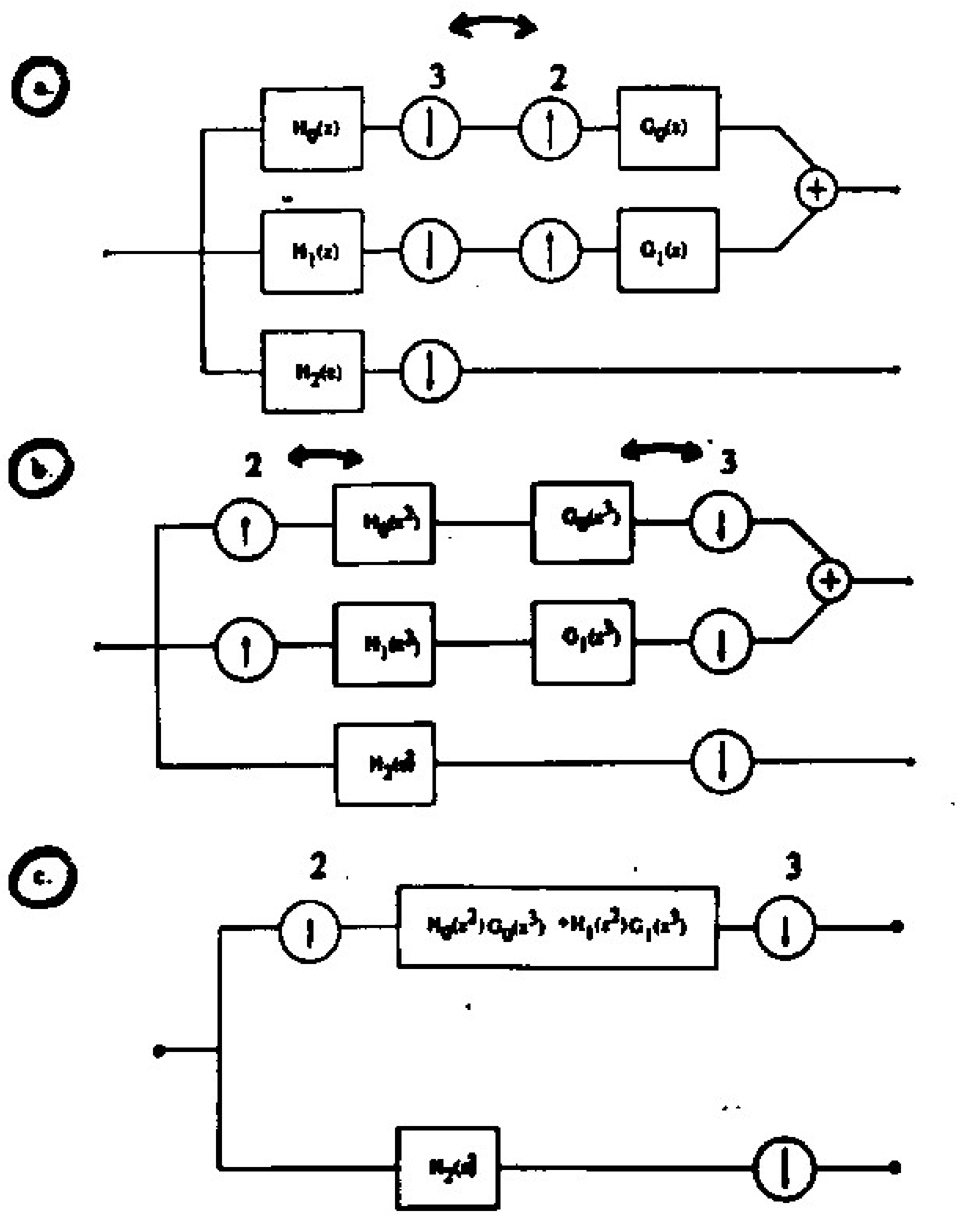
$Q = LCM\{q_i\}$

$p'_i = \frac{Q}{q_i} \cdot p_i$

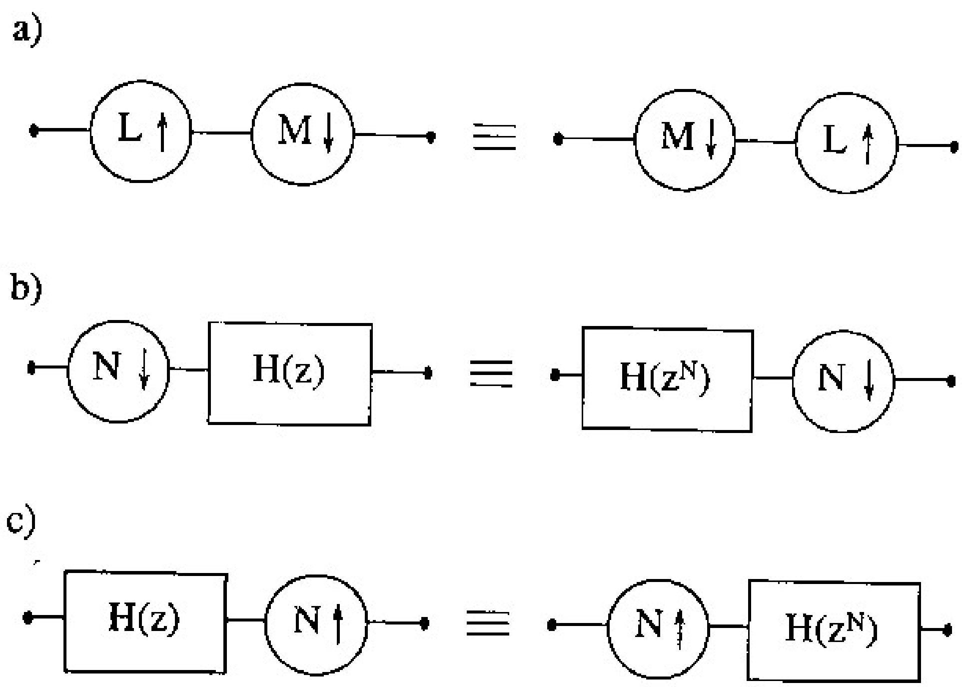


A solution when $(p_i, q_i) = 1$)





Works always in a tree fashion



Some useful identities in multirate signal processing (H(z) is rational)

(a) upsampling by **L** and subsampling by **M** can be interchanged if and only if **L** and **M** are relatively prime

(b) a filter after a subsamping by **N** can be represented in its upsampled version after the upsampling

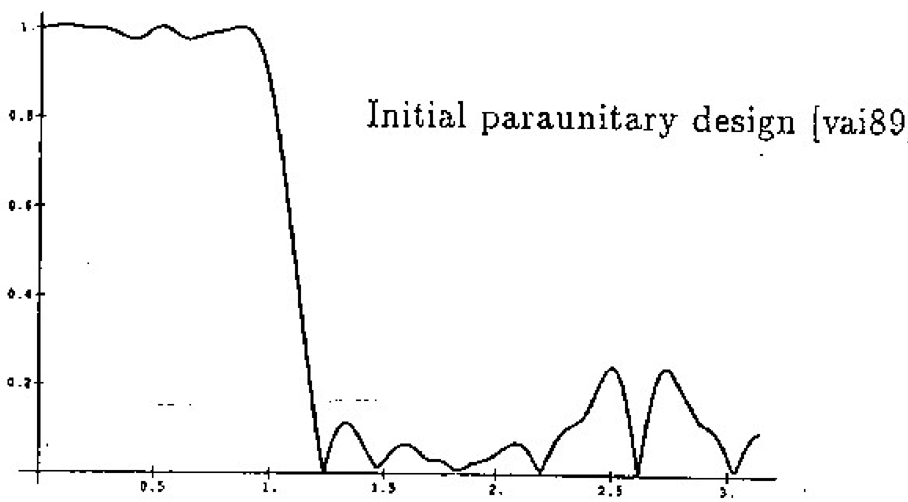
(c) a filter in front of an upsampling by **N** can be represented in its upsampled version after the upsampling

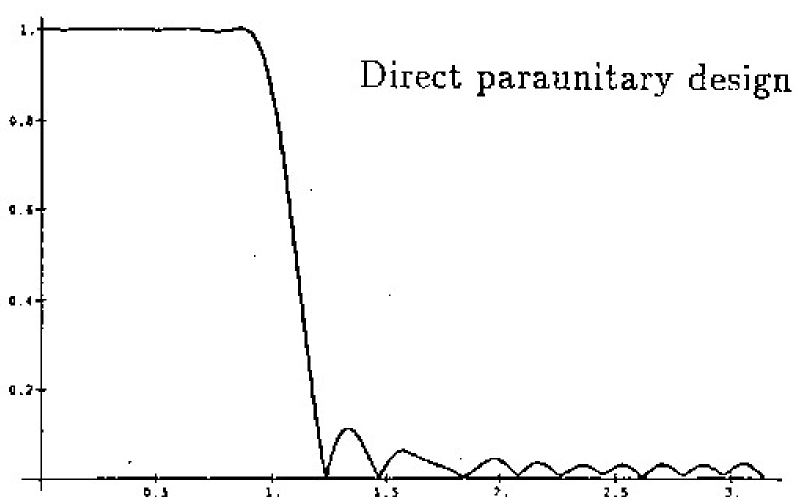
Example of unequal bandwidth filter bank design

2 channel bank: 4 lattice coefficients, 8 taps

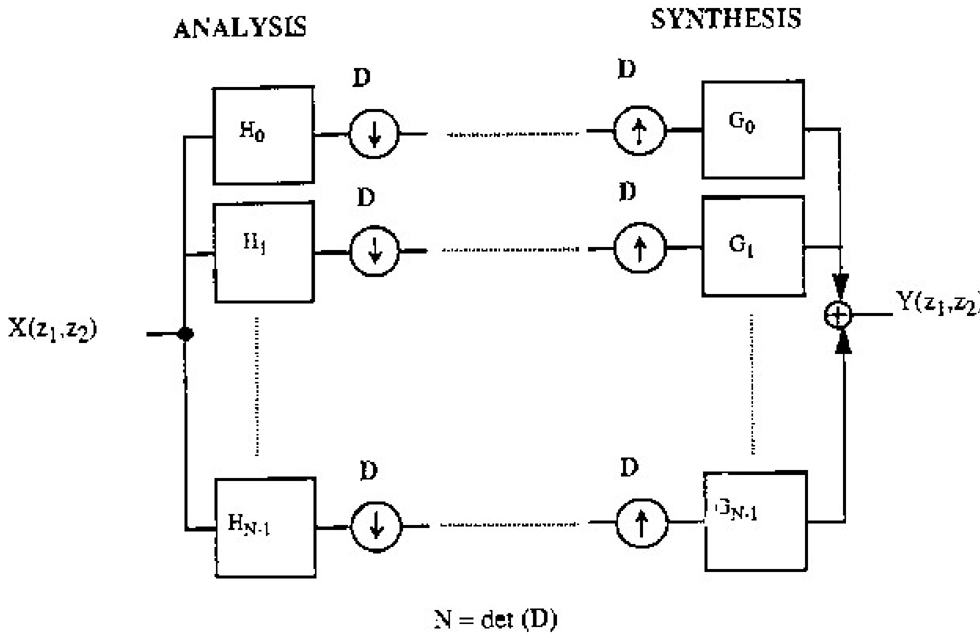
3 channel bank: 10 lattice coefficients, 15 taps

Resulting 1/3, 2/3 bank: 50 taps





Multidimensional perfect reconstruction filter bank



- Arbitrary sampling lattices **D**
- Results on aliasing cancellation
- Results on perfect reconstruction
- Paraunitary and linear phase solutions
- Non-separable filters allow more freedom

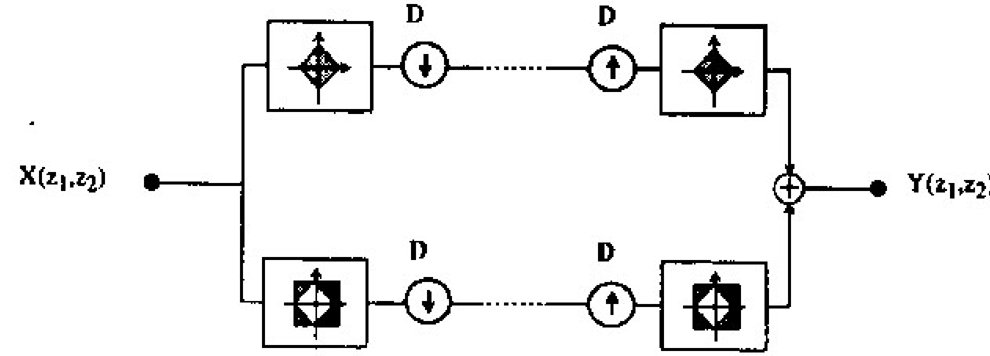
  Ex: 4 channel lin. phase and paraunitary

- Factorization very hard...

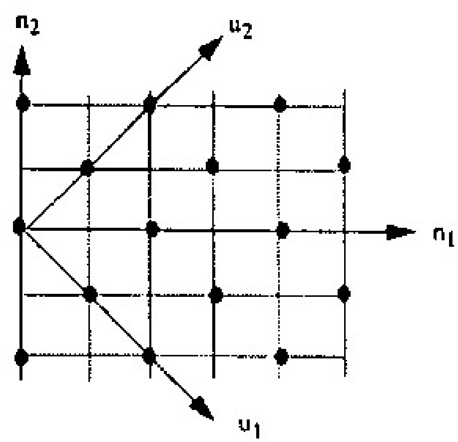
QUINCUNX CASE

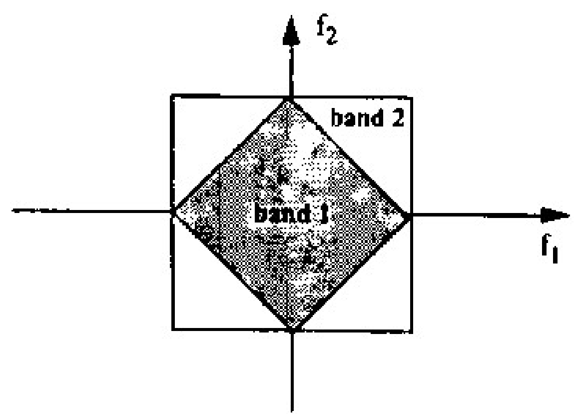
- **We split the signal in two channels:**

$$D = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$



- **The sampling structure and the desired frequency splitting:**

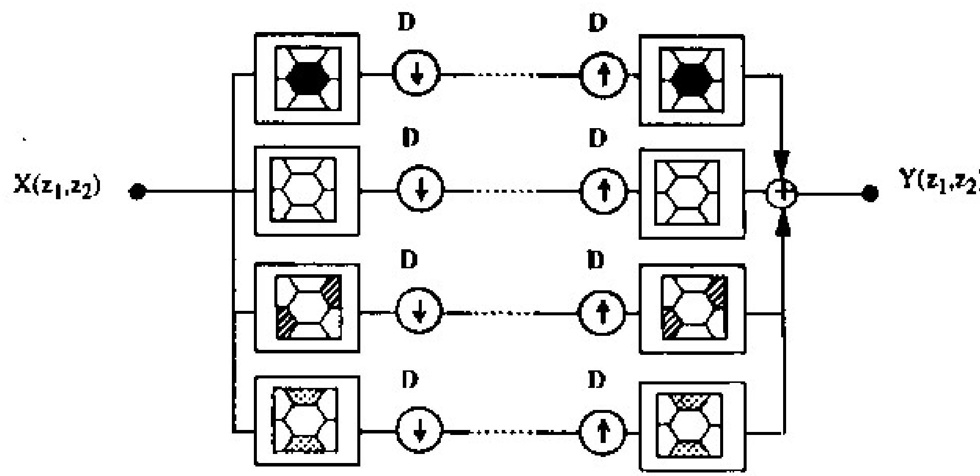




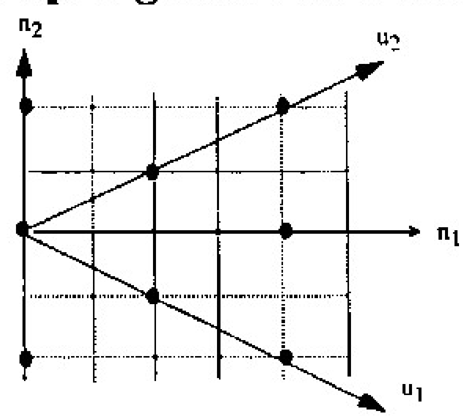
HEXAGONAL CASE

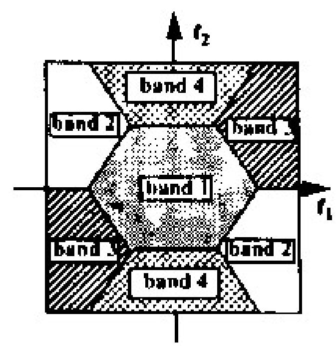
- **We split the signal in four channels:**

$$D = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}$$

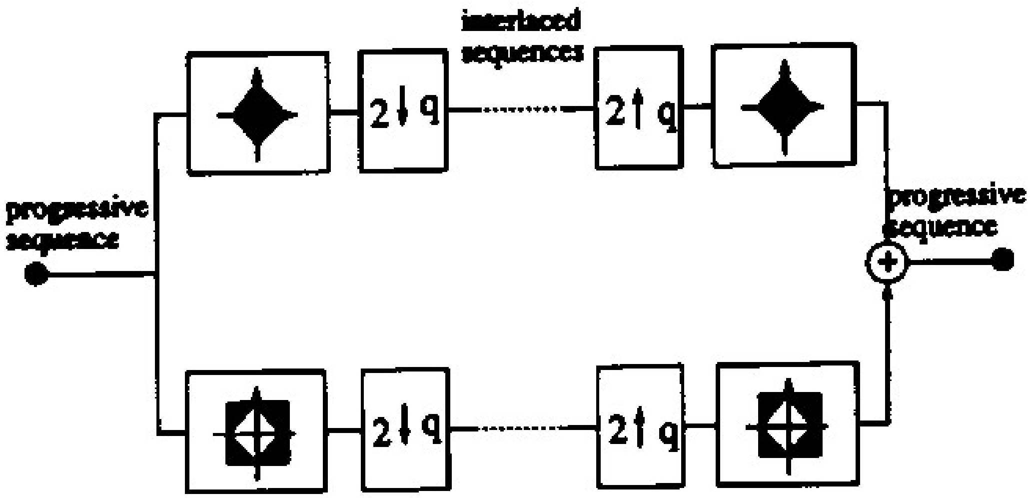


- **The sampling structure and the desired frequency splitting:**

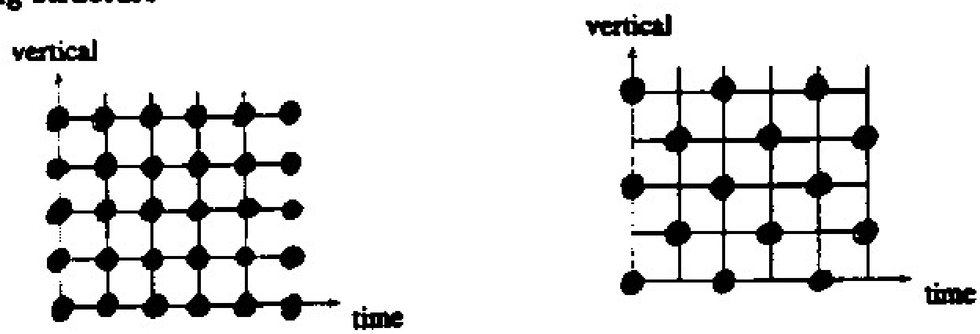




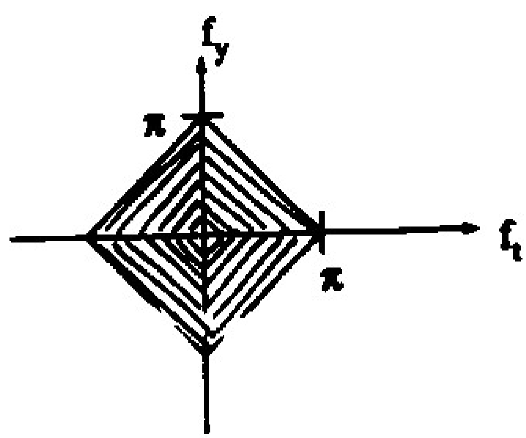
PROGRESSIVE TO INTERLACED AND ... BACK!



Sampling structure



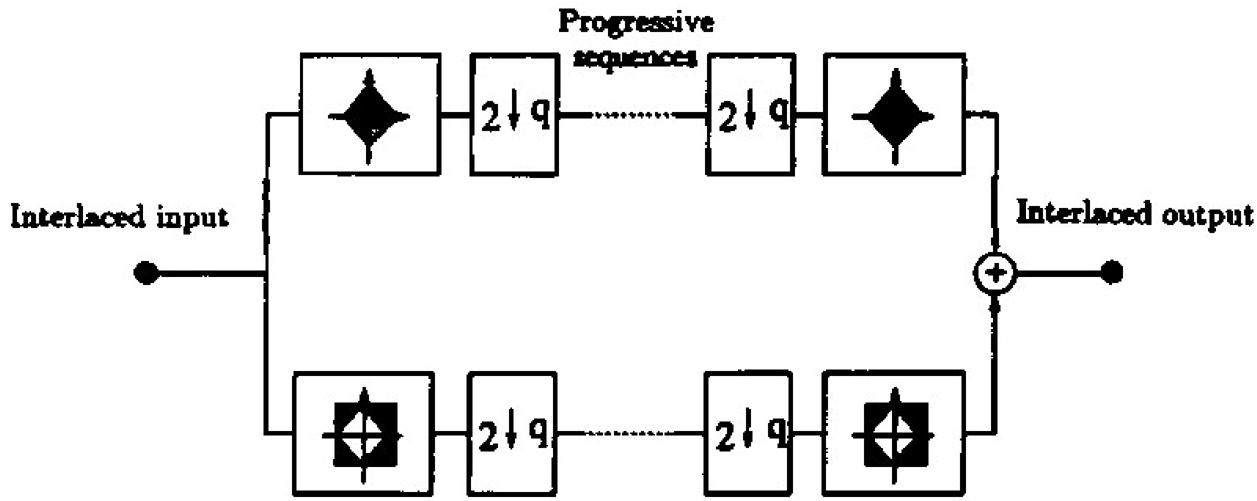
Filter characteristic



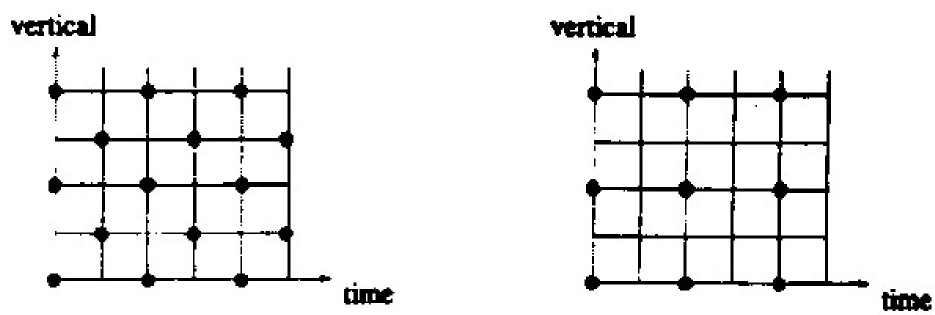
Useful for:

compatible lowpass channel

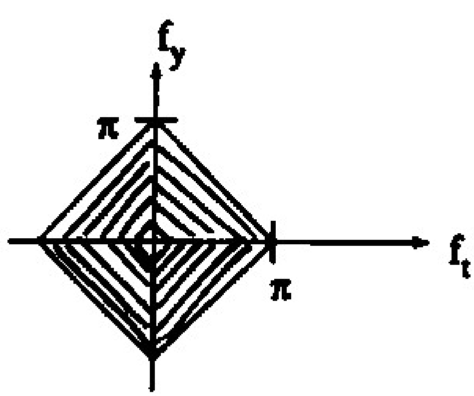
INTERLACED TO PROGRESSIVE AND ... BACK!



Sampling structure



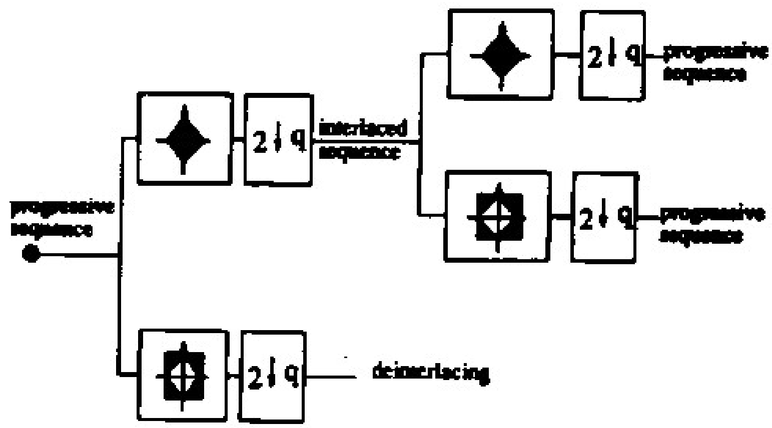
Filter characteristic



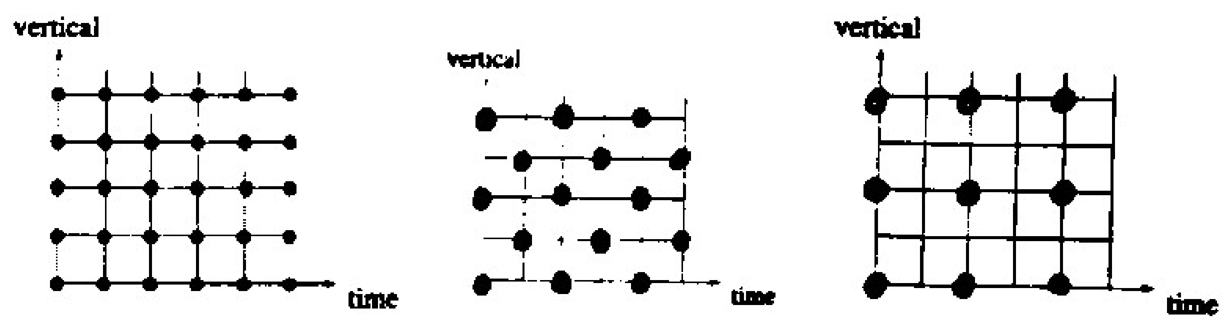
Useful for:

derive progressive channels for motion processing

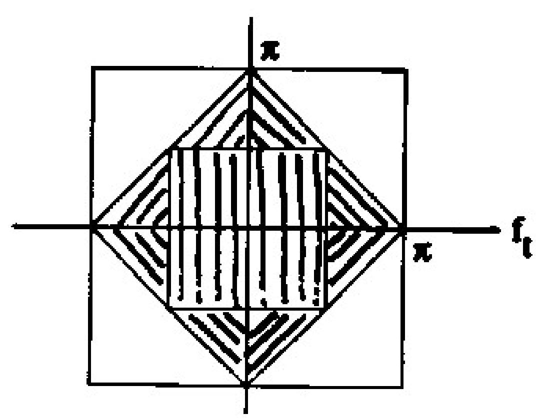
PROGRESSIVE TO INTERLACED TO PROGRESSIVE!



Sampling structure



Filter characteristic



Useful for:

compatibility and coding

Different size:

use the following elementary filters:

$$\begin{pmatrix} & 1 & \\ b & a & b \\ & 1 & \end{pmatrix}$$

$$\begin{pmatrix} & & 1 & & \\ & b+c/a & a & b+c/a & \\ bc/a & c & d & c & bc/a \\ & b+c/a & a & b+c/a & \\ & & 1 & & \end{pmatrix}$$

Cascading of the corresponding polyphase matrices creates linear phase filters with perfect reconstruction

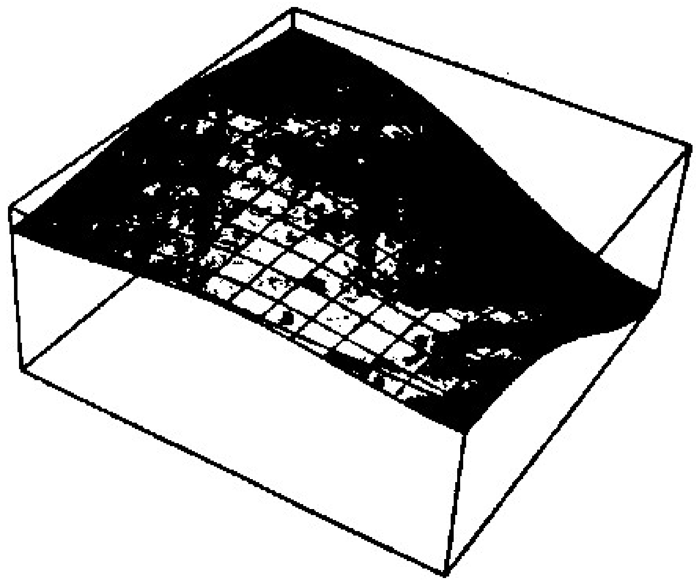
Note the top-bottom and left-right symmetry

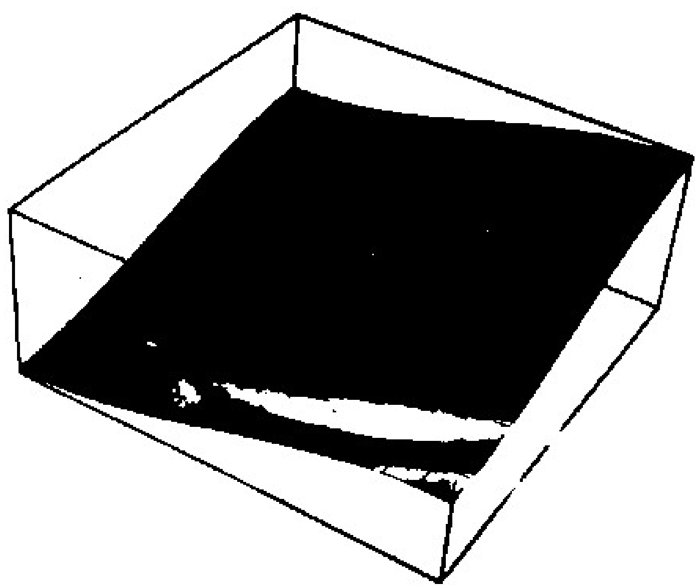
Additional rotational symmetry: $b = 1$ and $c = a$

Smallest example:

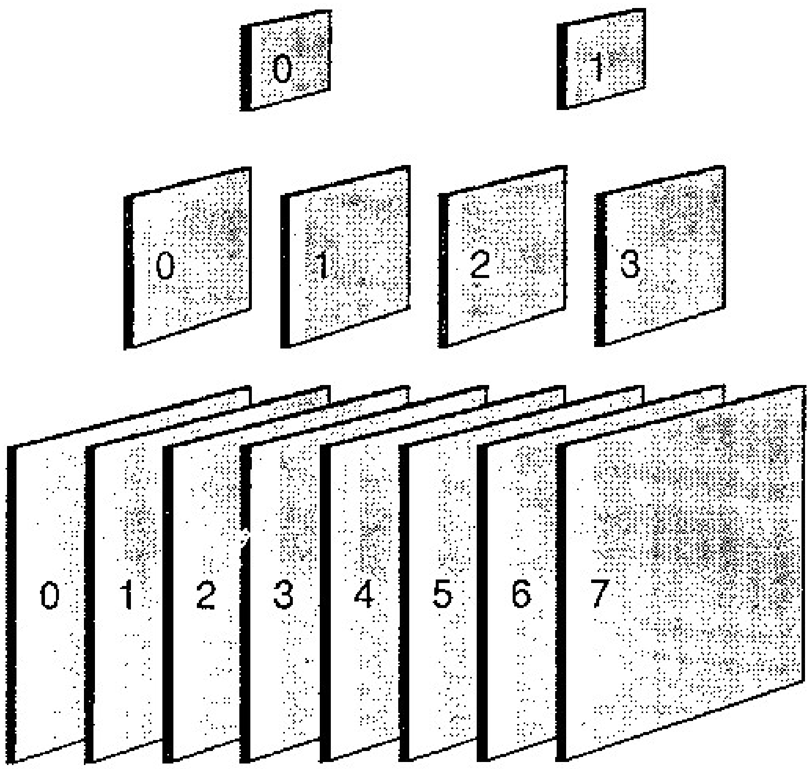
$$\begin{pmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{pmatrix} \quad \begin{pmatrix} & & 1 & & \\ & 2 & -4 & 2 & \\ 1 & -4 & -28 & -4 & 1 \\ & 2 & -4 & 2 & \\ & & 1 & & \end{pmatrix}$$

Magnitude of the frequency response:



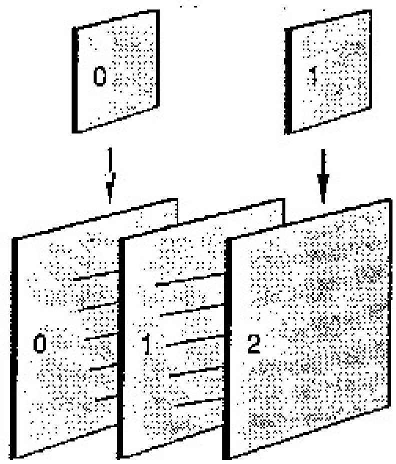


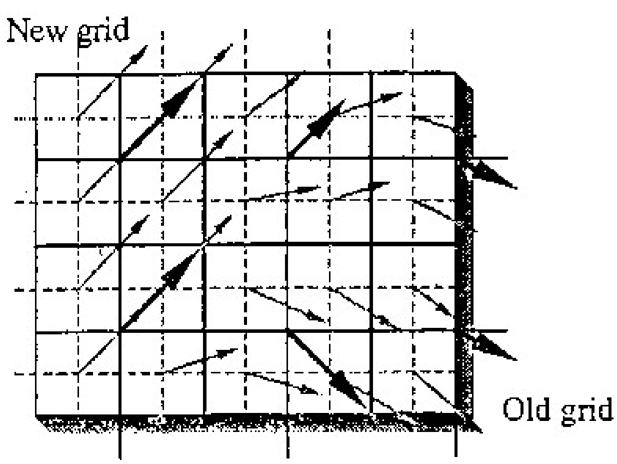
A Multiresolution Approach to Coding of Digital HDTV



The spatio-temporal pyramid structure. Each level is obtained from a lower one by spatial anti-aliasing filtering, followed by spatial and temporal decimation.

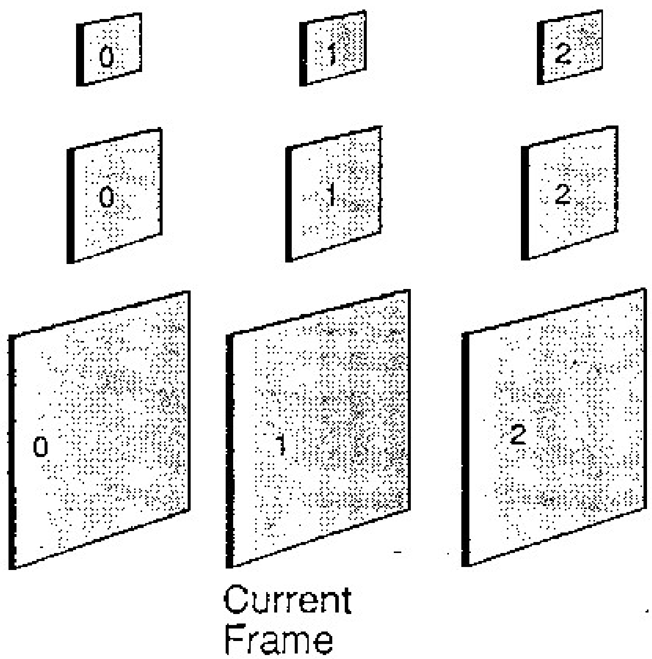
Interpolation as a stepwise refinement. First spatial resolution is increased, temporal interpolation follows yielding the next level in the pyramid.





The motion field is resampled in the new finer grid, giving an initial estimate for the search.

Hierarchical Motion Estimation

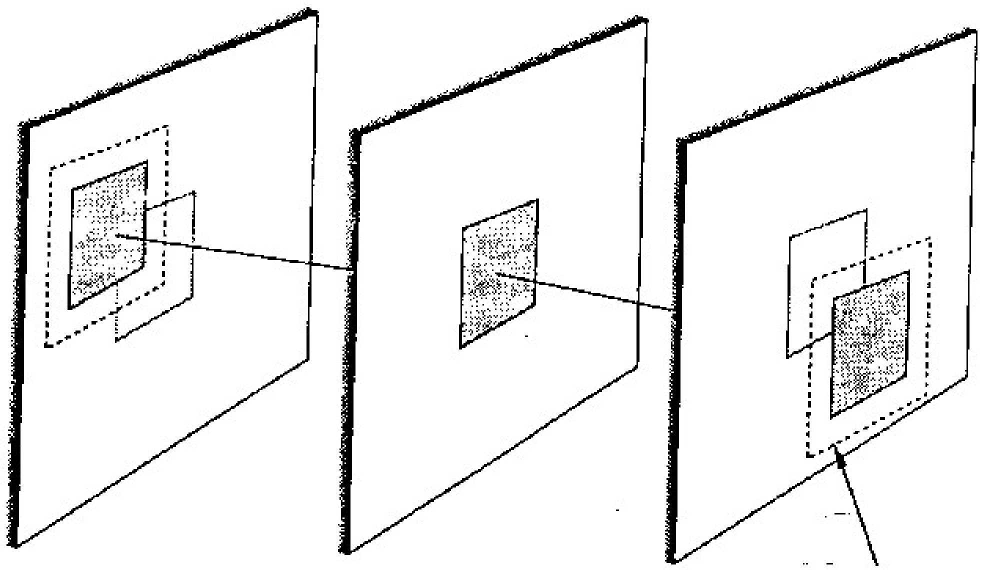


Previous and next frames at varying resolutions are used for the motion estimation. The search starts at the top level, i.e. the coarsest resolution. Results from the coarser resolution are used as initial estimates.

Hierarchical Motion Estimation

We use a hierarchical blockmatching technique modified for motion based interpolation.

1. Both previous and next frames are used in the estimation.
2. A constrained window full-search is performed at the coarsest resolution.
3. At finer levels, a constrained search is performed within a window offset by the initial coarse estimates.
4. At the finest resolution, interpolation mode is also selected.
   - Backward, forward, or averaged mode is decided.
   - Interpolation mode is coded along with motion vector.



The symmetric search is performed in a limited region around the initial estimates obtained from the prior coarser resolution search. Displacement resulting in the "best" interpolated block is selected.

Entropy Coding

A DCT-based coder is used. The top layer of the pyramid is encoded intra-frame. All remaining layers are formed by interpolation errors with varying degrees of perceptual significance.

The final layer contains over 75% of the samples. Therefore lower resolution subchannels can be coded at relatively high rates.

Temporal and spatial interpolation errors are also of perceptually different nature. Temporal errors can be tollerated without apparent loss of visual quality, and thus can be coded at lower bitrates.

Bit allocation to various layers (in bits per pixel, overall denotes contribution to total bitrate):

Layer	Spatial	Temporal	Overall
1	3.0	N/A	0.05
2	2.1	0.55	0.16
3	2.2	0.45	1.33

- Finite memory system is robust to channel errors.
- Motion based interpolation superior to recursive prediction
  - Ability to use future solves occlusion problems.
  - Irregular motion, including scene cuts, can be handled.
  - Even large errors are highly localized, hence invisible.
- 3-D pyramid provides adaptive bit allocation for temporal and spatial frequencies.
- High quality compatible subchannels are provided at no extra cost.

Conclusions and Directions

Multiresolution signal representation:

- successive approximation
- compatibility
- coding

Hierarchical methods for complexity and robustness

- hierarchical or multiresolution motion estimation

Framework:

- wavelet signal analysis
- orthogonal or biorthogonal bases
- multiresolution pyramids

Non-linear processing:

- need for robustness
- pyramids are more flexible

Broad range of problems and applications can be seen in a common framework!