Wavelet Video Coding –
Principles, Applications and Standardization

Mihaela van der Schaar

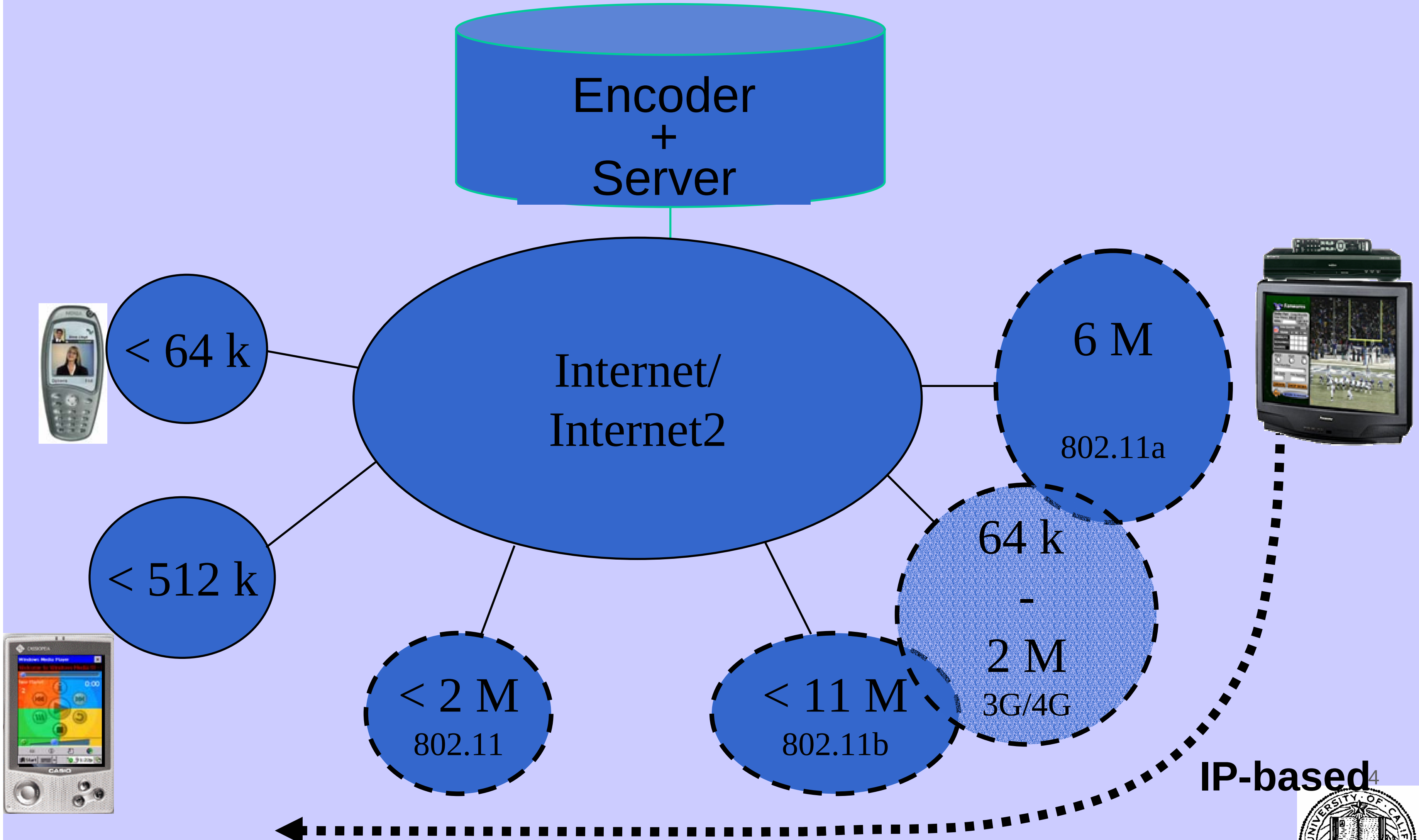
Electrical and Computer Engineering Department
University of California Davis

Outline

- Introduction
- Scalable coding – principles (review)
- Basic principles of wavelets (review)
- Motion Compensated Wavelet Coding – basic principles and classification
- Motion Compensation Temporal Filtering (MCTF)
- Overcomplete Motion Compensated Wavelet Coding
- Encoding of spatio-temporal wavelet coefficients
- Scalable coding of motion information
- Error resilience aspects
- Current status in MPEG standardization
- Comparisons with state-of-the-art non-scalable coding techniques

Introduction

Challenges for ubiquitous multimedia communication



Sample of concrete problems/questions

- **Signal processing**
  - compression efficiency versus quality of signal reproduction (rate-distortion tradeoffs)
  - compression efficiency versus robustness to losses

- **Networking**
  - realistic channel models for effective joint source/channel coding
  - source-channel interface control strategies for efficient network resource usage and high quality signal reproduction

- **Computer Architecture**
  - compression efficiency versus computational complexity

Possible solution: compression meets the network

- Do not require the transport mechanism to be flawless (modulation, channel coding, transmission protocol etc.), just design the coding system and transmission *jointly*

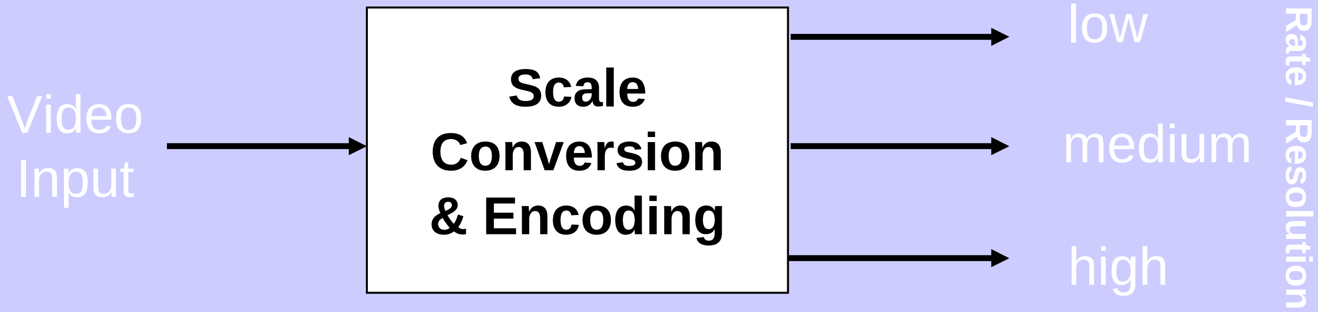
- Do not design for worst-case scenario - just adapt on the fly based on the network and device characteristics

Hence:

- **A. Scalable Coding**
- **B. Adaptive Streaming**

Principles of Scalable Coding

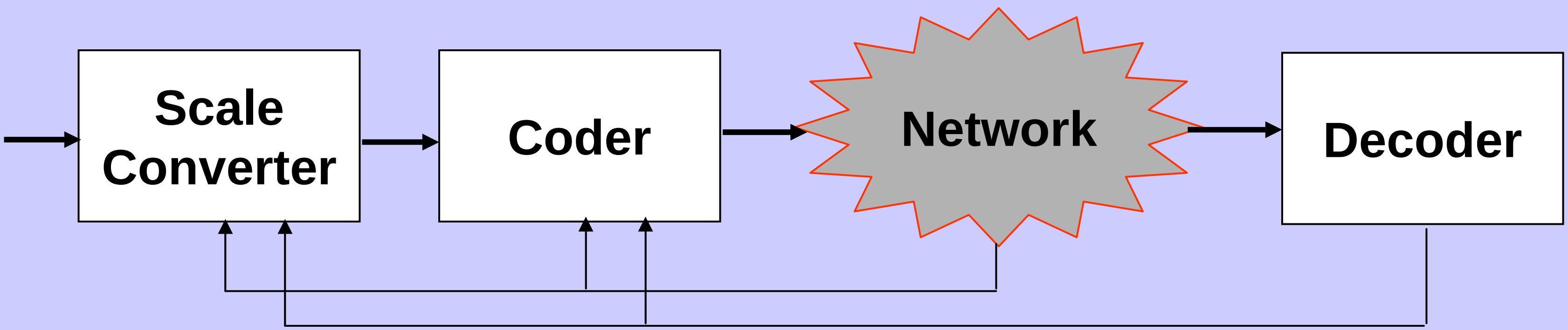
- Encoding of video signal with different resolution scales



- Downscaling of video signal by
  - Coding noise insertion – **SNR Scalability**
  - Spatial subsampling – **Spatial Scalability**
  - Sharpness reduction – **Frequency Scalability**
  - Temporal subsampling – **Temporal Scalability**
  - Selection of content – **Content related Scalability**

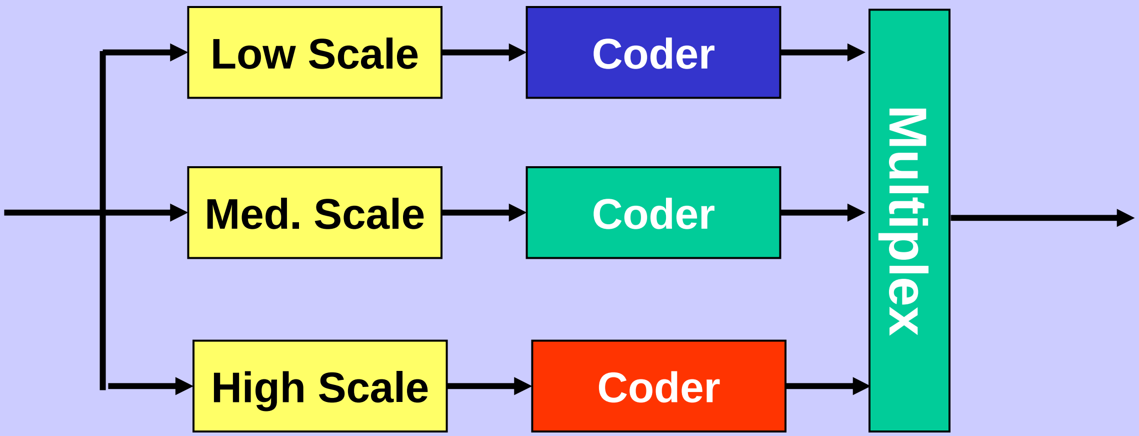
The Simple Way – Advance Scaling

- Requires feedback about channel / decoder status
- Only point-to-point connection supported
- Example : Stream switching



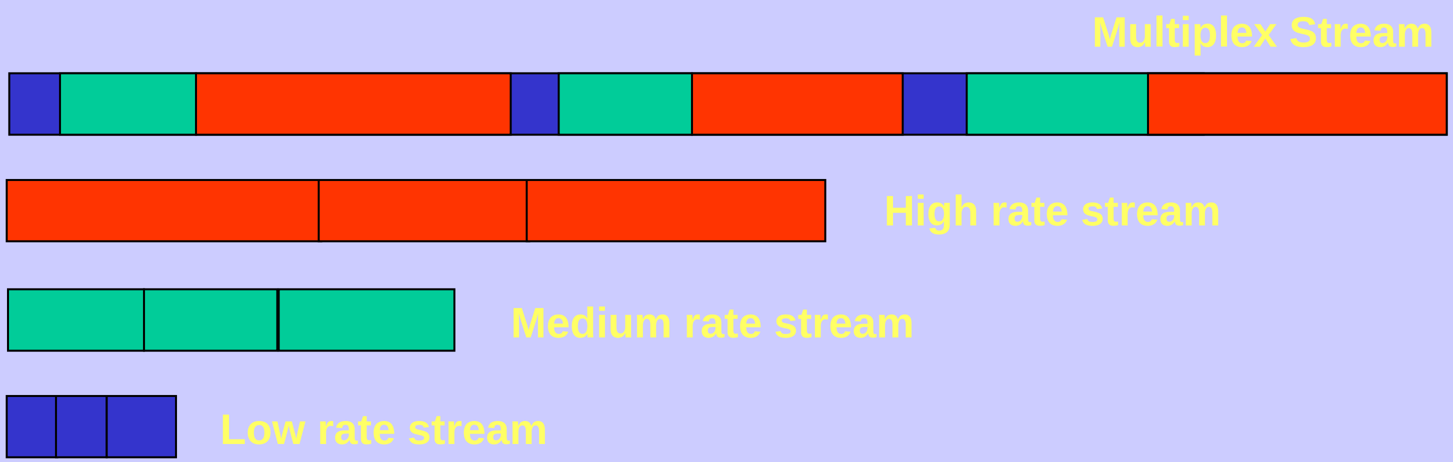
The Parallel Way - Simulcast

- Run independent encoders in parallel
- Requires a priori knowledge about network and decoder capabilities to select optimum scaling levels
- Point-to-multipoint connections possible



Simulcast

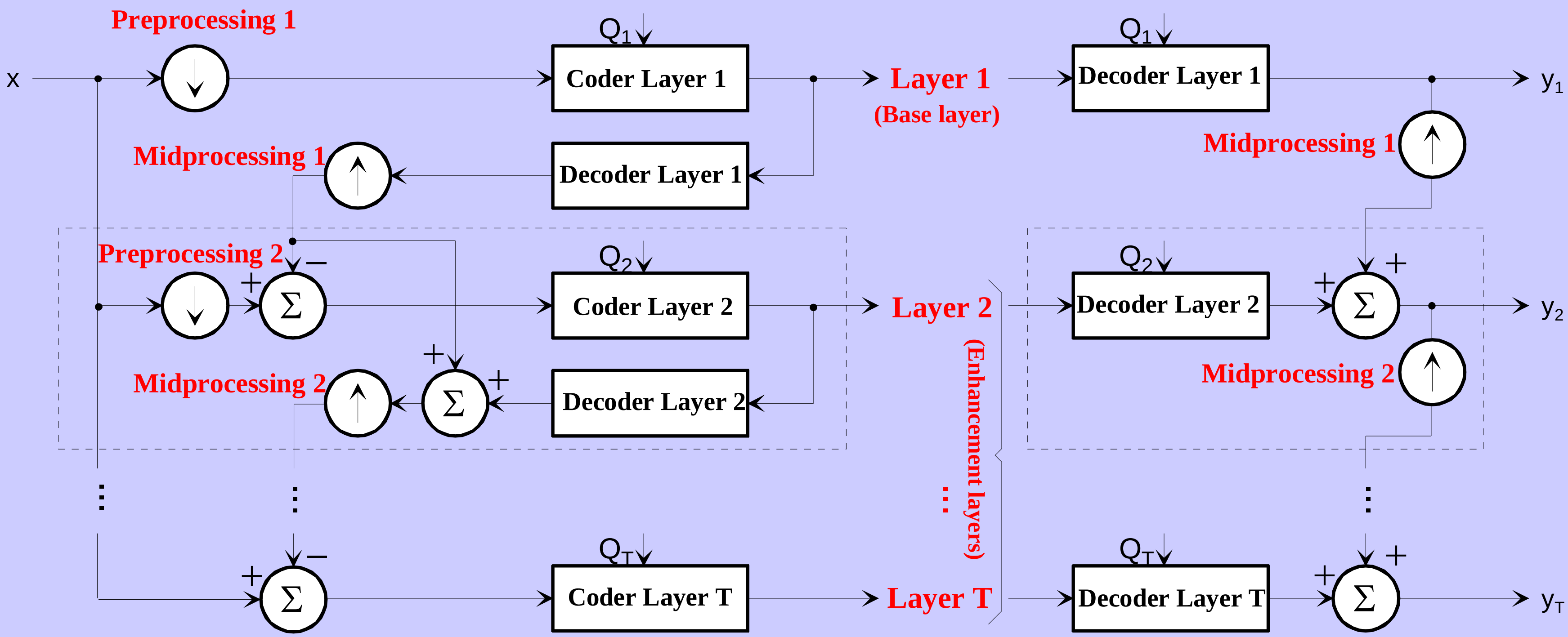
- Multiplexed transmission of streams



- Loss in efficiency due to multiple streams
  - Can cause network overload
  - Restricted number of scales

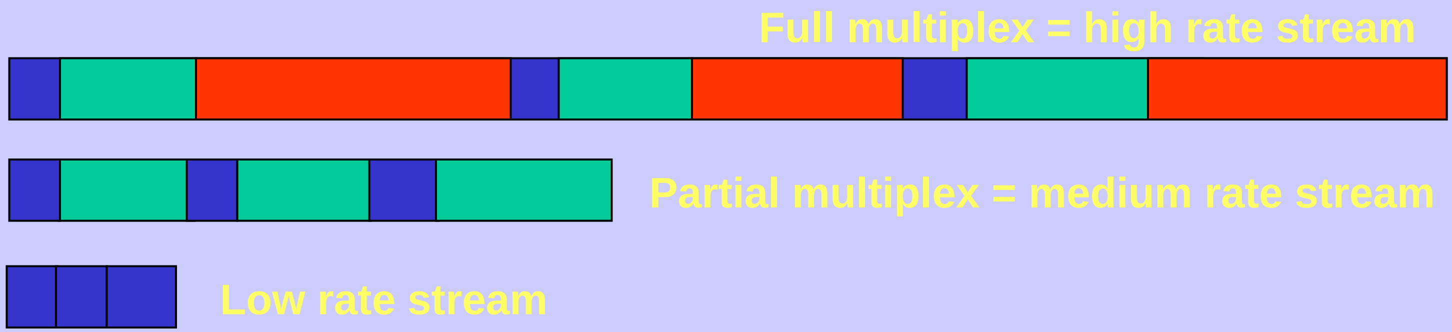
The Embedded Way – Layered Coding

- "Chain of layers" - information from low resolution utilized to encode next-higher resolution



Layered Coding

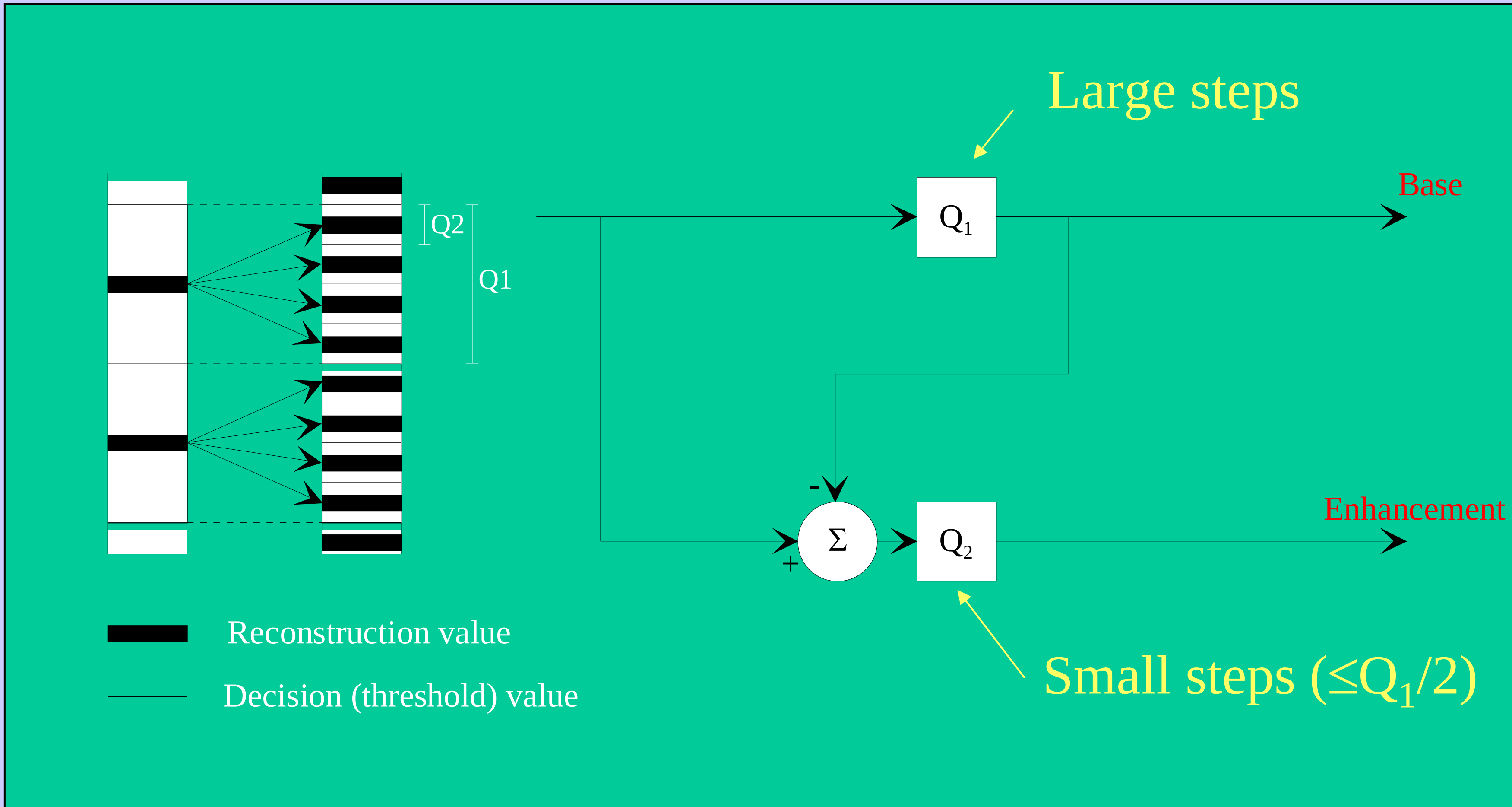
- Layered coding supports embedded streams
  - Re-configuration of bit stream for reconstruction with different spatial/temporal/quality resolution



- Possible loss in efficiency depends on coding scheme
- In theory, arbitrary number of scales could be achieved

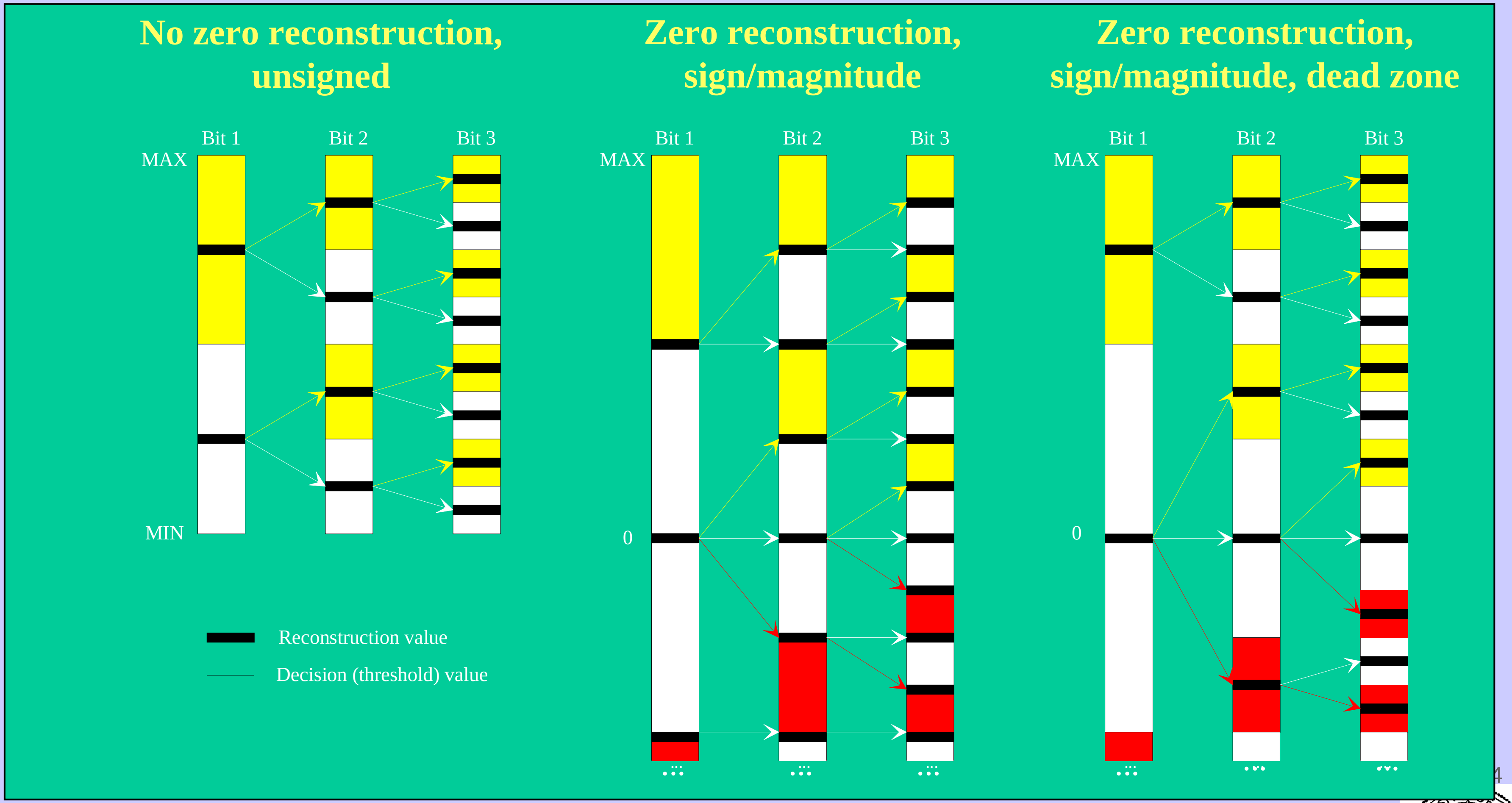
SNR Scalability – Re-quantisation

- Example : 2-stage quantizer



SNR Scalability – Bit-plane Coding

- Quantization related to bit planes



SNR Scalability – Bit-plane Coding

- Magnitude of MSB encoded by run-length or binary entropy coding
- Sign and remaining bits encoded binary, conditional on MSB

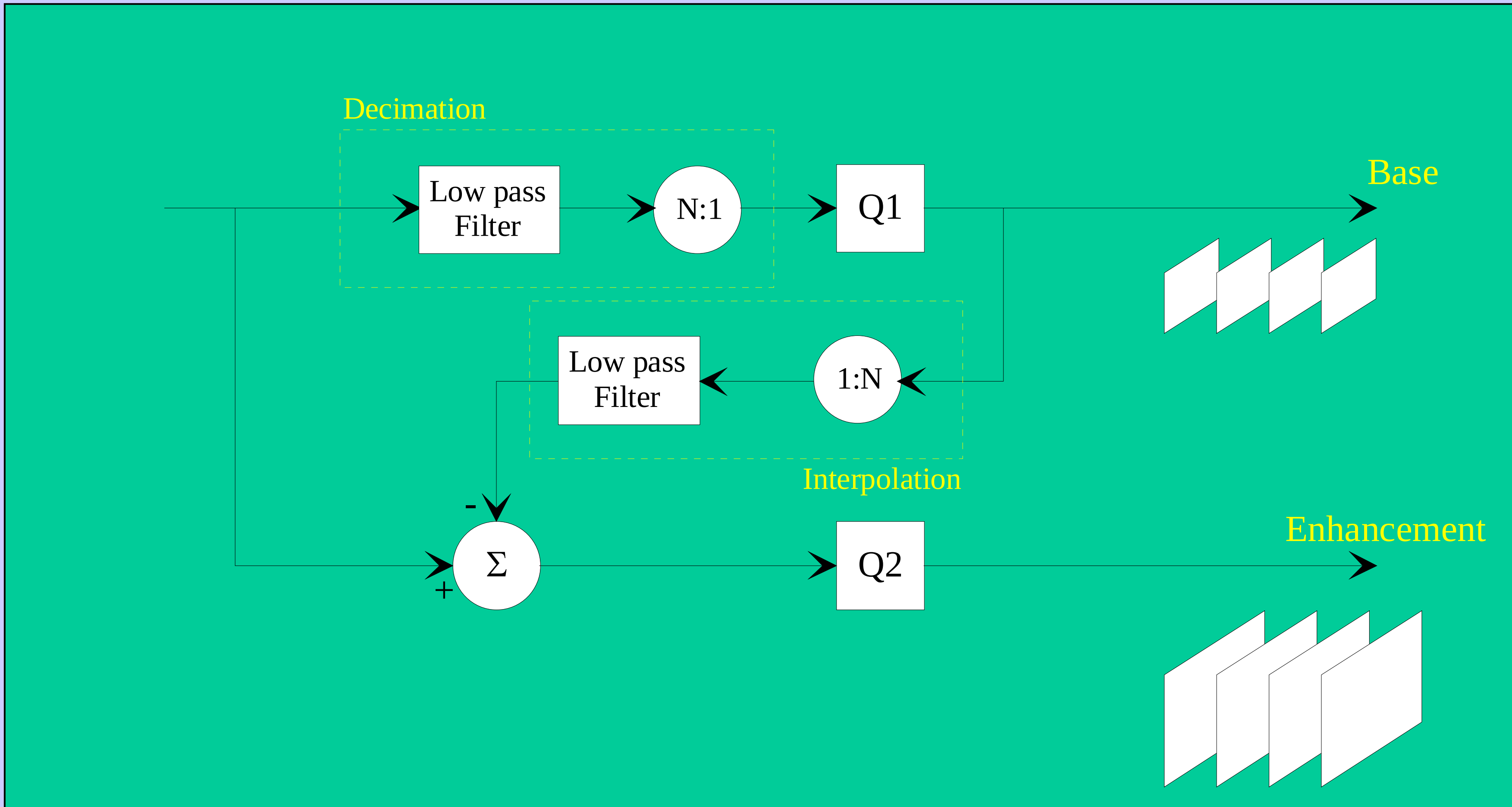
Sample

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Run-length code
Bit 1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	4,9
Bit 2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2,10
Bit 3	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	3,5,2
Bit 4	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	3,4,1
Bit 5	1	0	1	1	1	0	0	0	1	1	0	1	0	0	1	0,1,1,0,2
Sign	1	0	0	0	1	0	1	0	1	0	0	1	0	0	1	

binary coded

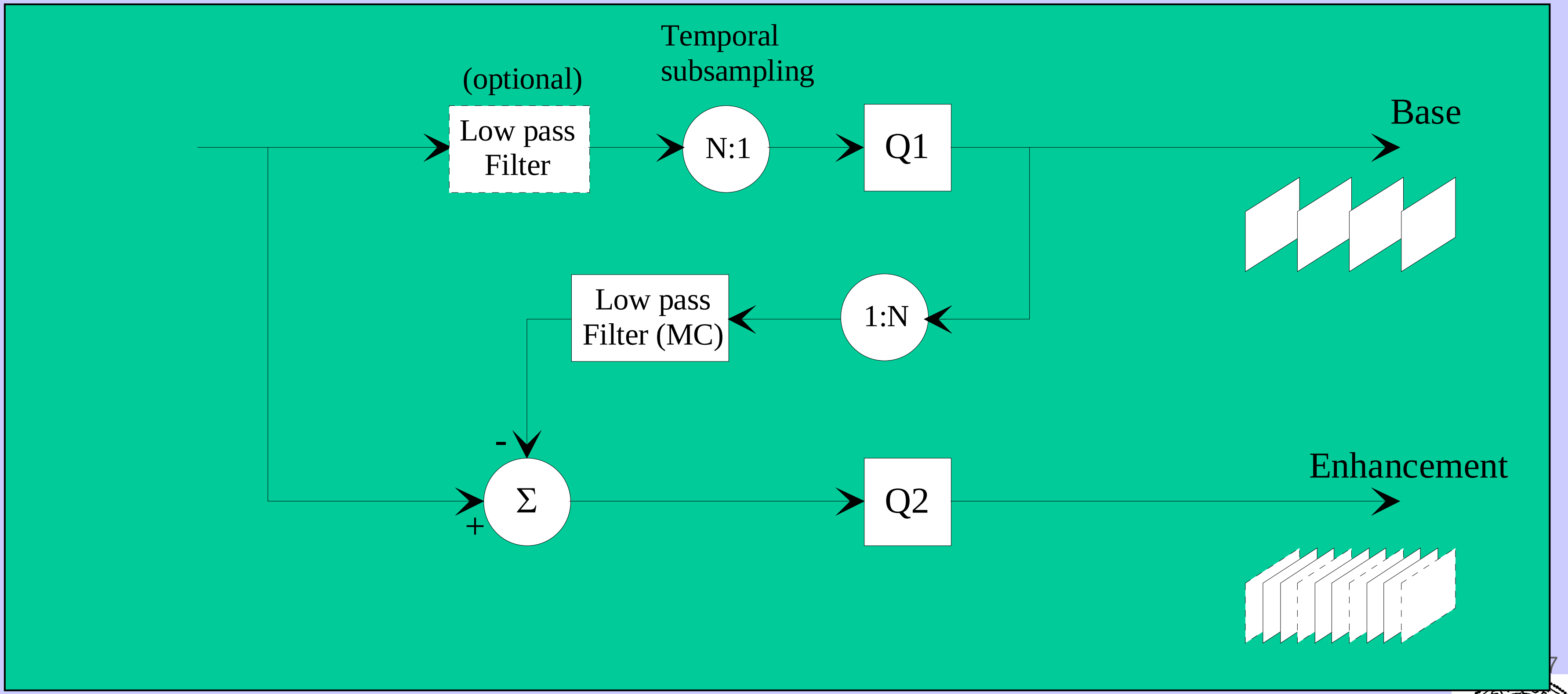
Spatial Scalability

- Base-to enhancement prediction



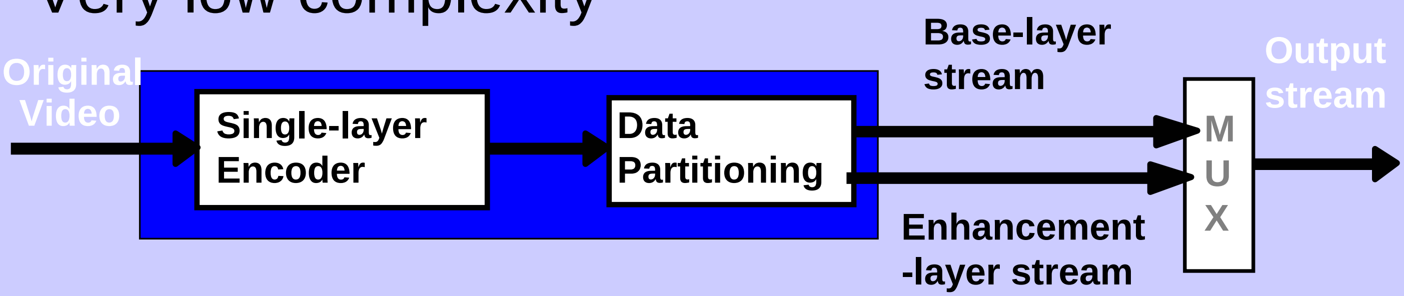
Temporal Scalability

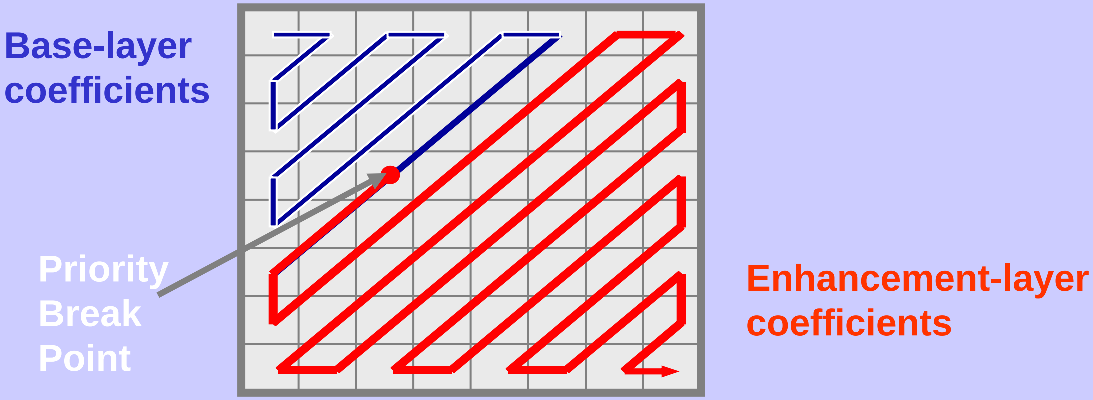
- Temporal downsampling with temporal anti-alias filter or by frame skipping
- Temporal upsampling by MC prediction



Frequency Scalability / "Data Partitioning"

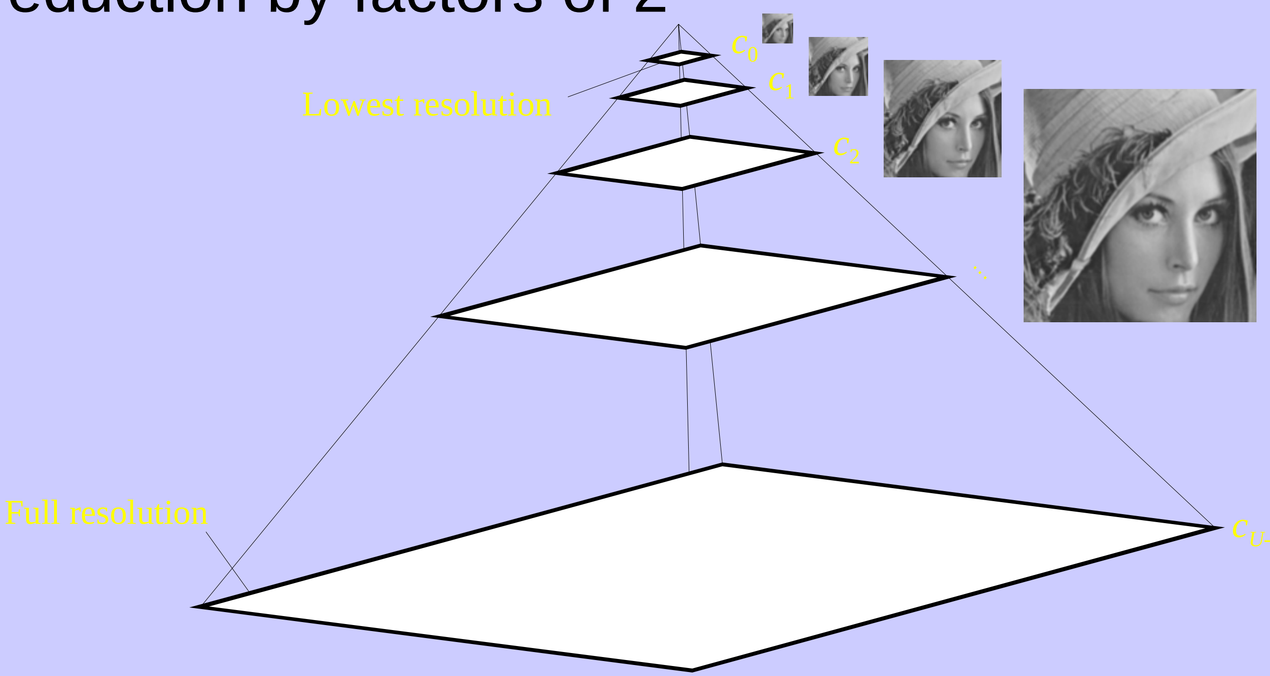
- Popular in context of Transform Coding
- Allocation of coefficients to different layers depending on frequency
- Very low complexity





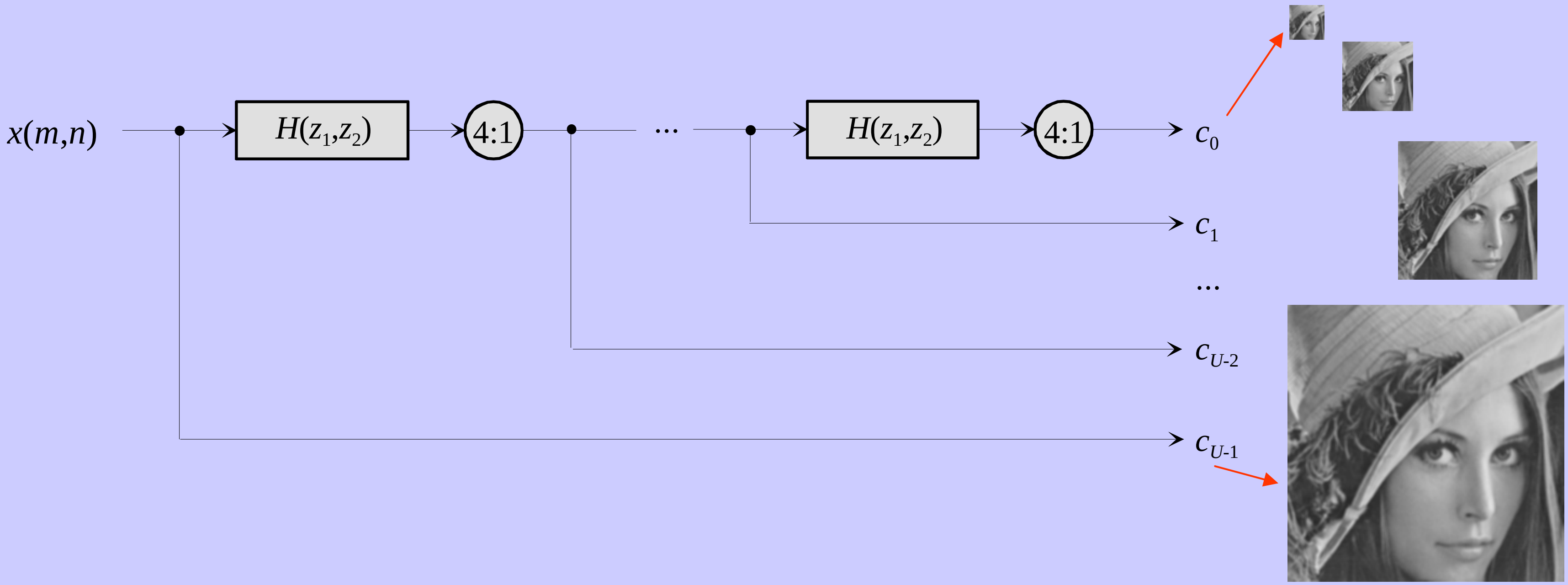
Multiresolution Concepts

- Generate different resolution levels by successive down/upsampling operations
- Resolution pyramids example : Spatial resolution reduction by factors of 2



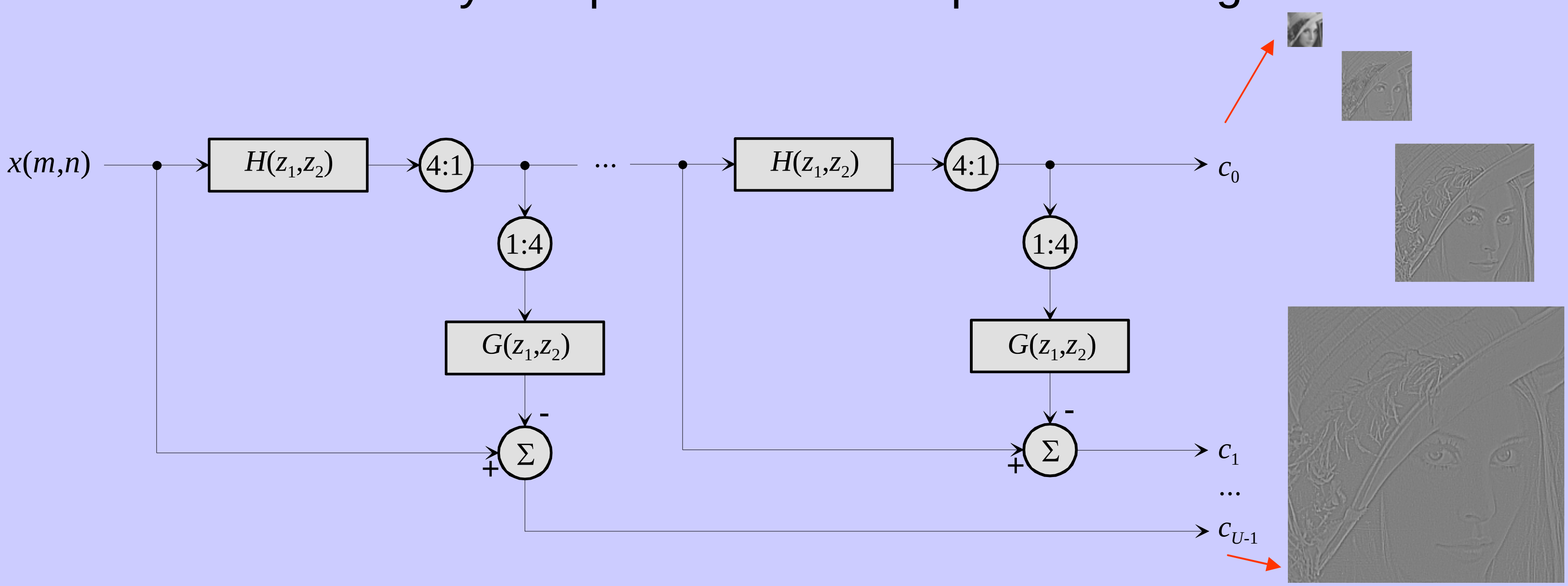
Multiresolution Concepts – Pyramids

- **Gaussian Pyramid**
  - Each layer is self-contained
  - Corresponds to Simulcast concept
  - More samples to be encoded



Multiresolution Concepts – Pyramids

- **Laplacian Pyramid** (Differential Pyramid)
  - All lower-resolution layers required to reconstruct high-resolution layers
  - Corresponds to Layered Coding concept
  - Not critically sampled – more samples than original



Multiresolution Concepts – Pyramids

- **Advantages :**
  - Pyramids can be combined with any coding scheme for the different resolution levels
  - Downsampling can be made alias-free

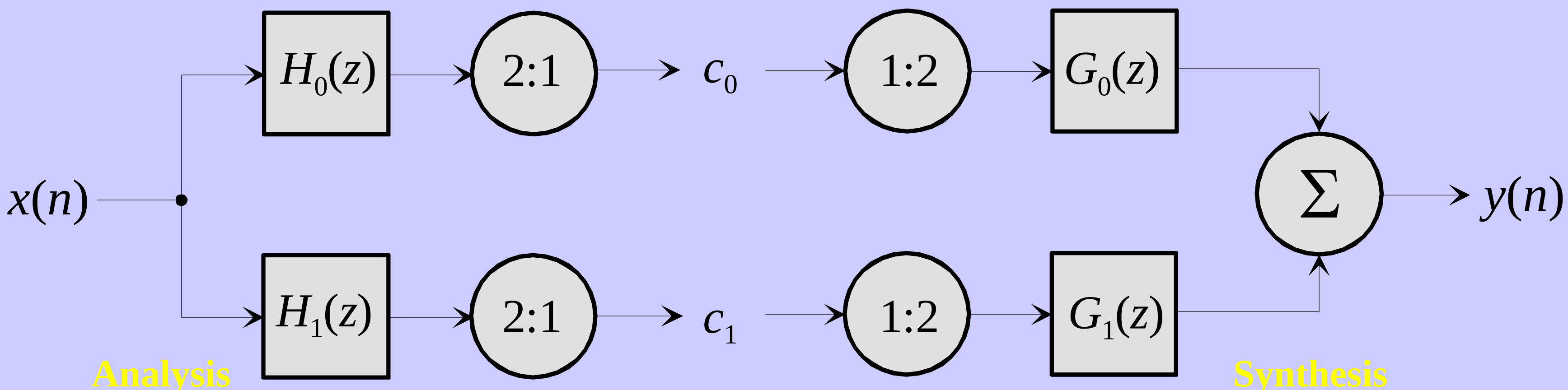
- **Disadvantages :**
  - Number of pixels higher than in original signal
  - Higher data rate than one-layer coding

- **Possible solution :**
  - Critically sampled pyramids (Wavelets)
  - Disadvantage : Downsampled signals bear alias

Basic Principles of Wavelets

Filter Pairs

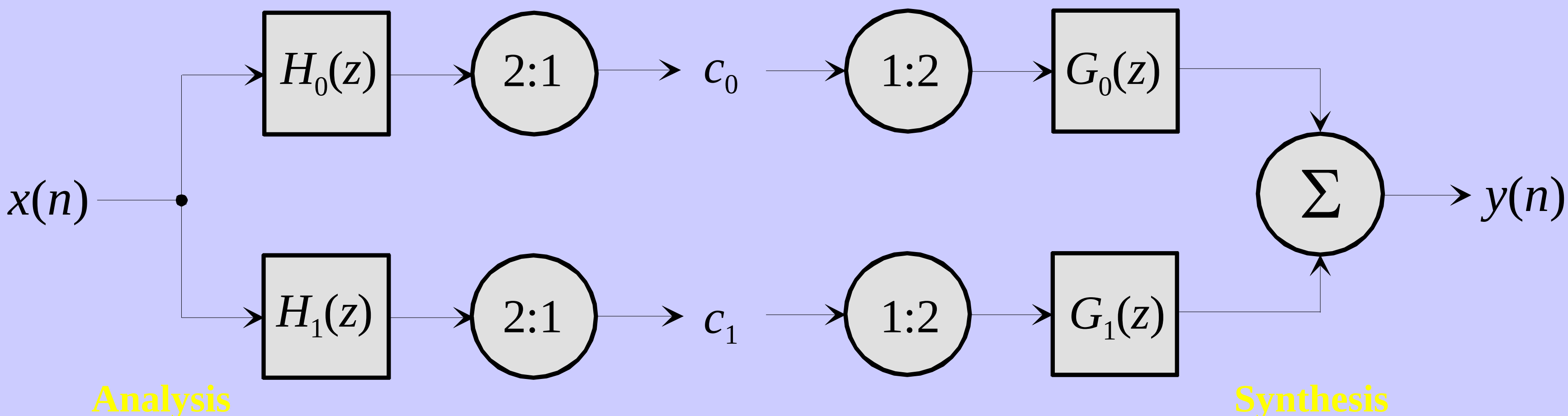
- Critically sampled filter bank with 2 bands



- Analysis low-/highpass filter pairs H_0/H_1
- Synthesis low-/highpass filter pairs G_0/G_1
- Number of samples c in frequency bands equal to total number of samples in signal x

Filter Pairs

- Perfect reconstruction is possible



$$Y(z) = \frac{1}{2}\boxed{\left[H_0(z)\cdot G_0(z) + H_1(z)\cdot G_1(z)\right]}\cdot X(z) \qquad =2\cdot z^{-k}$$

$$+\frac{1}{2}\boxed{\left[H_0(-z)\cdot G_0(z) + H_1(-z)\cdot G_1(z)\right]}\cdot X(-z). \qquad =0$$

- Subsampled signals *c* usually bear alias !

Biorthogonality Principle

- Perfect reconstruction conditions

$$Y(z)=\frac{1}{2}\left[H_0(z)\cdot G_0(z)+H_1(z)\cdot G_1(z)\right]\cdot X(z)$$

$$+\frac{1}{2}\underbrace{\left[H_0(-z)\cdot G_0(z)+H_1(-z)\cdot G_1(z)\right]}\cdot X(-z).$$

$$G_0(z)=z^k\cdot H_1(-z)$$

$$G_1(z)=-z^k\cdot H_0(-z)$$

$$\Rightarrow H_0(-z)\cdot z^k\cdot H_1(-z)$$

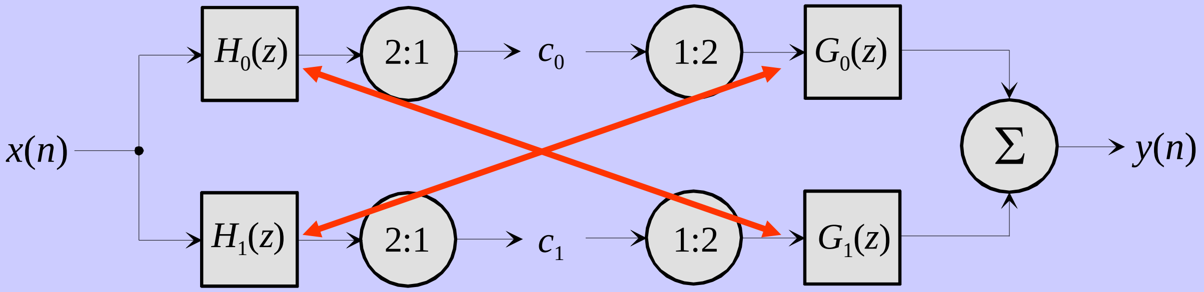
$$-H_1(-z)\cdot z^k\cdot H_0(-z)=0$$

$$Y(z)=\frac{1}{2}\underbrace{\left[H_0(z)\cdot H_1(-z)-H_1(z)\cdot H_0(-z)\right]}_{=2\cdot z^{-k}}\cdot X(z)\cdot z^k$$

$$\Rightarrow P(z)-P(-z)=2\cdot z^{-k} \quad \text{with} \quad P(z)=H_0(z)\cdot H_1(-z)$$

Biorthogonality Principle

- $H_0(z)/G_1(-z)$ and $H_1(z)/G_0(-z)$ constitute orthogonal pairs
- Low-/Highpass transfer functions not symmetric
- Linear phase or non-linear phase filters possible
- Low-/Highpass impulse responses may have different length

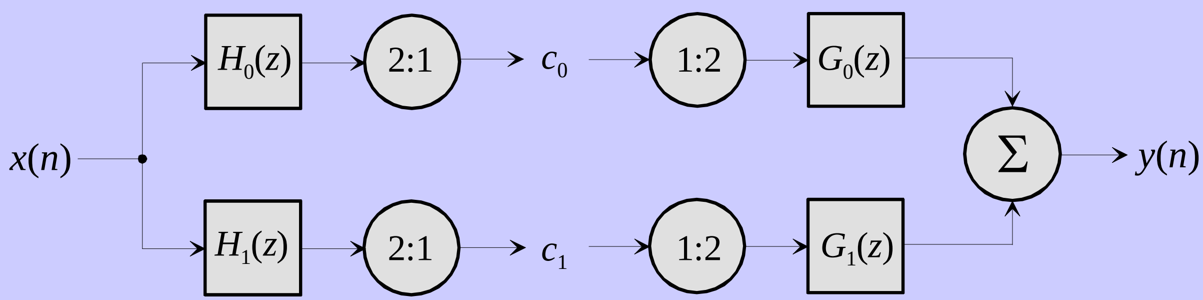


Biorthogonality Principle

- A simple biorthogonal filter pair (5/3 integer)

$$H_0^{(5/3)}(z)=\frac{1}{8}\left(-z^2+2\cdot z^1+6+2\cdot z^{-1}-z^{-2}\right)$$

$$G_0^{(5/3)}(z)=\frac{1}{2}\left(z^{-1}+2+z^{-1}\right)$$

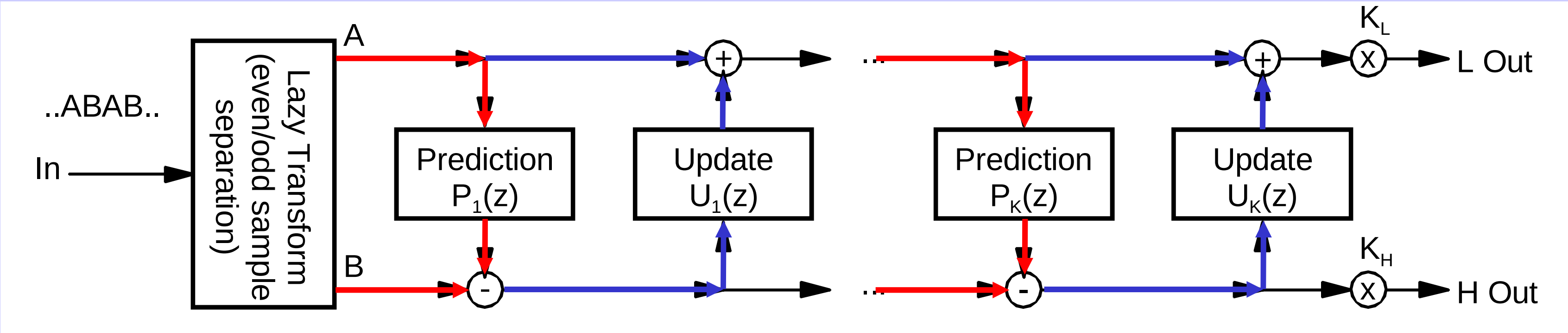


$$H_1^{(5/3)}(z)=\frac{1}{2}\left(-1+2\cdot z^{-1}-z^{-2}\right)$$

$$G_1^{(5/3)}(z)=\frac{1}{8}\left(-z^1+2+6\cdot z^{-1}+2\cdot z^{-2}-z^{-3}\right)$$

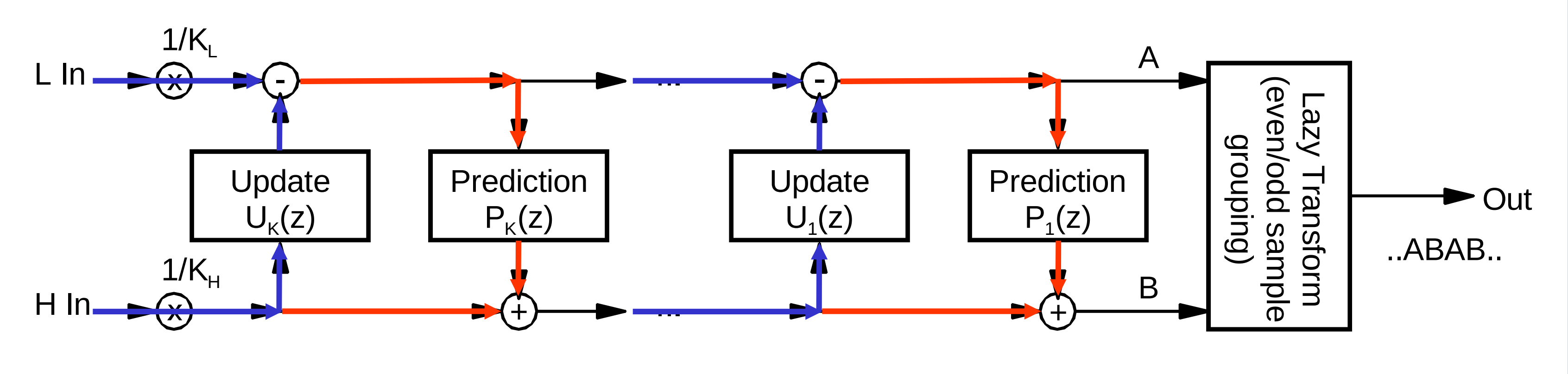
Lifting Filters

- Biorthogonal filter pairs can be factorized to be implementable in a "ladder structure"
- "Prediction" and "Update" steps using very short filter kernels are then iteratively performed
- "Lifting scheme" is most efficient implementation of wavelet filters available so far



Lifting Filters

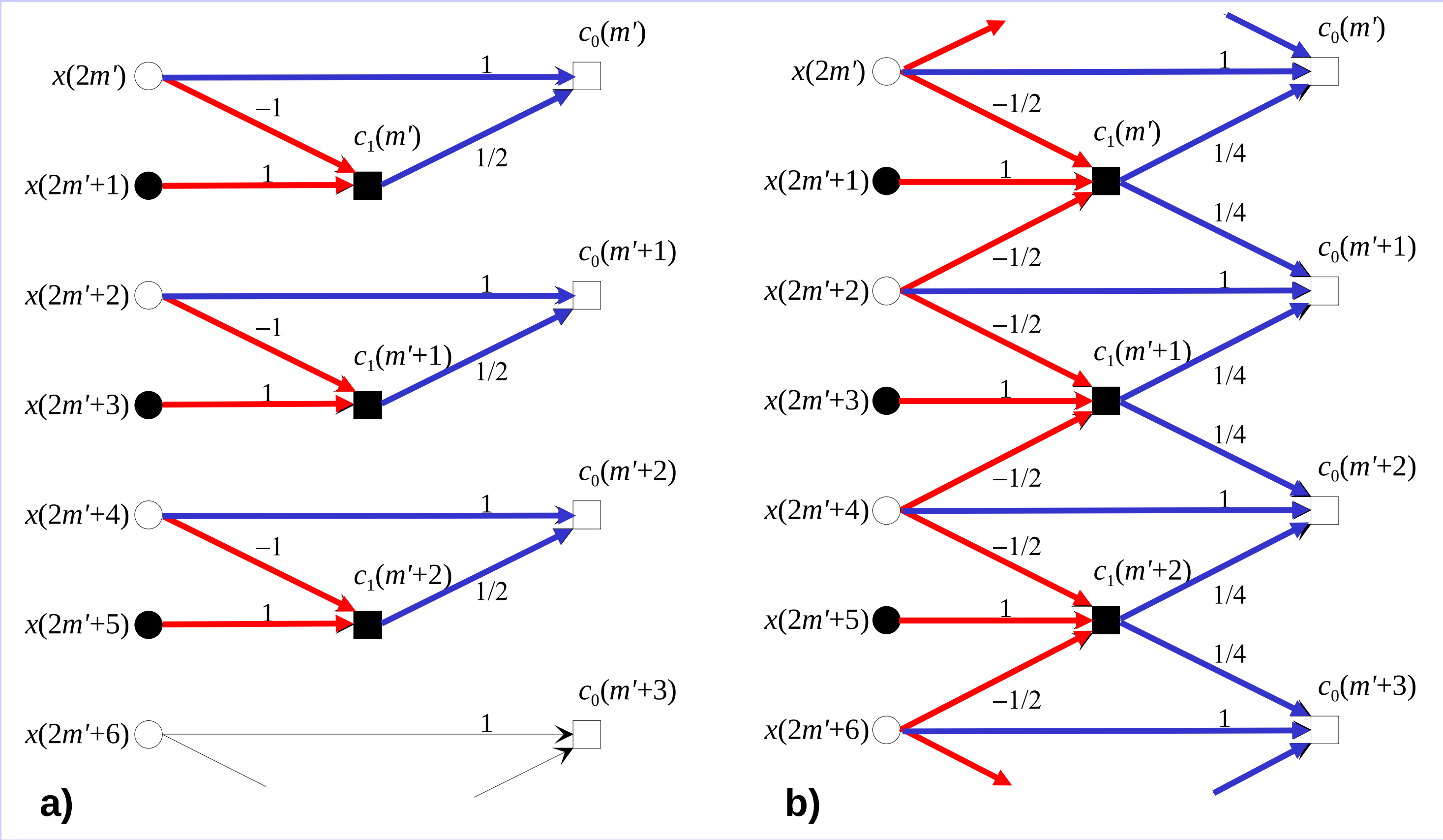
- Synthesis filter pair is implemented by inverse signal flow
- Perfect reconstruction is obvious



- Quantization of signals in the ladder branches gives integer realization of analysis and synthesis

Lifting Filters

- Signal flow diagrams of lifting implementations for (5/3) filters and Haar (2/2) filters



Motion Compensated Wavelet Coding – basic principles and classification

Wavelet Video Coding - Classification

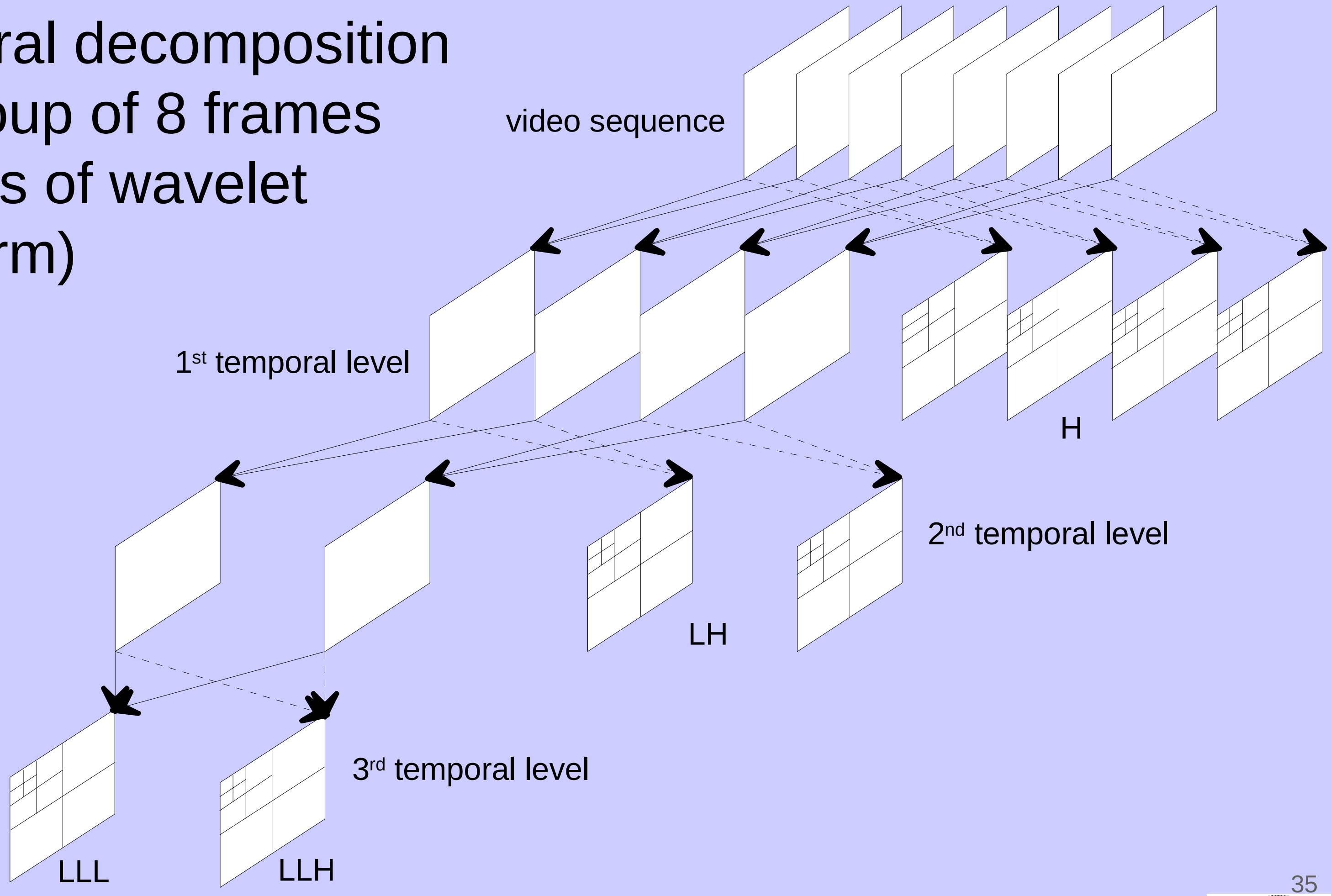
- Intraframe coding (e.g. MJPEG)
- 3D wavelet coding without MC
- Hybrid video coding using wavelet-based texture coding
- In-Band Motion Compensation Prediction
- Motion Compensated Temporal Filtering
- In-Band Motion Compensated Temporal Filtering

History

- Using transforms for interframe coding goes back to the 1970/1980s (e.g. Karlsson/Vetterli)
- Drawback was **lack of motion compensation** – first approch to filter over motion trajectories proposed by Kronander (1990)
- Solution **avoiding an overcomplete transform** developed by Ohm (1991,1994)
- Solution for **perfect reconstruction in case of half-pel motion** by Ohm/Rümmler (1997), Hsiang and Woods (1999)
- Different researchers proposed combination with **temporal axis lifting scheme** which makes virtually **any MC** possible : Pesquet/Bottreau, Luo/Li/Zhang, Secker/Taubman (2001)

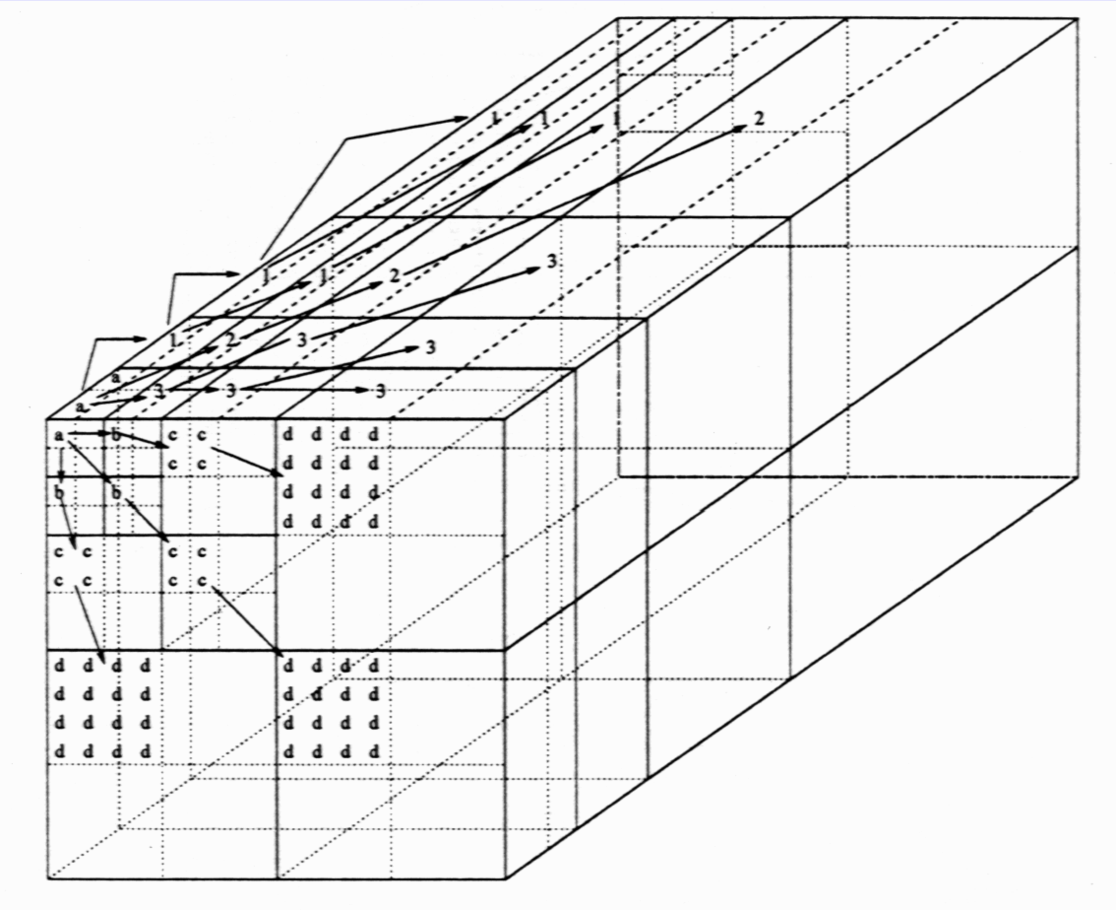
Three-dimensional Wavelet

- Temporal decomposition of a group of 8 frames (3 levels of wavelet transform)



Three-dimensional Wavelet Coding

- Extension of zero tree approach to temporal dimension
- Non-recursive coding structure



Examples:

- "3D SPIHT" by Pearlman et al.
- Layered Zero Coding (LZC) by Taubman and Zakhor (only constant displacement motion compensation)

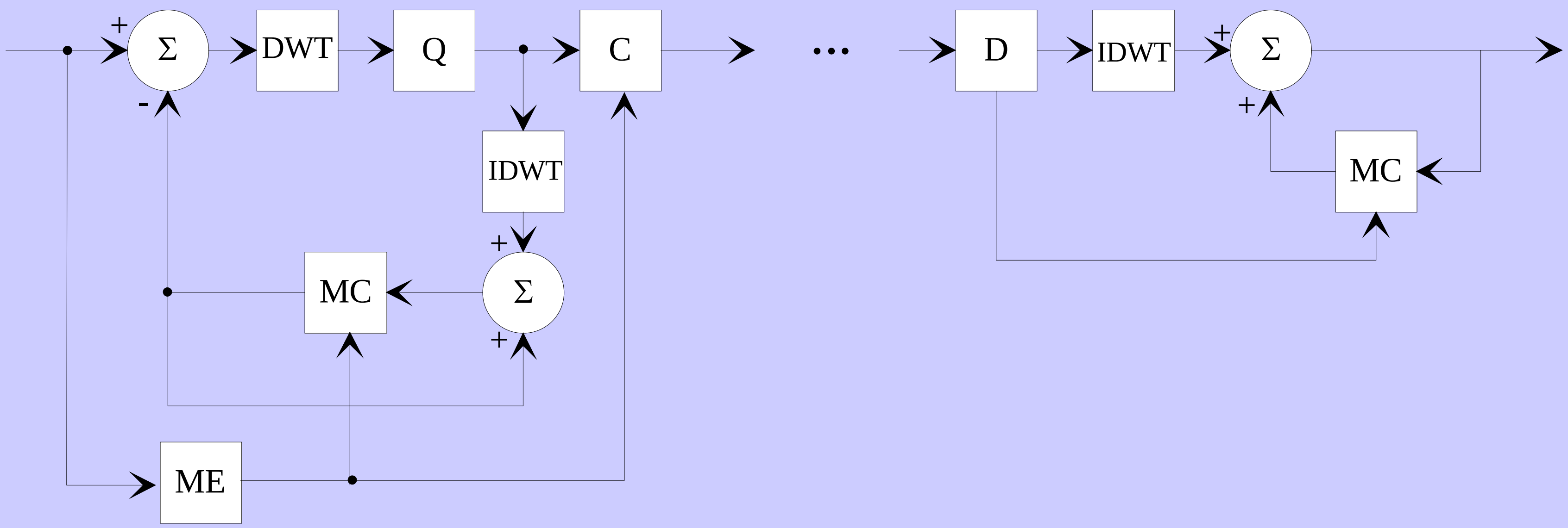
Wavelets and Motion Compensation

- **Motion compensation is key**
  - To achieve good compression performance
  - To guarantee visual quality – non MC/interframe coding with same SNR usually looks worse

- **Motion-compensated Wavelet video coding**
  - Temporal MC prediction followed by Wavelet Transform
  - Wavelet Transform followed by temporal MC prediction in wavelet domain
  - 3D Wavelet with MC

Hybrid Video Coding using Wavelets

- Replacement of DCT by Wavelet for 2D encoding in MC prediction loop



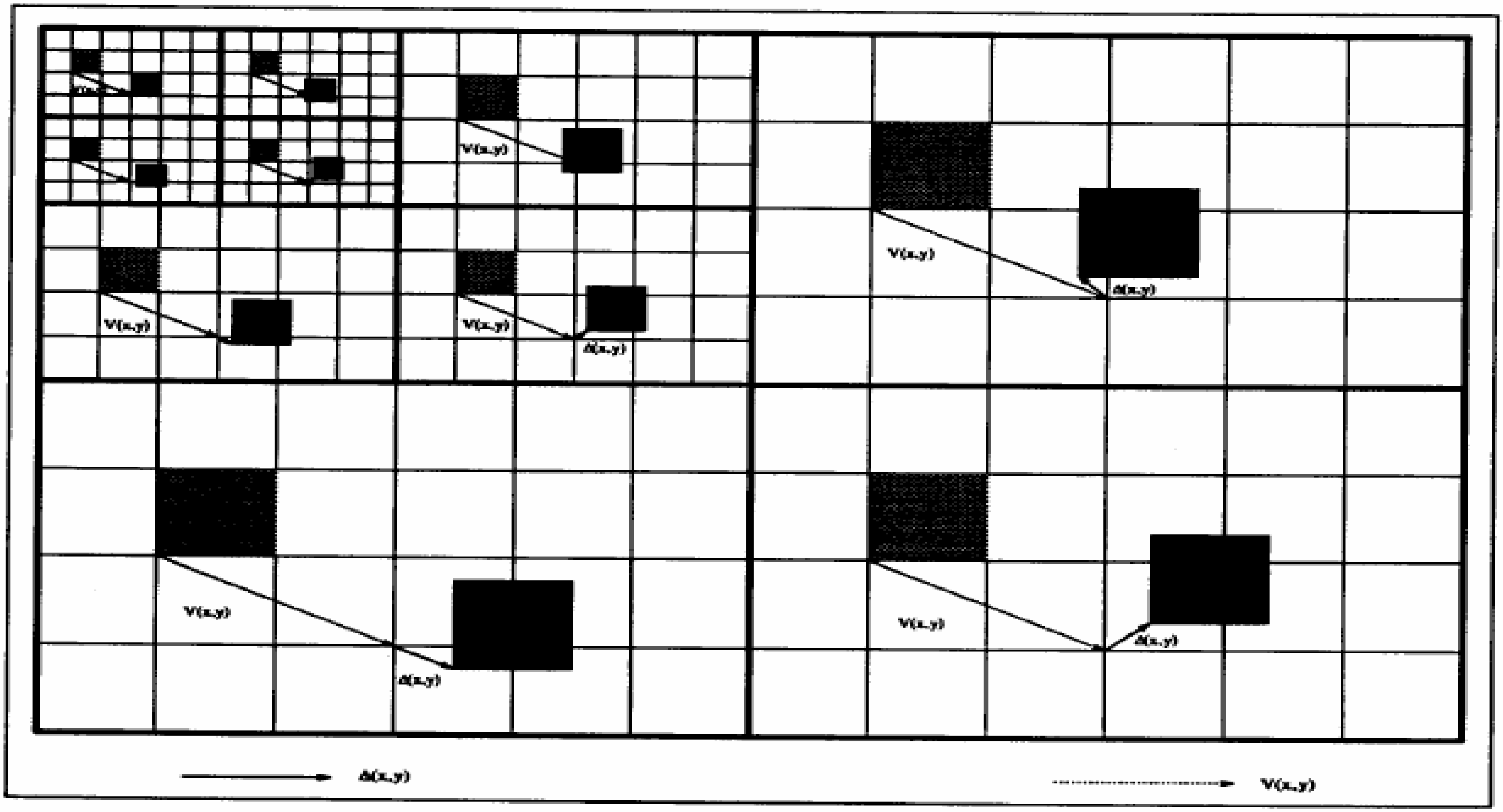
Hybrid Video Coding using Wavelets

- **Problems and possible solutions:**
  - Wavelet analysis is block-overlapping, discontinuities in motion vector field cause problems
    →Overlapping-block MC
  - Local switching between Intra/inter modes not block-wise
    →Symmetric extension at block discontinuities
- **Drift problem in MC loop is not solved**
  - This is not a real scalable solution

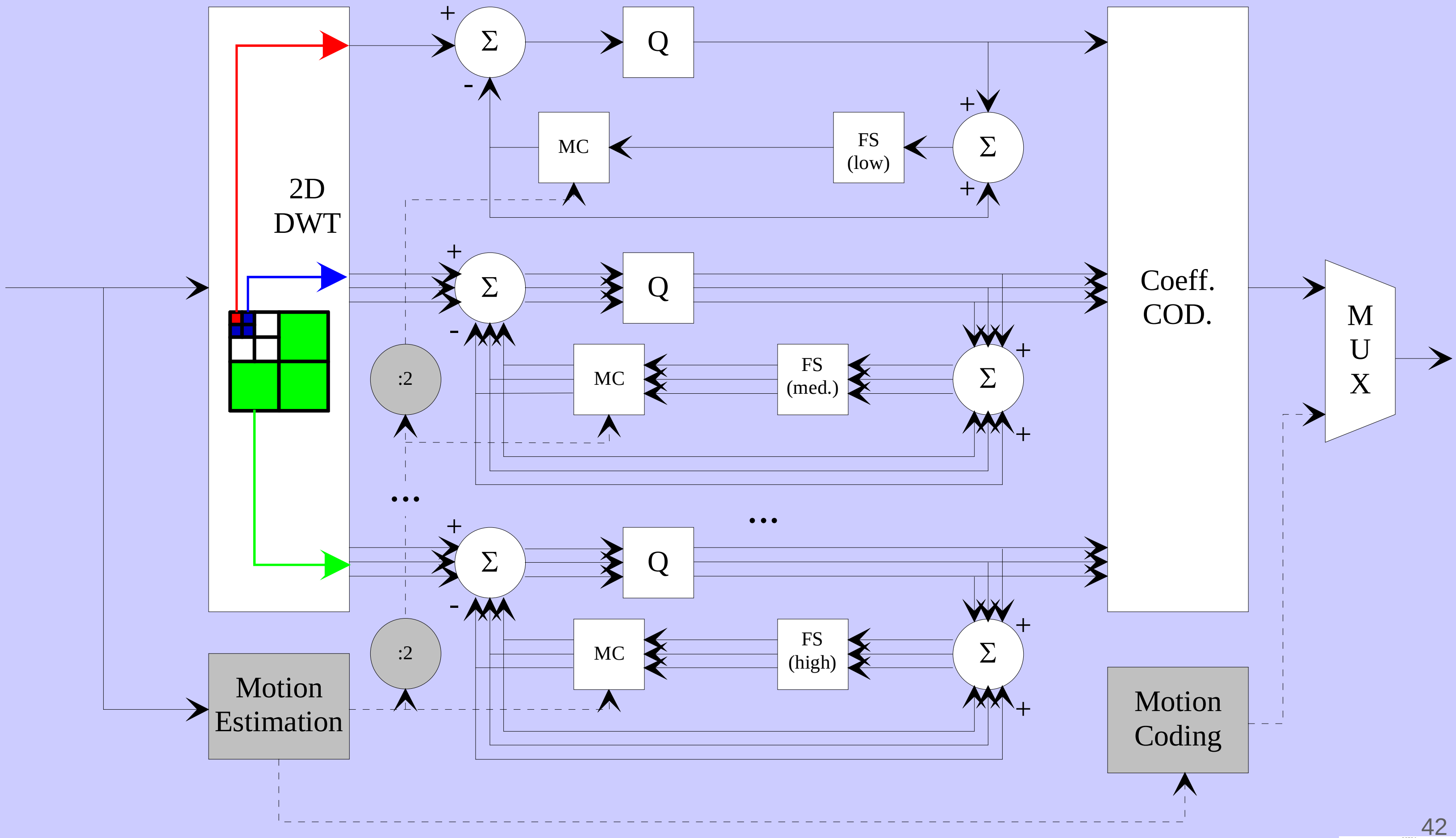
Motion compensation in the wavelet domain

- Multi-resolution nature of wavelet decomposition is ideal for providing spatially scalable video (QCIF, CIF, SD, and HD)
- Subbands are highly correlated in the temporal direction
  - Motion estimation and compensation can significantly reduce the temporal correlation
- Classical approach
  -
  - *Ref: Y. Zhang and S. Zafar, IEEE CSVT, Sept. 1992.*

Multi-Resolution Motion Compensation



MC in Wavelet Domain - Encoder



MC in Wavelet Domain

- Variable block size of the m-th layer subbands for M-level decomposition

$$p2^{M-m} \times p2^{M-m}$$

- Motion vector for each subband (j=1,..,3)

$$V_{i,j}^{(m)}(x,y) = V_{i,0}^{(m)}(x,y)2^{M-m} + \Delta_{i,j}^{(m)}(x,y)$$

i: frame number, j: subband index (j=0,…,3), m: layer number

- Adaptive search range for each subband

MC in Wavelet Domain – Advantages and Drawbacks

- **Multiple (separate) MC loops for wavelet bands**
  - one set of motion parameters may be used for all
- **No drift problem in spatial scalability**
  - Possible to skip higher frequency bands
- **Switching to "intra" coding mode without penalty**
  - Inverse DWT is applied to images (not differences)
- **Inefficiency of MC prediction in high bands**
  - Significant performance loss compared to ME/MC in spatial domain (e.g.1-2dB)
  - **The shift variant property** of wavelet decomposition

Motion-Compensated Temporal Filtering

Motion Compensated Haar Filters

- Non-orthonormal Haar filter basis

$$H_0(z) = \frac{1}{2}\left(1 + z_1^{\tilde{k}} \cdot z_2^{\tilde{l}} \cdot z_3^{-1}\right)$$

MC shift

Delay by one frame

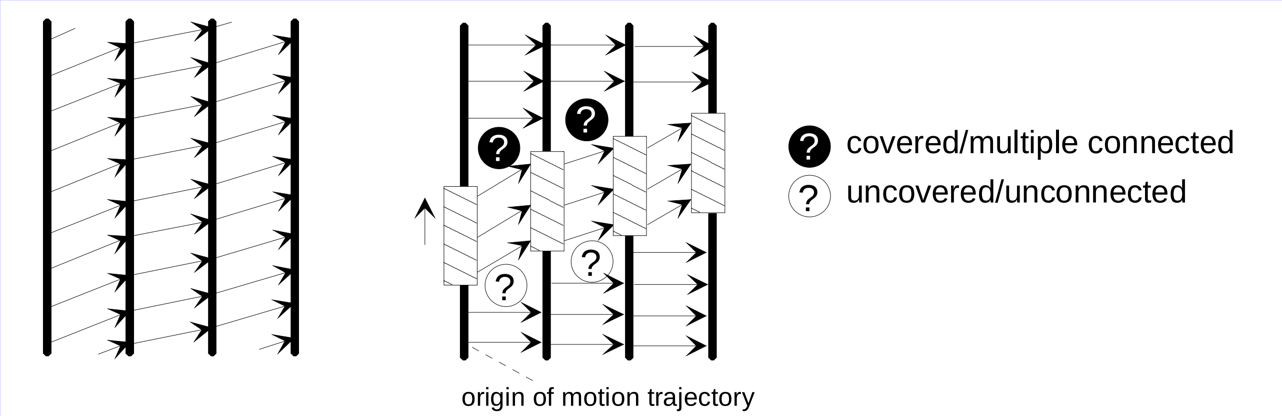
$$H_1(z) = -z_1^{k} \cdot z_2^{l} + z_3^{-1}.$$

Motion Compensated Haar Filters

- This motion-compensated filtering is no problem whenever unique sample-wise correspondences exist between two frames

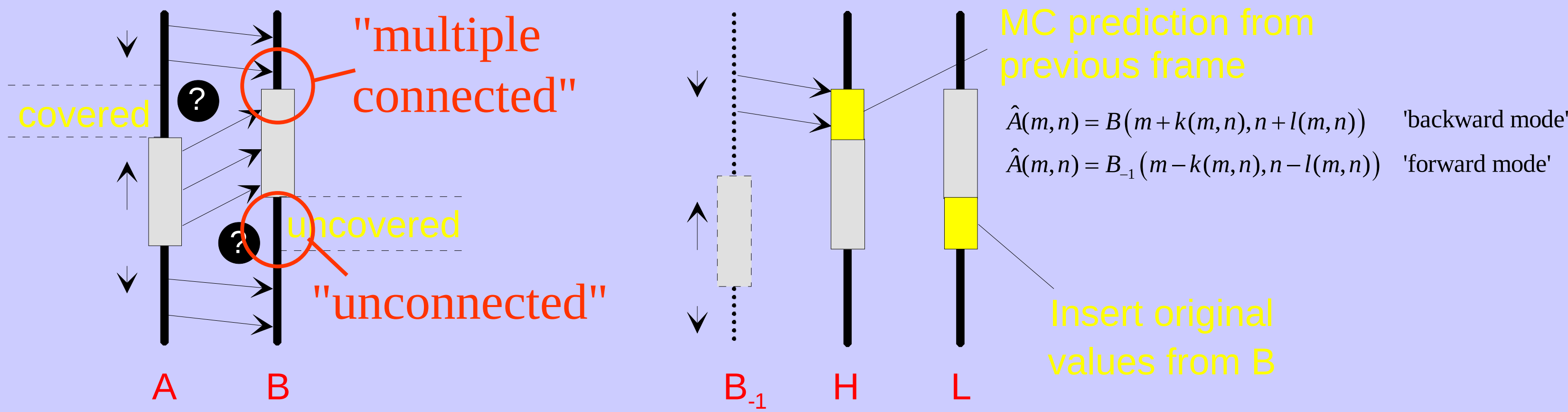
$$\begin{aligned}\tilde{k}(m_B,n_B)&=-k(m_A,n_A)\\ \tilde{l}(m_B,n_B)&=-l(m_A,n_A)\end{aligned}\qquad \text{with}\qquad \begin{cases} m_B=m_A+k(m_A,n_A)\\ n_B=n_A+l(m_A,n_A).\end{cases}$$

- Real motion vector fields are discontinuous, such that correspondences may not be unique



Motion Compensated Haar Filters

- Substitution technique for covered/uncovered areas allows perfect reconstruction at motion discontinuities



$$L(m,n) = 0.5 \cdot B(m,n) + 0.5 \cdot A\left(m + \tilde{k}(m,n), n + \tilde{l}(m,n)\right)$$

$$H(m,n) = A(m,n) - B\left(m + k(m,n), n + l(m,n)\right)$$

$$L(m,n) = B(m,n) \qquad \text{if 'unconnected'}$$

$$H(m,n) = A(m,n) - \hat{A}(m,n) \qquad \text{if 'multiple connected'}$$

Motion Compensated Haar Filters

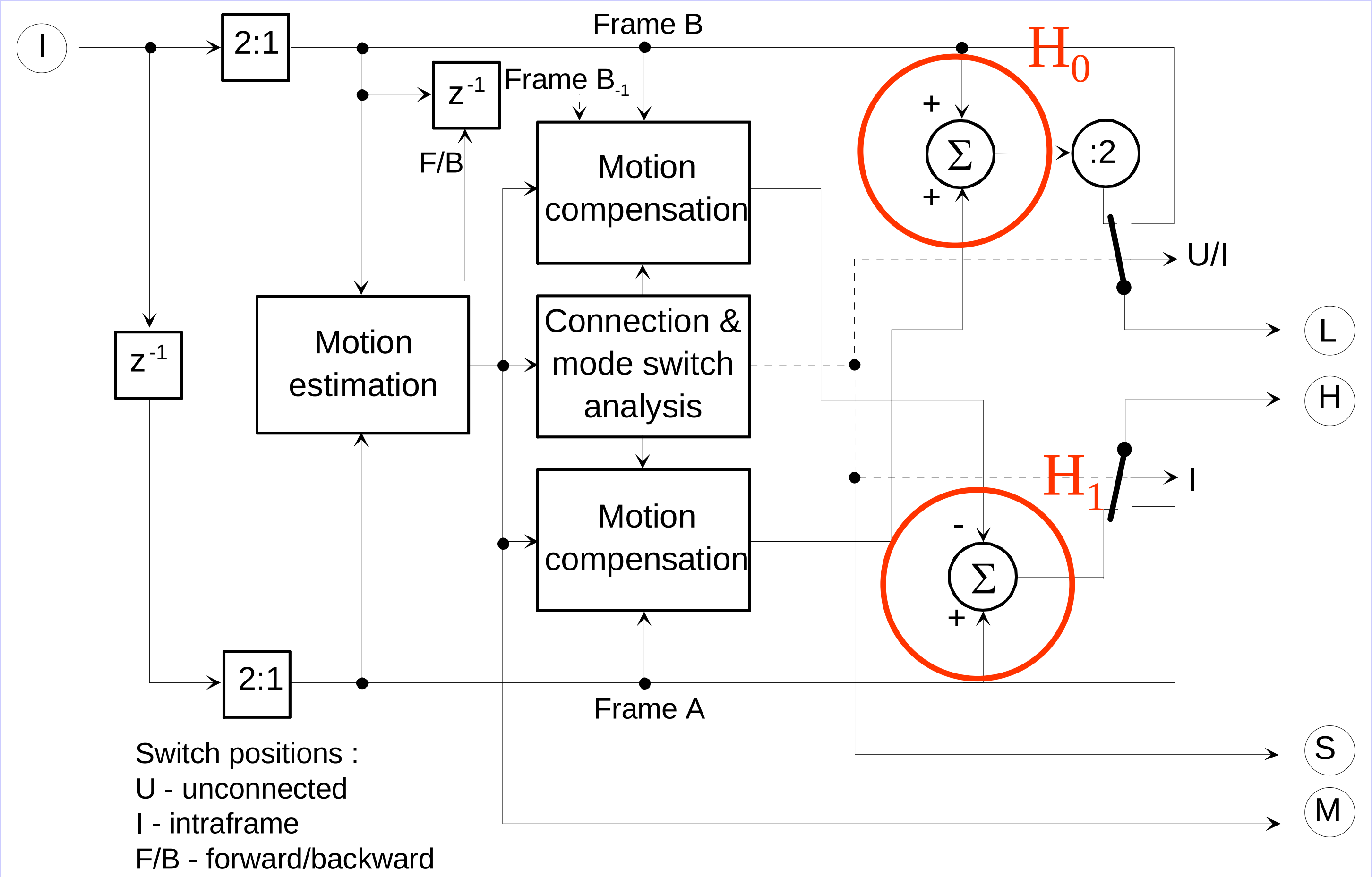
- Synthesis is straightforward in case of full-pixel correspondences

$$\tilde{A}(m,n) = L\left(m+k(m,n), n+l(m,n)\right) + 0.5H(m,n)$$
$$\tilde{B}(m,n) = L(m,n) - 0.5H\left(m+\tilde{k}(m,n), n+\tilde{l}(m,n)\right),$$

$$\tilde{B}(m,n) = L(m,n) \qquad \text{if 'unconnected'}$$
$$\tilde{A}(m,n) = \hat{A}(m,n) + H(m,n) \qquad \text{if 'multiple connected'}.$$

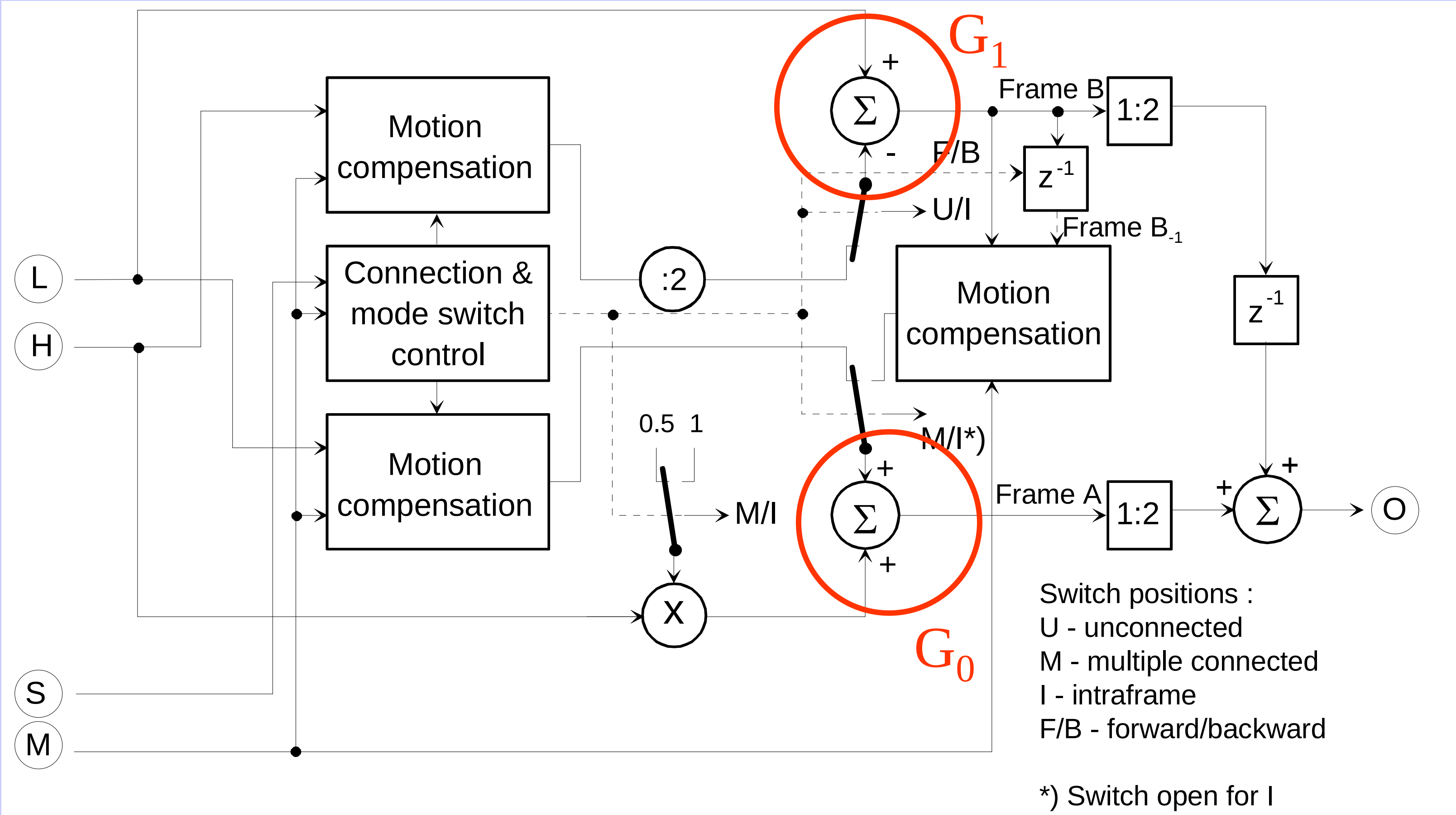
Motion Compensated Haar Filters

- 2-band temporal Haar analysis filter

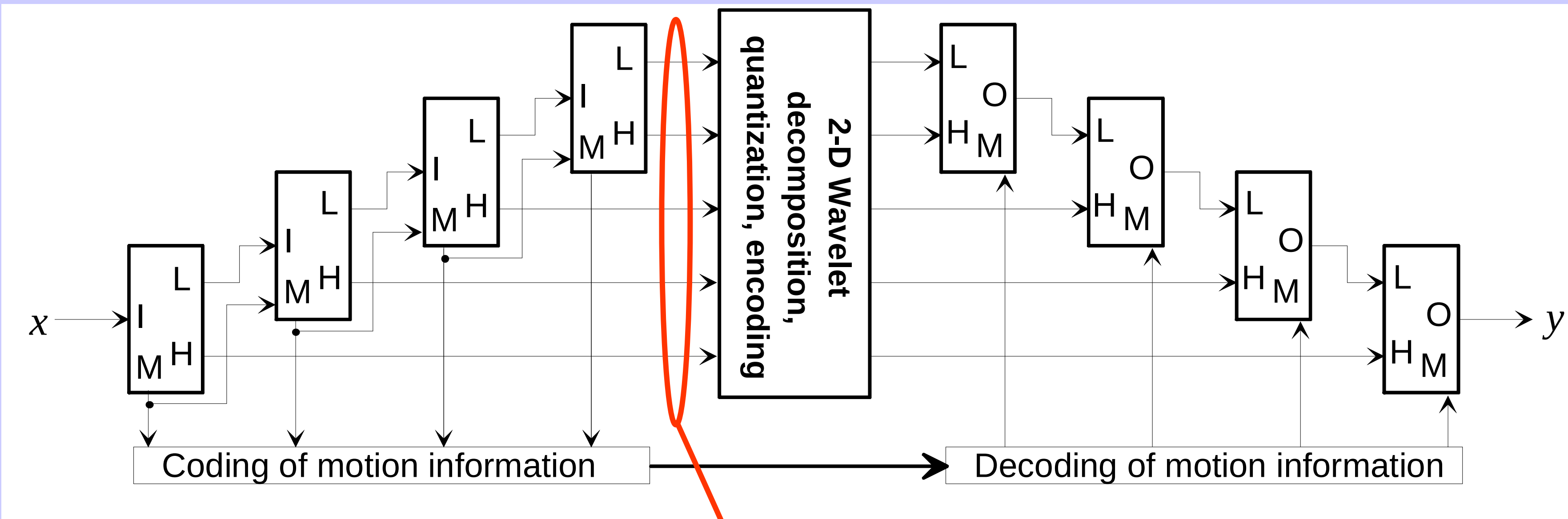


Motion Compensated Haar Filters

- 2-band temporal Haar synthesis filter



Motion Compensated Temporal Wavelet Tree

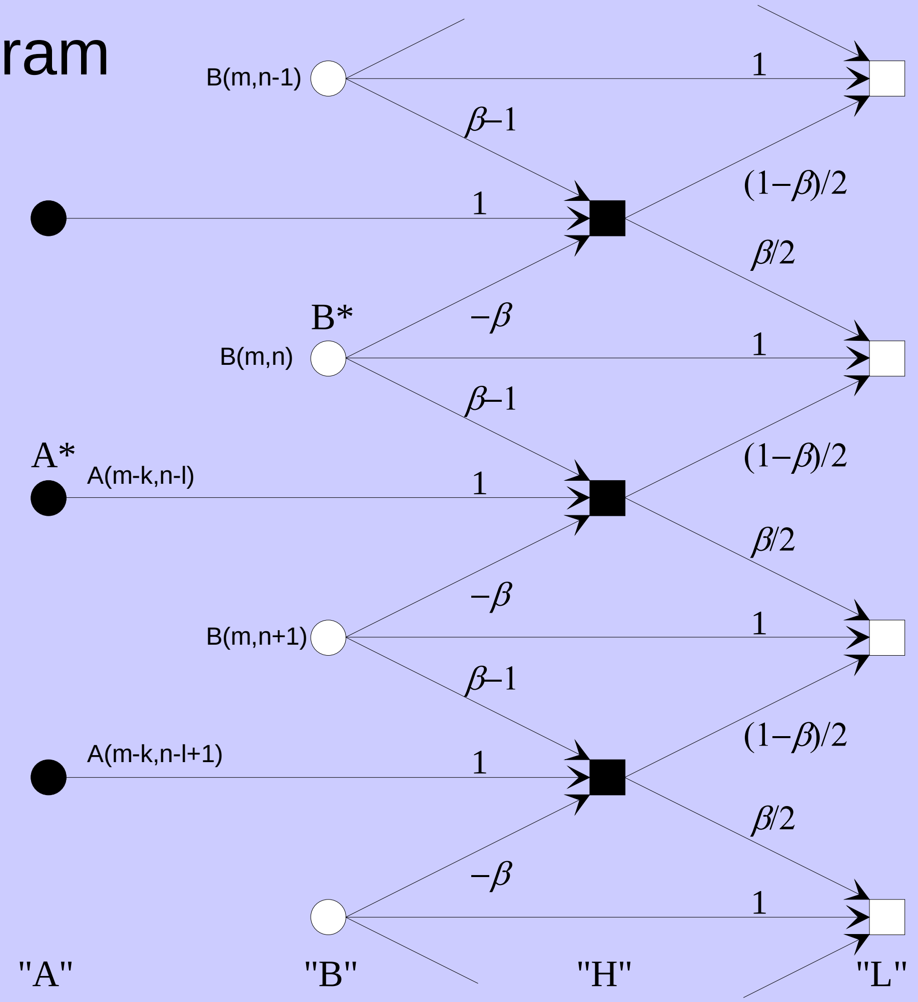


Scaling and Wavelet coefficients
from temporal analysis (arranged as 2D images)

Motion-compensated Lifting Filters

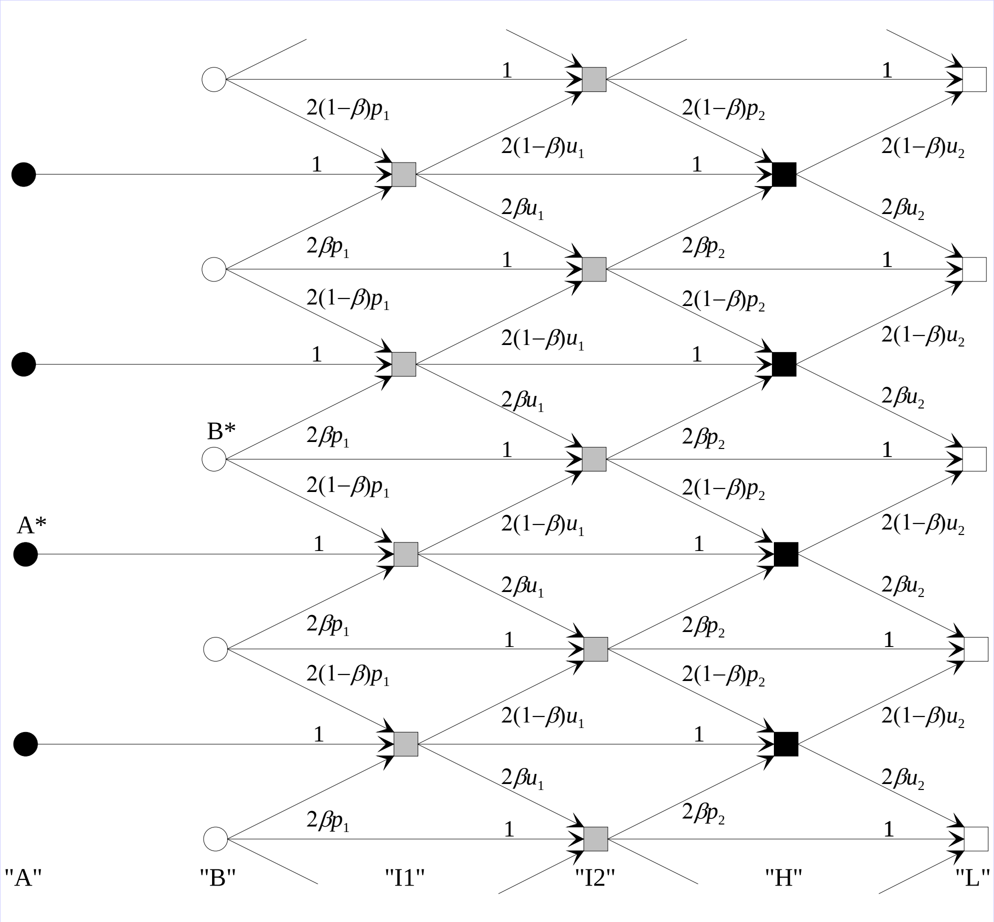
- Signal flow diagram

Vertical shift
by $l+\beta$
pixels



Motion-compensated Lifting Filters

- Extensible to longer interpolation filters, e.g. (9/7)



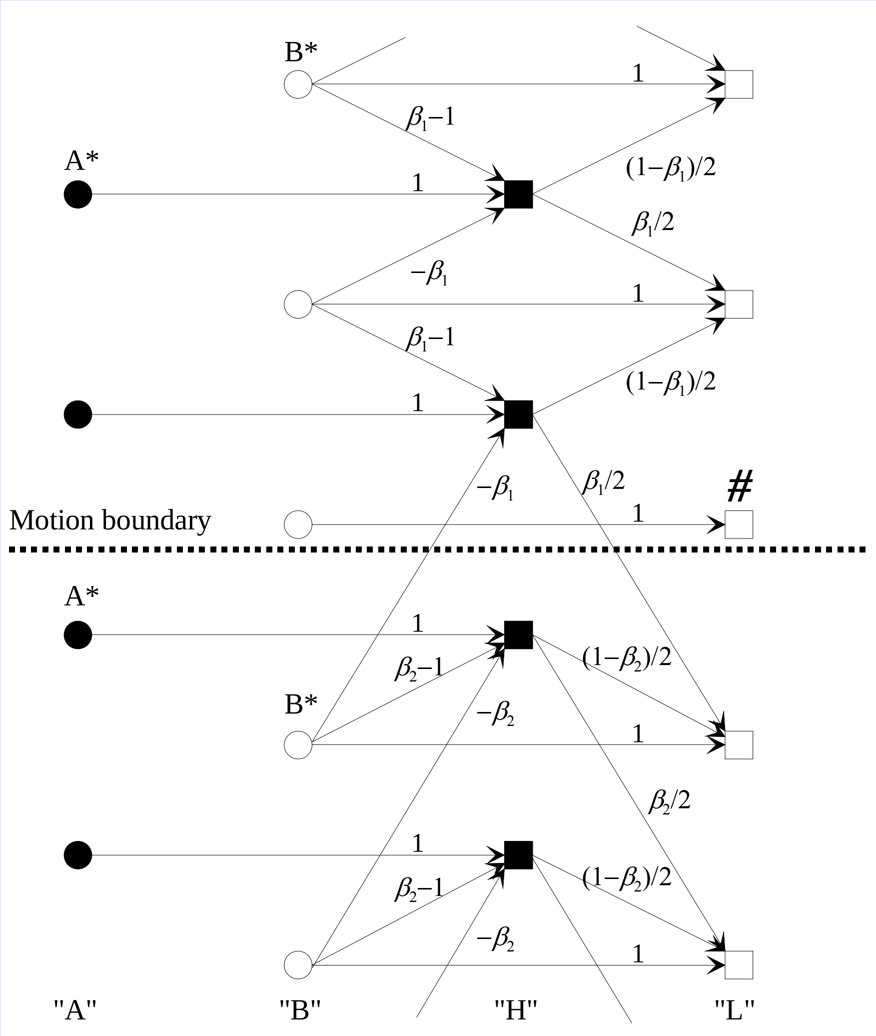
With β=0.5:
Equivalent to the half-pel P.R. method proposed in [Ohm,Rümmler 97] and used in [Hsiang, Woods 99]

Motion-compensated Lifting filters

- The principle is straightforwardly extensible to
  - longer wavelet filters
  - separable (or non-separable 2D filters)
  - change of a with any position (e.g. MC based on affine model, dense motion vector fields)
- Coincidence of motion correspondences in adjacent prediction and update steps must be observed
- Lifting implementation of temporal wavelet filtering also leads to an elegant interpretation of previous covered/uncovered pixel substitution
- Very efficient implementation

Motion-compensated Lifting filters

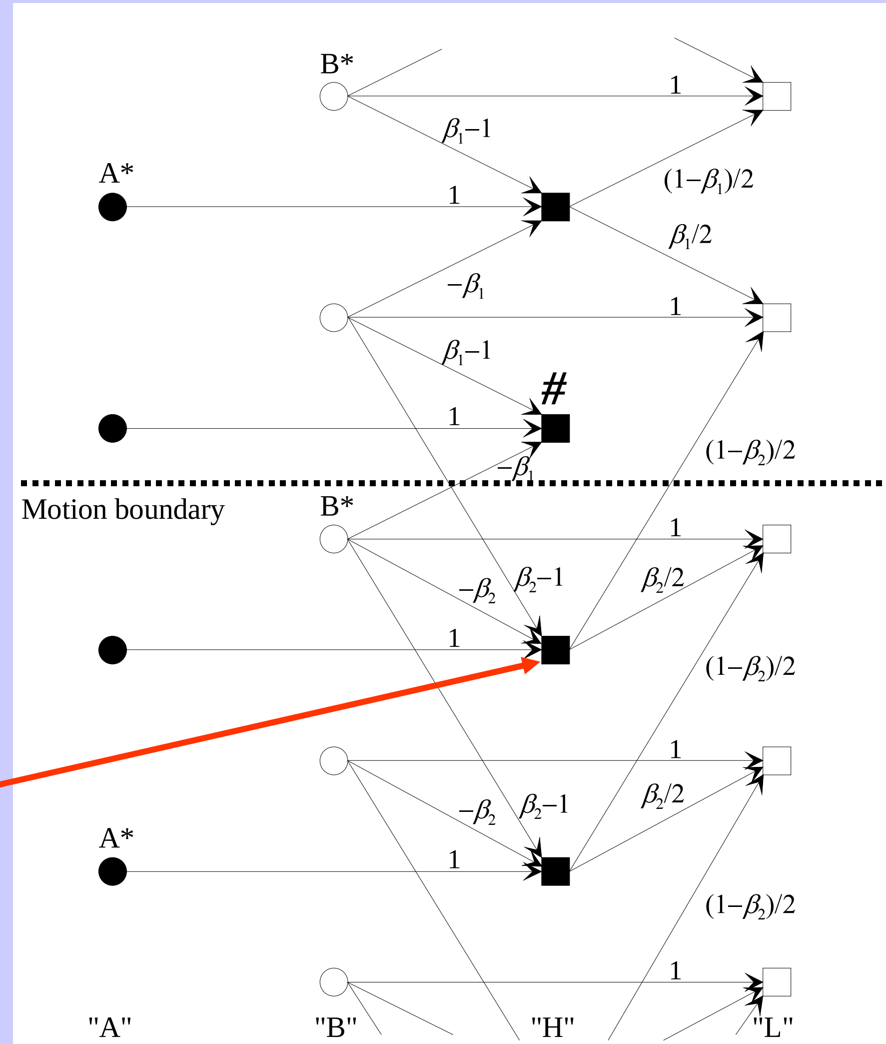
- Adaptation at motion boundaries : "uncovered/ unconnected" case
- Additional "lazy" pixel(s) in frame B



Motion-compensated Lifting filters

- Adaptation at motion boundaries : "Covered/multiple connected" case
- Additional prediction pixel(s) in frame A/H

This pixel might also take the 'unconnected' role!

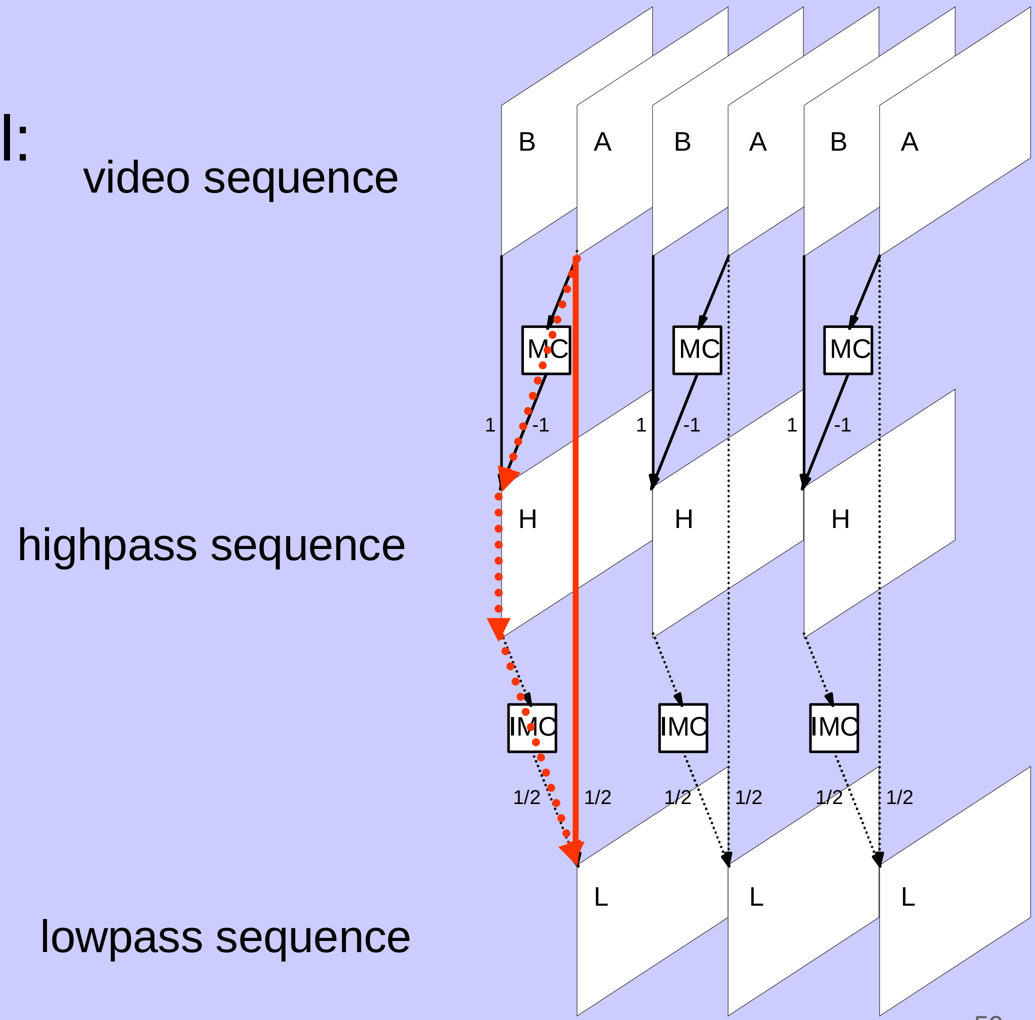


Motion-compensated Lifting filters

- **Frame-wise or localized implementation of intra coding is a key concept in MC prediction coders**
  - Switching to **intra mode** is applied whenever no motion correspondence can be found, e.g. scene changes or uncovered areas
- In MC temporal filtering
  - the equivalent is an **adaptation of wavelet tree depth**
  - but intra coding could also be applied individually in the prediction and update steps
- In general, **localized mode switching** can be included in a simple way in the lifting structure

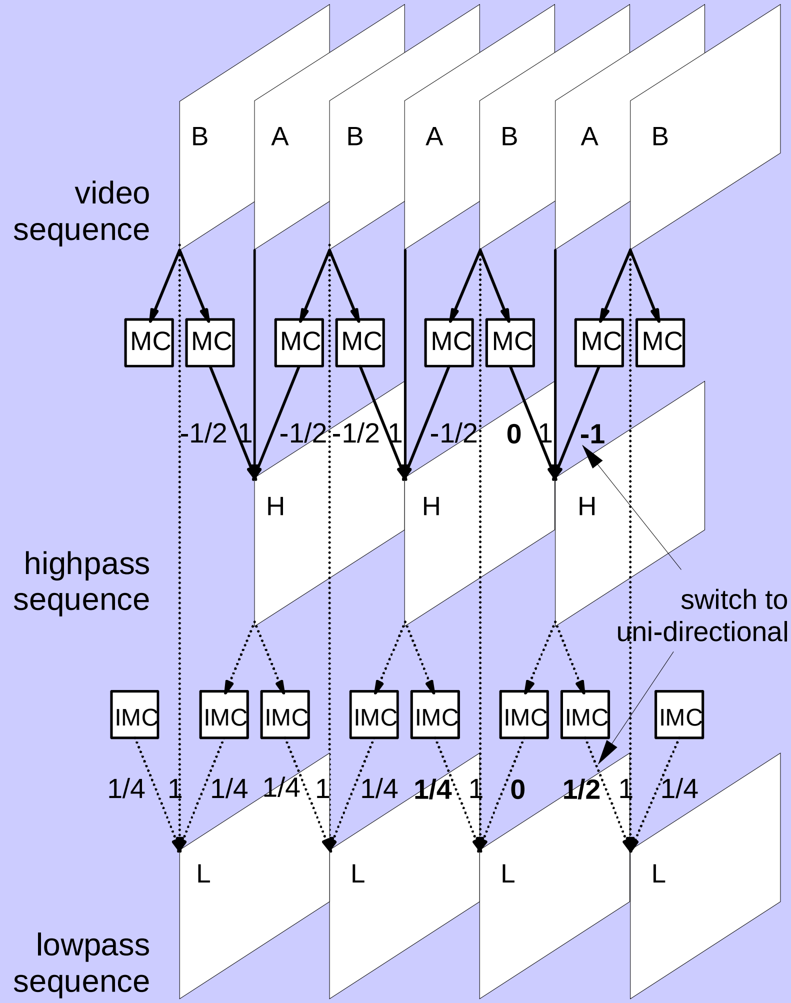
More Flexibility in MC Lifting Filters

- Different view of one transform level: Temporal-axis lifting filters, including 2D MC in cross paths
- MC and IMC should be related such that pixels from A correspond to L



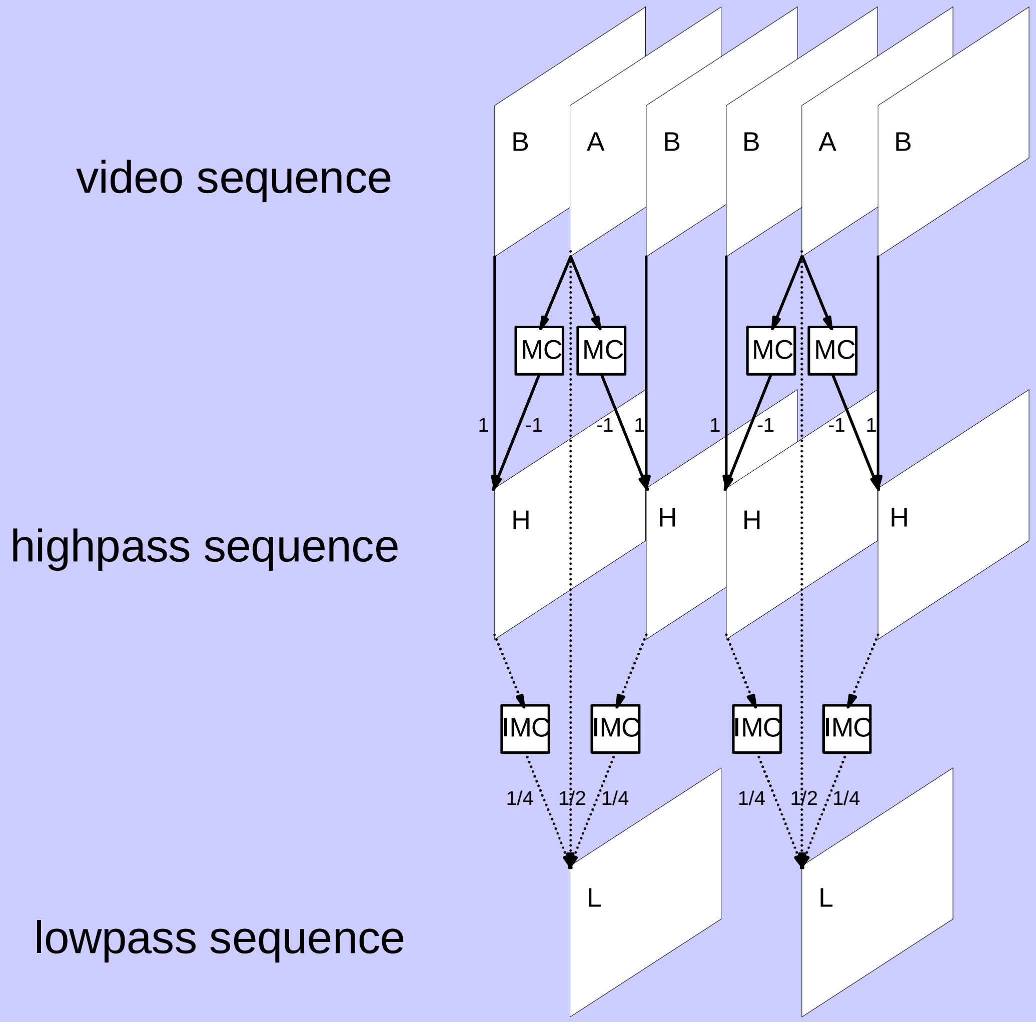
More Flexibility in MC Lifting Filters

- Extension to longer temporal filters (5/3)
- H frames equivalent to bidirectional prediction
- Forward/backward switching possible
- Better coding efficiency
- No temporal blocking
- More memory
- Higher delay
- More motion vectors



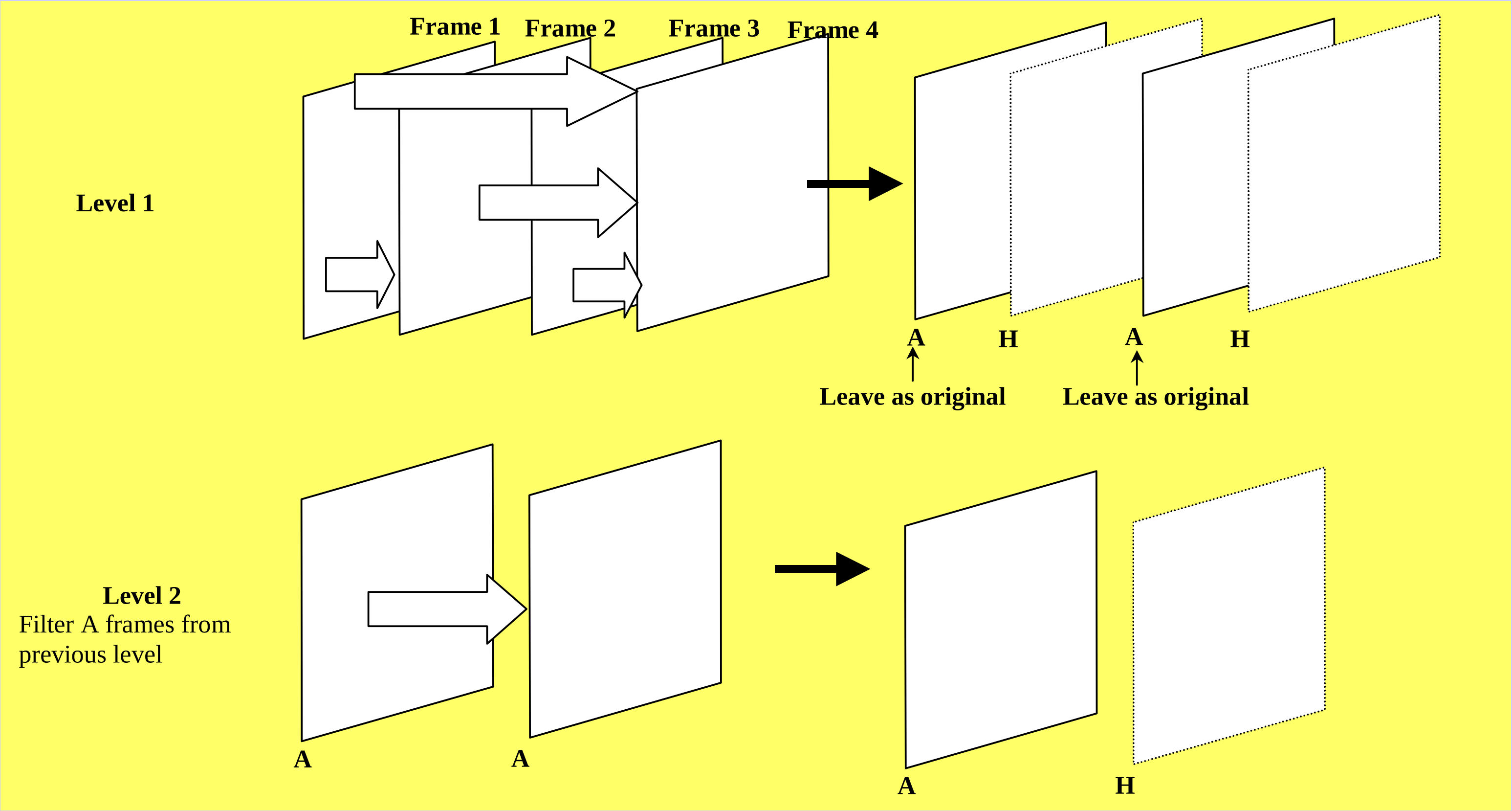
More Flexibility in MC Lifting Filters

- Non-dyadic decomposition
- Temporal block units of 3 frames
- E.g. 30/10 Hz temporal scalability
- Can be extended to bidirectional MC in prediction step



Low-Delay modes in MC Temporal Filtering

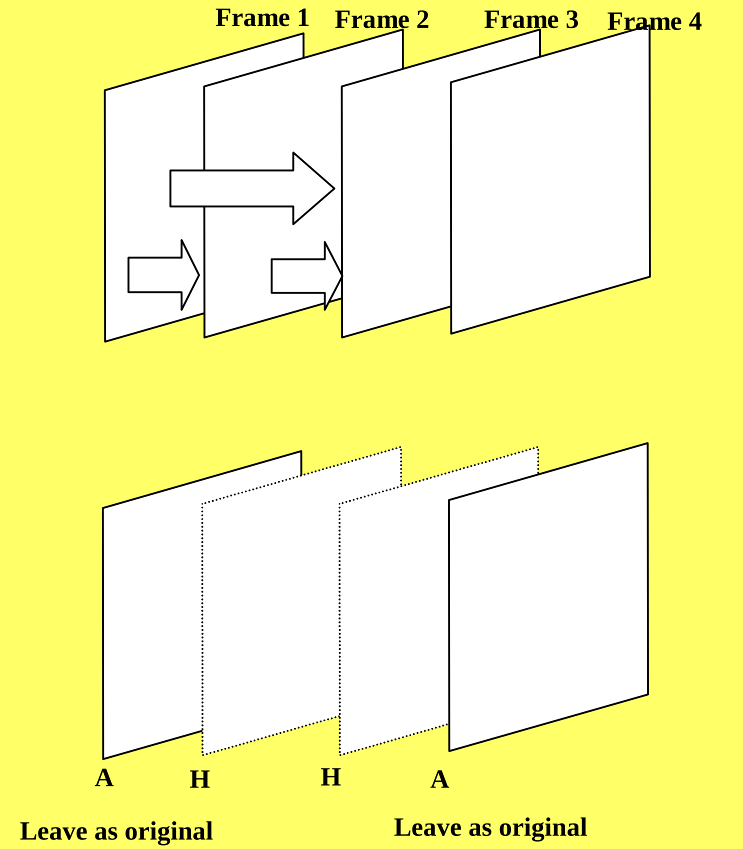
- Temporal pyramid decomposition with omission of update step ("A" frames left as originals)



Low-Delay modes in MC Temporal Filtering

"A" frames can be inserted at arbitrary locations ->

the sequence can be decoded at non-dyadic lower frame rates

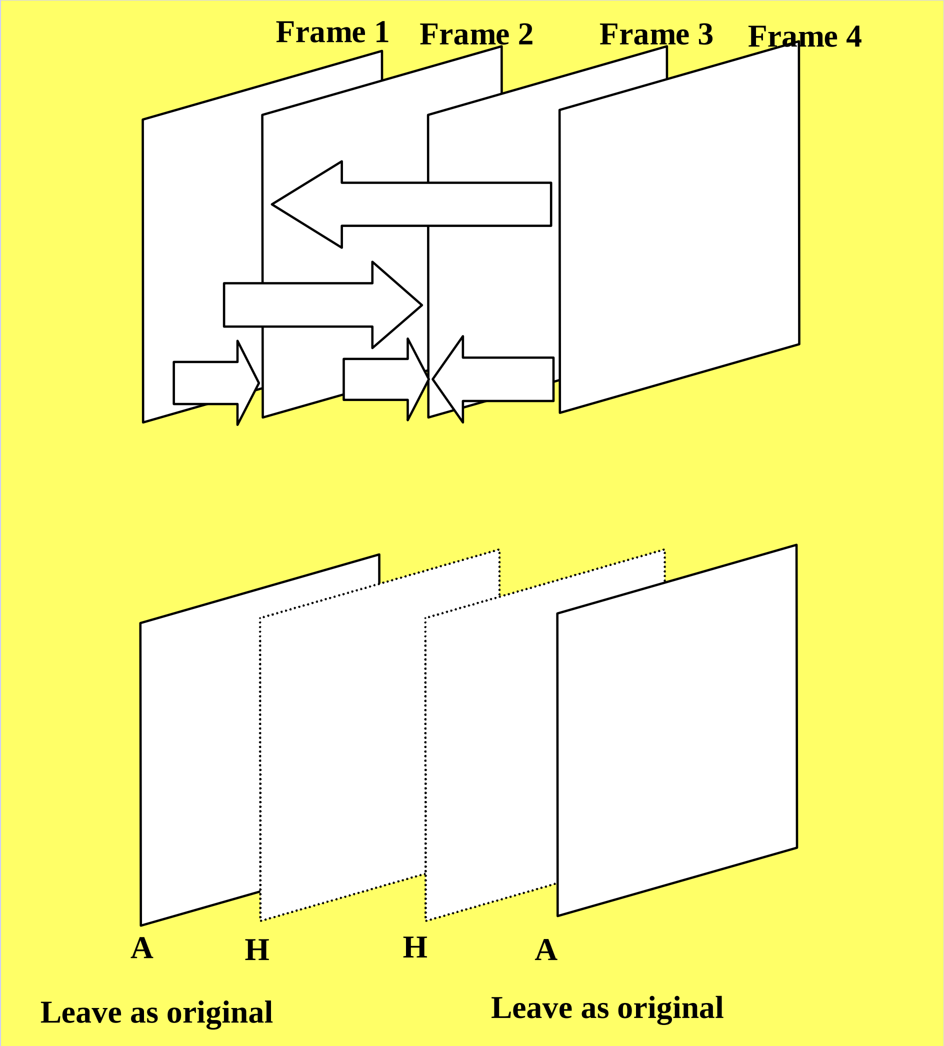


Modified MCTF Scheme

Low-Delay modes in MC Temporal Filtering

- **A frames allow implementation of a low-delay mode**
  - **A frames can be encoded and transmitted immediately, but must be stored for future reference**
  - **H frames can be encoded and transmitted immediately in any of the schemes**
- **Disadvantage :**
  - **lower coding efficiency**
  - **May be compensated by improved prediction**

Low-Delay modes in MC Temporal Filtering



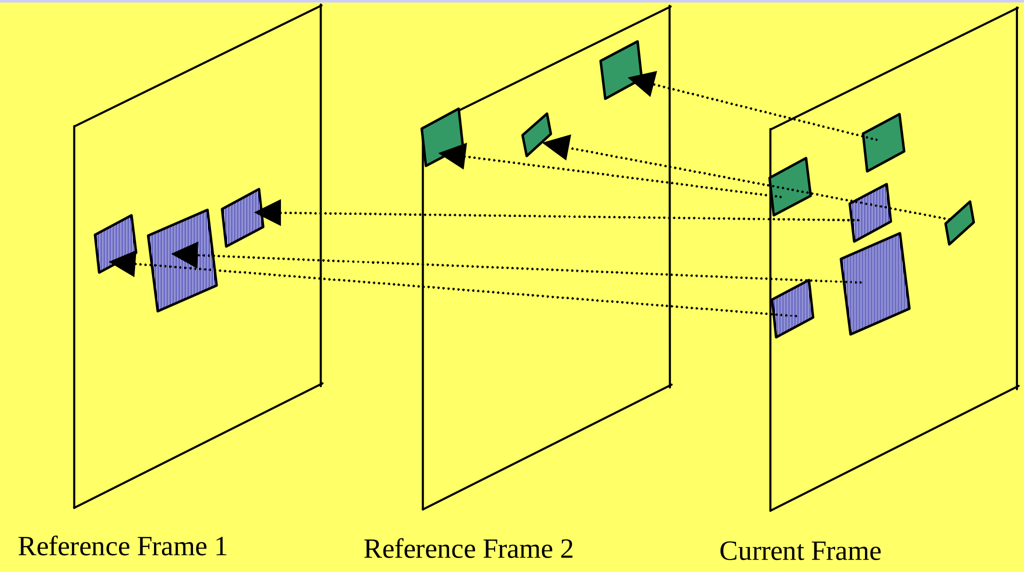
Inclusion of bi-directional prediction

Choice between 3 modes:

- Use backward prediction
- Use forward prediction
- Use the average block of the backward and forward predictions while filtering

Low-Delay modes in MC Temporal Filtering

- Prediction step can be enriched by selecting multiple reference frames



Advantages

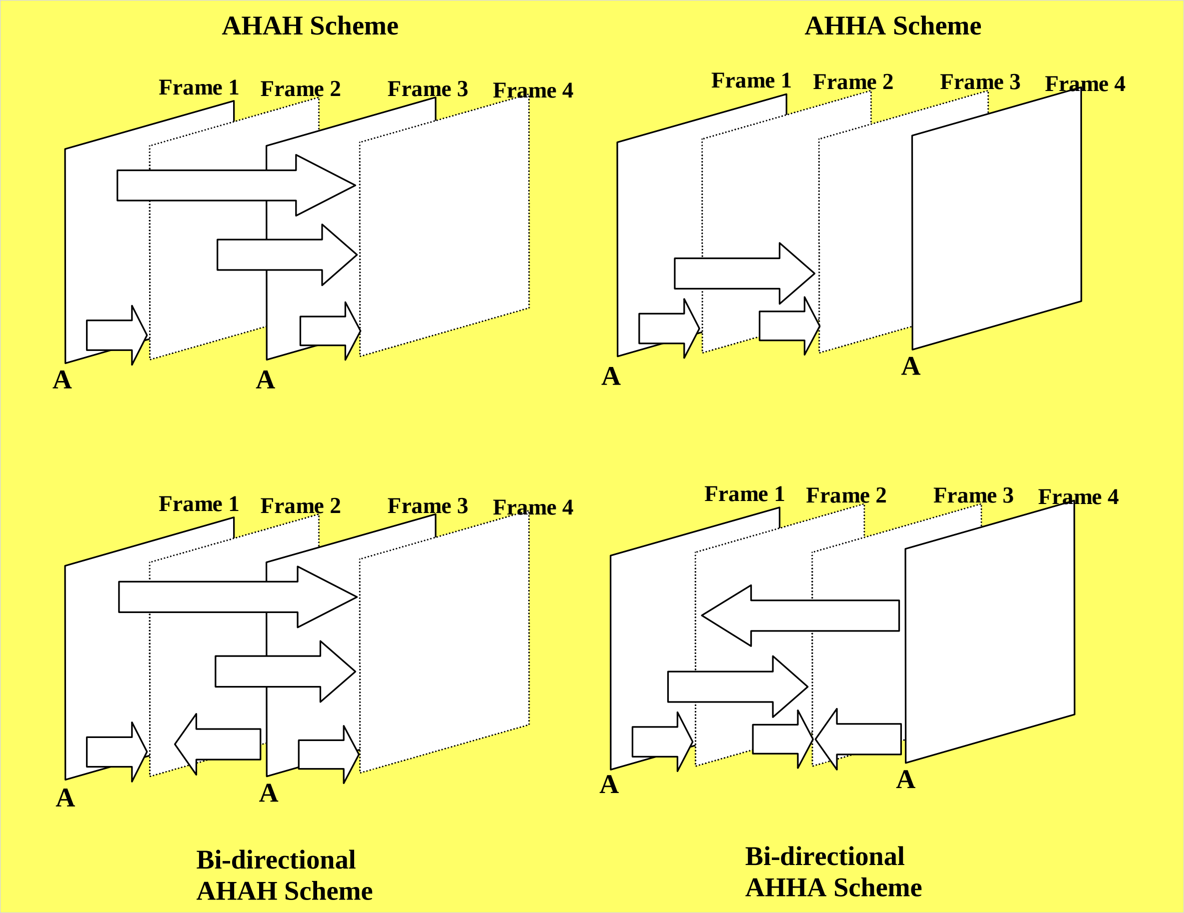
- Improved coding efficiency
- Easy to incorporate the advanced MC & ME options used by predictive coders (H.264/AVC, MPEG-4 etc.)
- Reduced no. of unconnected pixels

Disadvantages

- Sacrifice Temporal Scalability
- Prediction drift can become a problem when decoding at lower bit-rates

Low-Delay modes in MC Temporal Filtering

- Different configurations at any level of pyramid

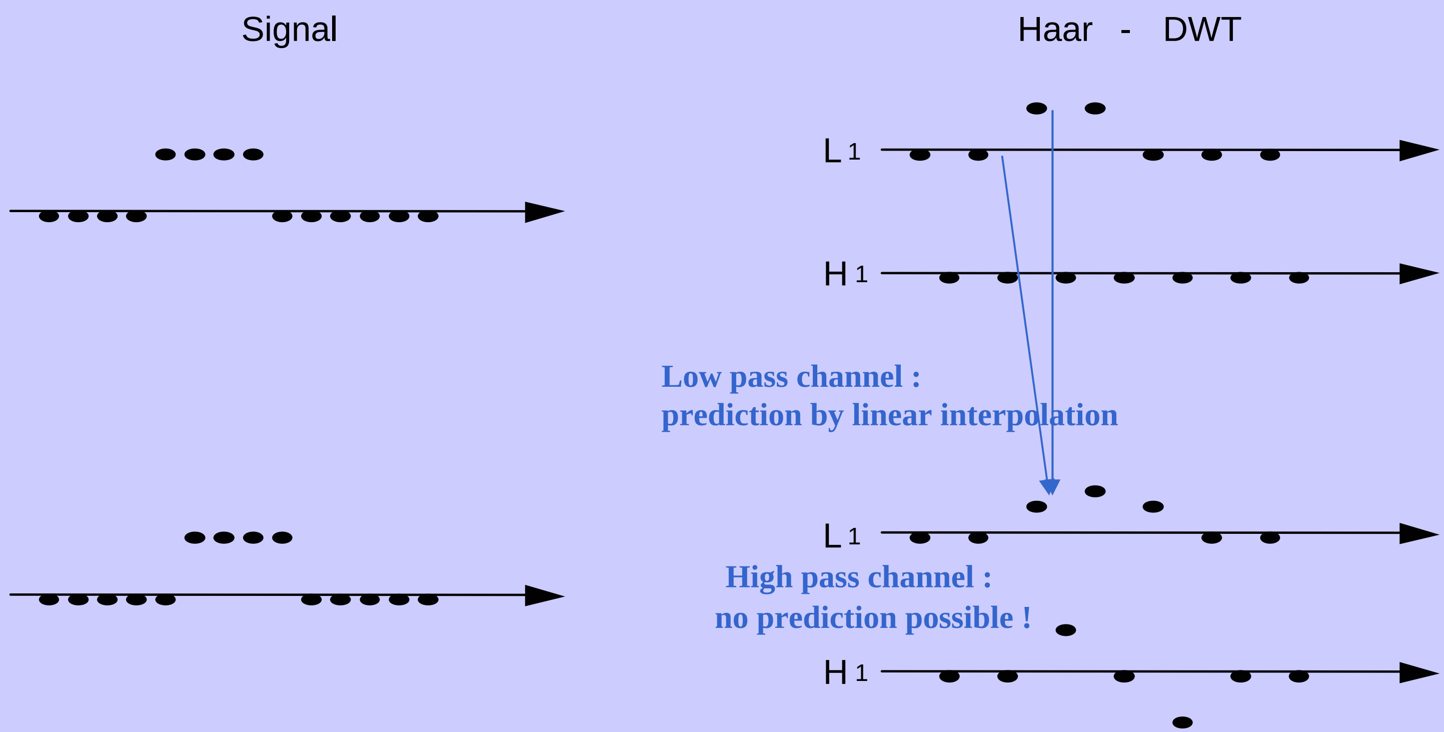


Overcomplete
Motion Compensated Wavelet Coding

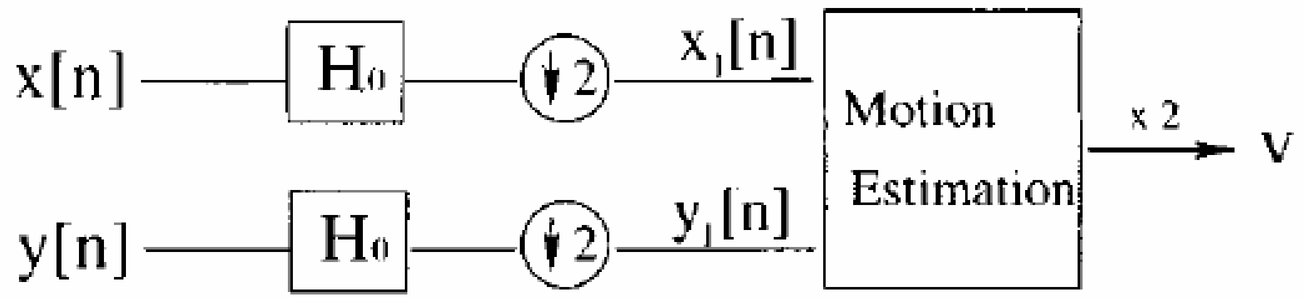
- **Shift-variant property of wavelets**
- **Frame theory - overcomplete wavelets**
- **Low band shifting method**
- **Inband motion compensated temporal filtering**
- **Simulation results**

The Shift Variance Property of Wavelets

- Haar filter output of step edges



The Shift Variance Property of Wavelets



$$X_1(\omega) = \frac{1}{2}\left(H_0\left(\frac{\omega}{2}\right)X\left(\frac{\omega}{2}\right) + H_0\left(\frac{\omega}{2}+\pi\right)X\left(\frac{\omega}{2}+\pi\right)\right)$$

$$Y_1(\omega) = \frac{1}{2}\left(H_0\left(\frac{\omega}{2}\right)Y\left(\frac{\omega}{2}\right) + H_0\left(\frac{\omega}{2}+\pi\right)Y\left(\frac{\omega}{2}+\pi\right)\right)$$

The Shift Variance Property of Wavelets

- Suppose $y[n] = x[n - v]$.
- By substituting $Y(\omega) = X(\omega)e^{-j\omega v}$

$$Y_1(\omega) = \frac{1}{2}\left(H_0\left(\frac{\omega}{2}\right)X\left(\frac{\omega}{2}\right)e^{-j(\omega/2)v} + H_0\left(\frac{\omega}{2}+\pi\right)X\left(\frac{\omega}{2}+\pi\right)e^{-j((\omega/2)+\pi)v}\right).$$

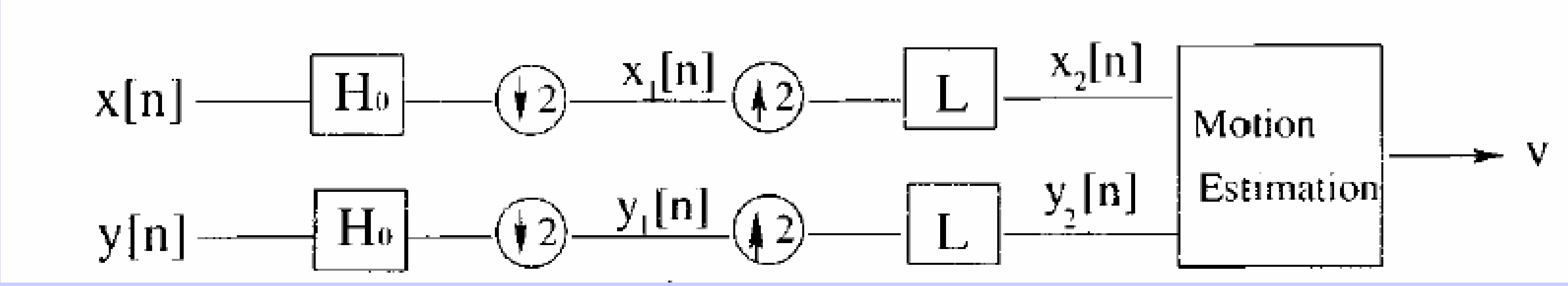
- Hence,

$$Y_1(\omega) - X_1(\omega)e^{-j(\omega/2)v} = \frac{1}{2}H_0\left(\frac{\omega}{2}+\pi\right)X\left(\frac{\omega}{2}+\pi\right)\left(1-e^{-j\pi v}\right)e^{-j\omega/2}$$

Aliasing components (zero only when v = even)

Optimal Aliasing Reduction Filter Approach

- In order to minimize the aliasing in wavelet domain ME/MC (X.Yang, K. Ramchandran, IEEE-TIP, May, 2000)



- L : aliasing reduction filter

Optimal Aliasing Reduction Filter Approach

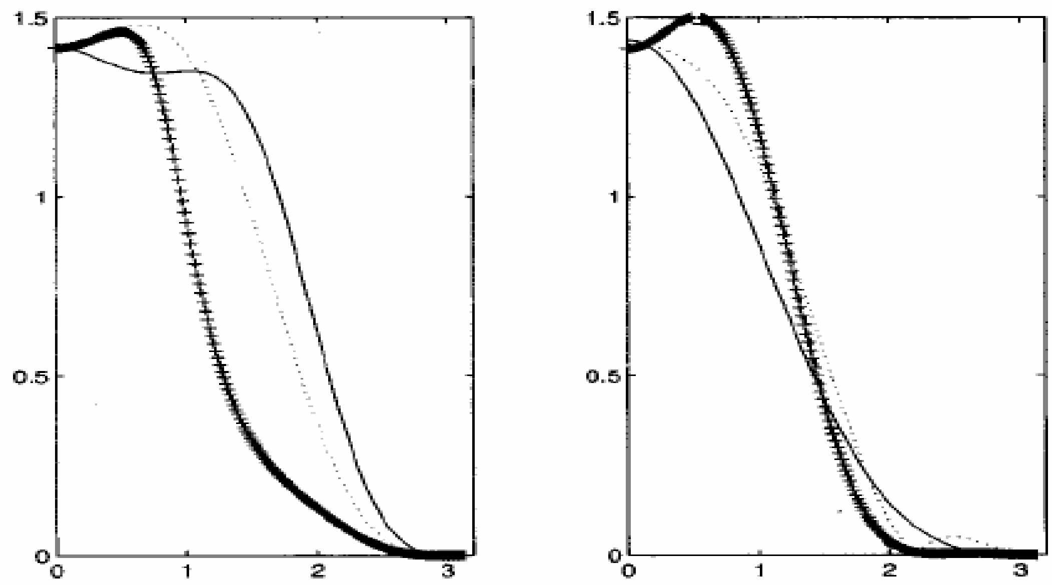


Fig. 3. Various filter responses in our design of $L(\omega)$, for Daubechies 9-7 filter bank and AR(1) input process with $\rho = 0.95$. On the left, the solid curve is $H_0(\omega)$, "..." is $G_0(\omega)$, and "+" is the optimal IIR filter given by (9). On the right, the solid, "...," "-.," and "+" curves are our optimized FIR filters at lengths five, seven, nine, and 11, respectively. Note that the last two curves are almost overlapped.

Optimal Aliasing Reduction Filter Approach

- Still not efficient as motion estimation in spatial domain

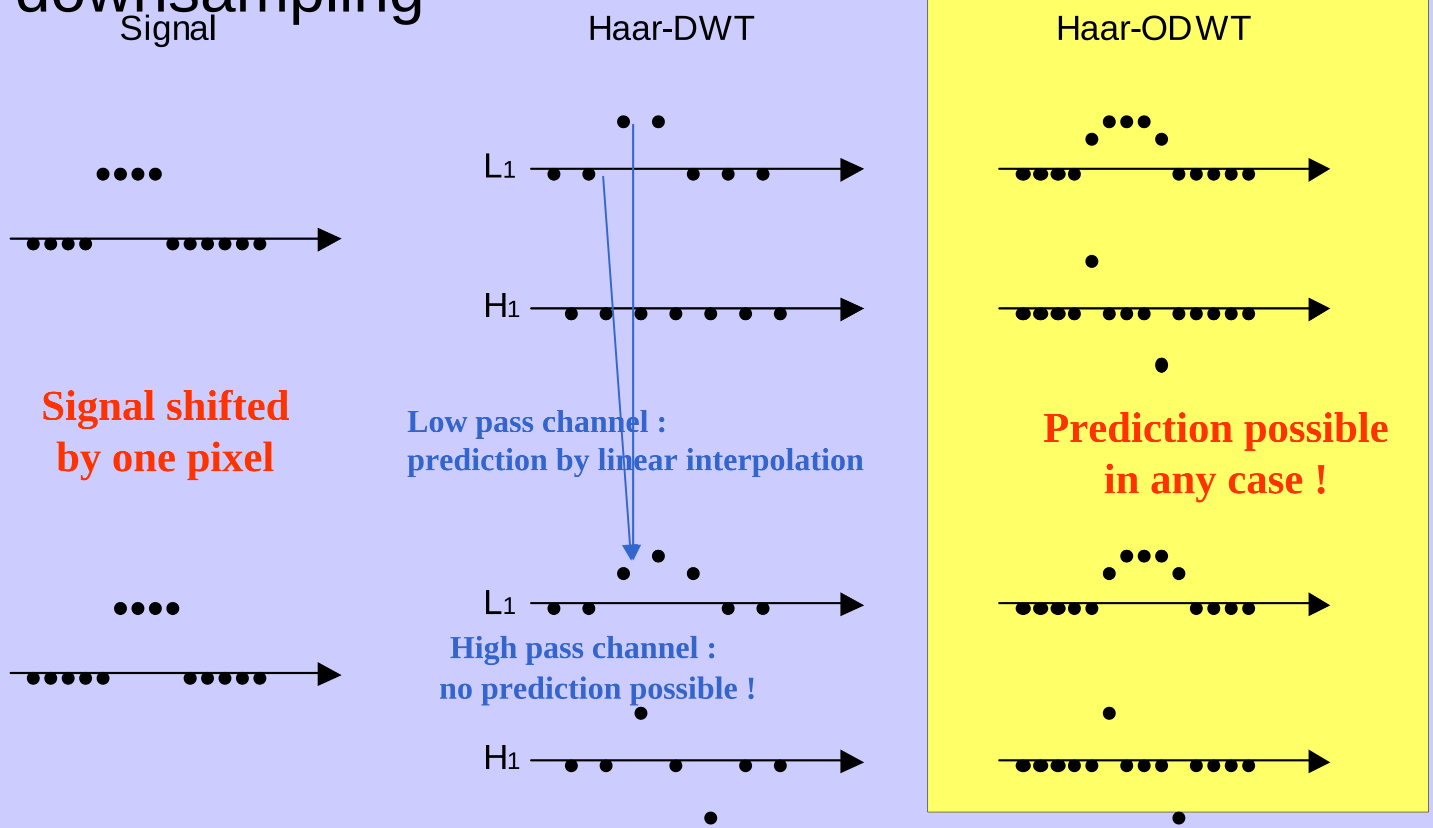
- Any ultimate solution ? → Shift invariant Overcomplete Wavelets

Frame Theory – Overcomplete Wavelets

- Properties of redundant frame
  - Noise reduction
  - More sparse representation → matching pursuit
  - Redundant representation → multiple description coding
  - Shift invariant property
    - Improved motion estimation/compensation in wavelet domain
    - Only motion references need to be overcomplete
    - Texture coding is still in complete wavelet domain

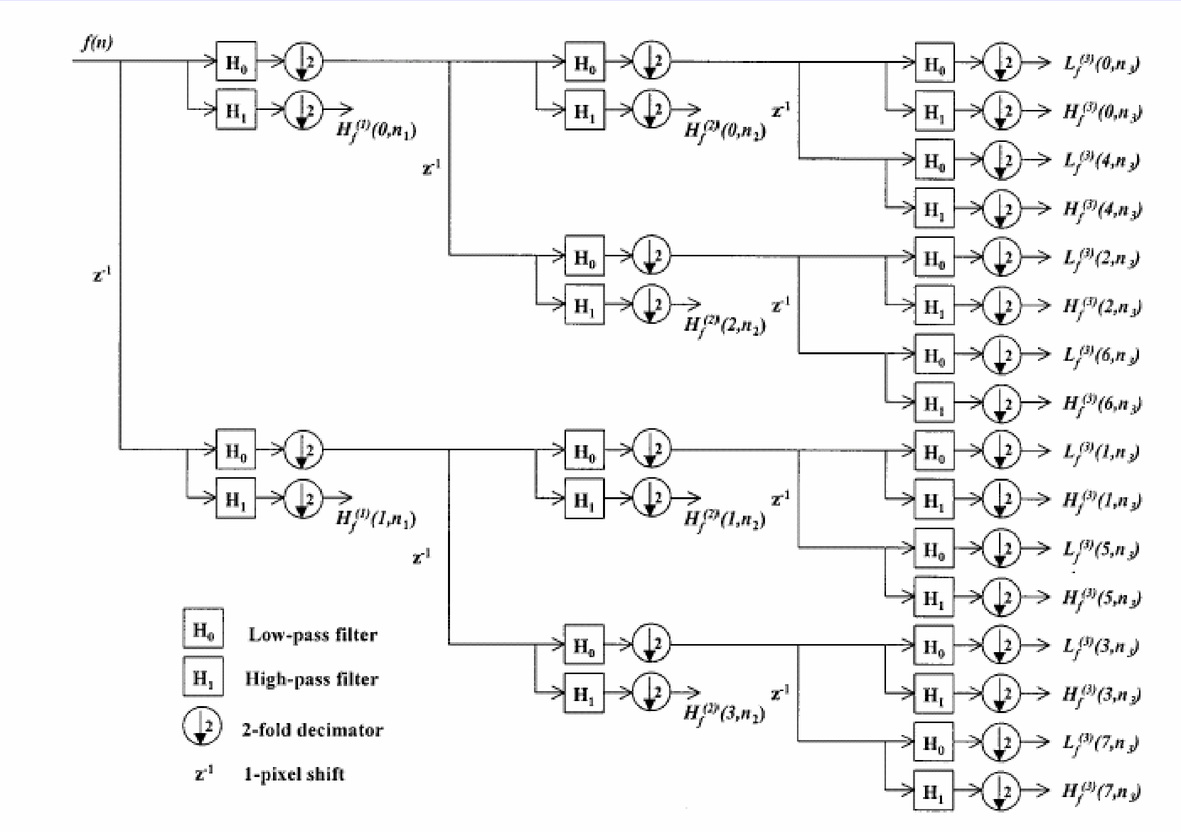
Shift Invariance of Overcomplete Wavelets

- Overcomplete representation without downsampling

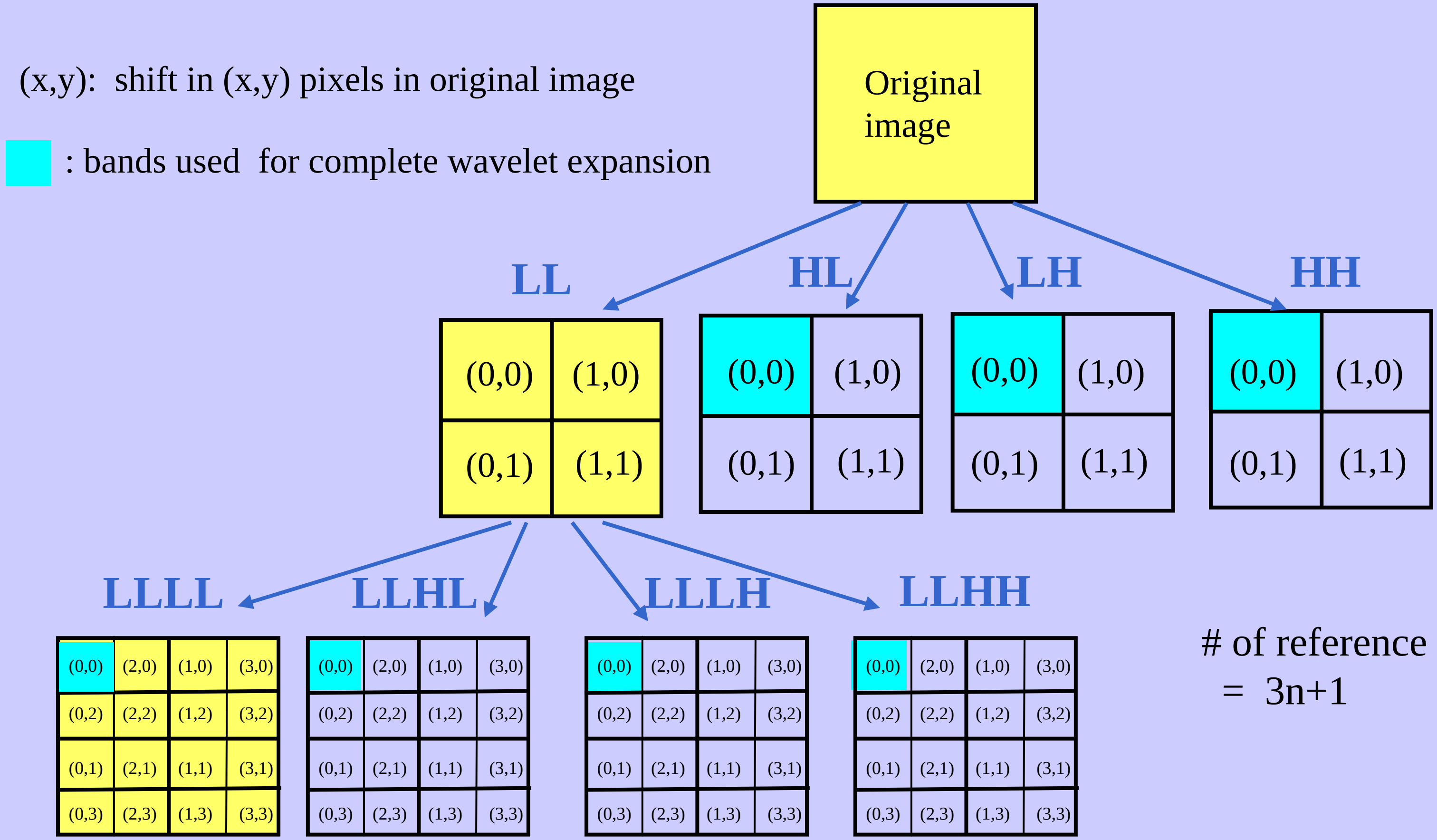


Low-Band-Shift Method

- Optimal way of generating overcomplete wavelet coefficients for every shift

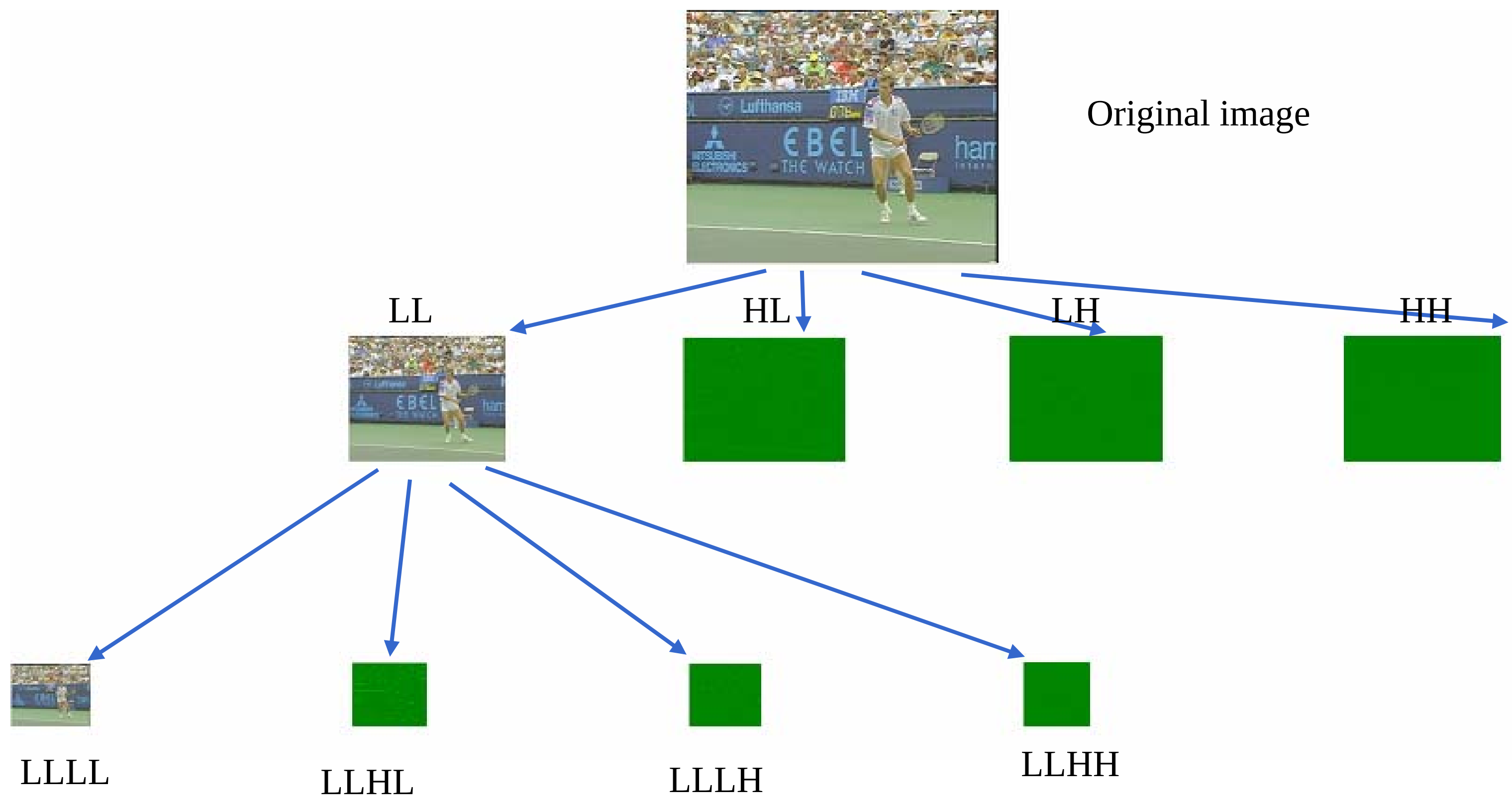


Low Band Shift Method for 2-D

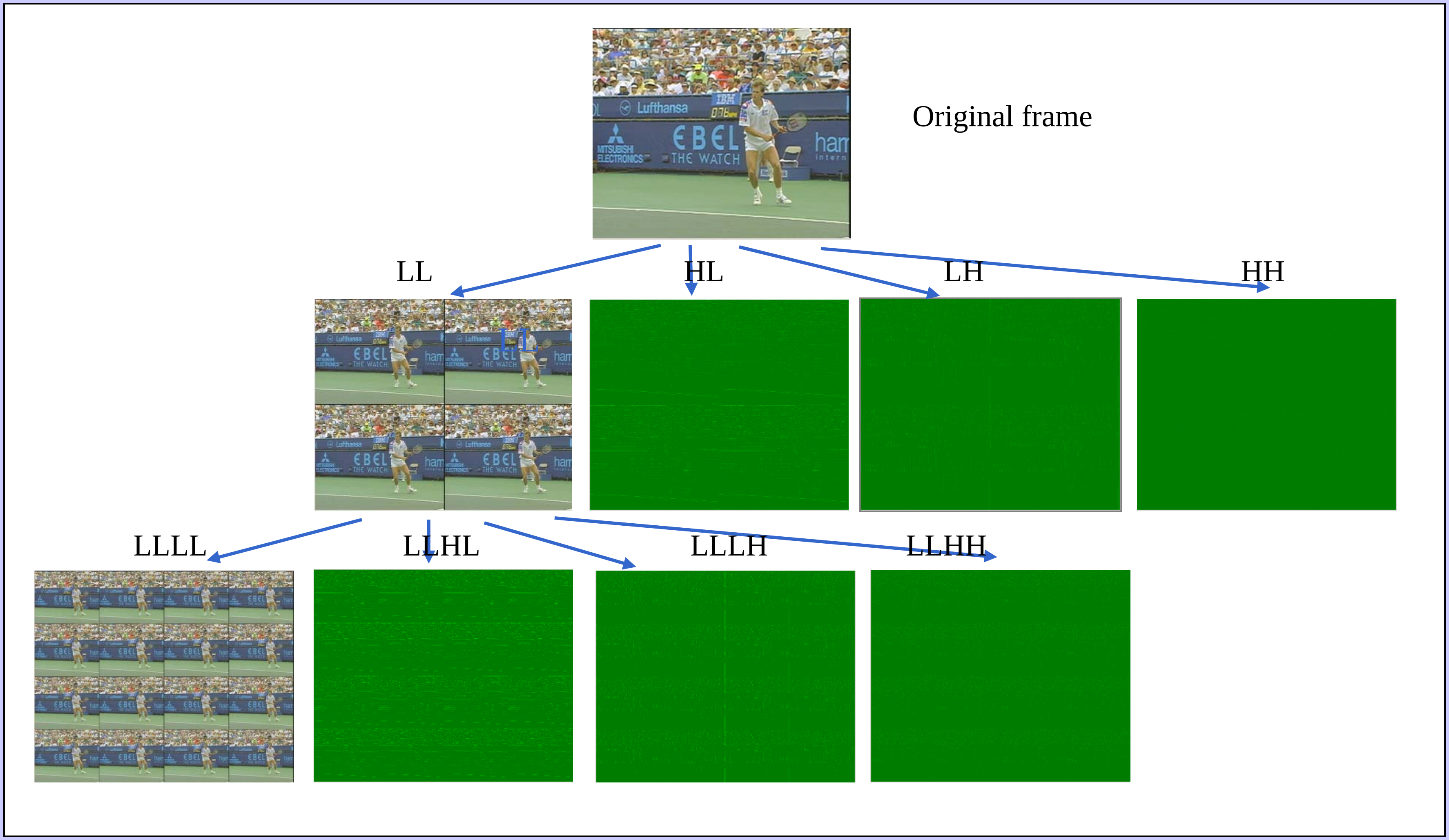


# of reference frames
= 3n+1

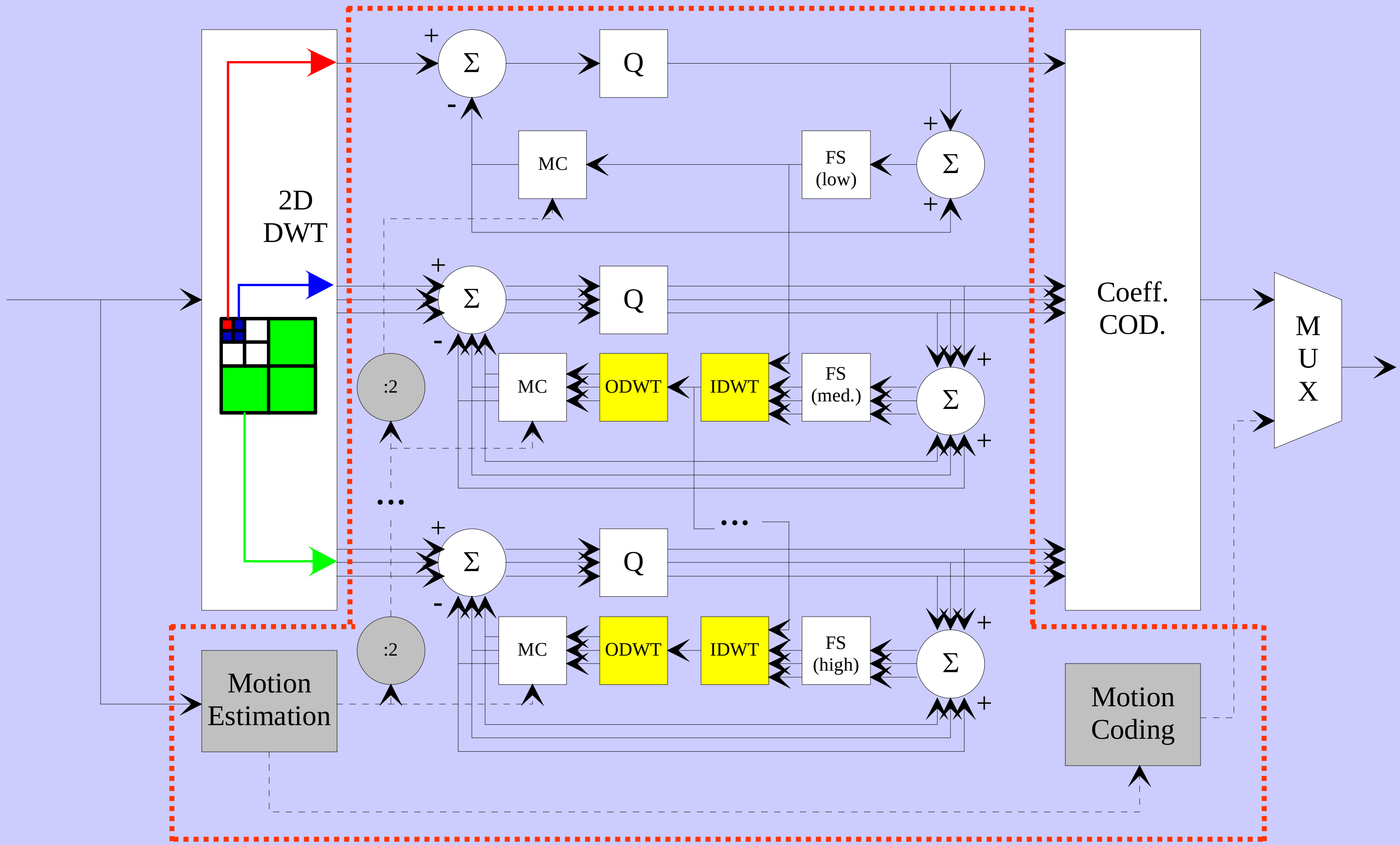
Conventional Wavelet Transform



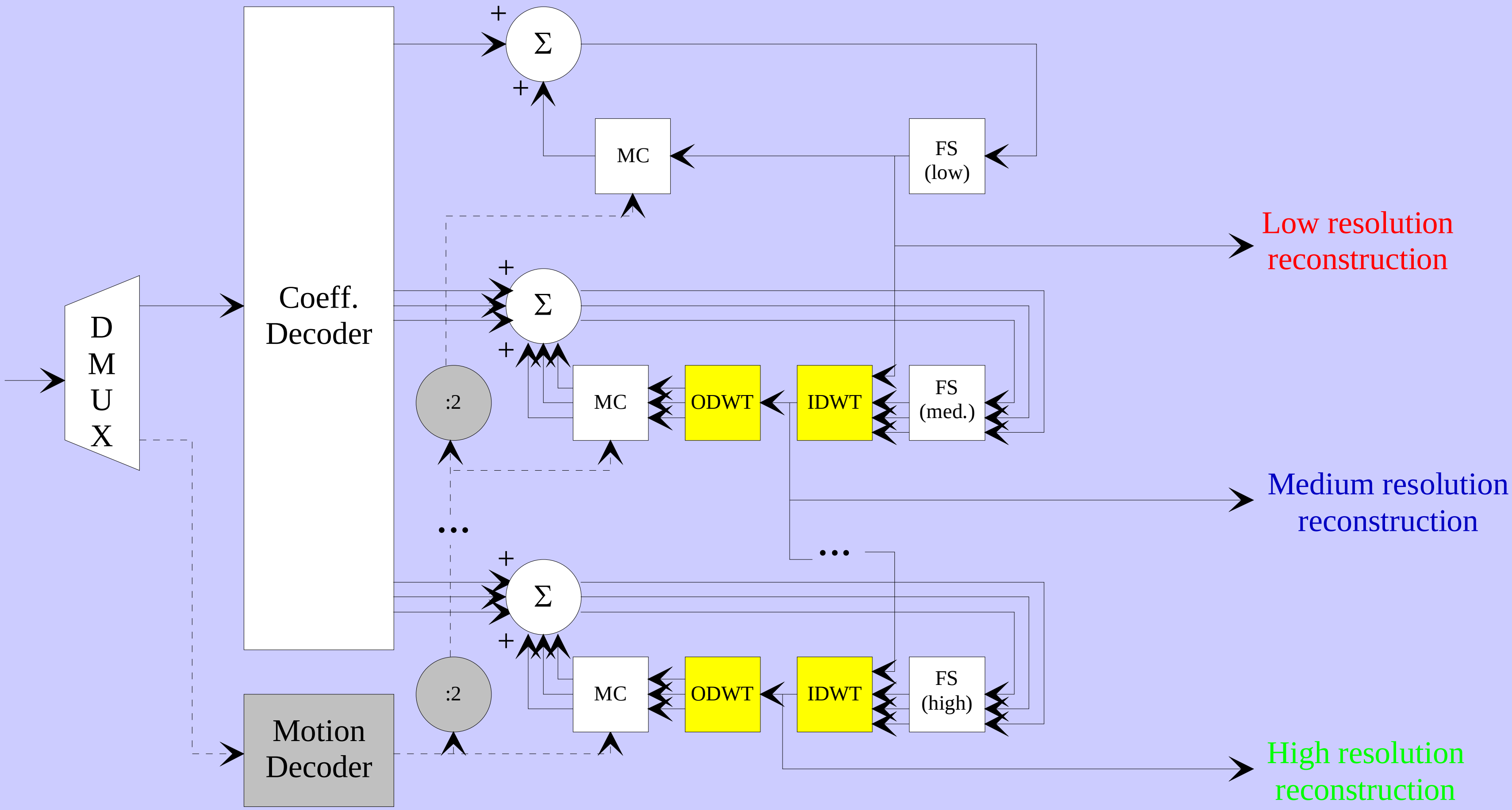
Overcomplete Wavelet Transform by Low-Band Shift Method



Overcomplete Wavelet MC Coding - Coder



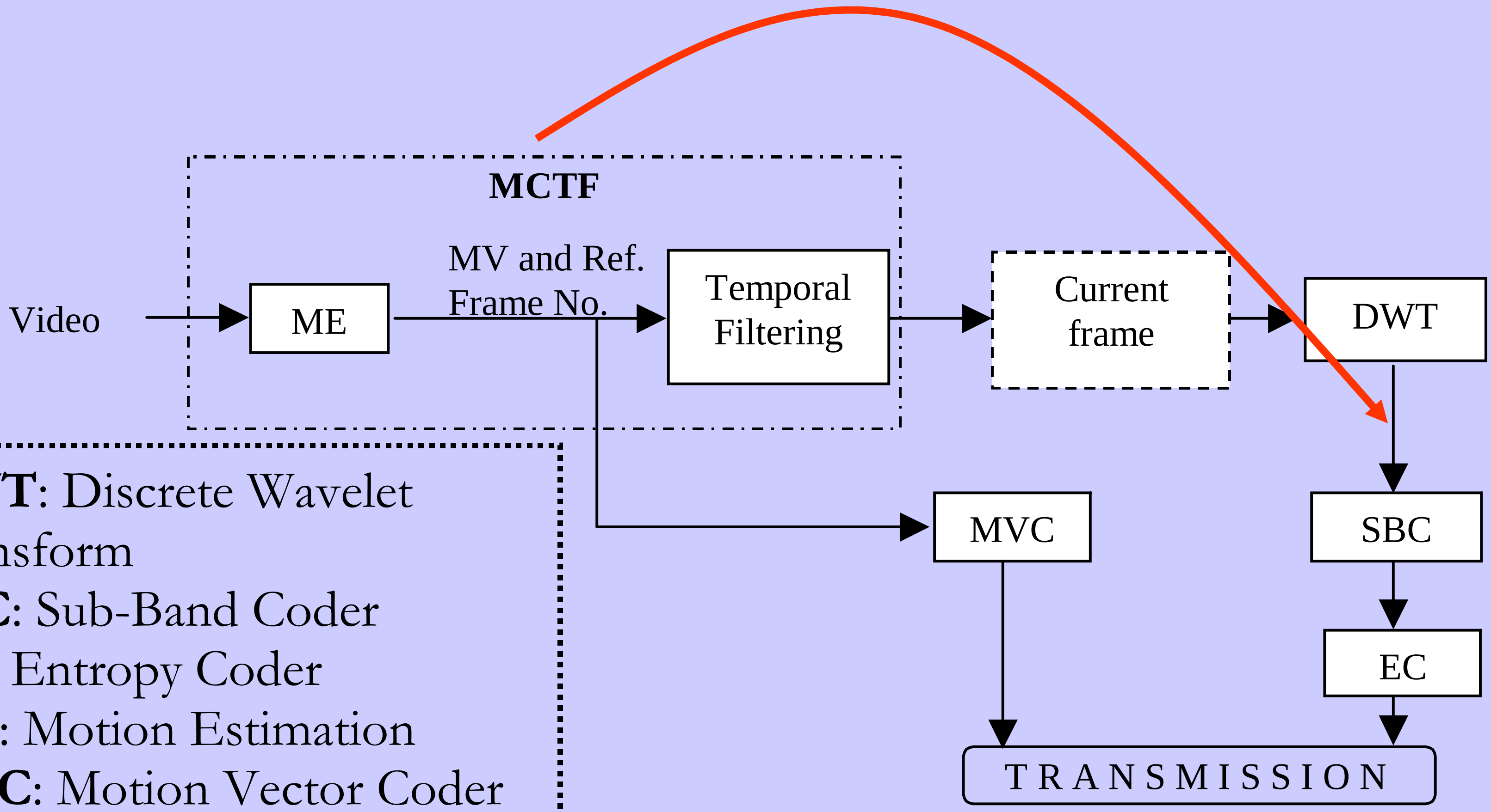
Overcomplete Wavelet MC Coding - Decoder



Overcomplete Wavelet MC Coding

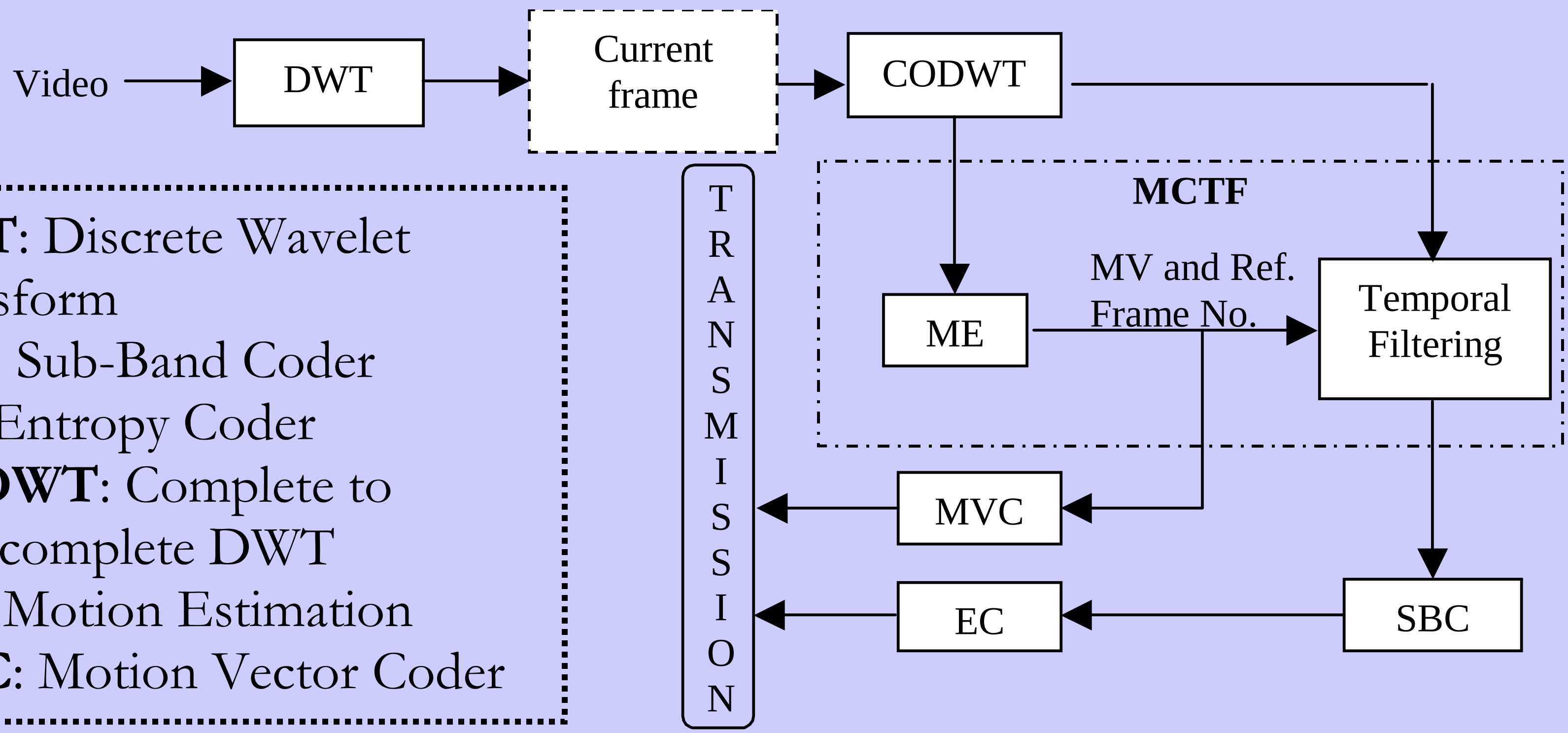
- **ODWT is Wavelet without subsampling**
  - More samples than original, like Pyramid representation
- **Allows Wavelet domain MC for high frequency components**
  - signal does not bear frequency-inversion alias component
- **Still only necessary to encode critically sampled coefficients**
  - Overcomplete transform coefficients can be generated locally within the decoder
- **Still does not resolve the drift in SNR scalability**
  - may be solved by multiple loops in each wavelet band
- Solution: In-Band MCTF (IBMCTF)

Spatial-Domain MCTF (SDMCTF)



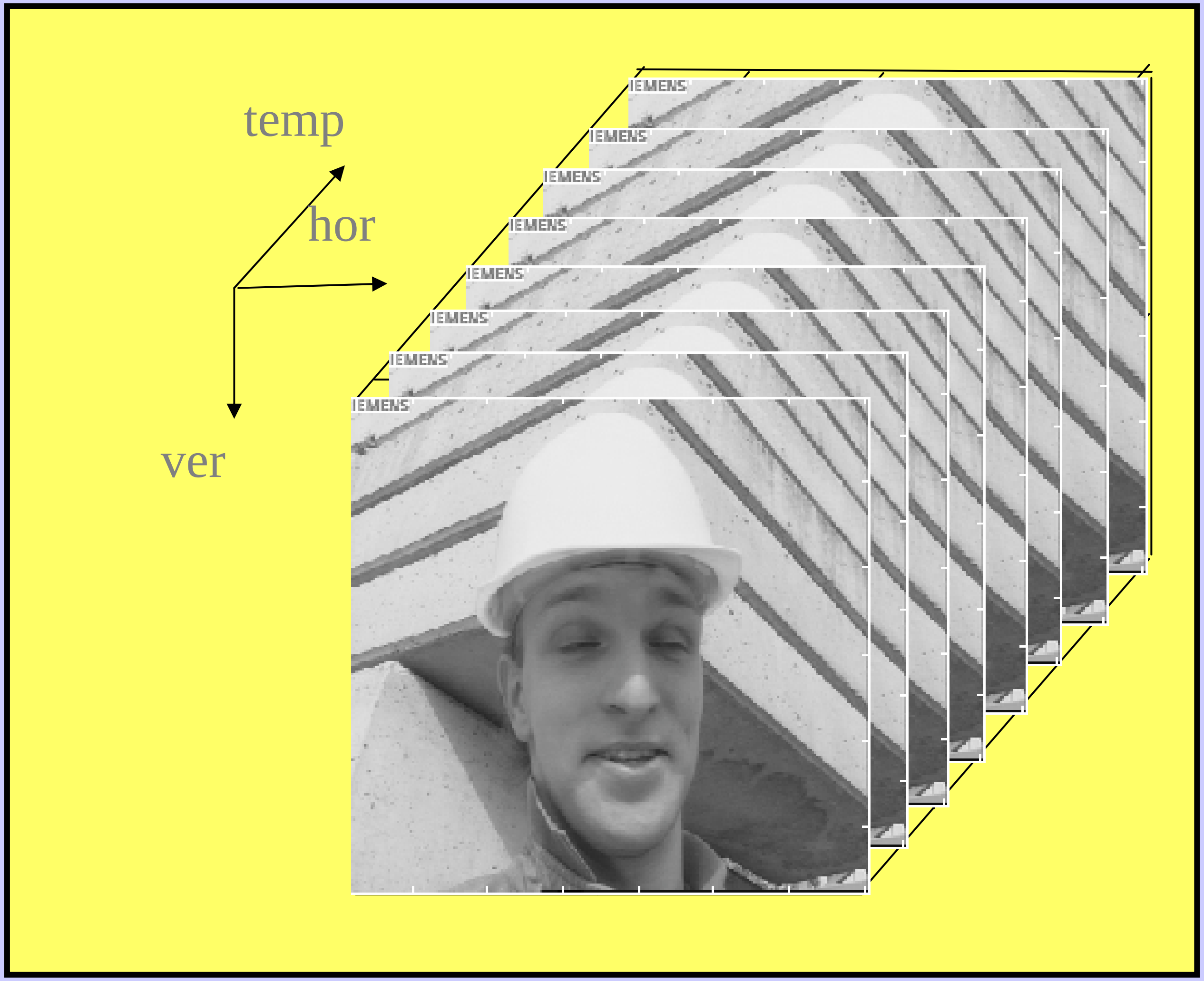
DWT: Discrete Wavelet Transform
SBC: Sub-Band Coder
EC: Entropy Coder
ME: Motion Estimation
MVC: Motion Vector Coder

In-band MCTF (IBMCTF)

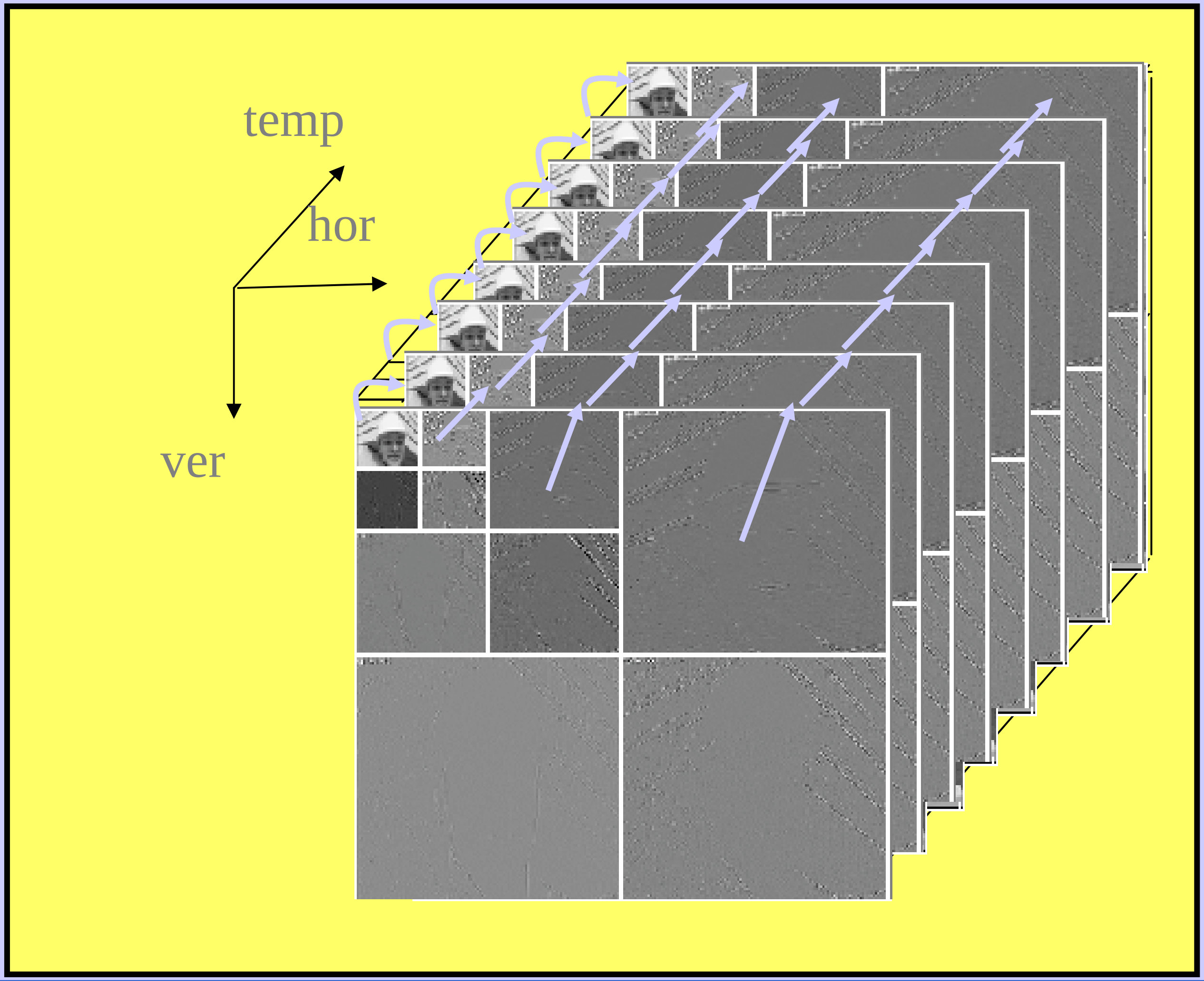


DWT: Discrete Wavelet Transform
SBC: Sub-Band Coder
EC: Entropy Coder
CODWT: Complete to Overcomplete DWT
ME: Motion Estimation
MVC: Motion Vector Coder

IBMCTF: concept

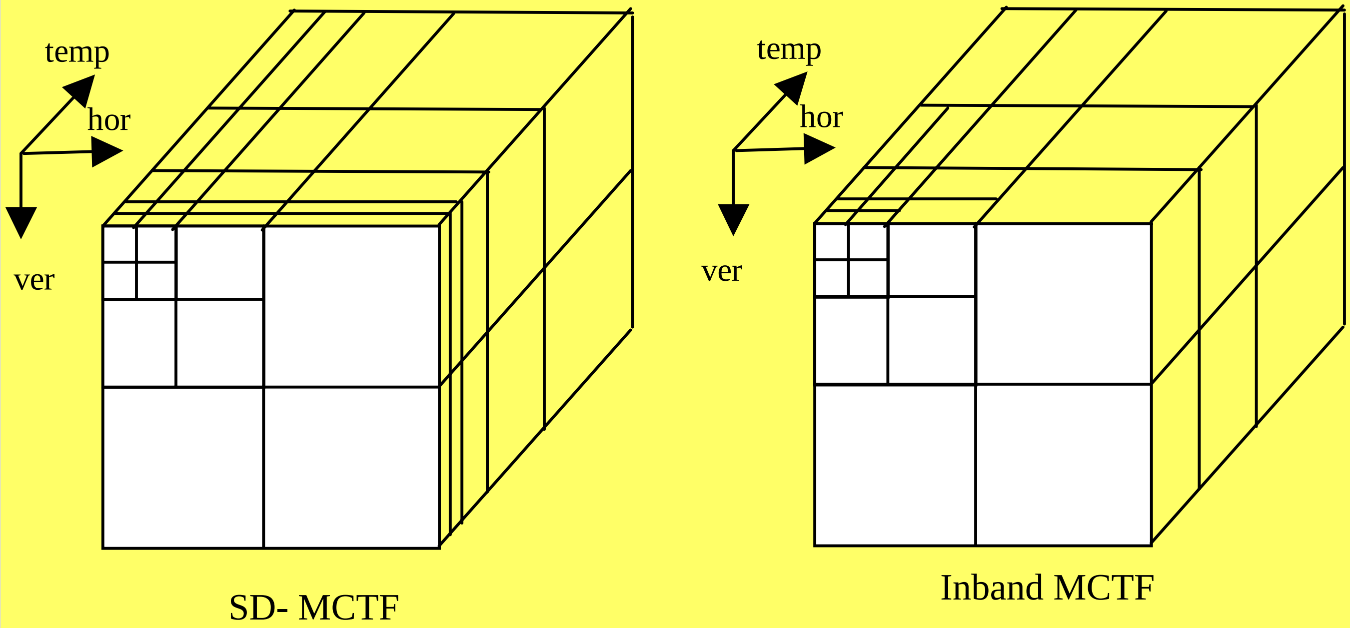


IBMCTF Wavelet Video

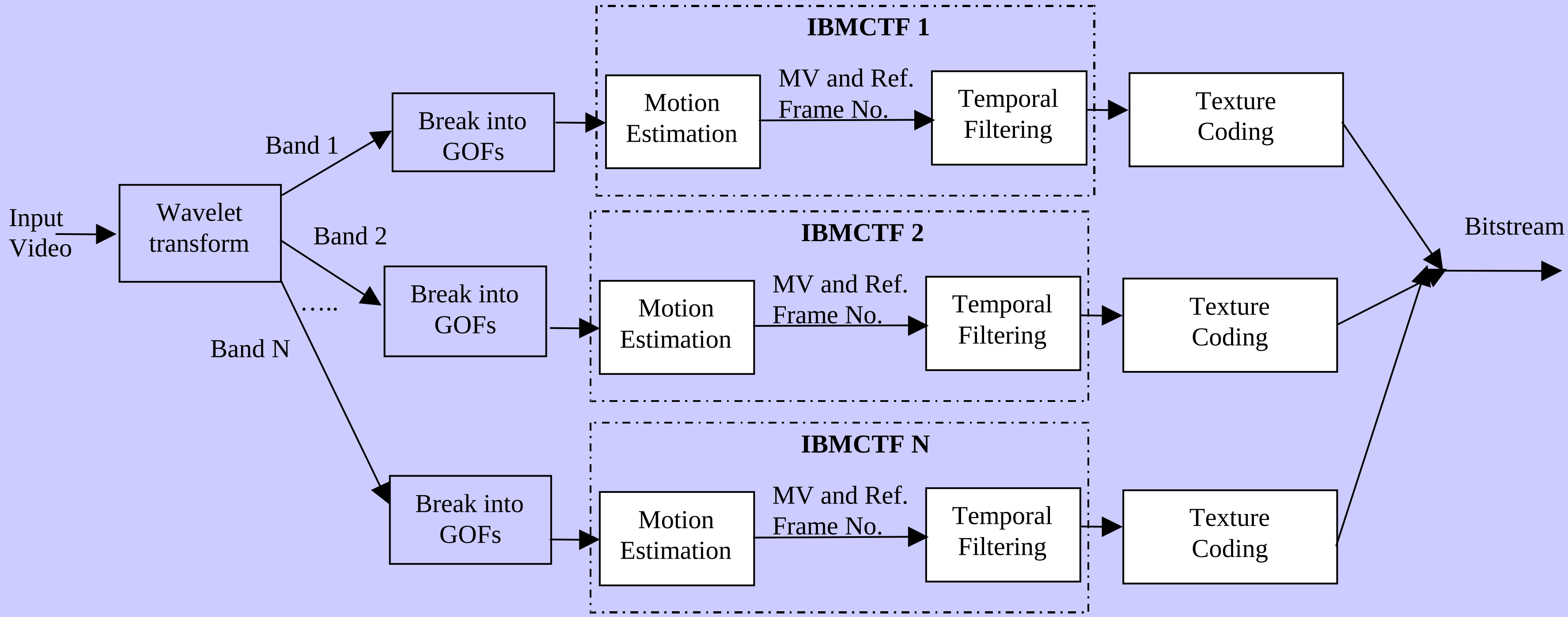


For efficient IBMCTF, ME should be performed in overcomplete wavelet domain

3-D decomposition structure



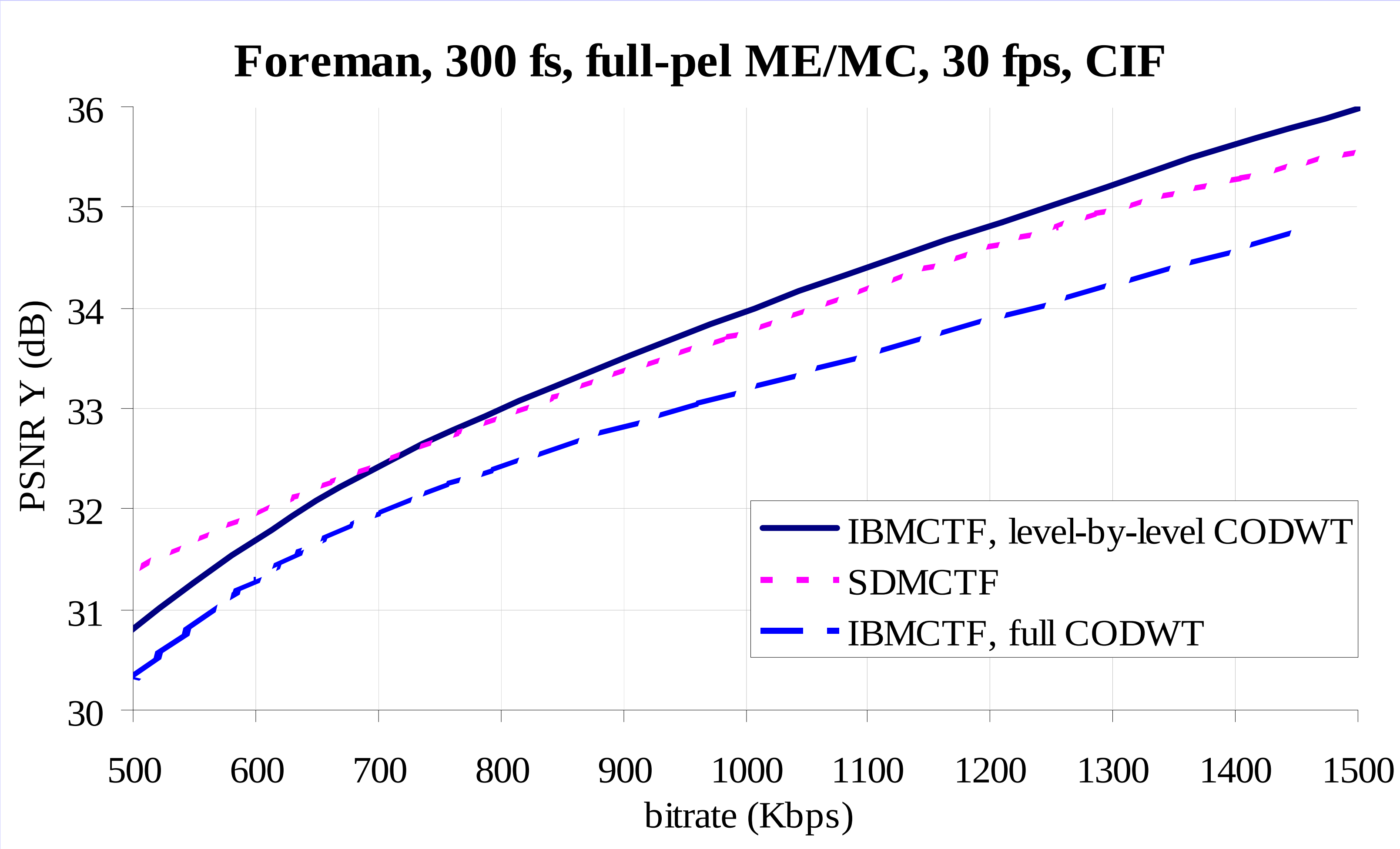
Block diagram of IBMCTF coder



IBMCTF coding

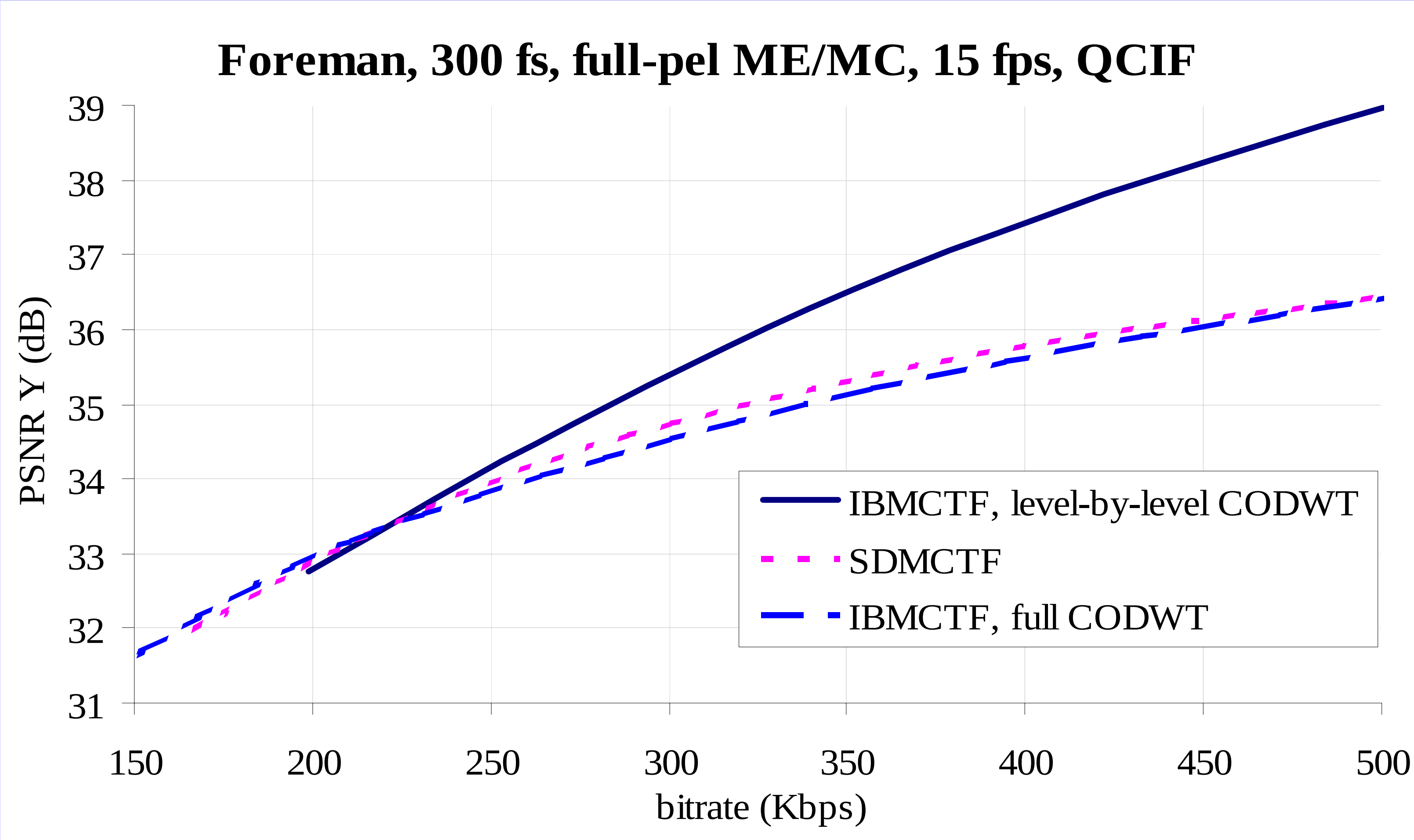
- **Allows Wavelet domain MC using shift-invariant overcomplete waveltes by Low-Band Shift method**
- **Still only necessary to encode critically sampled coefficients**
- **Advantages in spatial scalability**
- **Resolve the drift in SNR scalability**
- **Adaptive processing for each subband**
  - **Different ME accuracy, interpolation filter, temporal filter taps, etc.**
- **Very general framework which can be combined with other existing techniques (intra mode, UMCTF, etc)**

Results



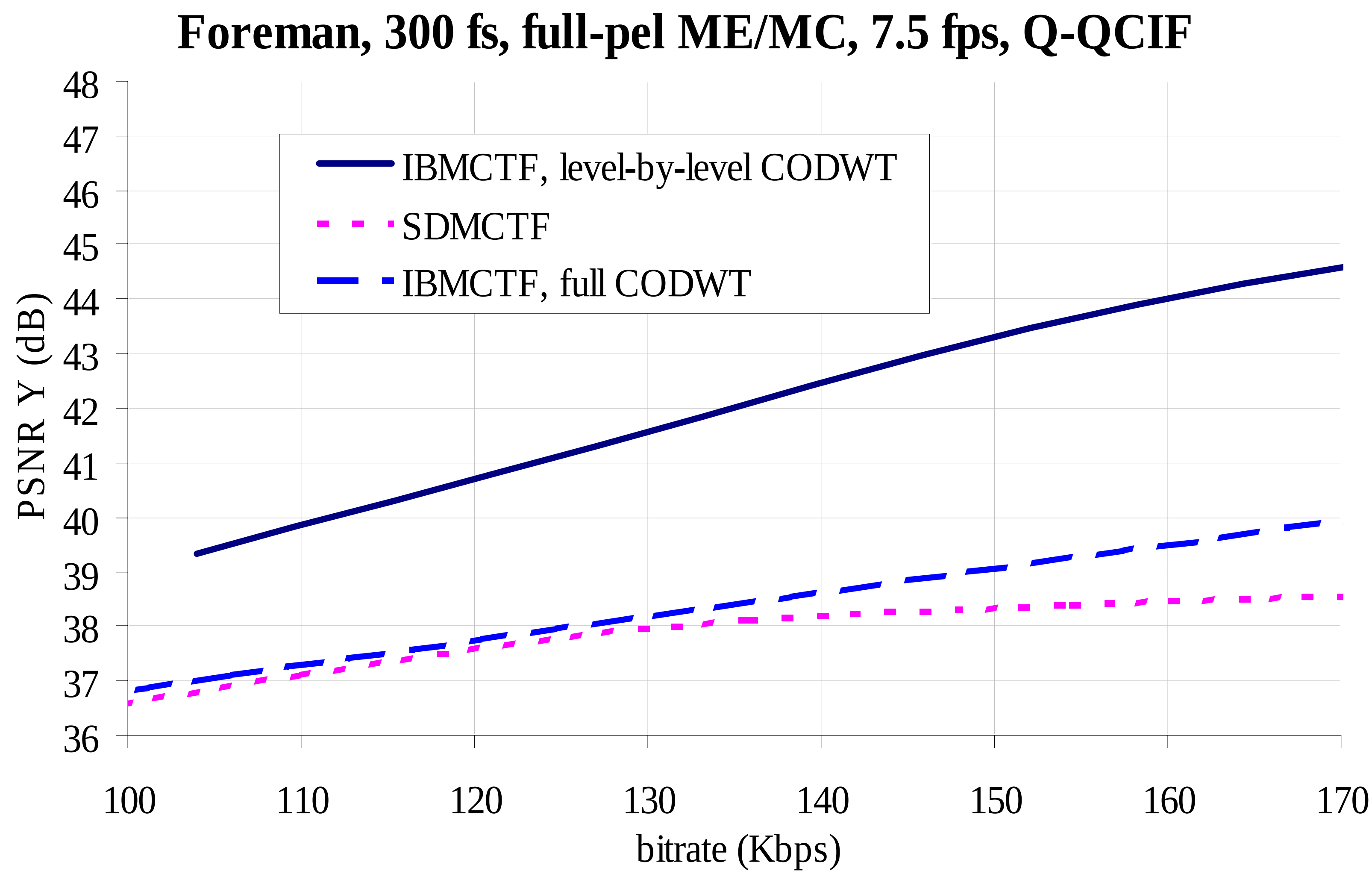
Foreman, 300 fs, full-pel ME/MC, 30 fps, CIF
PSNR Y (dB)
36
35
34
33
32
31
30
500
600
700
800
900
1000
1100
1200
1300
1400
1500
bitrate (Kbps)
IBMCTF, level-by-level CODWT
SDMCTF
IBMCTF, full CODWT

Results



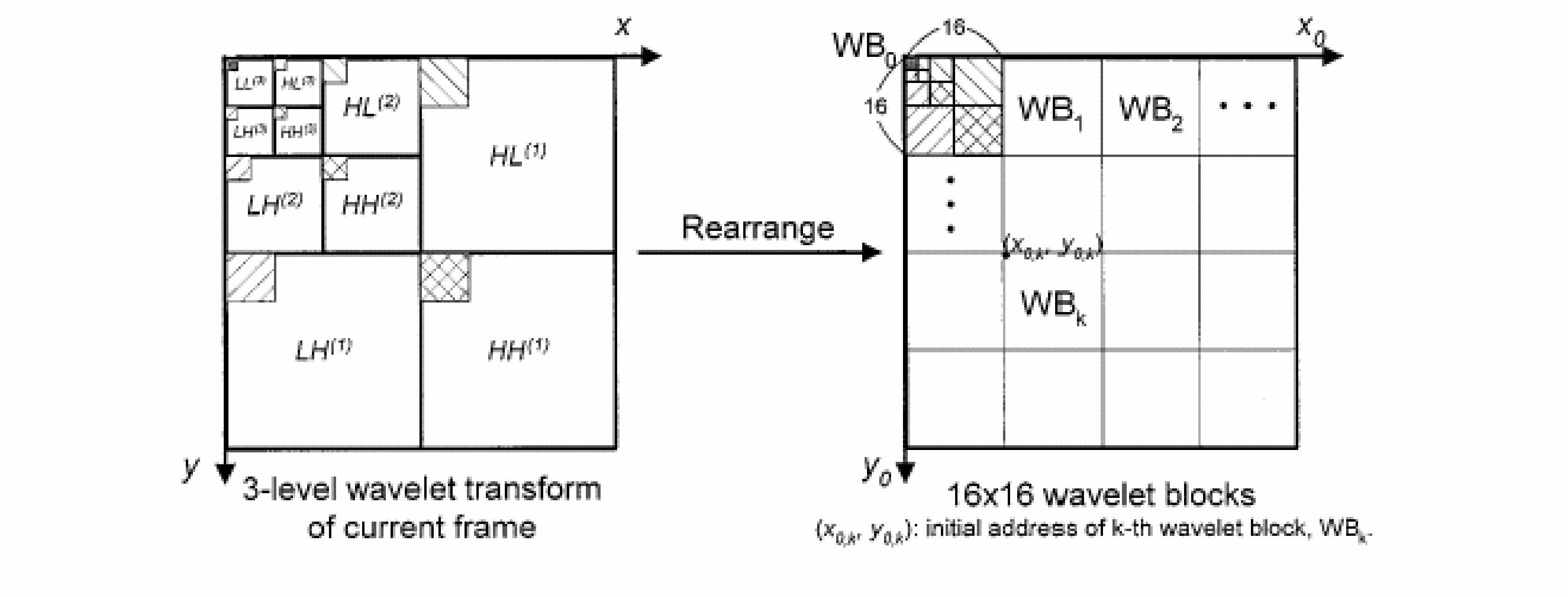
Foreman, 300 fs, full-pel ME/MC, 15 fps, QCIF
PSNR Y (dB)
39
38
37
36
35
34
33
32
31
150
200
250
300
350
400
450
500
bitrate (Kbps)
IBMCTF, level-by-level CODWT
SDMCTF
IBMCTF, full CODWT

Results



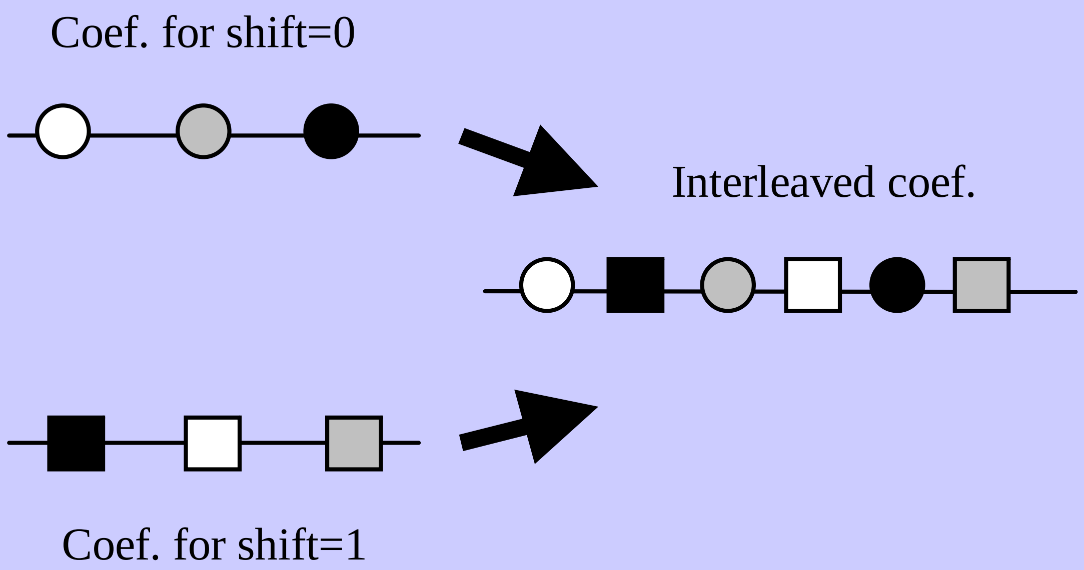
Foreman, 300 fs, full-pel ME/MC, 7.5 fps, Q-QCIF
IBMCTF, level-by-level CODWT
SDMCTF
IBMCTF, full CODWT
PSNR Y (dB)
48
47
46
45
44
43
42
41
40
39
38
37
36
100
110
120
130
140
150
160
170
bitrate (Kbps)

Generation of Wavelet Blocks

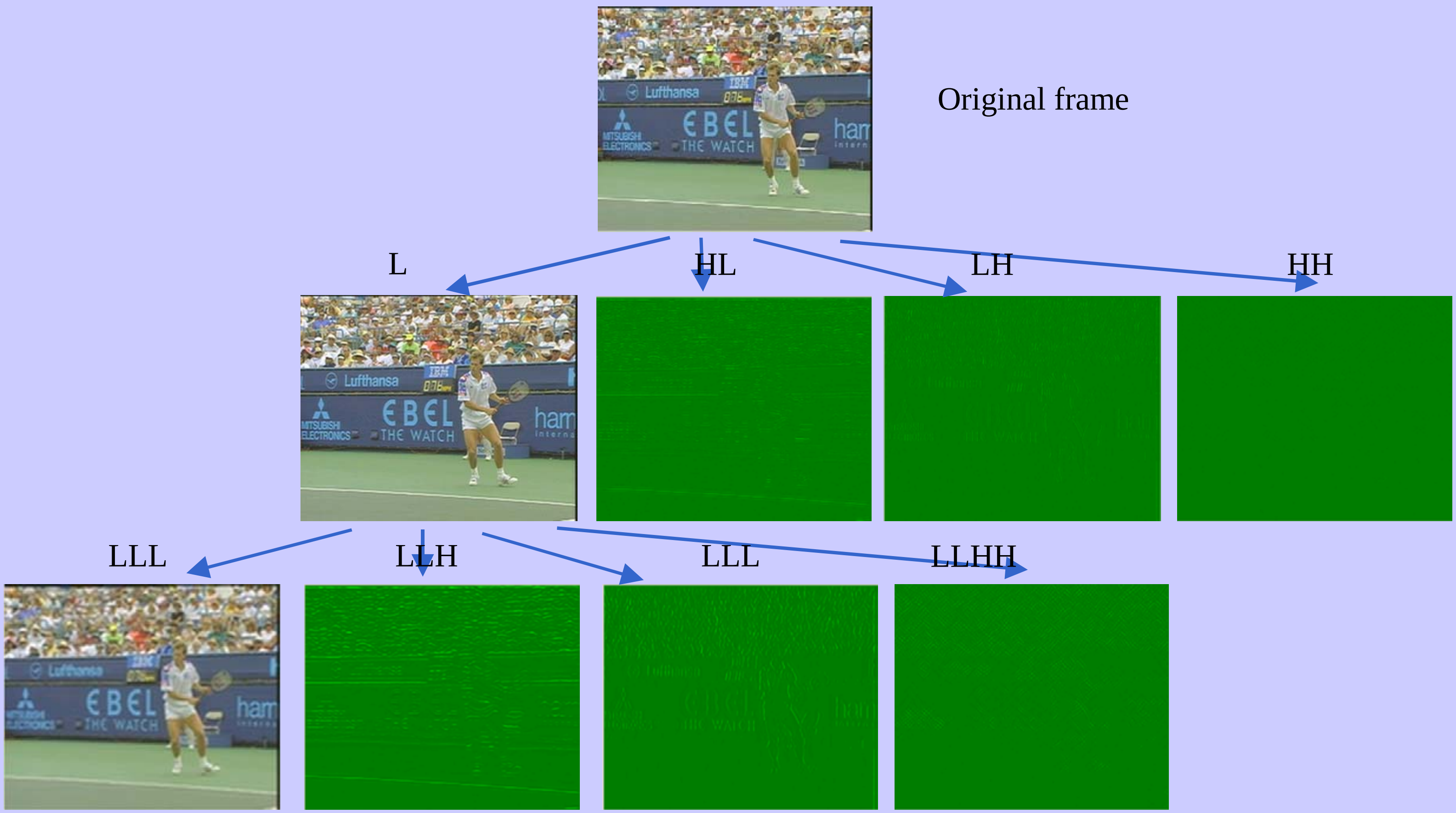
- Wavelet block provides a direct association between the wavelet coefficients and what they represent spatially
- 

- No motion vector overhead because the number of the motion vector to be coded is the same as that of SDMCTF
- Perfectly aligned with tree structure entropy coder
  - Entropy based motion estimation criterion !!

Proposed interleaving of overcomplete wavelet coefficients



Overcomplete Wavelet Transform with Interleaving



Advantages of Interleaving

- Interleaving algorithm enables *optimal* sub-pixel accuracy motion estimation and compensation in IBMCTF
- By interleaving, any existing ME module (HVSBM, FSBM, Intra Mode, etc) with any fractional pel accuracy can be used
- Can be easily used for MCTF framework with any fractional pel accuracy using lifting structure

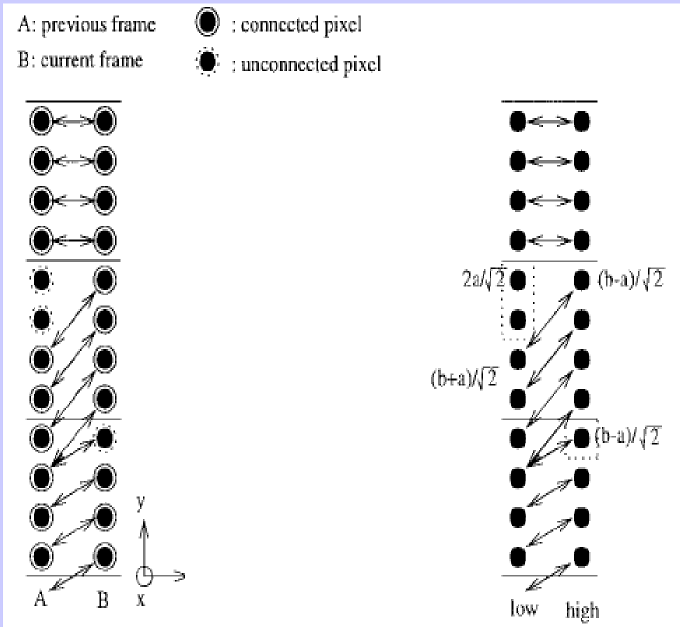
3-D Lifting Structure for IBMCTF

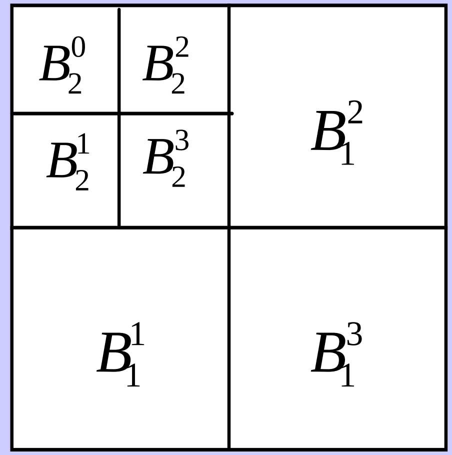
Direct extension of SD-MCTF lifting to IBMCTF:

$$H_j^i[m,n] = \left(B_j^i[m,n] - \tilde{A}_j^i\left[m - d_j^i(m), n - d_j^i(n)\right]\right)/\sqrt{2} \quad ,i = 0,\ldots,3$$

Interpolation operation for frame A_j^i is not optimal
(no cross-phase dependencies incorporated)

3-D Lifting Structure for IBMCTF

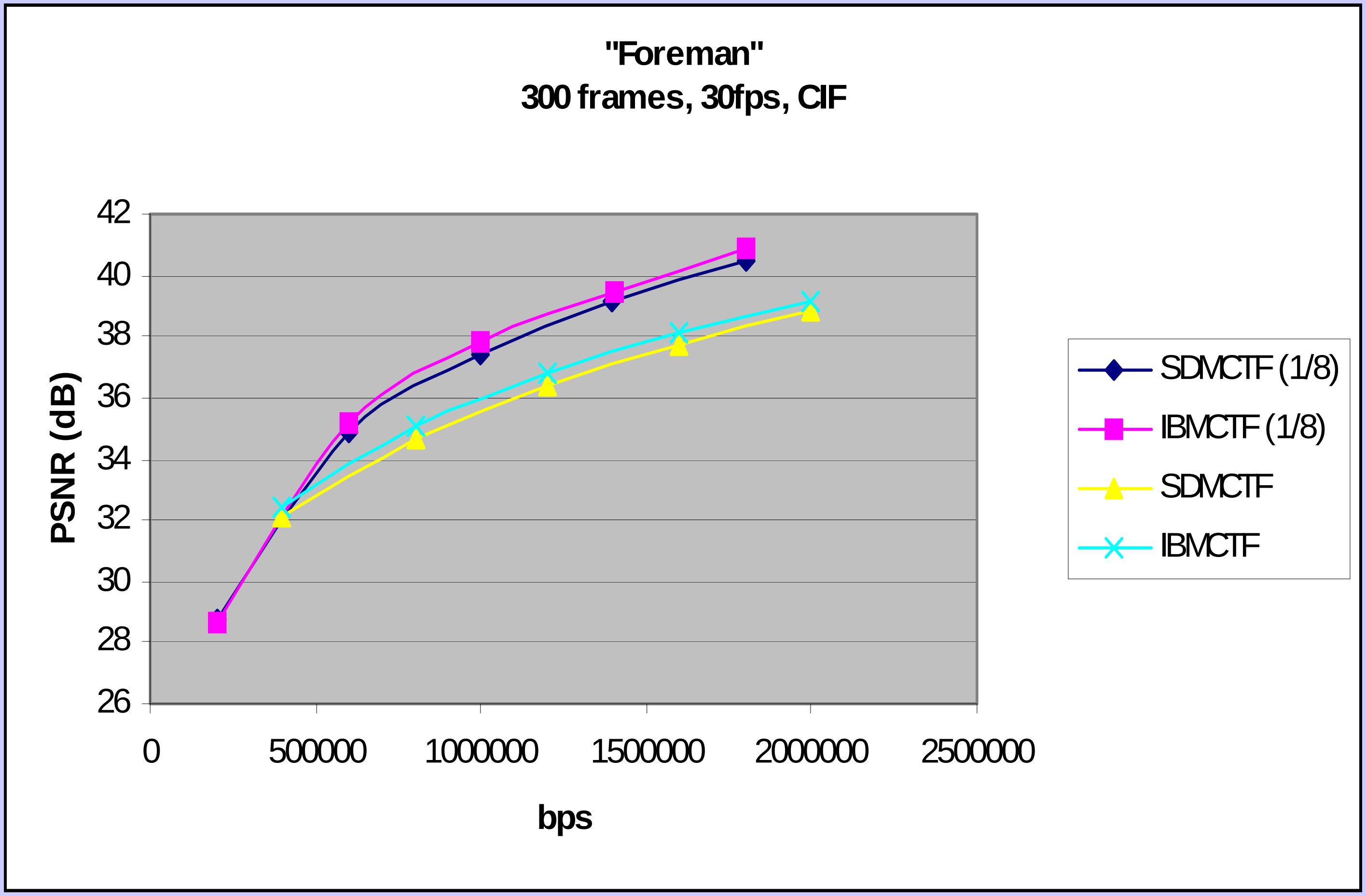




$$H_j^i[m,n] = \left(B_j^i[m,n] - LBS\_\tilde{A}_j^i\left[2^j m - d_m, 2^j n - d_n\right]\right) \big/ \sqrt{2} \quad , i = 0,\ldots,3$$

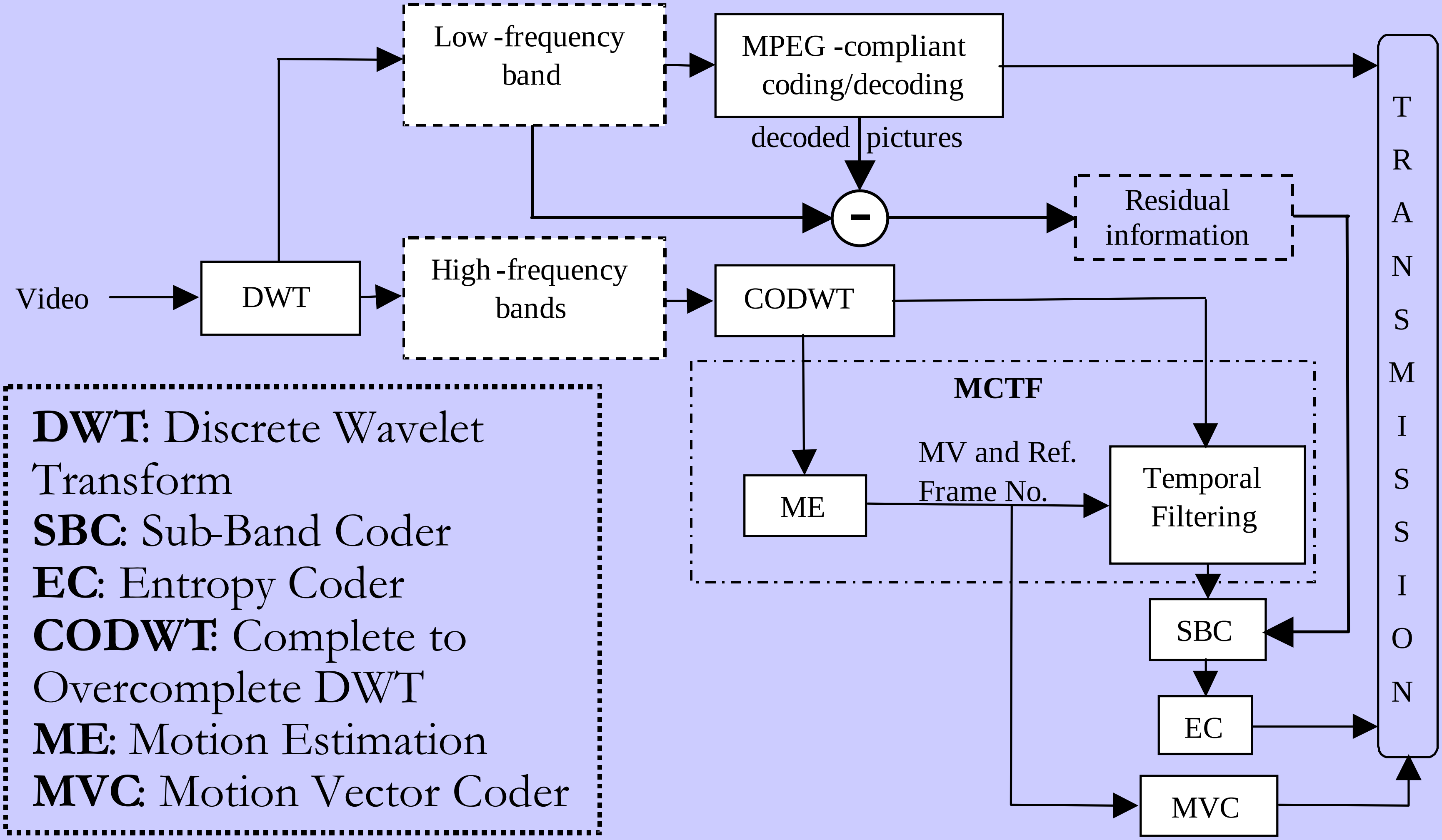
$$L_j^i\left[m - \bar{d}_j^i(m), n - \bar{d}_j^i(n)\right] = LBS\_\tilde{H}_j^i\left[2^j m - \bar{d}_m + d_m, n - \bar{d}_n + d_n\right]$$
$$+ \sqrt{2} A_j^i\left[m - \bar{d}_j^i(m), n - \bar{d}_j^i(n)\right] \qquad i = 0,..,3$$

Results



"Foreman"
300 frames, 30fps, CIF
PSNR (dB)
42
40
38
36
34
32
30
28
26
0
500000
1000000
1500000
2000000
2500000
bps
SDMCTF (1/8)
IBMCTF (1/8)
SDMCTF
IBMCTF

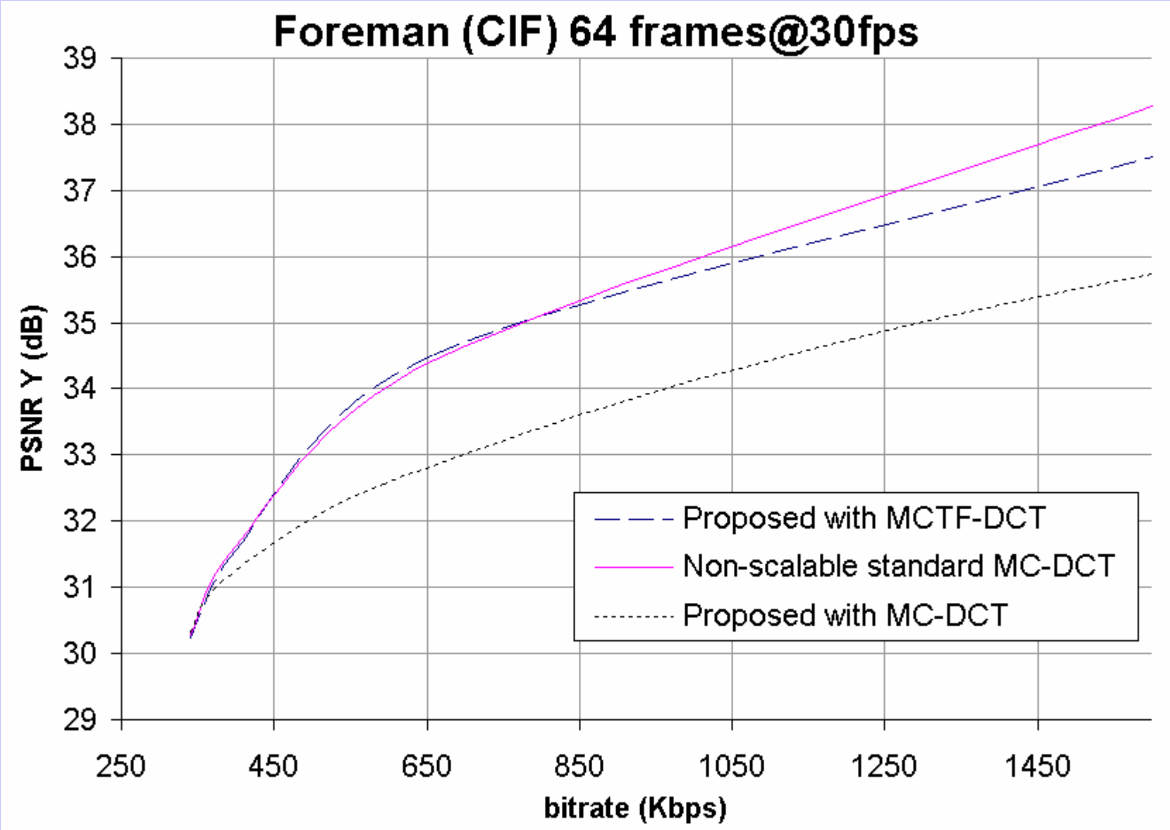
Overcomplete wavelet coding using standard-compliant DCT base-layers



DWT: Discrete Wavelet Transform
SBC: Sub-Band Coder
EC: Entropy Coder
CODWT: Complete to Overcomplete DWT
ME: Motion Estimation
MVC: Motion Vector Coder

Proposed by Andreopoulos, van der Schaar, et al – ICIP 2003

Results


Foreman (CIF) 64 frames@30fps
PSNR Y (dB)
39
38
37
36
35
34
33
32
31
30
29
250
450
650
850
1050
1250
1450
bitrate (Kbps)
Proposed with MCTF-DCT
Non-scalable standard MC-DCT
Proposed with MC-DCT

Current Status in MPEG Standardization

MPEG's Scalable Coding History

- **Development of scalable video coding** solutions has a **long history in MPEG**, starting from MPEG-2
  - Spatial, temporal and SNR scalability with at most 3 levels
  - MPEG-4 Fine granularity scalability
- So far, all standardized solutions have shown **deficiency in coding performance which is** mainly due to **recursive MC structure**
  - Drift occurs when not all information is available
  - Drift-free structures are less coding efficient

MPEG's Interframe Wavelet Coding Exploration

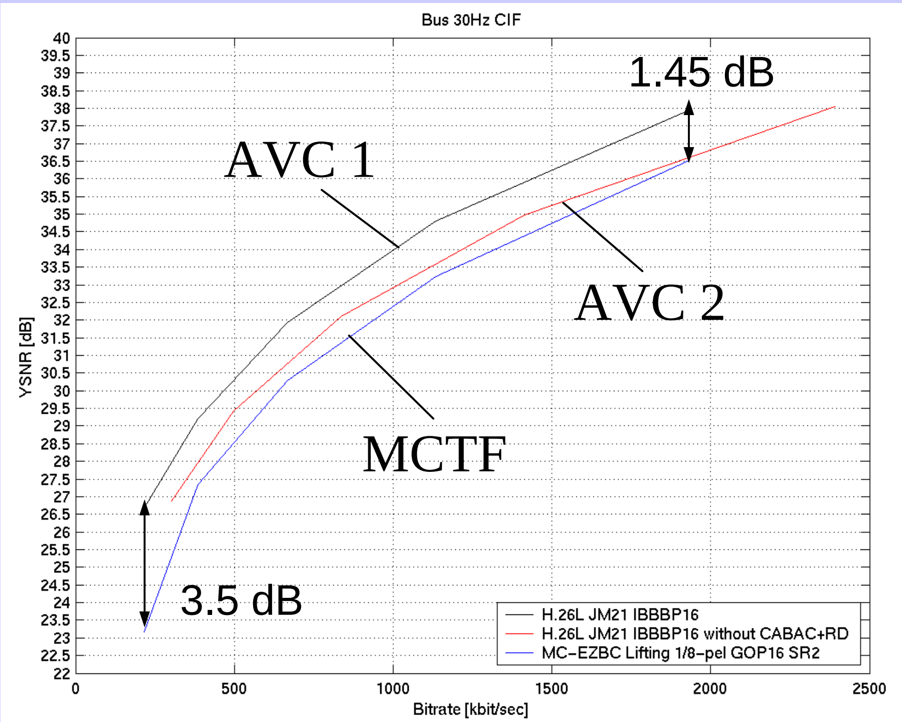
- New embedded wavelet solutions were proposed In the Digital Cinema Call for Proposals and in the Call for Proposals on improved coding efficiency (both due July 2001)
- At Pattaya meeting (Dec. 2001), MPEG started an Adhoc Group to explore Interframe Wavelet Coding
- Different methods were investigated
  - **MC prediction with intraframe (2D) wavelet**
  - **In-band MC prediction based on overcomplete 2D wavelet decomposition**
  - **3D (spatio/temporal) wavelet coding based on MCTF**
- 3D (t+2D, 2D+t) showed most promising, providing excellent coding efficiency while being fully scalable in temporal, spatial and quality resolution
- Experimental software was used

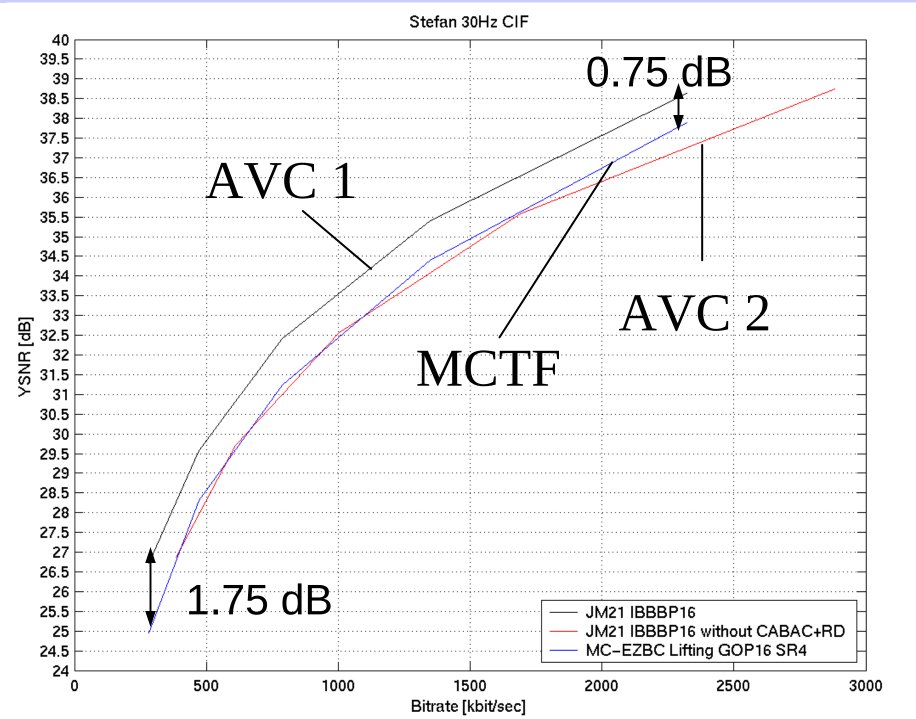
MPEG's Interframe Wavelet Coding Exploration

- The Interframe Wavelet exploration was successfully completed in October 2002
- 9 Call for Evidence on Scalable Coding Advances - July 2003
- 24 Call for Proposal Responses – Mach 2004

SNR Results from MPEG's Intraframe Wavelet Coding Exploration

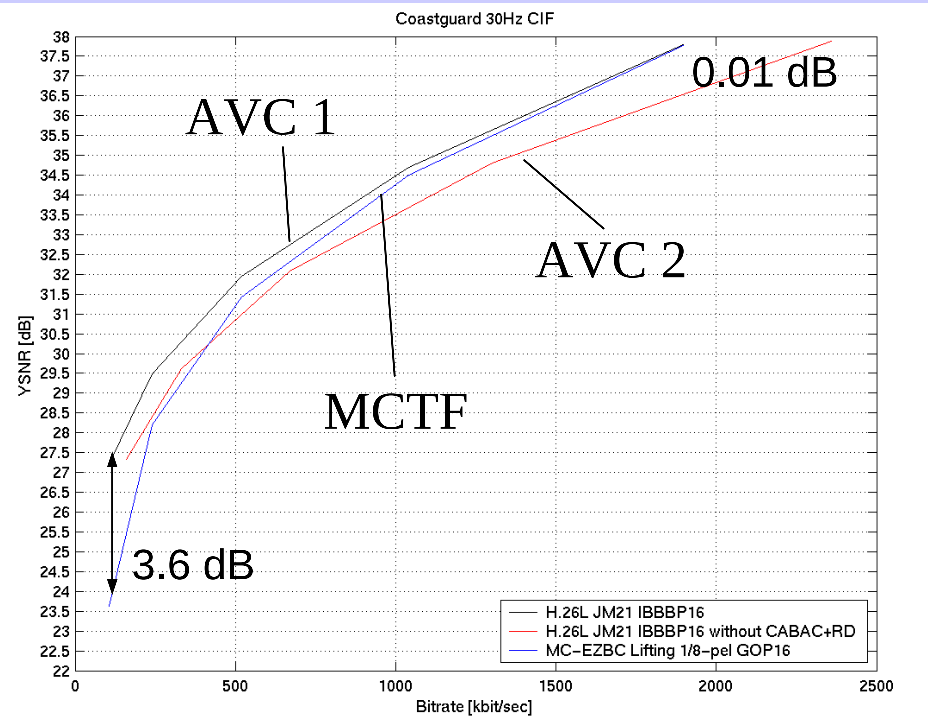
- Some less good results (out of 10 sequences)

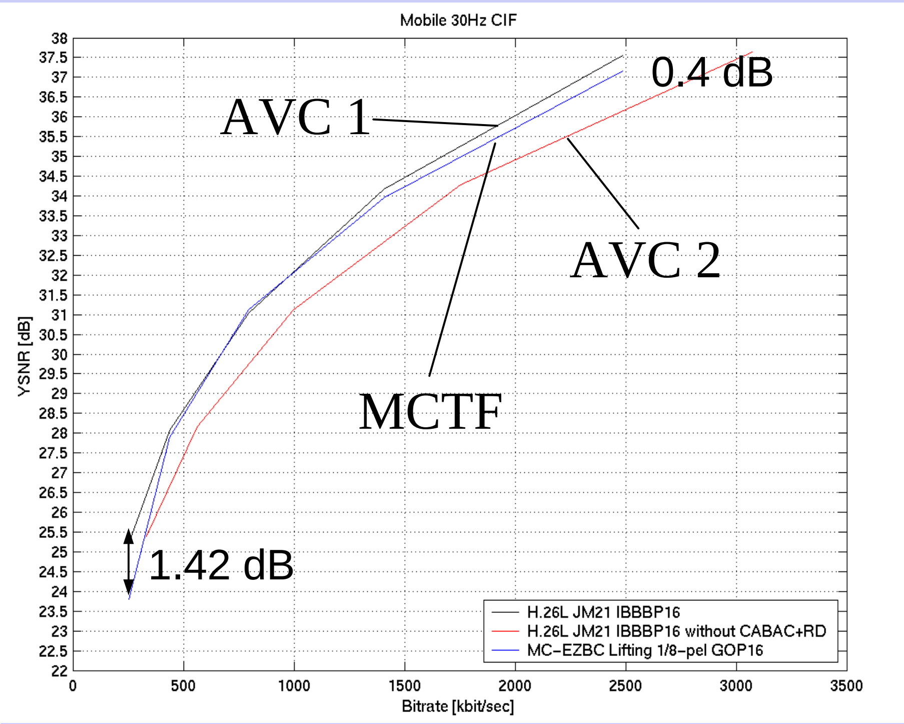




SNR Results from MPEG's Intraframe Wavelet Coding Exploration

- Some more good results (out of 10 sequences)





Acknowledgements

- Jens Ohm
- Yiannis Andreopoulos
- Jong Ye
- Konstantin Hanke
- Claudia Mayer
- Thomas Rusert