Scalable Video Coding in H.264/AVC

1. **Introduction – Potentials and Applications**
2. **Scalability Extension of H.264/AVC**
   - 2.1 Scalability Operation and High-Level Syntax
   - 2.2 Temporal Scalability
   - 2.3 SNR/Fidelity/Quality Scalability
   - 2.4 Spatial Scalability
3. **Profiles, Levels, and Systems Support**

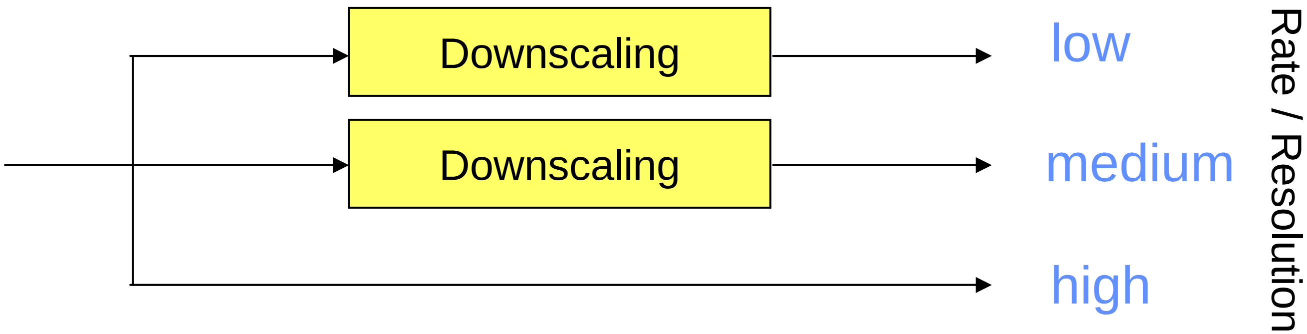
Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

1. Introduction –
Potentials and Applications

Fraunhofer Institut
Nachrichtentechnik
Heinrich-Hertz-Institut

Scalability

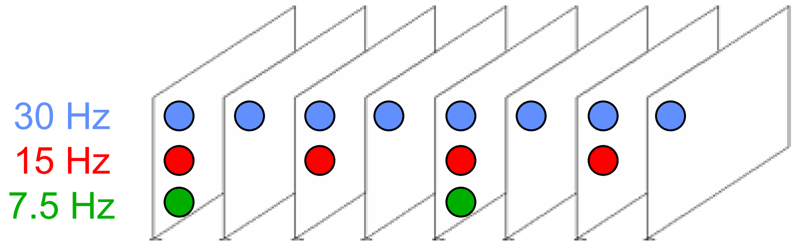
- **Basic approach : provide video signal with different scales / rate levels**



- **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 scalability)

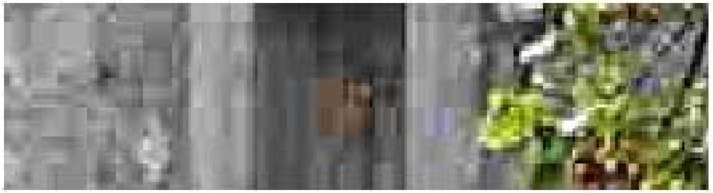
Scalability of Video - Modalities

- Temporal: change of frame rate



- Spatial: change of frame size

- Fidelity: change of quality (e.g. SNR)



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Scalability of Video - Standardization

- Video coding standards since 1994 have included tools for scalable coding
  - MPEG-2 = H.262: *S*, *T* (2 layers), *F*
  - H.263+: *S*, *T* (2 layers), *F*
  - MPEG-4 Part 2 Visual: *S*, *T*, *F* (FGS), *O*
- But*:* except for *T*, this had to be paid by a tremendous additional rate cost and was never accepted by the market

Latest Standardization of SVC

- **April 2004**: **MPEG Call for Proposals on Scalable Video Coding Technology**
  ⇨ 16 submissions: 12 Wavelet-based,
  4 Extensions of H.264/MPEG4-AVC
- **Until October 2004**: **Competitive Phase**
  ⇨ 2 Candidates: MSRA proposal (Wavelet-based)
  HHI proposal (Extension of H.264/MPEG4-AVC)
- **October 2004**: HHI proposal adopted as starting point
- **October 2004 – January 2005**: Discussion of Continuation of Joint Video Team (JVT) between **VCEG** and **MPEG**
  ⇨ Joint Standards Project – **SVC (H.264 / AVC AMD 2)**
- **JVT Chairs**: Jens-Rainer Ohm (RWTH Aachen), Gary Sullivan (Microsoft), Ajay Luthra (Motorola), and Thomas Wiegand (HHI)
- **JD-1: January 2005**
- **Final Standard (phase 1)**: **July 2007**
- **SVC for bit-depth and color formats (possible phase 2): 2008?**

Definition of Scalability in SVC

- H.264/AVC coding without the scalability extension is referred to as single-layer H.264/AVC coding
- Rate-distortion performance comparable to single-layer H.264/AVC means that the same visual reproduction quality is typically achieved at +/– 10% bit rate.
- ***Efficient Scalability****: a functionality for removal of parts of the bitstream while achieving an R-D performance at any supported spatial, temporal, or SNR resolution that is comparable to single-layer H.264/AVC coding at that particular resolution.*

Supported Functionalities and Potential Applications

- **One single encoding process to produce the scalable bitstream**
- **Partial decoding of the scalable bitstream allows**
  - Graceful degradation when “lower priority” parts of bitstream get lost
  - Bit rate adaptation
  - Format adaptation
  - Power adaptation
- **Potential Applications**
  - Compact representation of video signal at various resolutions allows efficient transmission and storage (upload of signal for distribution, erosion storage)
  - Unicast transmission services with varying channel conditions (throughput, errors) or device types (supported spatio-temporal resolution by decoder, display and power) – simplification of adaptation
  - Multicast or broadcast transmission services with a diversity of uncertainties as of the unicast transmission above – avoidance of simulcast

New Scalability Ecosystem

- **Prior scalable video coding standards: MPEG-2, H.263, MPEG-4 failed because of technical shortcomings and (see e.g. DVB)**
  - Graceful degradation not really supported by transmission channels
  - Bit rate adaptation not needed by fixed-throughput channels
  - Power adaptation not needed: no handheld devices
  - Format adaptation not needed: just one single format – standard definition
- **Today, the situation has changed**
  - Internet and mobile transmission are becoming primary distribution mechanisms
    - ➢ Shared resource systems with varying throughput and errors
    - ➢ *Graceful degradation, bit rate adaptation, power adaptation*
  - A variety of terminals and displays from QCIF, QVGA, VGA, SD, to HD
    - ➢ Backwards compatible extensions important
    - ➢ *Format adaptation*: QVGA ➔ VGA, 720p/1080i ➔ 1080p

SVC for Mobile TV

- **Two major technical problems in mobile TV**
  - Coverage and loss of signal
  - Backwards-compatible extension towards multiple services at different spatial resolution
- **Coverage and loss of signal**
  - Graceful degradation through temporal (if possible), fidelity (mainly), and spatial (potentially) scalability
  - Use of hierarchical modulation and unequal loss protection
- **Backwards-compatible extension**
  - Example: Introduction of QVGA enhancing VGA services through spatial scalability
  - Graceful degradation comes as by-product

Format Enhancements in Mobile TV

- **Spatial and temporal format extension**
- **Introduction of QVGA services**

- **Backwards-compatible introduction of VGA services**

SVC for Backwards-Compatible HD Format Extension

- Since January 2005, **H.264/AVC-based HD broadcasting** is being introduced into the market with millions of receivers being shipped (DirecTV, Echostar, BSkyB,...)
- The **HD formats** are either **720p50/60 or 1080i25/30**
- For **premium services** (e.g. sports), broadcasters want to broadcast **1080p50/60** while maintaining the existing customer base
- **Broadcasting of 1080p50/60 SVC bitstreams** enhances video resolution from 720p50/60 or 1080i25/30
- Provide **enhanced set-top box** with additional 1080p50/60 SVC (and H.264/AVC) capability to premium customers without the need to exchange recently shipped set-top box

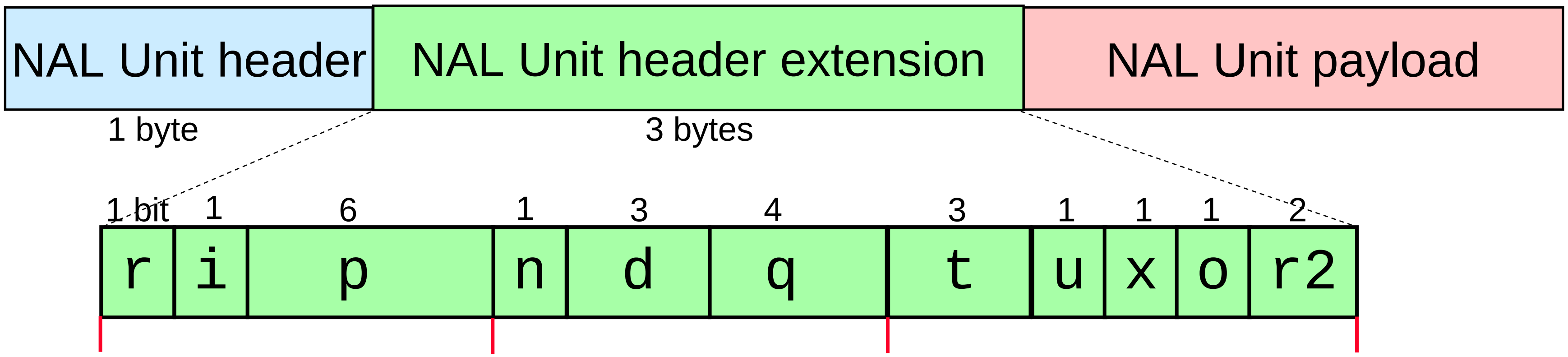
2.. Scalabe Video Coding in H.264 / AVC

Fraunhofer Institut
Nachrichtentechnik
Heinrich-Hertz-Institut

2.1 Scalability operation and high-level syntax

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

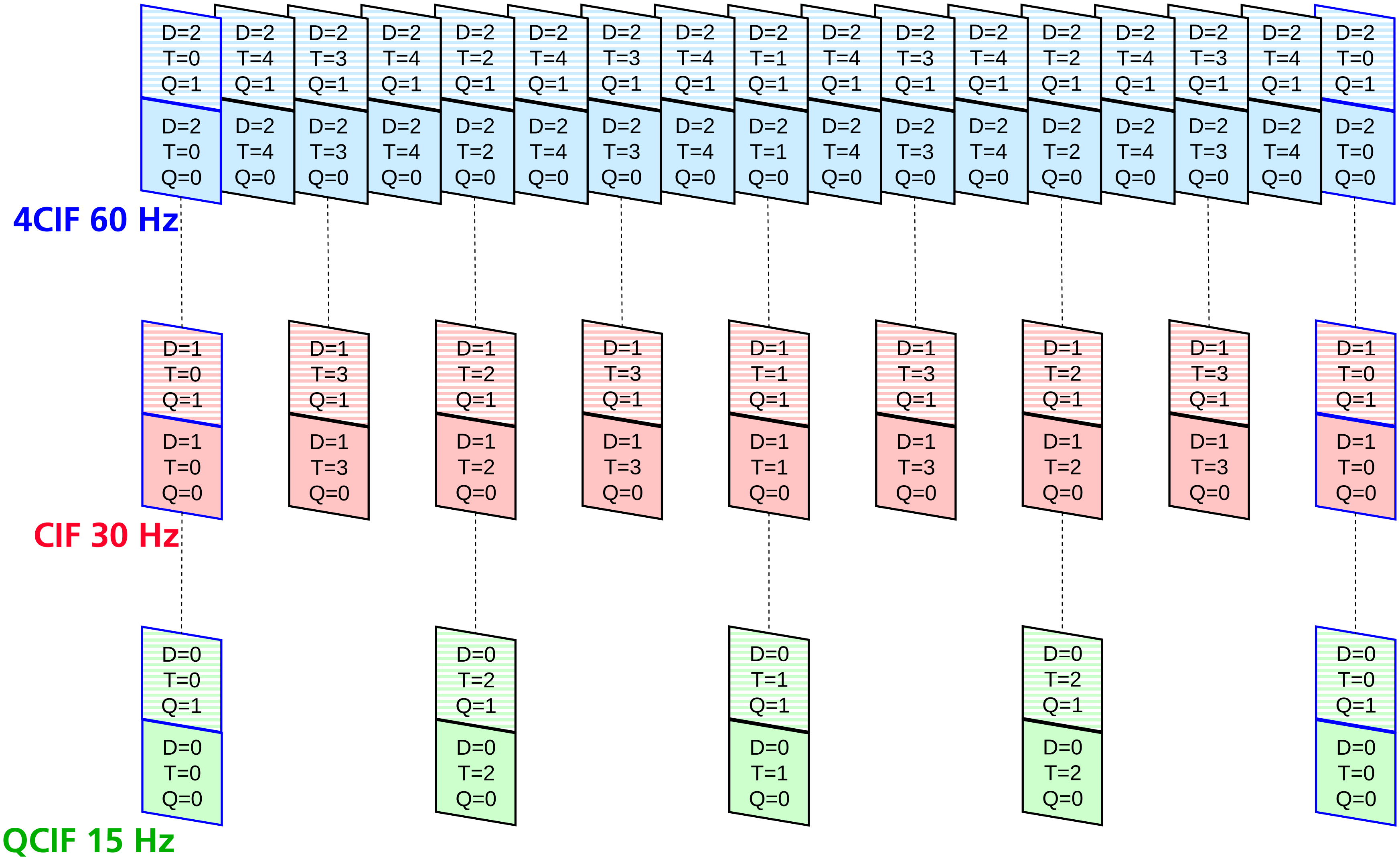
SVC NAL Unit Structure



- `r` **(reserved_one_bit)**
- `i` **(idr_flag)**
  - Specifies that current AU is IDR AU
- `p` **(priority_id)**
  - Priority information for easy stream manipulation
- `n` **(no_inter_layer_pred_flag)**
  - Indicates whether inter-layer prediction is used
- `d` **(dependency_id)**
  - Indicates a layer with separate SPS
  - Indicates spatial or CGS layer
- `q` **(quality_id)**
  - Indicates the quality refinement layer
- `t` **(temporal_id)**
  - Indicates temporal resolution
- `u` **(use_base_prediction_flag)**
  - Use base representation of reference pictures for motion-compensated prediction
- `x` **(discardable_flag)**
  - NAL unit is not required for decoding higher layers (with larger dependency_id)
- `o` **(output_flag)**
  - Specifies whether the decoded picture is output
- `r2` **(reserved_three_2bits)**
  - Two reserved bits having fixed value of 11

- **NAL unit header + NAL unit header extension = 32 bit (1 word)**
- **NAL unit header extension cannot accumulate start code**

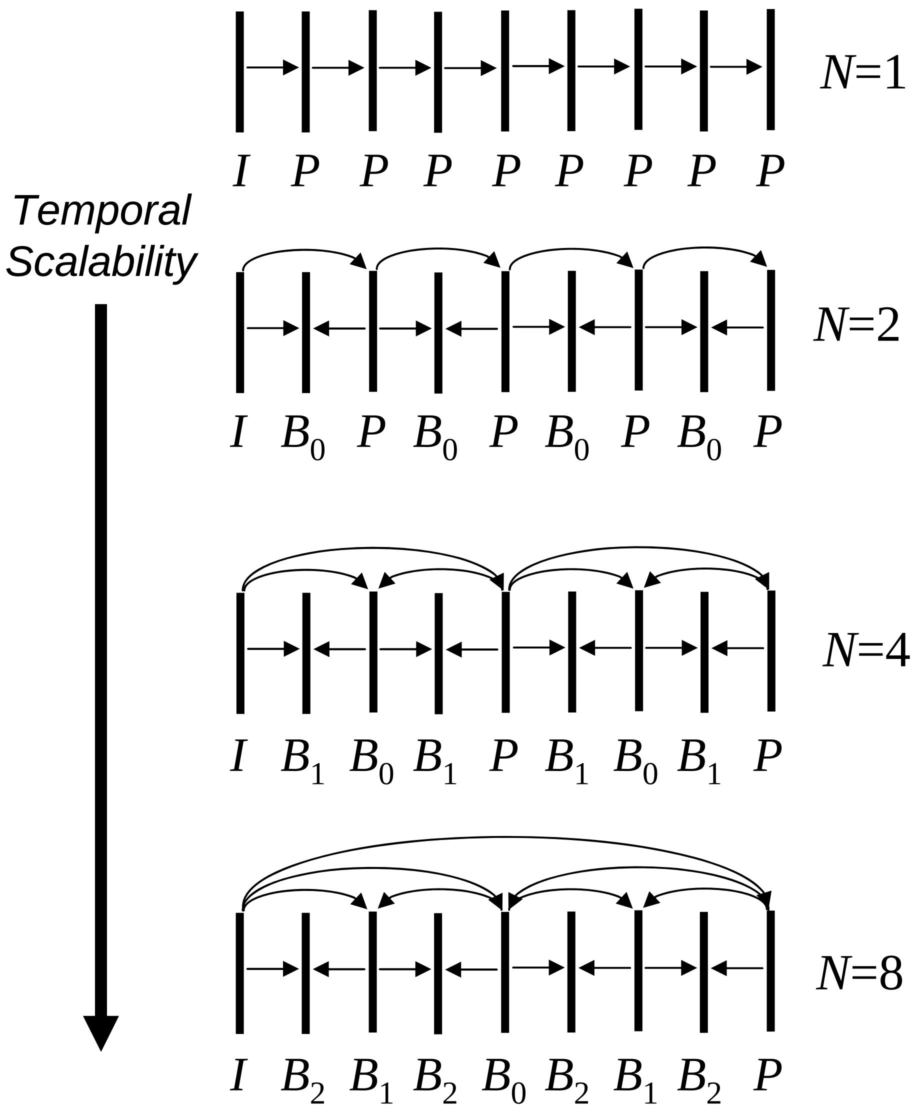
Assignment of (d,q,t) information



2.2 Temporal Scalability

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

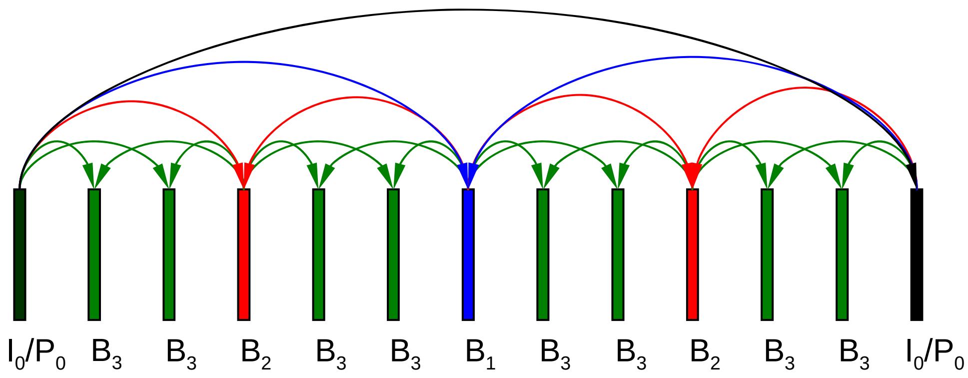
Temporal Decomposition of Video



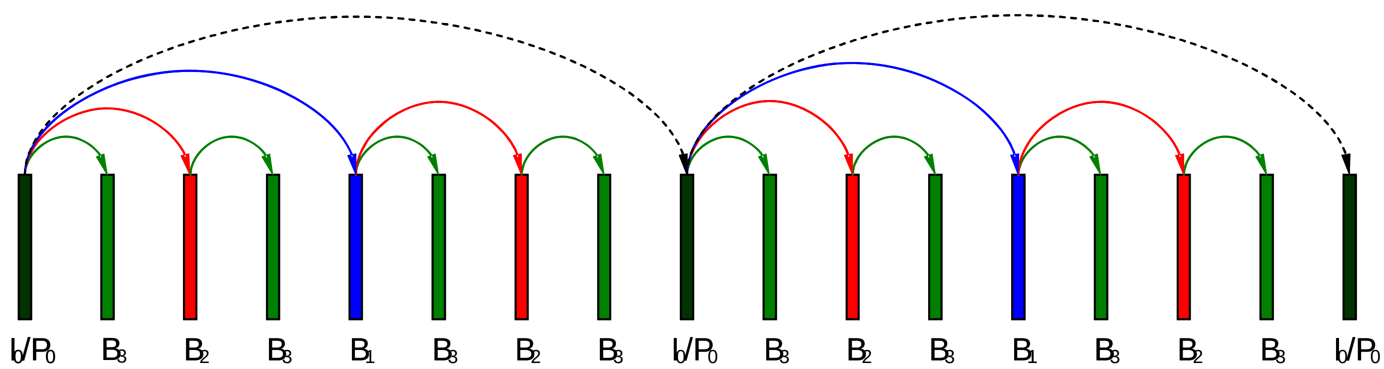
- **General rule:**
  Layers with higher temporal resolution shall not be used for prediction of layers with lower temporal resolution
- **Hierarchical prediction structures**
  - efficient method for enabling temporal scalability
  - not restricted to dyadic temporal scalability
  - can be combined with multiple reference pictures
  - delay can be controlled by restricting the use of future pictures for MCP

Hierarchical prediction structures

- **Non-dyadic temporal scalability**



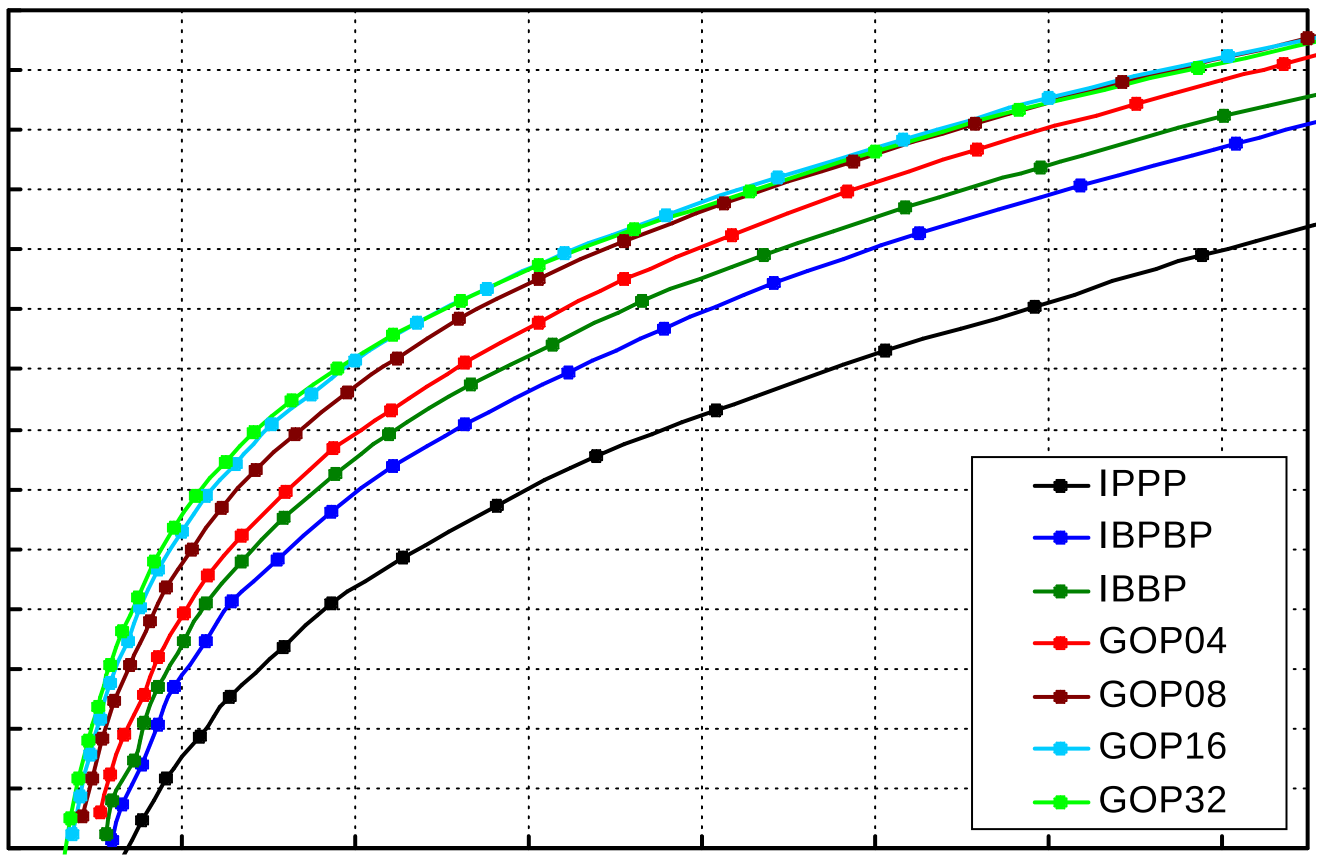
- **Low-delay prediction structure (structural delay is 0)**



SVC Extensions for Temporal Scalability

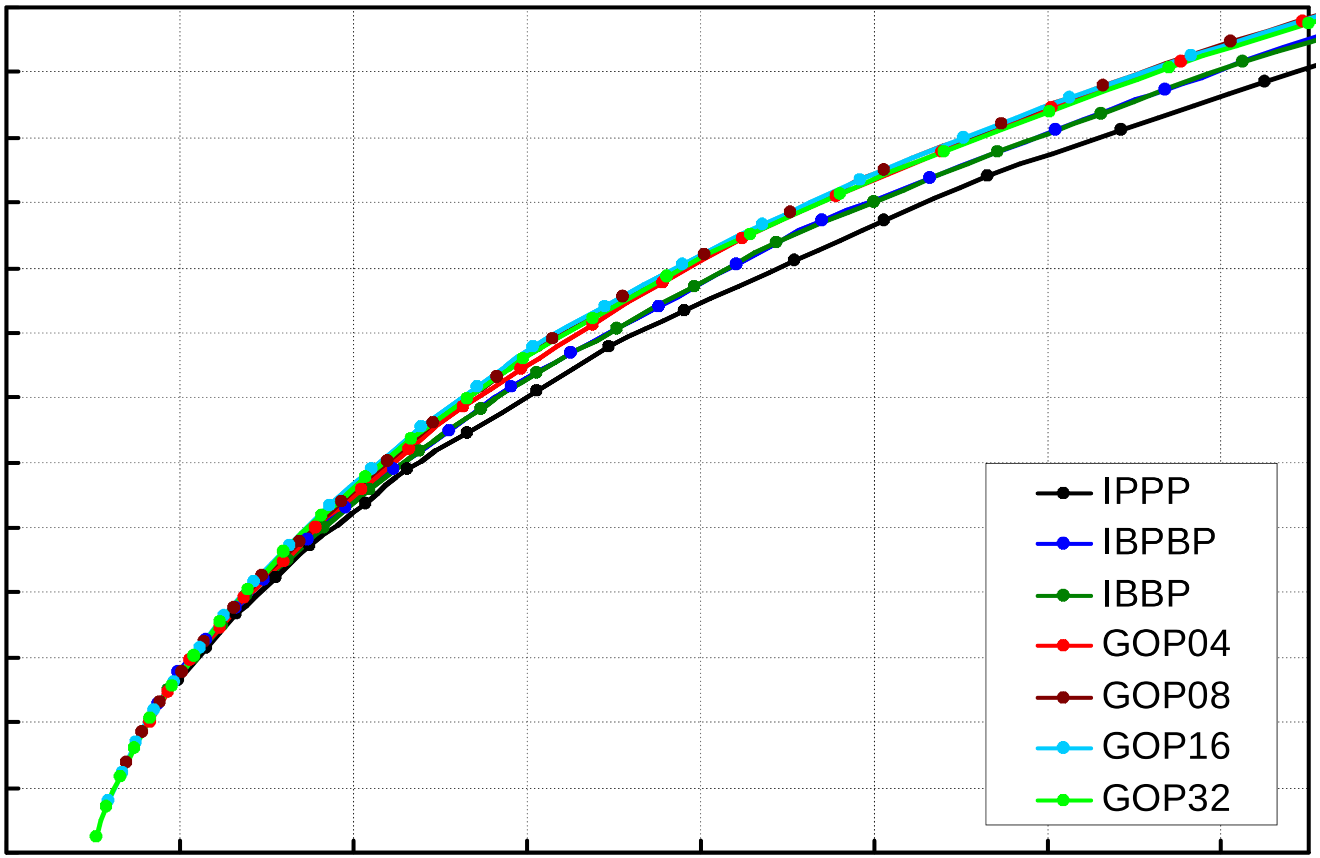
- Temporal scalability already supported in H.264/AVC through multiple reference picture syntax:
  - Decoupling of coding and display order
  - Reference picture list reordering (RPLR)
  - Memory management control operation (MMCO)
- RPLR and MMCO must be set by encoder such that dropping of a lower layer does not cause non-conforming bit-streams
  - Low memory requirements: Usage of long-term pictures
- Signaling of scalability structure in H.264/AVC through sub-sequence SEI messages
- SVC provides enhanced scalability signaling support by a NAL unit header extension or separate SVC NAL units succeeding H.264/AVC NAL units

Impact on Coding Efficiency



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Impact on Coding Efficiency



Impact on Coding Efficiency

Sequence Rugby, CIF 25Hz, 500 kbit/s, H.264/AVC

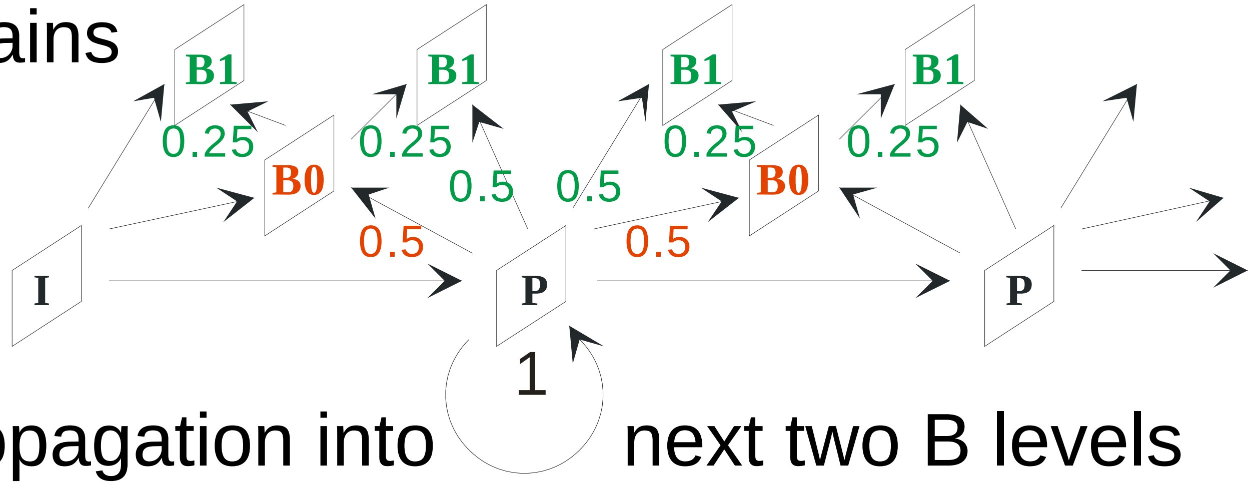
IBBPBBP... Coding

Hierarchical B Pictures

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Encoding of Hierarchical B Pictures

- Quantization error propagation through the pyramid explains necessity to decrease the quality

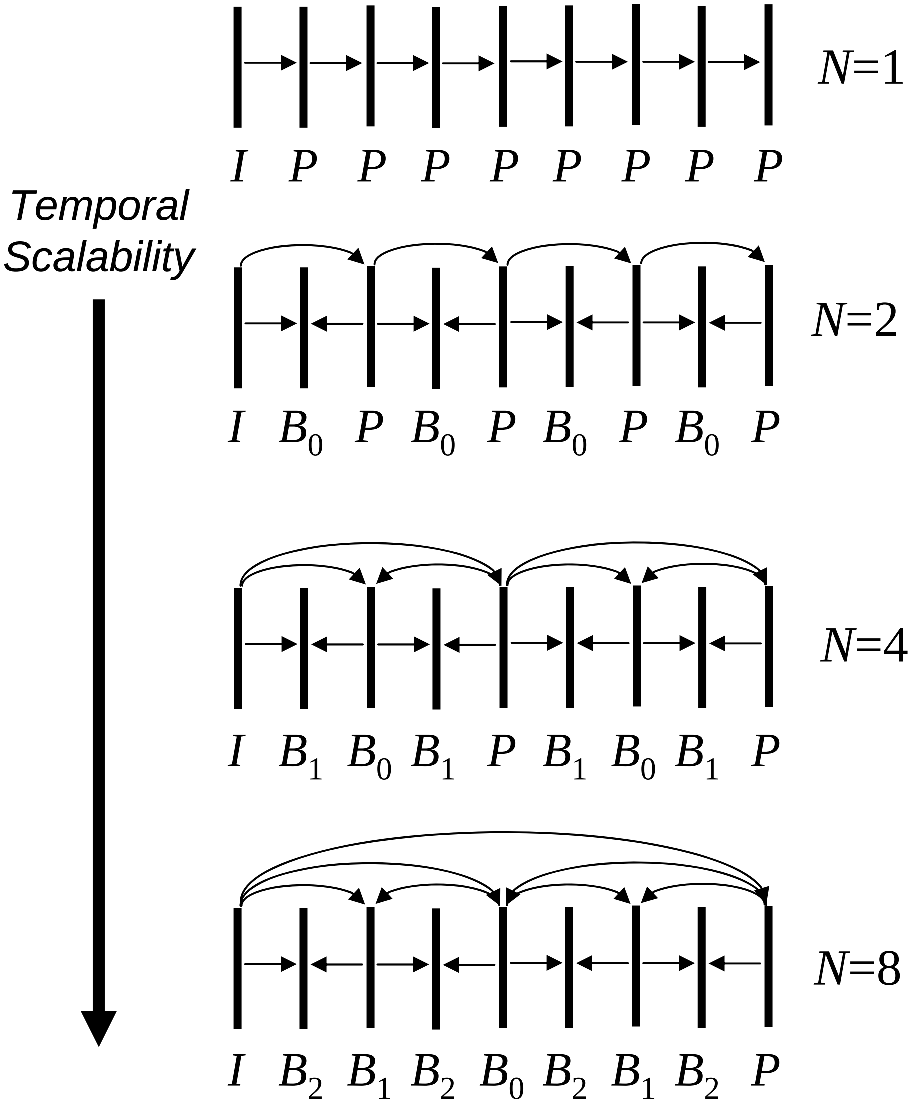


- Example: Propagation into next two B levels

$$\underbrace{\underbrace{\underbrace{1}_{1=1.5^0} + 2\cdot(0.5)^2}_{1.5=1.5^1} + 2\cdot(0.5)^2 + 4\cdot(0.25)^2}_{2.25=1.5^2}$$

- Quantization step size should be changed in next level by $\sqrt{1.5} \approx 1.225$

Encoding of Hierarchical B Pictures

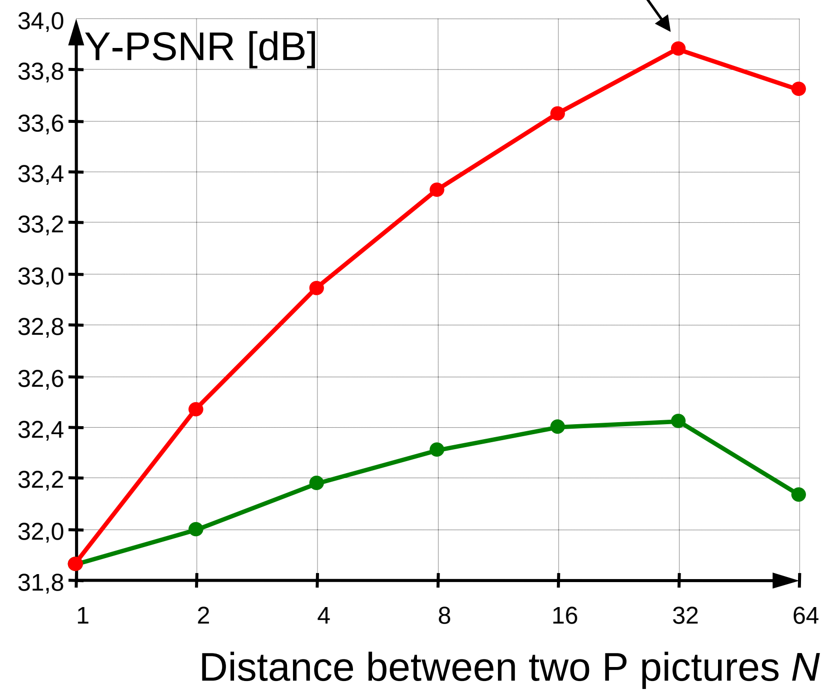


Video Coding Experiment with H.264/MPEG4-AVC

- Foreman, CIF 30Hz @ 150 kbit/s
- Performance as a function of N

Cascaded QP assignment

$\mathrm{QP}(P) \approx \mathrm{QP}(B_0) - 3 \approx \mathrm{QP}(B_1) - 4 \approx \mathrm{QP}(B_2) - 5$



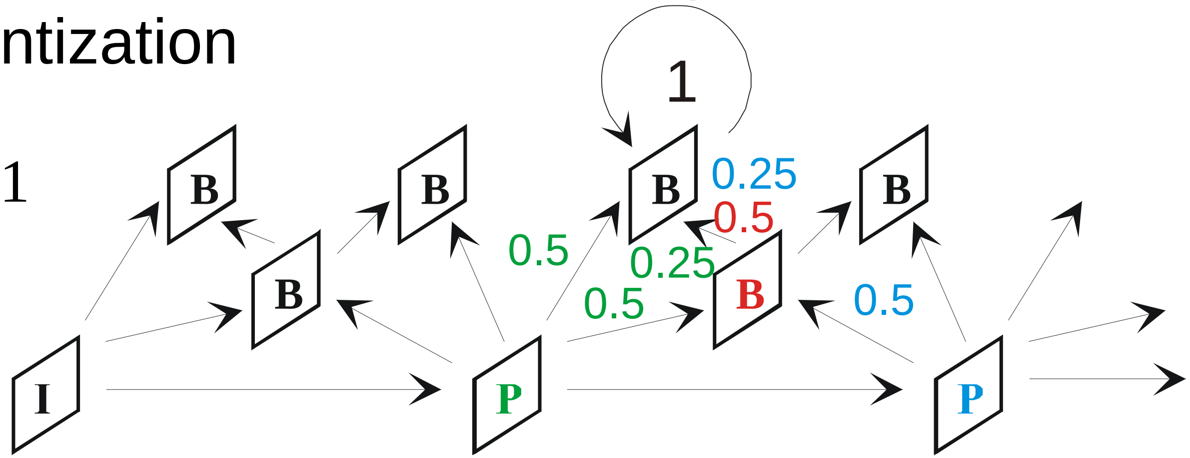
Influence of Prediction References

- When different prediction references are used at encoder and decoder (data loss or open-loop coding), the drift penalty stays small for hierarchical B pictures with optimum quantization

$$\Delta = 1$$

$$\Delta = 1/\sqrt{1.5}$$

$$\Delta = 1/\sqrt{2.25}$$



$$1^2 + \frac{0.5^2}{1.5} + \frac{0.5^2 + 0.25^2}{2.25} + \frac{0.25^2}{2.25} = 1 + 0.167 + \underbrace{0.1388 + 0.0277}_{=0.167}$$

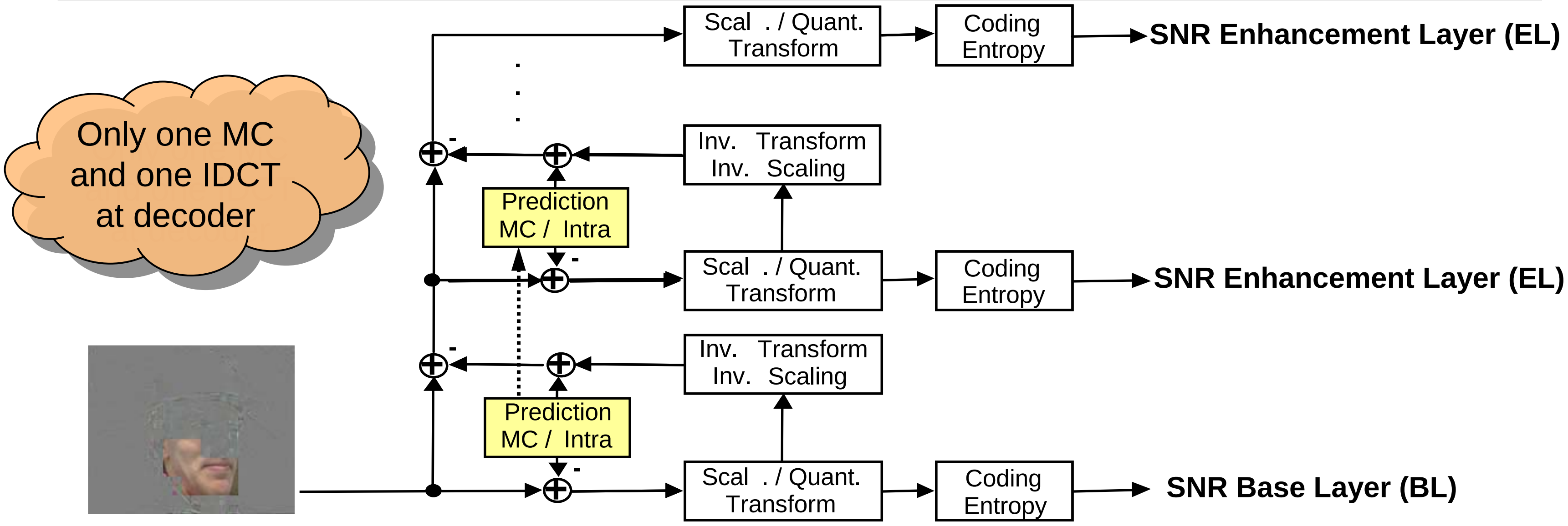
- Even though difference between open and closed loop is small, the loop should be closed at different layers

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

2.3 SNR/Fidelity/Quality Scalability

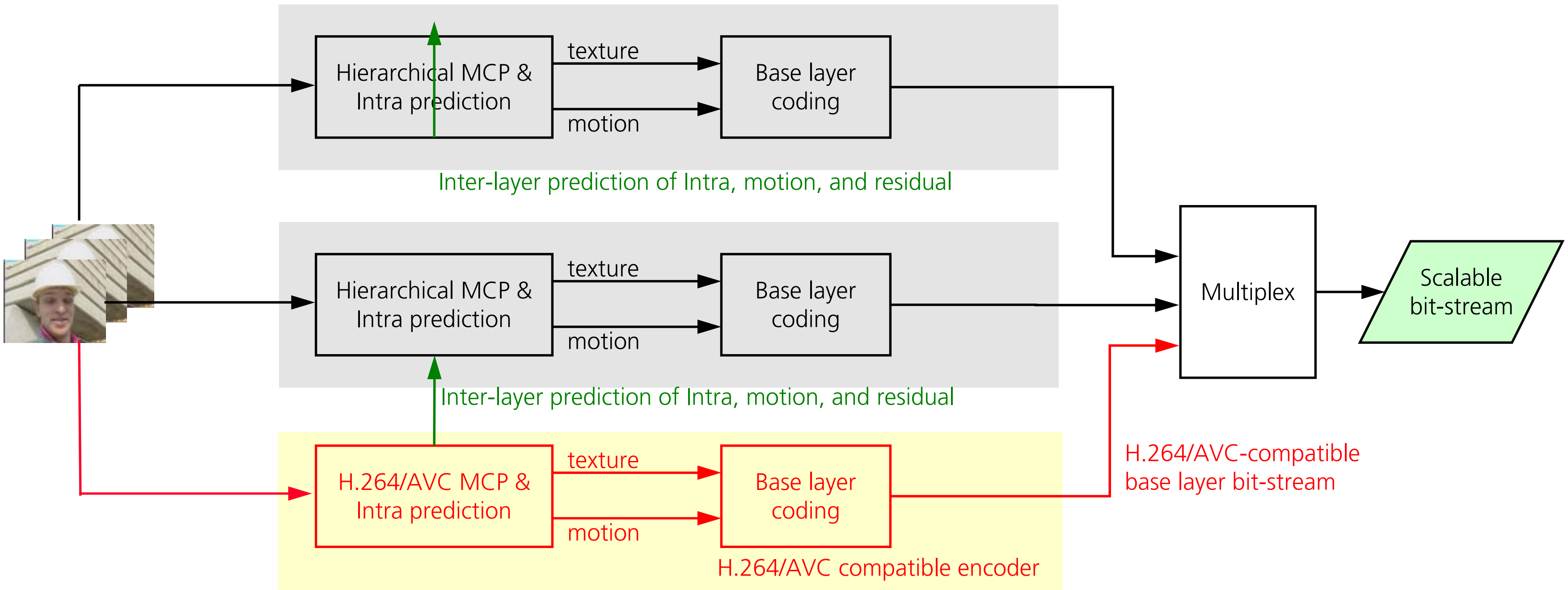
Fraunhofer Institut
Nachrichtentechnik
Heinrich-Hertz-Institut

Layered Coding for SNR Scalability



- Same resolution of pictures over all layers
- Intra prediction only in base layer
- Residual coding of the quantization error between original pictures and their lower layer reconstruction
- Motion vector refinement possible
- Can be converted to H.264/AVC without transcoding at any bit rate

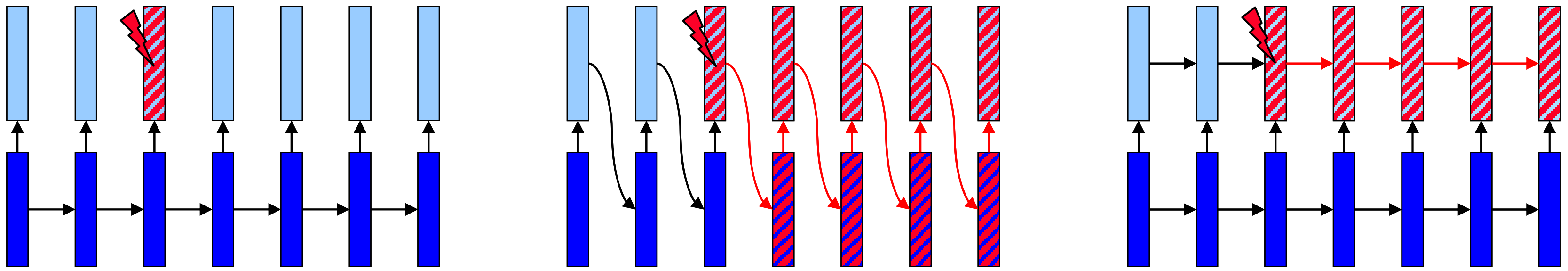
Coarse-grain SNR scalability (CGS)



- **Extreme case of (extended) spatial scalability**
  - resolution ratio is equal to 1, no cropping
  - no upsampling (motion, texture) required
  - singe-loop decoding !
- **Special case: Requantization**
  - identical motion data for all layers, always use residual prediction

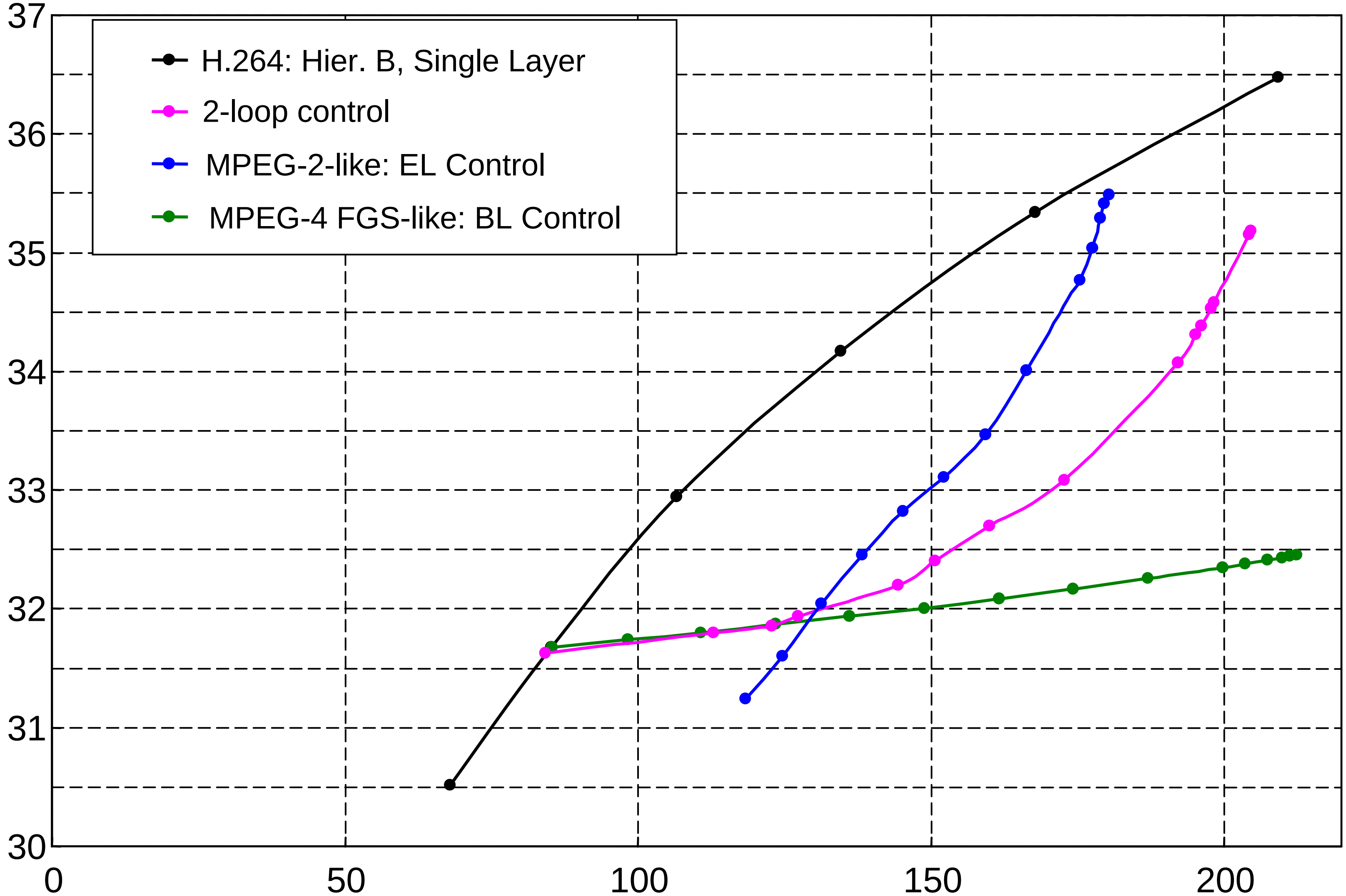
Drift in Past SNR Scalable Coding

- **Source of drift:**
  Non-synchronized motion compensation loops in encoder and decoder due to loss of enhancement layer information



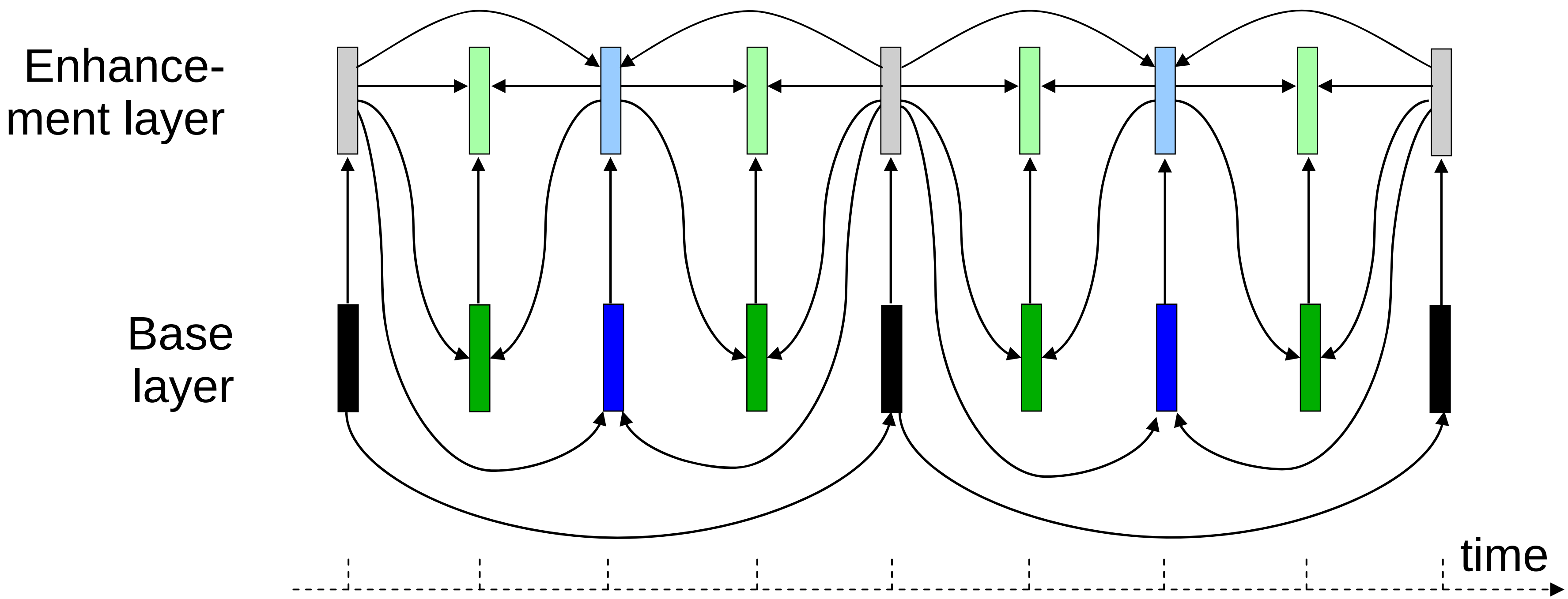
- **BL only control**
  - MPEG-4 FGS
  - No drift
  - Low complexity
  - Efficient BL and inefficient EL coding
- **EL only control**
  - MPEG-2 SNR
  - Drift in both BL and EL
  - Low complexity
  - Inefficient BL and efficient EL coding (when latter complete)
- **2-loop control**
  - MPEG-4 spatial used as SNR
  - No drift in BL
  - Drift in EL
  - High complexity
  - Efficient BL and medium efficient EL coding

SNR Scalability Results: Past Codecs

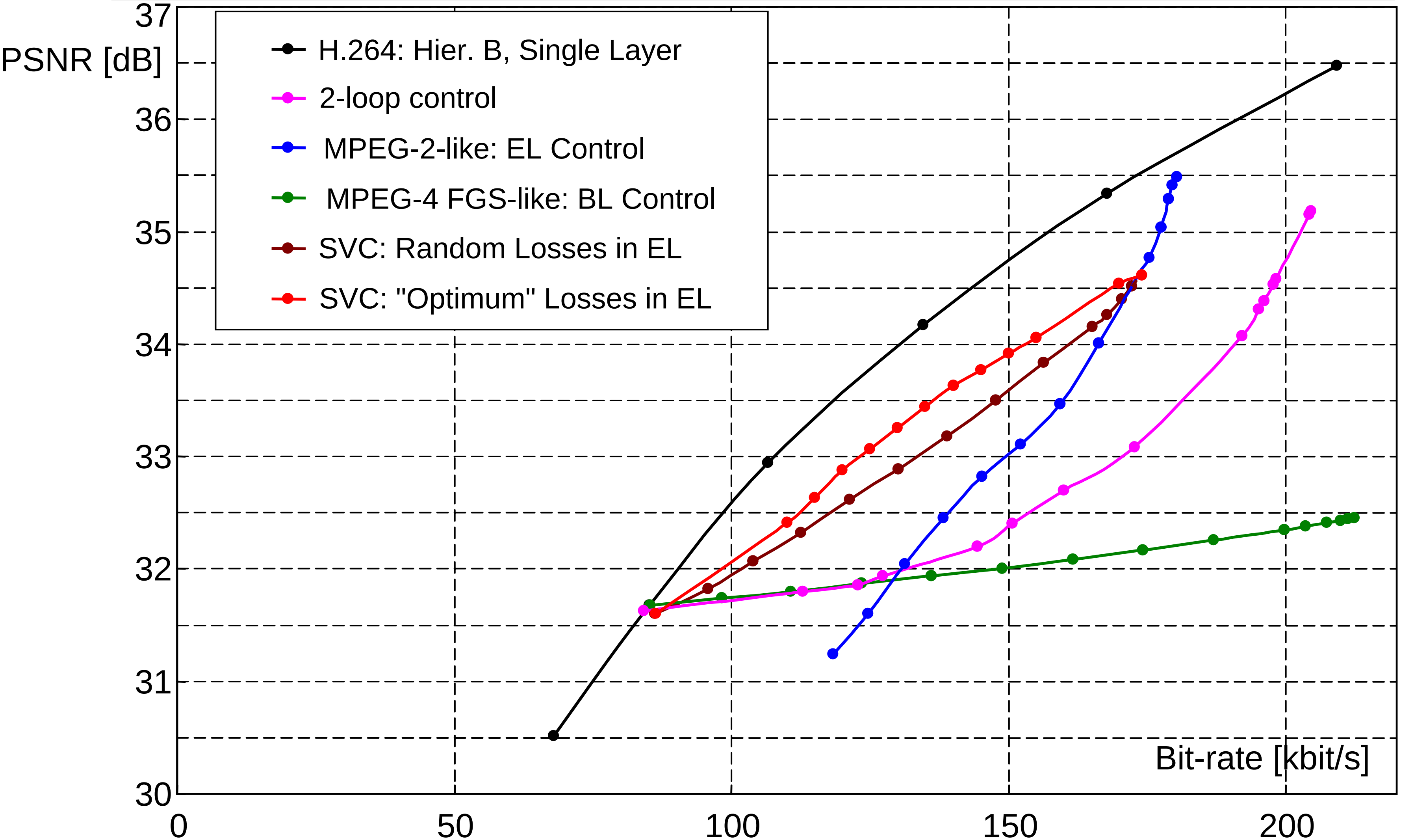


Drift Control in SVC

- Adaptive BL/EL only encoder control for hierarchical prediction structures
- Normal pictures: use EL reconstruction as reference for MCP
- Closed-loop pictures: use BL reconstruction as reference for MCP at encoder side
- Single motion compensation loop decoding

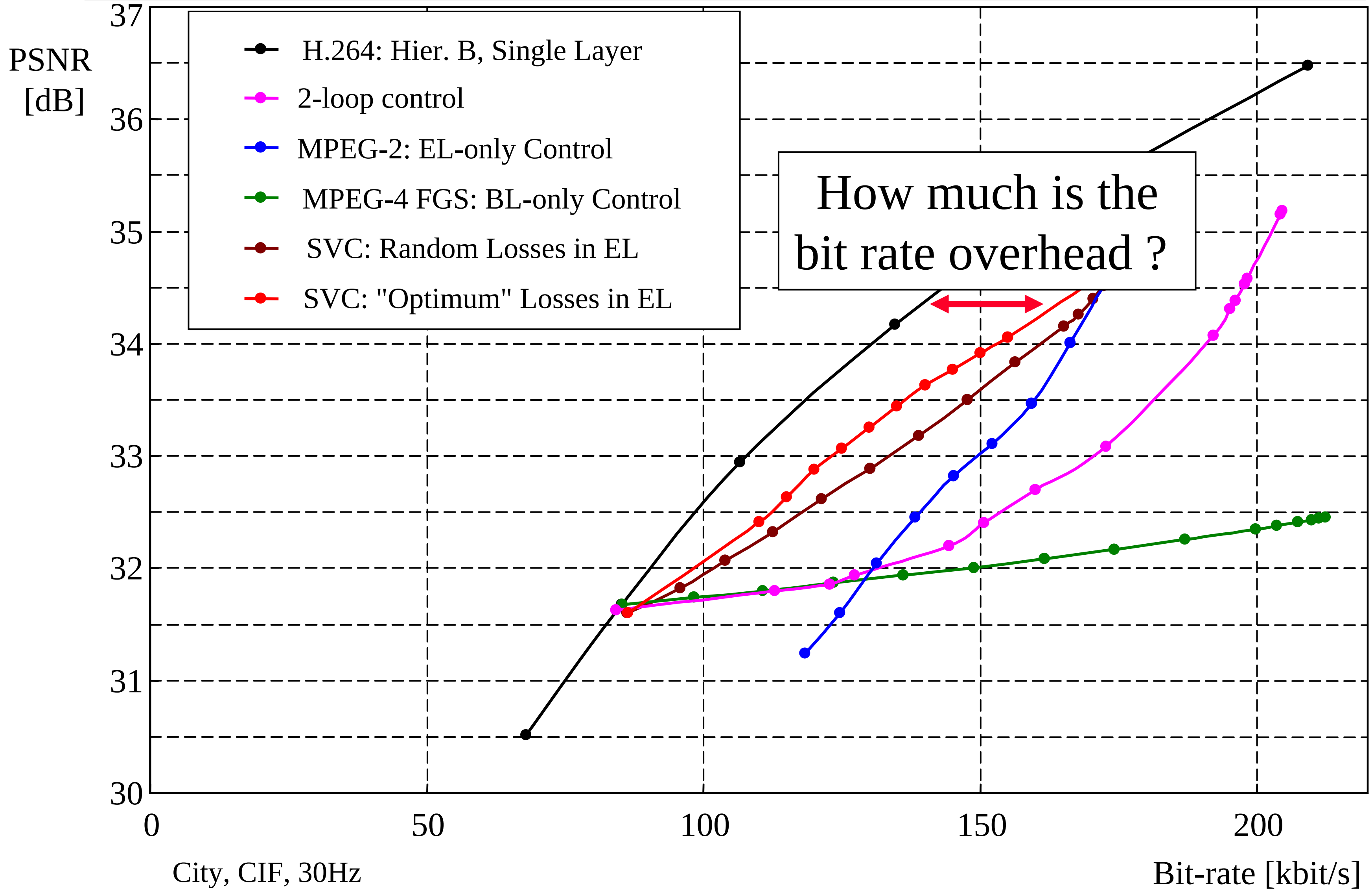


SNR Scalability Results: New SVC



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

SNR Scalability Results: New SVC



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Coder Control

- **Constrained problem**:

$$\min_{\boldsymbol{p}} D(\boldsymbol{p}) \quad \text{s.t.} \quad R(\boldsymbol{p}) \le R_T$$

D Distortion
R Rate
R_T Target rate
$\boldsymbol{p}$ Parameter Vector

- **Unconstrained Lagrangian formulation**:

$$\boldsymbol{p}' = \arg\min_{\boldsymbol{p}} \{D(\boldsymbol{p}) + \lambda \cdot R(\boldsymbol{p})\}$$

with λ controlling the rate-distortion trade-off

- **Rate-Constrained Mode Decision**

$$D_2(M|QP) + \lambda_M R(M|QP)$$

M Evaluated macroblock mode out of a set of possible modes
QP Value of quantizer for transform coefficients
λ_M Lagrange parameter for mode decision
D_2 Sum of squared differences (luminance & chrominance)
R Number of bits associated with header, motion, transform coefficients

- **Rate-Constrained Motion Estimation**

$$D_1(\boldsymbol{m}, \mathrm{D}) + \lambda_D R(\boldsymbol{m}, D)$$

$\boldsymbol{m}$ Motion vector containing spatial displacement
Δ Picture reference parameter
λ_D Lagrange parameter for motion estimation
D_1 Sum of absolute differences (luminance)
R Number of bits associated with motion information

Encoder Optimization: JSVM encoder control

- **Lagrangian bit-allocation techniques**: bottom-up encoding process
- **Encode base layer:**
  - Coding parameters are optimized for the base layer

$$\boldsymbol{p}_0' = \arg\min_{\{\boldsymbol{p}_0\}} \left[D_0(\boldsymbol{p}_0) + \lambda_0 \cdot R_0(\boldsymbol{p}_0) \right]$$

- **Encode enhancement layer**
  - Coding parameters are optimized for the enhancement layer

$$\boldsymbol{p}_1' = \arg\min_{\{\boldsymbol{p}_1 | \boldsymbol{p}_0\}} \left[D_1(\boldsymbol{p}_1 \mid \boldsymbol{p}_0) + \lambda_1 \cdot R_1(\boldsymbol{p}_1 \mid \boldsymbol{p}_0) \right]$$

- All enhancement layer decisions are conditioned on already determined base layer coding parameters

Encoder Optimization: Joint Control

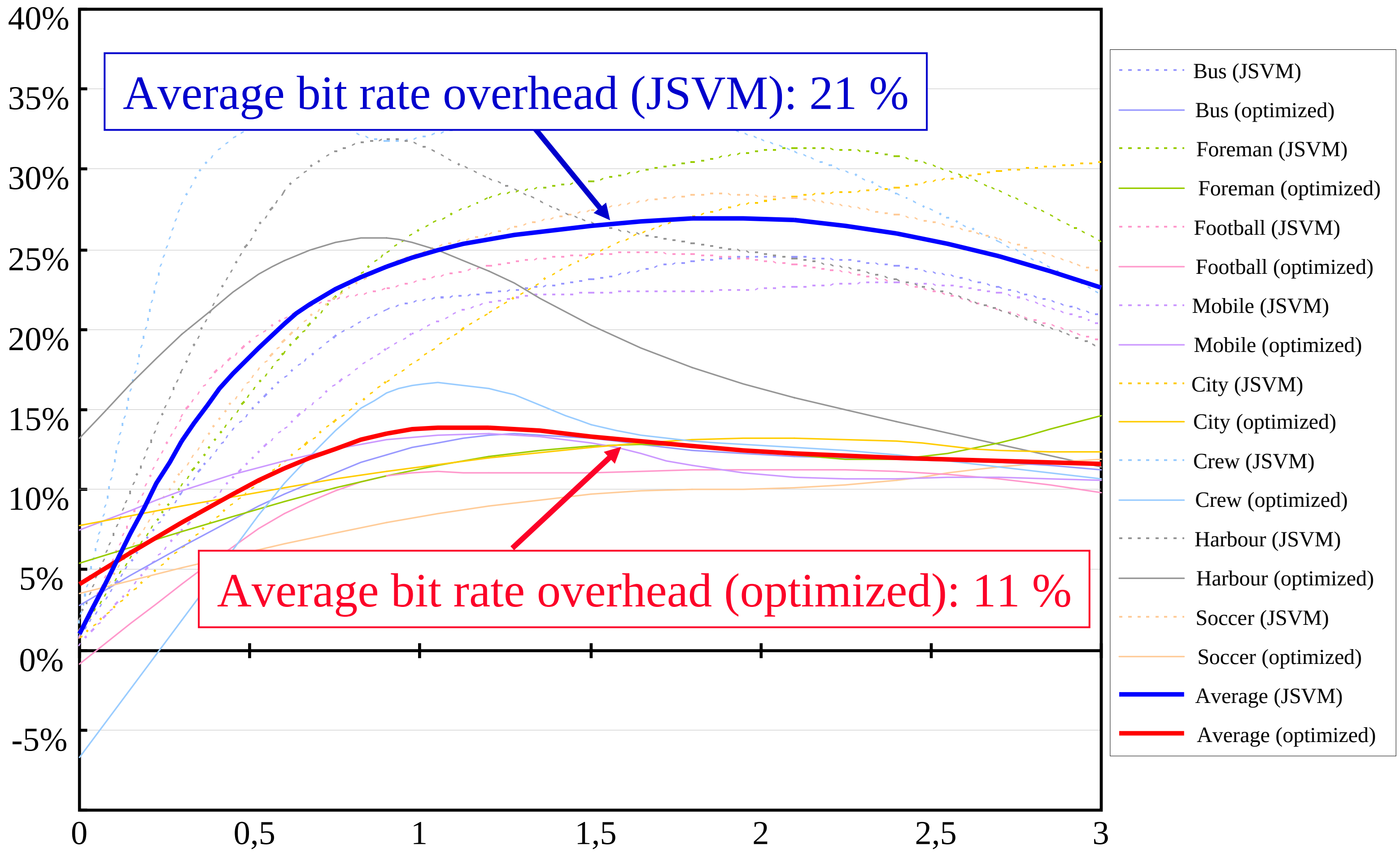
- **Jointly optimize base and enhancement layer coding parameters**
- **Consider enhancement layer during base layer encoding**
- **Modified Lagrangian approach for base layer encoding**

$$\boldsymbol{p}_0 = \arg\min_{\{\boldsymbol{p}_0, \boldsymbol{p}_1 | \boldsymbol{p}_0\}} \left[(1-w) \cdot \left(D_0(\boldsymbol{p}_0) + \lambda_0 \cdot R_0(\boldsymbol{p}_0) \right) + w \cdot \left(D_1(\boldsymbol{p}_1 | \boldsymbol{p}_0) + \lambda_1 \cdot R_1(\boldsymbol{p}_1 | \boldsymbol{p}_0) \right) \right]$$

- **Weighting factor** w
  - $w = 0$: Current JSVM encoder control
  - $w = 1$: Single-loop encoder control (base layer is not controlled)
  - $0 < w < 1$: Partly consider enhancement layer during base layer encoding
- **Enhancement layer encoding is not modified**

Results for SNR Scalability

Average bit rate overhead relative to single layer coding

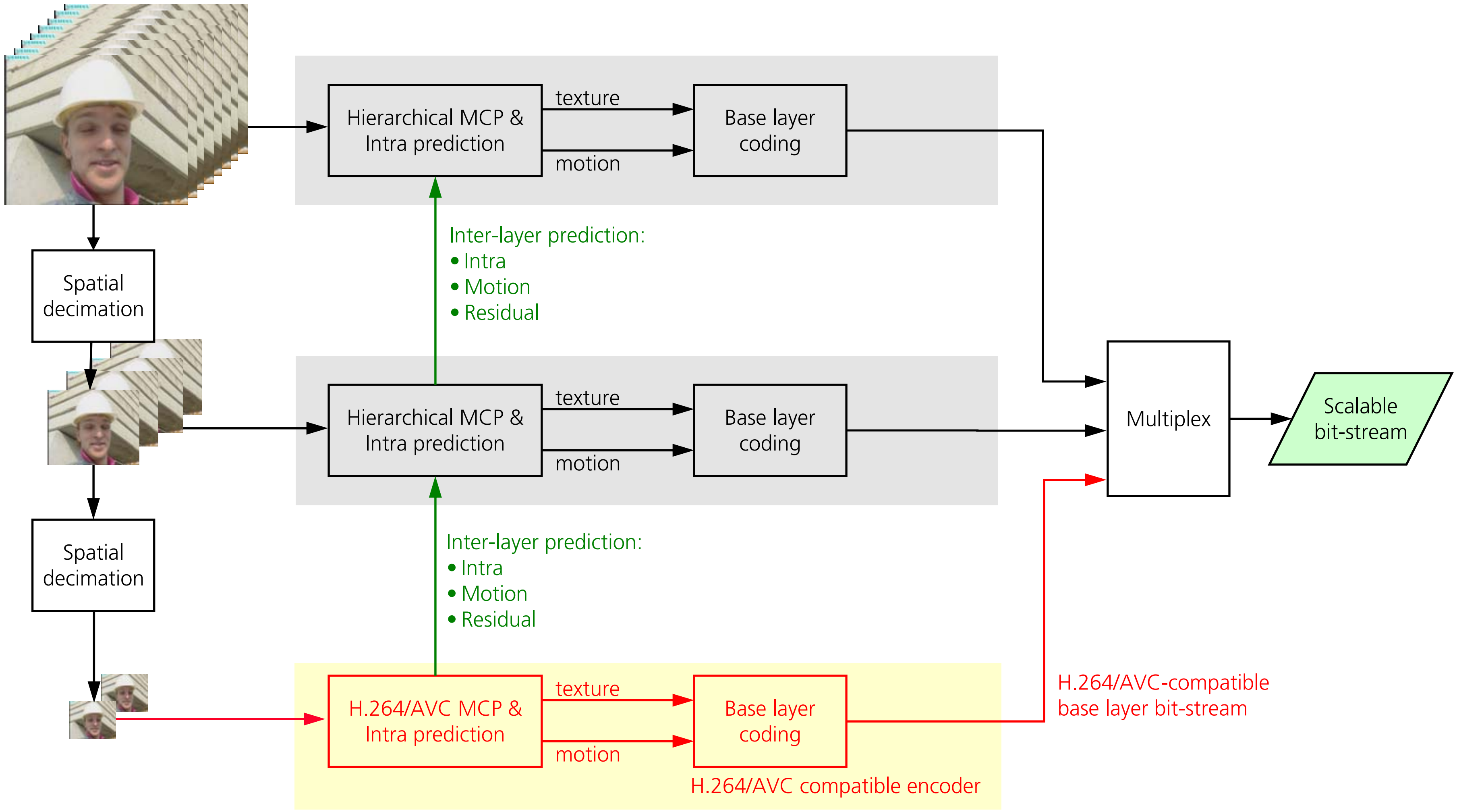


PSNR difference [dB] relative to JSVM base layer

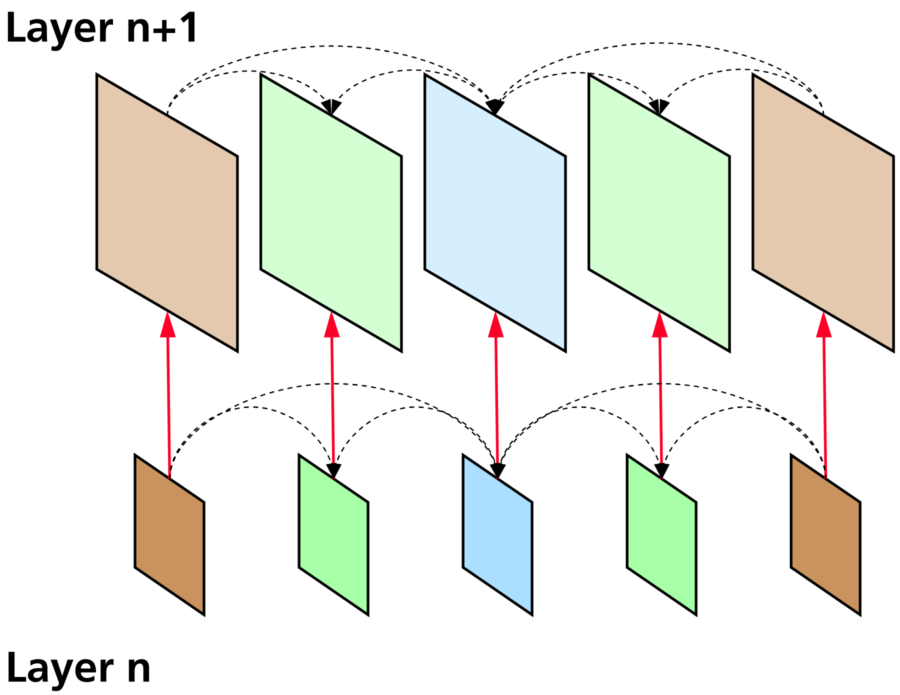
2.4 Spatial Scalability

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Typical Encoding for Spatial Scalability



Layered Coding for Spatial Scalability

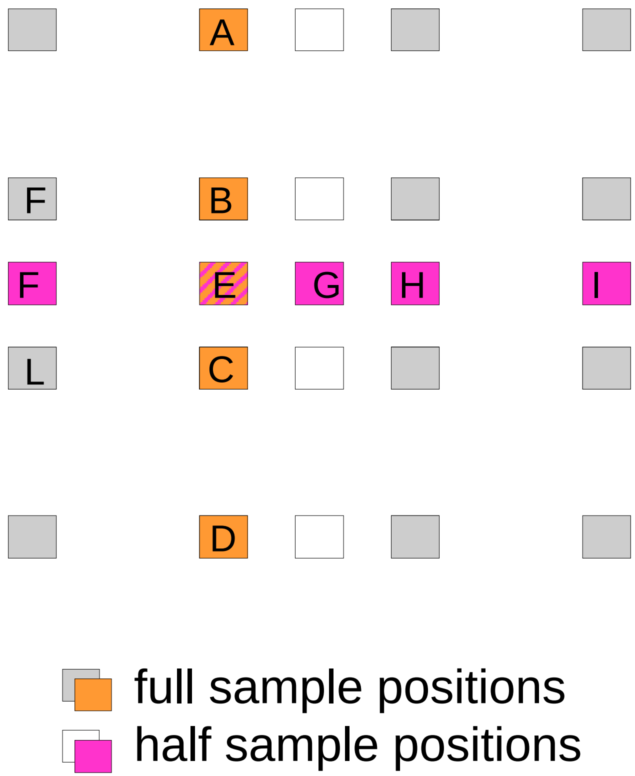


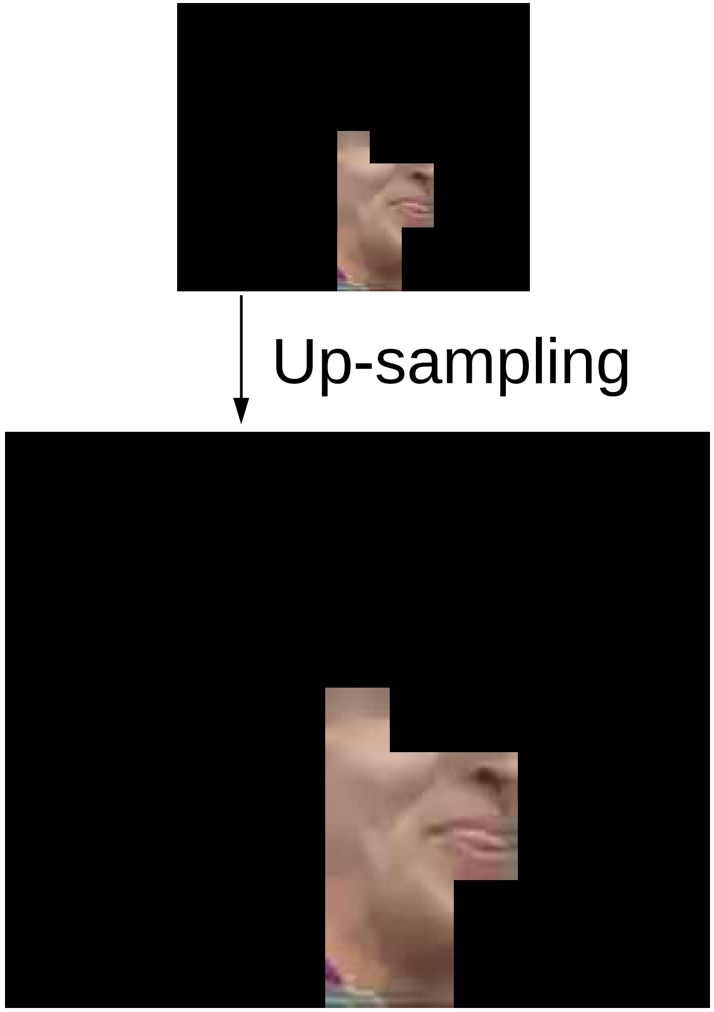
- **Layered coding**
  - Oversampled pyramid for each resolution:
    e.g. QCIF, CIF, 4CIF, 16CIF
  - MC prediction structures of all layers are aligned
- **Inter-layer prediction:**
  Switchable prediction (with upsampling)
  - Prediction of intra macroblocks (cp. MPEG-2, H.263, MPEG-4)
  - No prediction of inter MBs (unlike MPEG-2 etc.)
  - Prediction of partitioning and motion information (new in SVC)
  - Prediction of residual data (new in SVC)

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Intra Prediction

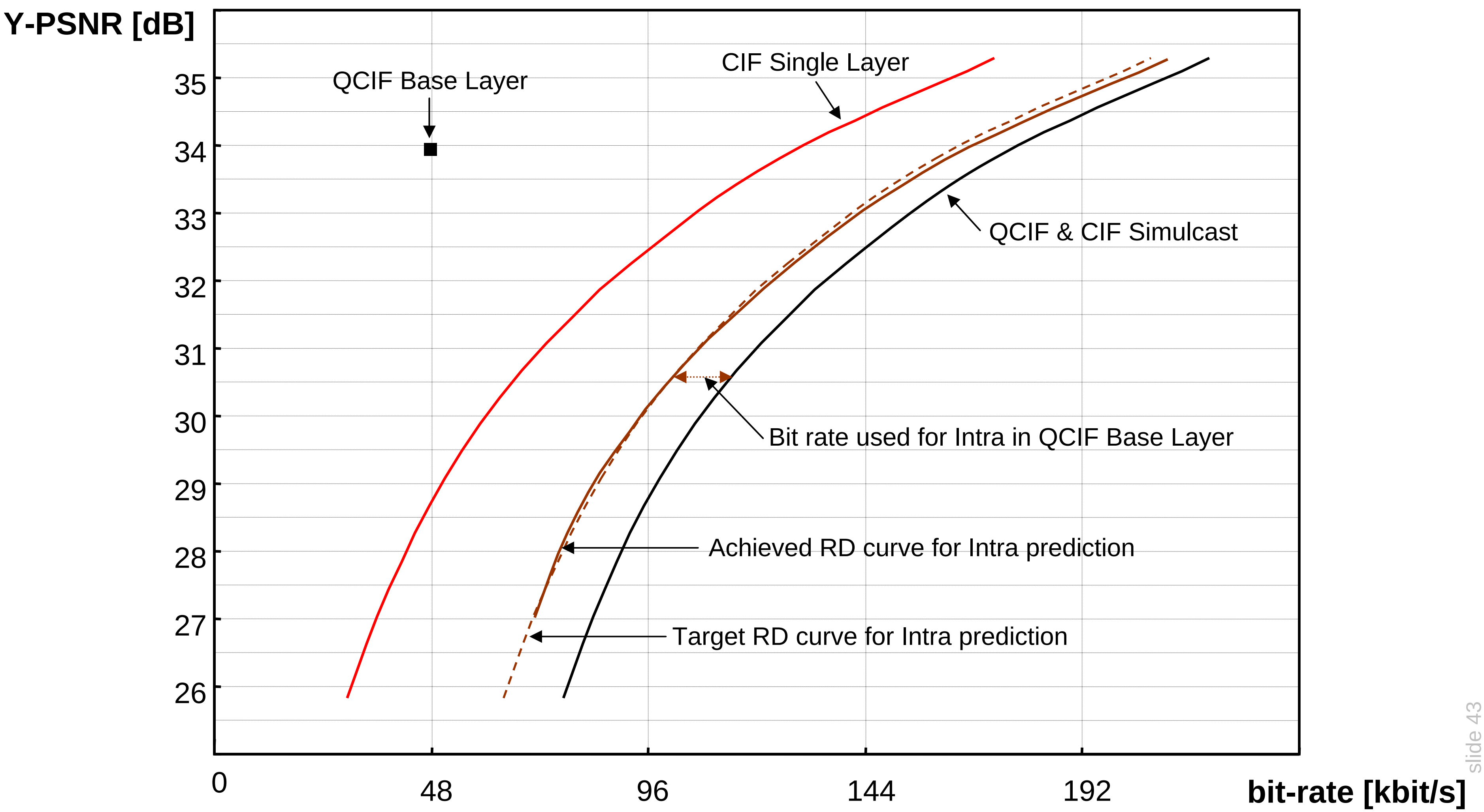
4-tap filter: [-3, 19, 19, -3]





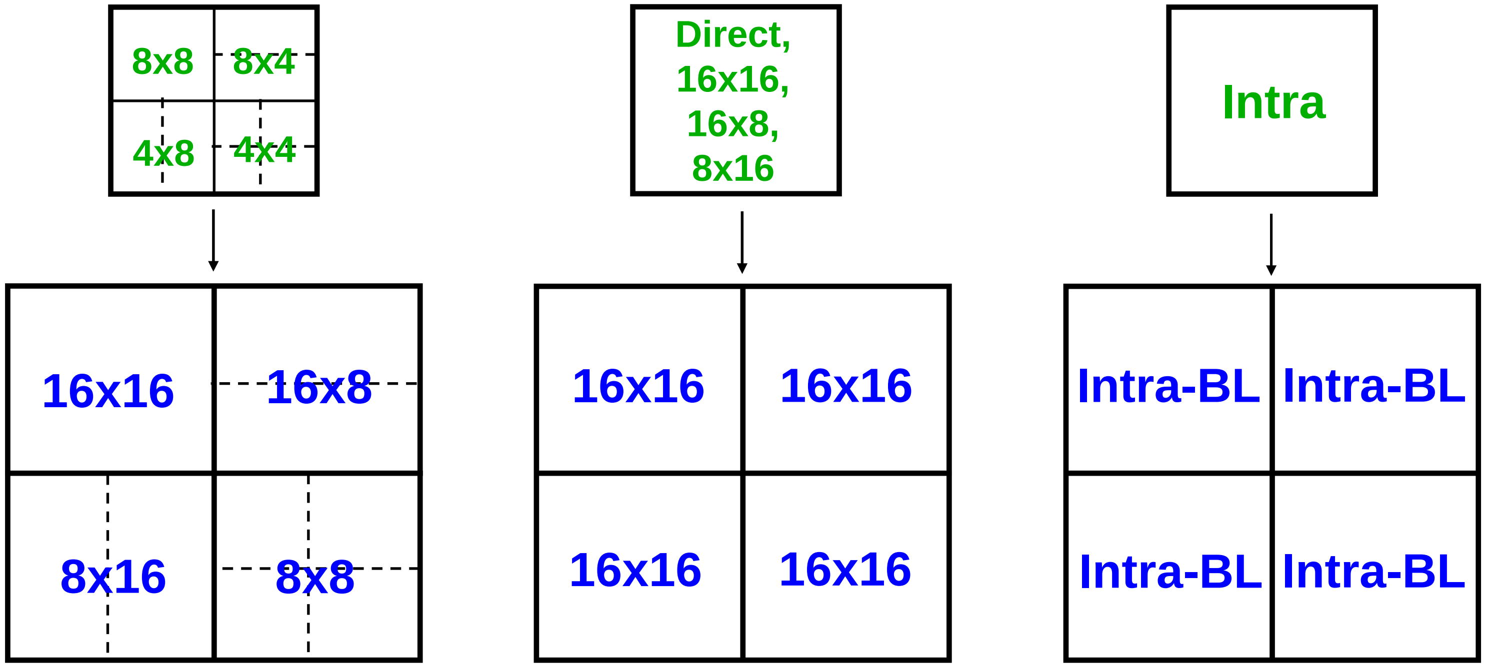
Spatial Prediction of Data

Spatial Scalability: Foreman, 150 pics, QCIF 15Hz @ 48 kbit/s -> CIF 15 Hz

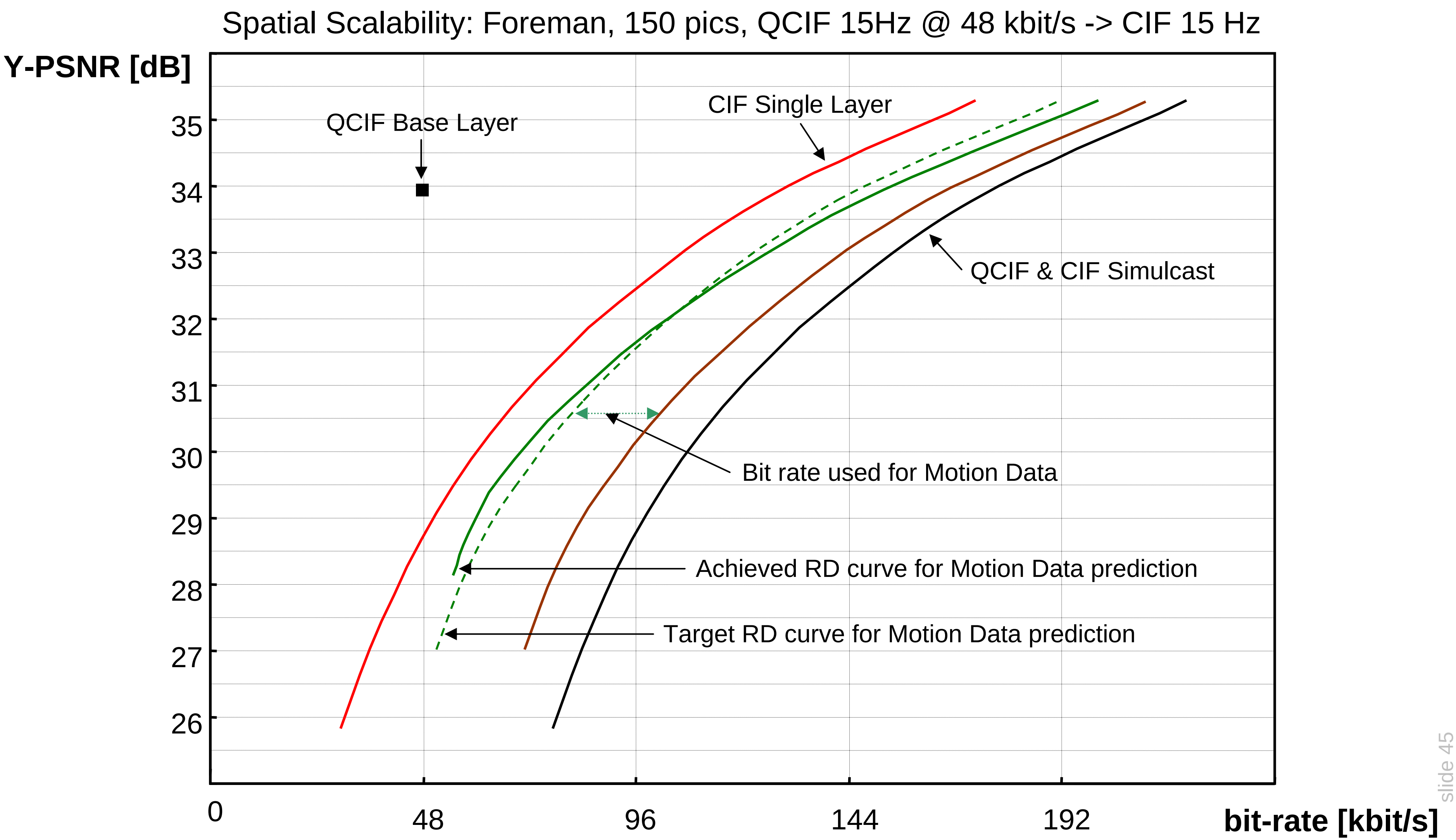


Spatial Prediction of Motion Data

- Upsample macroblock partitioning as switchable partitioning predictor
- Multiply motion vectors by 2 and use them as switchable predictors (keep list 0, list 1, bi-predictive and reference indices information)



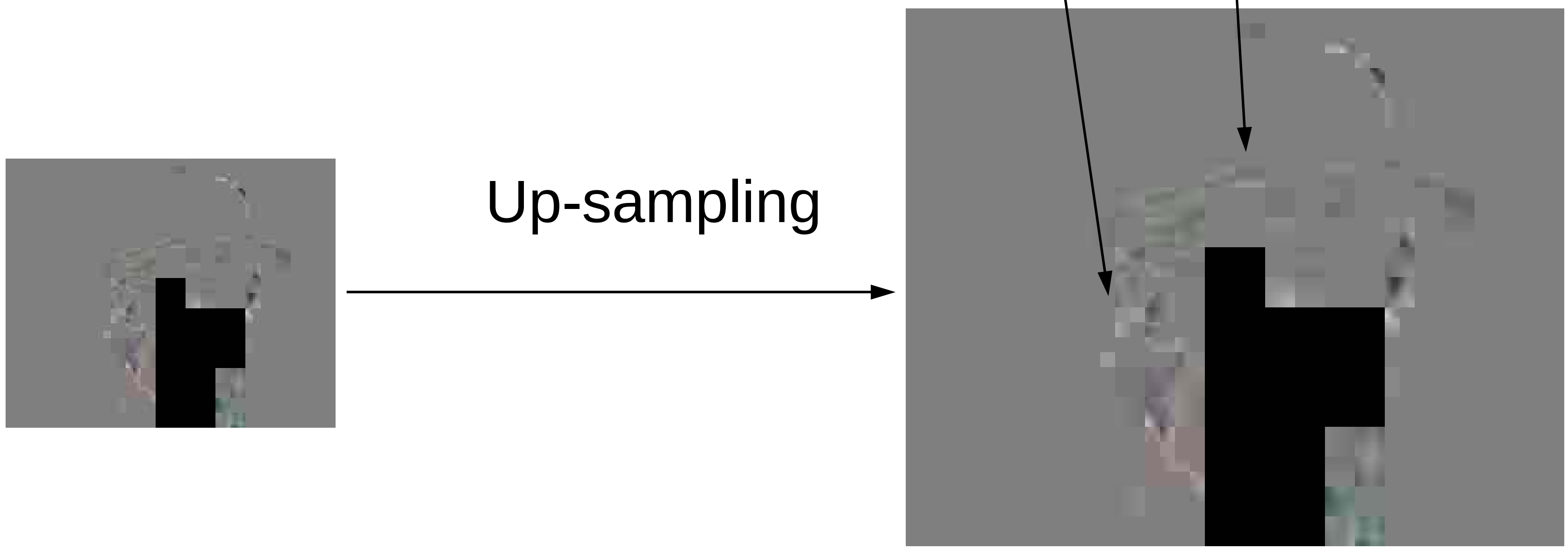
Spatial Prediction of Data



Residual Prediction

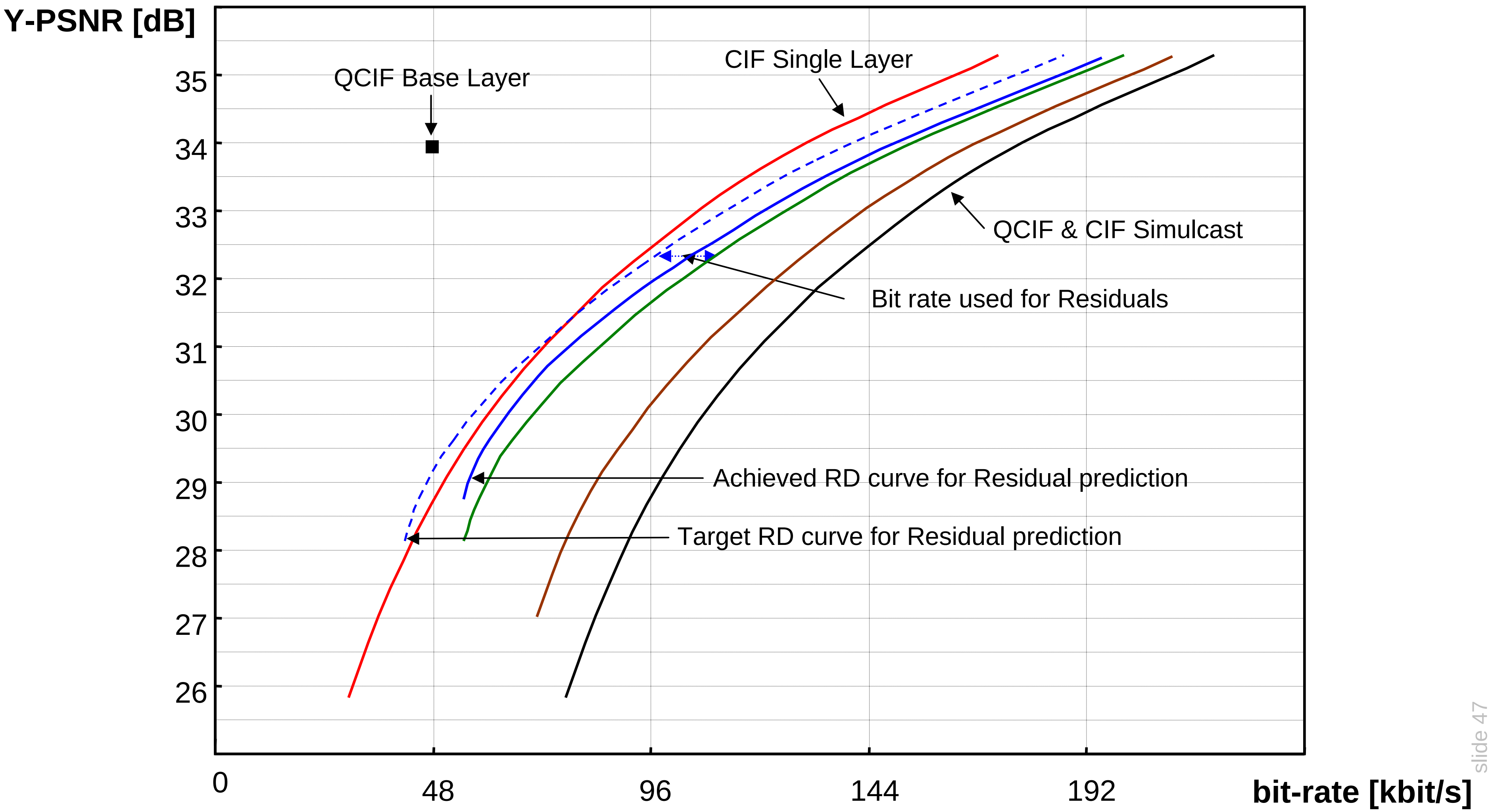
Block-wise bi-linear up-sampling filter

- Block-wise: block boundaries of 4x4 or 8x8 blocks
- bi-linear: small 4x4 block size



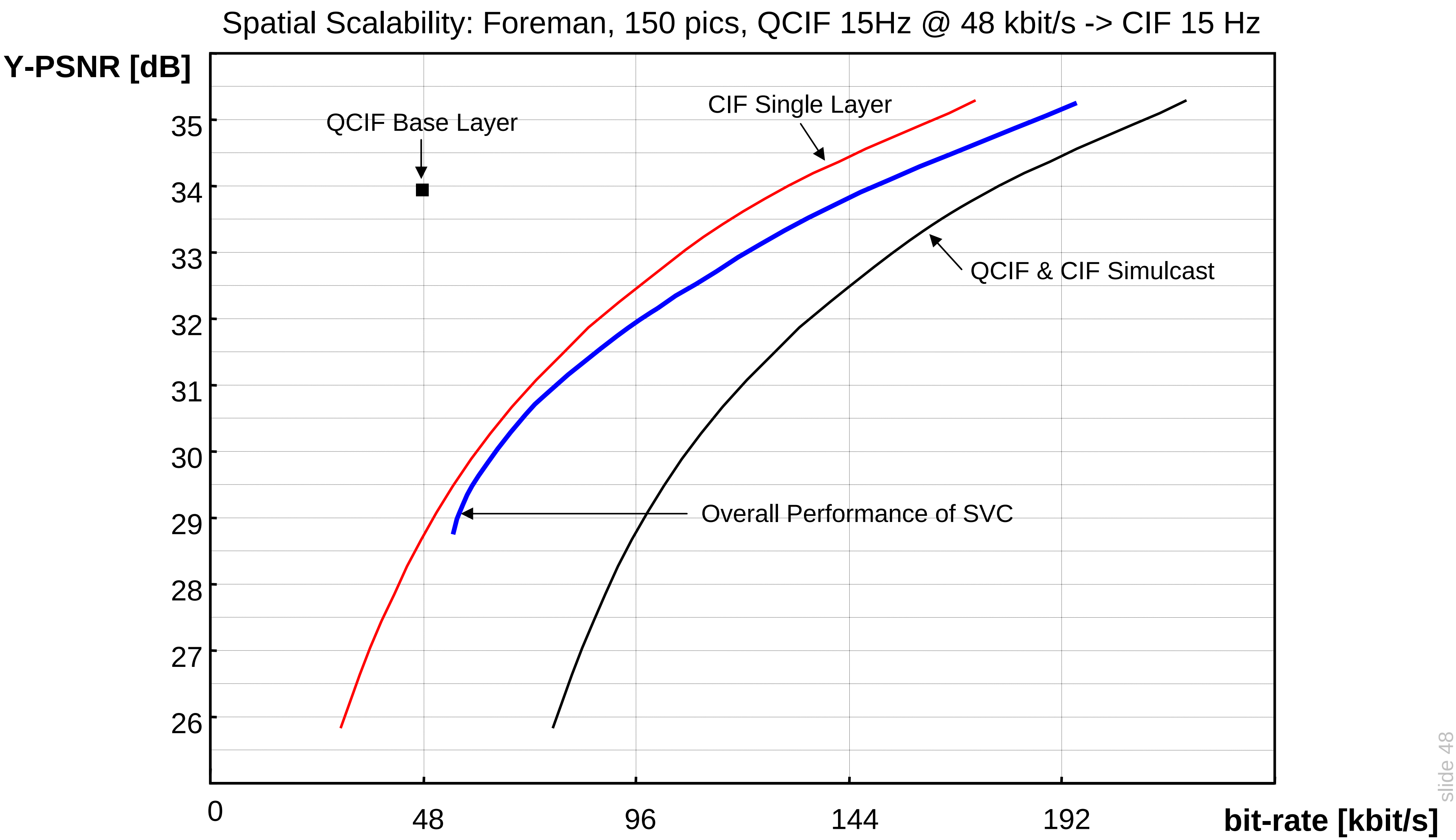
Spatial Prediction of Data

Spatial Scalability: Foreman, 150 pics, QCIF 15Hz @ 48 kbit/s -> CIF 15 Hz



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Spatial Prediction of Data

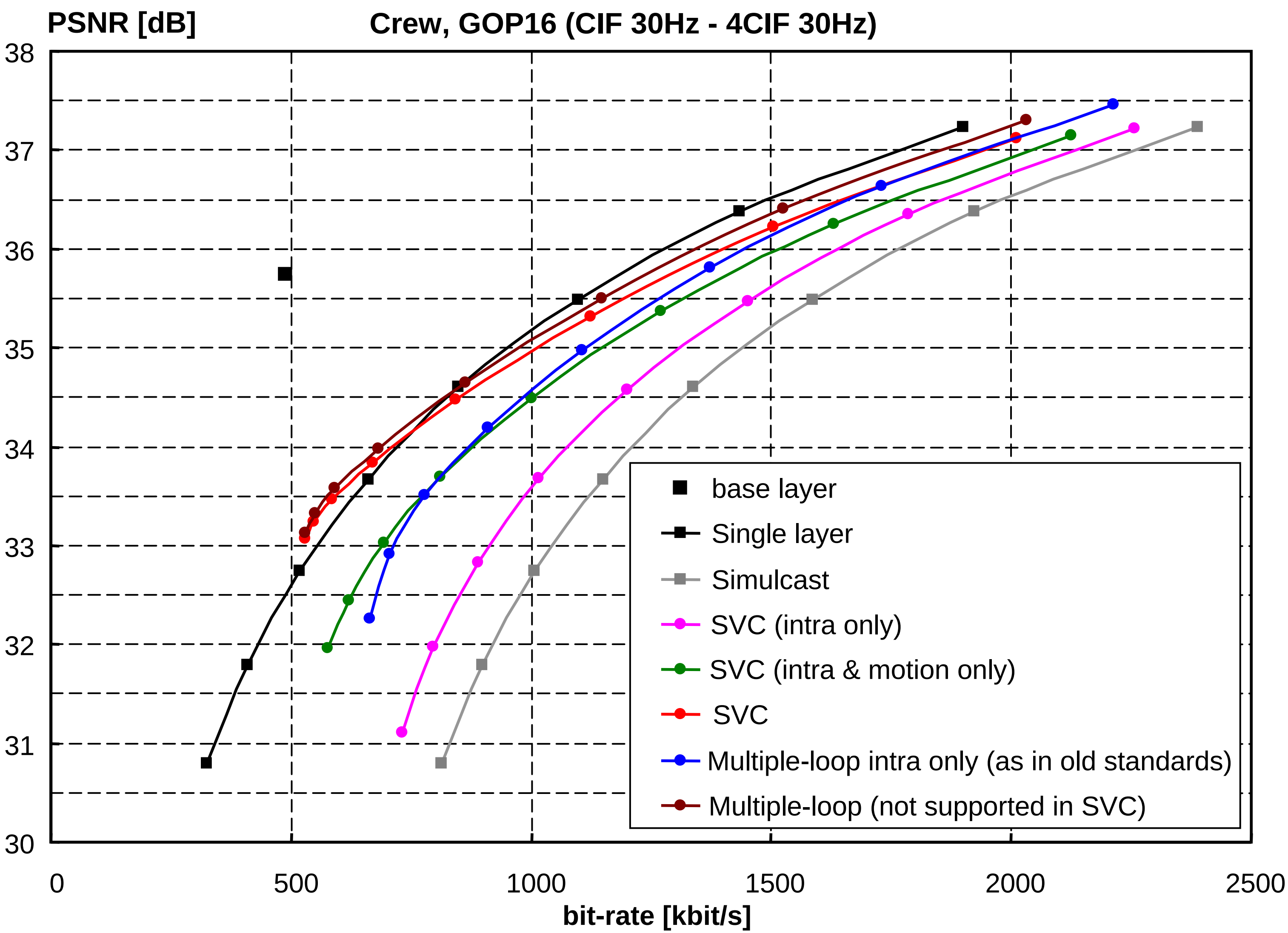


Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

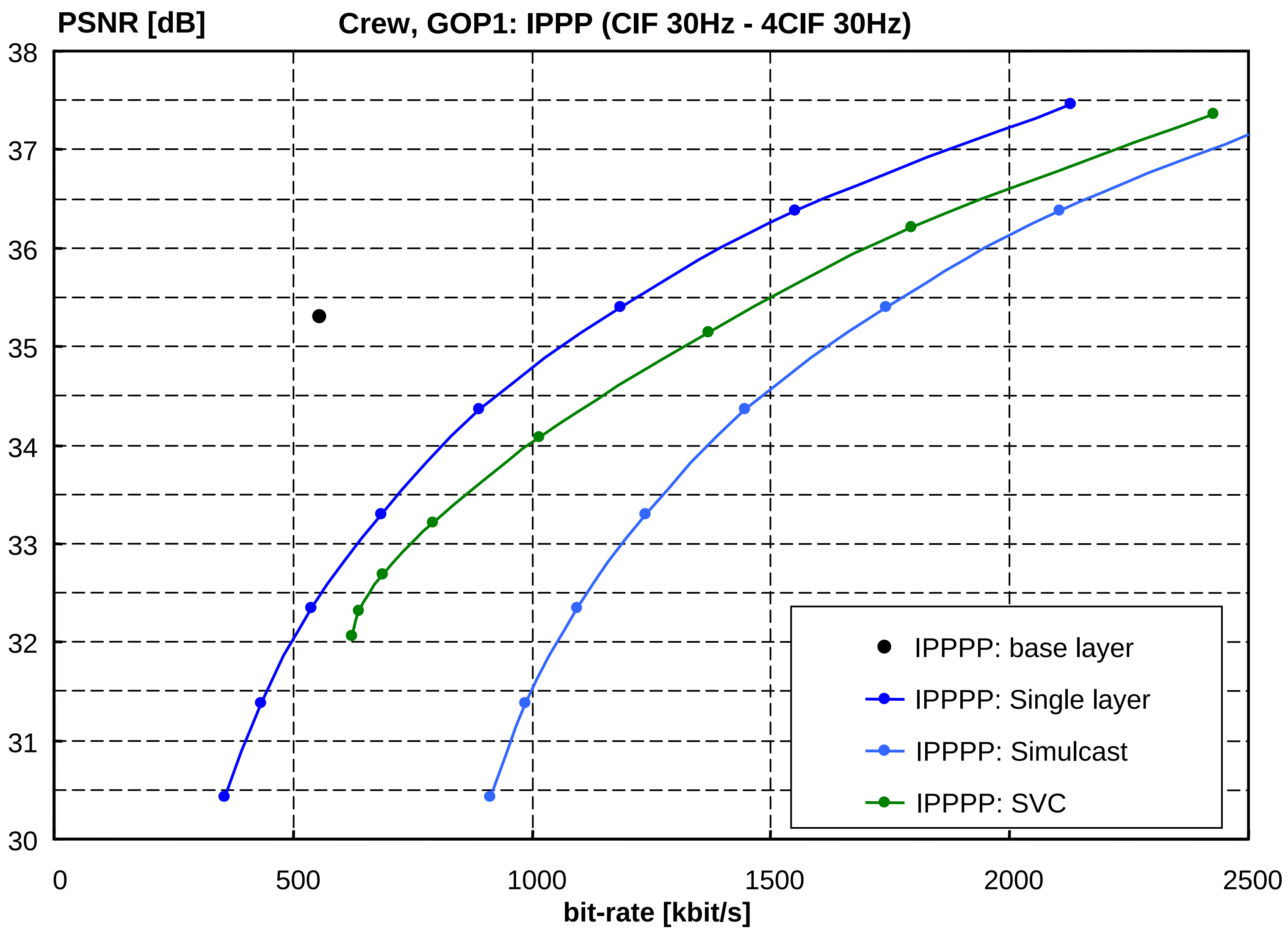
Single-loop Decoding

- **MPEG-2, H.263, MPEG-4**
  - reconstructed base layer pictures are used for inter-layer prediction
  - requires multiple motion-compensation loops
- **SVC**
  - inter-layer intra prediction is restricted to base layer macroblock that are coded in intra mode
  - single motion compensation loop (including deblocking) is sufficient at decoder side
  - only pictures of highest layer are stored in the decoded picture buffer
  - encoder should be operated in multiple-loop mode for avoidance of drift and best compression results
  - additionally required complexity for supporting spatial scalability is smaller than for MPEG-2, H.263, MPEG-4
- **Impact on Coding Efficiency**
  - Minor impact (0 – 0.5 dB)
  - New inter-layer prediction concepts (motion & residual) in comparison to MPEG-2, H.263, MPEG-4

Coding Efficiency of Different Tools



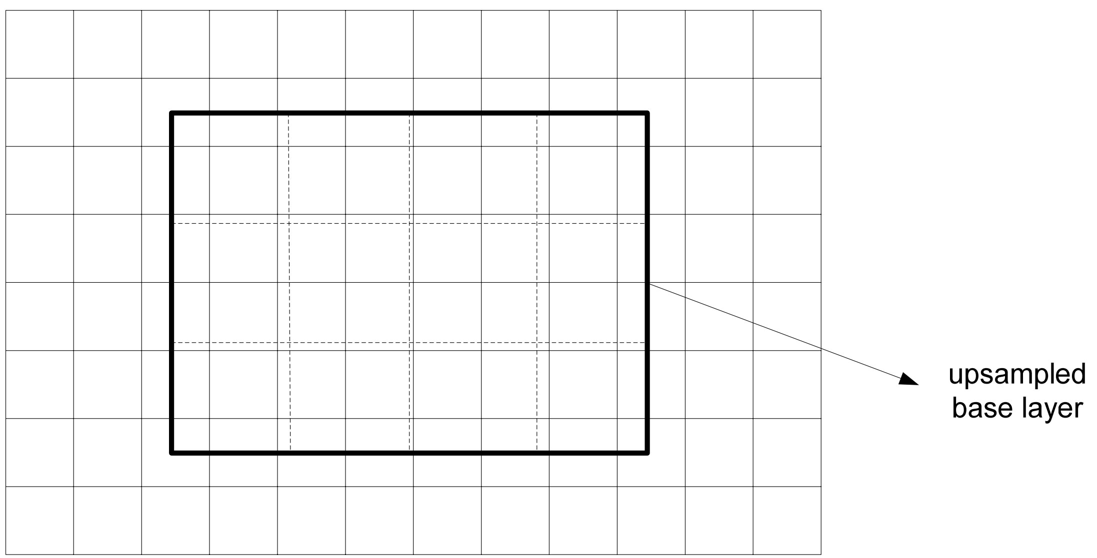
Impact of Temporal Prediction Structure



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Extended Spatial Scalability

- **In the typical case, macroblock boundaries over subsequent spatial resolutions are aligned (e.g. QCIF -> CIF -> 4CIF)**
- **When the up-sampling factor is non-dyadic, or when parts of the image are cropped in lower resolution, this is no longer the case**

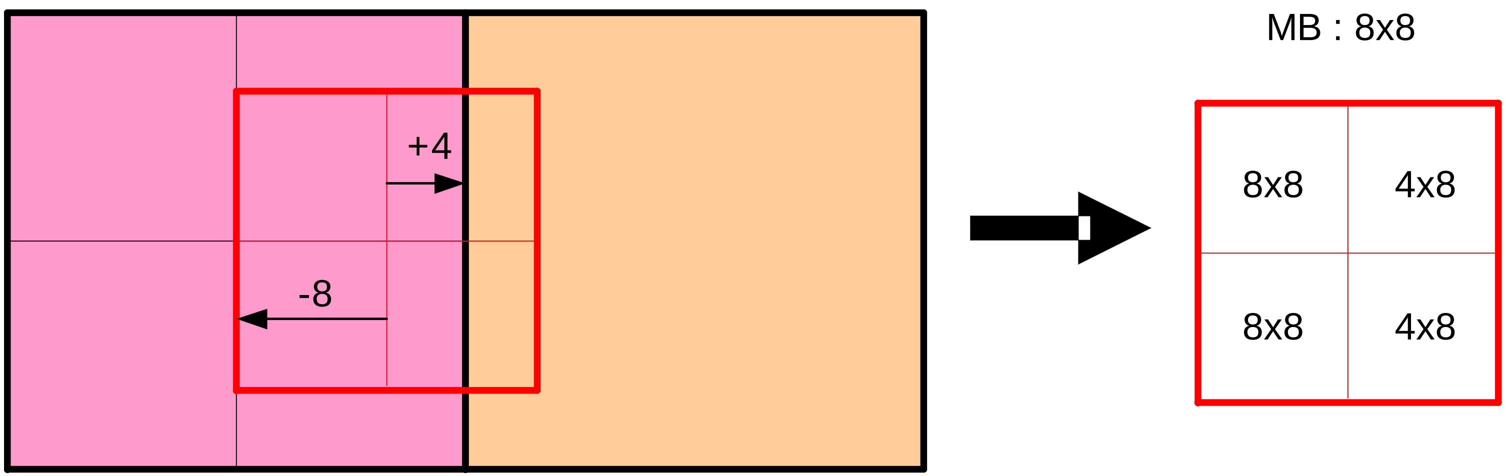


Extended Spatial Scalability in SVC

- **Arbitrary resolution ratios**
  - e.g. SD -> 720p, 720p -> 1080p
- **Cropping of base layer signal**
  - enhancement layer contains new image regions
  - e.g. 4:3 -> 16:9
- **Cropping of enhancement layer signal**
  - enhancement layer zooms out a region of the base layer
  - surveillance: higher quality picture of interesting area
- **Modification on a picture basis**
  - resolution ratio
  - cropping of base and enhancement layer

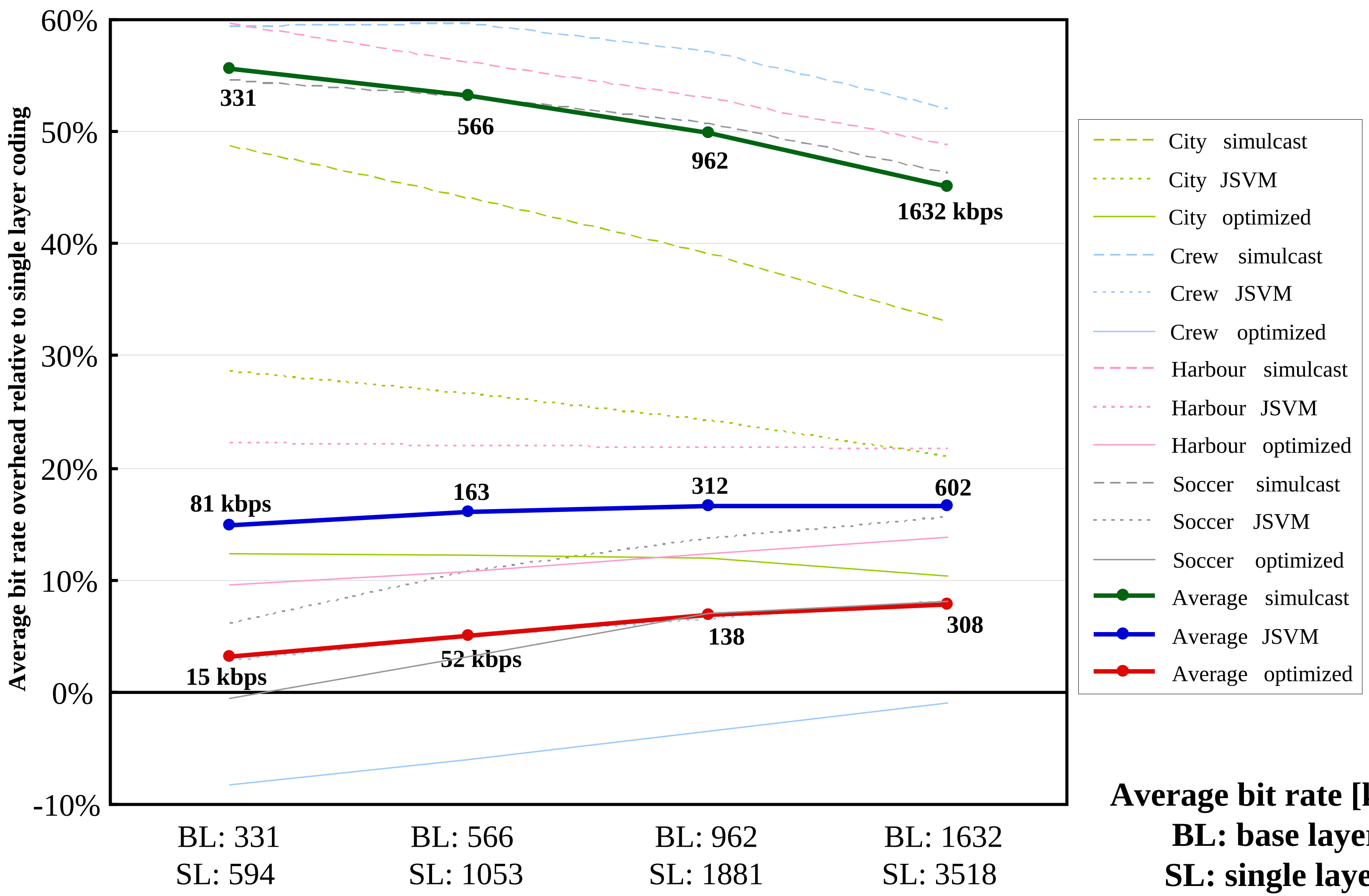
Extended Spatial Scalability in SVC

- **Generalization of inter-layer prediction concepts**
  - Upsampling filters for intra prediction
  - Upsampling filters for residual prediction
  - Inheritance of macroblock modes and partitioning
  - Inheritance of motion vectors



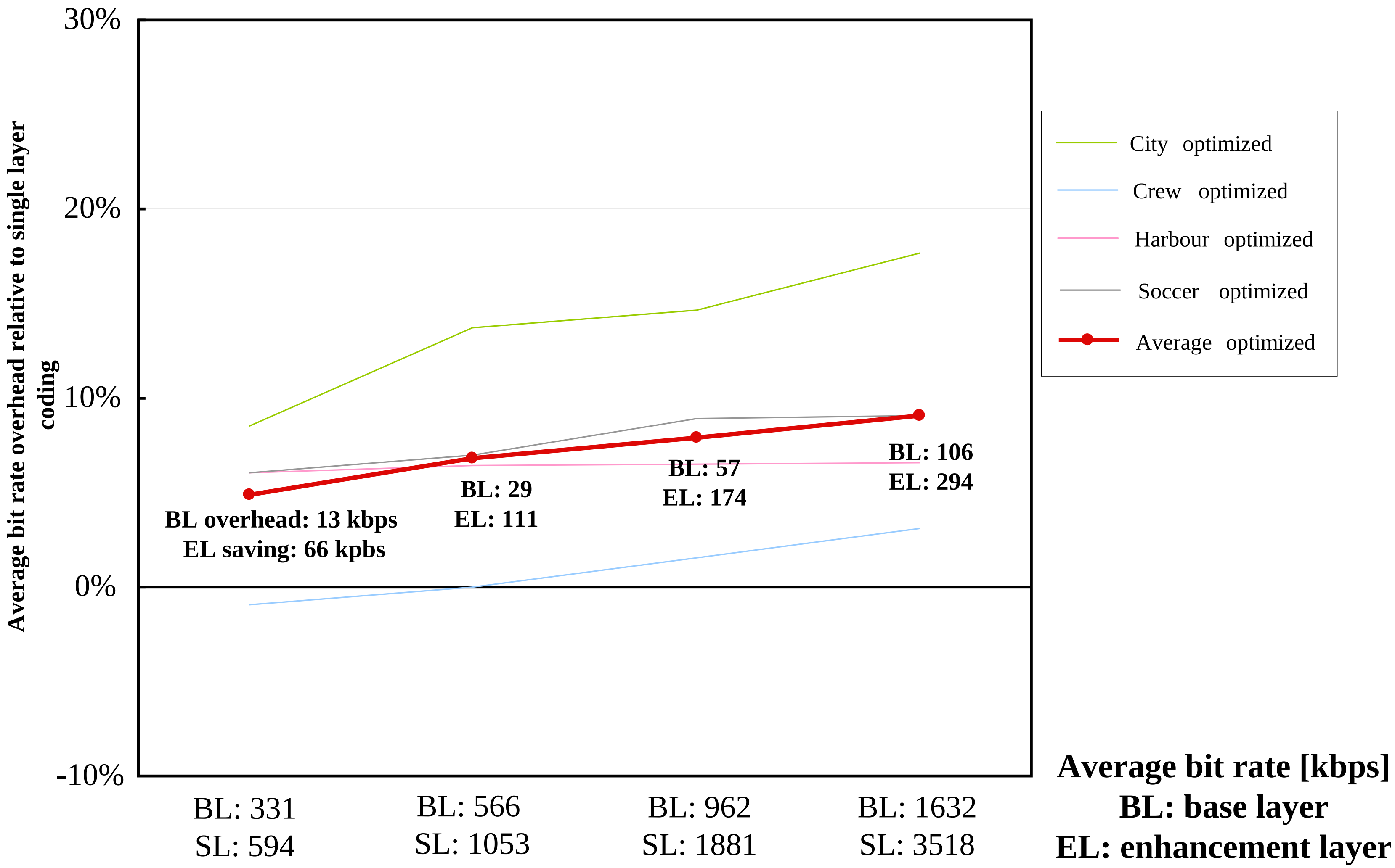
Example: Derivation of MB partitioning

Enhancement Layer Bit rate Overhead for CIF/30Hz ➔ 4CIF/30Hz Spatial Scalability

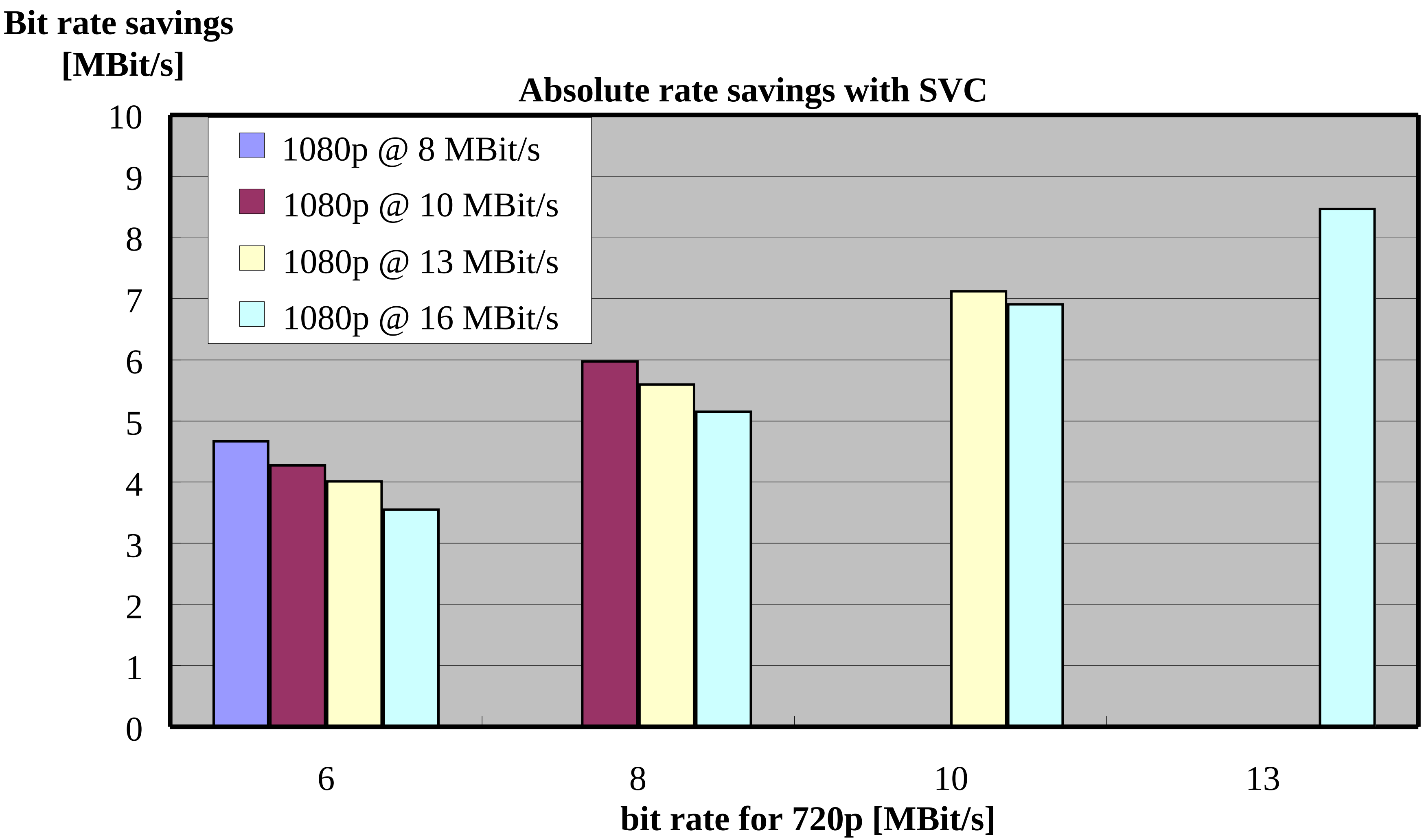


Average bit rate [kbps]
BL: base layer
SL: single layer

Base Layer Bit rate Overhead in Joint Optimization for CIF/30Hz ➔ 4CIF/30Hz Spatial Scalability



720p to 1080p: Absolute Rate Savings Relative to Simulcast



3. Profiles, Levels, and Systems Support

Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

H.264/AVC Profiles I

- H.264/AVC before SVC contained 11 profiles:
  - Baseline, Extended, Main, High,
  - High 10, High 4:2:2, High 4:4:4 Predictive,
  - High 10 Intra, High 4:2:2 Intra, High 4:4:4 Intra, CAVLC 4:4:4 Intra
- Baseline profile (e.g., Videoconferencing & Wireless)
  - I and P picture types (not B)
  - In-loop deblocking filter
  - 1/4-sample motion compensation
  - Tree-structured motion segmentation down to 4x4 block size
  - VLC-based entropy coding (CAVLC)
  - Some enhanced error resilience features
    - Flexible macroblock ordering/arbitrary slice ordering
    - Redundant slices
  - **Note**: No support for interlaced video in Baseline

H.264/AVC Profiles II

- Extended profile (esp. Streaming/Internet)
  - All Baseline features
  - B pictures
  - Adaptive weighting for B and P picture prediction
  - Picture and MB-level frame/field switching
  - *More error resilience: Data partitioning*
  - *SP/SI switching pictures*
  - Note: Extended *is* a superset of Baseline (but not of Main)
- Main profile (esp. Broadcast/Entertainment)
  - All Baseline features *except enhanced error resilience features*
  - B pictures
  - Adaptive weighting for B and P picture prediction
  - Picture and MB-level frame/field switching
  - *CABAC*
  - **Note**: Main is not exactly a superset of Baseline

H.264/AVC Profiles III

- High profile (esp. Broadcast/Entertainment)
  - All Main profile features
  - 8x8 transform and intra prediction
  - Frequency weighting for 4x4 and 8x8 transform
  - Note: High *is* a superset of Main profile
- Extensions of High profile tools:
  - High 10: bit-depth of 10 bit with High profile tools
  - High 4:2:2: color sampling of 4:2:2 and 10 bit
  - High 4:4:4 Predictive: 4:4:4 and 14 bits
- Restrictions High 10, High 4:2:2, and High 4:4:4 predictive to intra-only tools specify High 10 Intra, High 4:2:2 Intra, and High 4:4:4 Intra profiles, respectively
- Restriction of High 4:4:4 Intra to CAVLC entropy coding specifies CAVLC 4:4:4 Intra profile

SVC Profiles

- **Scalable Baseline**
  - H.264/AVC base layer: “restricted” Baseline profile
  - SVC enhancement layer: I, P, and restricted B slices, CAVLC and CABAC (from level 2.1), spatial factors 1:1, 1:1.5, 1:2
- **Scalable High**
  - H.264/AVC base layer: High profile
  - SVC enhancement layer: I, P, and B slices, CAVCL and CABAC, all spatial factors
- **Scalable High Intra**
  - H.264/AVC base layer: High Intra profile
  - SVC enhancement layer: I slices, CAVCL and CABAC, all spatial factors

H.264/AVC Levels

Level	Max Picture Size (example)	Max Frame Rate	Max Bit Rate
1	QCIF (176x144)	15	64 kbps
1b	QCIF	15	128 kbps
1.1	CIF (352x288)	7.5	192 kbps
1.2	CIF	15	384 kbps
1.3	CIF	30	768 kbps
2	CIF	30	2 Mbps
2.1	625 HHR (352x576)	25	4 Mbps
2.2	625 SD (720x576)	12.5	4 Mbps
3	625 SD	25	10 Mbps
3.1	720p HD (1280x720)	30	14 Mbps
3.2	SXGA (1280x1024)	42	20 Mbps
4	2Kx1K (2048x1024)	30	25 Mbps
4.1	2Kx1K	30	62.5 Mbps
4.2	2Kx1080	60	62.5 Mbps
5	3672x1536	26.7	135 Mbps
5.1	4096x2304	26.7	240 Mbps

Baseline Profile

3GPP

DVB-H

T-DMB

Resolution increase towards High Profile

DVB-C/S/T/IPDC

Blu-Ray Disc
HD DVD

SVC Levels

- **Levels for all scalable profiles are identical**
- **Numbers are kept from H.264/AVC**
- **A 2-layer scalable bit-stream is categorized into a level by**
  - Number of macroblocks of the enhancement layer
  - Overall bit-rate and buffer sizes
- **A 2+X-layer scalable bit-stream is categorized by**
  - Let the layers be labelled as 0 ... 2+X-1 with layer 0 being the base layer
  - Number of macroblocks of layer 2+X-1 plus number of macroblocks of layers 0 ... X-1 times 0.5
  - Overall bit-rate and buffer sizes

Systems Support

- **File format (MPEG):**
  - Referred to as the SVC File Format
  - New amendment 2 to the AVC file format progressed in July to FPDAM
  - Can be finalized in January 2008
- **RTP payload (IETF):**
  - Builds on top of H.264/AVC RTP payload (RFC 3984)
  - Will become a new RFC
  - Can be finalized by the end of 2007 (with WG last call after November IETF meeting)
- **H.222.0 | MPEG-2 Systems (MPEG - common text with SG16):**
  - New amendment 3 to MPEG-2 Systems (2006 edition) progressed in July to PDAM
  - Enables the carriage of scalable video data within MPEG-2 program and transport streams
  - Can be finalized in April 2008

Summary I

- **SVC is an efficient and straight-forward extension of H.264/AVC**
- **Mainly uses H.264/AVC building blocks**
- **Complexity is similar to single layer H.264/AVC decoding (single motion-compensation loop decoding in SVC)**
- **Coding efficiency is significantly better than previous scalable video coding standards**

Summary II

- **H.264/AVC hierarchical B or P pictures allow temporal scalability**
- **Scalability**
    - Hierarchical picture structures enhance scalability
    - Layered representation of pictures
  - **SNR scalability**
    - Layers: residual coding of H.264/MPEG4-AVC
    - Medium granularity of scalability by proper adaptation
  - **Spatial scalability**
    - Oversampled pyramid and independent temporal decomposition in each spatial layer
    - Inter-layer prediction: Intra, motion, residuals
    - One MC loop decoding through restriction of prediction from lower layer reconstructed signals to Intra-only

Summary III

- Scalable Video Coding is the next step video applications
  - SVC realized through layered extension of H.264/AVC
  - Power adaptation: decode appropriate part of bit-stream
  - Graceful degradation when the “right” parts of the bitstream get lost
  - Format adaptation: backwards compatible extension in mobile TV: QVGA⇨VGA
  - Interlaced and other ratios than 2:1 are also specified, e.g.. 1.5:1 for 720p to 1080p
- What’s next in SVC standardization?
  - Bit-depth scalability (8bit 4:2:0 -> 10bit 4:2:0)
  - Color format scalability (4:2:0 -> 4:4:4)

Further Reading

- Special issue in IEEE Transactions on Circuits and Systems for Video Technology on H.264/AVC: July 2003

  (T. Wiegand, G. J. Sullivan, G. Bjøntegaard, A. Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003)*

- Special issue in IEEE Transactions on Circuits and Systems for Video Technology on H.264/AVC: September 2007

  (H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", to appear in IEEE Transactions on Circuits and Systems for Video Technology, September 2007)*

- Free download of standard (published by both ITU-T and ISO/IEC as so-called "twin-texts"): http://www.itu.int/rec/T-REC-H.264/en

  ITU-T and ISO/IEC JTC 1, "Advanced Video Coding for Generic Audiovisual Services", ITU-T Recommendation H.264 & ISO/IEC 14496-10 (MPEG4-AVC), Version 1, May, 2003; Version 2, January 2004; Version 3 (with FRExt), Sept. 2004; Version 4, July, 2005, Version 5, July 2007 (with Intra profiles and SVC).

*Papers can be downloaded at iphome.hhi.de/wiegand/