Scalable Video Coding

Jens-Rainer Ohm

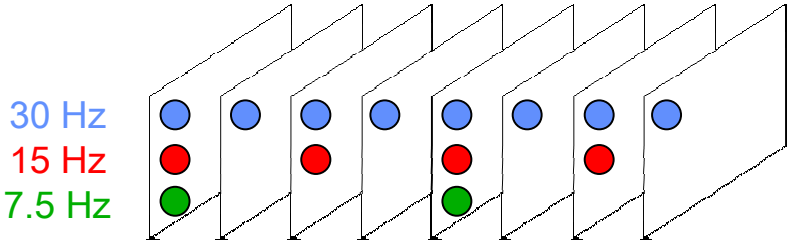
RWTH Aachen University
Chair and Institute of Communications Engineering
ohm@ient.rwth-aachen.de
http://www.ient.rwth-aachen.de

Outline

1. Introduction – Scalability and Applications
2. SVC – Technical Solution
3. Profiles, Levels, and Systems Support
4. Conclusions

1. Introduction – Potentials and Applications

- **Temporal**: change of frame rate



- **Spatial**: change of frame size


QCIF


CIF


TV


HDTV

- **Fidelity**: change of quality (a.k.a. SNR)


High rate


Low rate

- **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)
- **But*:* except for *T*, this had large additional rate cost and was never accepted by the market**
- **New SVC: Amendment to AVC standard**
  - **ISO/IEC 14496-10:2005/Amd.3 (finalized July 2007)**
  - **to be published in new edition**
  - **Goal: Efficient Scalability**
  - **Only moderate complexity increase, moderate rate increase as compared to single layer coding, supporting big variability of S,T,F layers**

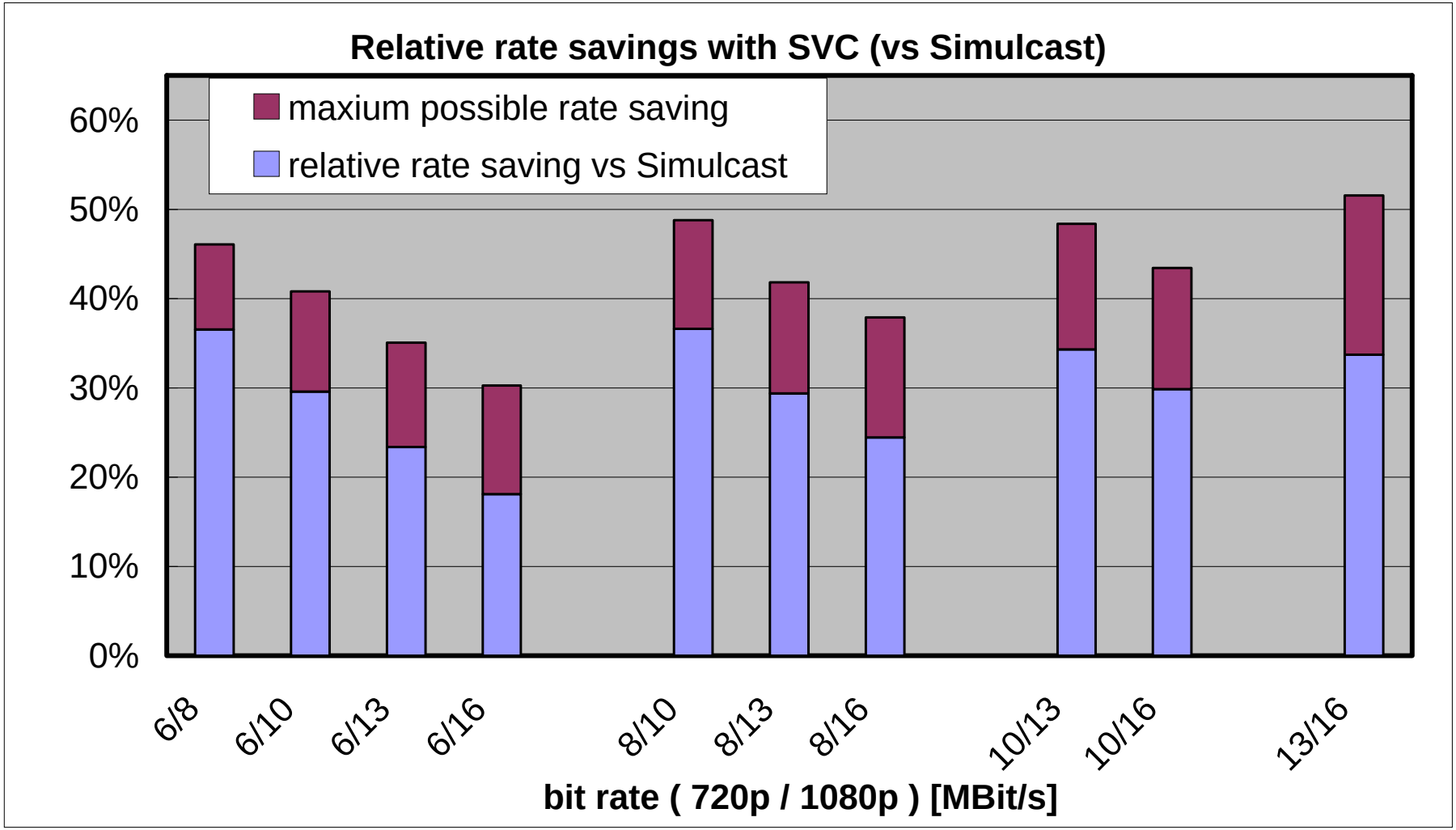
Supported Functionalities and Potential Applications

- **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 and multicast transmission services with varying
    - channel conditions (throughput, errors)
    - device types (supported spatio-temporal resolution by decoder, display and power) –
- **Simplification of adaptation (e.g. compared to transcoding)**

Example: Backwards-Compatible HD Format Extension

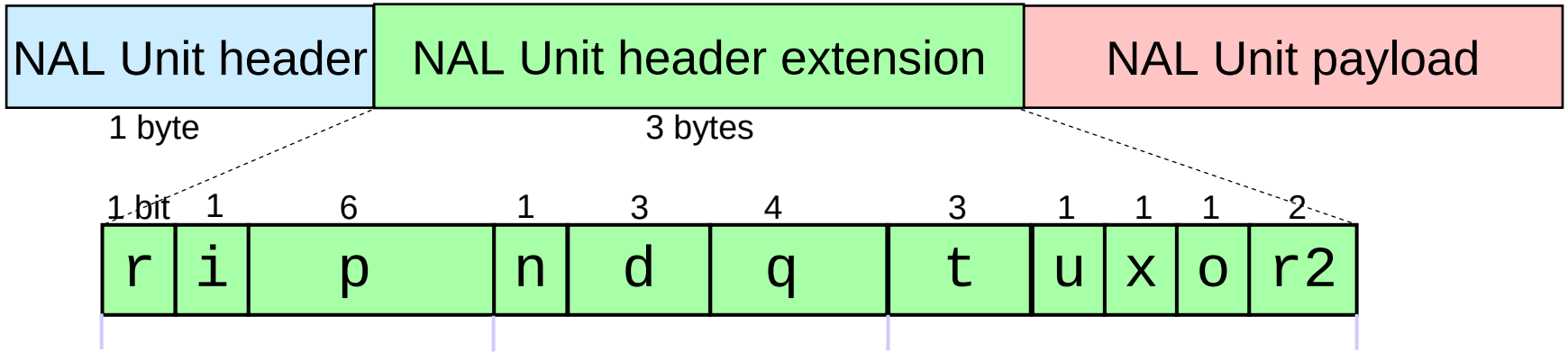
- First generation HD formats are either 720p50/60 or 1080i25/30
- Broadcasters want to introduce 1080p50/60 while maintaining the existing customer base (and devices)
- Broadcasting of 1080p50/60 SVC bitstreams enhances video resolution from 720p50/60 or 1080i25/30
- Provide enhanced set-top box with additional 1080p50/60 SVC (and AVC base layer 720p/1080i) capability without the need to exchange recently shipped set-top box

720p to 1080p: % Bit Savings Relative to Simulcast



2. SVC: Technical Solution

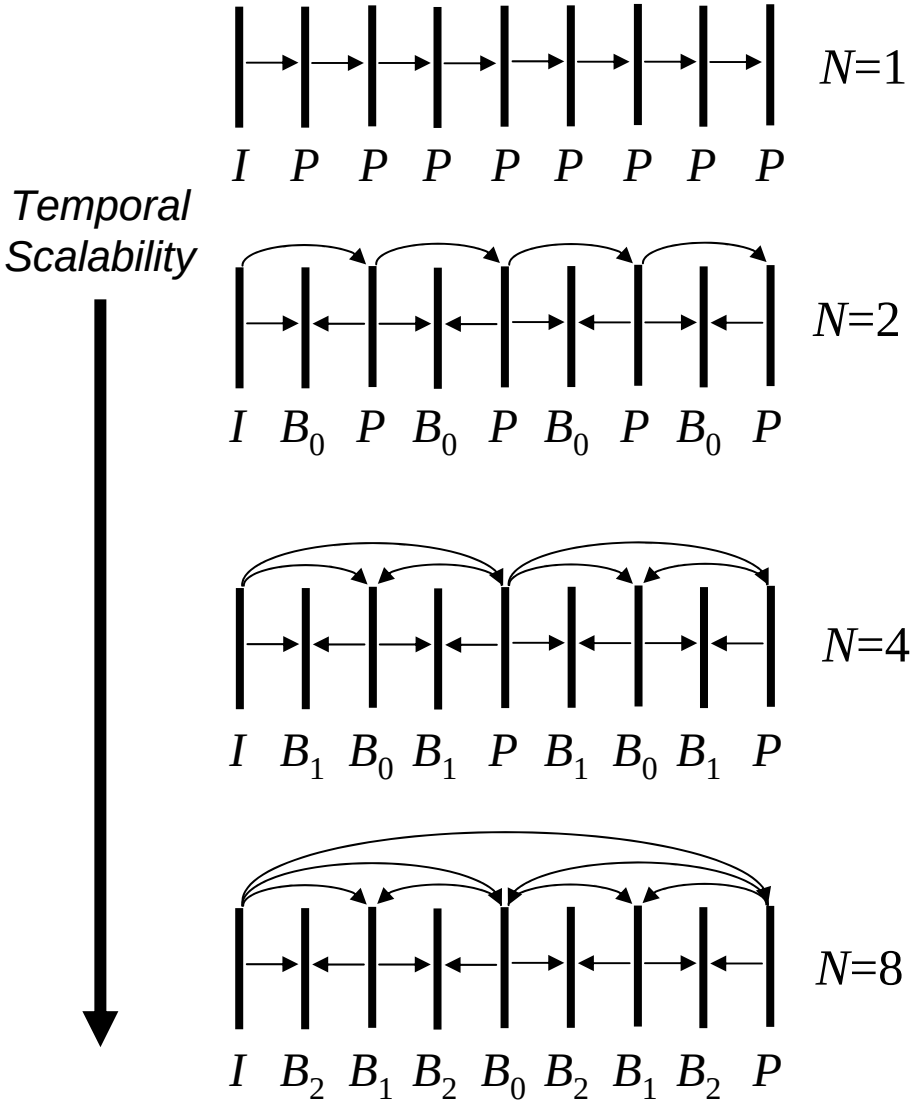
Network Abstraction Layer



- **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

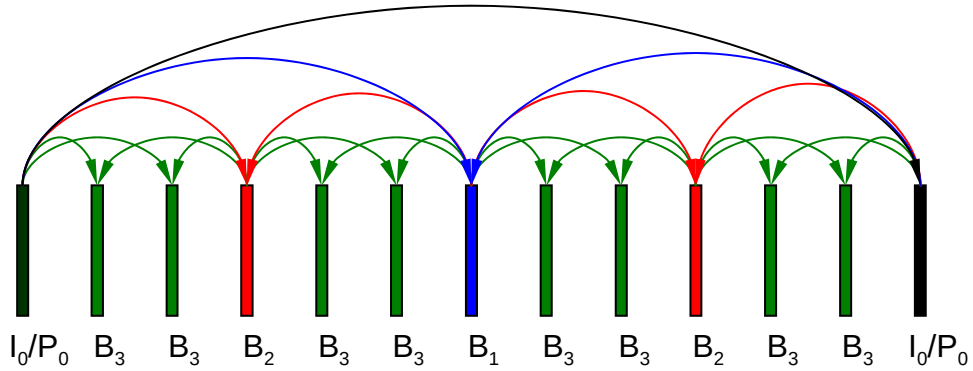
Temporal Decomposition of Video



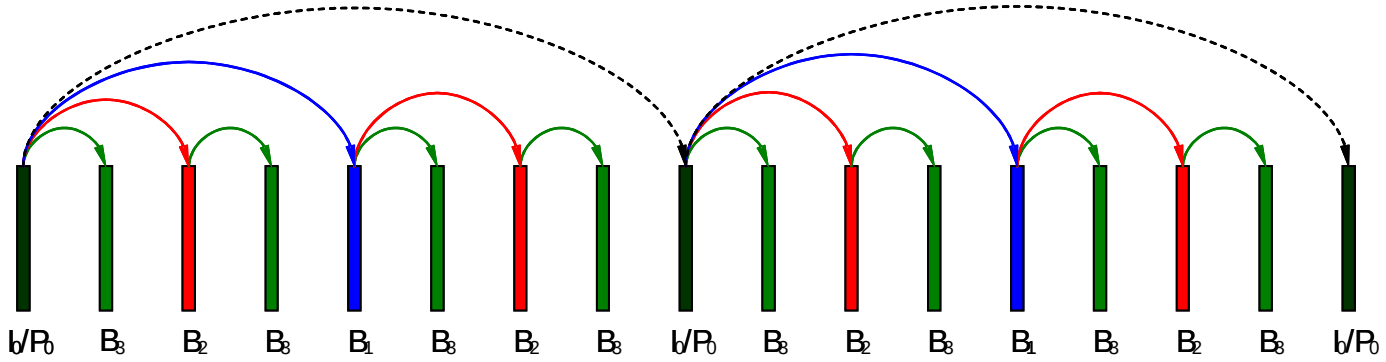
- Hierarchical prediction structures
  - efficient method for enabling temporal scalability
  - not restricted to dyadic temporal scalability
  - can be combined with multiple reference pictures
  - delay can be controlled by restricting the use of future pictures for MCP

Hierarchical prediction structures

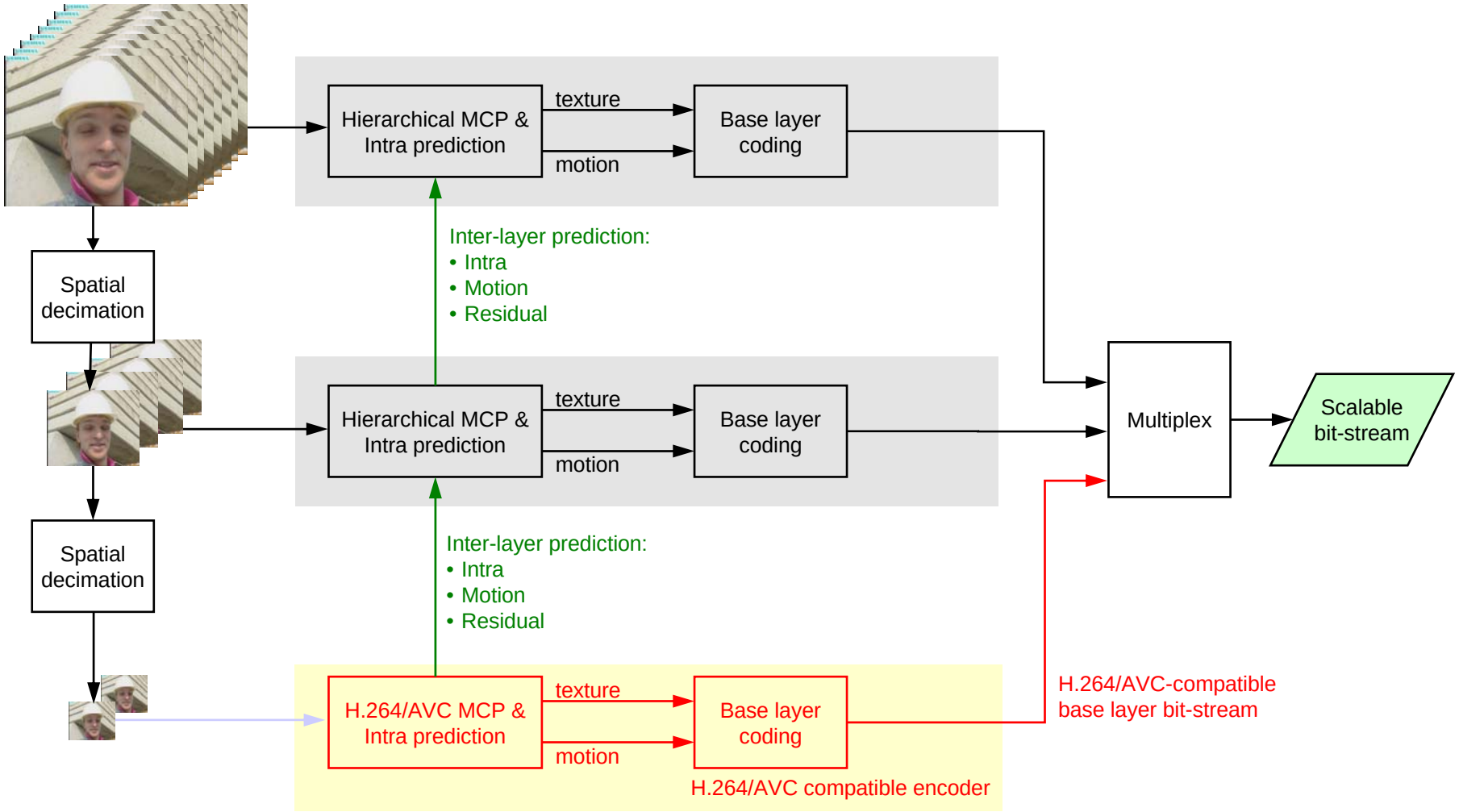
- Non-dyadic temporal scalability



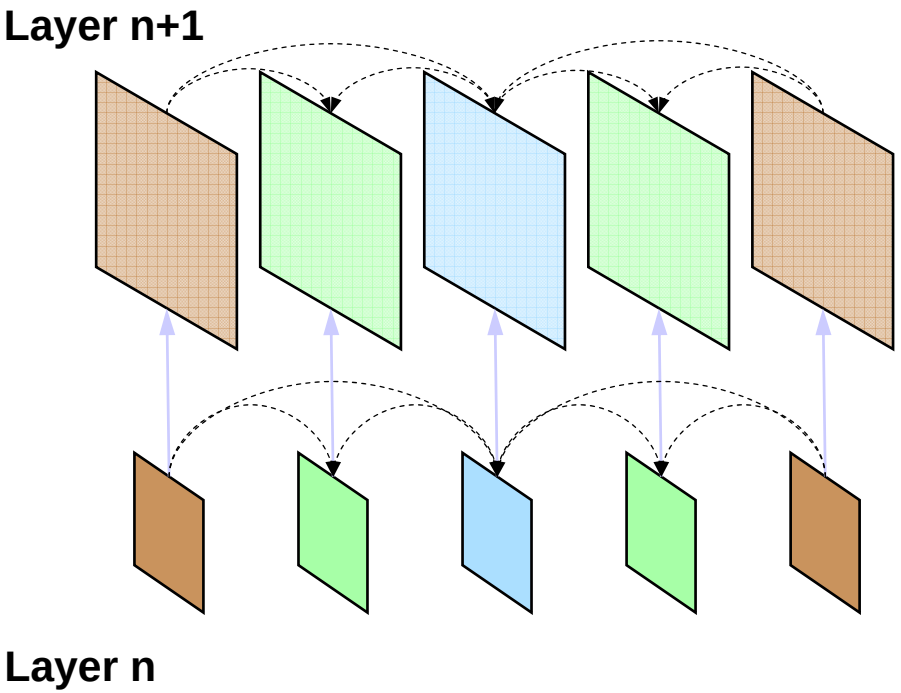
- Low-delay prediction structure (structural delay is 0)



Typical Encoding for Spatial Scalability

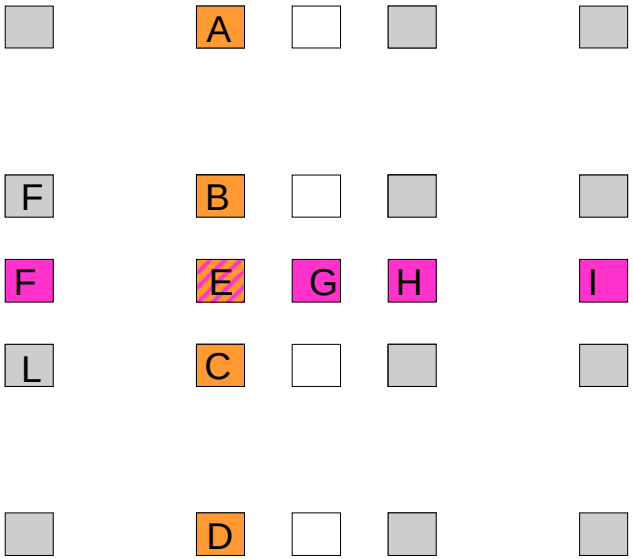


Layered Coding for Spatial Scalability

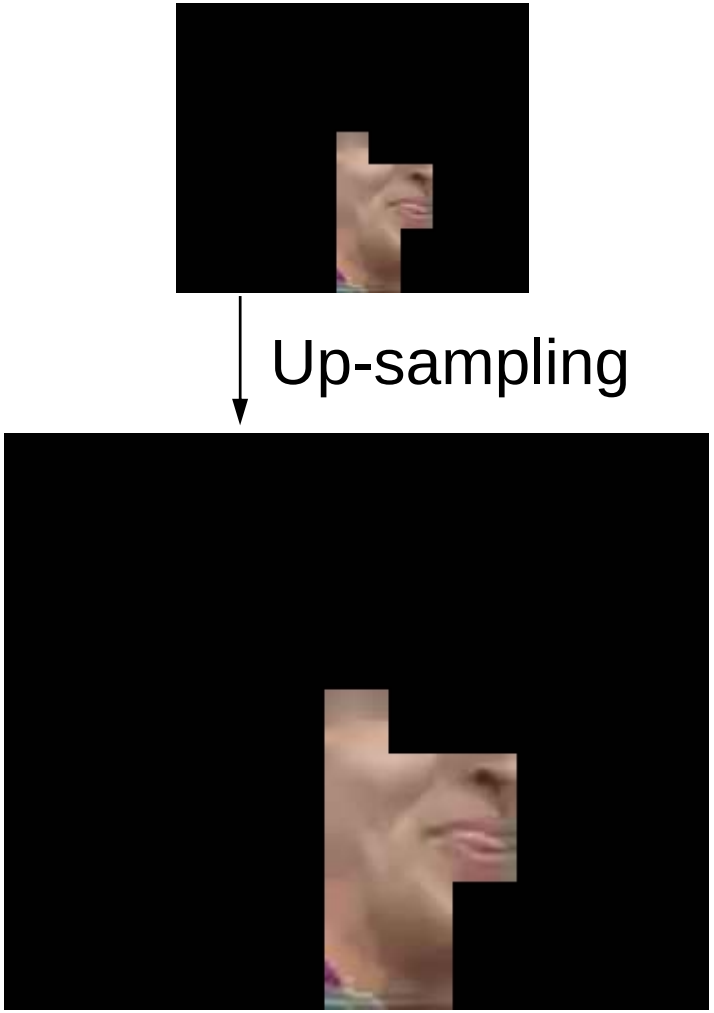


- Layered coding
  - Oversampled pyramid for each resolution:
    e.g. QCIF, CIF, 4CIF, 16CIF
  - MC prediction structures of all layers are aligned
- Inter-layer prediction:
  Switchable prediction (with upsampling)
  - Prediction of intra macroblocks
    (cp. MPEG-2, H.263, MPEG-4)
  - No prediction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

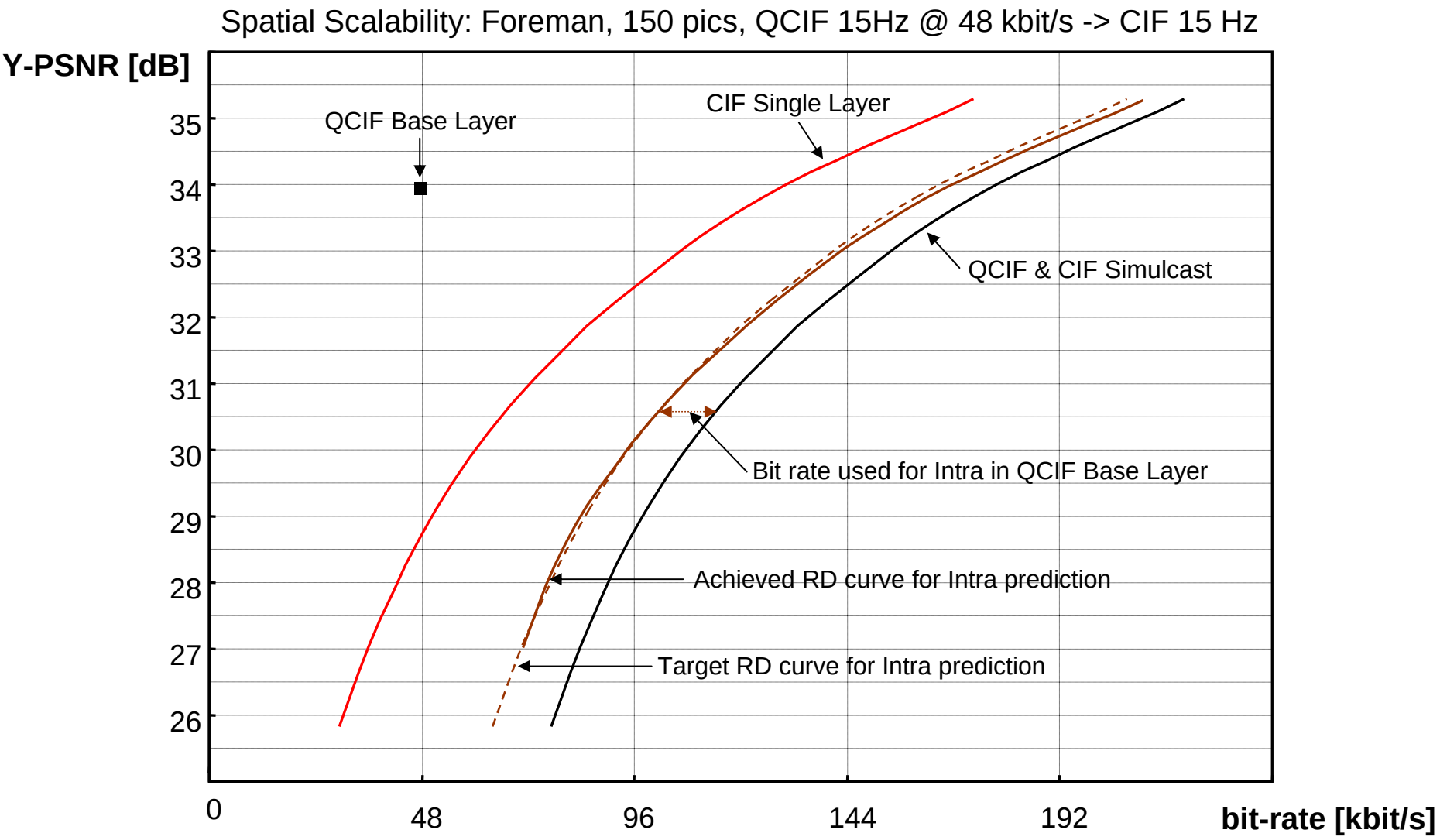
4-tap filter: [-3, 19, 19, -3]



full sample positions
half sample positions

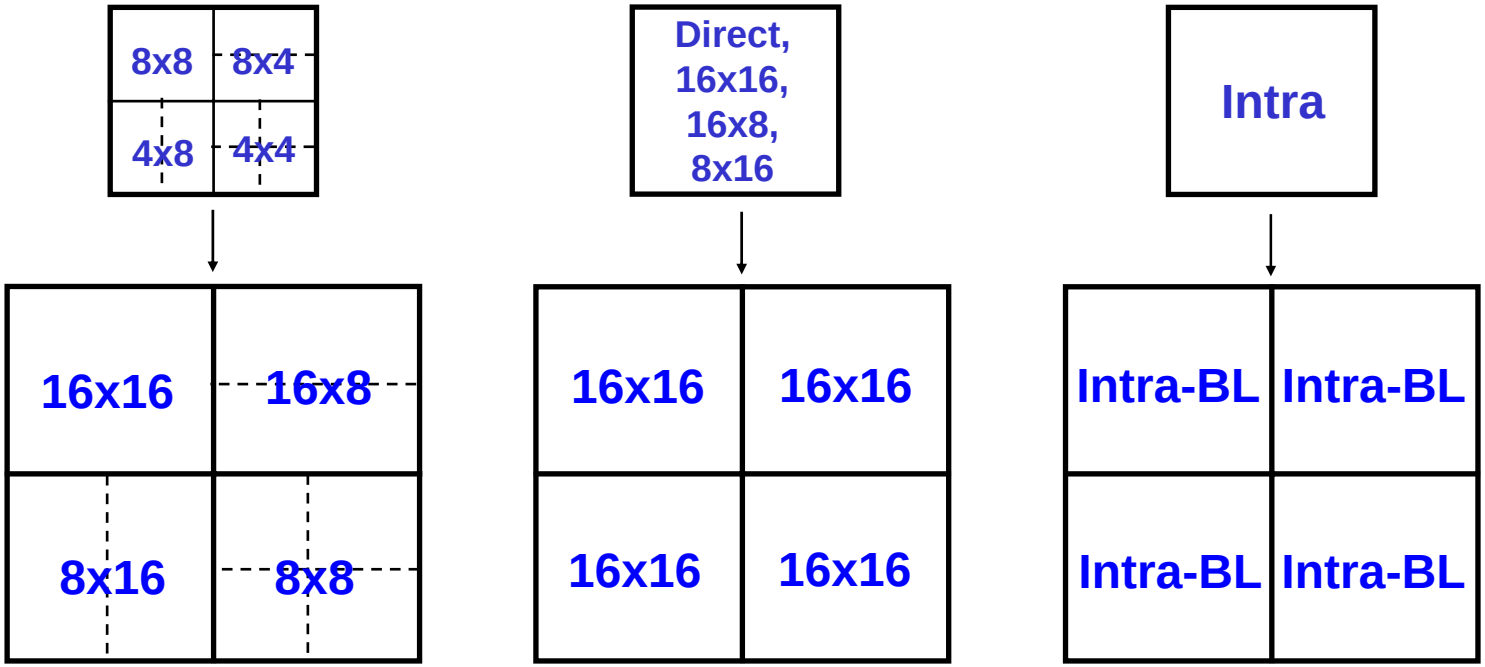


Spatial Prediction of Data

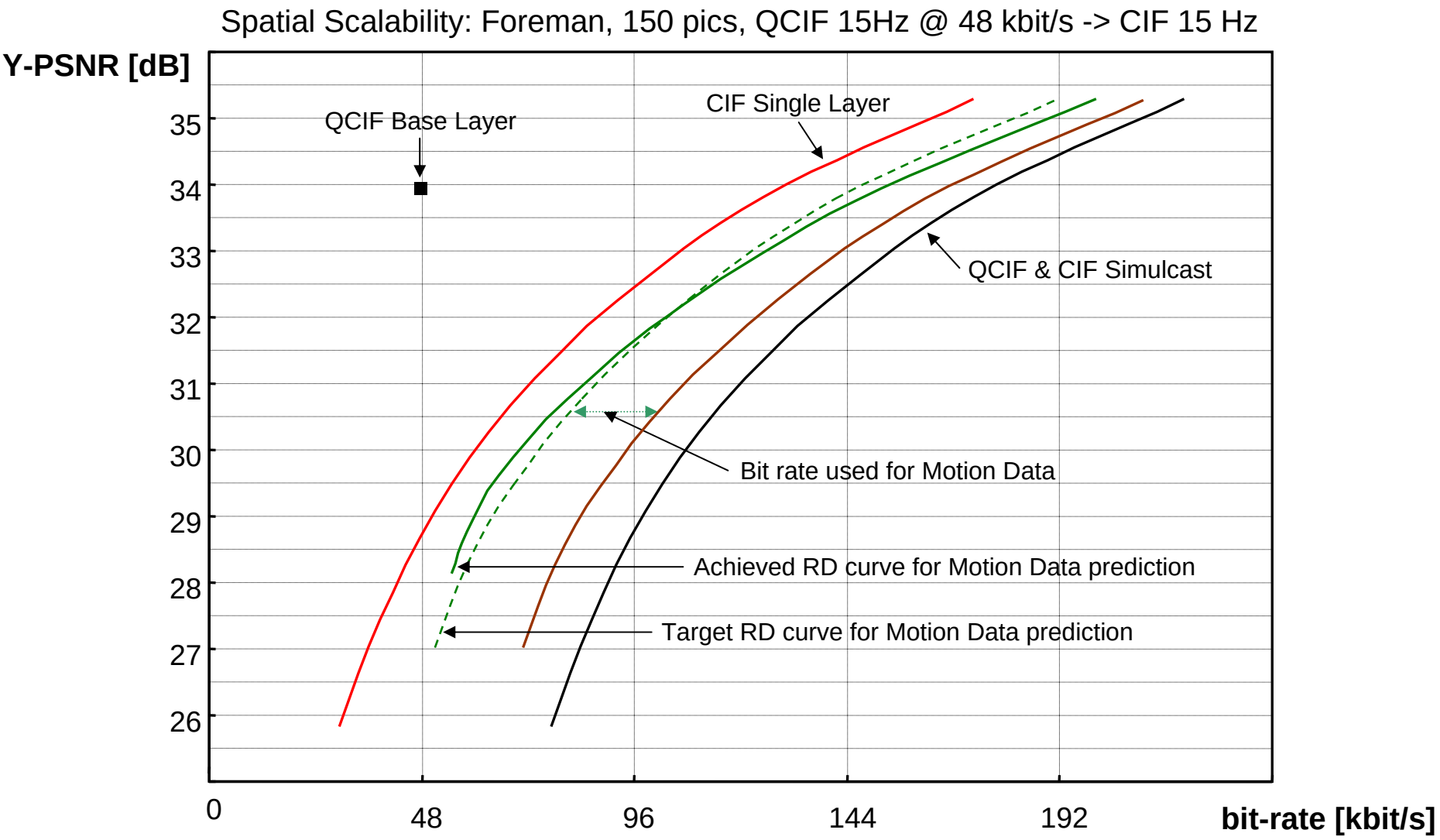


Spatial Prediction of Motion Data

- **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)**

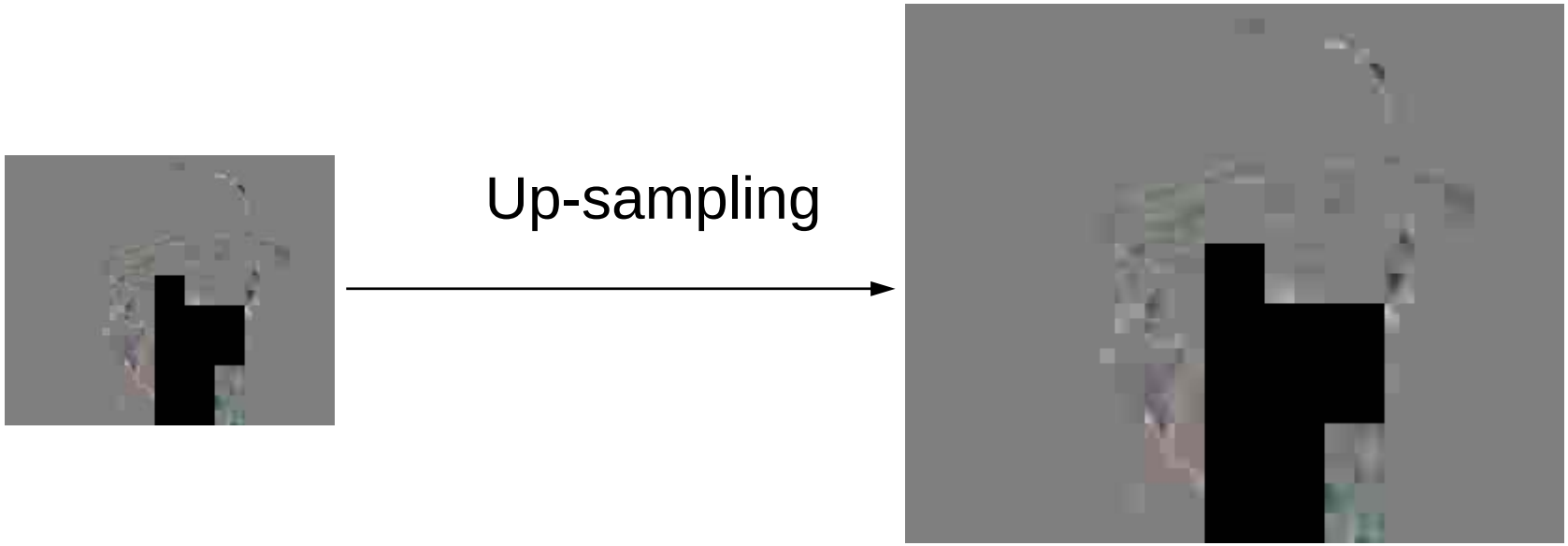


Spatial Prediction of Data

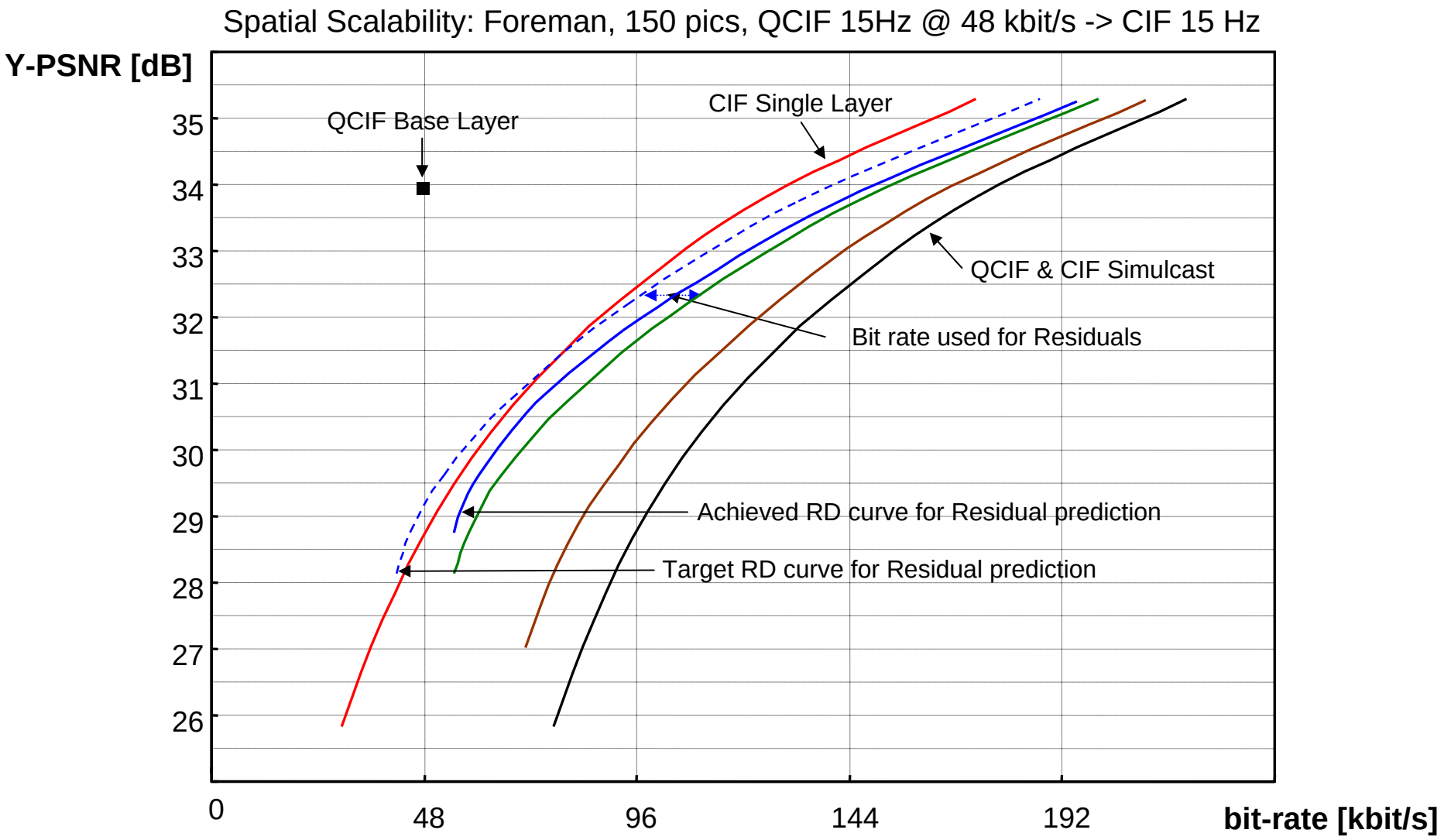


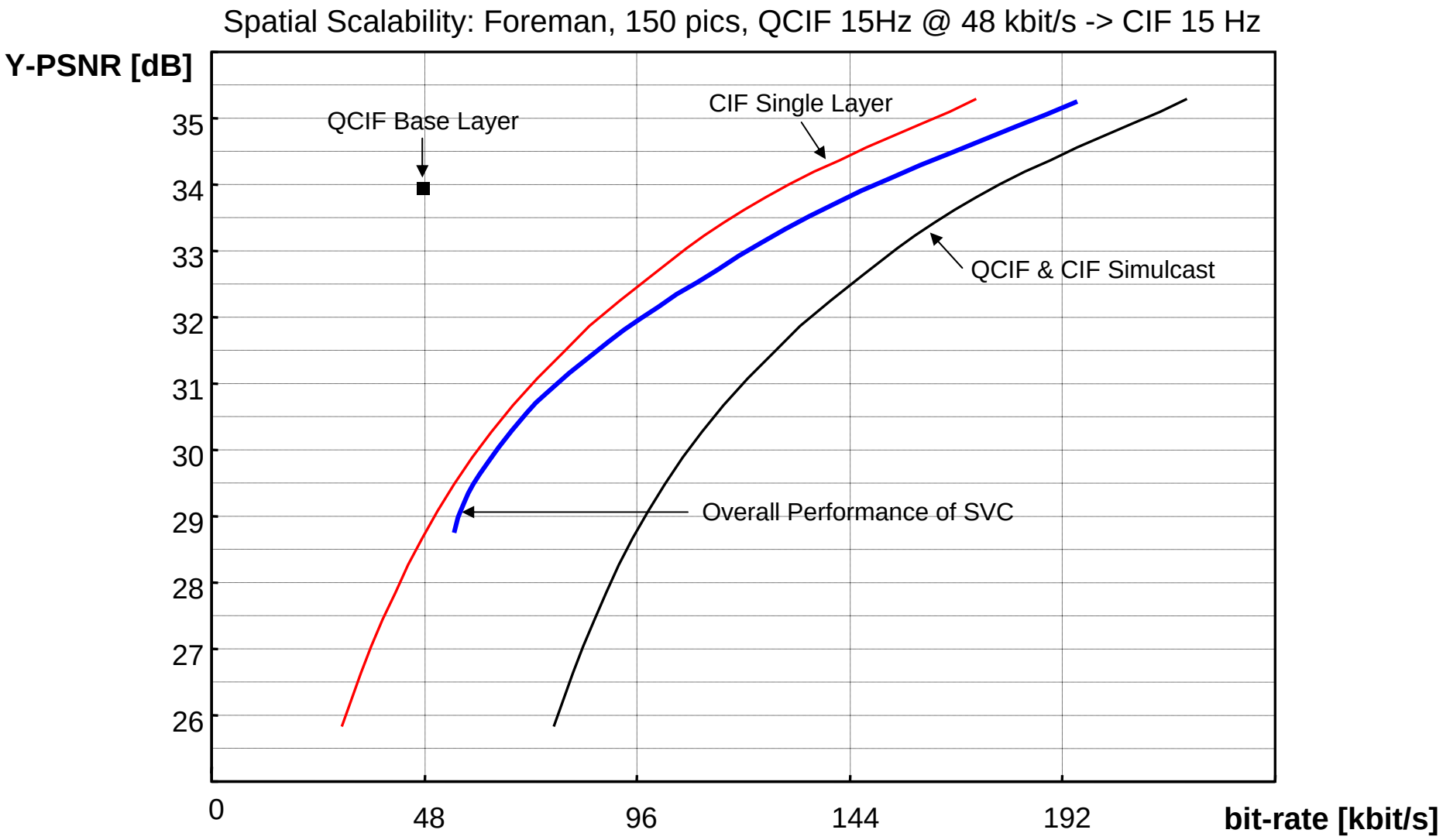
Bi-linear up-sampling filter

- Block-wise processing
- Special processing at block boundaries



Spatial Prediction of Data



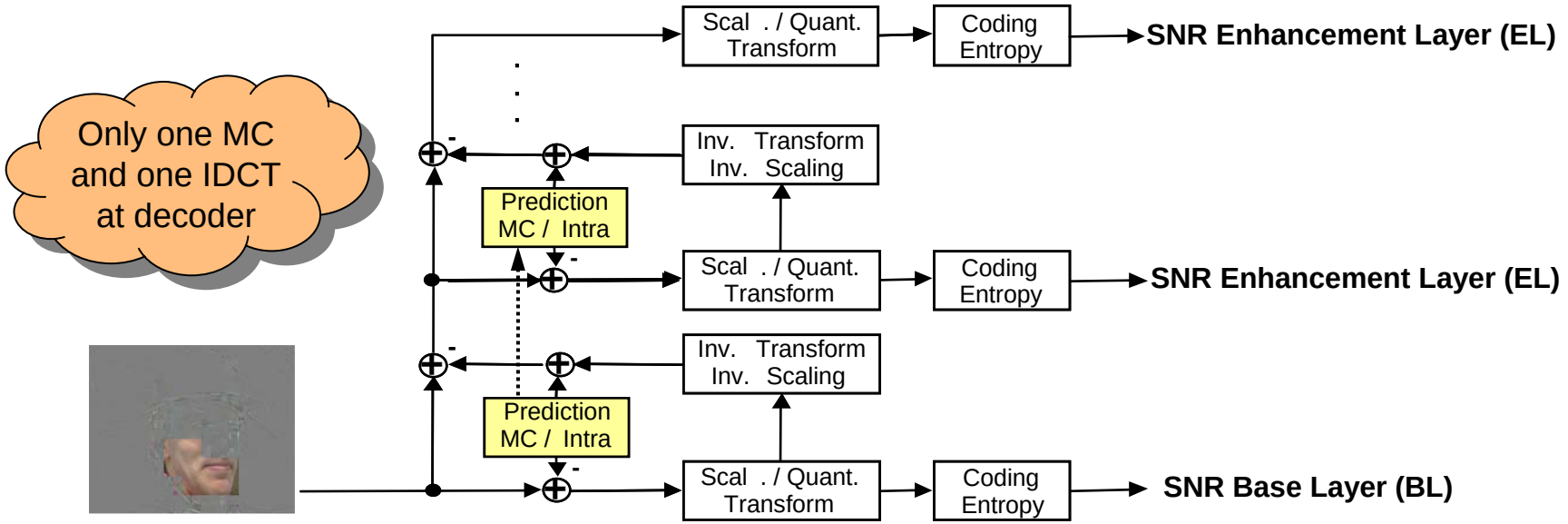

Spatial Scalability: Foreman, 150 pics, QCIF 15Hz @ 48 kbit/s -> CIF 15 Hz
Y-PSNR [dB]
35
34
33
32
31
30
29
28
27
26
0
48
96
144
192
bit-rate [kbit/s]
QCIF Base Layer
CIF Single Layer
QCIF & CIF Simulcast
Overall Performance of SVC

- 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

SNR Scalability in SVC – 2 Methods

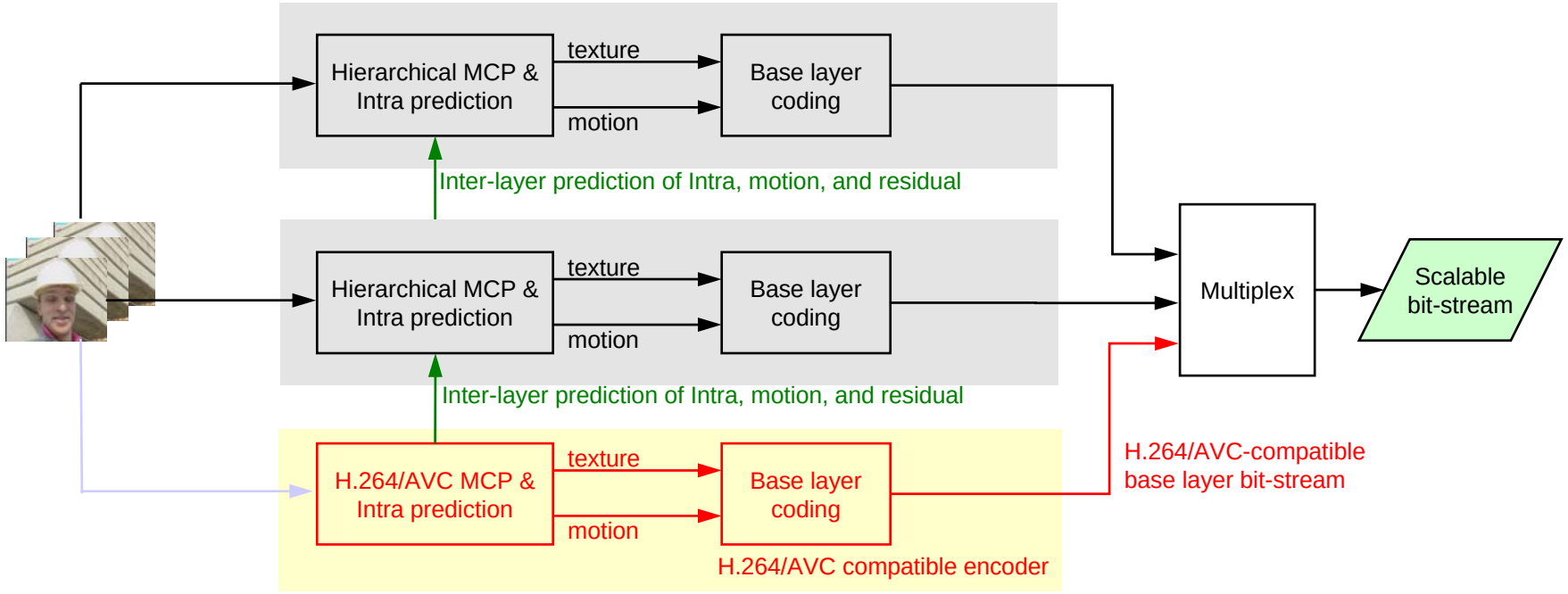
- Coarse-grain SNR scalability (CGS)
  - layered approach similar to spatial scalability, i.e., works like spatial scalability with spatial resolution factor being equal to 1
  - All enhancement layer NAL units are required to be present in the bitstream
  - designed in order to support a few selected SNR points with typically at least 50% bit-rate increase from one layer to the next
- Medium-grain SNR scalability (MGS)
  - Combination of simple CGS syntax with advanced motion-compensated prediction structure
  - MGS NAL units are discardable
  - Uses adaptive prediction from either base layer or enhancement layer

Layered Coding for SNR Scalability



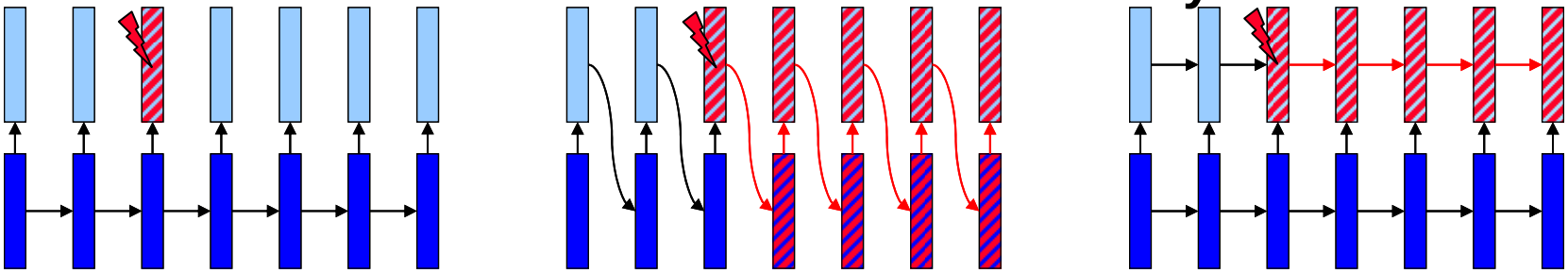
- **Same resolution of pictures over all layers**
- **Intra prediction only in base layer**
- **Residual coding of the quantization error between original pictures and their lower layer reconstruction**
- **Motion vector refinement possible**
- **Can be converted to AVC without transcoding at any bit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

Coarse-grain SNR scalability (CGS)



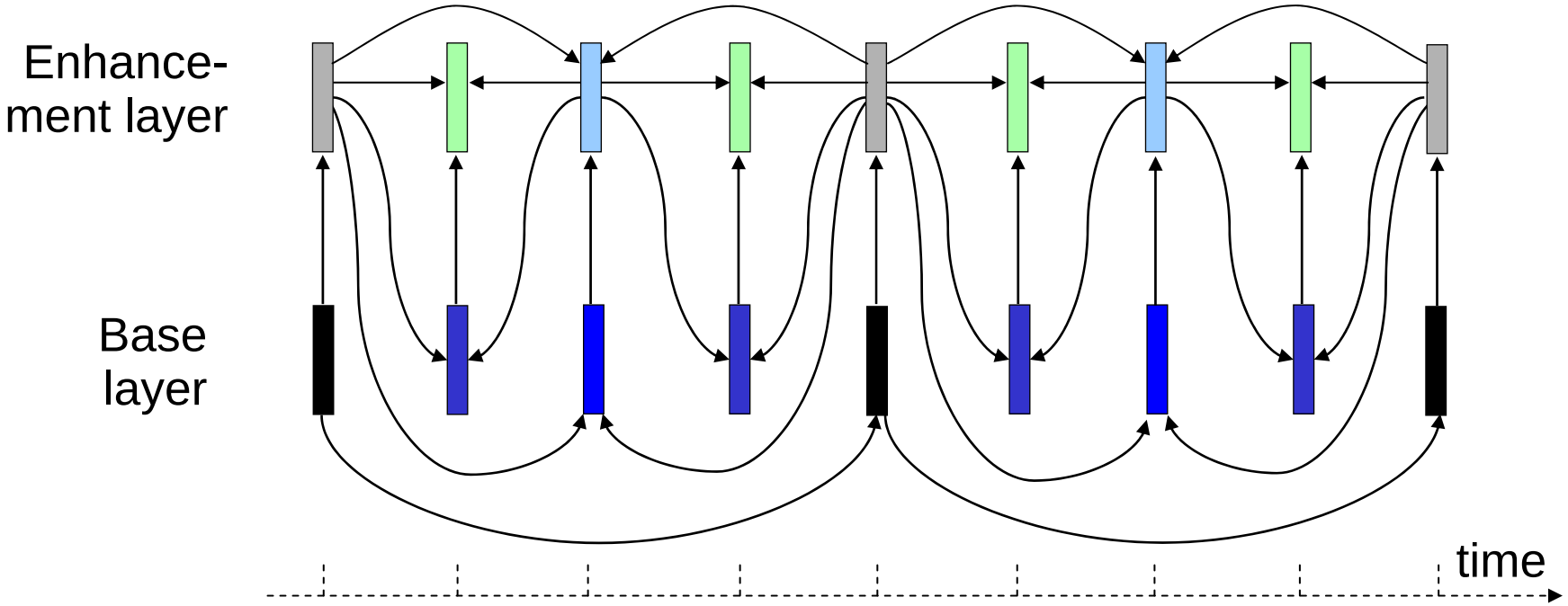
- Extreme case of (extended) spatial scalability
  - resolution ratio is equal to 1, no cropping
  - no upsampling (motion, texture) required
  - singe-loop decoding !
- Special case: Requantization
  - identical motion data for all layers, always use residual prediction

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

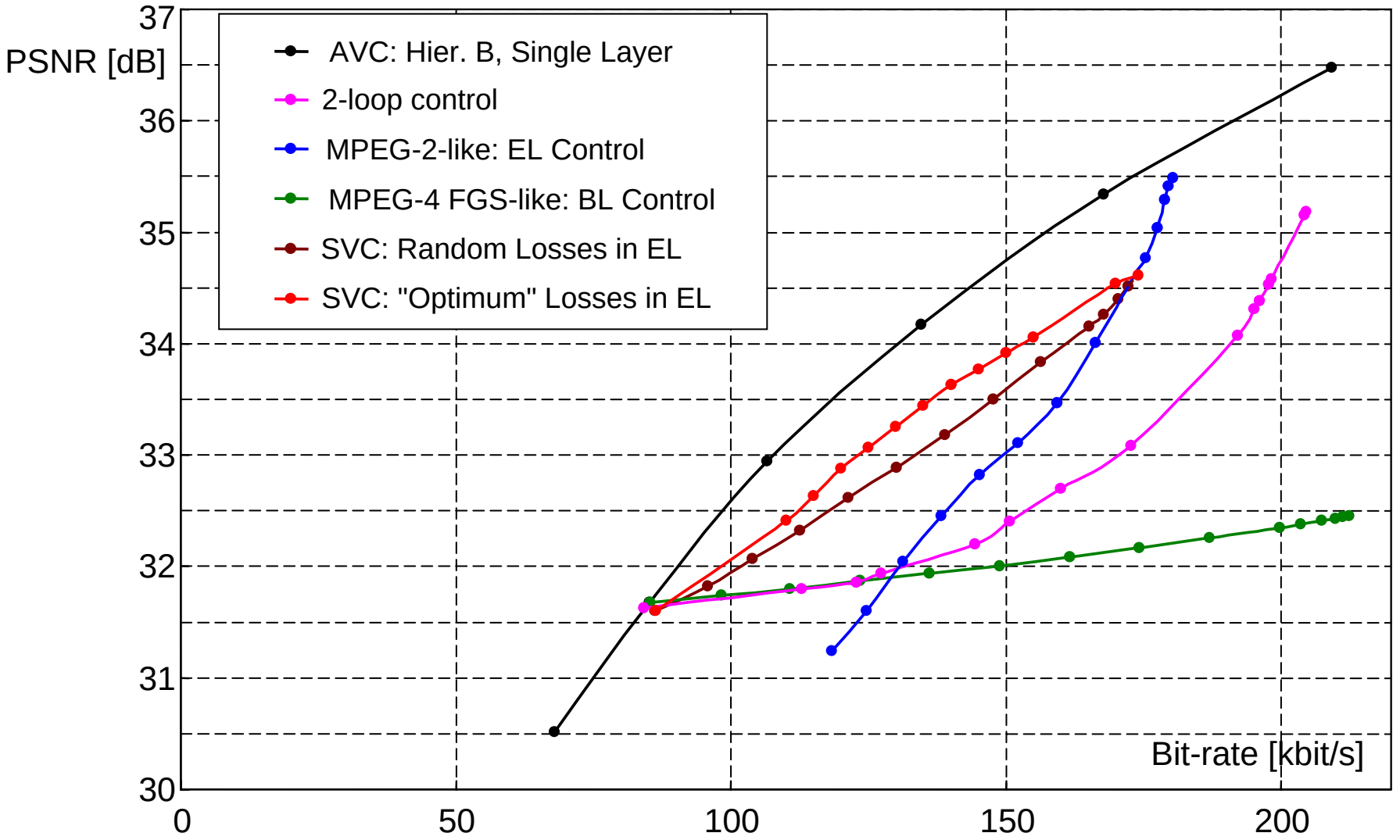


- **BL only control**
  - MPEG-4 FGS
  - No drift
  - Low complexity
  - Efficient BL and inefficient EL coding
- **EL only control**
  - MPEG-2 SNR
  - Drift in both BL and EL
  - Low complexity
  - Inefficient BL and efficient EL coding (when latter complete)
- **2-loop control**
  - MPEG-4 spatial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

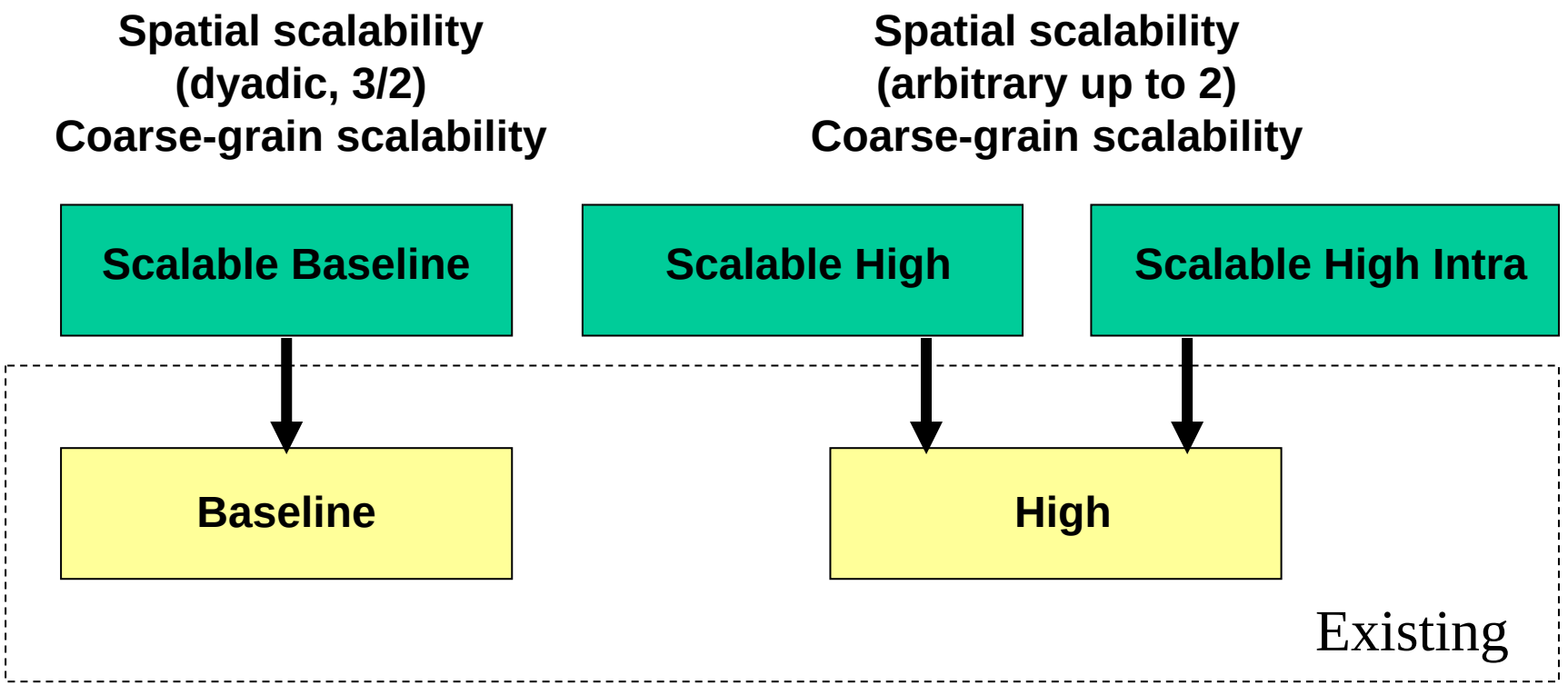
- **Adaptive BL/EL only encoder control for hierarchical prediction structures**
- **Normal pictures: use EL reconstruction as reference for MCP**
- **Closed-loop pictures: use BL reconstruction as reference for MCP at encoder side**
- **Single motion compensation loop decoding**



SNR Scalability Results: New SVC



3. Profiles, Levels, and Systems Support



**Level sizes identical for all profiles,
same as previously defined inAVC**

- Scalable Baseline
  - 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
  - AVC base layer: High profile
  - SVC enhancement layer: I, P, and B slices, CAVCL and CABAC, all spatial factors
- Scalable High Intra
  - AVC base layer: High Intra profile
  - SVC enhancement layer: I slices, CAVCL and CABAC, all spatial factors

SVC Levels

- Levels for all scalable profiles are identical
- Numbers are kept from 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

- 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 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)
- MPEG-2 Systems:
  - 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

4. Conclusions

The new SVC standard allows better compression and more flexibility than any scalable video coding known before.

But (as usual) more progress is to be expected – most likely by further optimization of encoders and proper adaptation for various applications.

Thank you for your attention!

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 FREext), Sept. 2004; Version 4, July, 2005, Version 5, July 2007 (with Intra profiles and SVC).

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