ECE 634: Digital Video Systems
Scalable coding: 3/23/17

Professor Amy Reibman
MSEE 356
reibman@purdue.edu

http://engineering.purdue.edu/~reibman/ece634/index.html

Scalability Outline

- Introduction: Heterogeneous clients
  - Simulcast
  - Transcoding
  - Scalability
- Definition of scalability
- Four (or more) types of scalability
- Evolution of the standards

Heterogeneity

- Many heterogeneous clients
  - Different bandwidth requirements
  - Different decoding complexity and power constraints
  - Different screen sizes

- Heterogeneous networks
  - Different rates on different networks
    - Mobile phone
    - Corporate LAN
  - Dynamically varying rates
    - Congestion in the network
    - Distance to base station

Simulcast and Transcoding

- Simulcast
  - Compress video once for each client capability
  - To support a range of possible clients requires bandwidth for storage/transmission at each possible rate
- Transcoding
  - Compress video once; reduce bit-rate based on client capability
  - Simplest scenario: decode and re-encode
  - Also possible to reduce complexity by careful design; however, it almost always involves more than VLC
  - To support a range of possible clients requires transcoding to each possible rate

Scalable Video Coding

- Definition
  - Ability to recover acceptable image/video by decoding only parts of the bitstream
- Ideal goal is an *embedded bitstream*
  - Truncate at any arbitrary rate

- Practical video coder
  - Layered coder: base layer provides basic quality, successive layers refine the quality incrementally
  - Fine granularity (FGS)
- To be useful, a scalable solution needs to be more efficient than Simulcast or Transcoding

Functionality provided by Scalability

- Graceful degradation if the less important parts of the bitstream are not decoded (lost, discarded)
- Bit-rate adaptation to match the channel throughput
- Format adaptation for backwards compatible extensions
- Power adaptation for a trade-off between decoding time and quality

- Avoiding loss of important units due to congestion
- Overall error robustness

Considerations for scalability

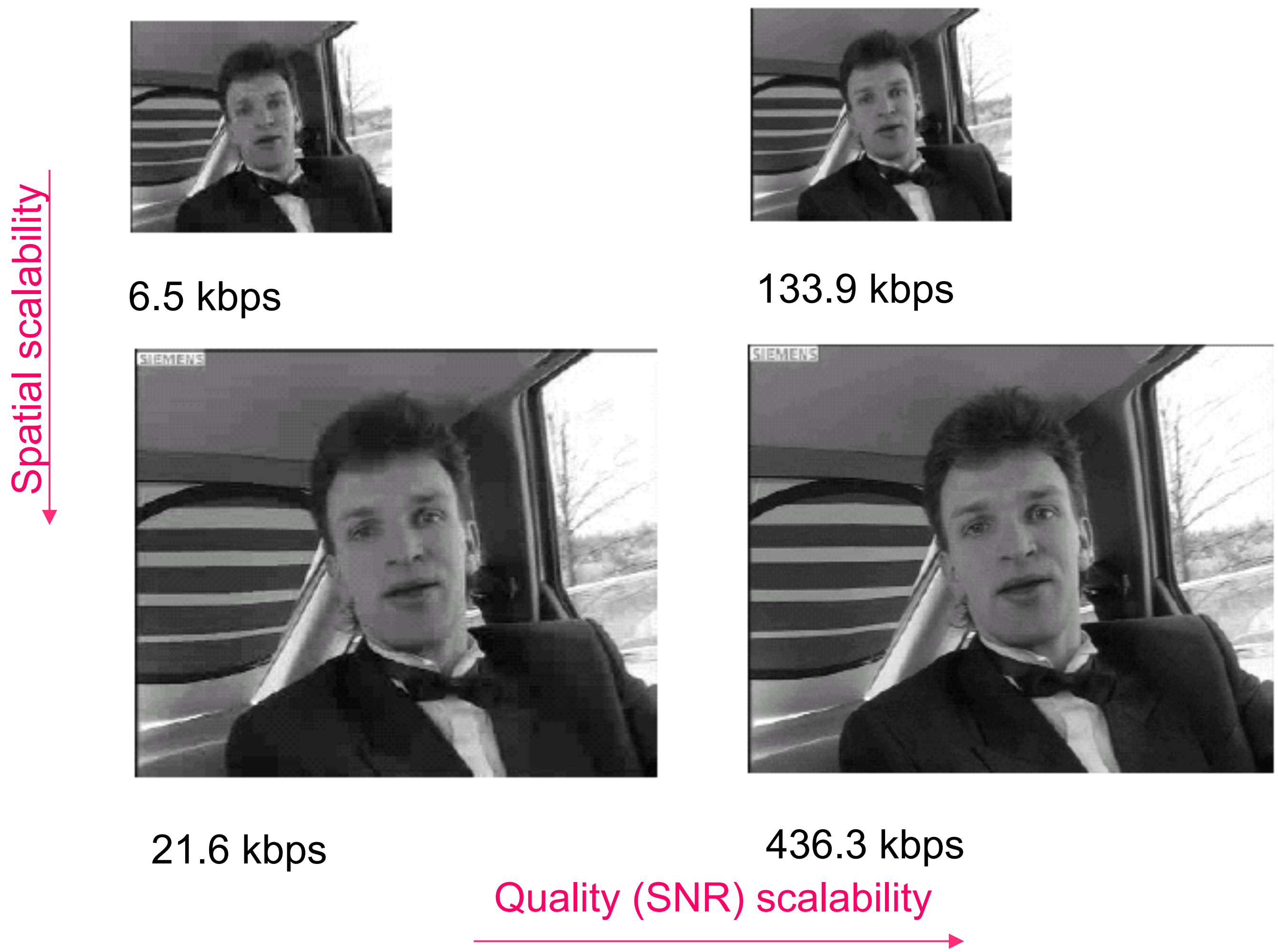
- Compression efficiency
- Decoder complexity
- Resilience to losses
- Flexible partitioning for rate adaptation
  - Range of rate partitioning (ratio of base rate to total rate)
  - Number of partitions (finely granular, or a few discrete levels)
- Compatibility with standards
- Ease of prioritization

- Prediction structure and method of partitioning controls most of these!

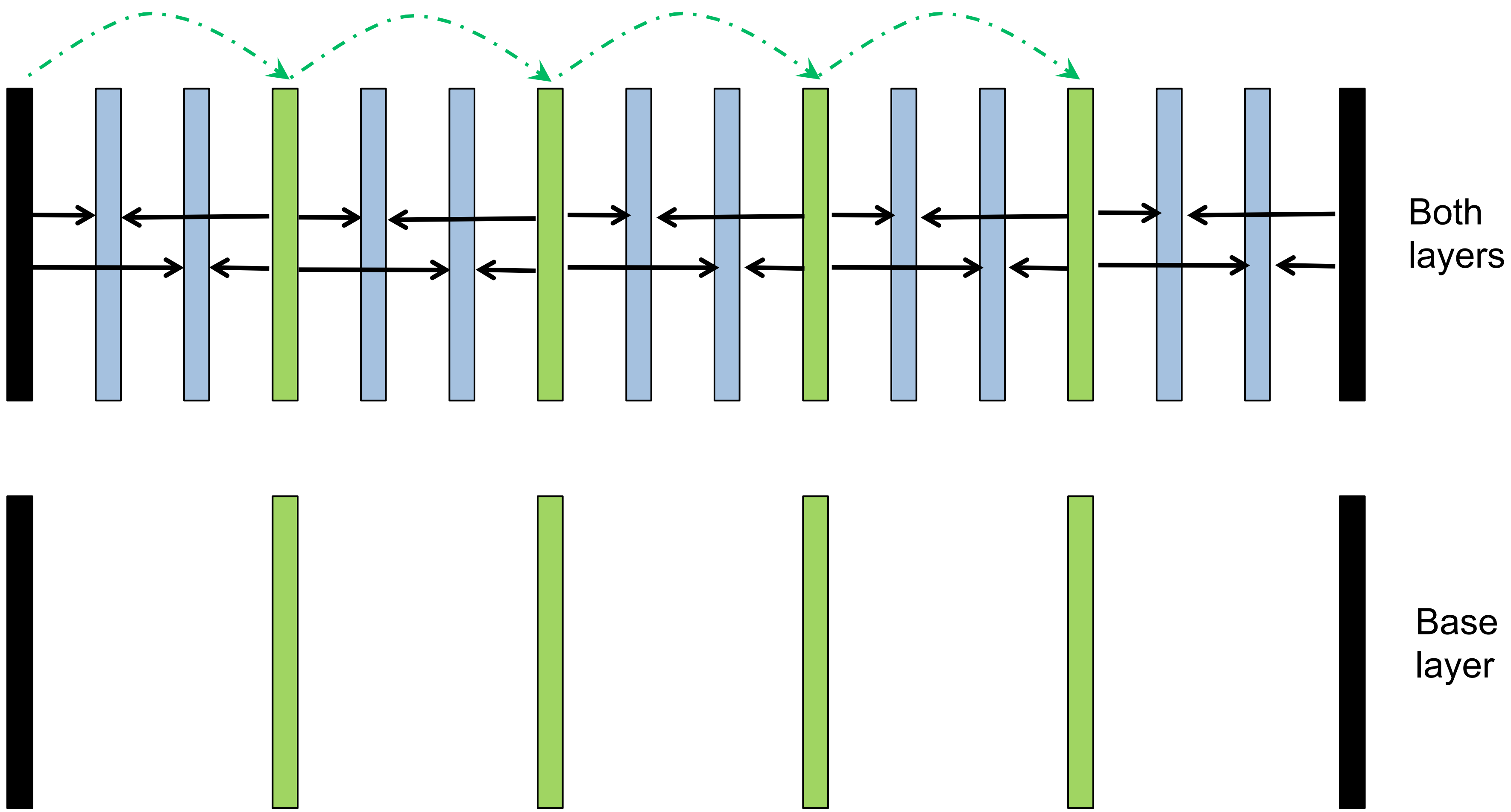
Scalability partitioning methods

- Temporal scalability (frame rate)
- Spatial scalability (picture size)
- Fidelity/SNR/Amplitude/Quality scalability
- Frequency scalability (transform coefficients)
- Object-based or ROI scalability (content)

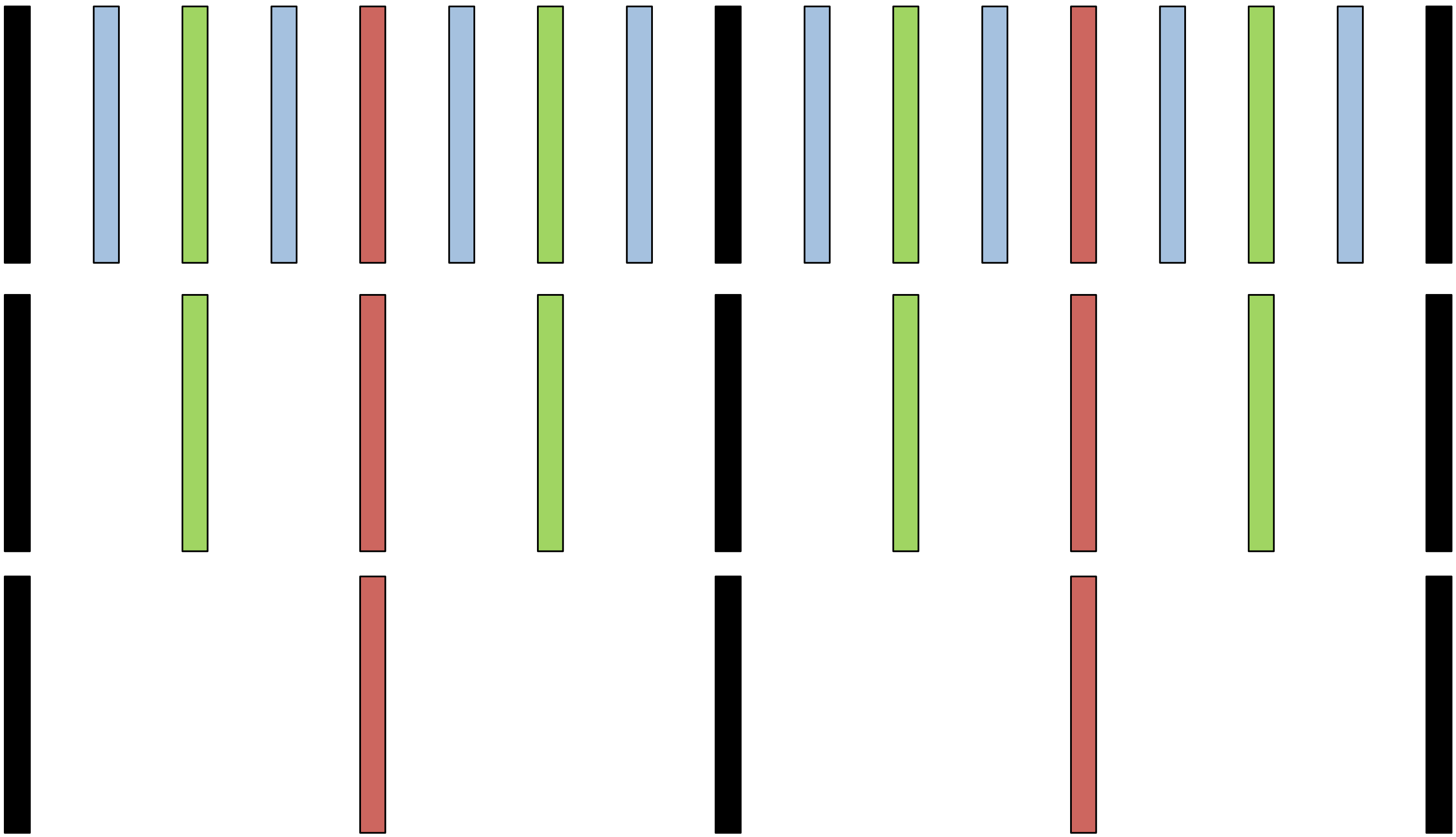
Illustration of Scalable Coding



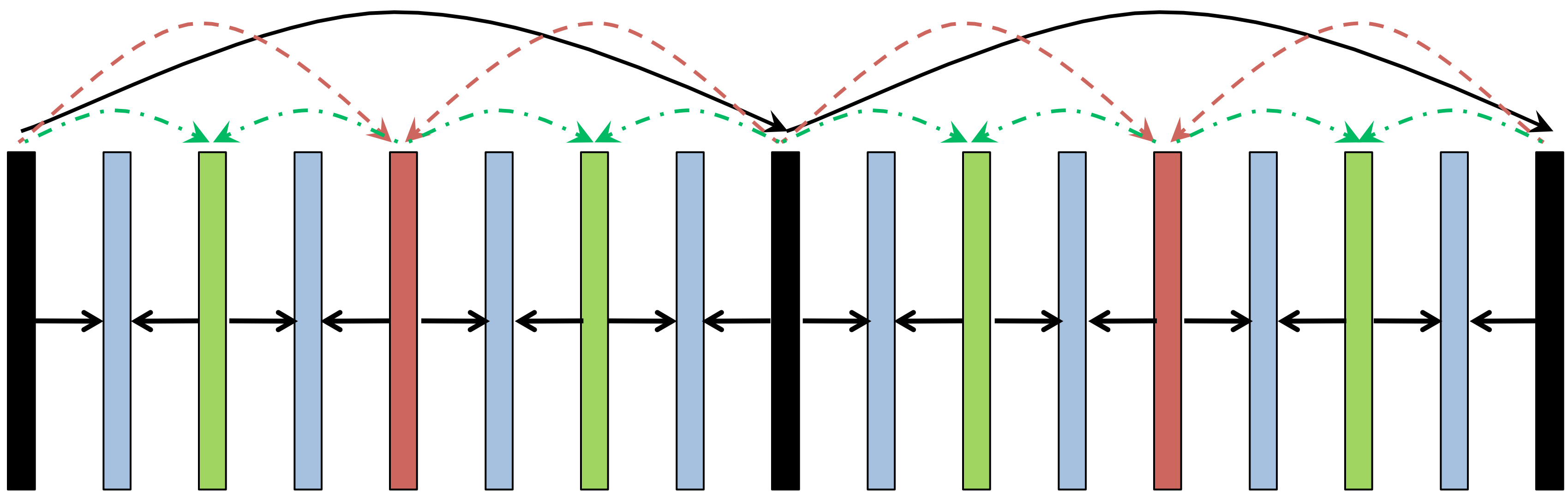
MPEG-1,2,4, H.263
Temporal Scalability



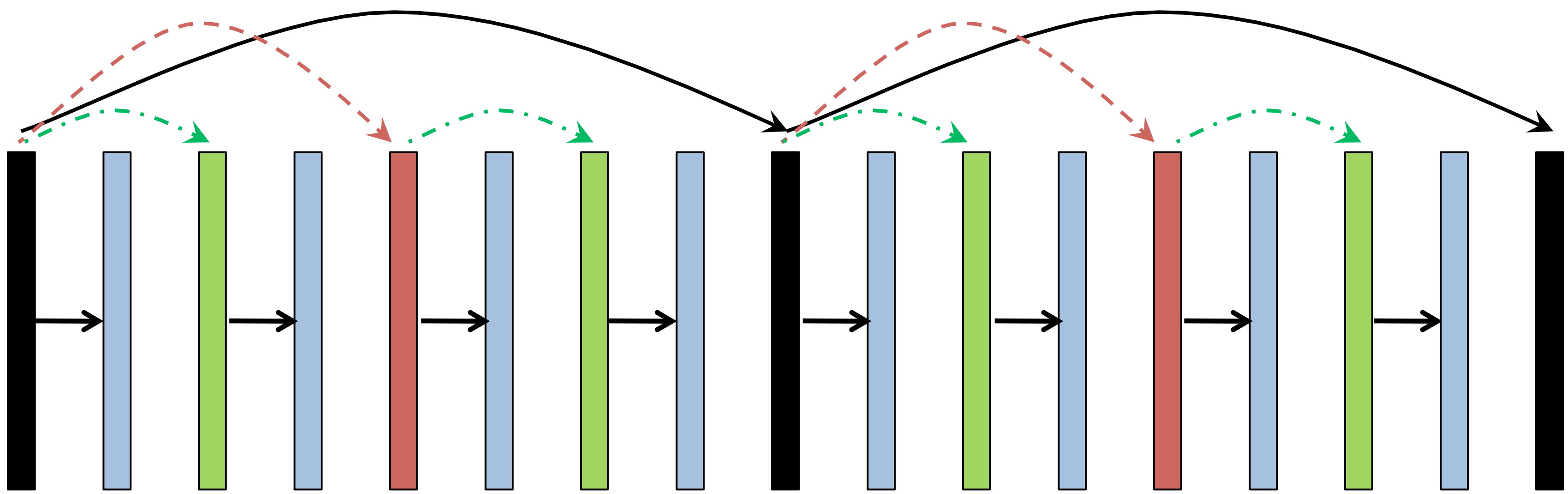
H.264: Temporal Scalability with Hierarchical prediction



Temporal Scalability with Hierarchical B pictures



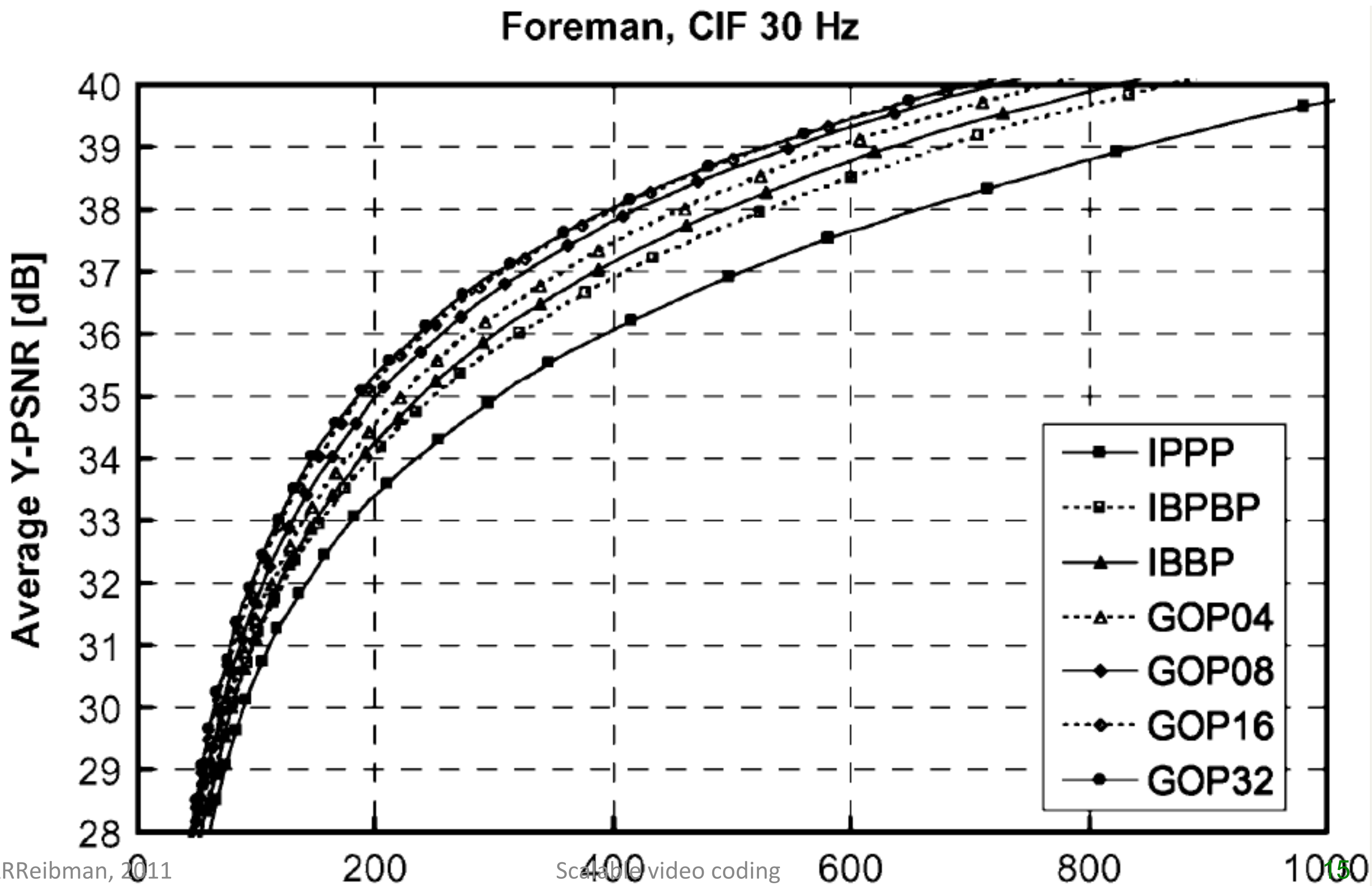
Temporal Scalability with Hierarchical prediction and Zero delay



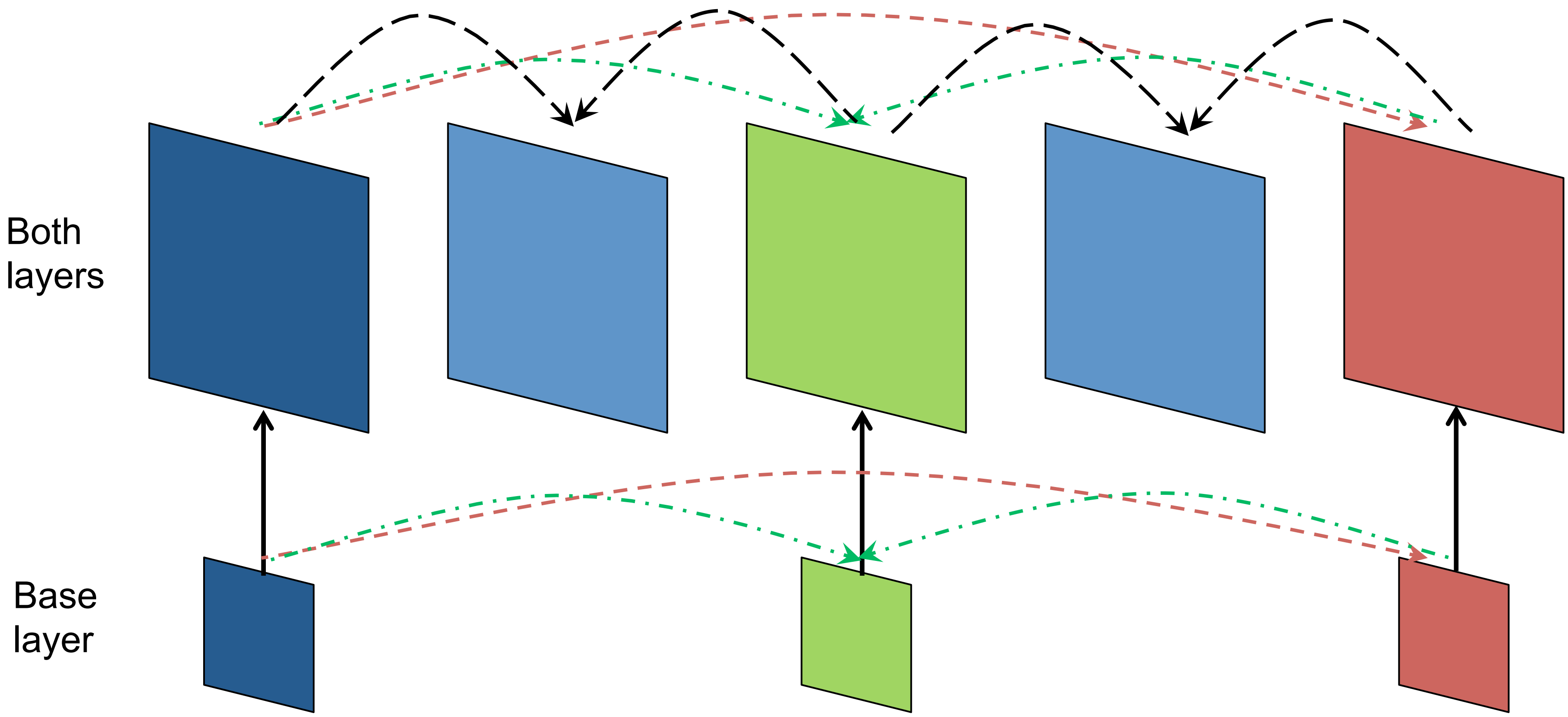
Comments about Temporal Scalability

- MPEG-1, MPEG-2, MPEG-4, and H.263+ all had capability for Temporal scalability through B-frames
  - These all require added delay at encoder/decoder
- H.264 added flexible temporal prediction, enabling more flexible temporal scalability
  - This can be implemented with or without added delay

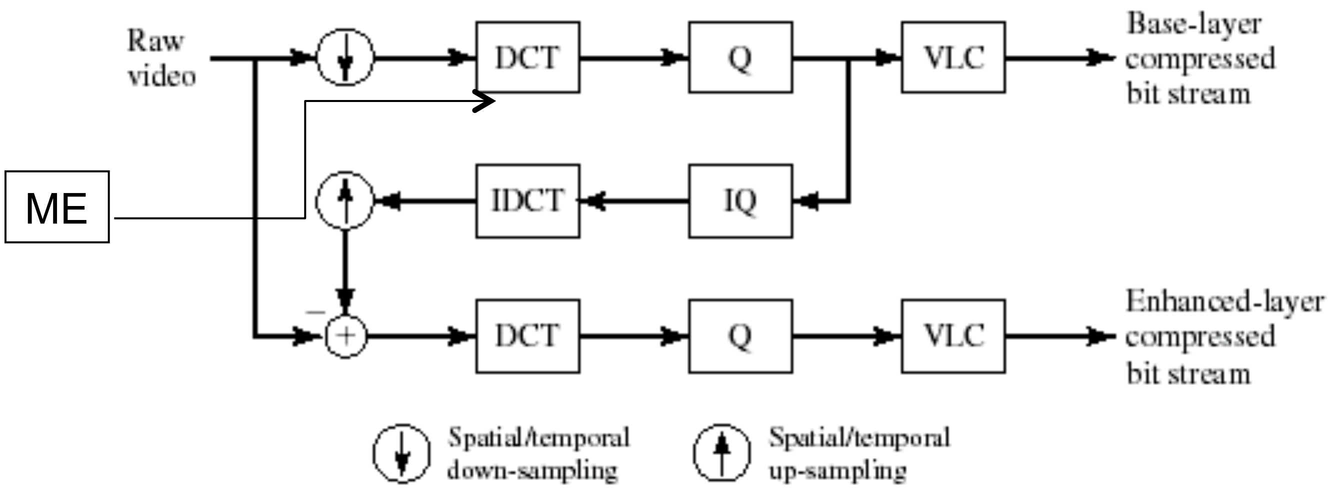
Efficiency H.264 Temporal Scalability



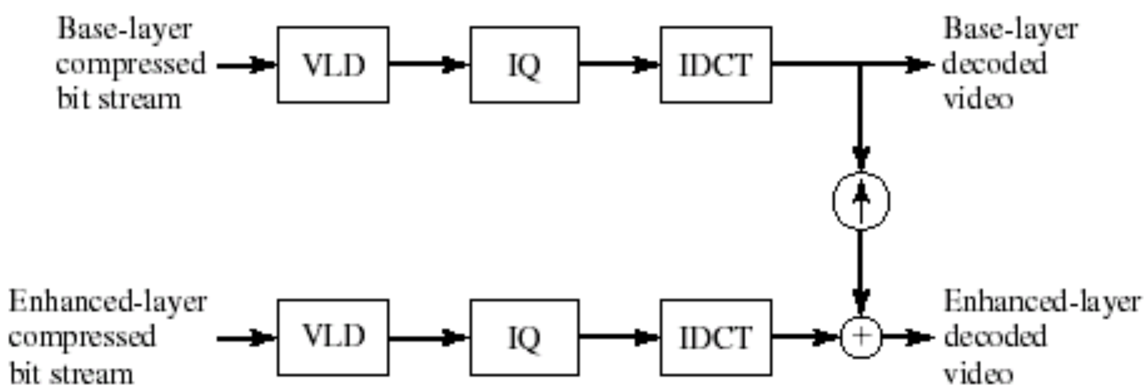
Spatial Scalability



Spatial Scalability Through Down/Up Sampling



(a)



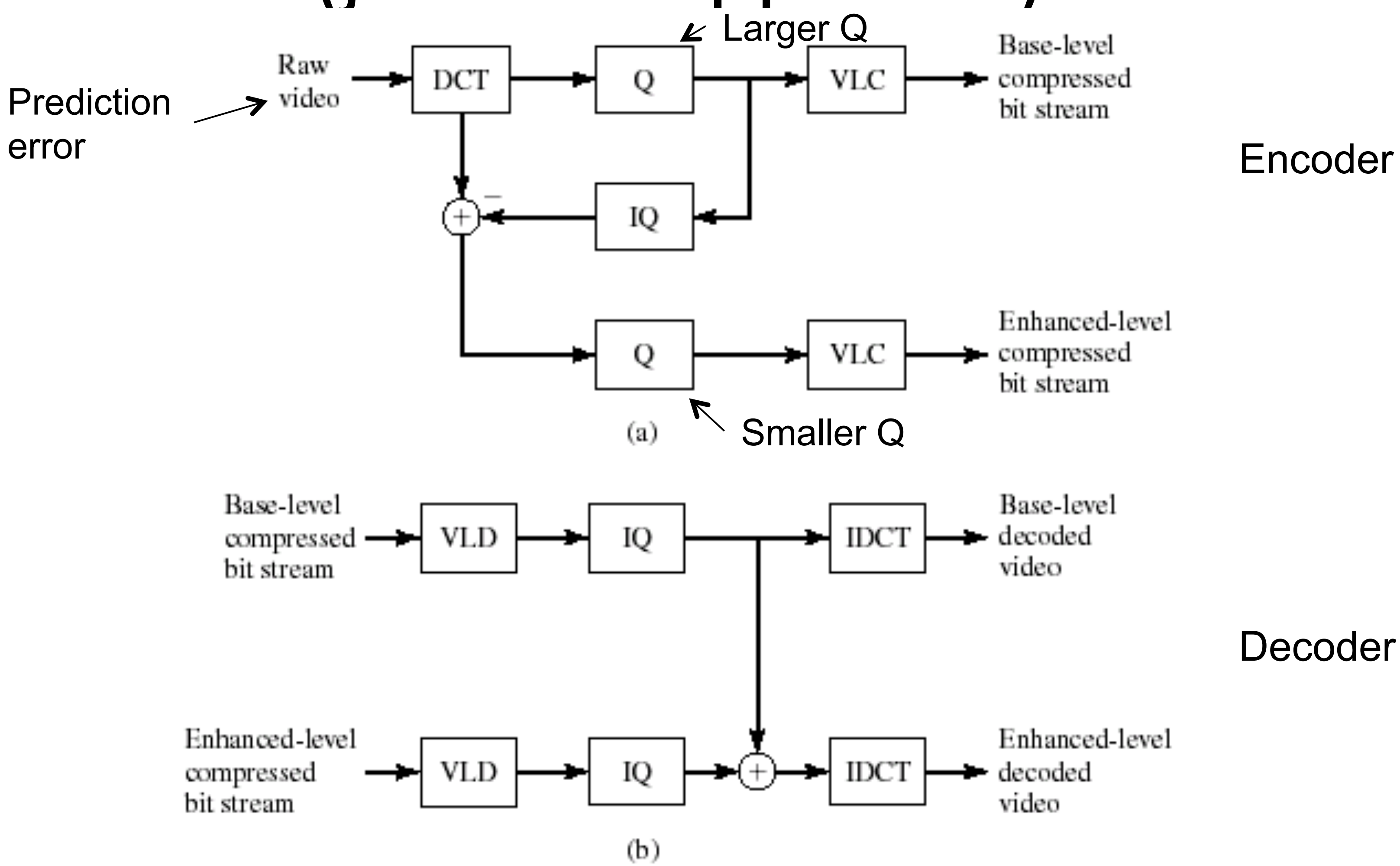
(b)

SNR Scalability

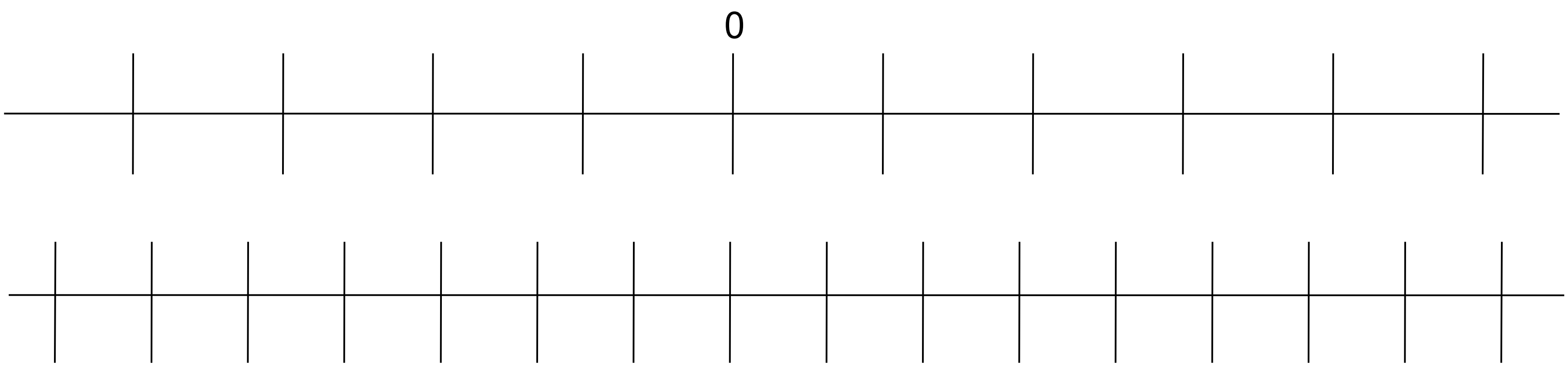
- SNR (or quality, or fidelity, or amplitude) scalability is just spatial scalability where each layer has the same spatial resolution
- Quality in each layer differs because of the quantization level

- Only the base layer can do intra-coding
- Enhancement layer(s) code the residual (between original and lower layer)

Quality (SNR) Scalability By Multistage Stage Quantization (just one approach)

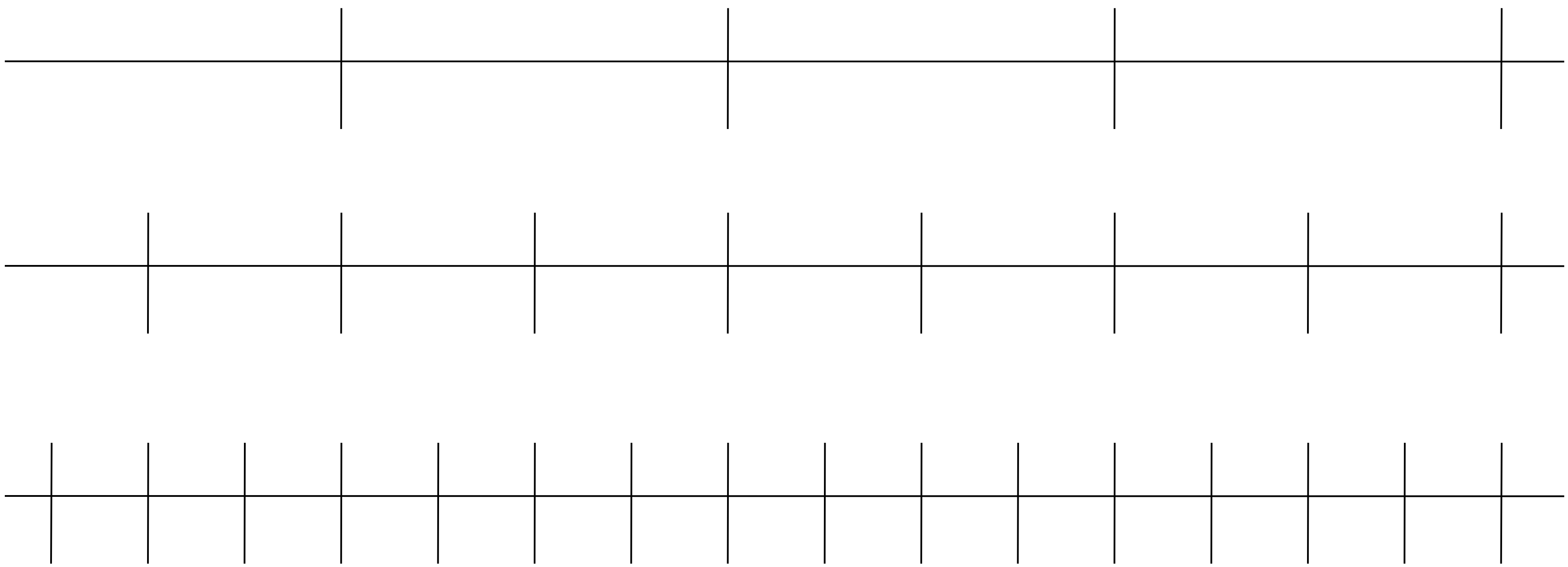


Multi-Stage Quantization

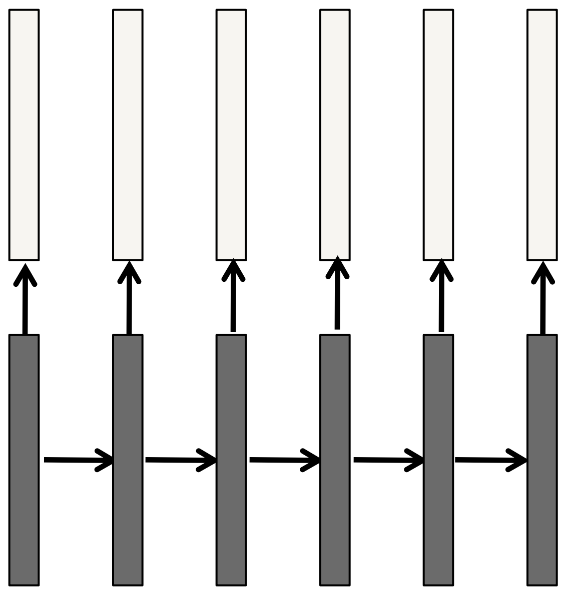


Bitplane coding

- Special case of multistage quantization, where successive step sizes differ by a factor of 2

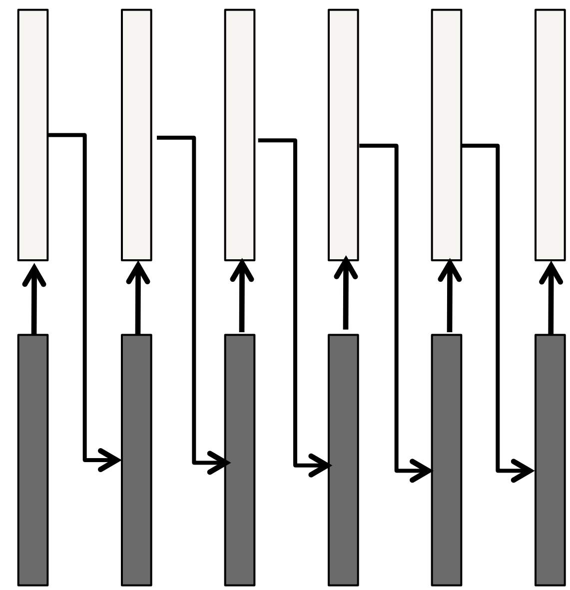


Prediction structures for scalability (1 and 2)



Enhancement layer is predicted only from same frame in base layer

MPEG-2 Spatial Scalability (1)
MPEG-4 FGS
VERY INEFFICIENT!!
No drift in base layer



Enhancement layer is used to predict base layer

MPEG-2 SNR scalability
Errors propagate into base layer
Efficient enhancement layer

Prediction strategies

- Predict from the base layer only:
  - Can be implemented with bit plane coding (MPEG4 FGS)
  - No mismatch at decoder
  - Low prediction accuracy if the base layer use large Q

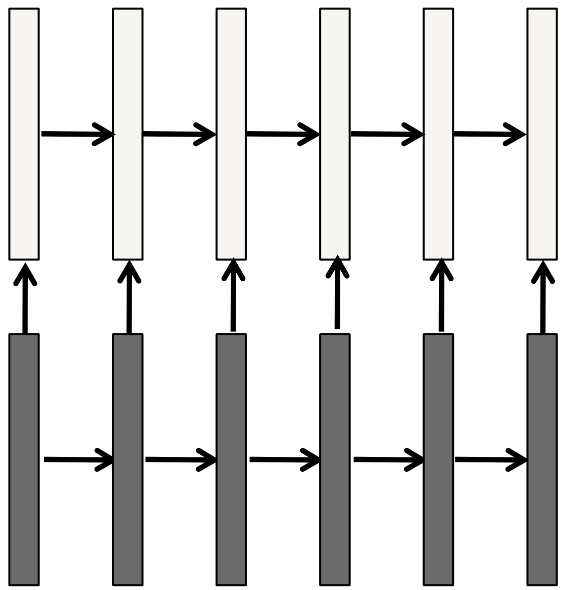
- Predict from the current layer:
  - Mismatch at decoder receiving only lower layers!
  - When the prediction requires unavailable information, this is called "drift"

Allow both intra-layer and inter-layer prediction

- Inter-layer prediction
  - Predict from the lower layer (higher Q), quantize the error using lower Q (same as before)
- Intra-layer prediction
  - Predict from previous frame (or previous blocks of the current frame) of the current layer (lower Q), quantize the error using the same lower Q

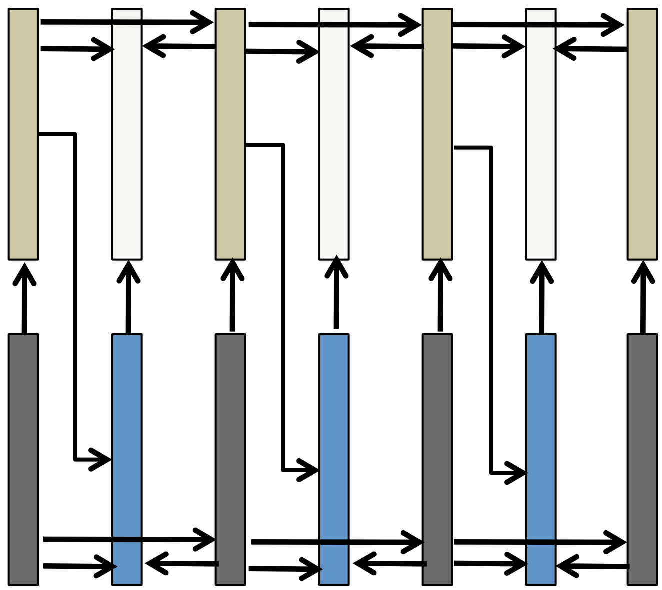
- Choose which ever is better in RD sense (H.264/SVC quality scalability)

Prediction structures for scalability (3 and 4)



2-loop control
Both base and enhancement layers use their own prediction loop

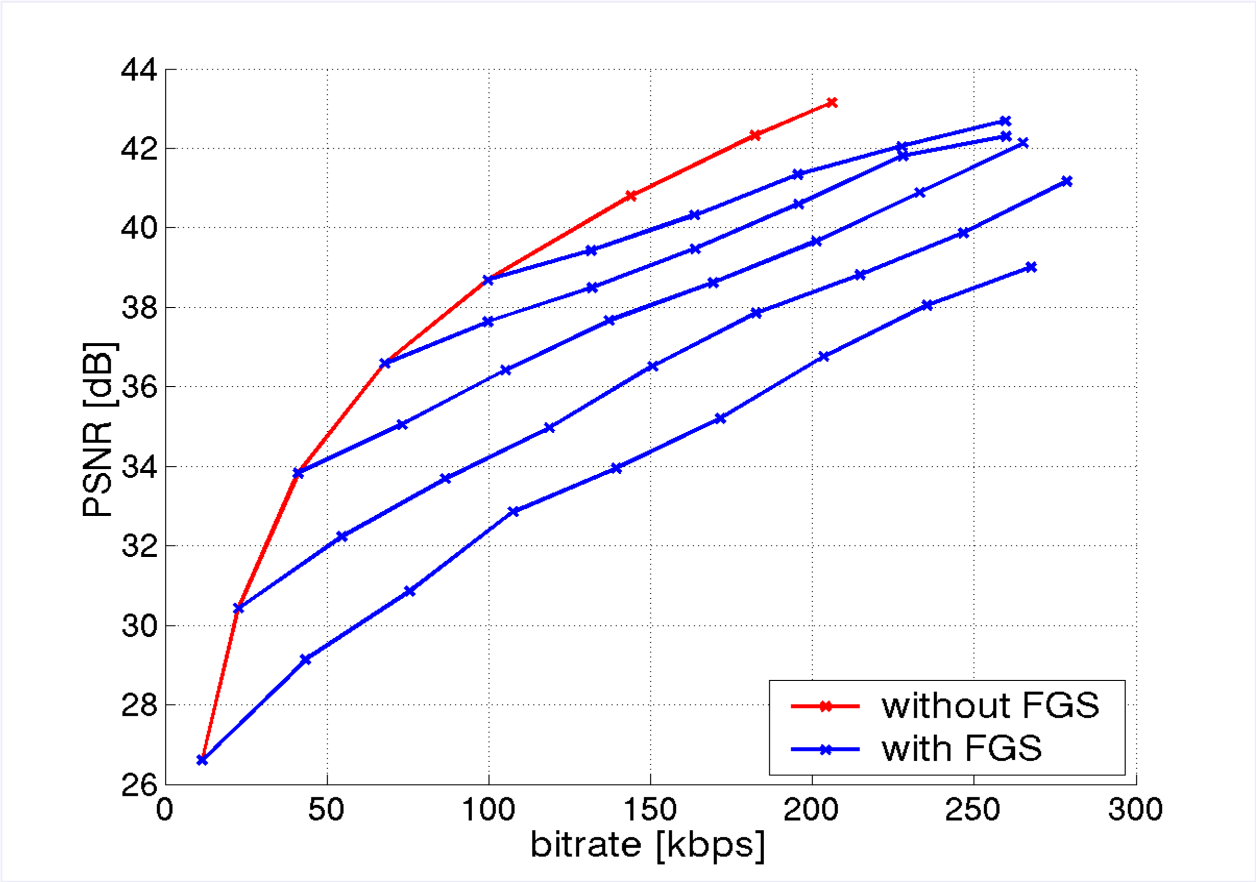
MPEG-2 Spatial Scalability (2)
No drift in base layer
High complexity; reasonably efficient



Adaptive prediction
Base: hierarchical temporal prediction
Enhance: non-key frames predict using enhancement; key frames from key frames

H.264 scalability
Single-loop decoder; low complexity

Inefficiency of predicting only from the base layer (MPEG-4 FGS)

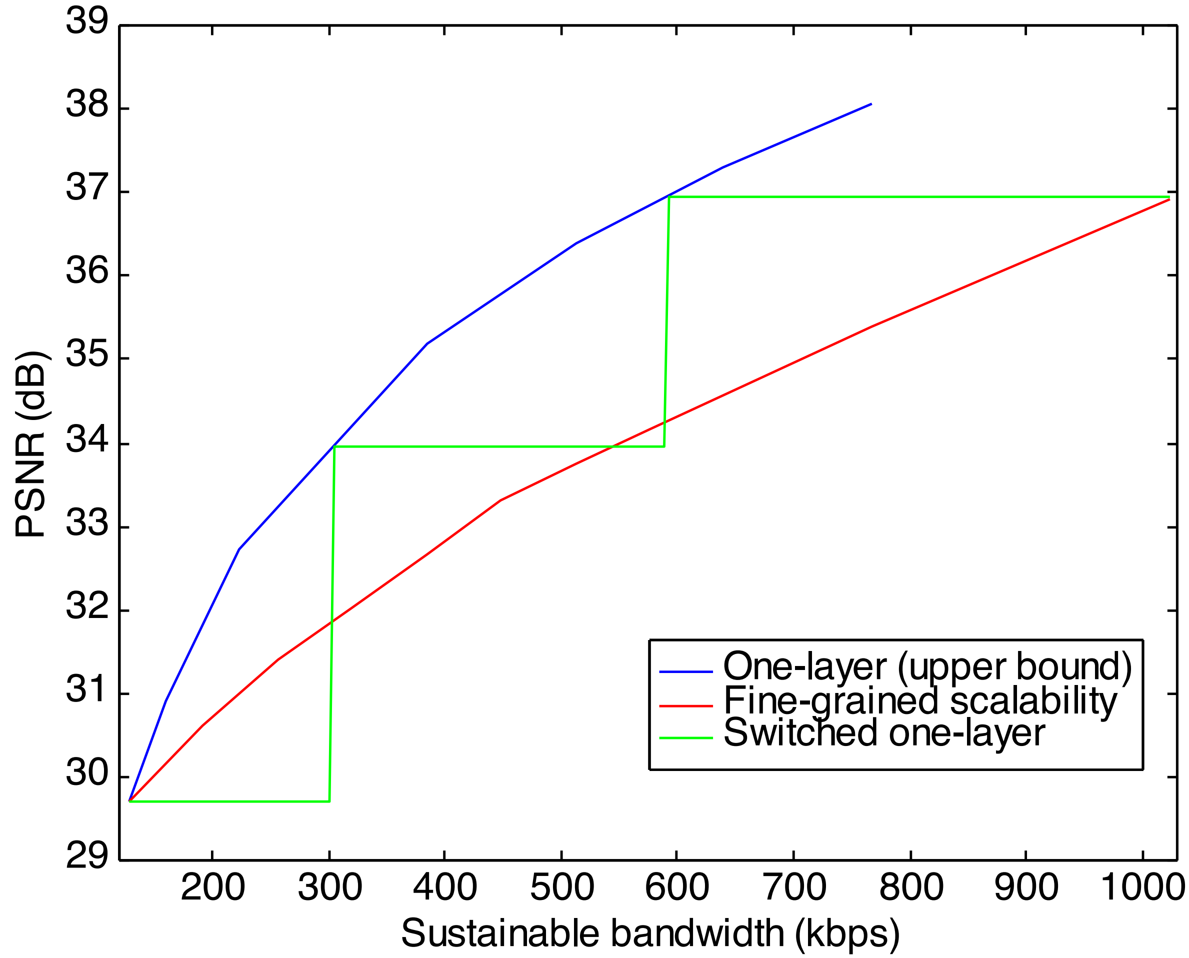


Each blue curve is obtained with MPEG4 FGS using different base-layer rate

Example: Simulcast vs. FG Scalability

- Assume minimum sustainable throughput
  - 128 kbps
- Assume known maximum possible throughput
  - 1024 kbps
- Assume equally probable rates between min and max
- Choose 3 rates for storing simulcast one-layer video
  - Switch between different one-layer videos depending on channel rate
  - Rate of all 3 videos must sum to 1024 kbps
- Compare average video quality of one-layer videos to average video quality of Fine-Grained Scalability

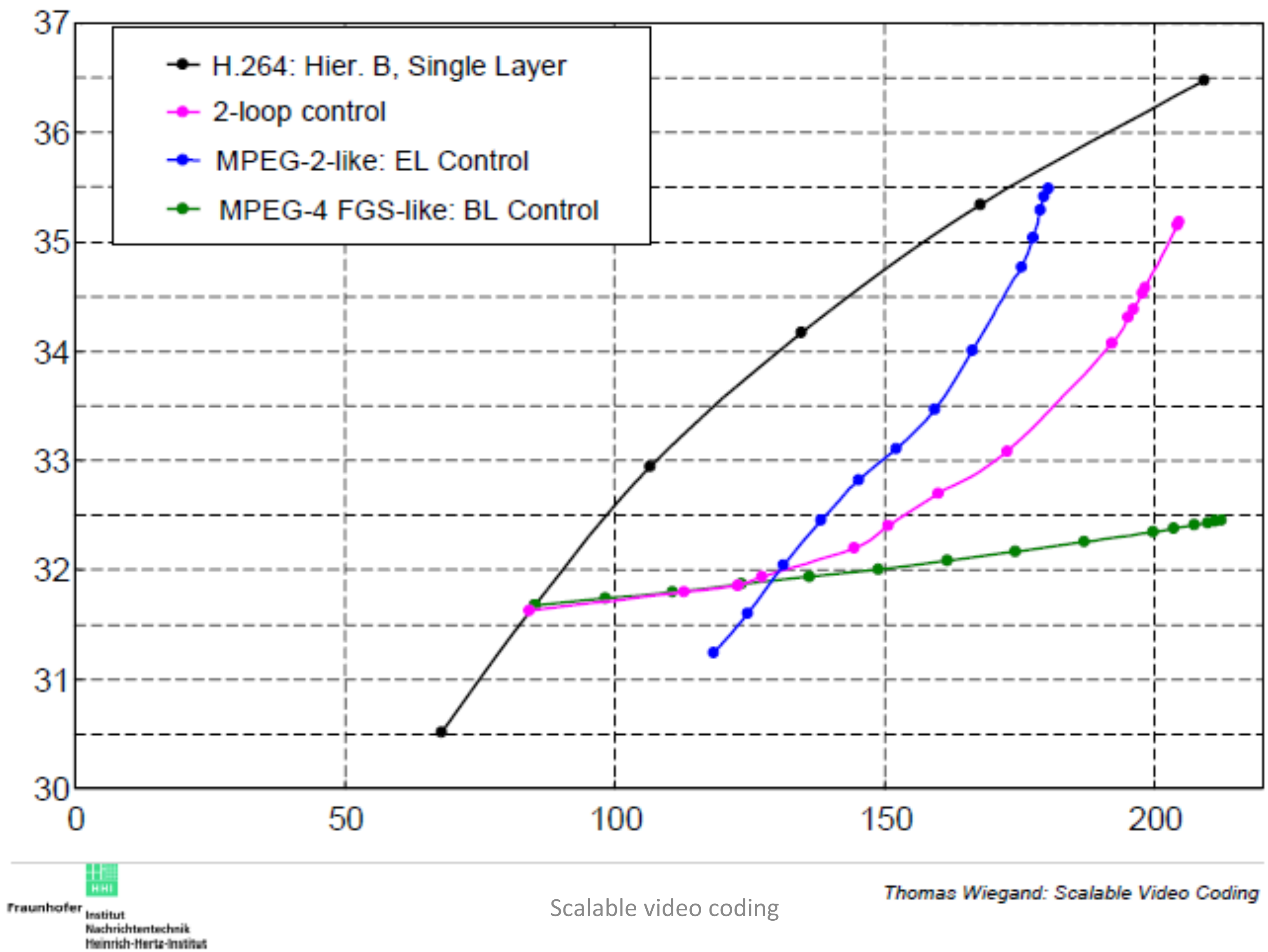
One-layer(s) vs. FG Scalability



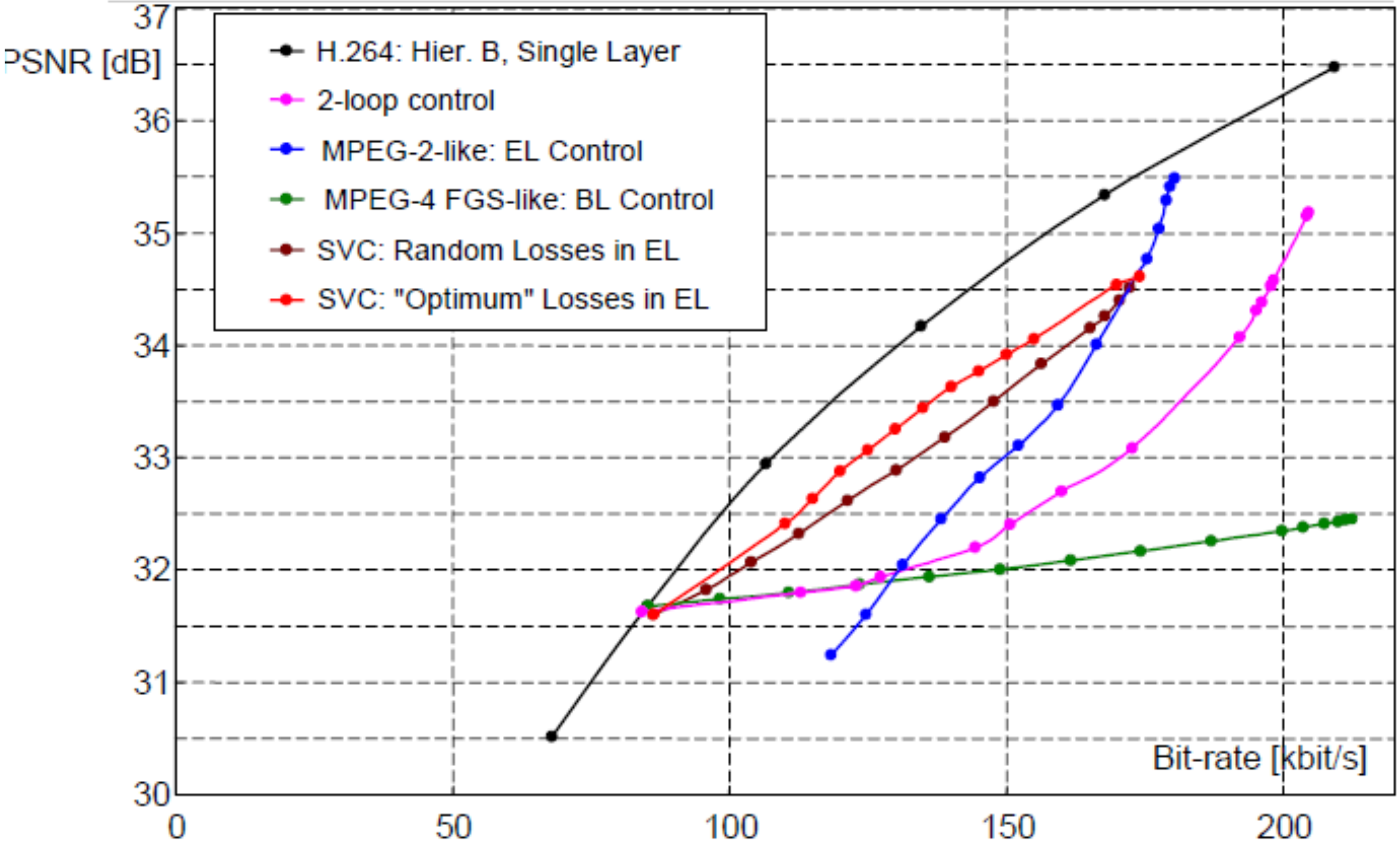
Average PSNR for switched one-layer is more than 1 dB better than average PSNR for FG Scalability

(due to prediction inefficiencies of FGS)

SNR scalability: Before H.264 SVC



SNR scalability: with H.264 SVC



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Thomas Wiegand: Scalable Video Coding

Frequency scalability (also called Data Partitioning)

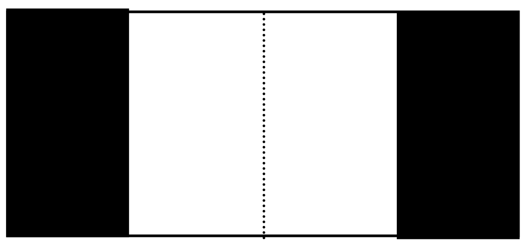
- Base layer: low frequencies of DCT
- Enhancement layer: remaining high frequencies of DCT

- Standardized in MPEG-2
- A breakpoint included in the bitstream made it very easy to partition

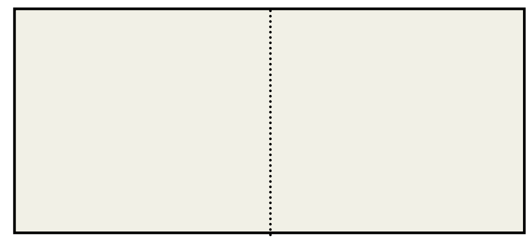
- One encoder prediction loop → missing the high frequencies means strong drift

Frequency scalability: Effect of lost information

Two blocks at encoder:



Two blocks at decoder:

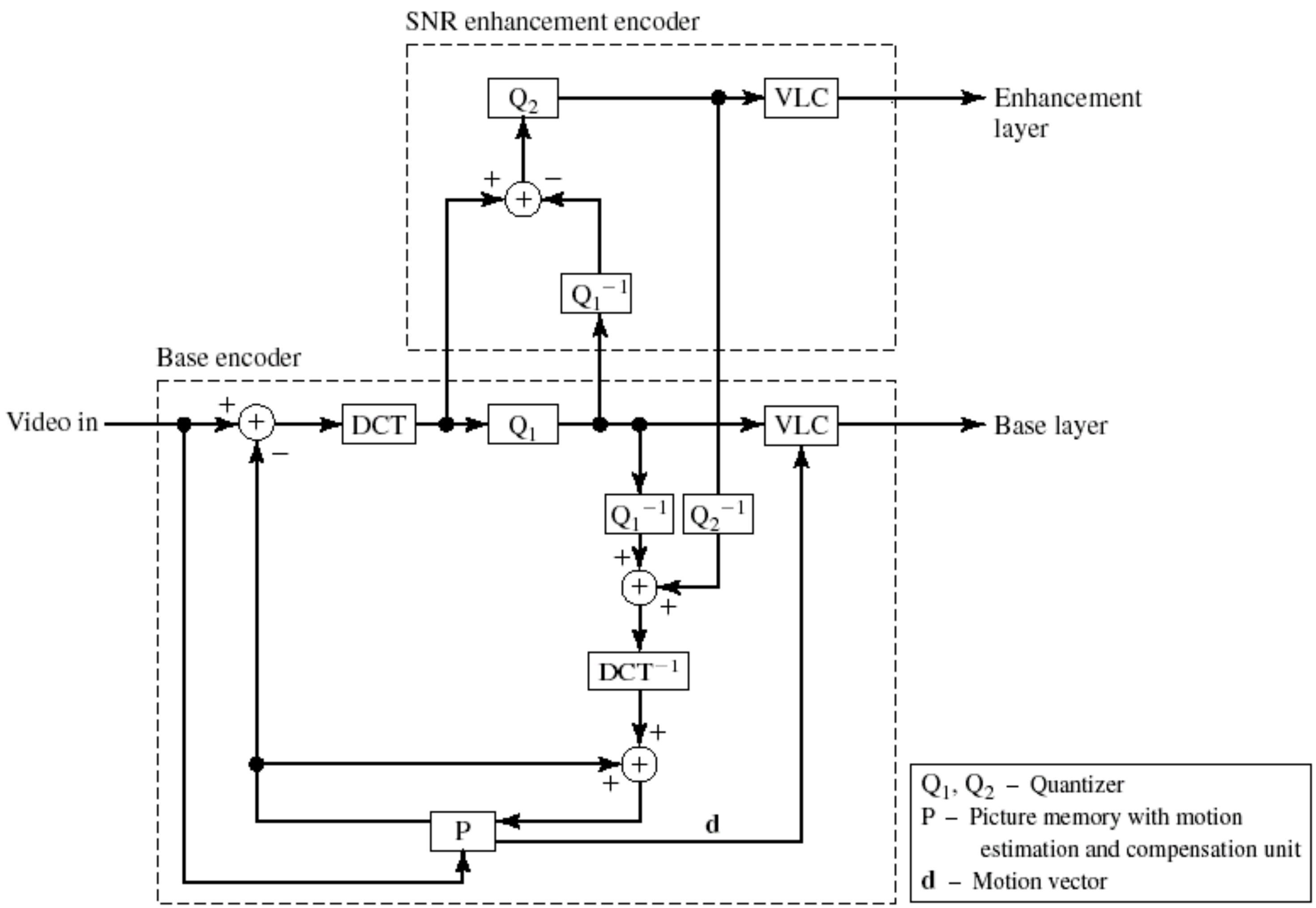


- Errors from previous frame propagate into current frame
- Motion causes error to spread, not just spatially, but *in frequency*
- Prediction method affects degree of propagation

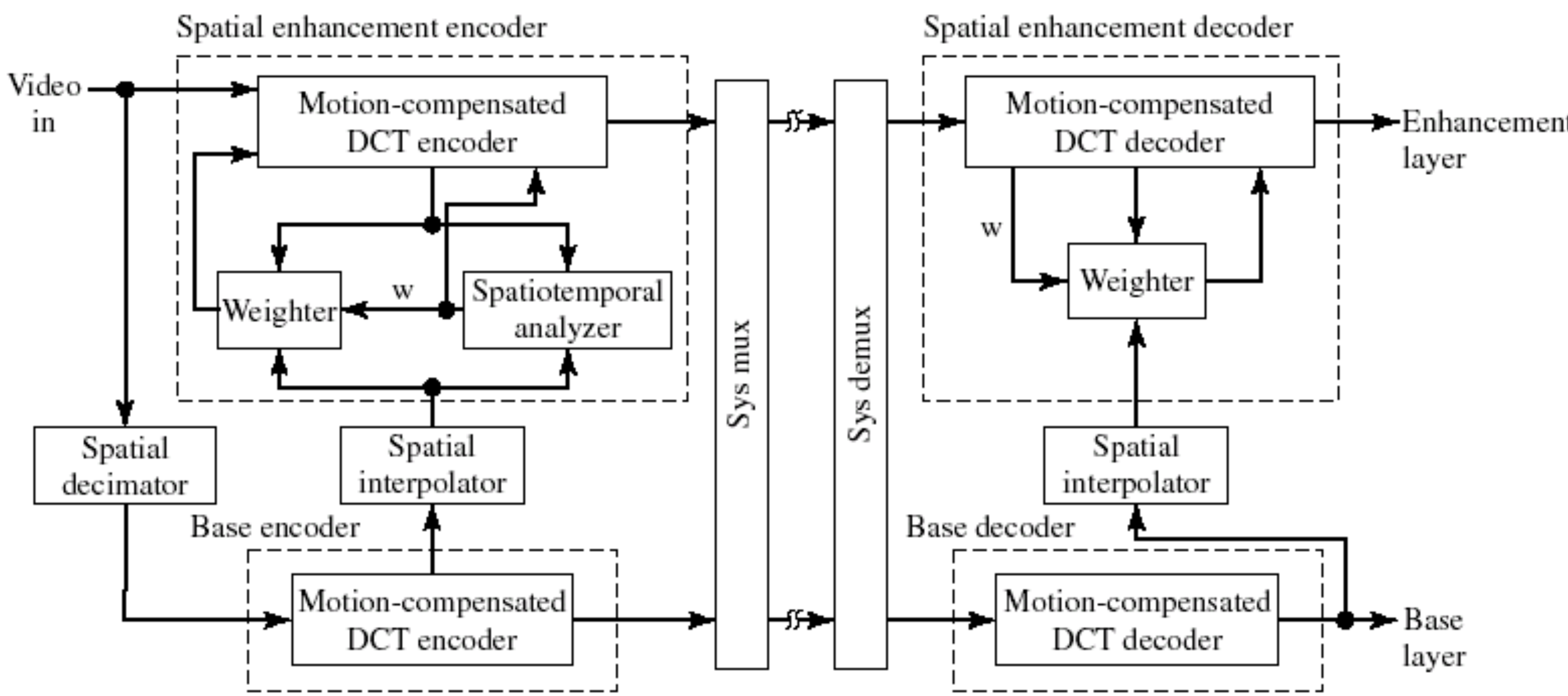
MPEG-2 Scalability:
First standard that offers scalability

- Data partition
  - All headers, MVs, first few DCT coefficients in the base layer
  - Can be implemented at the bit stream level
  - Simple
- SNR scalability
  - Base layer includes coarsely quantized DCT coefficients
  - Enhancement layer further quantizes the base layer quantization error
  - Relatively simple
- Spatial scalability
  - Complex
- Temporal scalability
  - Simple; two layers only
- Drift problem:
  - If the encoder's base layer information for a current frame depends on the enhancement layer information for a previous frame
  - Exist in the data partition and SNR scalability modes

MPEG-2 SNR Scalability Encoder



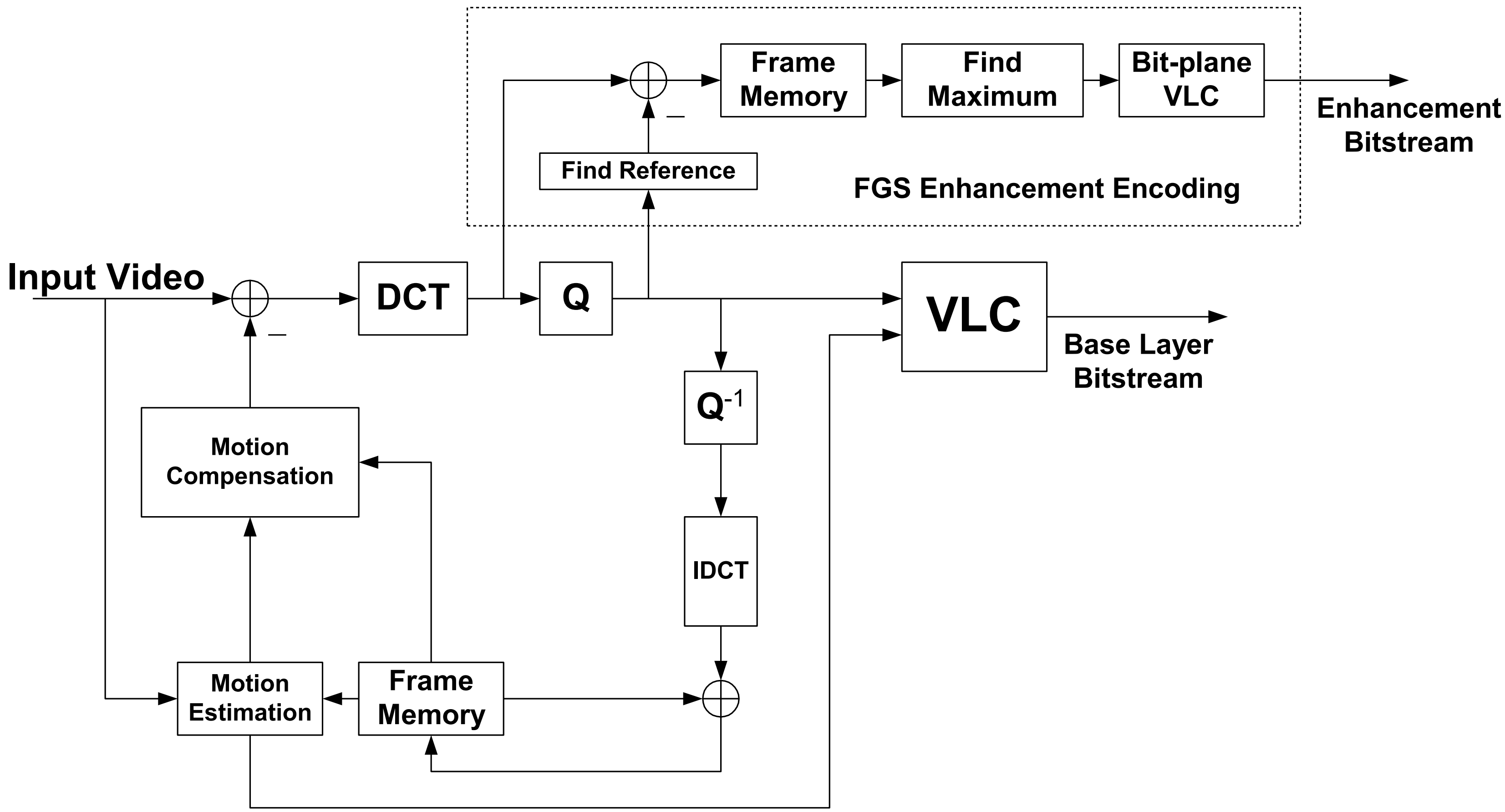
MPEG-2 Spatial Scalability Codec



Fine Granularity Scalability in MPEG-4

- MPEG-4 achieves fine granularity quality scalability through bit-plane coding
  - The DCT coefficients are represented losslessly in binary bits
  - The bit planes are coded successively, from the most significant bit to the least.
  - The bit plane within each block is coded using run-length coding.
- Temporal scalability is accomplished by combining I, B, and P-frames
- Spatial scalability is achieved by spatial down/up sampling

Fine-Grained Scalability encoder



Encode once, decode to any bandwidth

H.264 SVC (Scalable Video Coding)

- An optimized H.264 SVC encoder has an average overhead bit-rate of about 11% **
- A good trade-off between efficiency and error-propagation/drift
- Complexity is similar to single-layer H.264 decoding
- Uses only a single motion-compensation loop at the decoder

- Predicts not only residual (DCT) information, but also predict motion information and macroblock modes
- **Compared to a H.264 base profile without CABAC

Scalable HEVC

- High level syntax common to SHVC, MVHEVC, and future extensions, in Annex F of HEVC
- SHVC decoder and conformance requirements in Annex H of HEVC
- Test Model (encoder) for SHVC

- Interlayer SHVC allows any parameter changed
  - Spatial resolution, bit depth, color gamut
  - Texture resampling, motion-field resampling, color mapping

Scalable Video Coding Using Wavelet Transforms

- Wavelet-based image coding:
  - Full frame image transform (as opposed to block-based transform)
  - Bit plane coding of the transform coefficients can lead to embedded bitstreams
  - EZW → SPIHT → JPEG2000
- Wavelet-based video coding
  - Temporal filtering with and without motion compensation
    - Using MC limits the range of scalability
  - Can achieve temporal, spatial, and quality scalability simultaneously
  - Still an active research activity!?

Additional Reading

- Wang, Ostermann, Zhang "Video Processing and Communications": Sec. 11.1, 11.2, 11.3
- R. Aravind, M. R. Civanlar, and A. R. Reibman, "Packet loss resilience of MPEG-2 scalable video coding algorithms", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 5, pp. 426-435, October 1996.
- H. Schwarz, D. Marpe, T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Trans. CSVT, September 2007 http://iphome.hhi.de/wiegand/assets/pdfs/DIC_SVC_07.pdf