Scalable Video Model 3.0

Thomas Wiegand

wiegand@hhi.de

Fraunhofer
Institut
Nachrichtentechnik
Heinrich-Hertz-Institut

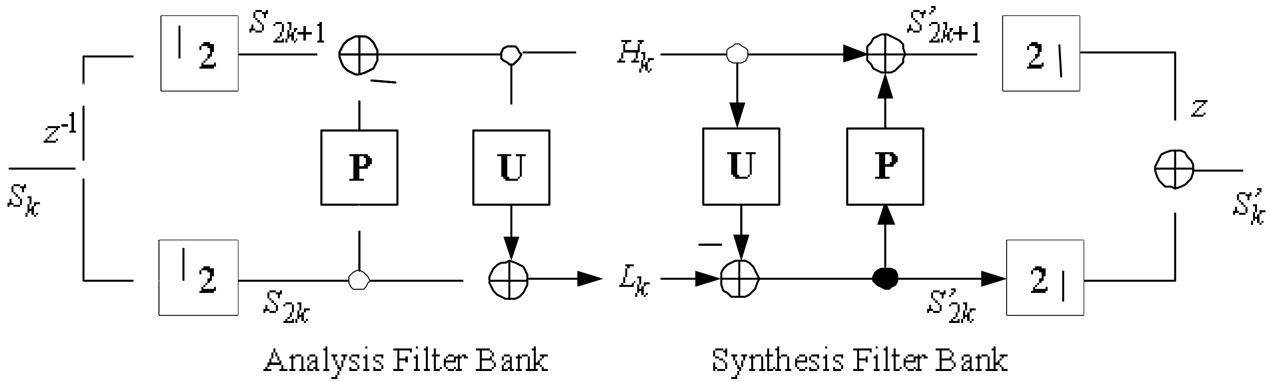
Outline

- Introduction
- Motion-compensated temporal filtering (MCTF)
- MCTF extension of H.264/MPEG-4 AVC
- Spatial Scalability
- SNR Scalability
- Combined scalability
- Summary

Introduction

- **Recent advances in MCTF-based video coding**
  - Incorporation of motion compensation into the lifting steps of a temporal filter bank
  - Lifting scheme is invertible
  - Motion compensation with any motion model possible to incorporate
    - Variable block-size motion compensation
    - Sub-sample accurate motion vectors
- **Realization as an extension of H.264/MPEG-4 AVC**
  - Highly efficient motion model of H.264/MPEG-4 AVC
  - Lifting steps are similar to motion compensation in B slices
  - Block-based residual coding
  - Open-loop structure of the analysis filter bank offers the possibility to efficiently incorporate scalability

MCTF using the Lifting Representation



Obtaining the high-pass (prediction residual) pictures

$$H_k = S_{2k+1} - \mathbf{P}(S_{2k})$$

Obtaining the low-pass pictures

$$L_k = S_{2k} + \mathbf{U}(S_{2k+1} - \mathbf{P}(S_{2k})) = \tfrac{1}{2} S_{2k} + \mathbf{U}(S_{2k+1})$$

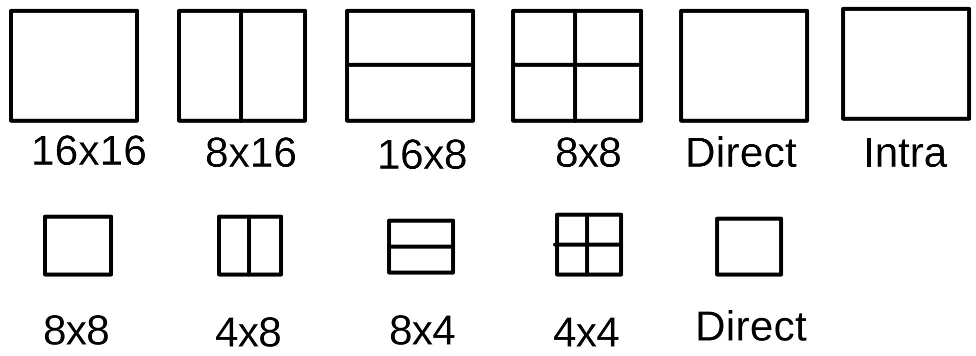
$$\mathbf{U}(\mathbf{P}(s)) = s/2$$

slide 4

Motion model of H.264/MPEG-4 AVC

Macroblock modes

Sub-macroblock modes



- Variable block-size motion model
- Quarter-sample accurate motion vectors
- Reference picture selection for each macroblock partition
- Selection of predictive and bi-predictive coding for each macroblock partition
  - Predictive: One list of reference pictures - $\mathbf{m}_{P0}$ and r_{P0}
  - Bi-predictive: Two lists of reference pictures - $\mathbf{m}_{P0}$, r_{P0}, $\mathbf{m}_{P1}$, and r_{P1}

Prediction and Update operators

- Adaptive switching between the lifting representations of the Haar and the 5/3 spline filters on a block-basis
- Incorporation of multiple reference picture capabilities
- Incorporation of intra macroblock modes

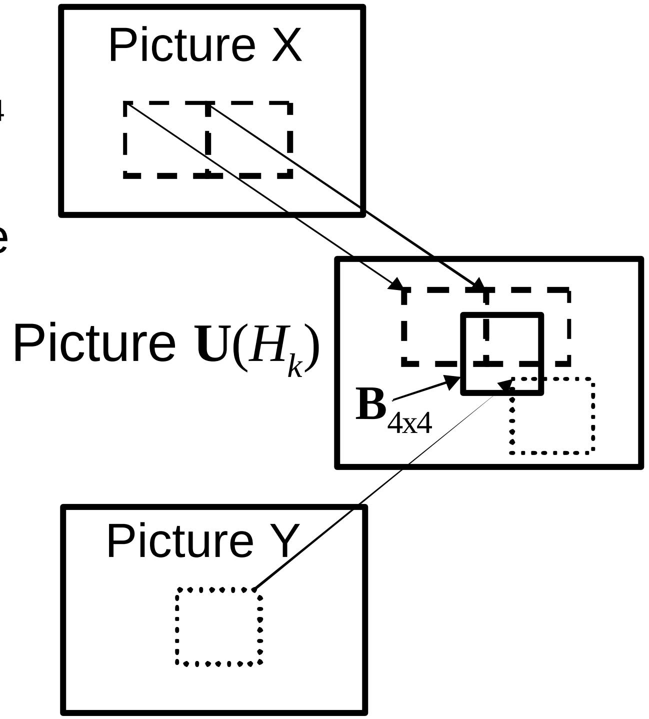
$$\mathbf{P}\left(S[\mathbf{x},2k]\right)=\tfrac{1}{2}\left(w_0 \cdot S[\mathbf{x}+\mathbf{m}_{\mathrm{P}0},\, 2k-2r_{\mathrm{P}0}]+w_1 \cdot S[\mathbf{x}+\mathbf{m}_{\mathrm{P}1},\, 2k+2r_{\mathrm{P}1}+2]\right)$$

$$\mathbf{U}\left(H[\mathbf{x},k]\right)=\tfrac{1}{4}\left(w_0 \cdot H[\mathbf{x}+\mathbf{m}_{\mathrm{U}0},\, k+r_{\mathrm{U}0}]+w_1 \cdot H[\mathbf{x}+\mathbf{m}_{\mathrm{U}1},\, k-r_{\mathrm{U}1}-1]\right)$$

w_0	w_1	Transform	Prediction / Update
1	1	5/3 spline wavelet	Bi-predictive / update
2	0	Haar wavelet	Predictive / update
0	2		
0	0	Identity transform	Intra mode / no update

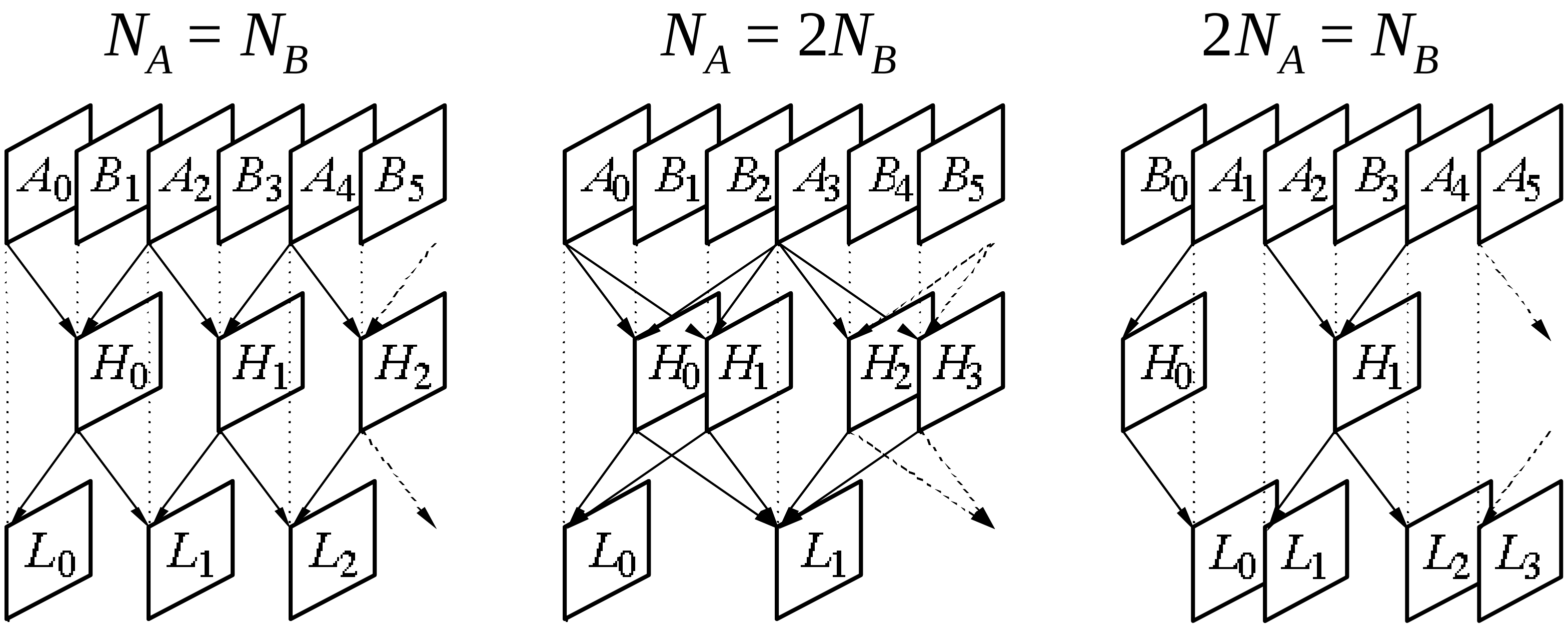
Derivation of Update Operators

- Design goal: derive motion vectors and reference picture indices that can be input to H.264/MPEG-4 AVC motion compensation without having to change it at highest coding efficiency
- For each 4x4 luma block $\mathbf{B}_{4x4}$ in the picture $\mathbf{U}(H_k)$, derive $\mathbf{m}_{U0}$, r_{U0}, $\mathbf{m}_{U1}$, and r_{U1} as follows
  1. Evaluate all $\mathbf{m}_{P0}$ and $\mathbf{m}_{P1}$ that point into $\mathbf{B}_{4x4}$
  2. Select those $\mathbf{m}_{P0}$ and $\mathbf{m}_{P1}$ that use maximum number of samples for reference out of $\mathbf{B}_{4x4}$
  3. Set $\mathbf{m}_{U0} = -\mathbf{m}_{P0}$ and $\mathbf{m}_{U1} = -\mathbf{m}_{P1}$
  4. Set r_{U0} and r_{U1} to point to those pictures into which MC is conducted using $\mathbf{m}_{P0}$ and $\mathbf{m}_{P1}$, respectively
  5. Harmonize derived $\mathbf{m}_{U0}$, r_{U0}, $\mathbf{m}_{U1}$, and r_{U1} with H.264/MPEG-4 AVC syntax



Temporal Coding Structure

- Group of N_0 input pictures partitioned into two sets: Set 1: N_A ($0 < N_A < N_0$), Set 2: $N_B = N_0 - N_A$
- Set 1: pictures A_k, Set 2: pictures B_k
- Pictures H_k are spatially shift-aligned with pictures B_k
- Pictures L_k are spatially shift-aligned with pictures A_k



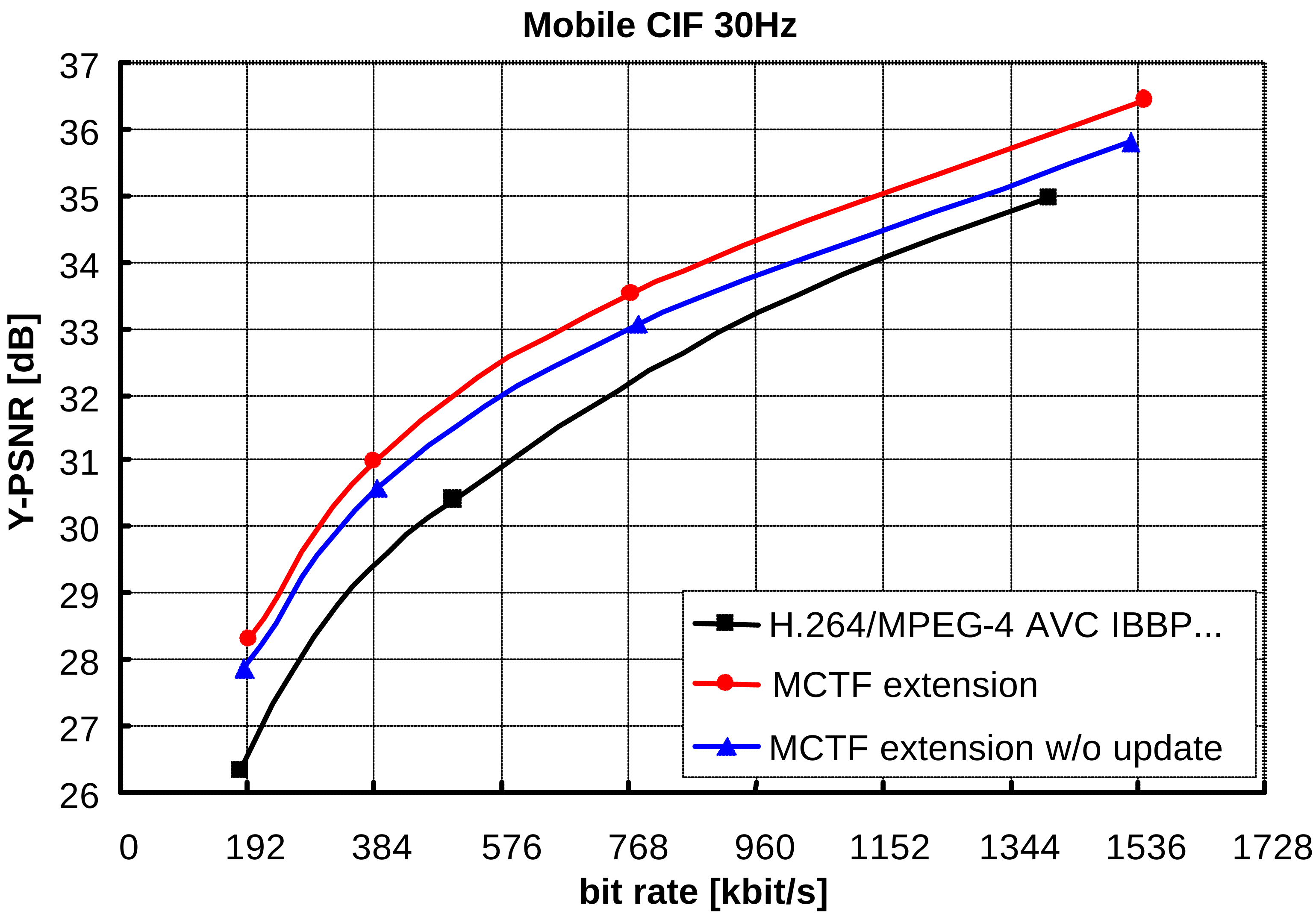
Temporal Decomposition Structure

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

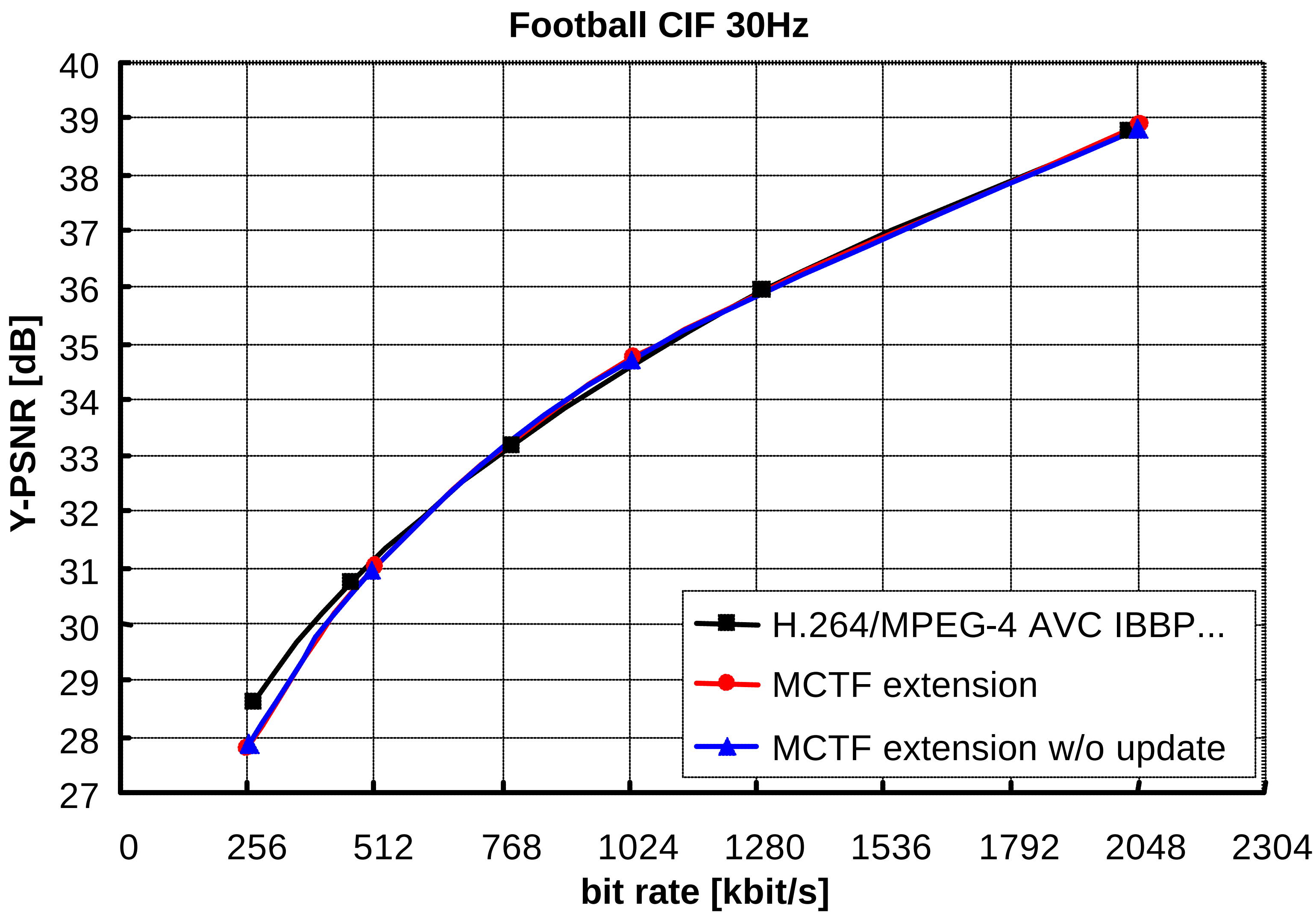
Impacts on H.264/MPEG-4 AVC

- **Syntax is not changed**
  - High-pass pictures and motion fields are coded as B pictures
  - Low-pass pictures are coded as I or P pictures
- **Decoding process**
  - Motion-compensated update with extended bit-depth
  - Derivation process for update motion fields
    - Inversion of prediction motion fields
    - Block-wise motion compatible with the H.264/MPEG-4 AVC
  - Without update: H.264/MPEG-4 AVC compatible coder with open-loop control
- **Encoding process**
  - Lagrangian methods as in the H.264/MPEG-4 AVC test model
  - Cascading of quantization parameters according to the scaling factors of the analysis filter bank

Simulation Results for the MCTF Extension

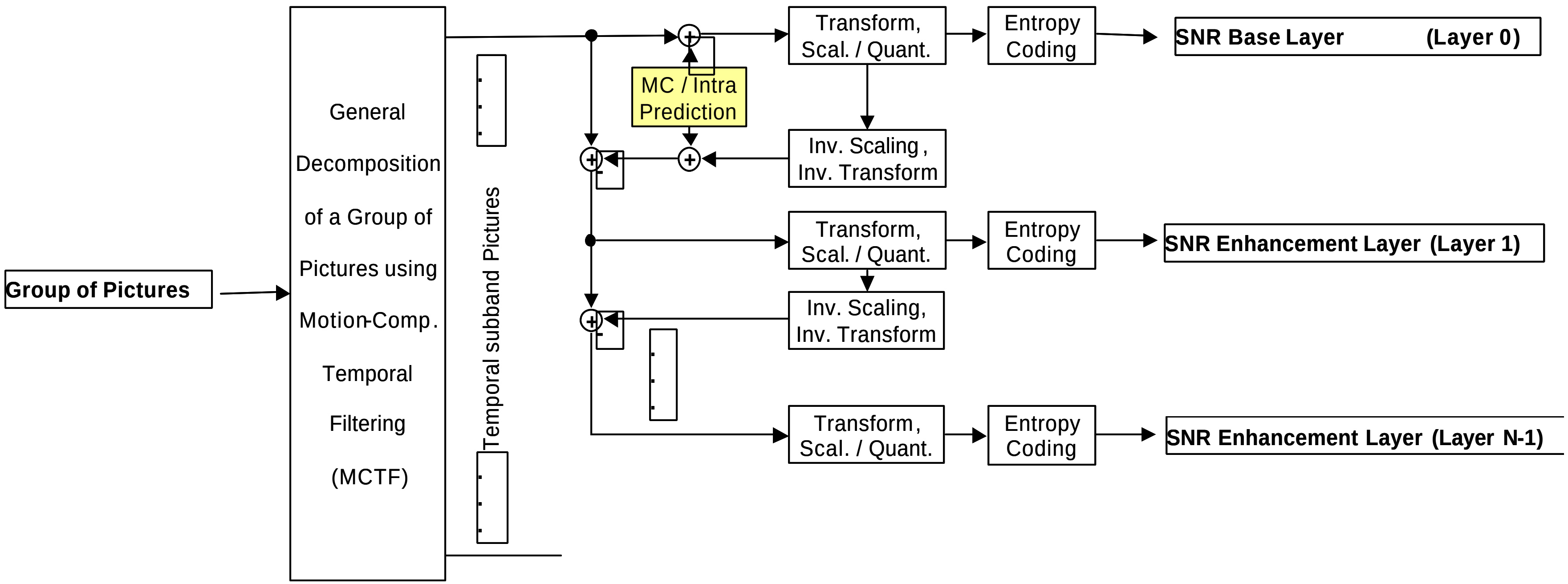


Simulation Results for the MCTF Extension

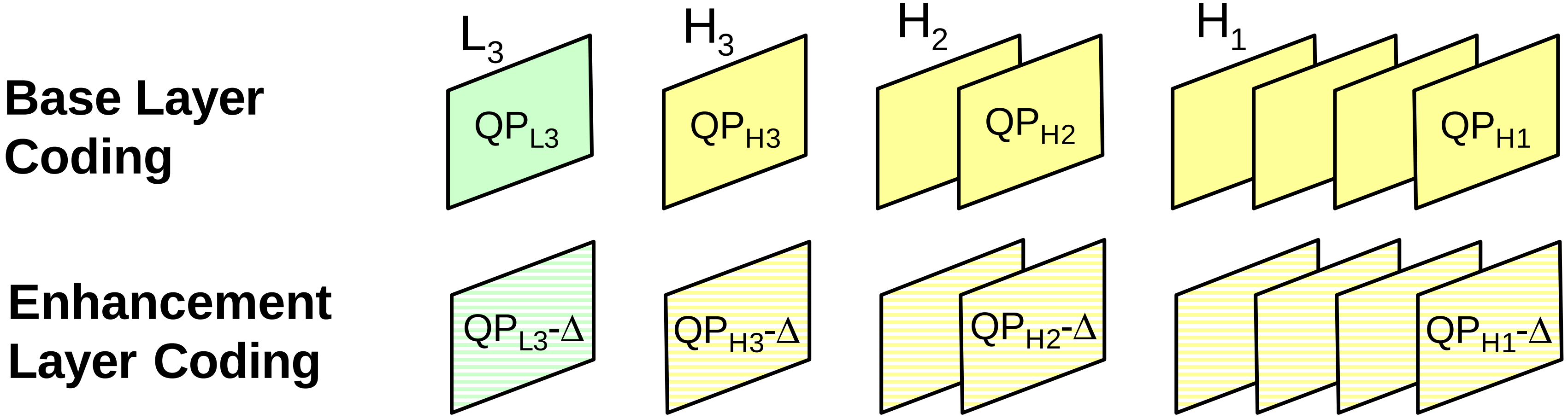


SNR-Scalable Extension

- **Layered representation of subband pictures**
  - Residual coding of the quantization error between the original subband pictures and their base layer reconstruction
  - Efficient for coarse grains of scalable SNR layers



Adjustment of the Quantization Parameter

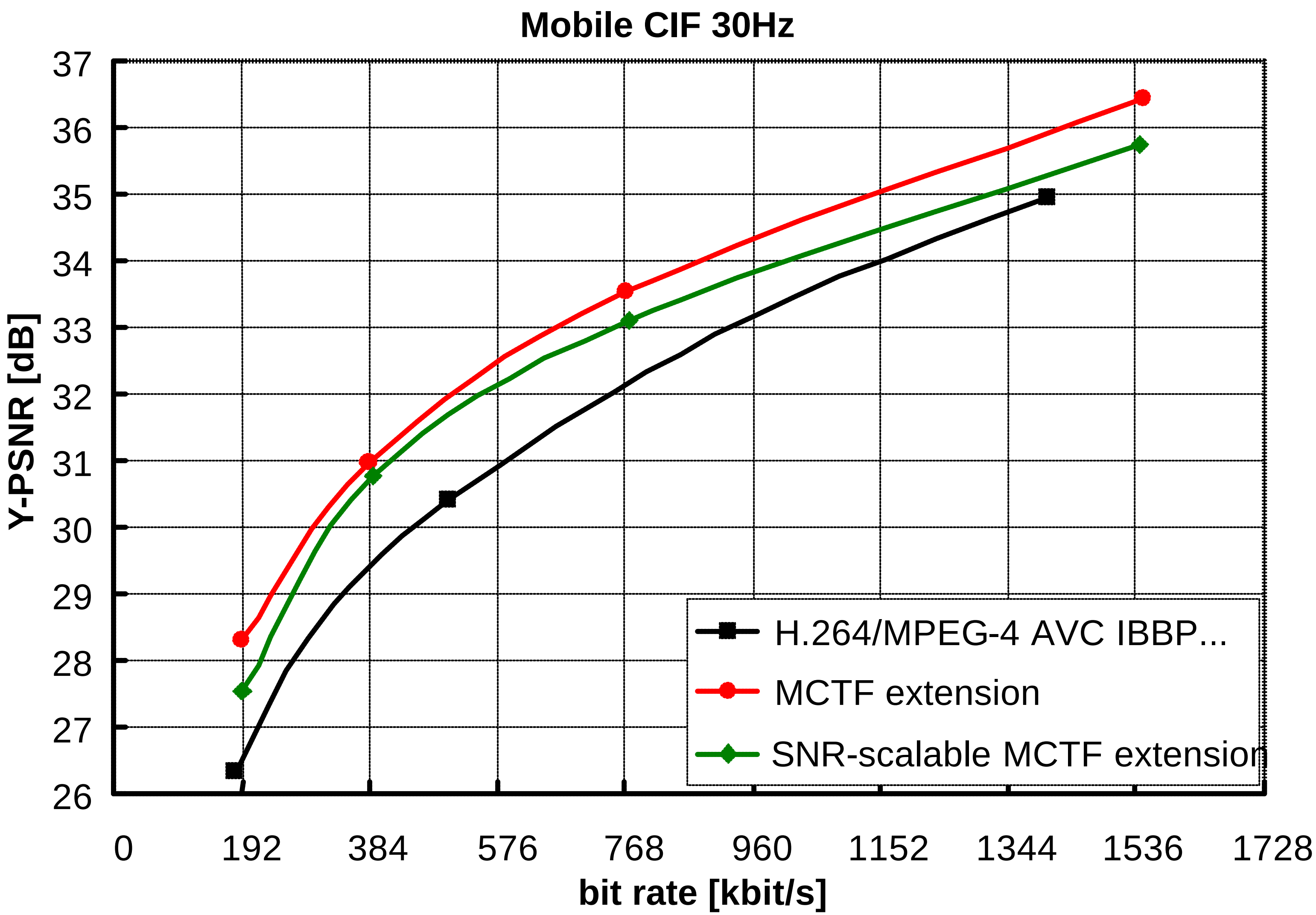


- Enhancement Layer
  - Difference between original and reconstructed subband pictures after decoding the base layer (or a previous enhancement layer)
  - Difference pictures are coded using the residual picture syntax
- Small performance losses if the quantization parameter is decreased by a value of 6 from one layer to the next
  - Doubling of the bit-rate (approximately)

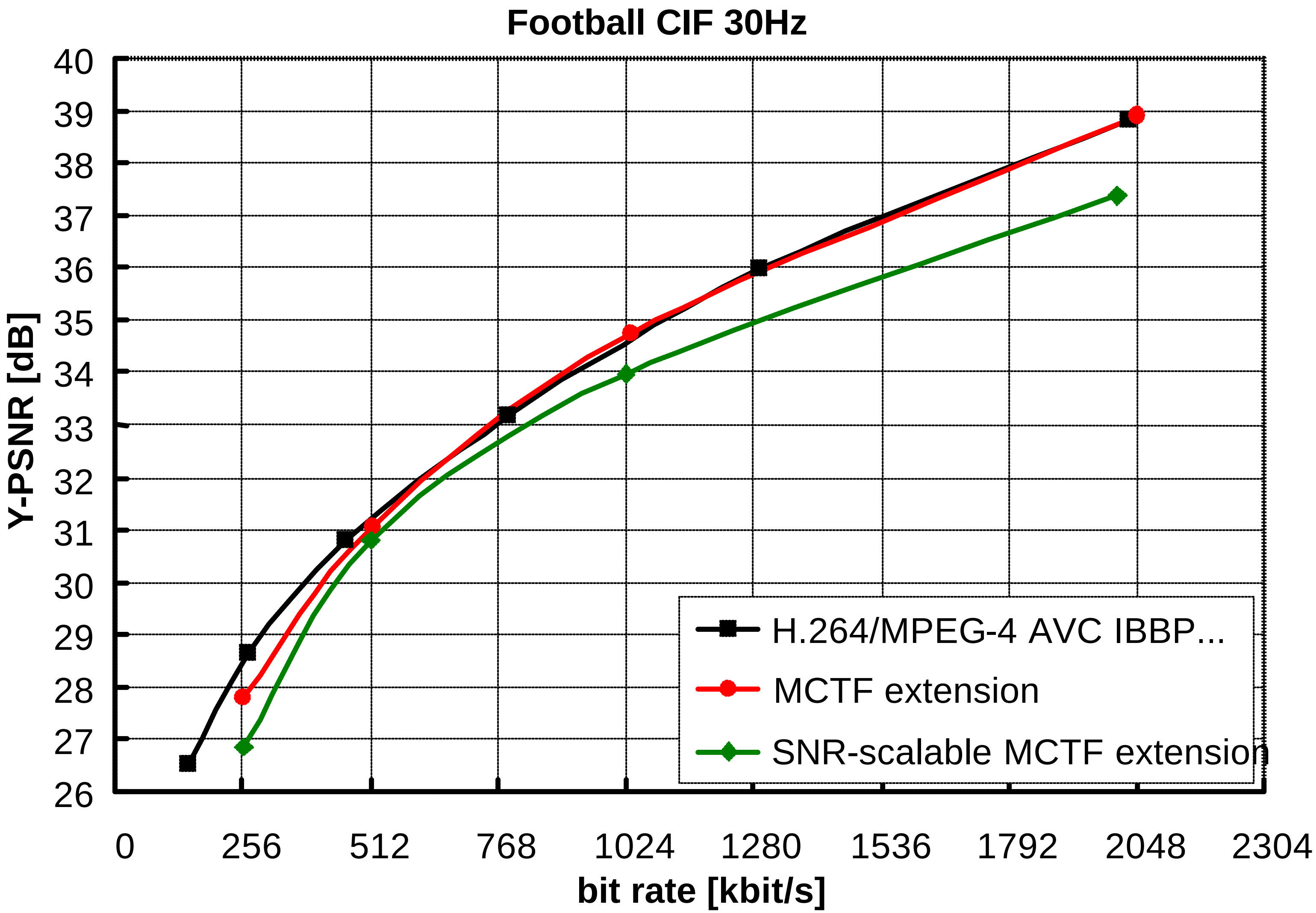
Impacts on H.264/MPEG-4 AVC

- **Syntax**
  - Additional slice type for coding of residual pictures
  - Re-use of parts of the macroblock syntax: CBP, residual coding
- **Decoding process**
  - Importing data from subordinate SNR layers
  - Motion-compensation after adding up the residual signals
- **Encoding process**
  - Consider entire range of rate points
  - Carefully trade-off bit-rates for motion and residual data
  - Experience shows closed to single layer efficiency for all supported bit-rates

Results for Layered SNR-Scalability



Results for Layered SNR-Scalability

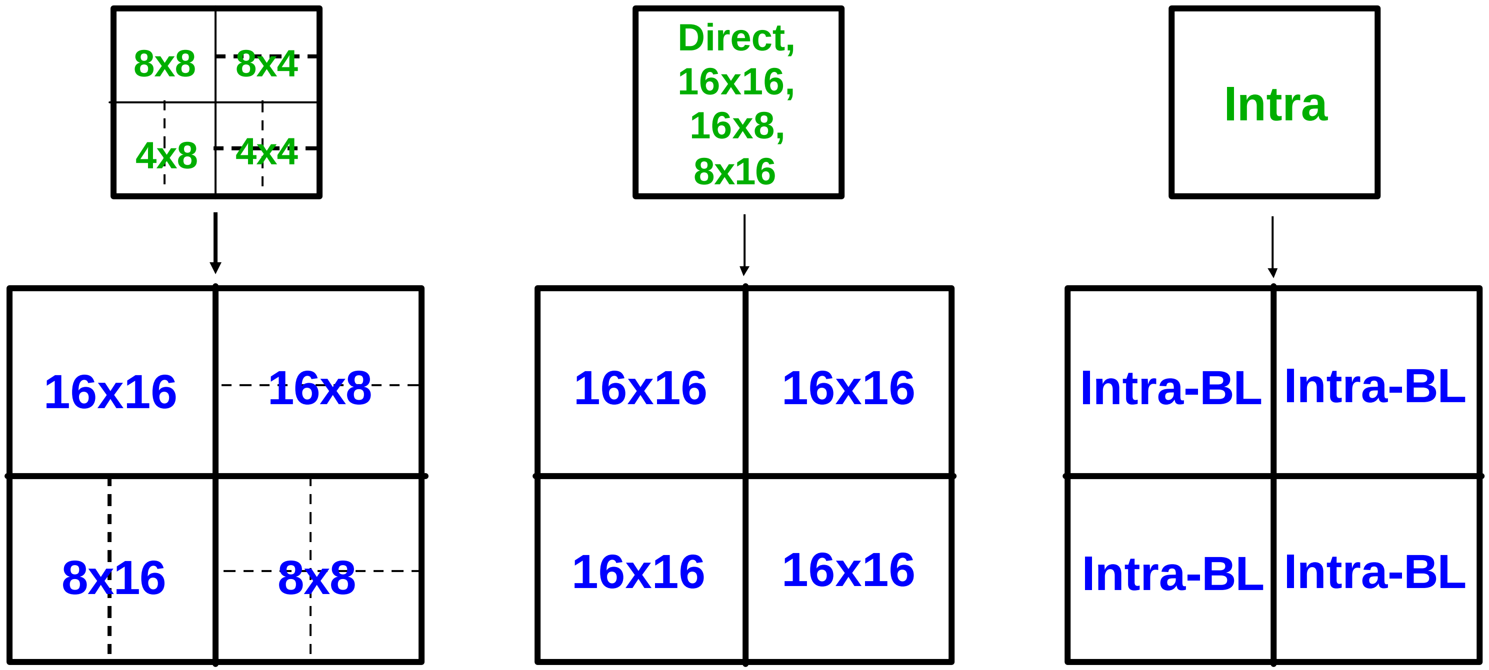


Spatial Scalability

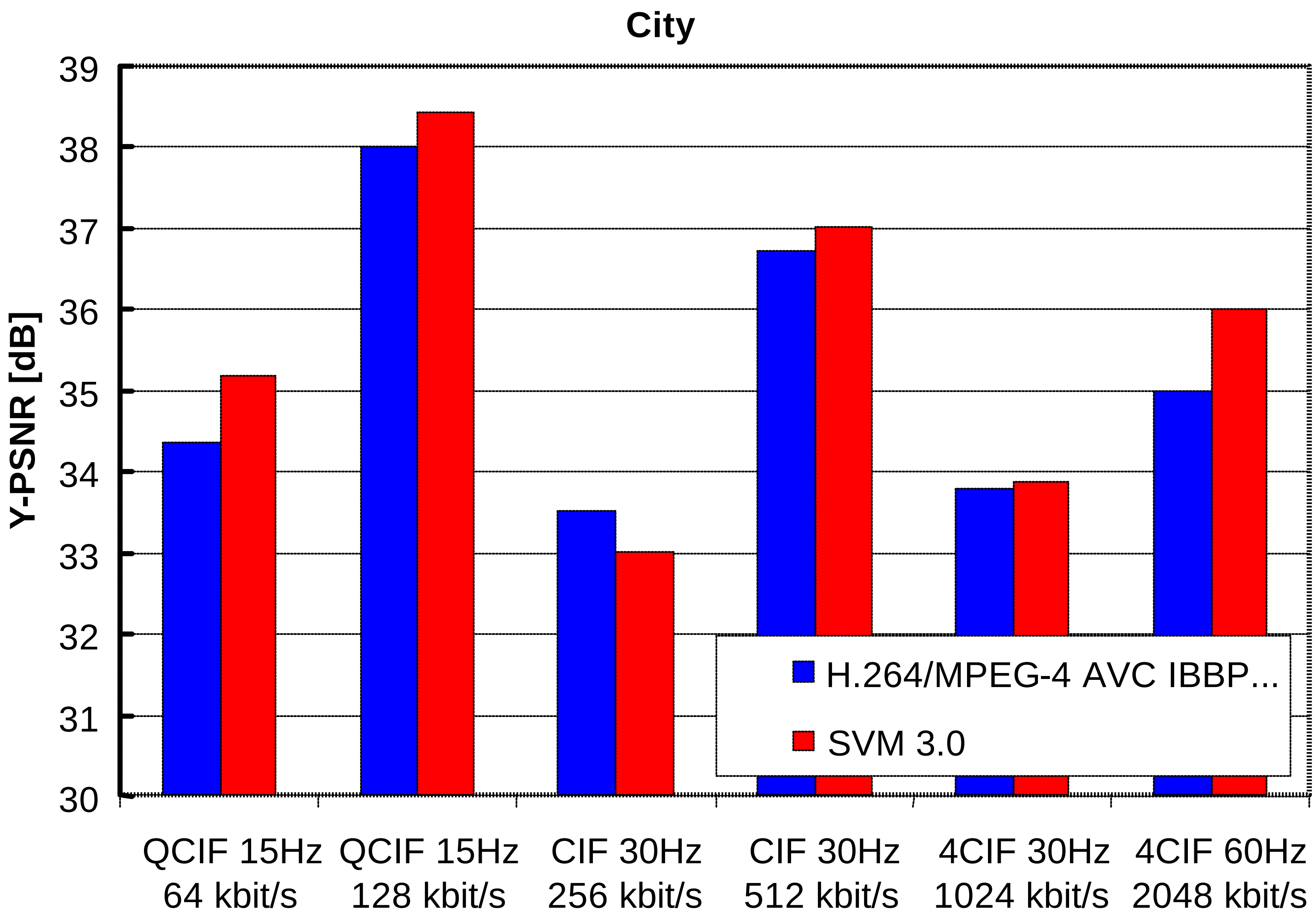
- **Lowest spatial resolution is H.264/MPEG-4 AVC conforming: MCTF without update step**
- **Oversampled pyramid with independent MCTF for each resolution: QCIF, CIF, 4CIF, 16CIF, ...**
- **Switchable prediction mechanisms for conveying information from lower spatial layer to the next**
  - Motion information prediction by defining two additional macroblock modes and a switchable motion vector prediction
  - Prediction of residual information (high-pass signals) using the upsampled residual signal of the lower resolution layer
  - Intra prediction using the reconstructed signal of the lower resolution layer

Spatial Prediction of Data

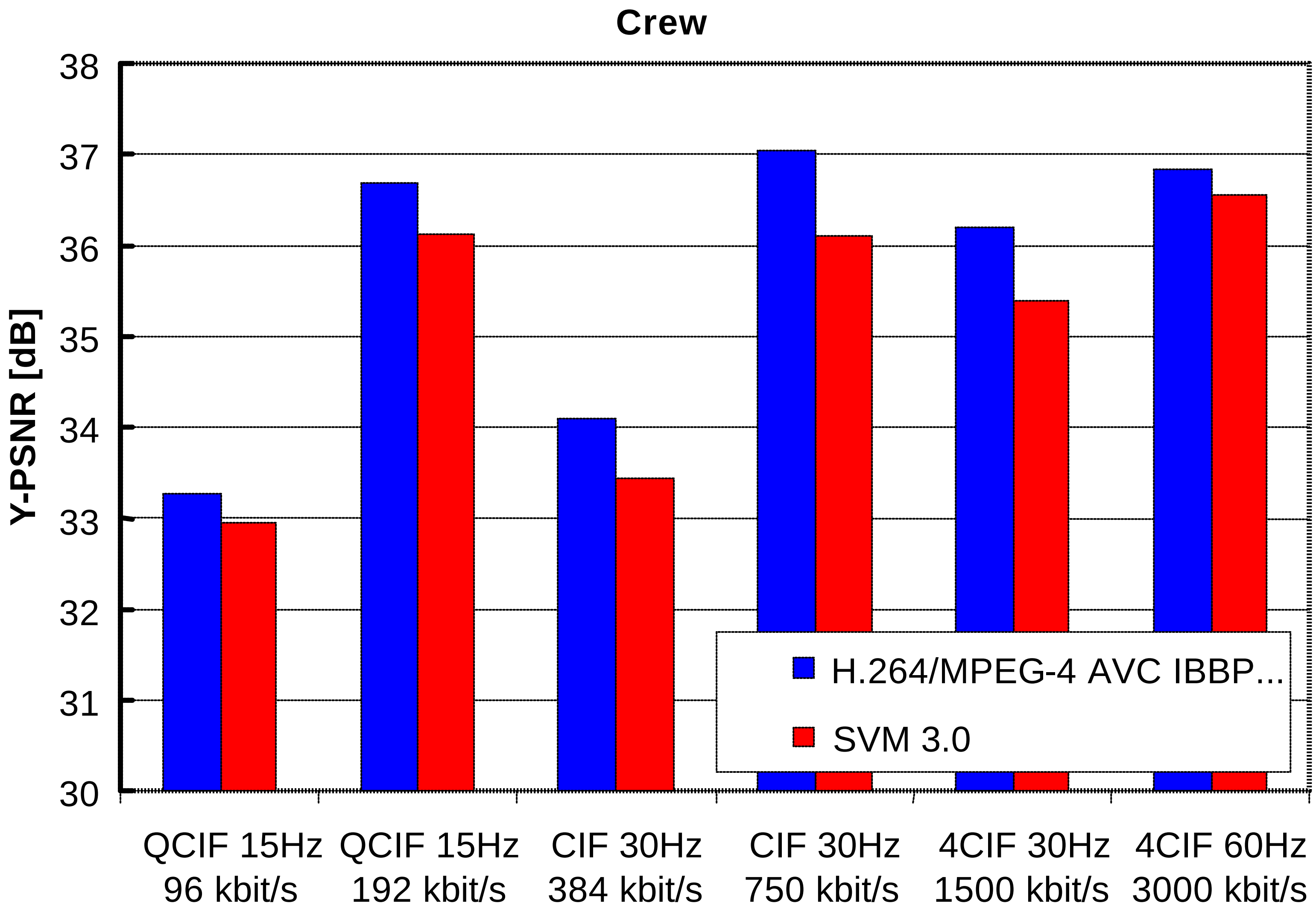
- Upsample partitioning and motion vectors and use them for prediction (keep list 0, list 1, bi-predictive and reference indices information)
- Block-wise upsample residual picture using bi-linear filter
- Upsample reconstructed intra macroblock using half-pel filter of H.264/MPEG-4 AVC



Results for Layered Combined Scalability



Results for Layered Combined Scalability



Fine Granular Scalability

- **Layered representation of subband pictures**
  - Non-scalable base layer representation
    - Motion information
    - Base layer representation of intra and residual data
    - Minimal acceptable reconstruction quality
  - Quality scalable enhancement layer representations
    - Residue between original and base layer representation
    - FGS packets can be truncated at arbitrarily points
    - FGS packet: refinement signals corresponds to a bisection of the quantization step size
    - Refinement signals are directly coded in the transform domain
    - Single inverse transform at the decoder side

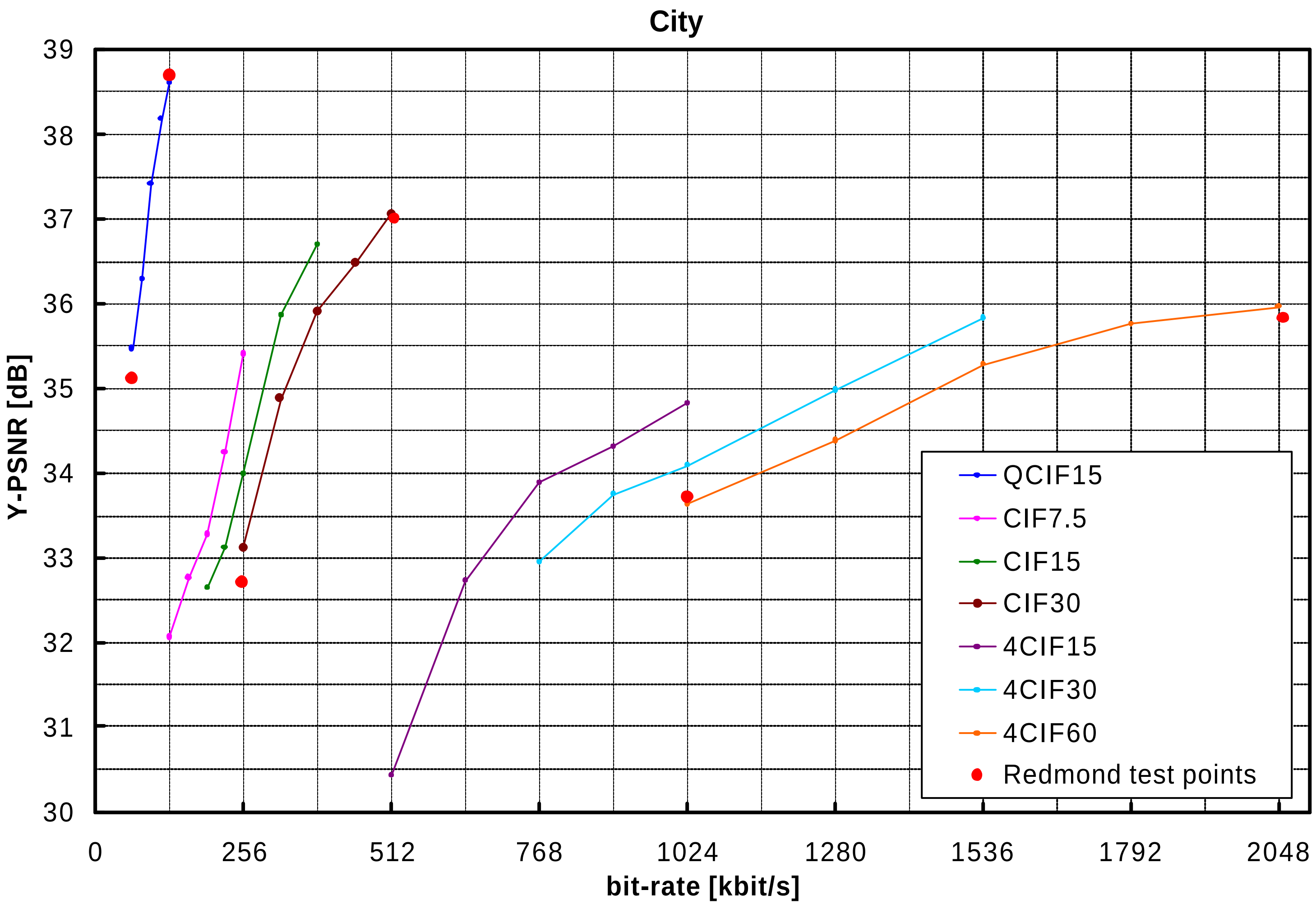
Fine Granular Scalability

- **Coding of enhancement layer packets**
  - Re-quantization of transform coefficients (similar to Redmond HHI proposal)
  - Usage of CABAC contexts as specified in H.264/MPEG-4 AVC with minor modifications
  - 6 CABAC contexts added
  - Modification of the transmission order of transform coefficient levels
  - Differentiation between significant and non-significant transform coefficient level
  - Non-significant transform coefficient levels are basically coded using CABAC
  - For significant transform coefficients, only a refinement symbol (-1, 0, 1) is transmitted

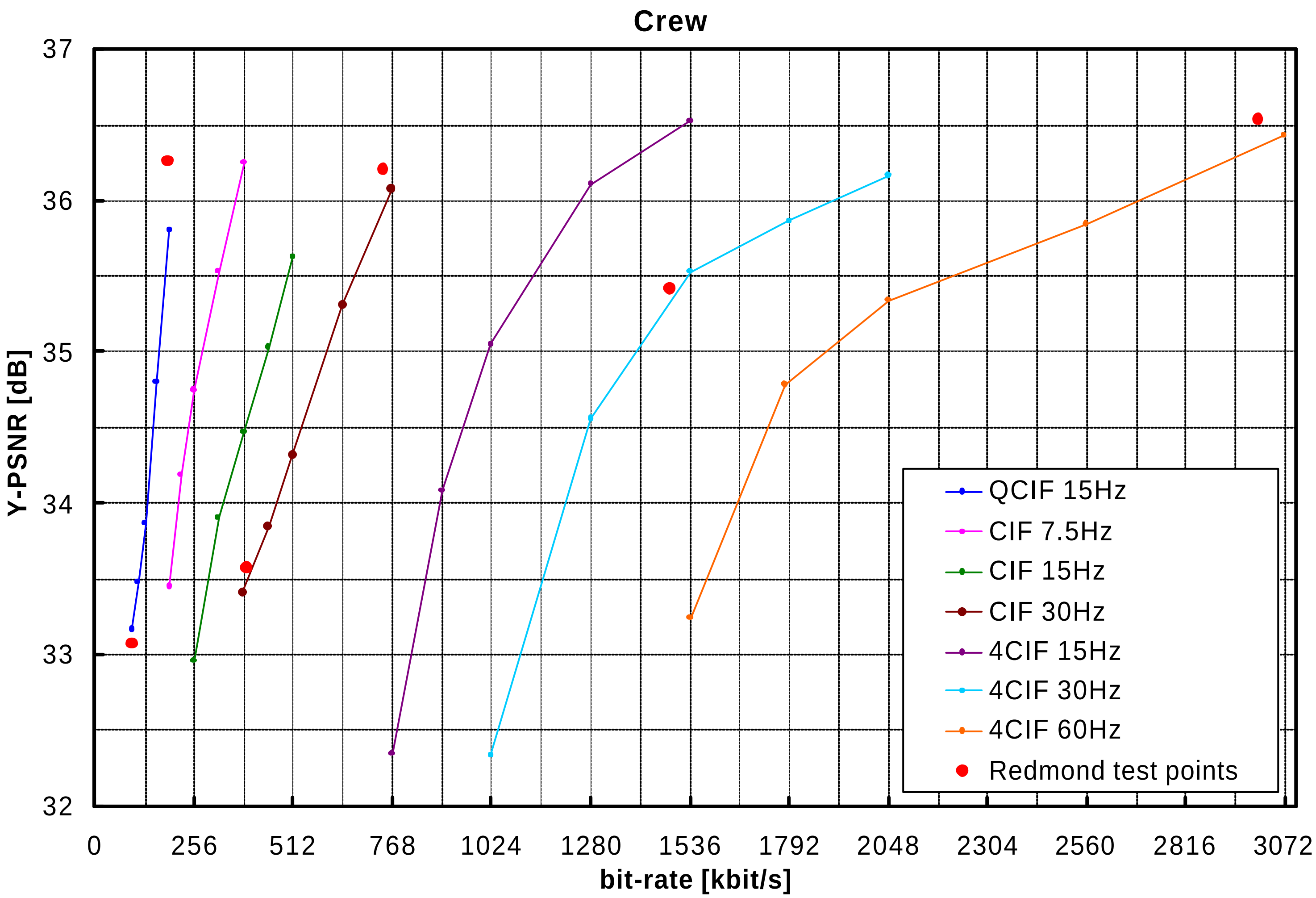
Fine Granular Scalability

- **Three scan for coding of transform coefficients**
  - 1. non-significant transform coefficients of significant blocks
  - 2. refinement symbols for already significant coefficients
  - 3. coefficients of non-significant blocks
- **Scanning pattern for each of the three scans**
  - Scanning from low to high frequency bands (zig-zag)
    - Mapping of 8x8 blocks onto four 4x4 blocks
  - Inside each band: first luma then chroma in raster scan
- **Coding of transform coefficient levels**
  - Non-significant coefficients
    - Coded Block Pattern, Coded Block Bit, (Delta QP, Transform Size)
    - CABAC symbols (SIG, LAST, SIGN, ABS)
  - Significant coefficients Refinement symbol (-1, 0, +1)

Simulation Results: City



Simulation Results: Crew



Summary

- **Straightforward MCTF extension of H.264/MPEG-4 AVC**
  - Addition of motion-compensated update steps
  - Usage of H.264/MPEG-4 AVC tools as specified in the standard
- **Scalability**
  - Layered representation of subband pictures
  - Adding of a slice type for coding residual signals
  - Layered SNR scalability or FGS
  - Usage of the residual coding tools of H.264/MPEG-4 AVC
  - One MC loop only when restricting prediction from base-layer reconstructed signals to intra
  - Temporal scalability is inherently provided by the MCTF scheme
  - Spatial scalability using a oversampled pyramid structure and independent temporal decomposition in each spatial layer with inter-layer prediction mechanisms