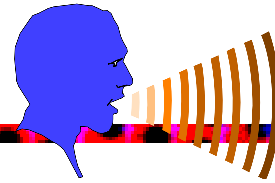
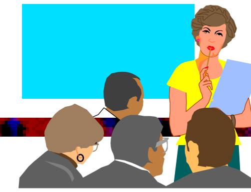


Video Processing & Communications

Scalable Video Coding

Yao Wang
Polytechnic Institute of NYU
Brooklyn, NY11201

(Modified from slides prepared by Amy Reibman)

Outline

- 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 storage/transmission at each possible rate
- Transcoding
  - Compress video once; transcode to a lower 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

Illustration of Scalable Coding

Spatial scalability

6.5 kbps

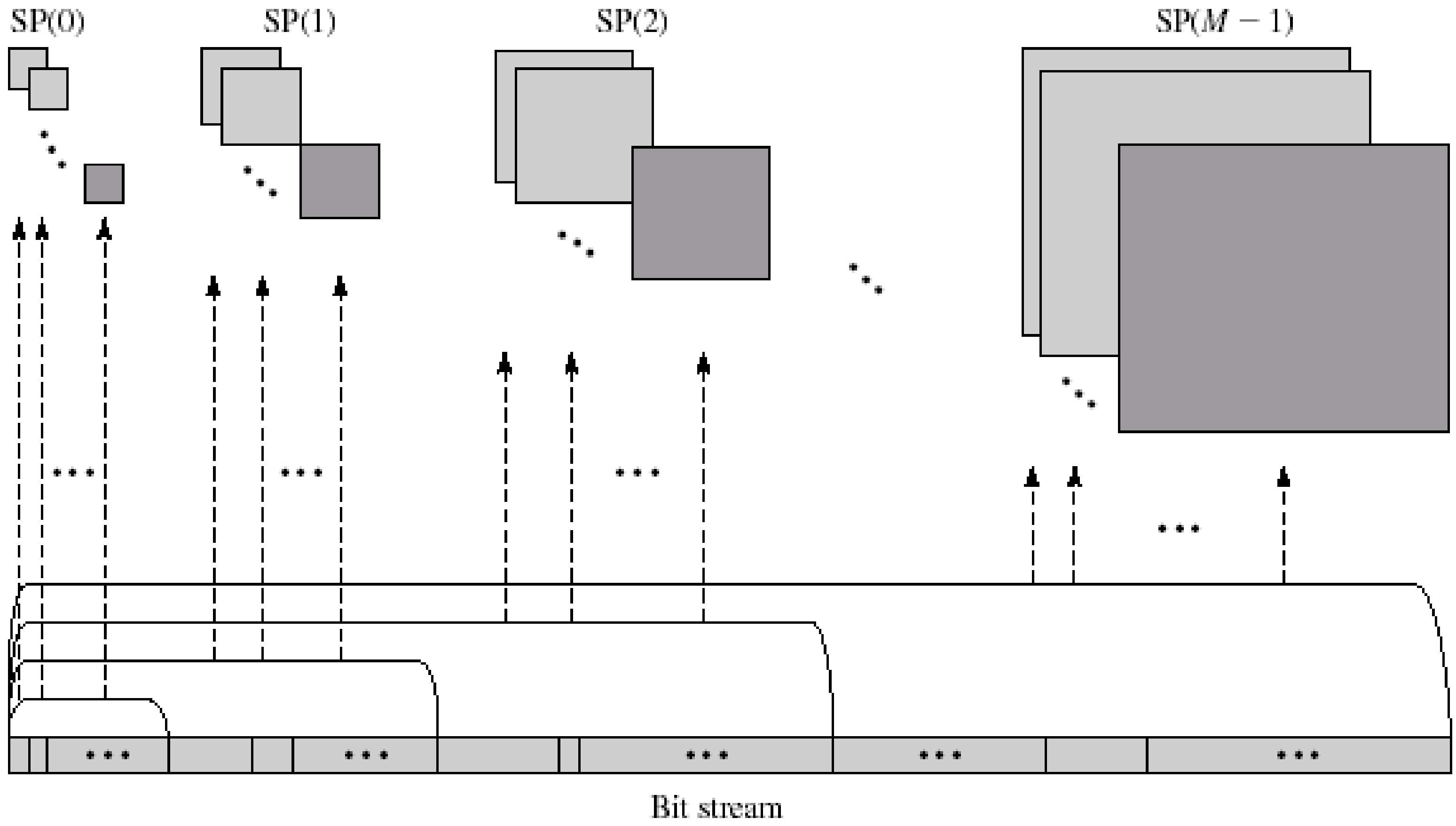
133.9 kbps

21.6 kbps

436.3 kbps

Amplitude (SNR or quality) scalability

Embedded Bit Stream



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): each layer is very thin
- 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 delivered or received or decoded (lost, discarded)
- Bit-rate adaptation at the sender or intermediate nodes to match the channel throughput
- Format adaptation for backwards compatible extensions
- Power adaptation for a trade-off between decoding time (power consumption) and quality
- Transport module can provide more protection against packet losses to lower layers (unequal error protection or UEP)
- Overall robustness to bandwidth fluctuation and packet losses

Design Considerations for Scalability

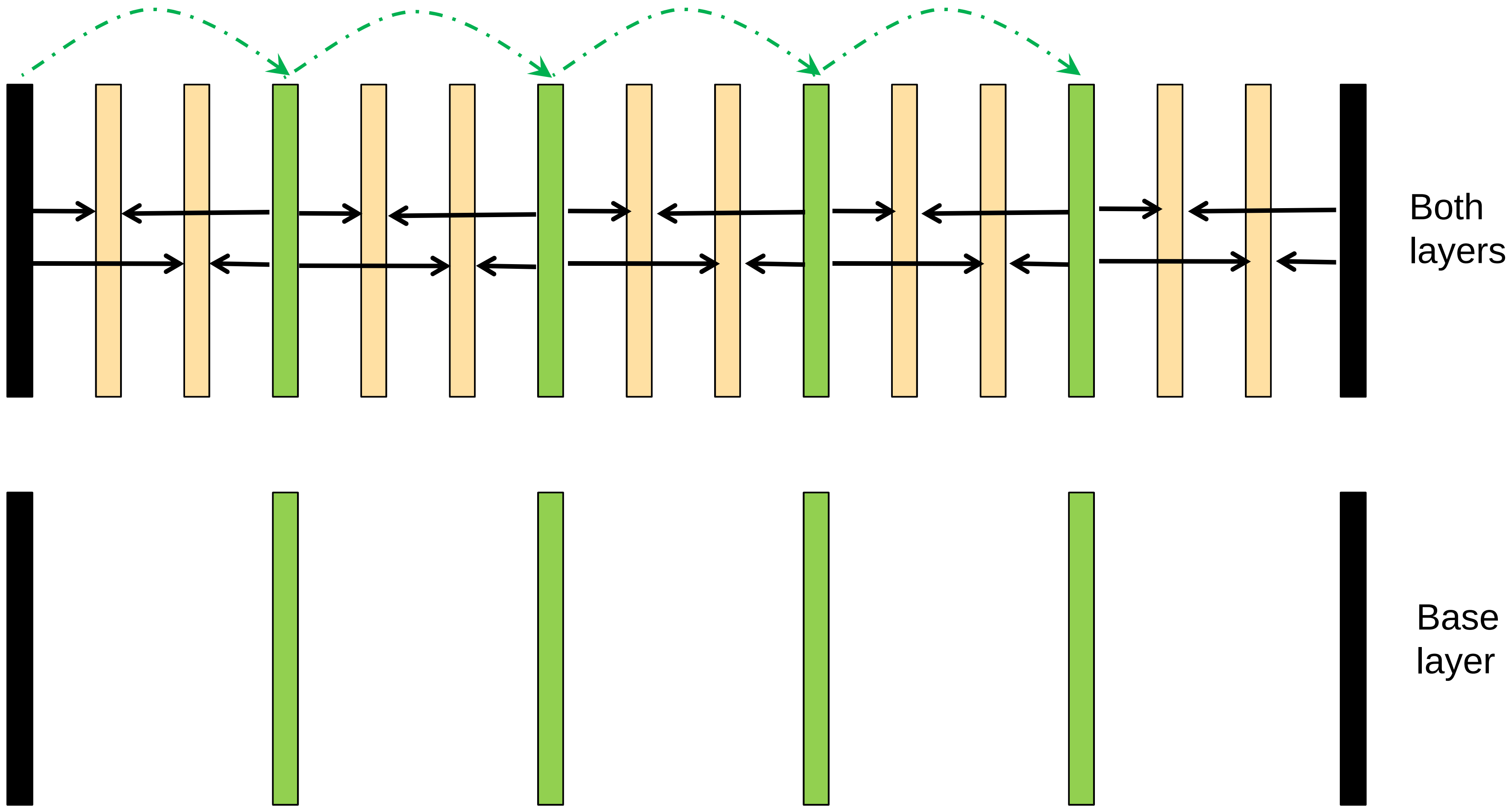
- Compression efficiency
- Encoder and decoder complexity
- Resilience to losses
- Flexible partitioning for rate adaptation
  - Range of rate partitioning (ratio of base rate to total rate)
  - Number of partitions (finely granular, or a few discrete levels)
- Compatibility with standards
- Ease of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

- Prediction structure controls most of these!

Scalability methods

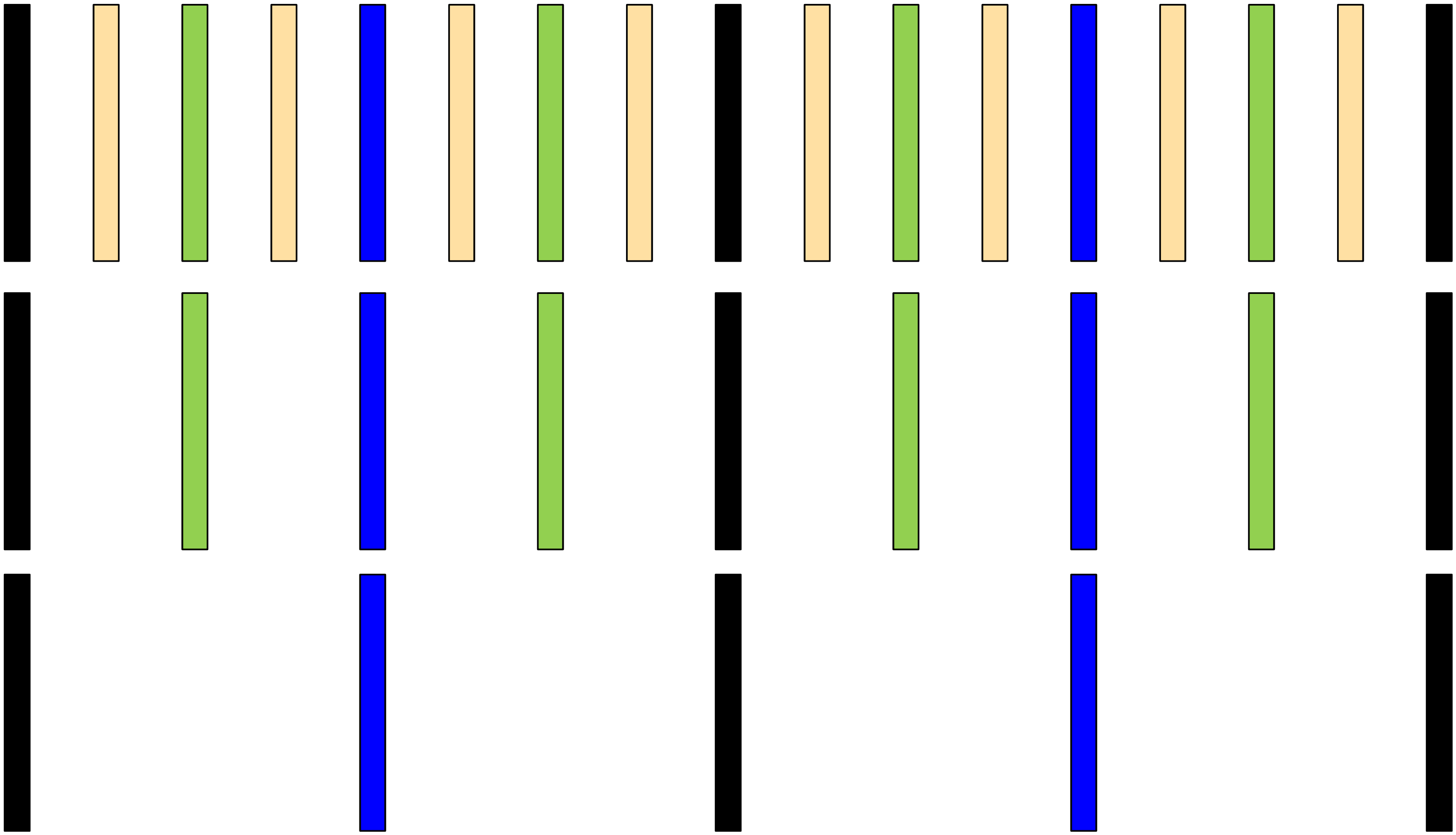
- Temporal scalability (frame rate)
- Spatial scalability (picture size)
- Amplitude (AKA SNR or Quality) scalability (quantization stepsize or QP)
- Frequency scalability (transform coefficients)
- Object-based or ROI scalability (content)

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

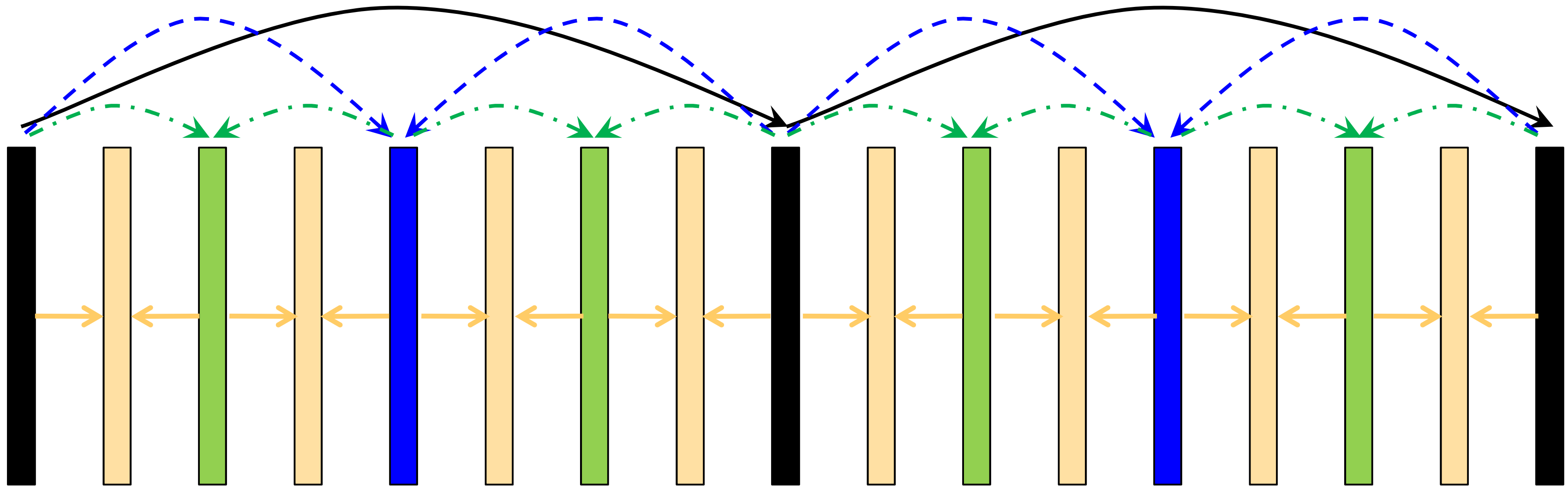


Can also be considered three layers: Layer 0: Black (I-frames), Layer 1: Green (P frames), Layer 2: brown (B-frames)

H.264: Temporal Scalability with Hierarchical prediction



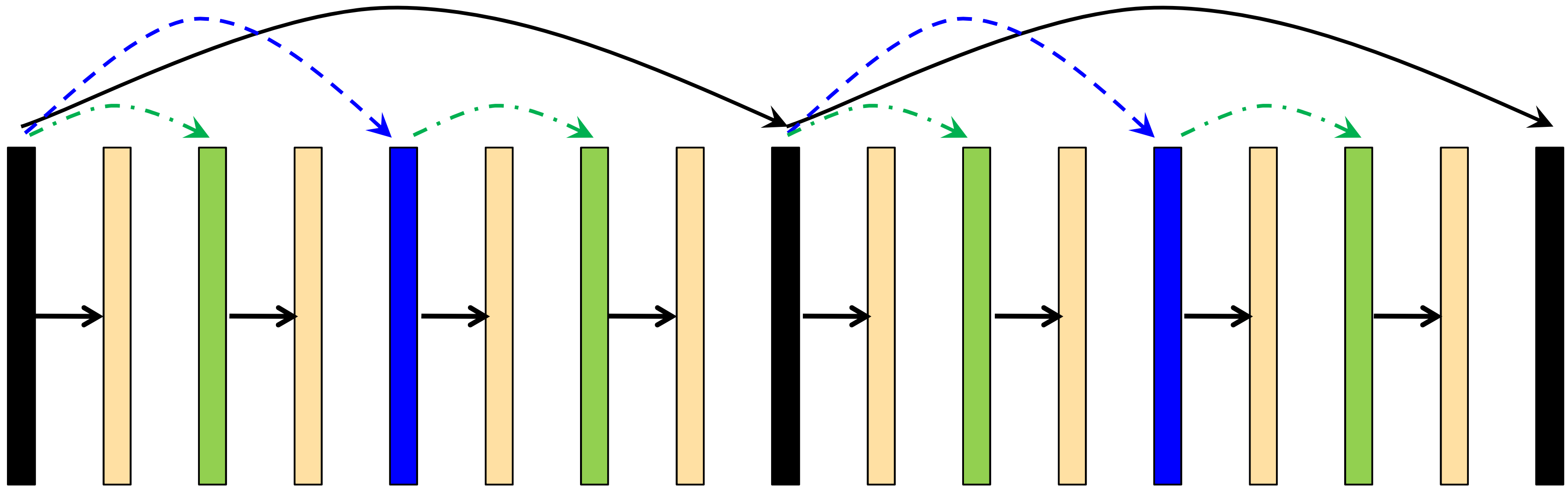
Temporal Scalability with Hierarchical B pictures



Problem: encoding delay = number of frames in a GOP (between black frames)

OK for non-realtime applications: live streaming, video-on-demand

Temporal Scalability with Hierarchical prediction and Zero delay (Hierarchical P)

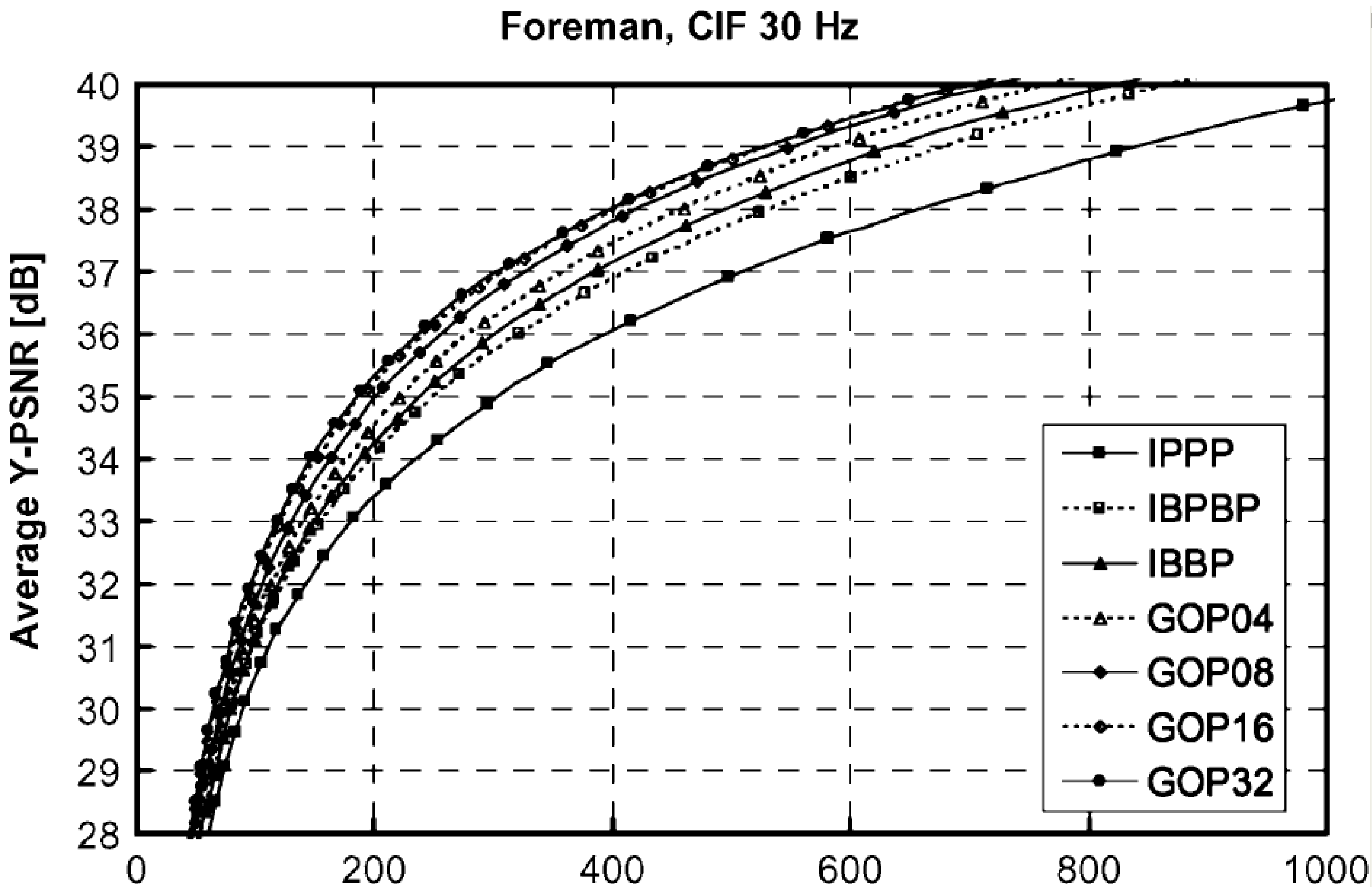


Good for realtime applications: chat or conferencing

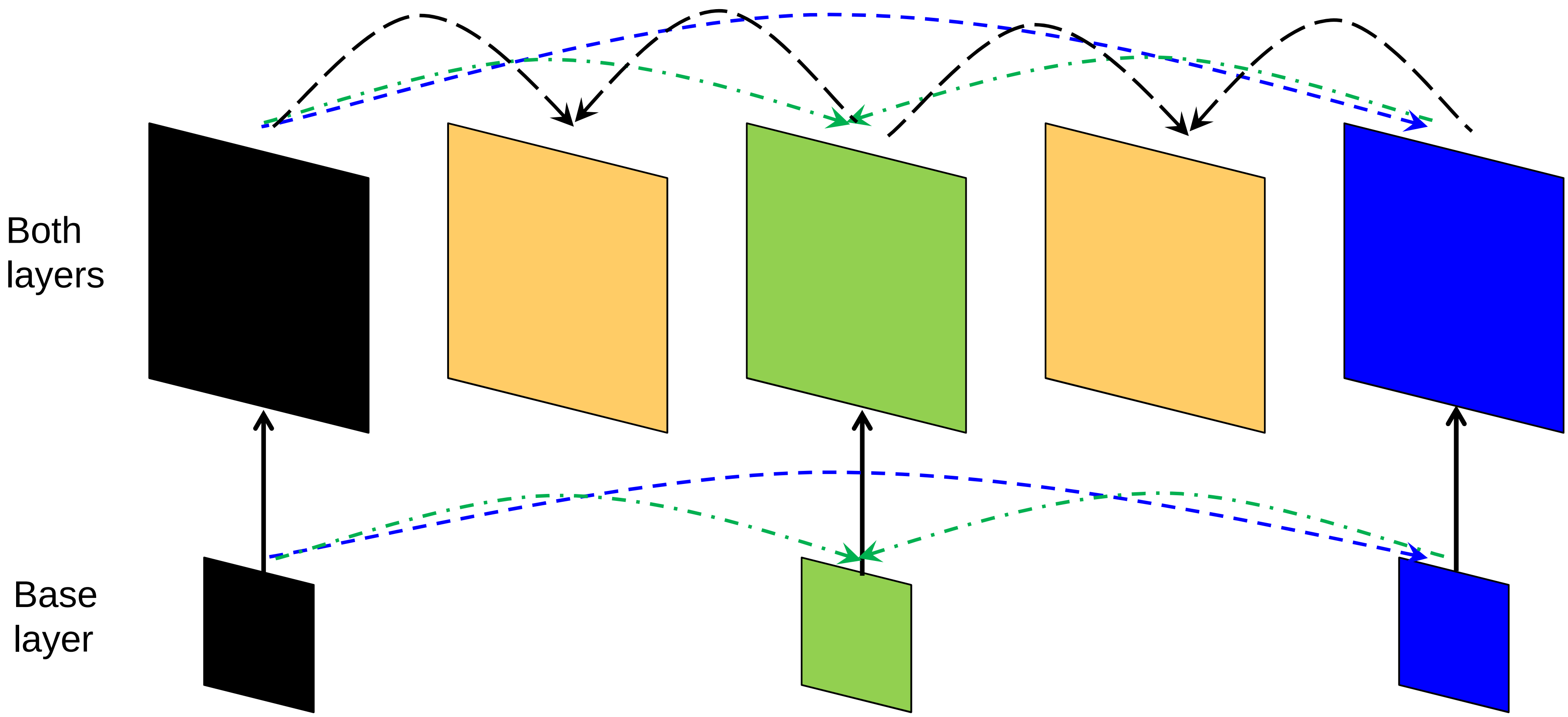
Comments about Temporal Scalability

- 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
  - Hierarchical B structure with large GOP size not only enables temporal scalability with many layers, but also generally improves coding efficiency over using IPP.. Structure.

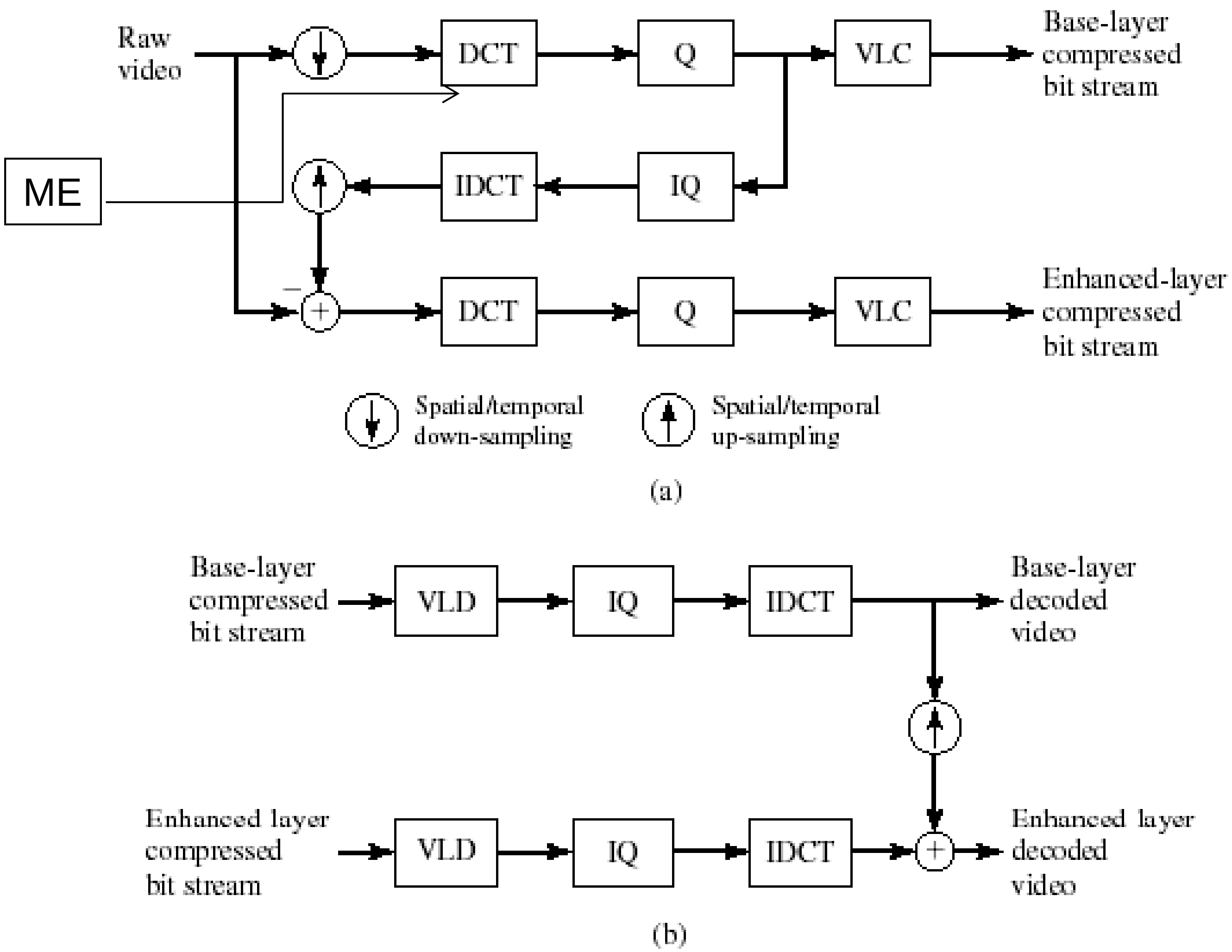
Efficiency of H.264 Temporal Scalability



Spatial and Temporal Scalability



Spatial Scalability
Through Down/Up Sampling

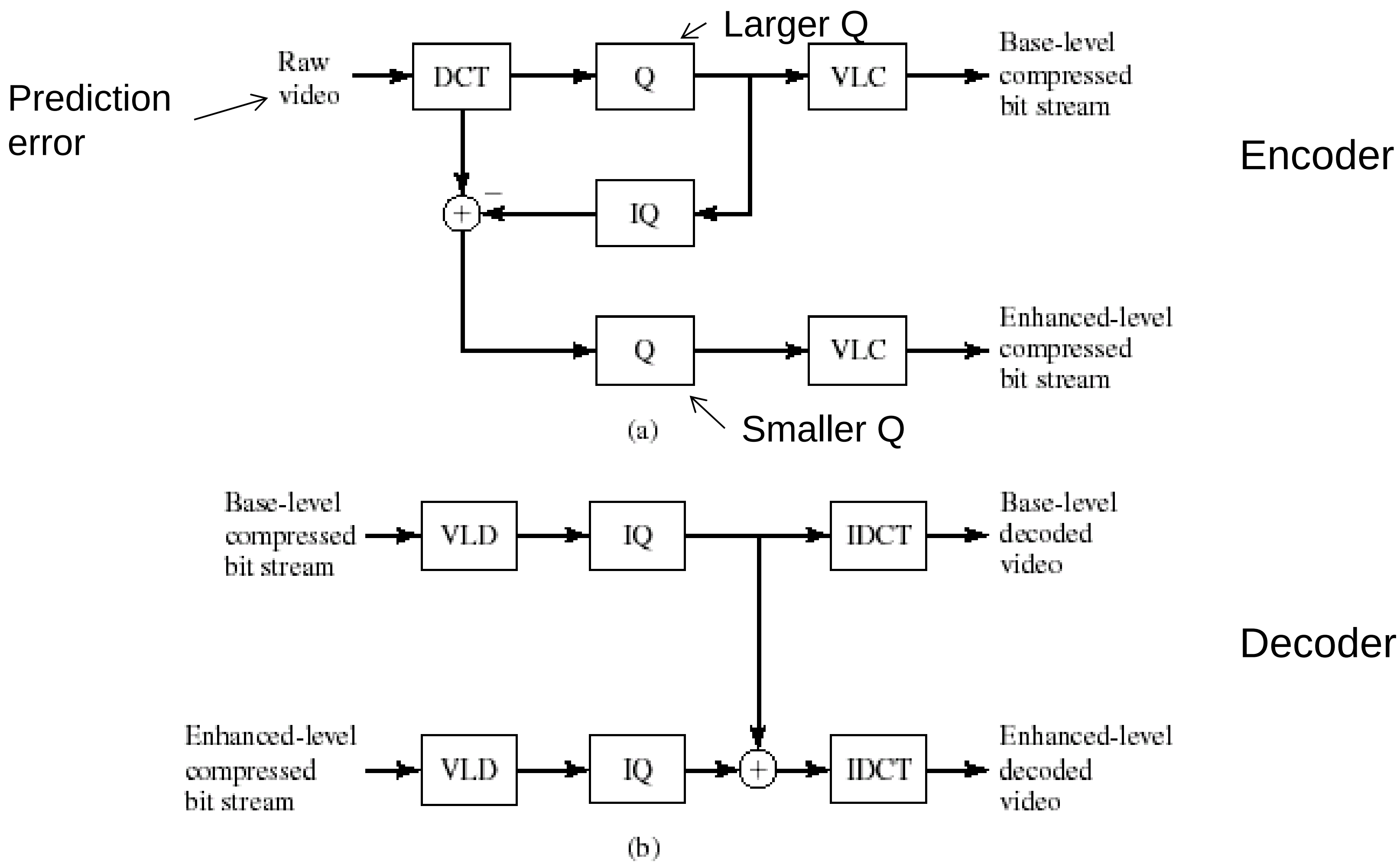


Amplitude Scalability

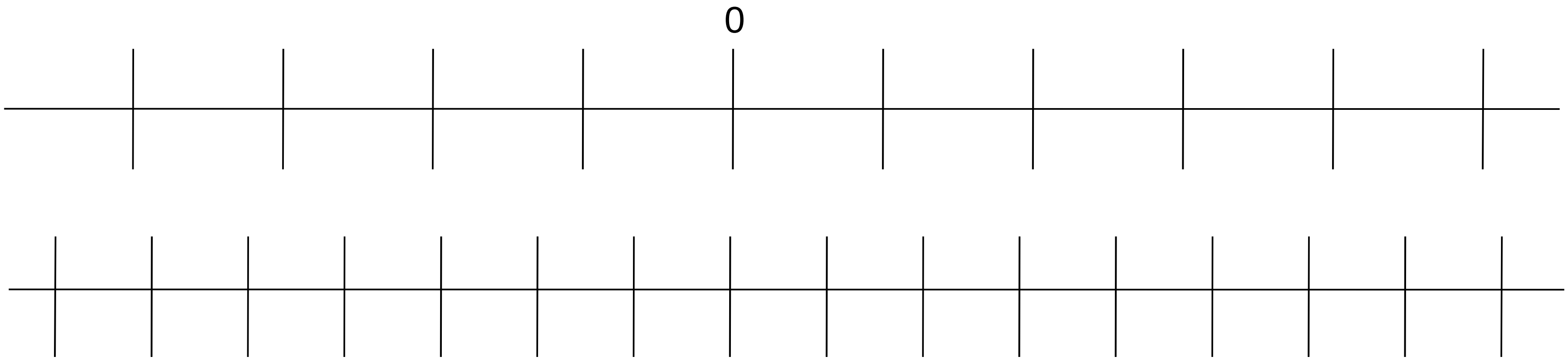
- Quality in each layer differs because of the quantization level

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

Amplitude (SNR) Scalability By Multistage Stage Quantization

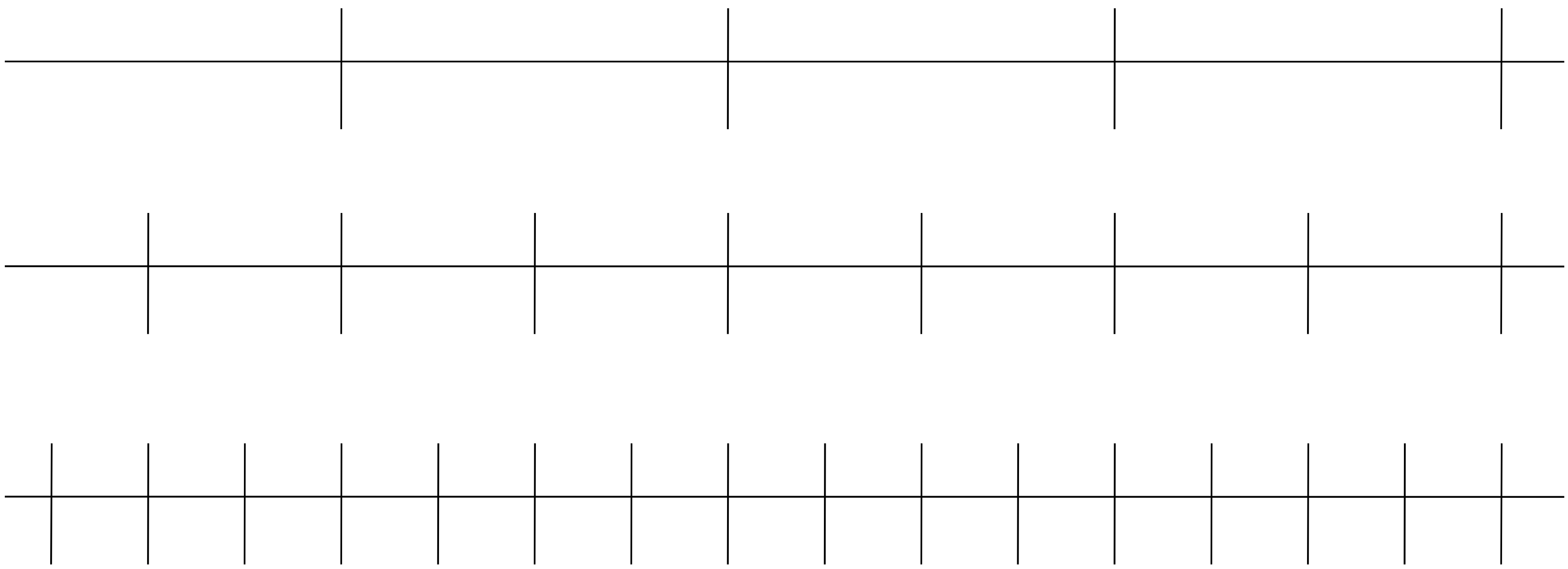


Multi-Stage Quantization



Bitplane coding

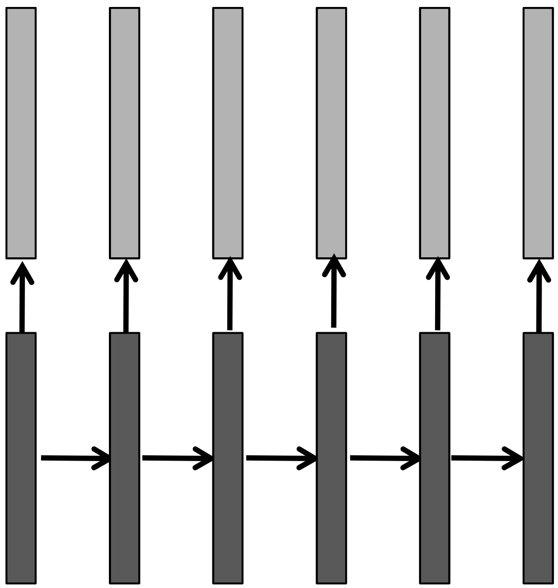
- Special case of multistage quantization, where successive step sizes differ by a factor of 2



Prediction strategies

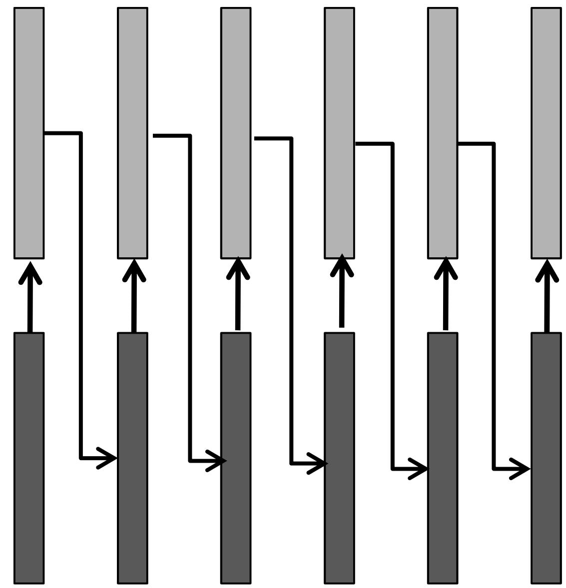
- Predict from the base layer only (Option 1):
  - Can be implemented with bit plane coding (MPEG4 FGS)
  - No mismatch at decoder
  - Low prediction accuracy if the base layer use large Q
- Predict from the highest layer (Option 2):
  - Mismatch at decoder receiving only lower layers!
  - When the prediction requires unavailable information, this is called “drift”
  - High prediction accuracy

Prediction structures for scalability (Options 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



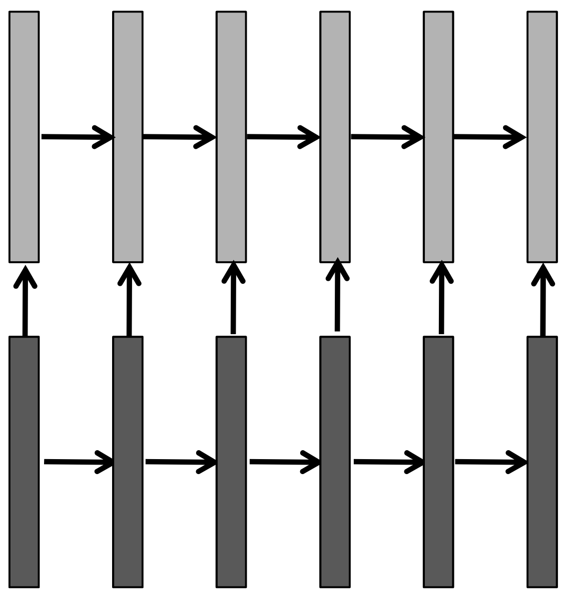
Enhancement layer is used to predict base layer

MPEG-2 SNR scalability
Errors propagate into base layer
More efficient

More Efficient Prediction Structures (Options 3 and 4)

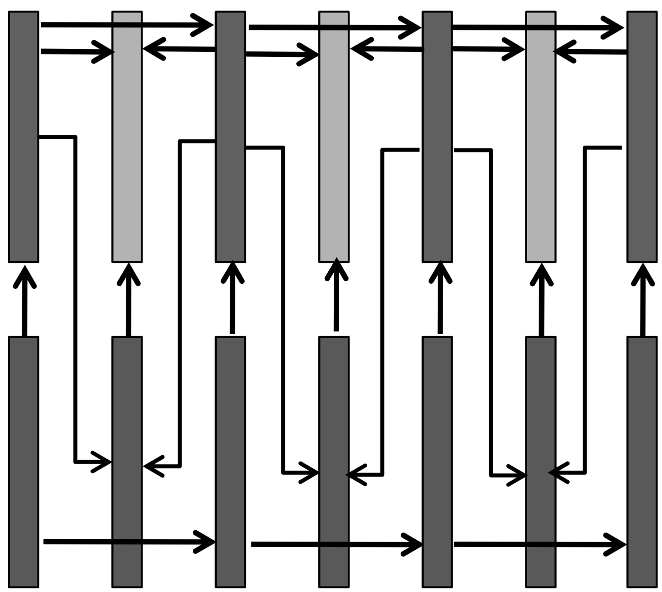
- Base layer predict from base layer; higher layer predict from either high layer or base layer (Two loop control) (Option 3)
- Allow base layer be predicted from enhancement layer; enhancement layer predict from enhancement layer (Option 4)

Prediction structures for scalability (Options 3 and 4)



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

MPEG-2 Spatial Scalability (2)
H.264 CGS
No drift in base layer
reasonably efficient



H.264 MGS:
Base: non-key frames predict using enhancement; key frames from base layer key frames
Enhancement: predict from enhancement
Tradeoff between efficiency and robustness

Allow both intra-layer and inter-layer prediction

- Inter-layer prediction
  - Predict from the same frame of the lower layer (higher Q), quantize the error using lower Q
- 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)

Frequency scalability
AKA Data Partitioning

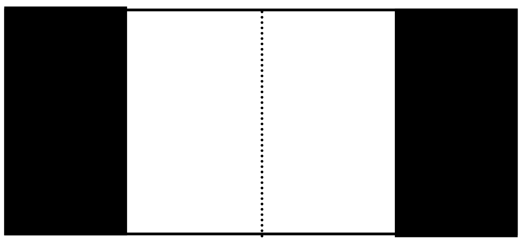
- Base layer: low frequencies of DCT
- Enhancement layer: remaining high frequencies of DCT

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

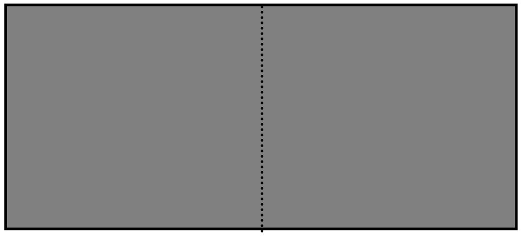
- One encoder prediction loop → missing the high frequencies means strong drift
  - (Prediction assumes all coefficients are available in the previous frame)

Frequency scalability: Effect of lost information

Two blocks at encoder:



Two blocks at decoder:

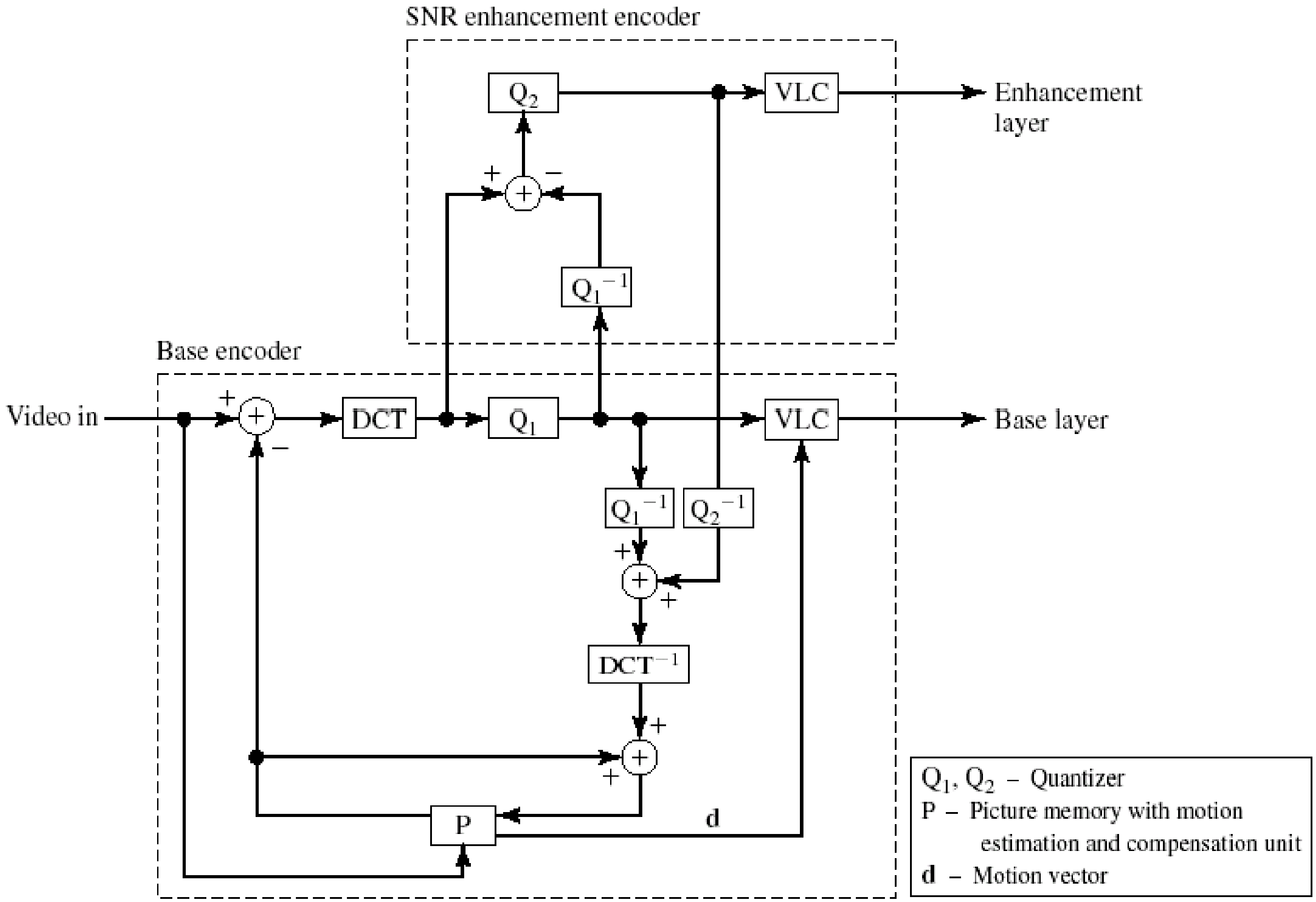


- Errors from previous frame propagate into current frame
- Motion causes error to spread, not just spatially, but *in frequency*
- Prediction method affects degree of propagation

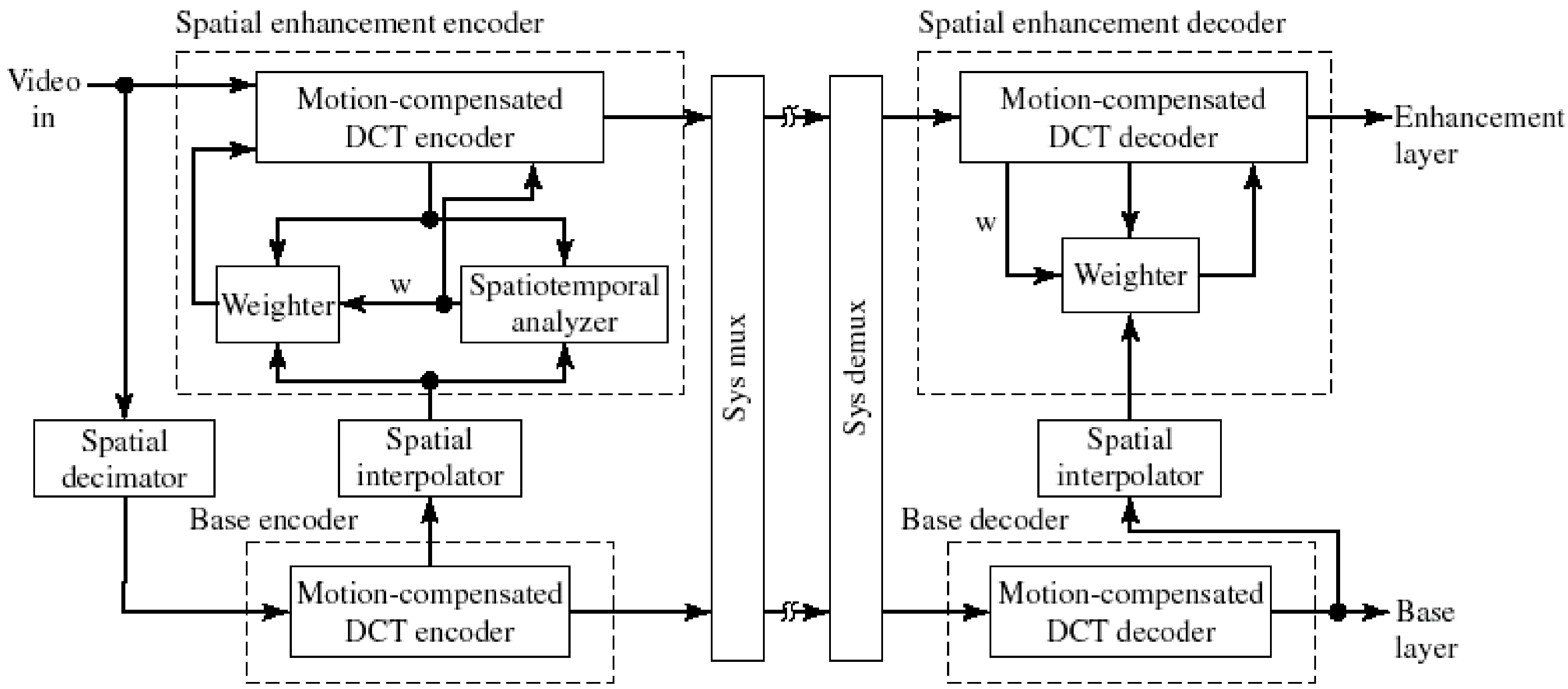
MPEG-2 Scalability: First standard that offers scalability

- 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
  - Predict from enhancement layer of previous frame
- Spatial scalability
  - Complex
  - Predict from previous frame of the same layer, or upsampled frame from lower layer
- 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



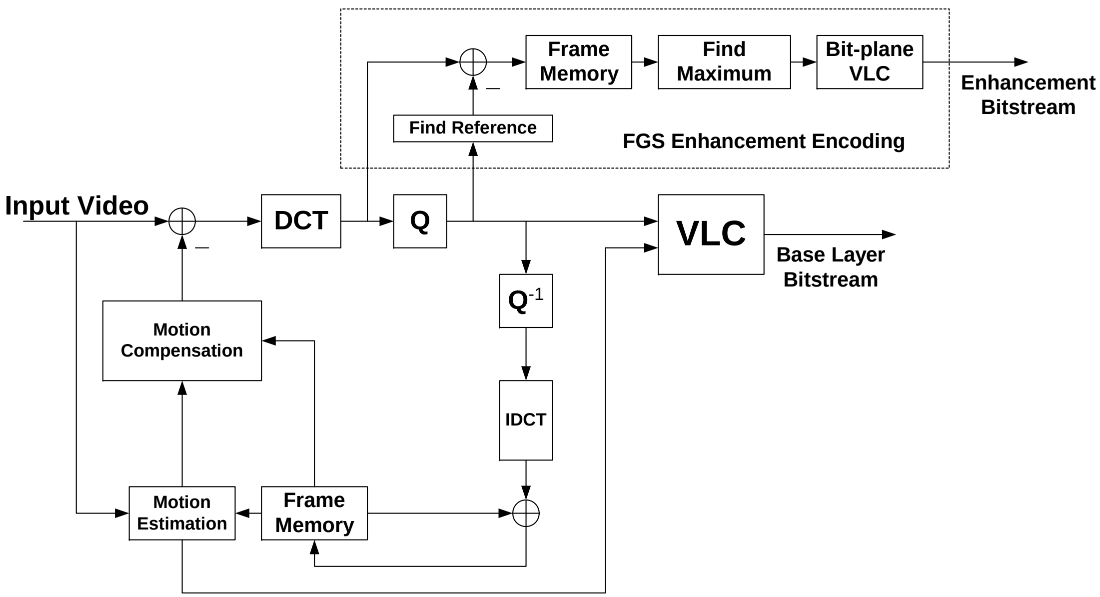
MPEG-2 Spatial Scalability Codec



Fine Granularity Scalability (FGS) in MPEG-4

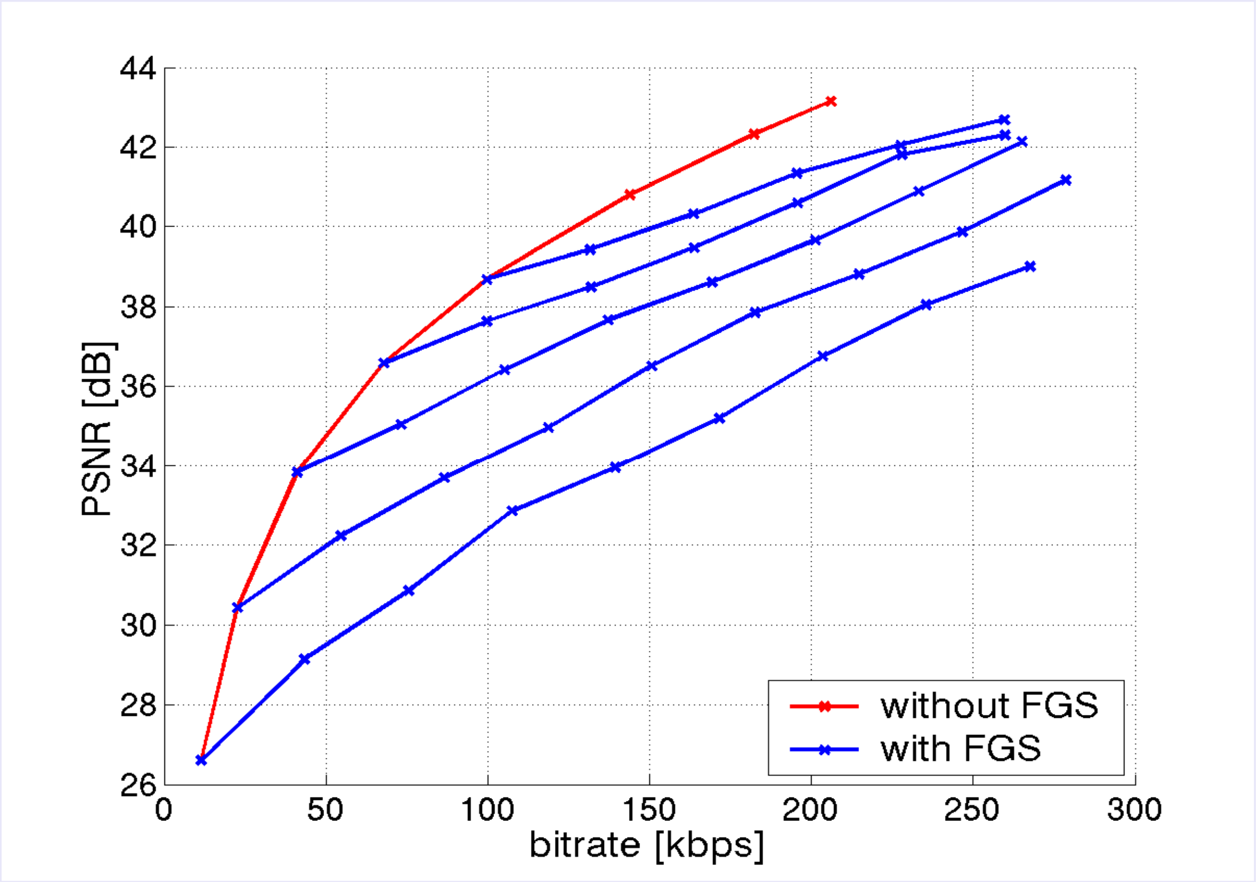
- MPEG-4 achieves fine granularity quality scalability through bit-plane coding
  - Base layer coded using a large QP on DCT coefficients
  - Quantization error for DCT coefficients are represented losslessly in binary bits
  - The bit planes are coded successively, from the most significant bit to the least.
  - The bit plane within each block is coded using run-length coding.
  - The same bit plane from all blocks forms one layer
  - Temporal prediction from base layer frames
  - Efficiency depends on base layer QP (or base layer rate)

Fine-Grained Scalability encoder



Encode once, decode to any bandwidth

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

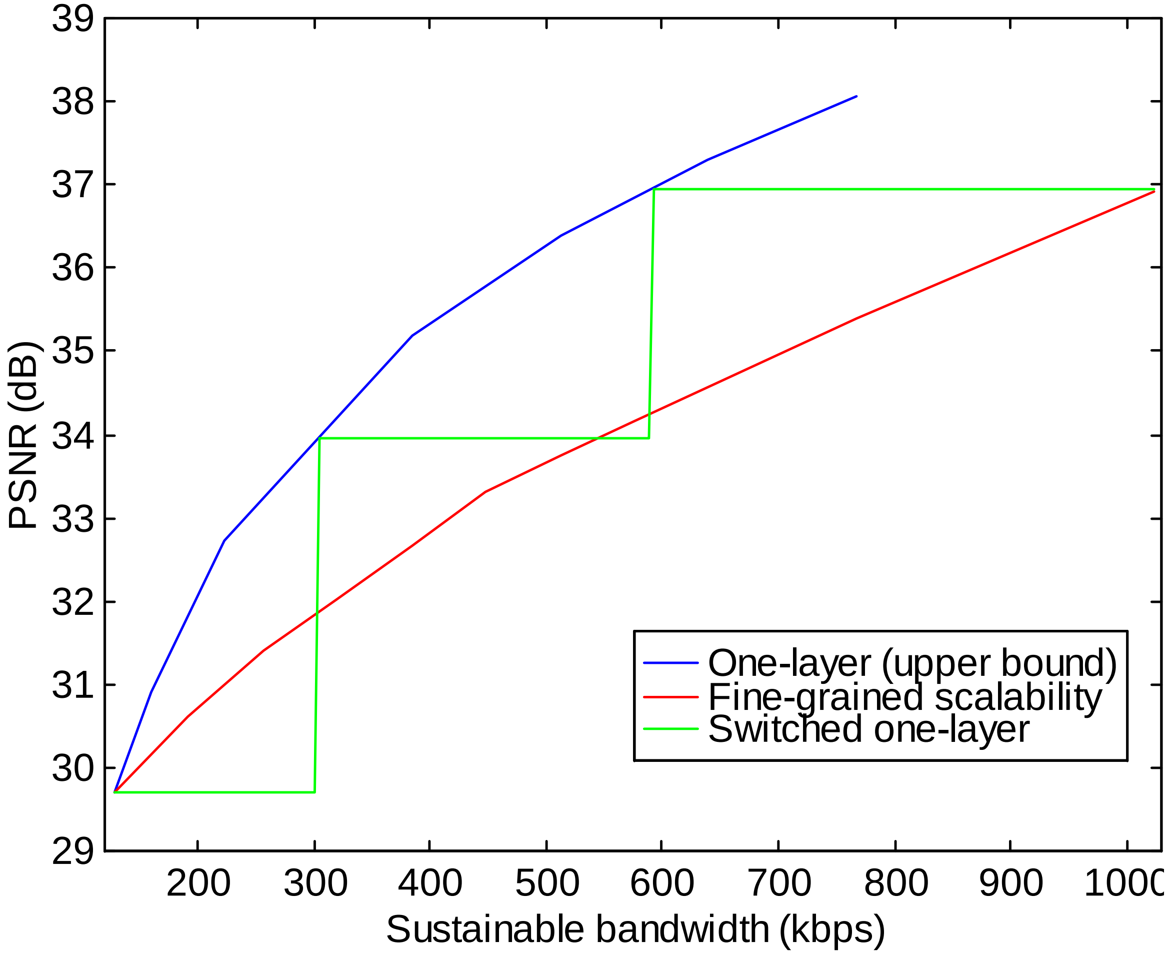


Each blue curve is obtained with MPEG4 FGS using different base-layer rate

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

Simulcast 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)

Temporal and Spatial Scalability of MPEG 4

- Temporal scalability is accomplished by combining I, B, and P-frames
- Spatial scalability is achieved by spatial down/up sampling

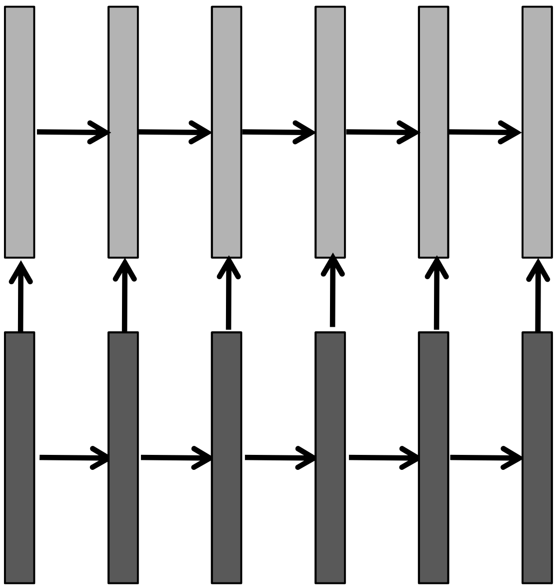
H.264 SVC (Scalable Video Coding)

- An optimized H.264/SVC encoder has an average overhead bit-rate of about 11% compared to non-scalable version (H.264/AVC)
- A good trade-off between efficiency and error-propagation/drift
- Decoding 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

SVC scalability modes

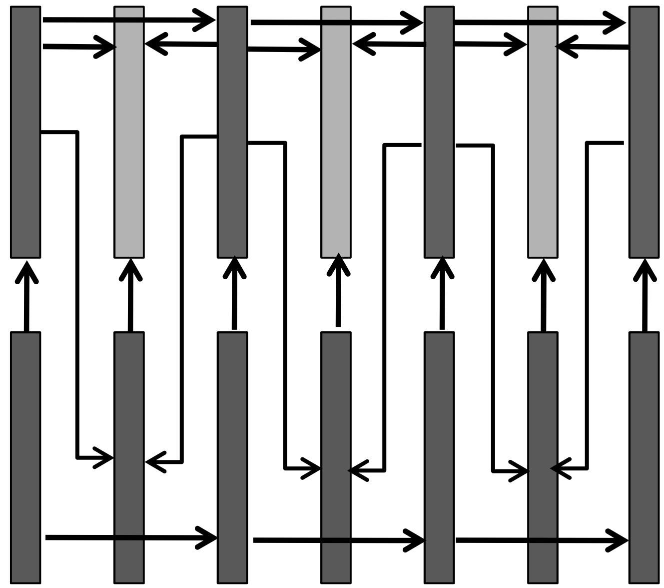
- Temporal scalability: using hierarchical B or hierarchical P structure.
  - No loss of coding efficiency when using hierarchical B
- Spatial scalability:
  - Using down/up sampling combined with switching between intra-layer and inter-layer prediction (CGS and MGS)
- Amplitude (quality) scalability
  - Same as spatial scalability where each layer has the same spatial resolution, but different QP
- QP cascading:
  - Using lower QP for lower spatial/temporal layers, increasing QP for higher spatial/temporal layers incrementally

Prediction structures for scalability (Options 3 and 4)



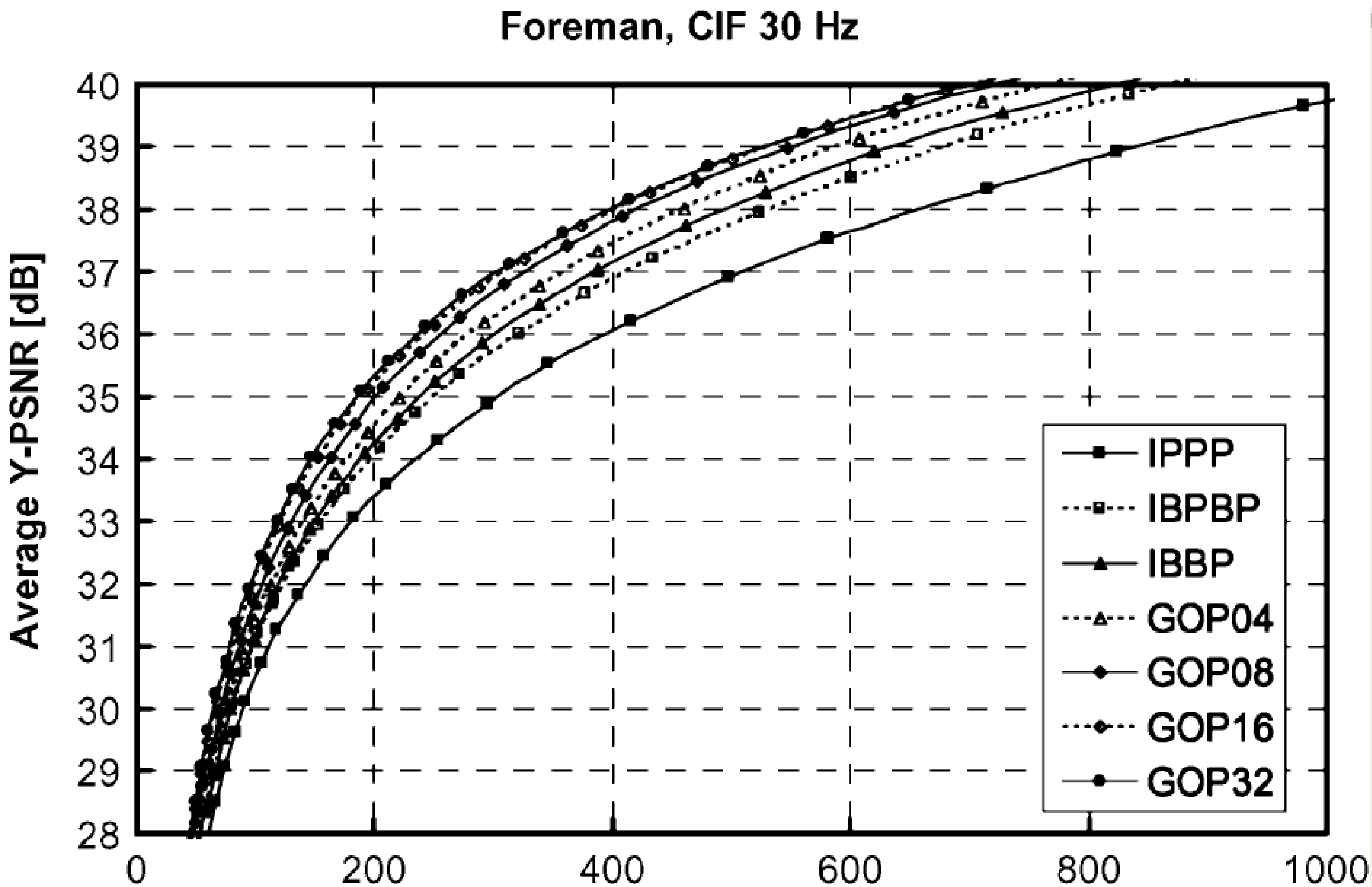
2-loop control
Both base and enhancement layers use their own prediction loop

MPEG-2 Spatial Scalability (2)
H.264 CGS
No drift in base layer
reasonably efficient

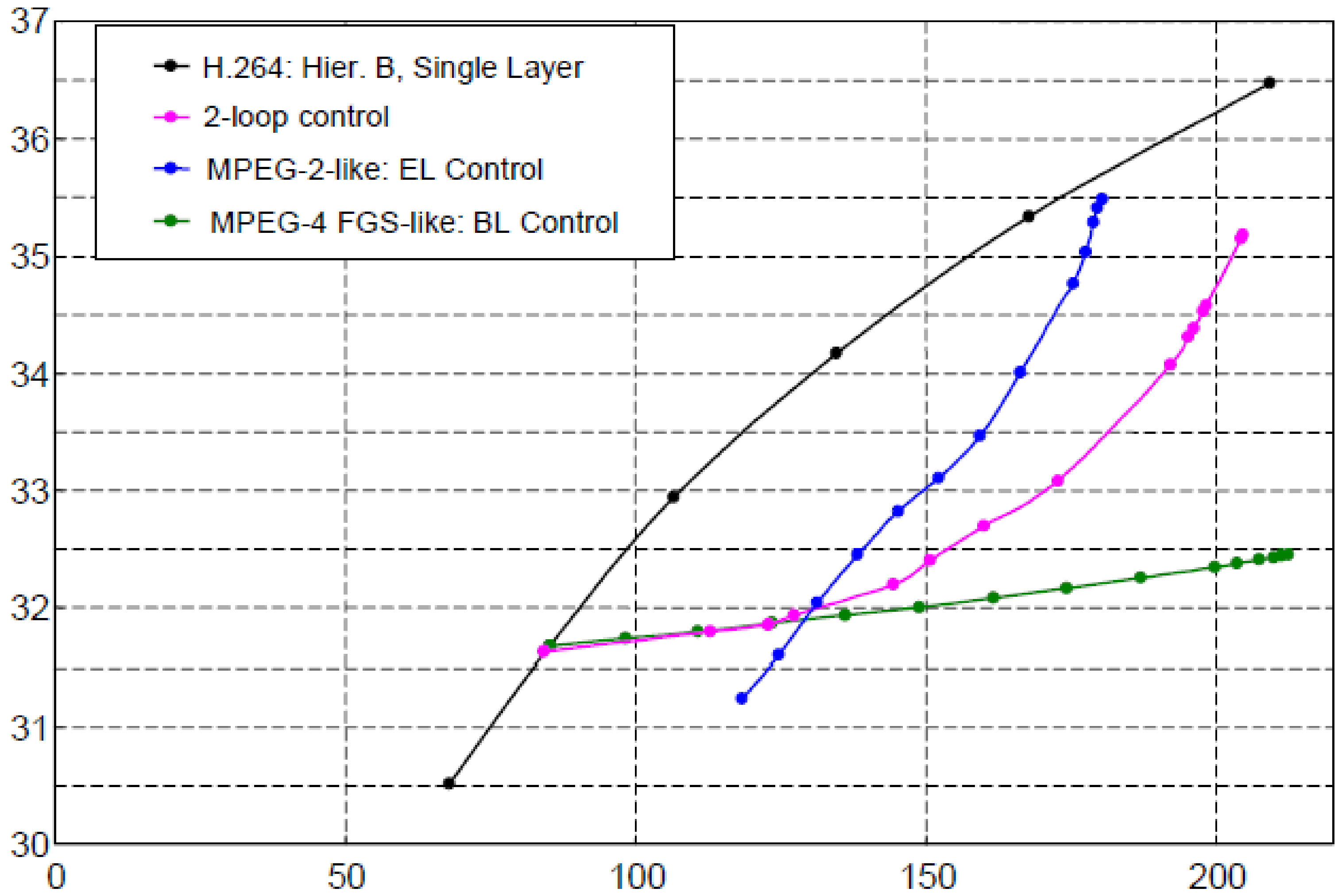


H.264 MGS:
Base: non-key frames predict using enhancement; key frames from base layer key frames
Enhancement: predict from enhancement
Tradeoff between efficiency and robustness

Efficiency of H.264 Temporal Scalability



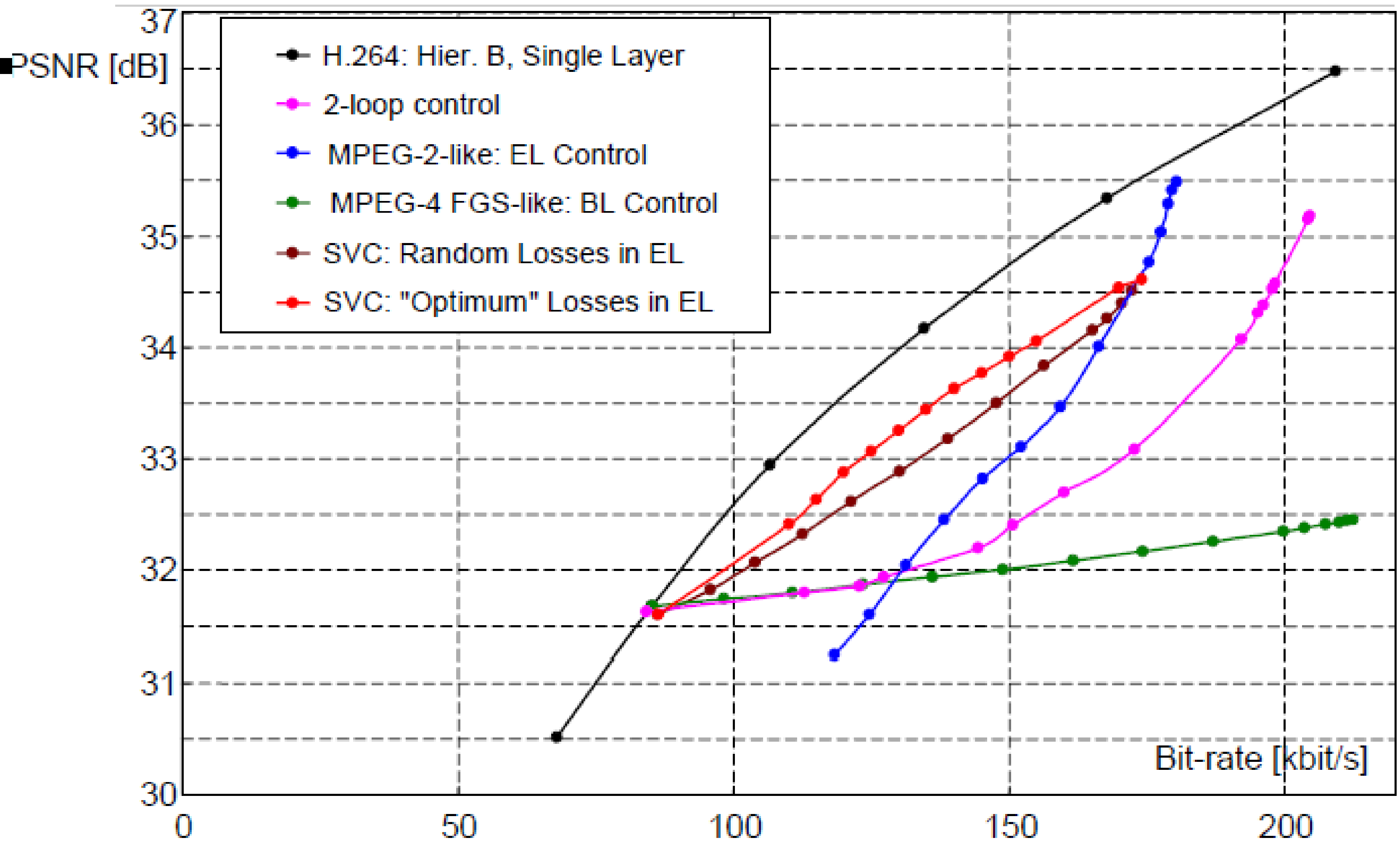
SNR scalability: Before H.264 SVC



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Thomas Wiegand: Scalable Video Coding

SNR scalability: with H.264 SVC



Fraunhofer Institut Nachrichtentechnik Heinrich-Hertz-Institut

Thomas Wiegand: Scalable Video Coding

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
  - So far has not outperformed block-based approach!

Homework and References

- Reading assignment: Sec. 11.1, 11.2, 11.3
- Written assignment
  - Prob. 11.3, 11.4

- Additional information:
- 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