

Scalable Extension of H.264 / AVC

*Heiko Schwarz, Detlev Marpe, and
Thomas Wiegand*



Fraunhofer
Institut
Nachrichtentechnik
Heinrich-Hertz-Institut

Overview

- **Basic codec components**
 - Block-adaptive motion-compensated temporal filtering (MCTF)
 - Block-based coding of subband pictures
 - Layered Coding with inter-layer prediction
 - Deblocking filter at the decoder side
- **Realization as an extension of H.264/AVC**
 - **Most components of H.264/AVC can be used as specified**
 - Syntax and semantics (with some minor extensions)
 - Network abstraction layer
 - Motion compensation / intra prediction
 - Residual coding
 - Entropy coding
 - Only a few components need to be adjusted to the subband structure
 - Current H.264/AVC can be used as base layer

MCTF: Lifting Representation

- Two-channel analysis-synthesis filterbank

$$h[\mathbf{x}, k] = s[\mathbf{x}, 2k + 1] - \mathbf{P}(s[\mathbf{x}, 2k + 1])$$

$$l[\mathbf{x}, k] = s[\mathbf{x}, 2k] + \mathbf{U}(s[\mathbf{x}, 2k])$$

- Motion-compensated prediction and update operators

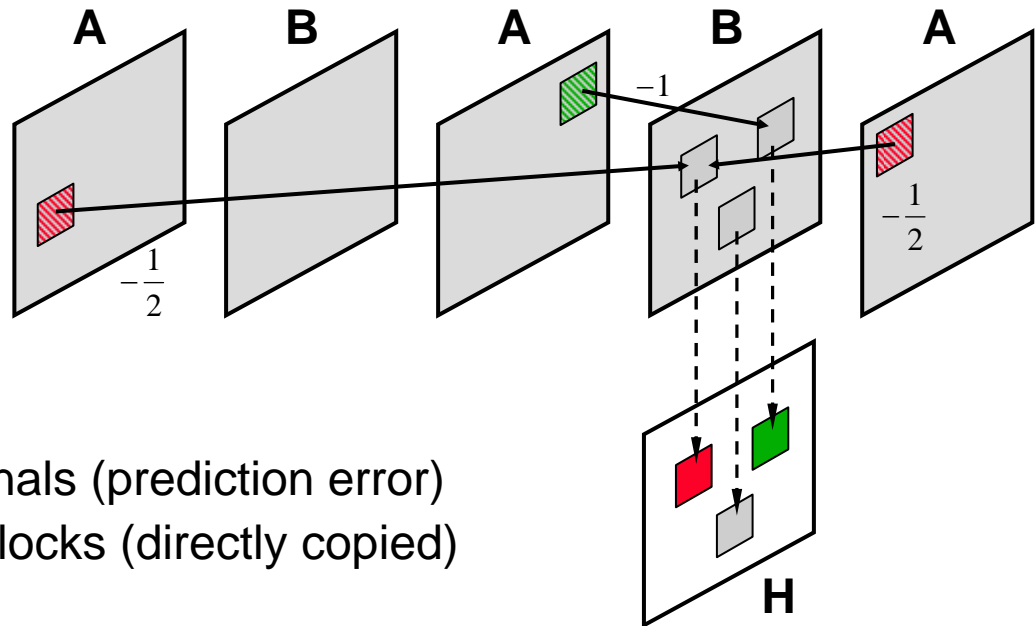
$$\mathbf{P}(s[\mathbf{x}, 2k + 1]) = \frac{1}{2} (w_0 \cdot s[\mathbf{x} + \mathbf{m}_{P0}, 2k - 2r_{P0}] + w_1 \cdot s[\mathbf{x} + \mathbf{m}_{P1}, 2k + 2r_{P1} + 2])$$

$$\mathbf{U}(s[\mathbf{x}, 2k]) = \frac{1}{4} (w_0 \cdot h[\mathbf{x} + \mathbf{m}_{U0}, k + r_{U0}] + w_1 \cdot h[\mathbf{x} + \mathbf{m}_{U1}, k - r_{U1} - 1])$$

w_0	w_1	Transform	Prediction/Update
1	1	5/3 spline wavelet	Bi-directional prediction/update
2	0	Haar wavelet	Uni-directional prediction/update
0	2		
0	0	Identity transform	Intra mode

MCTF: Motion-Compensated Prediction

- Block-adaptive switching between the lifting representations of
 - the Haar filter (uni-directional prediction) and
 - the 5/3 filter (bi-directional prediction)
- Multi-reference picture capabilities
- Intra Mode
- Motion-compensated prediction is **compliant to H.264/AVC**

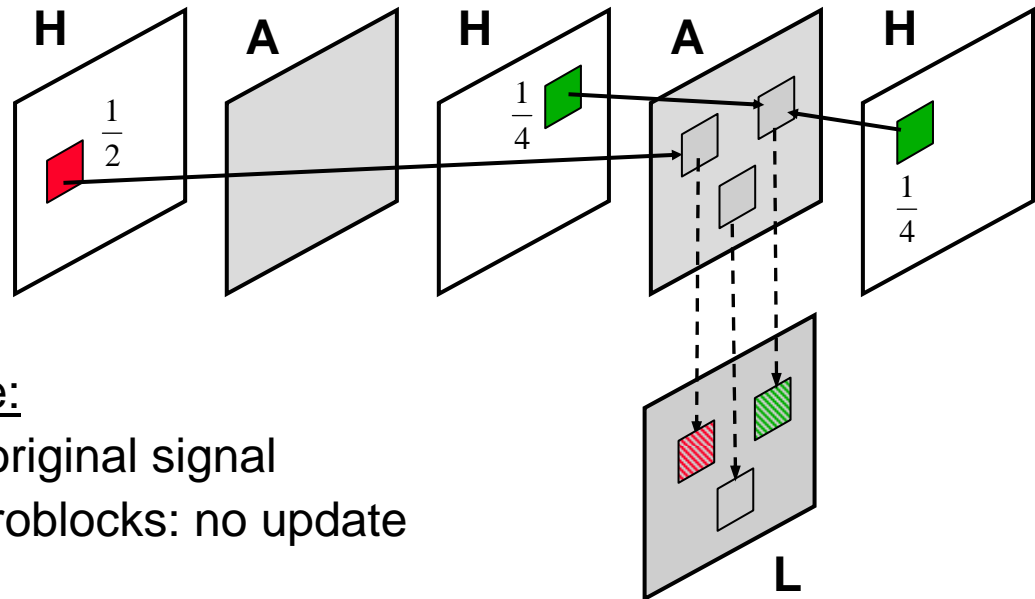


High-pass picture:

- Residual signals (prediction error)
- Intra macroblocks (directly copied)

MCTF: Motion-Compensated Update

- Update step: No directly inverse motion compensation
- Prediction data used in the update step are derived from prediction data used in the prediction step
- Uni-directional / bi-directional modes
- Multiple reference picture, Intra mode (no update)
- Motion-compensated update is **compliant to H.264/AVC**

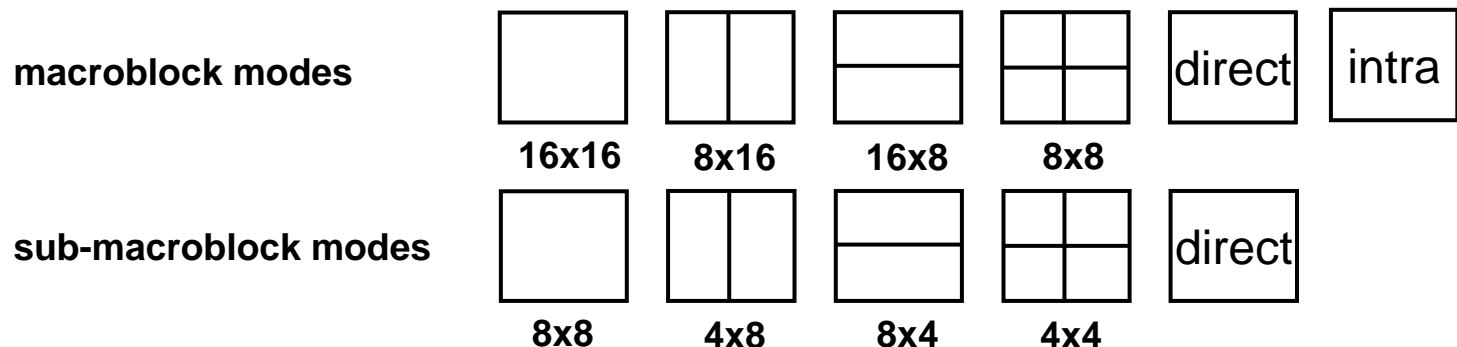


Low-pass picture:

- Updated original signal
- Intra macroblocks: no update

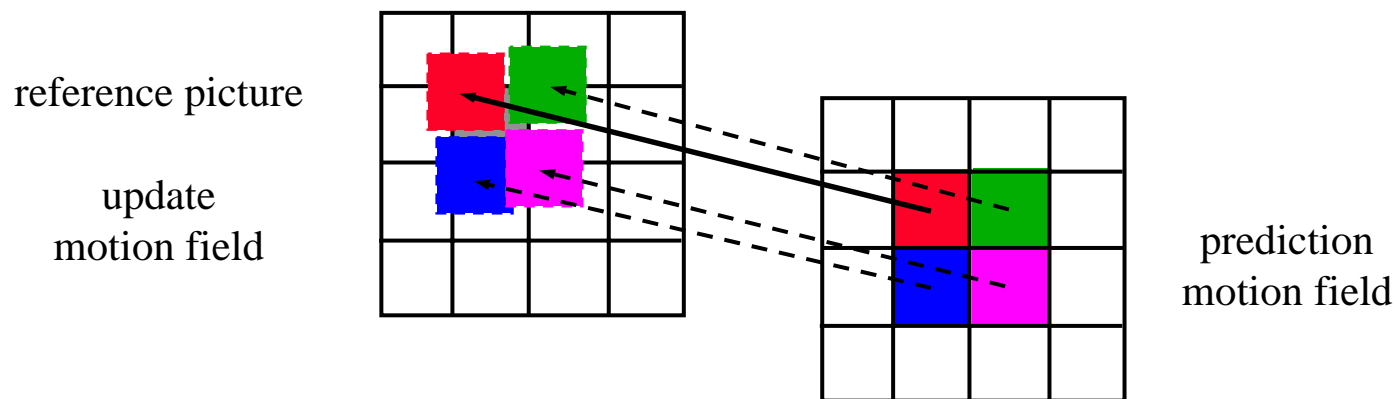
Motion / Prediction Description

- **Subset of the B slice syntax of H.264/AVC**
 - Reference pictures and prediction direction/method (L0, L1, or bi-predictive) can be chosen on macroblock partition level (for 16x16, 16x8, 8x16, or 8x8 blocks)
 - Direct modes: Spatial motion vector predictions
 - Only one intra macroblock mode



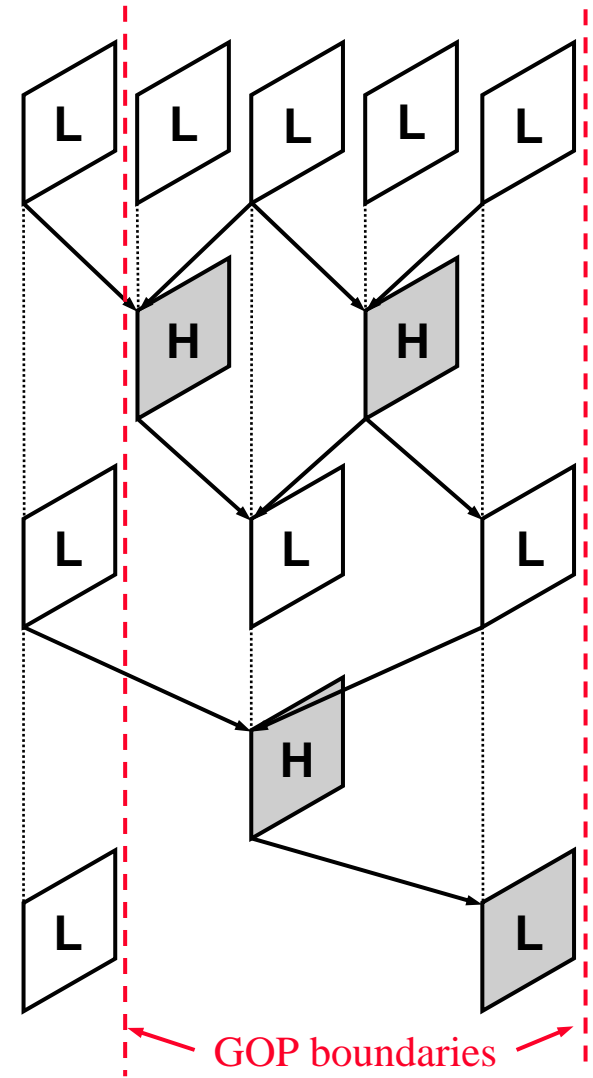
Derivation of Update Prediction Data

- **Principle** (based on 4x4 blocks)
 - For each 4x4 block of the update motion field
 - For each 4x4 block of the prediction motion fields
 - Store motion vector and number of covered samples
 - Update covered samples if motion vector was already stored
 - Choose inverse of the motion vector for which the largest number of covered samples was counted
 - Determine reference indices, (sub-)macroblock modes
 - Chose intra if number of covered samples is less than a threshold



Temporal Decomposition Structure

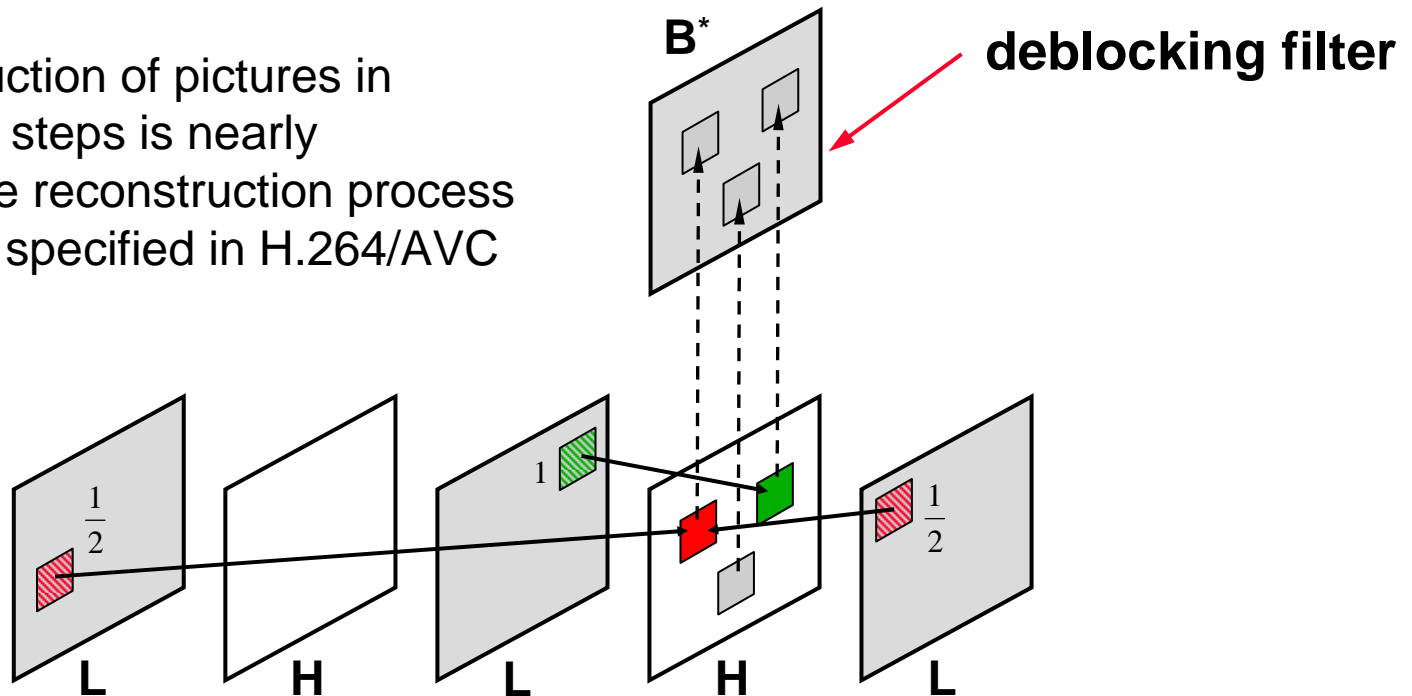
- Cascading of two-channel decompositions
- GOP structure with prediction across GOP boundaries
 - Reduction / prevention of temporal blocking
 - Disturbing subjective artifacts with independent GOP's
 - Smaller encoding-decoding delay as continuously decomposition
 - Delay 2^N frames (N decomposition stages)
 - Continuous filtering: $3(2^N-1)$ frames



Deblocking at the Decoder Side

- Deblocking filter is only applied to pictures that are reconstructed in the prediction steps at the decoder side
- Deblocking filter process is **compliant with H.264/AVC**

The reconstruction of pictures in the prediction steps is nearly identical to the reconstruction process of B slices as specified in H.264/AVC

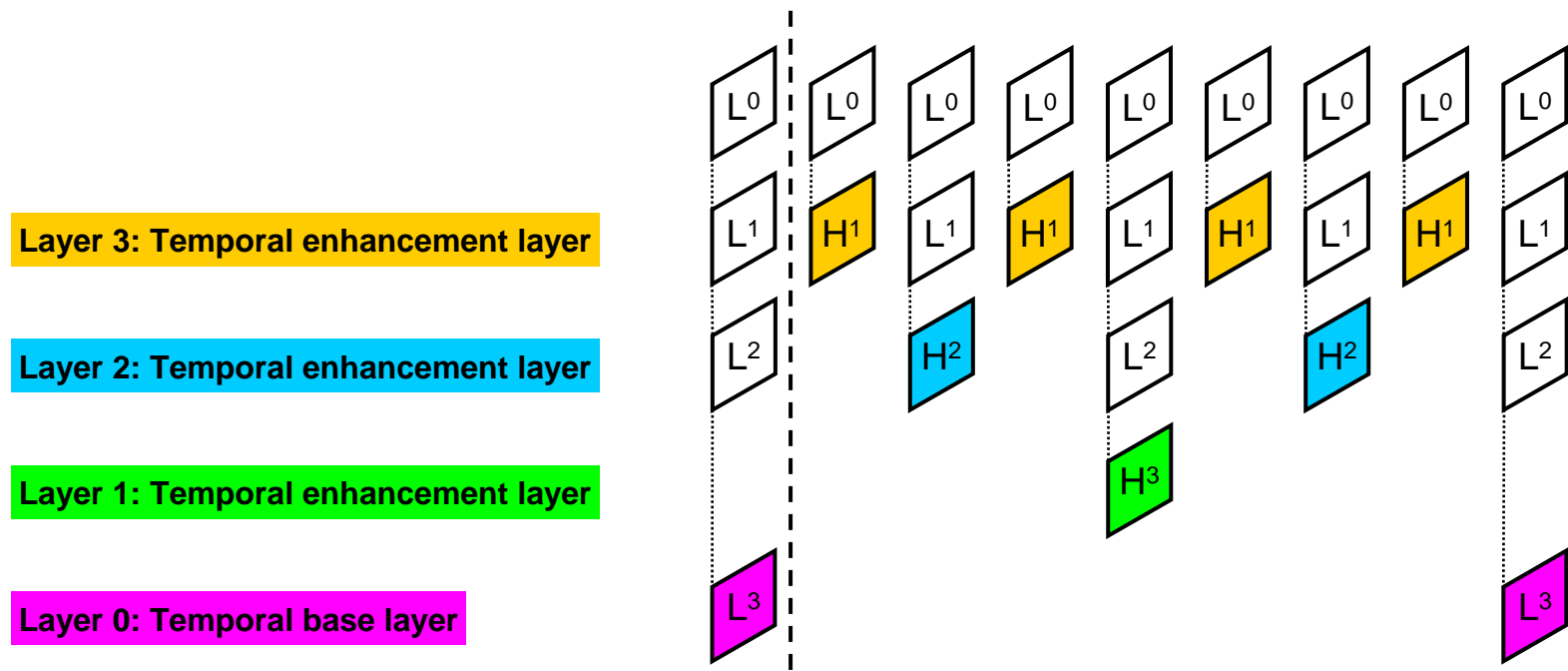


Coding of Subband Pictures

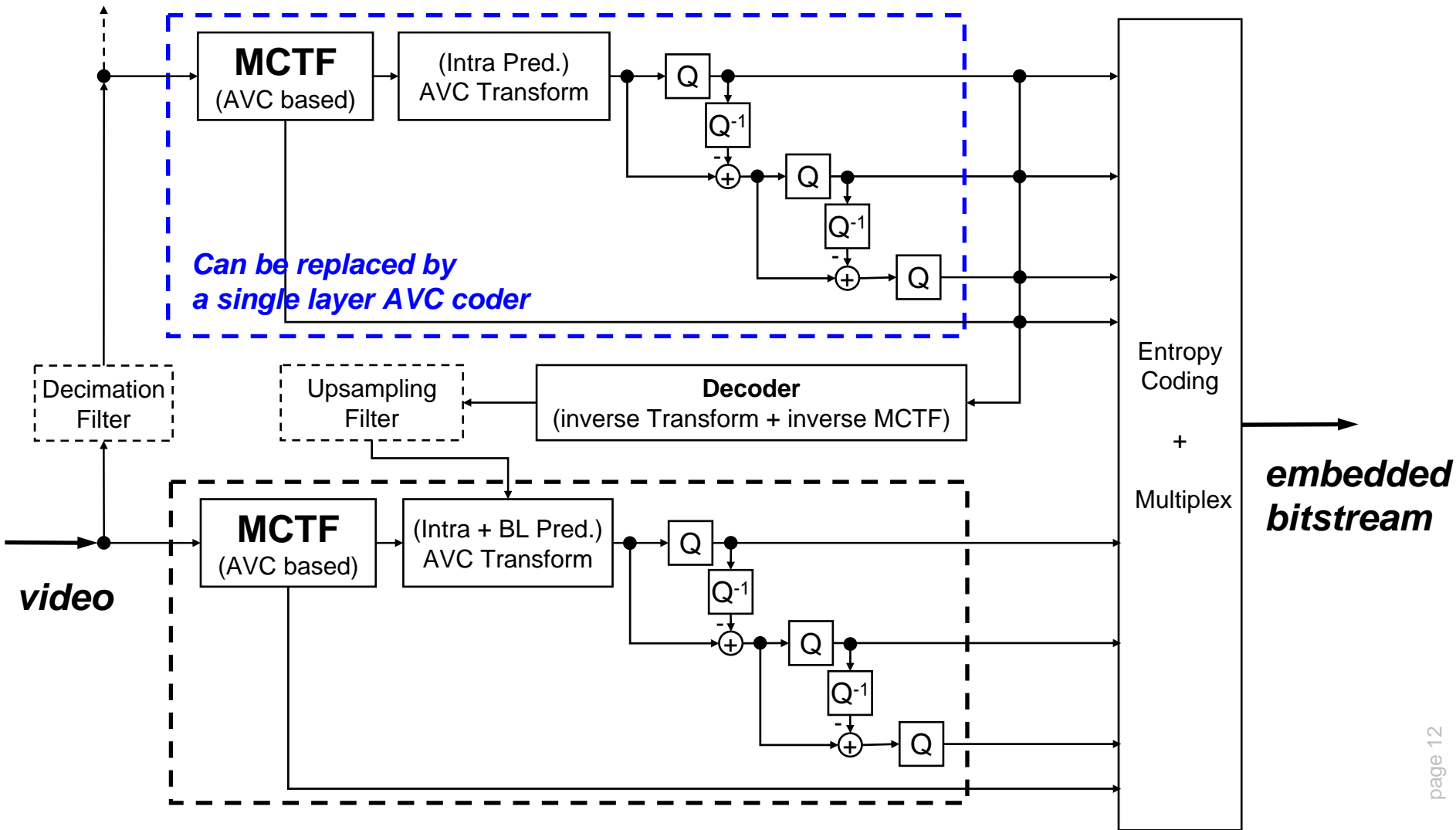
- **Prediction data (motion information)**
 - Subset of the **H.264/AVC slice layer syntax**
 - Macroblock mode, sub-macroblock modes, reference indices, motion vector differences
- **High-pass pictures**
 - Residual macroblocks: **Residual syntax of H.264/AVC**
 - Intra macroblocks: **Intra macroblock modes of H.264/AVC**
- **Low-pass pictures**
 - **Syntax of H.264/AVC**
 - Random access points: IDR (Intra) pictures in regular intervals
 - General: Prediction between successive low-pass pictures is allowed (P or B pictures / slices)

Temporal Scalability

- The MCTF framework inherently provides temporal scalability
- A similar degree of temporal scalability can be realized with standard H.264/AVC coding
 - Same coding structure without update step

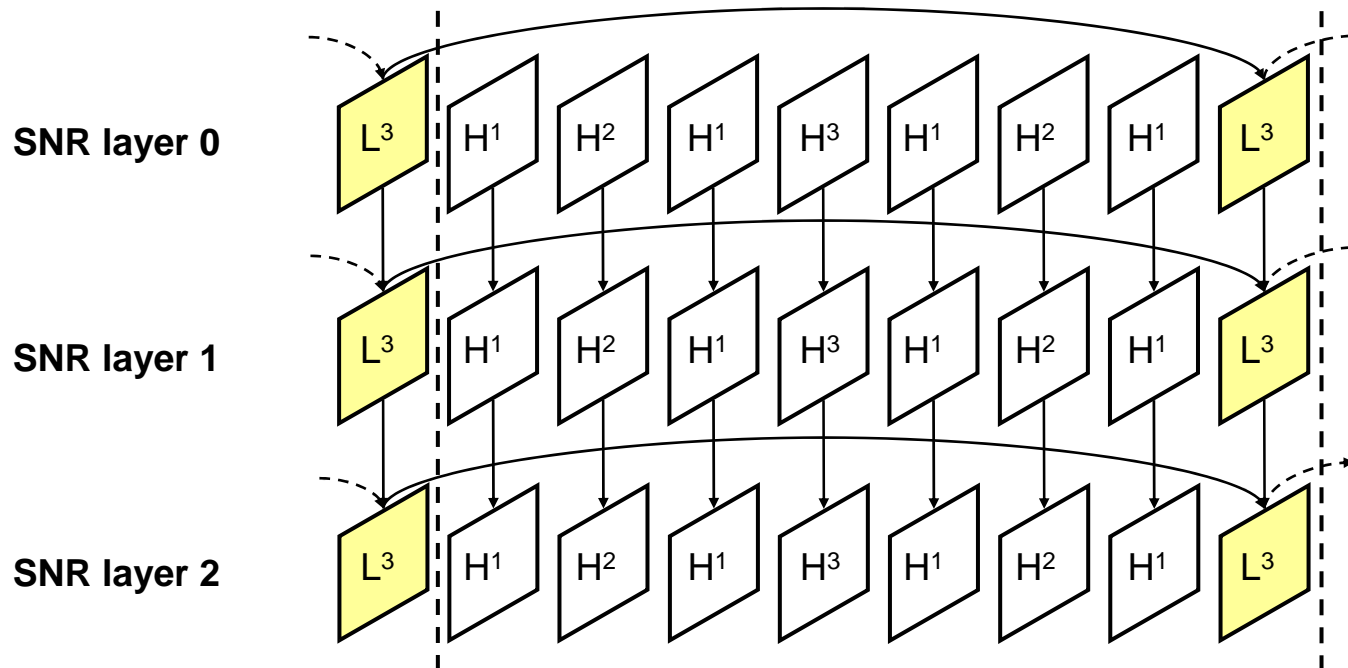


SNR and Spatial Scalability: Block Diagram



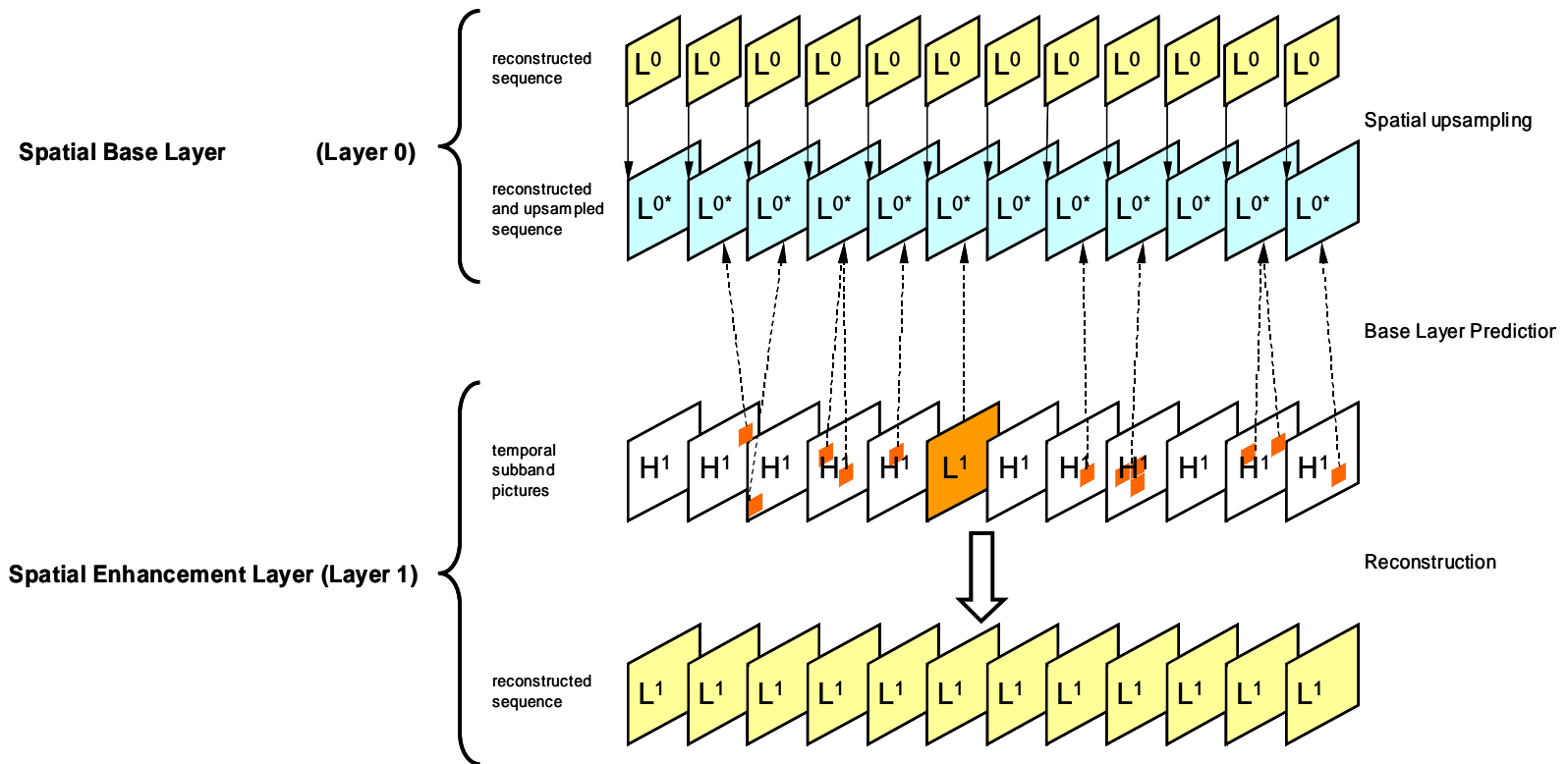
Basic SNR Scalability

- SNR scalability is provided by a layered representation of the subband pictures
- Macroblocks of high-pass pictures are predicted from the subordinate SNR layer (residual syntax of H.264)
- Low-pass pictures are coded using the H.264/AVC syntax with one additional intra macroblock mode (using base layer prediction)



Spatial and Enhanced SNR Scalability 1

- Additional prediction of intra macroblocks from the reconstructed pictures of the lower resolution spatial layer (low- and high-pass pictures)
- Interpolation using the H.264/AVC half-sample filter: $(1, -5, 20, 20, -5, 1) \gg 5$
- H.264/AVC can be used as spatial base layer

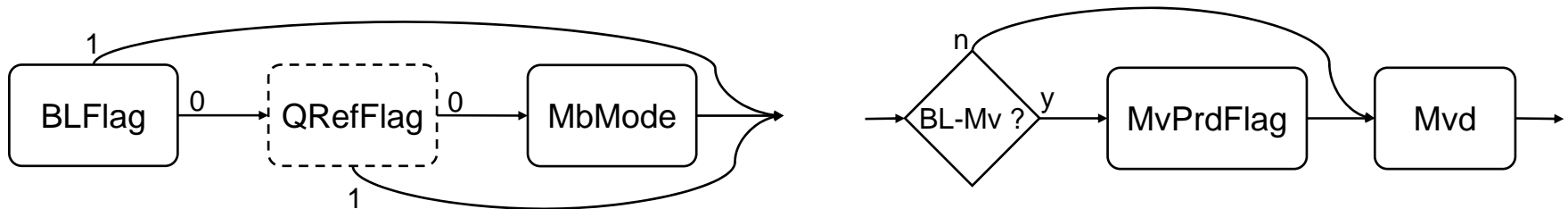


Spatial and Enhanced SNR Scalability 2

- **When base layer is coded using the MCTF extension of H.264/AVC**

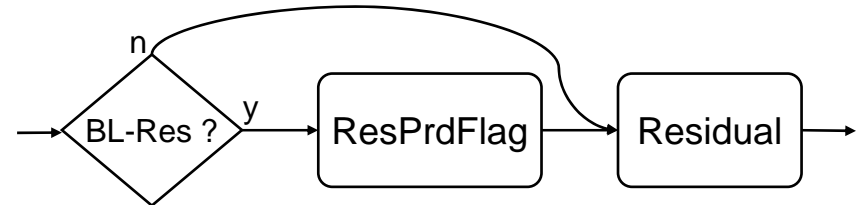
- Prediction of motion field / prediction information

- Base layer mode: Use (scaled) motion information of base layer
- Refinement mode: Base layer mode with quarter-pel refinement
- Switch between spatial and base layer motion vector predictor



- Prediction of (upsampled) prediction residuals

- Signaled by a flag per macroblock

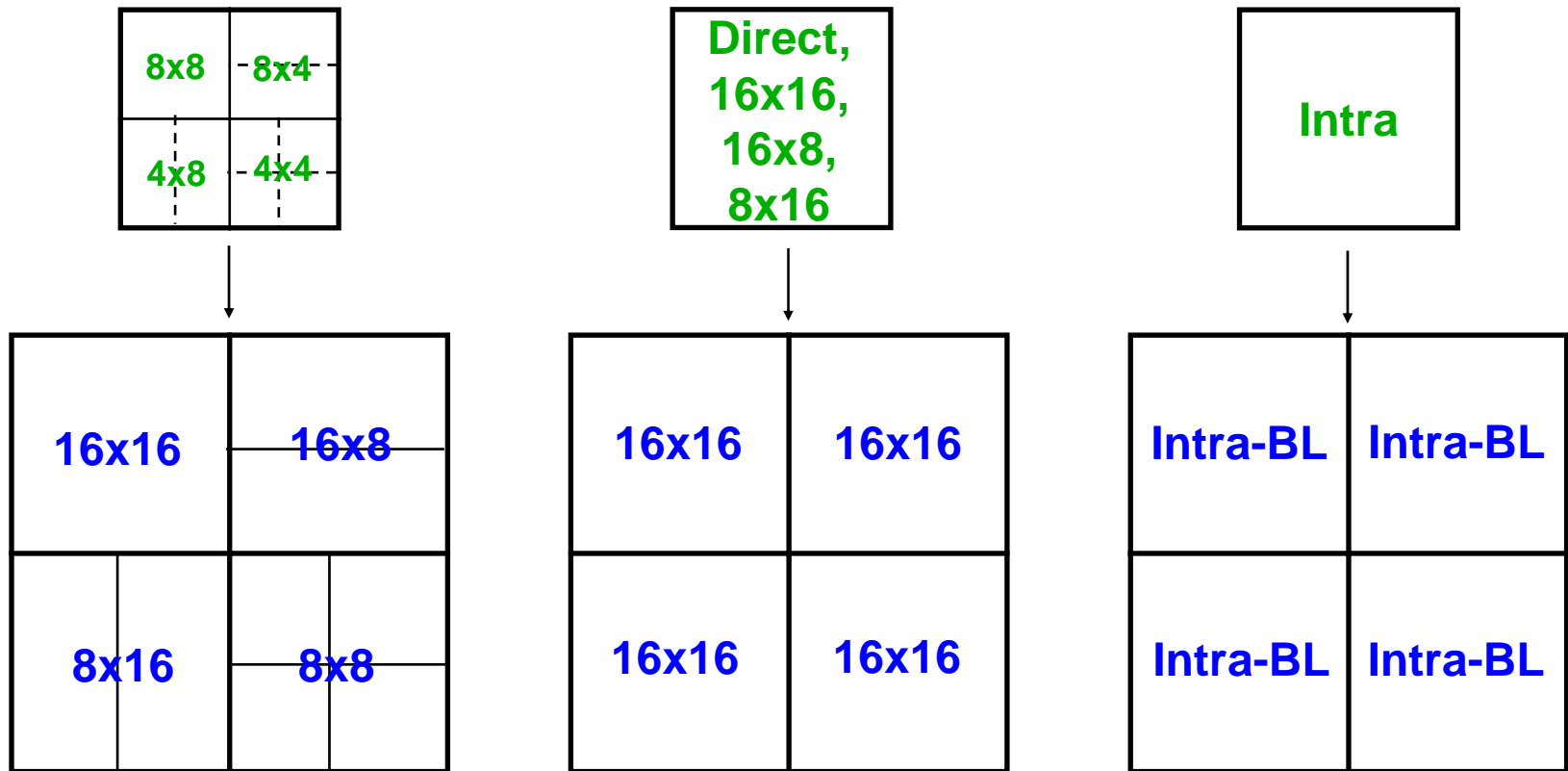


- Generalized spatial / SNR scalability

- General SNR scalability without upsampling and scaling of motion vectors

Upsampling of Motion / Prediction Data

- Upsample partitioning
- Use prediction method (list 0, list 1, bi-predictive) and reference indices
- Scaling of motion vectors by a factor of 2



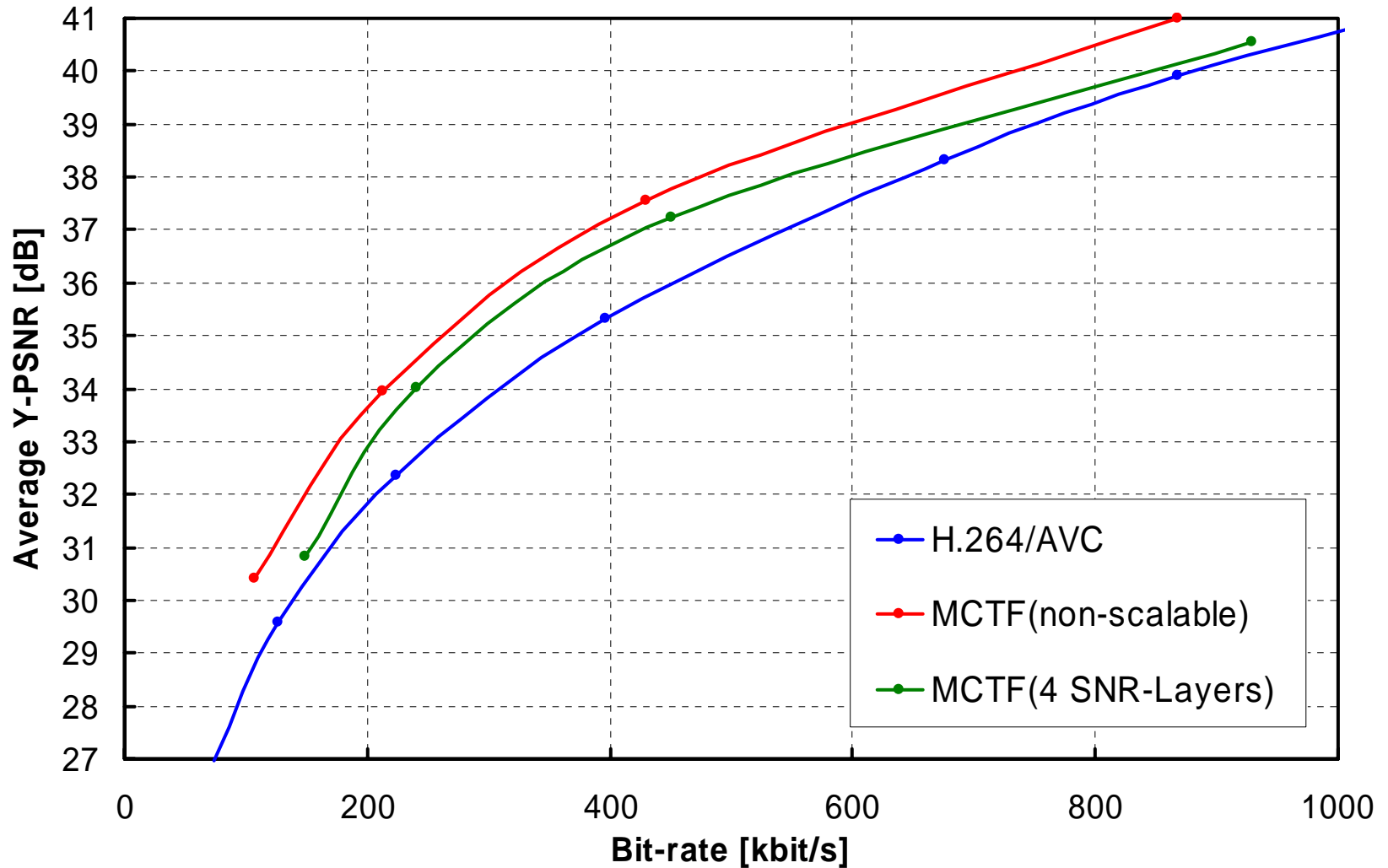
Coding Efficiency

Experiment 1

- Single layer coding
- SNR scalability only (4 layers)
- Reference: H.264/AVC with 2 B frames & CABAC

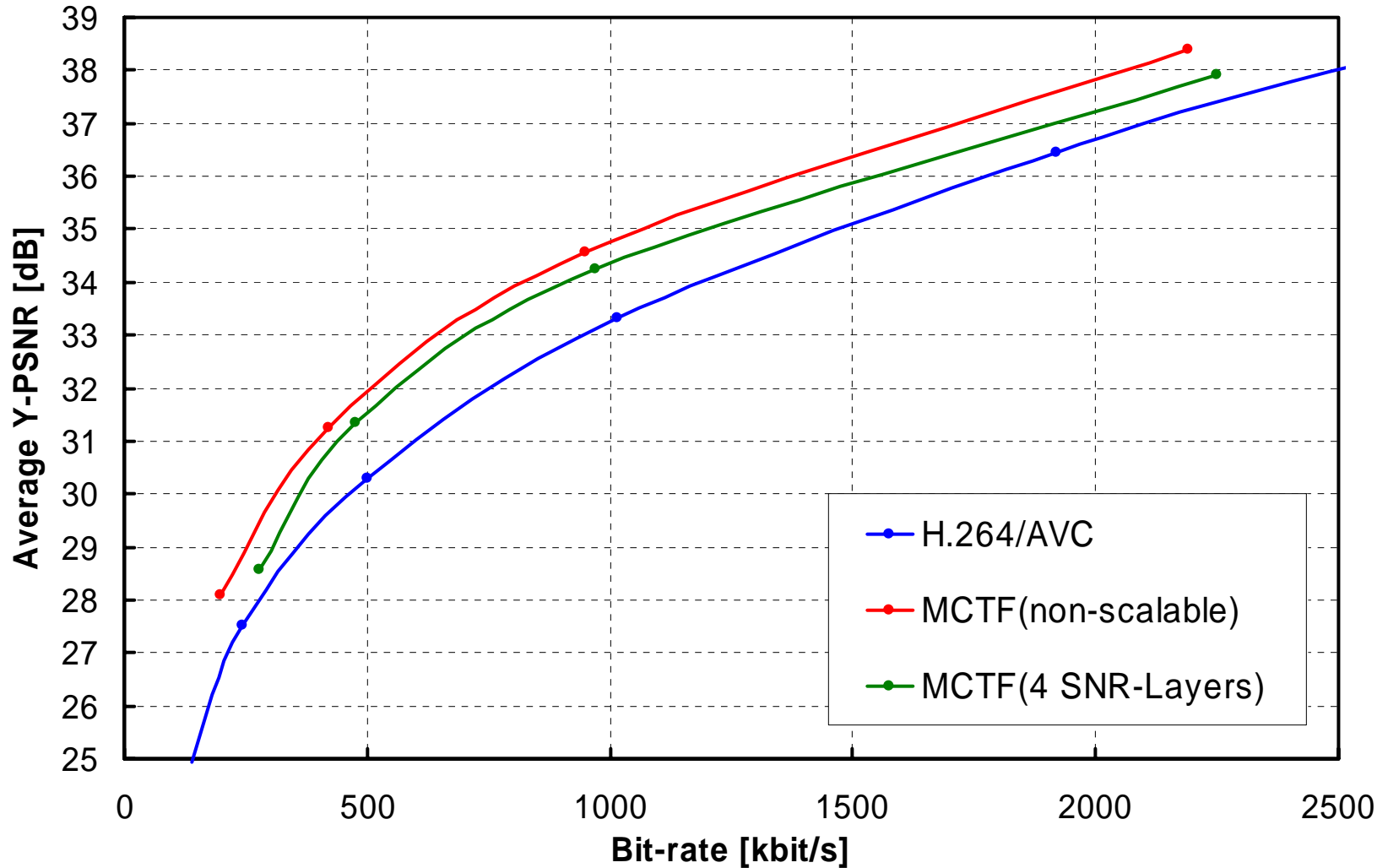
Coding Efficiency: Paris

Paris (CIF, 30Hz, 300 frames)



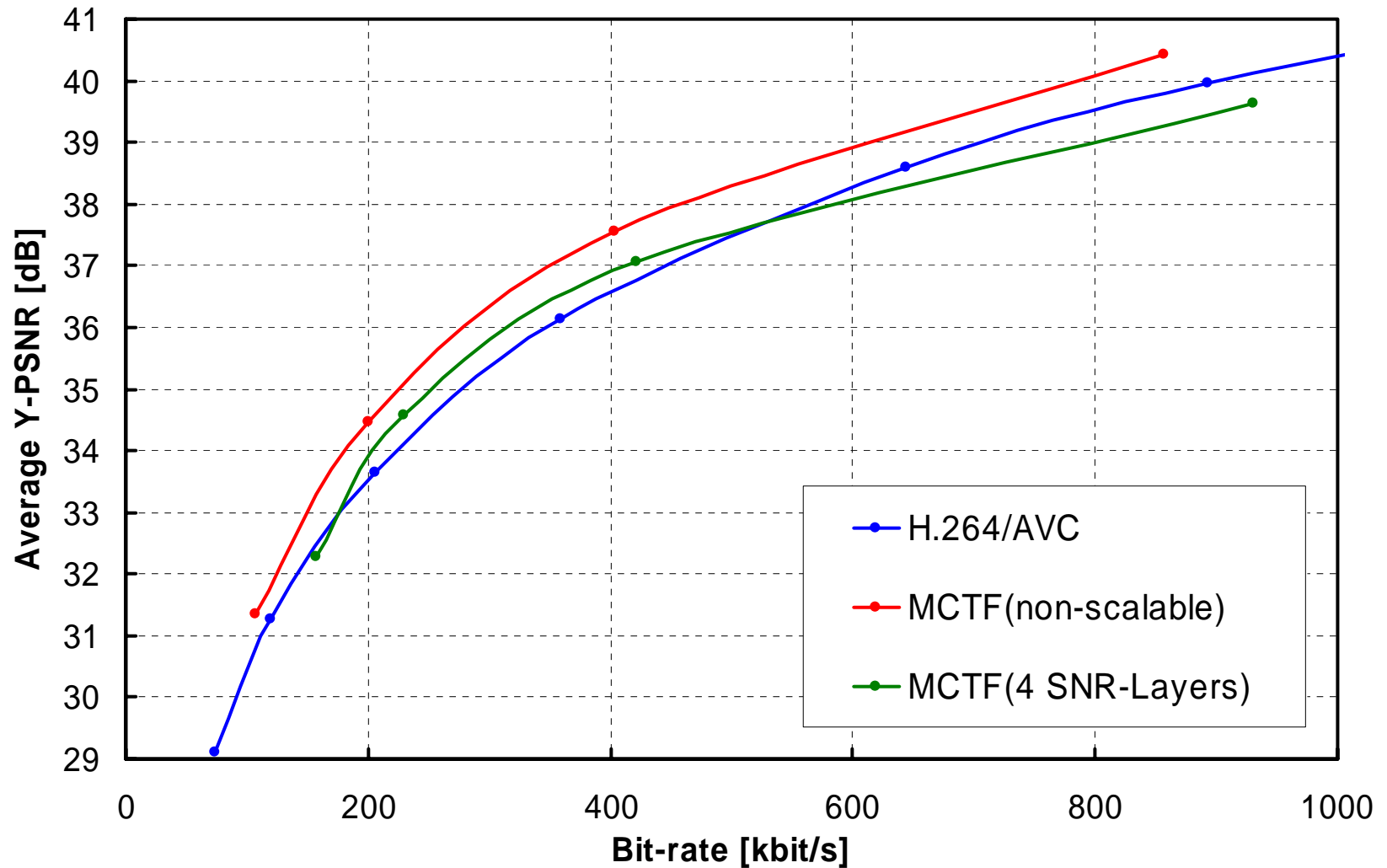
Coding Efficiency: Mobile

Mobile (CIF, 30Hz, 300 frames)



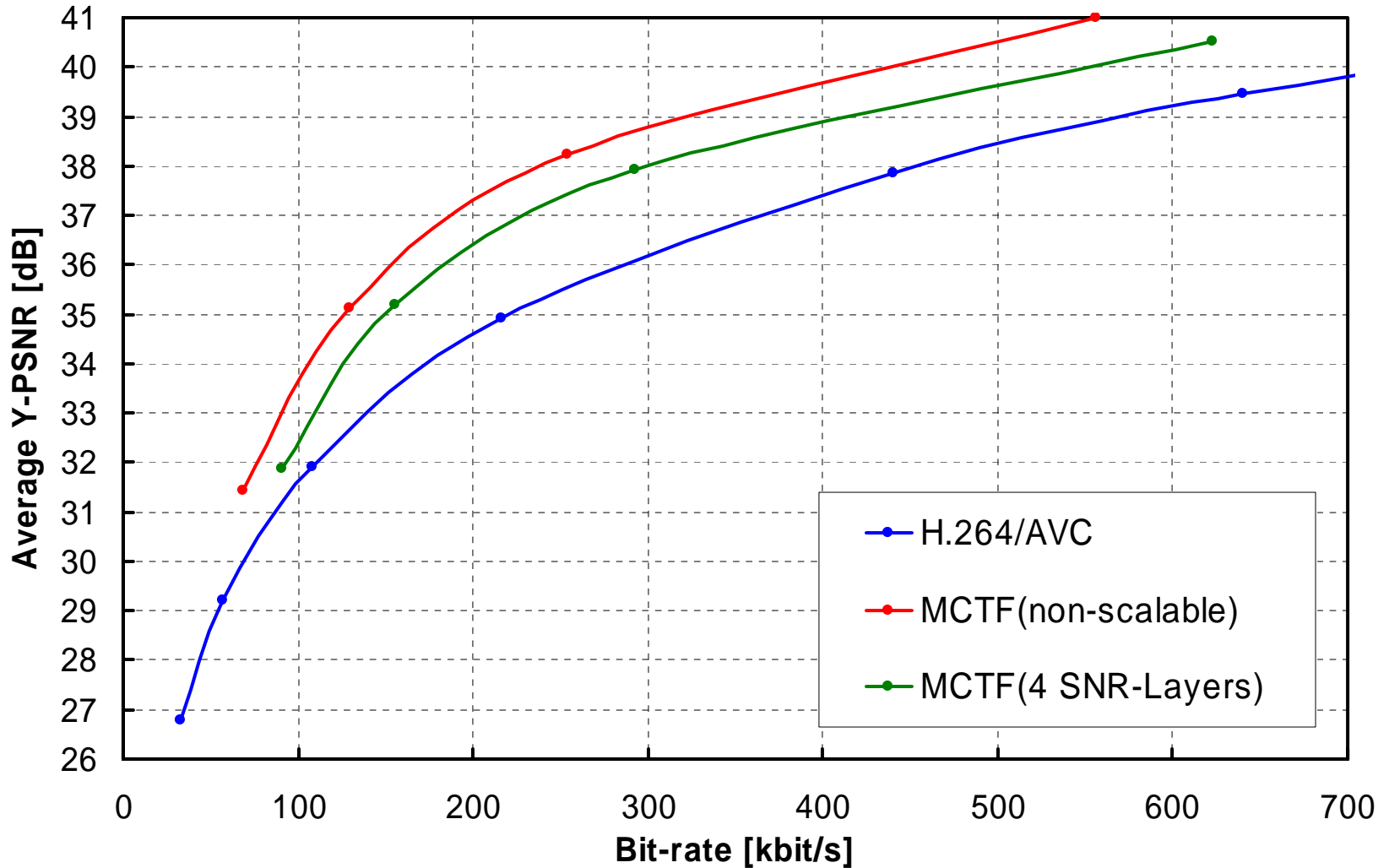
Coding Efficiency: Foreman

Foreman (CIF, 30Hz, 300 frames)



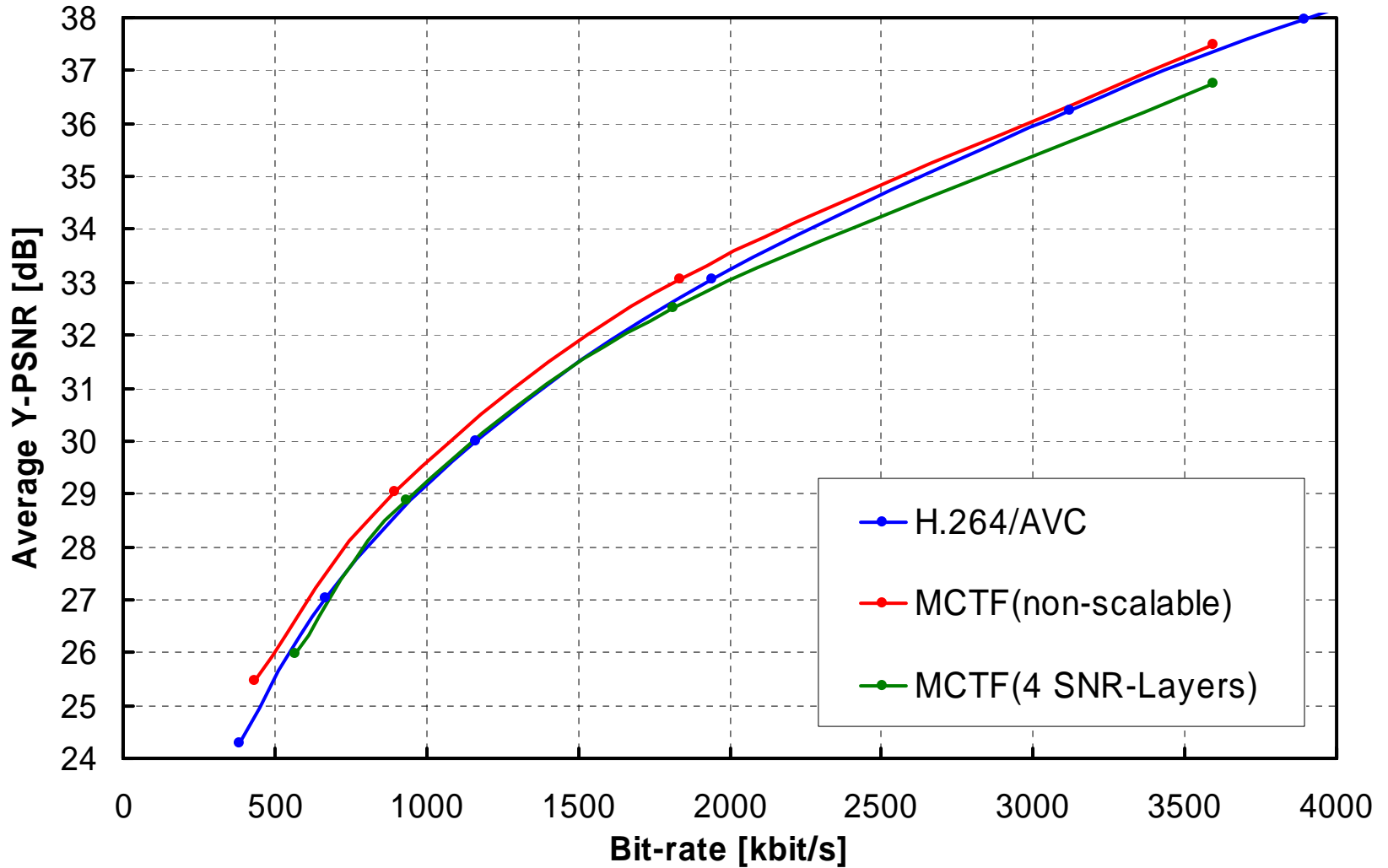
Coding Efficiency: Stockholm

Stockholm (CIF, 30Hz, 250 frames)



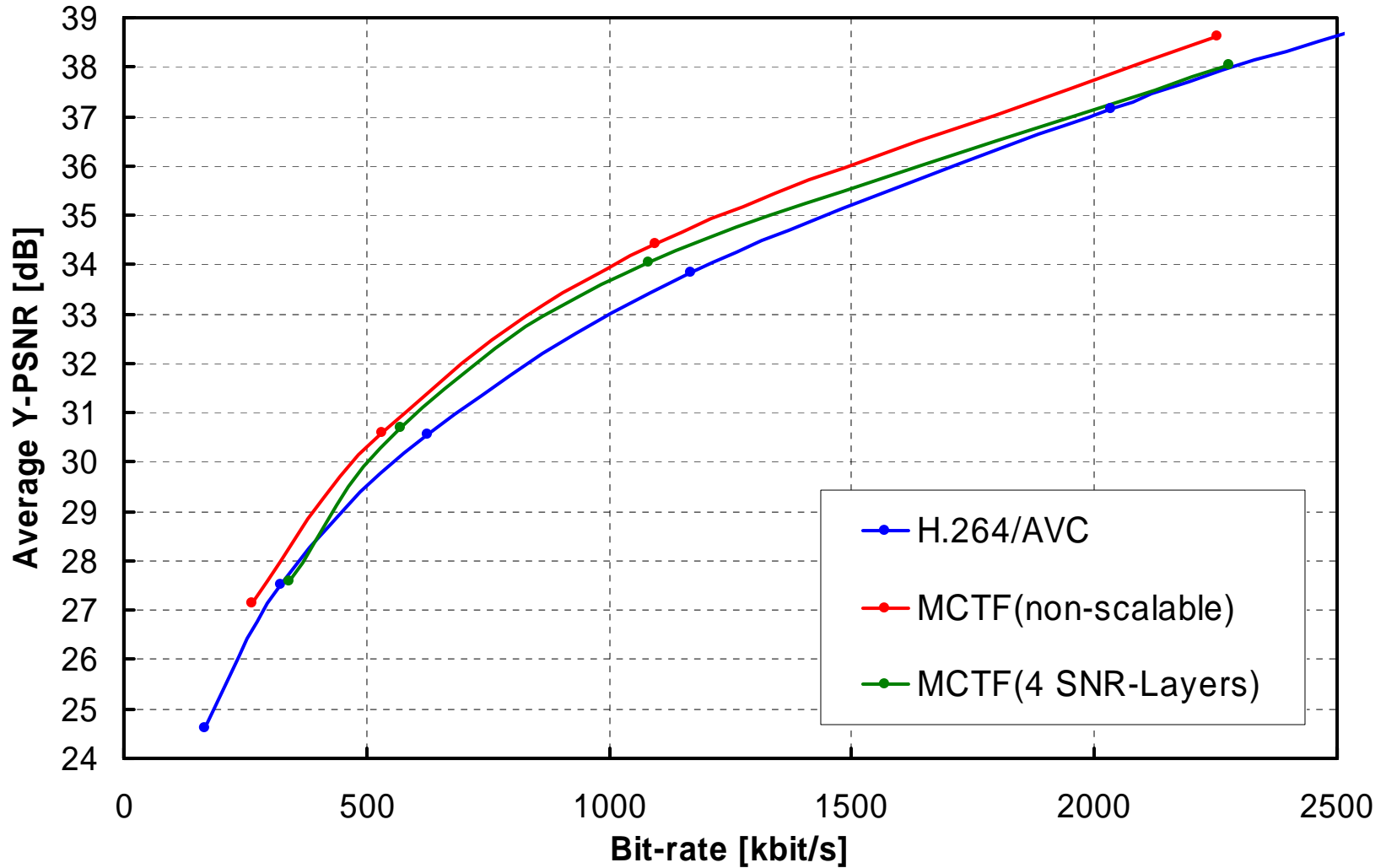
Coding Efficiency: Basketball

Basket (CIF, 30Hz, 249 frames)



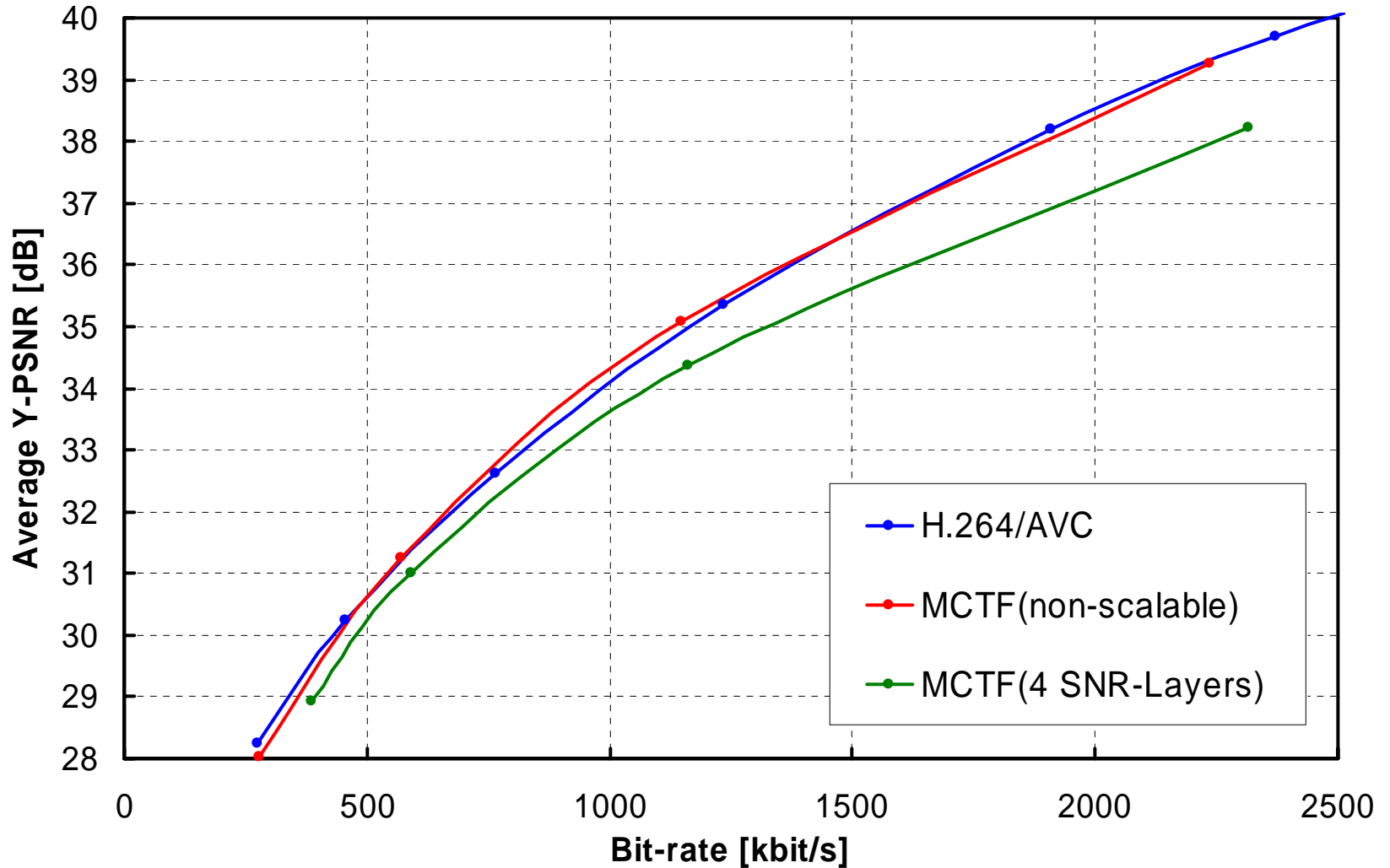
Coding Efficiency: Flowergarden

Flowergarden (CIF, 30Hz, 250 frames)



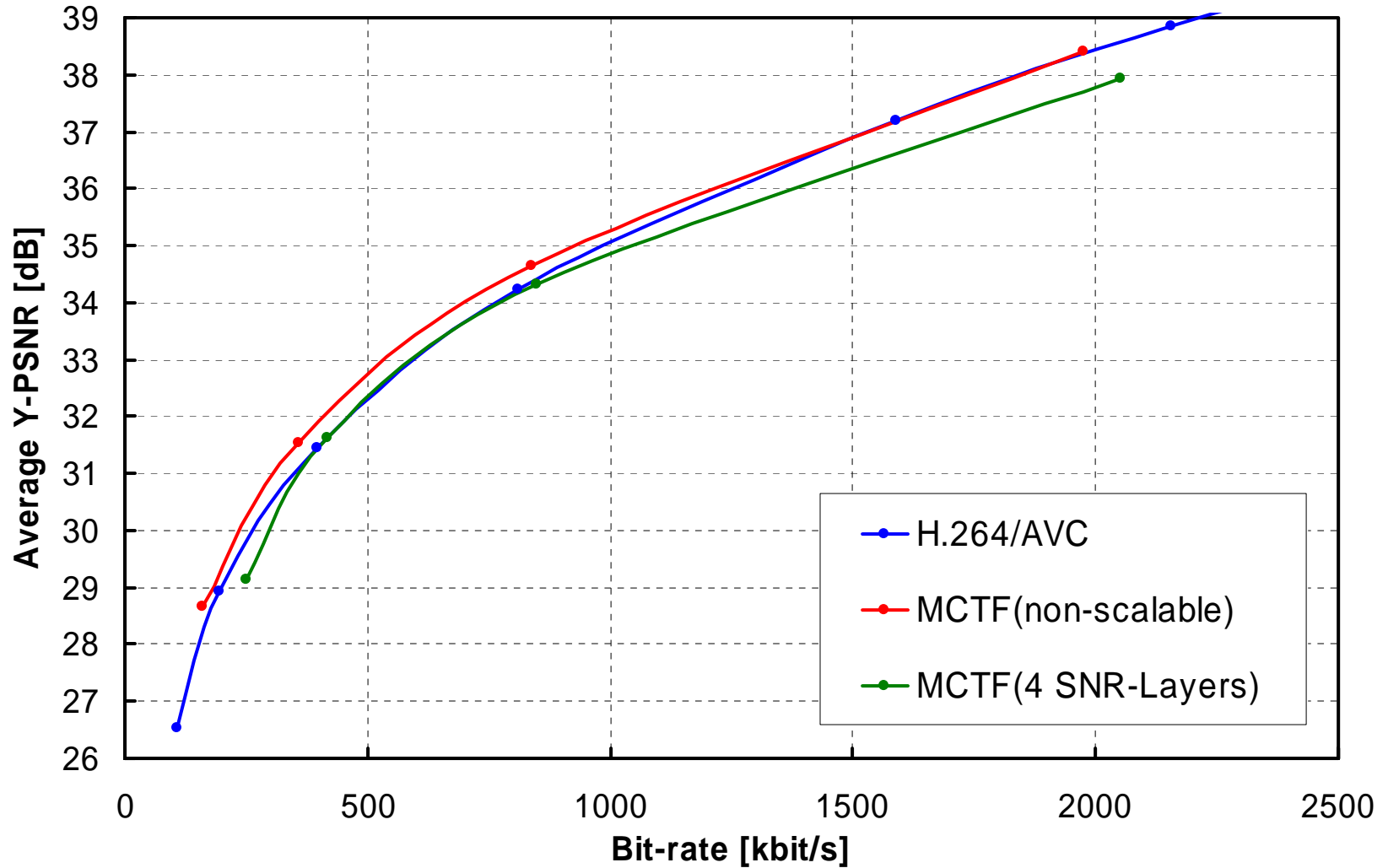
Coding Efficiency: Football

Football (CIF, 30Hz, 250 frames)



Coding Efficiency: Tempete

Tempete (CIF, 30Hz, 260 frames)

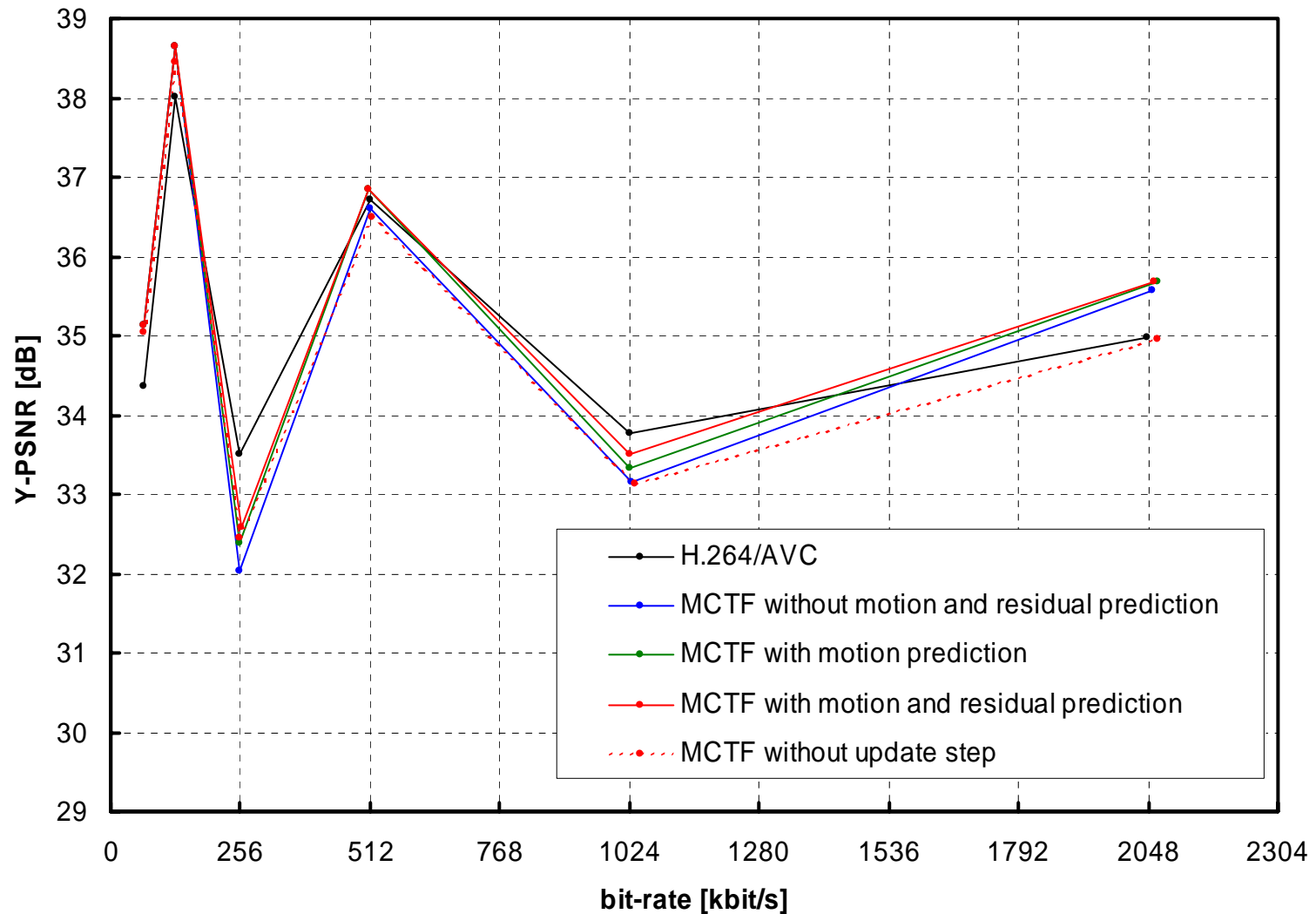


Coding Efficiency

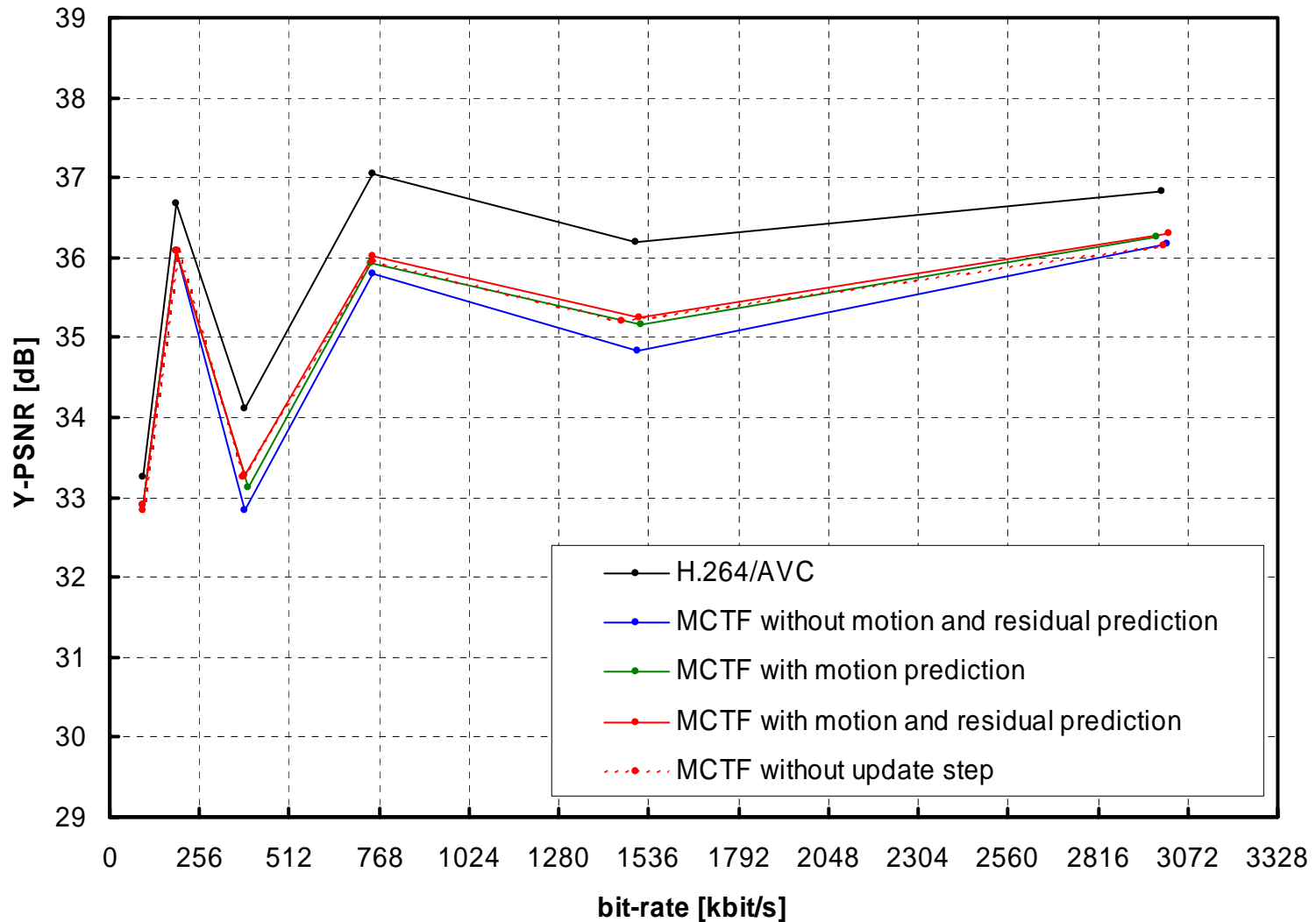
Experiment 2

- Combined spatial, SNR, and temporal scalability
- Sequences and test conditions as defined in the SVC Core Experiment
 - Test 1: 2 x QCIF 15Hz, 2 x CIF 30Hz, 4CIF 30Hz, 4CIF 60Hz
 - Test 2: QCIF 7.5Hz, QCIF 15Hz, 2 x CIF 15Hz, CIF 30 Hz
- Reference: H.264/AVC anchors

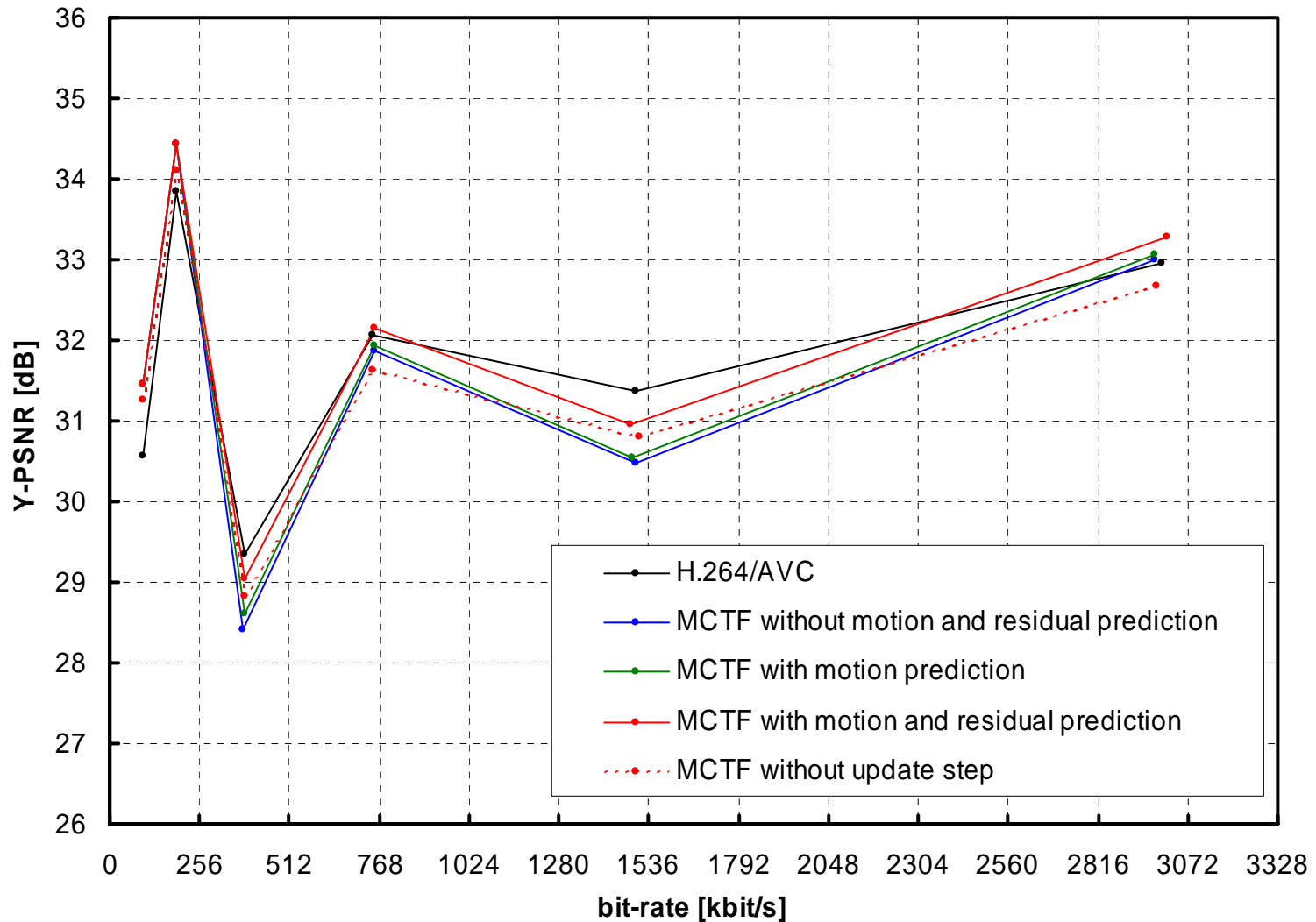
Coding Efficiency: City



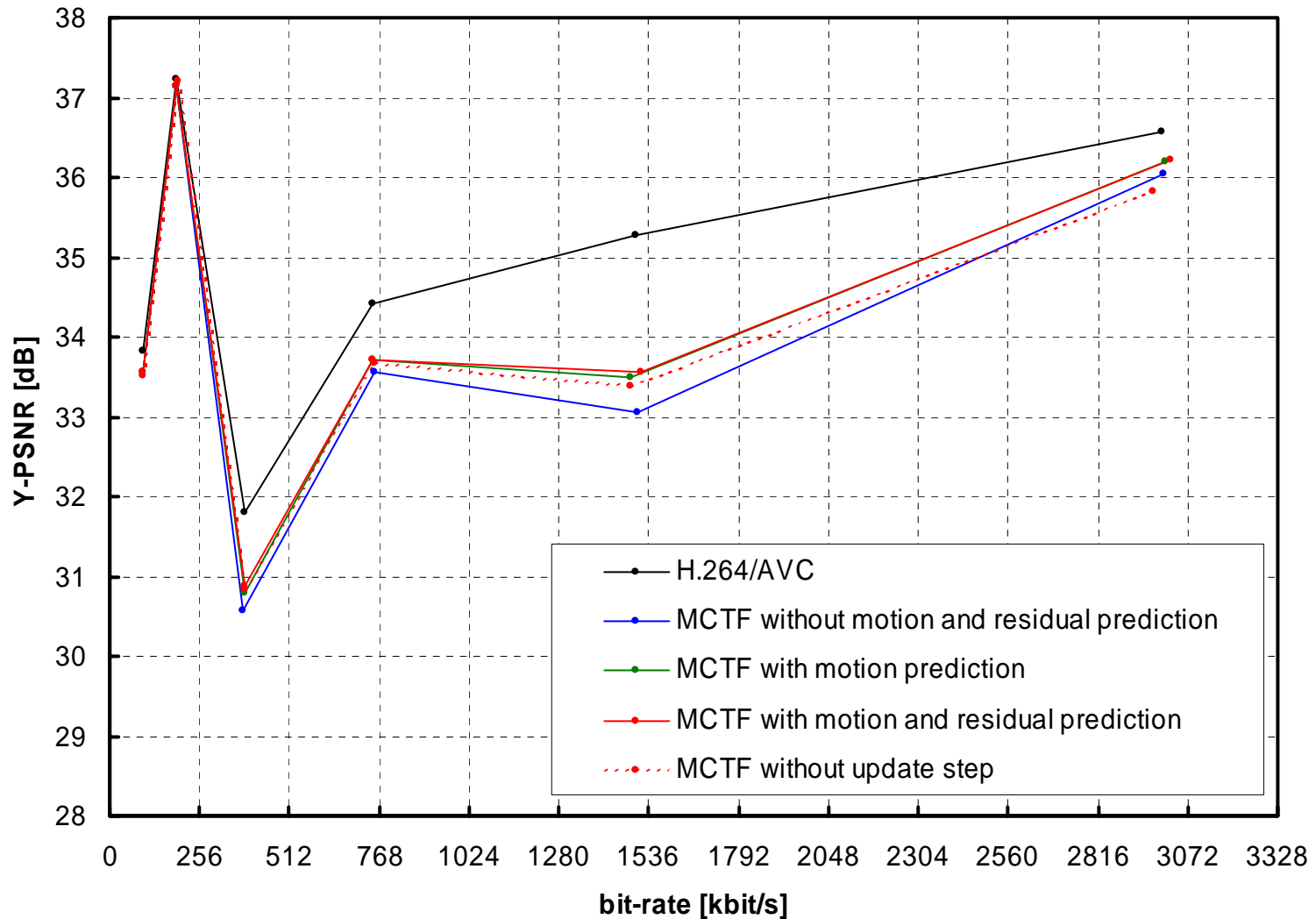
Coding Efficiency: Crew



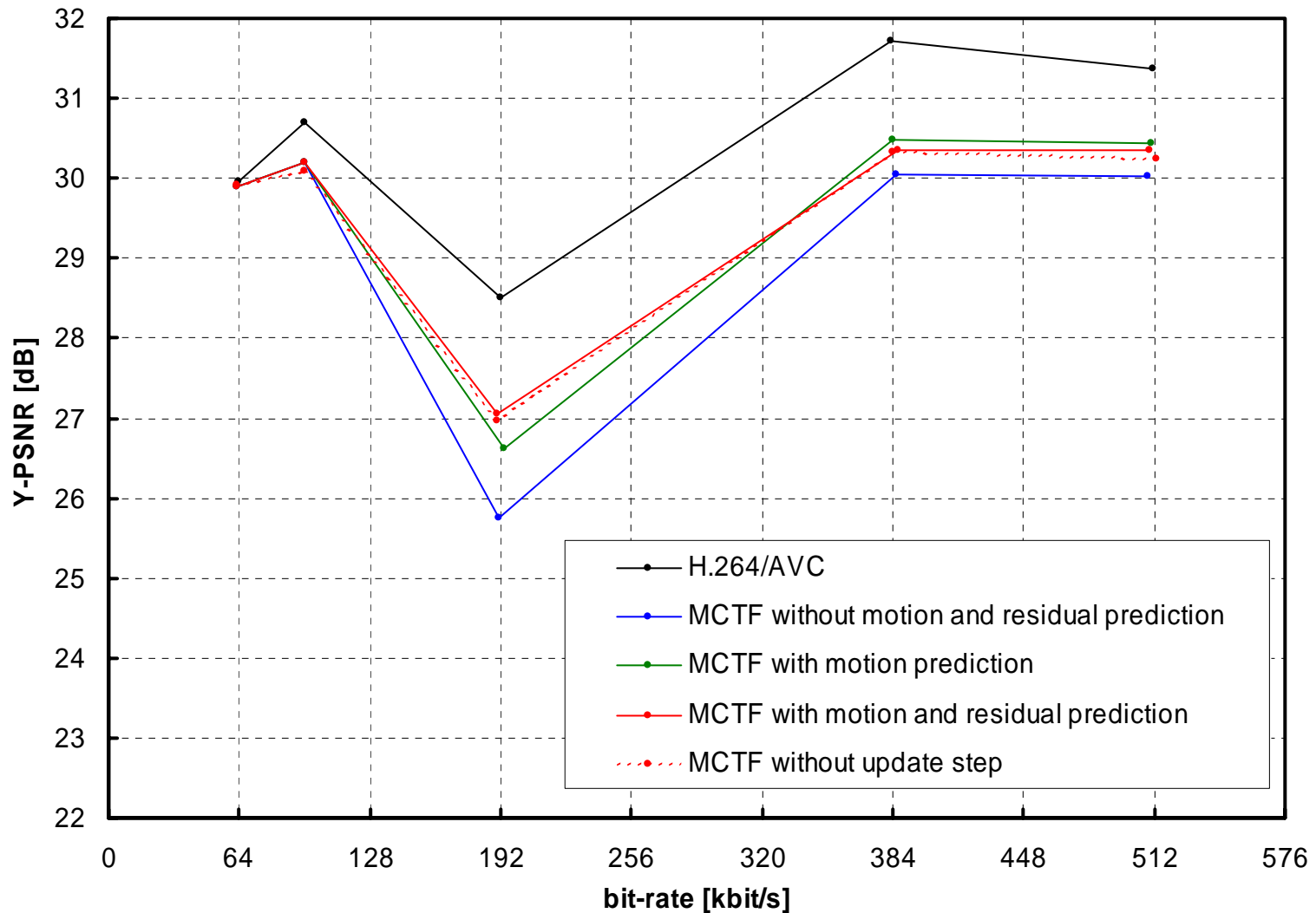
Coding Efficiency: Harbour



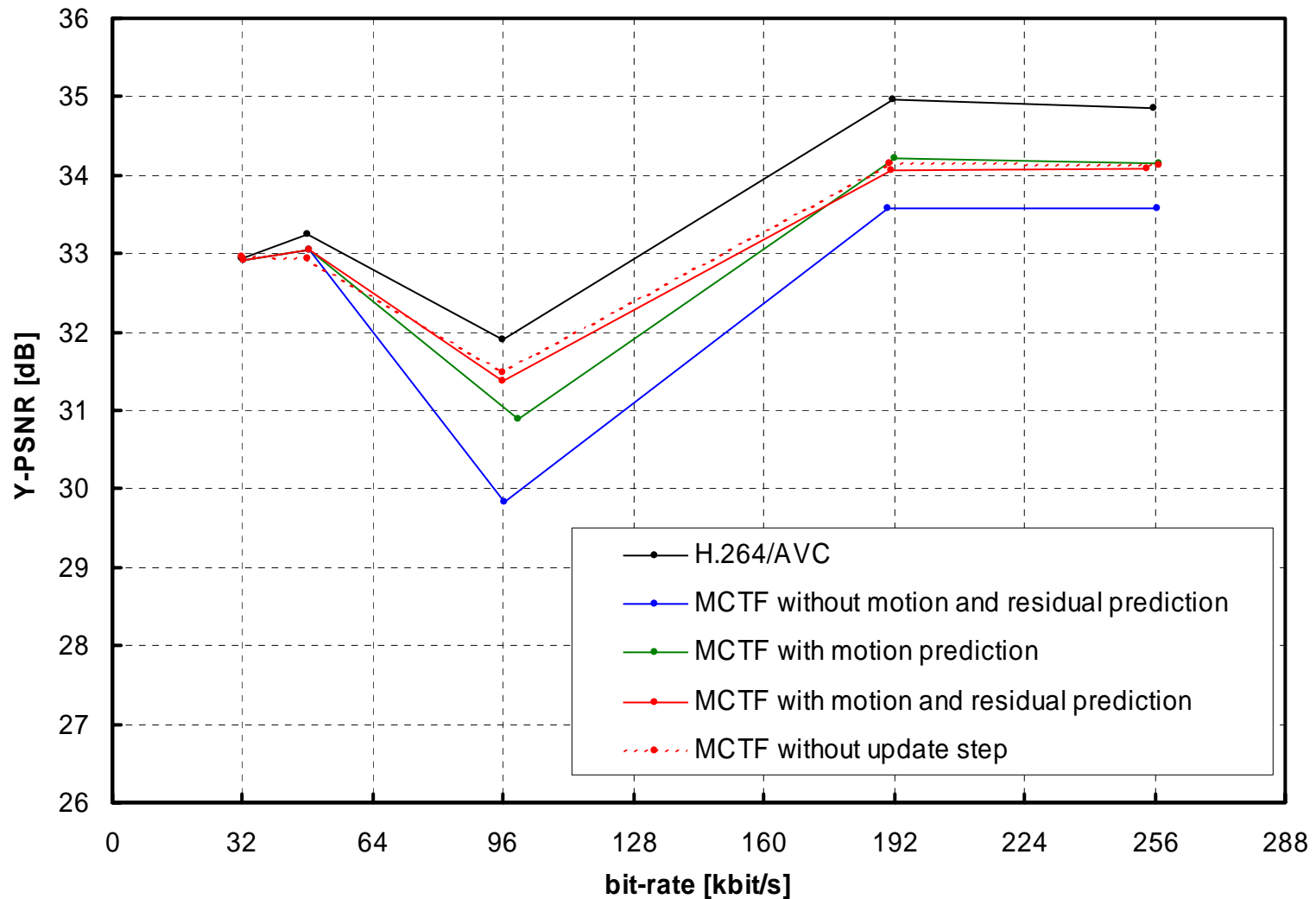
Coding Efficiency: Soccer



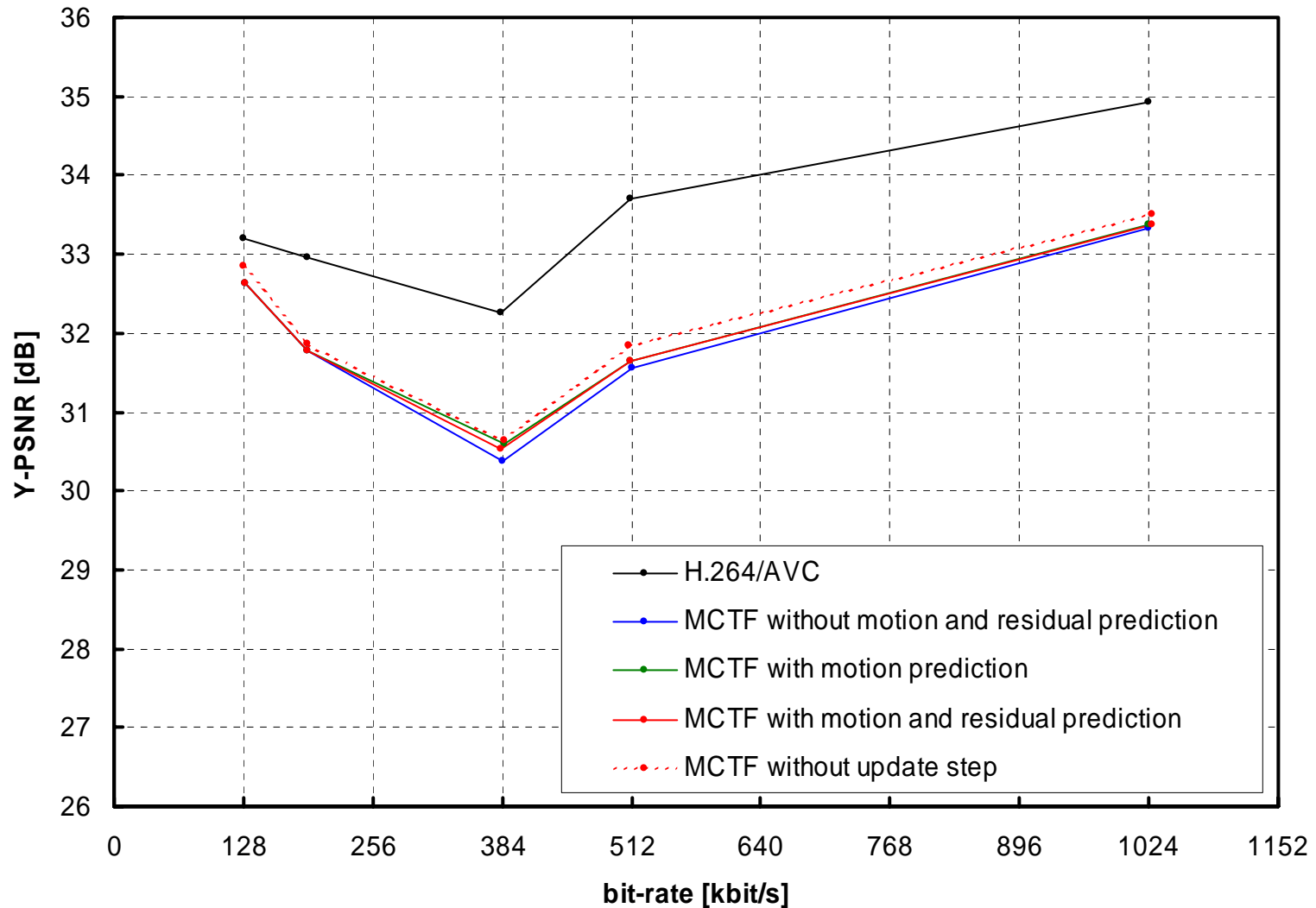
Coding Efficiency: Bus



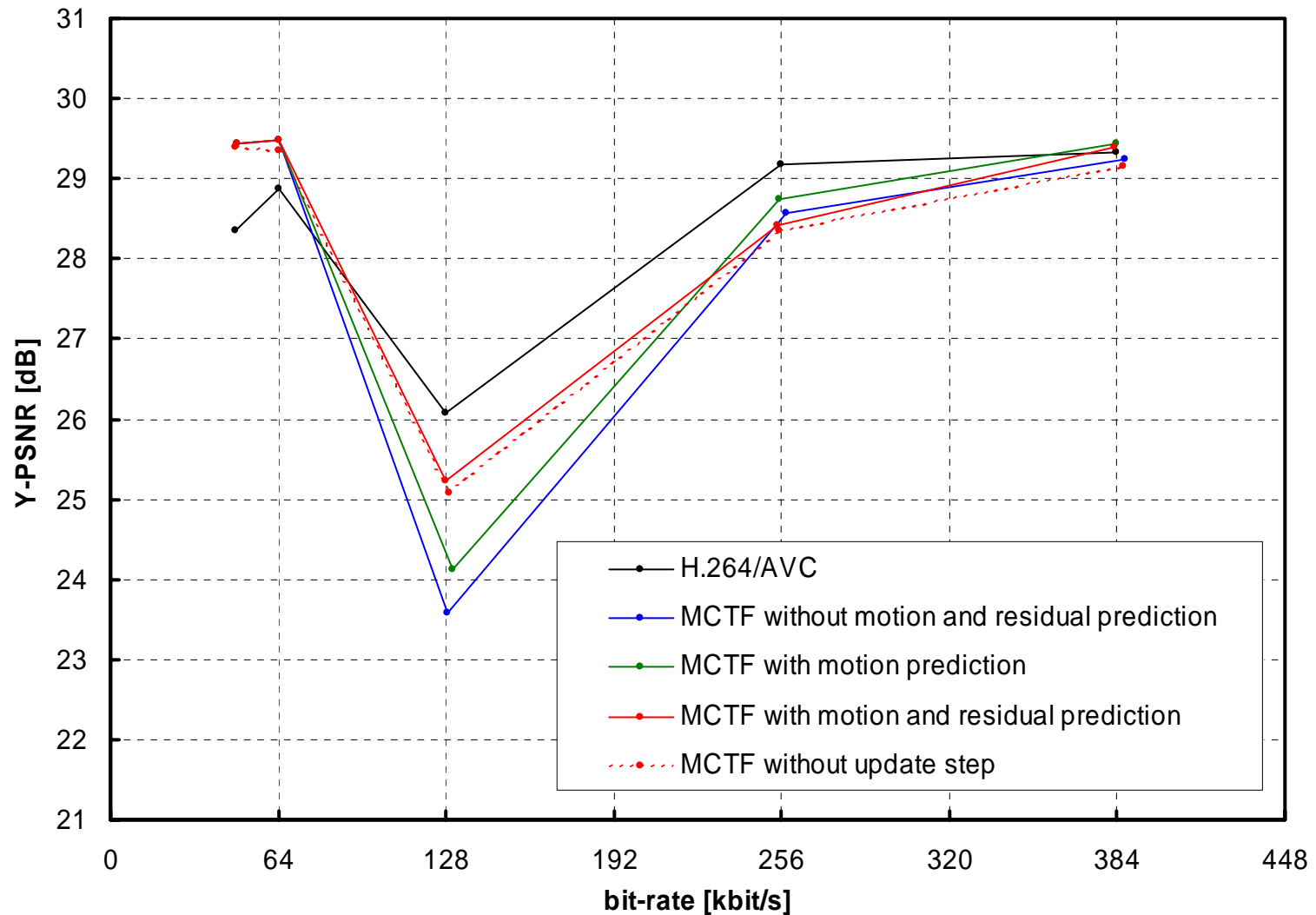
Coding Efficiency: Foreman



Coding Efficiency: Football



Coding Efficiency: Mobile



Summary

- **Subband extension of H.264/AVC**
 - Most components are used as specified in the standard
 - Motion compensation within the MCTF framework
 - Intra prediction, transform coding, syntax, entropy coding
 - H.264/AVC can be used as base layer
- **Flexible combined scalability**
 - Spatial scalability with arbitrary number of levels
 - Temporal scalability with arbitrary number of levels
 - SNR scalability with arbitrary number of level
- **Coding efficiency**
 - Dependent on the application requirements
 - Single layer version can provide a coding efficiency superior to the state-of-the-art H.264/AVC standard