Compact Lecture

# Multimedia Coding: Methods & Applications

## Part 4: Video Coding Basics

### 4.4 Scalable Video Coding

Dr. Klaus Illgner

Dr. Uwe Rauschenbach

# Motivation

- **Adaptation for different consumer <u>devices</u>**
  - HDTV (1920 x 1080 pixel)
  - TV (STB / PC) (CCIR 720 x 576 pixel)
  - PDA (CIF: 352 x 288)
  - Mobile phone (QCIF: 172 x 144)

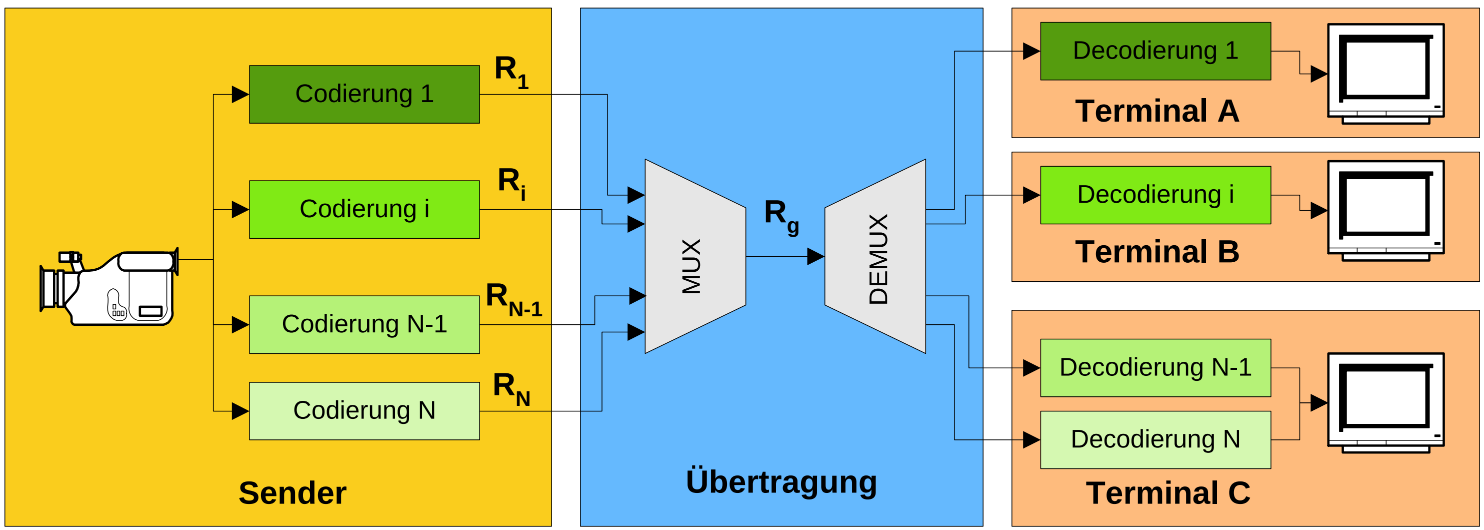

- **Adaptation to different <u>networks</u>**
  - WLAN 802.11b (2-6 Mbps)
  - UMTS (typical: 384 kbps) GPRS (64 kbps)
  - xDSL (700 – 6000 kbps)
  - ISDN (64 kbps)
- **Adaptation to <u>network conditions</u> (→ gracefull degradation)**
- **Adaptation to different <u>User Preferences</u>**

**→ various encodings of the same content have to prepared and transmitted, potentially in parallel → Simulcast**

# Characteristics -- Simulcast

**Aim:**

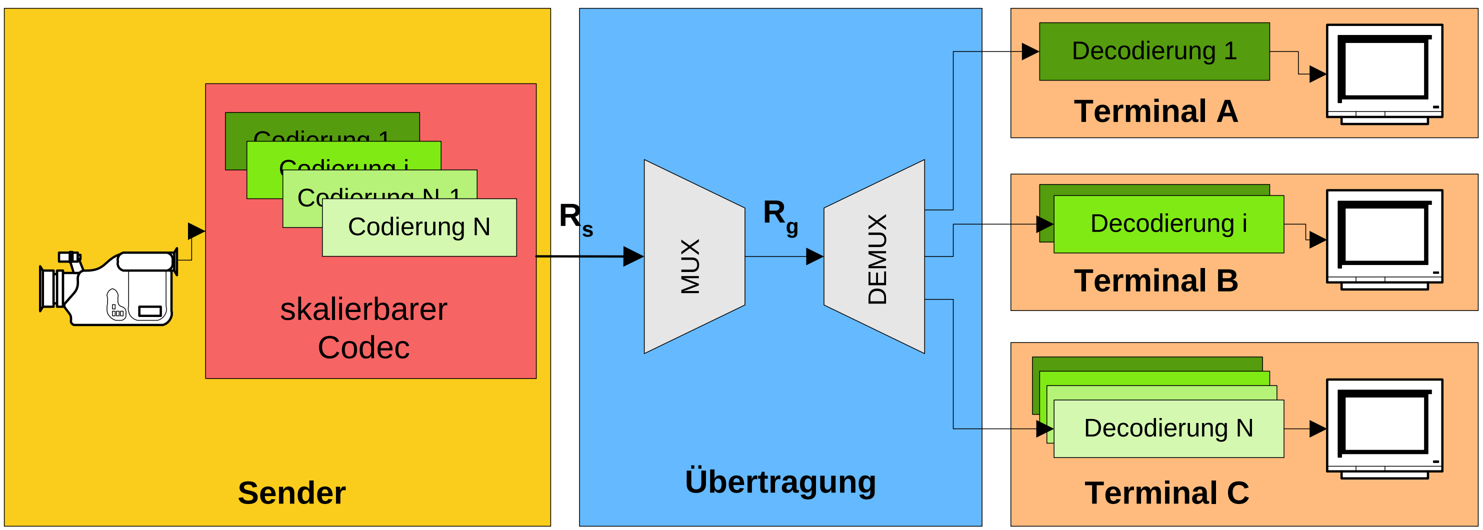
Simultaneous transmission of multiple encodings of the same data source with different quality levels (resolutions)



# Characteristics – Scalable Coding

**Aim:**

Simultaneous transmission of a single encoding of the same data source with different quality levels (resolutions)




**Distinct feature of scalable coding schemes:**

$$R_s < \sum R_i$$

# Concepts of Scalability

**Approach:**
images of lower „quality“ is implicitly contained in the image of higher „quality“.

**Graceful degradation:**

0 0 1 0 0 1 — low quality layer

0 1 1 0 1 0 0 1 0 1 1 0 0 0 1 0 0 1 1 0 0 1 0 0 — high quality layer

Reconstructing an image of **lower quality** by **partially** decoding of the bitstream

**Successive Refinement:**

low quality layer: 0 0 1 0 0 1

\+ refinement layer: 0 1 0 1 0 1 0 1 1 0 0 1 1 0 0 0

≡ high quality layer: 0 1 1 0 1 0 0 1 0 1 1 0 0 0 1 0 0 1 1 0 0 1 0 0

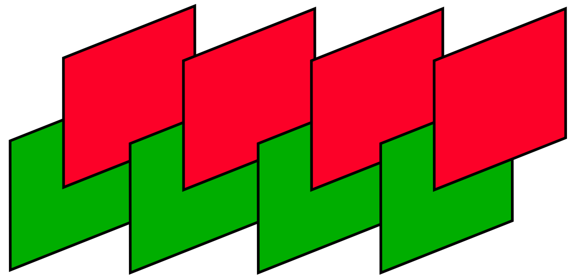
Reconstructing an image of **higher quality** by decoding an **complementing** bit stream

# Types of Scalability

## Temporal scalability

Enhancement layer

Base layer

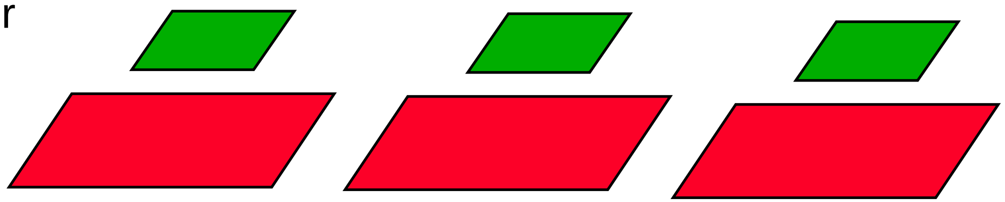


In this case:
Doubling the image frequency
e.g. B-frames

## Spatial scalability

Base layer

Enhancement layer



e.g. subband decomposition
or pyramids

## SNR scalability

e.g. different quantisers / „bitplane refinement“

## Hybrid scalability

Any combination of temporal, spatial and SNR scalability

# Performance Analysis

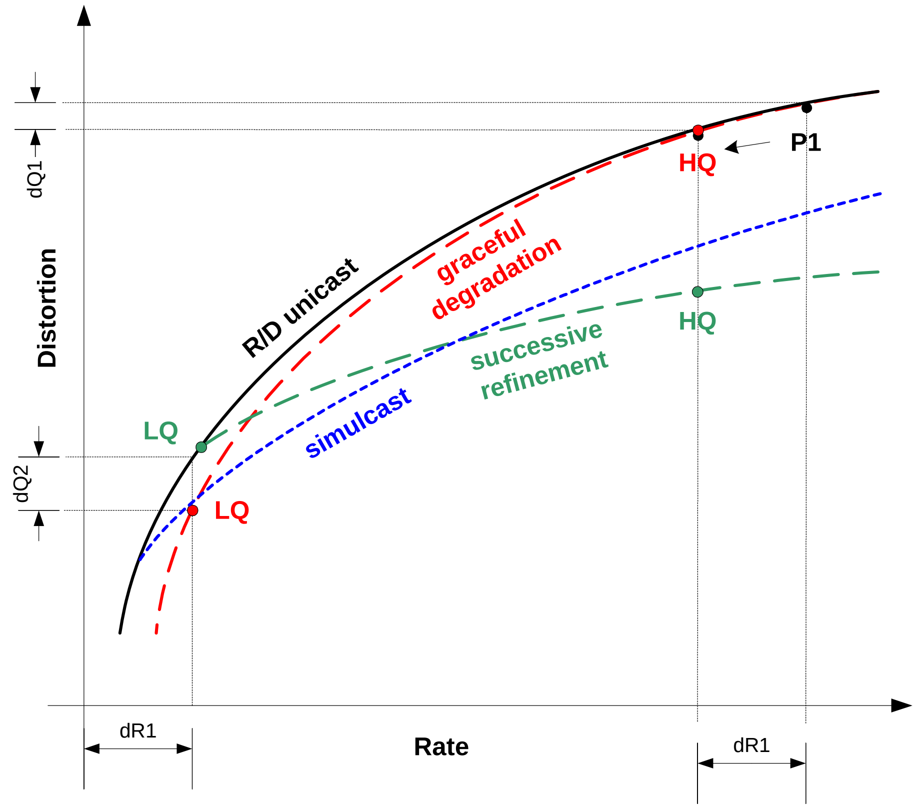
**Reference points:**

- N independent codings → Simulcast
- One „optimal“ coding → Unicast

**Observations:**

None of the established coding schemes gets close to the performance of a unicast encoding
(„R/D optimal encoding“)

$$\Delta R = \sum_{i=1}^{N} R_i - R_{HQ} > 0$$



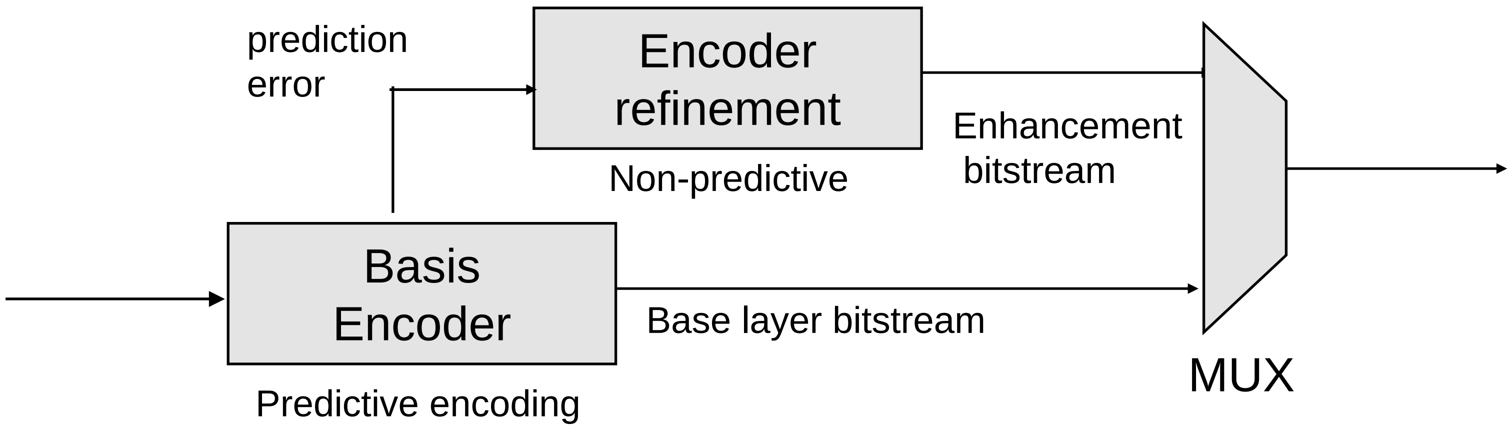

**Problem: Configuration**

- number of quality levels N
- Position of the points of operation
- degree of efficiency loss of scalable algorithms / simulcast operation compared to unicast

# Solution Concepts

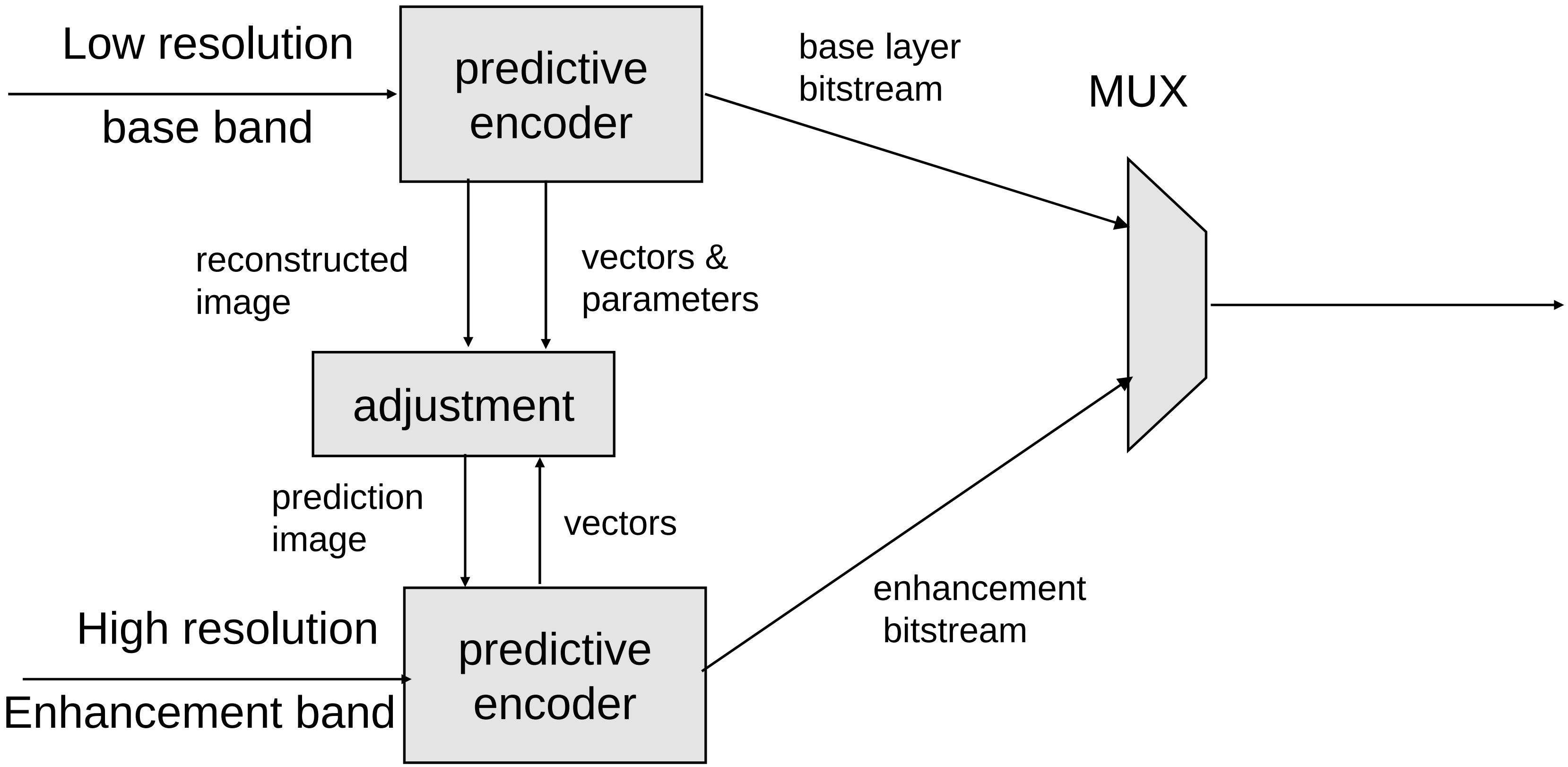
- **N independent Codings**
  - Simulcast
- **Refined coding of prediction error**
  - MPEG-4 FGS (Fine Grane Scalability) → Bitplane Coding
- **Independent encoding of the enhancement layers**
  - Temporal scalability →B-Frames
  - SNR scalability → bit plane coding
  - Pyramid Encoder
- **3D transform coding**

# MPEG-4 FGS



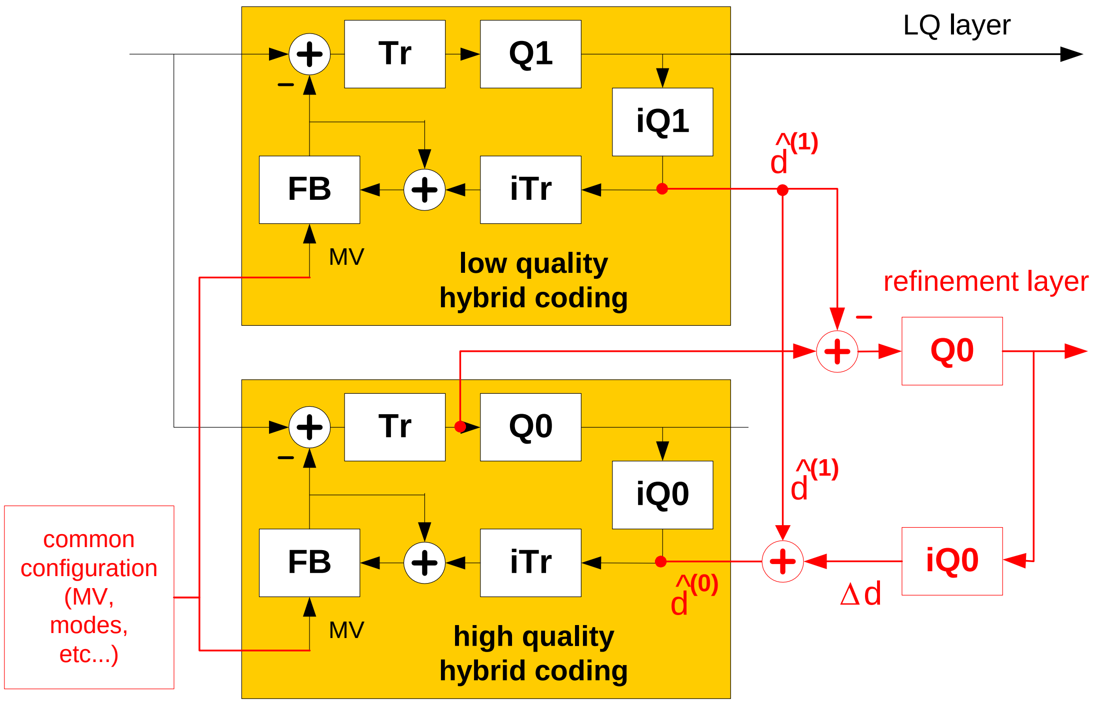

- applicable in principle to every encoder
- successive refinement of prediction error by bit plane coding (fine granularity)
- refinement information is being encoded as a still image (non predictive)
  → limited efficiency

# Independent Encoding of Enhancement Layers



# SNR Scalable Video Coding

**Advanced Simulcast**



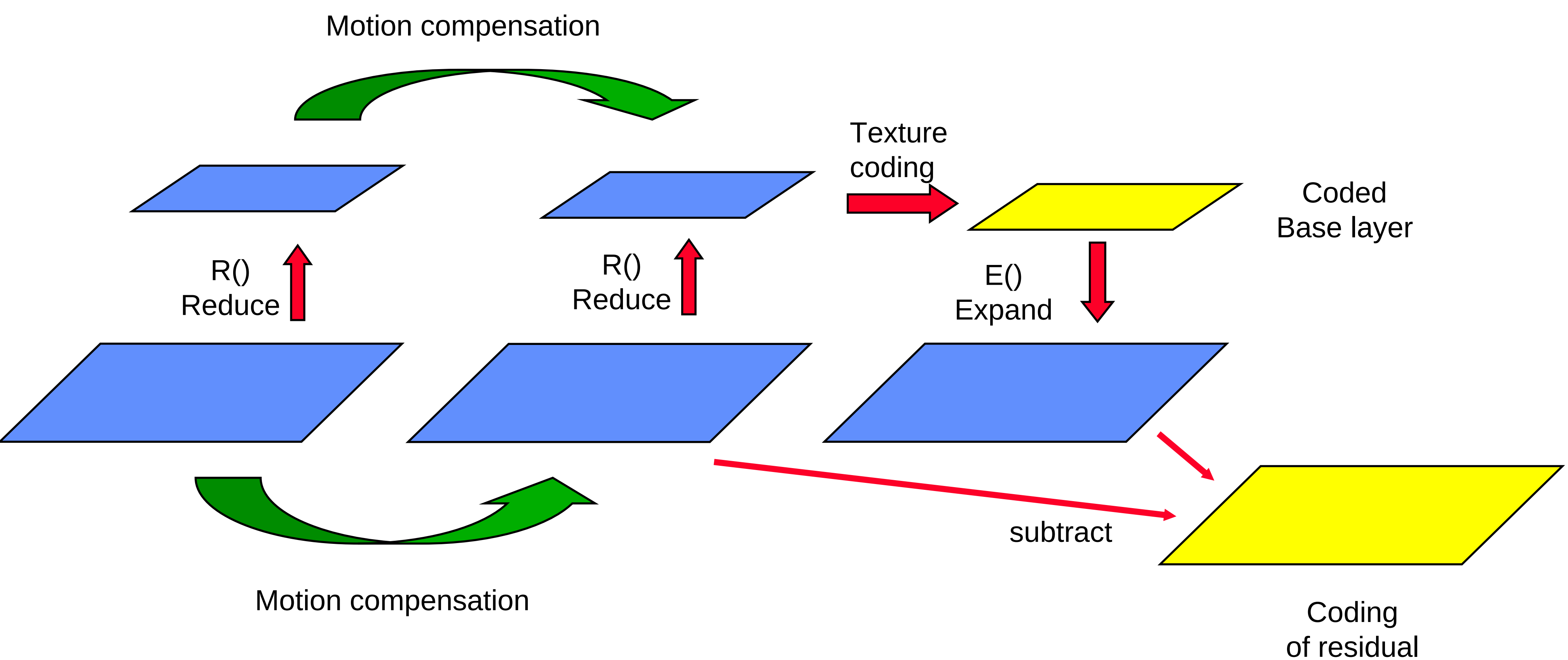

It holds:

$$\Delta d = Q^{(0)}\left(d^{(0)} - \hat{d}^{(1)}\right) = Q^{(0)}\left(d^{(0)}\right) - \hat{d}^{(1)} = \hat{d}^{(0)} - \hat{d}^{(1)}$$

**gain:** $H\left(\hat{d}^{(0)} - \hat{d}^{(1)}\right) < H\left(\hat{d}^{(0)}\right)$

Coding parameters of one layer

# Pyramid Encoder (1997)



# Joint Video Team

## Evaluation of proposals for SVC

- **5 proposals evaluated for fixing a technical starting point**
  - 1 x block-based, differential pyramid based on AVC
  - 2 x wavelet-based
  - 2 x hybrid wavelets with block based base layer
- **H.264 / AVC based approach had superior performance**
  - Selected as scalable video model (SVM)
  - SVM Software (reference software)
- **Characteristics of JVT Joint Scalable Video Model (JSVM 1.0)**
  - Standard compliant Base-Layer (AVC Main Profile)
  - Differential Pyramid (Bottom-Up approach)
  - MCTF with adaptive prediction
  - FGS functionality
  - 1 Decoder loop possible

**→ joint work group ITU-T / ISO MPEG developed an approach based on H.264**

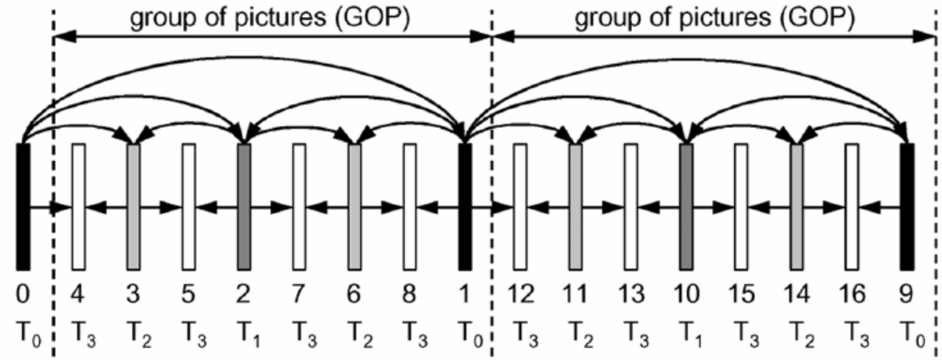
**→ standardised as H.264 amendment 3 SVC (2005)**

# H.264 AVC – SVC
# Temporal Scalability

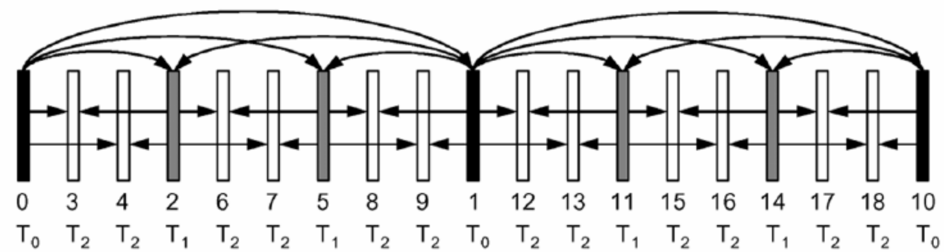
**Basic concept**

**Hierarchical B-pictures**

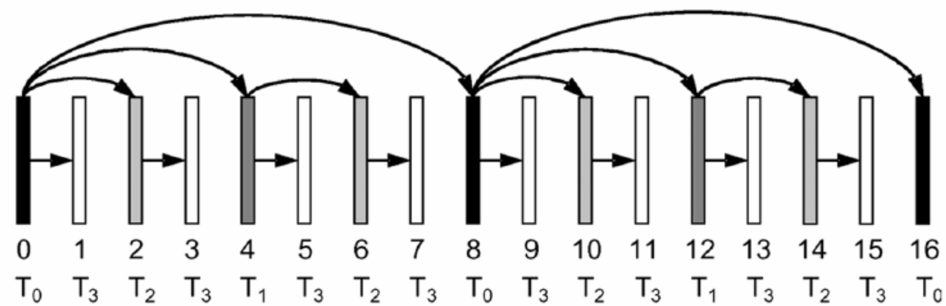
Dyadic representation




Non-dyadic representation

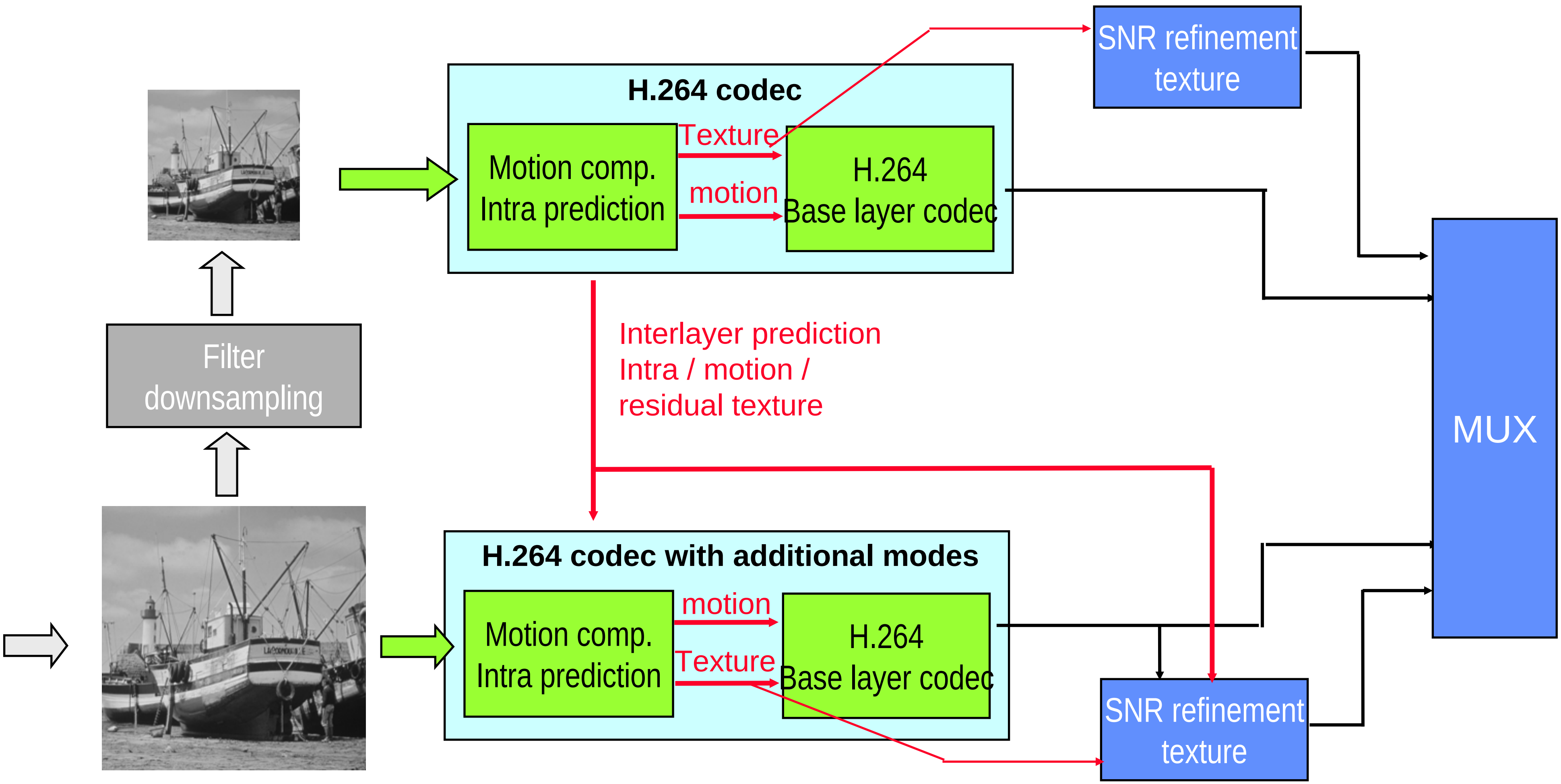



Zero delay structure




IEEE CSVT 17(9), 9/07

# H.264 AVC – SVC
# Spatial Scalability

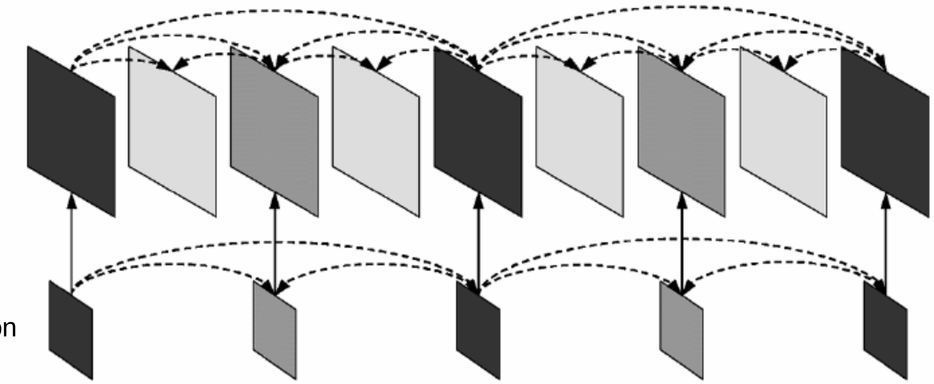


# Inter-Layer Prediction

Relationship between temporal resolution

and spatial layers for inter-layer prediction

$$\hat{g}_1^n = \alpha E\{\tilde{g}_0^n\} + (1-\alpha)\tilde{g}_1^{n-1}$$



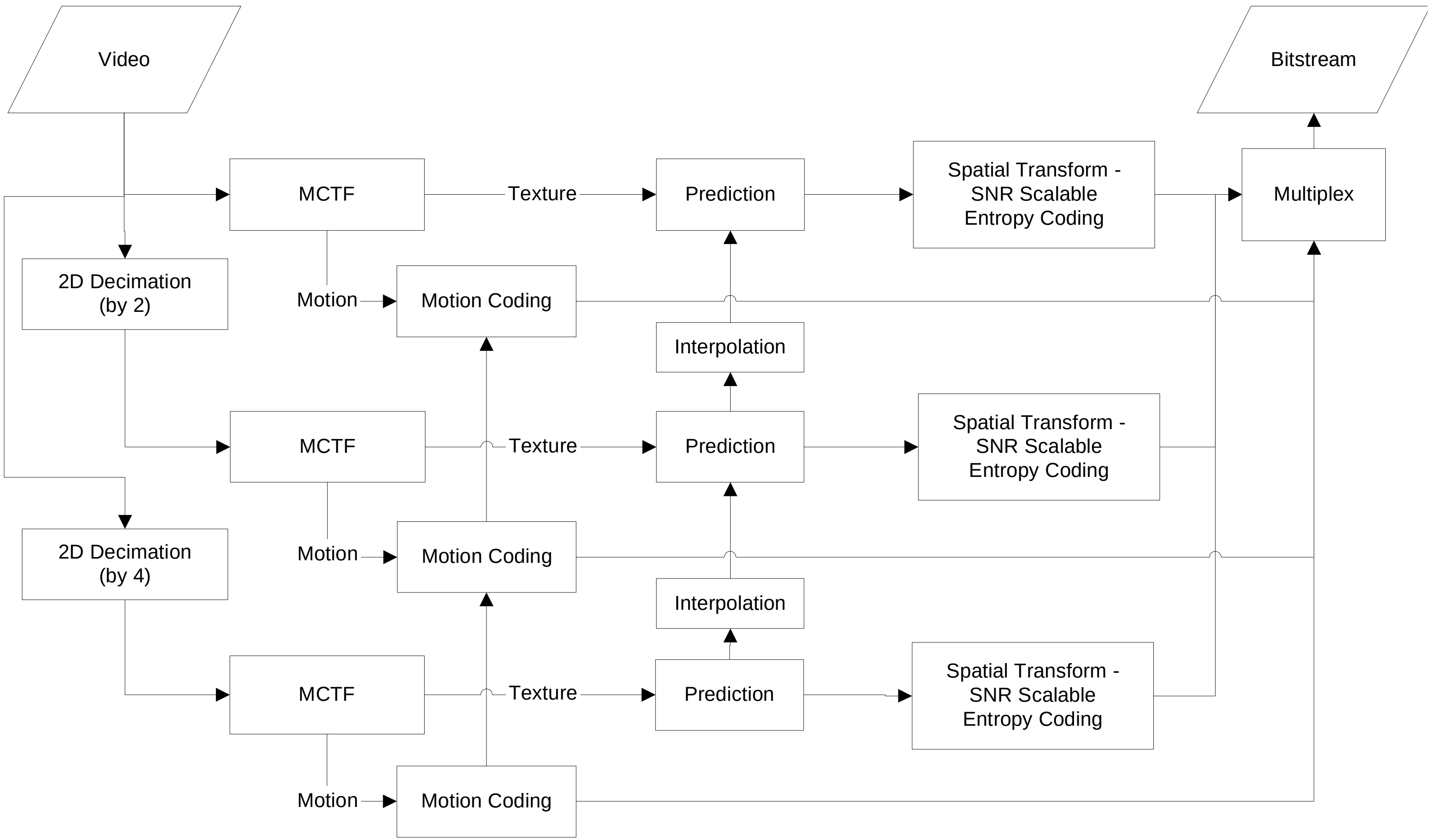
1: high spatial resolution
n: time index

Inter-layer motion prediction
- depends on mode of the corresponding block in the reference layer
- prediction of motion vectors

Inter-layer residual prediction
- applicable to all inter-coded macroblocks
- Block-wise bilinear upsampling

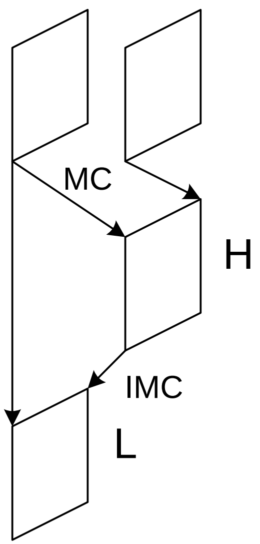
inter-layer intra prediction
- submacroblock in reference layer is intra-coded

# Wavelet based approach
# Architecture Overview

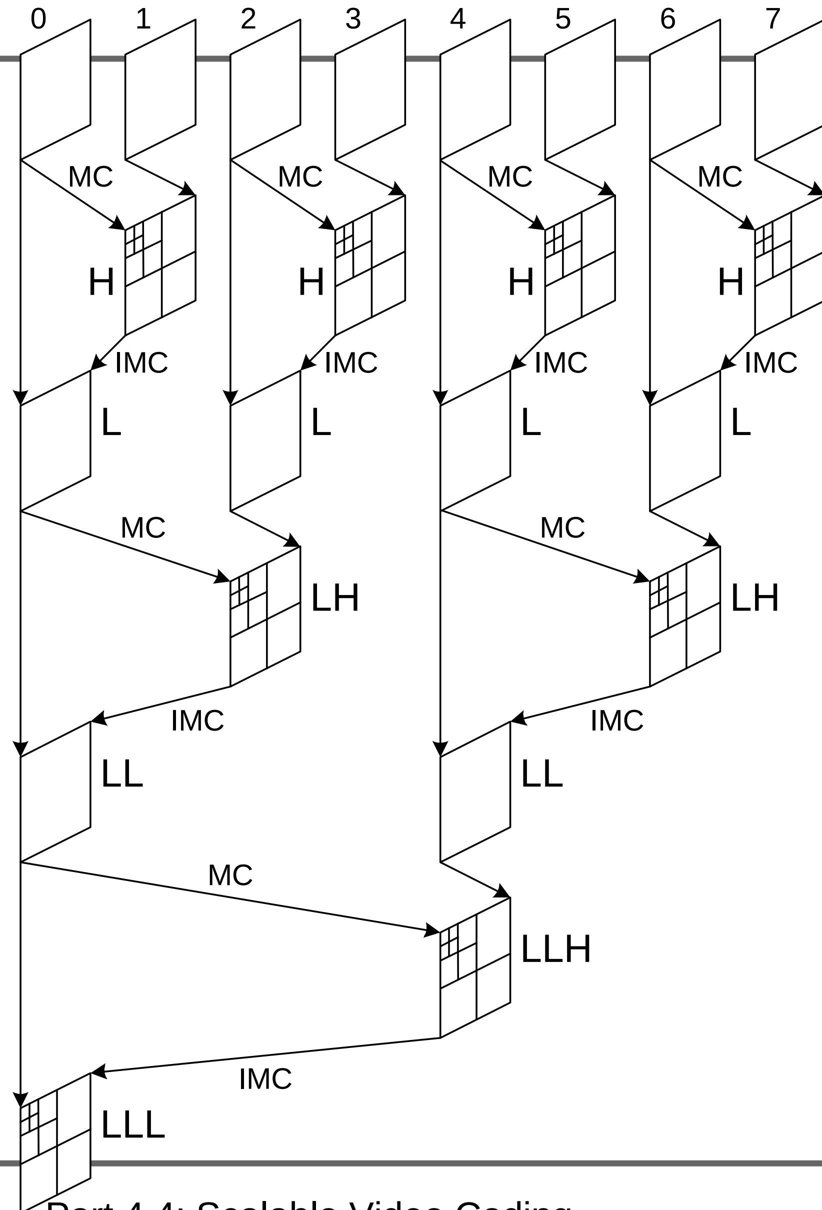


# Wavelet based approach
# Temporal Scalability

- **Motion compensated temporal filtering (MCTF)**
  - Low pass filtering to achieve lower temporal resolution




Basic element



# Wavelet Based Approach
# Spatial Scalability

- **H.264 / MPEG-4 AVC base layer**
- **Prediction of intra coded blocks from base layer**
- **Quality levels, e.g.**
  - QCIF @ 7.5 fps
  - QCIF @ 15 fps
  - CIF @ 15 fps
  - CIF @ 30 fps