Source coding alternatives for video transport over networks

Amy R. Reibman
AT&T Labs – Research
Florham Park, NJ

(*amy@research.att.com)*

Digital libraries in action

- How Raven’s QB Trent Dilfer prepares for Super Bowl
- “Show me every time the Giants blitzed during the playoffs”
- 90 hours of digital video, chopped into individual plays, hand-annotated with basis stats (down, who touched ball, result) and more – up to 30 labels per clip
- Vikings did it too, before 41-0 shutout!

Outline

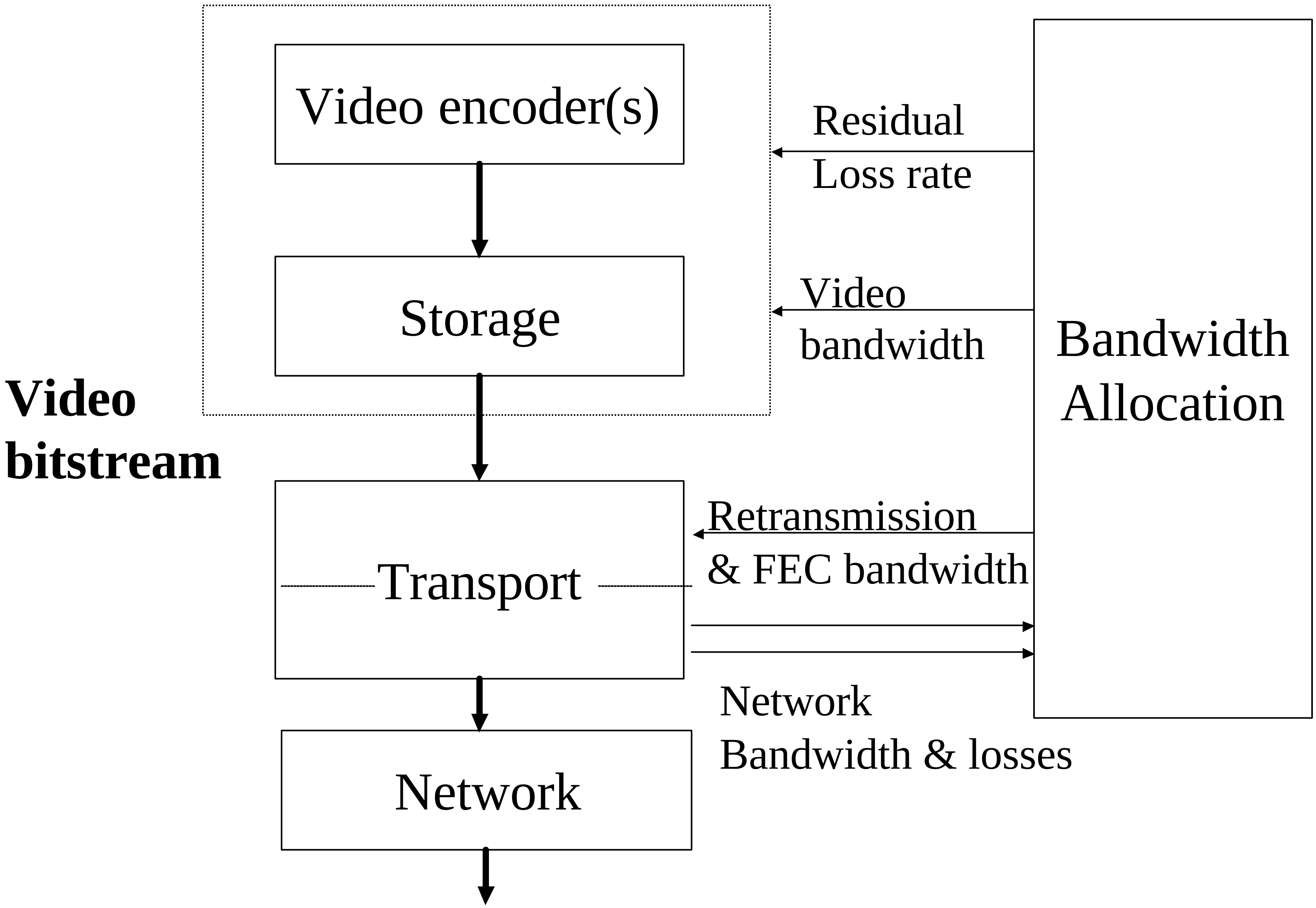
- Networked video introduction
  - Applications and simplified system
- Video coding basics
  - Tools for efficient compression
  - Impact of network impairments
  - Tools for less efficient compression
- Alternative source coding algorithms
  - Scalable (two-layer) video coding
  - Multiple Description video coding
- Which is best?!? When?!

Application scenarios

- Stored content on demand
- Personalized location-based ad insertion
- Security from remote location
- On-site uploads (real estate, construction, troubleshooting)
- Interactive communications

- Different requirements: loss, delay, throughput, storage
- Different networking solutions (wired, wireless)

Simplified system block diagram



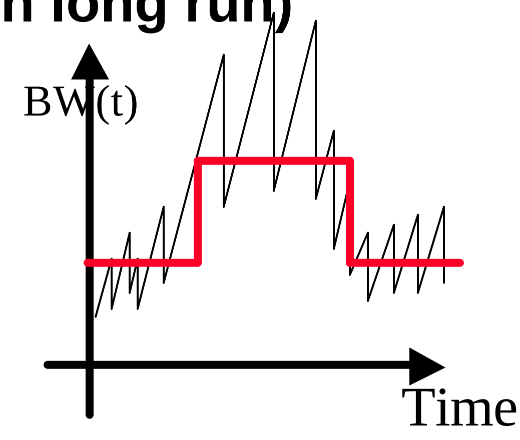
System components

- **Video encoder**
  - **Compression, resilience, flexibility of representation**
- **Storage**
  - **May provide a hard partition between encoder and transport/network**

- **Transport**
  - **Retransmission, FEC: reduce rate to decrease residual losses**
  - **Congestion control: adapts rate to fit connection**
- **Network**
  - **Introduces impairments: losses and delay**
- **Bandwidth allocation**
  - **Decide which components get how much bandwidth**

Important system issues

- **Algorithms to do bandwidth allocation**
  - **How much bandwidth for retransmission? FEC? Video-induced resilience? Actual video data?**
- **Algorithms for smoothing network rate into video rate**
- **Interaction between congestion control and available network bandwidth**
  - **(ie, being greedy may actually hurt in long run)**



Our focus

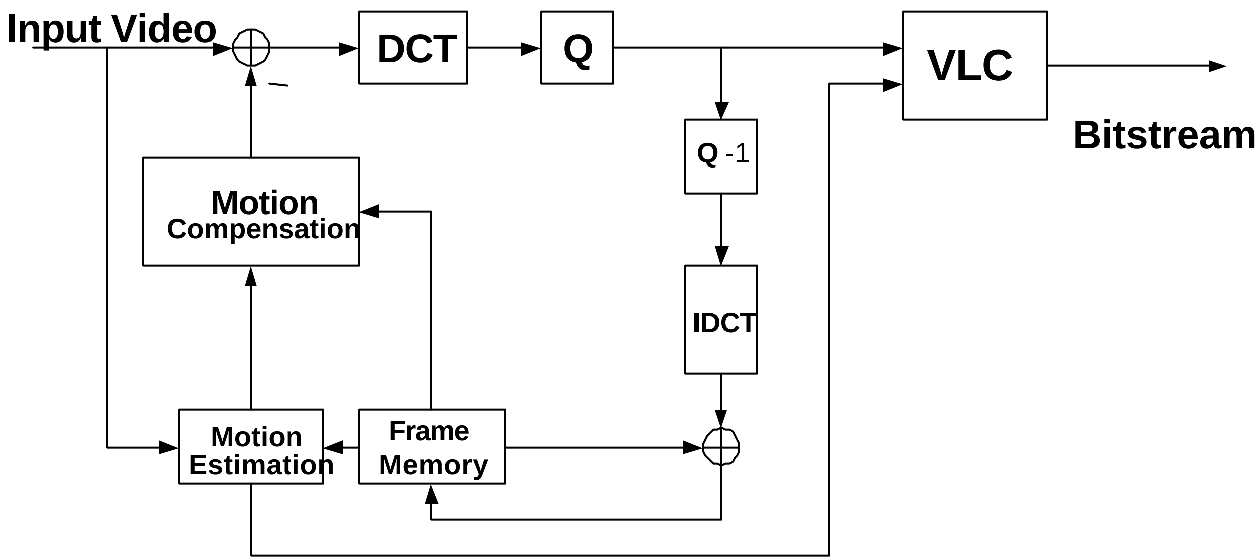
- **Assume bandwidth for video data is given, residual losses after transport may be present**
- **Consider impact on the video encoder design of:**
  - **Available bandwidth for video (BW)**
  - **Residual loss rate (LR)**
- **What video encoding strategies are best for different (BW, LR) pairs?**

- **How does lack of knowledge of (BW,LR) affect encoding strategy?**
- **How does lack of precision of (BW,LR) affect encoding strategy?**

Source coding options

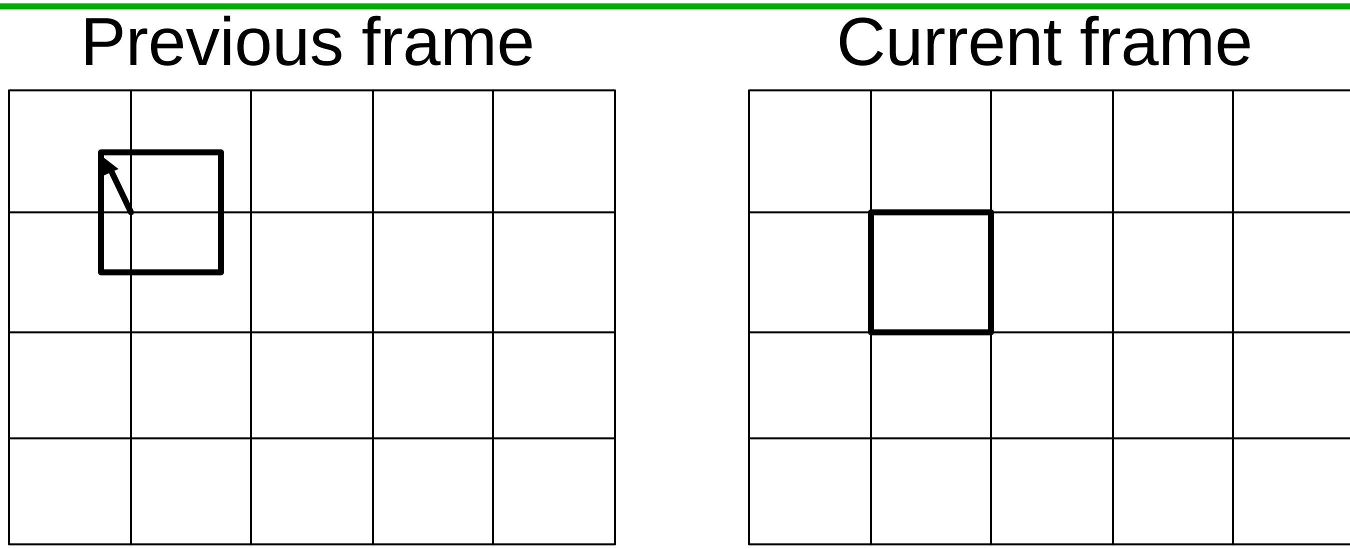
- (BW,LR) known, LR=0
  - Code as efficiently as possible into available BW!
- BW not exactly known at coding time, LR=0
  - Change representation of video to allow flexible adaptation to changing BW
  1) Scalable coding
  - Generally reduces compression efficiency
- LR != 0
  - Code less efficiently, but with more resilience to losses
  1) Reduce reliance on temporal prediction (intra-coding)
  2) Multiple description coding
  3) Scalable coding with prioritization

Video compression basics



- Lossless compression: VLC, DCT, temporal prediction
- Lossy compression in quantizer only
- DCT: block-based linear transform
- Temporal prediction: motion compensation

Motion Compensated Prediction (P)



- **Assumes rigid bodies move translationally; uniform illumination; no occlusion, uncovered**
- **Big win: Improves compression by factor of 5-10**
- **Motion-compensated interpolation (B-frames)**
  - **Incurs additional delay**
  - **Handles uncovered background and occlusion better**

Discrete Transforms

- Reorganize the image information before quantization (ie, introducing coding losses)
  - Eye is less sensitive to high frequency errors
  - Good energy compaction
- Discrete Cosine Transform
  - Block-based decorrelating transform
  - Nearly optimal for highly-correlated data
- Discrete Wavelet Transform
  - Not block-based
  - Natural spatial hierarchy
  - Better spatial localization of high frequencies
  - Not well suited to block-based motion-compensation

Variable Length Coding (VLC)

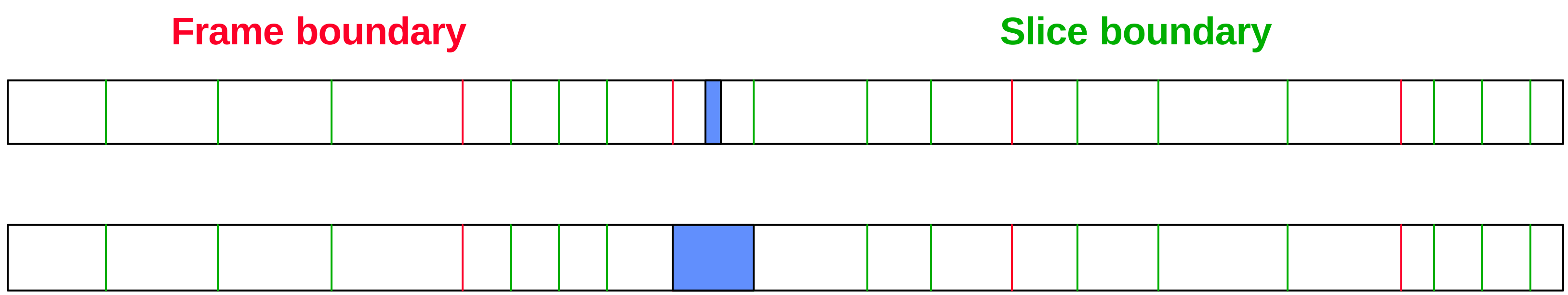
- More frequent values sent with fewer bits
- Prediction-error images have lots of zeros!

- Applied to
  - Macroblock address increments
  - Coded block patterns
  - DCT coefficients: (zero-runs, non-zero level) pairs
  - Motion vectors
- Good compression
- Poor error resilience

Impact of errors: spatial propagation

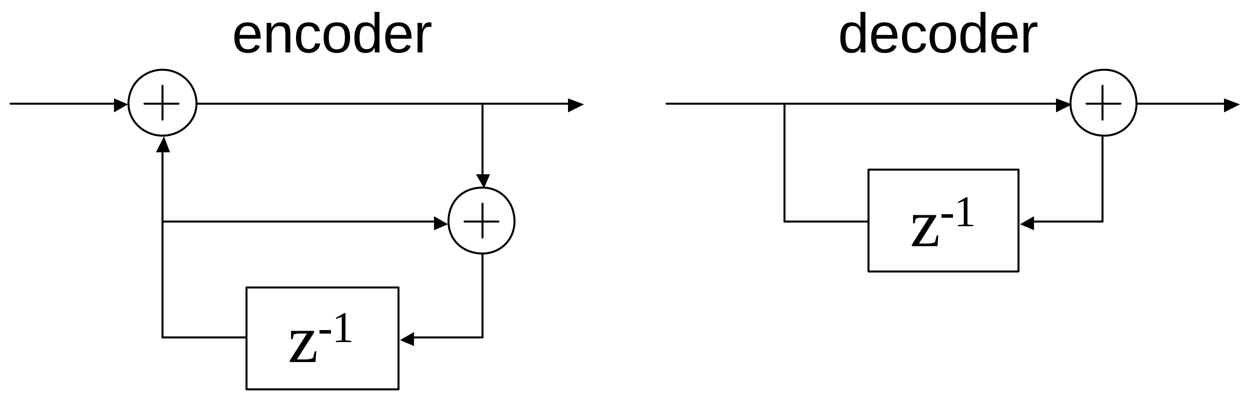
- **Spatial error propagation**
  - **Variable Length Coding loses synchronization**
  - **No re-entry into bitstream**

  - **Add slices a.k.a. synchronization codewords**
  - **More slices -> lower video quality**
  - **Fewer slices -> less error resilience**



  - **Reversible VLCs (helpful with bit errors only)**

Impact of errors: temporal propagation



- **Temporal error propagation**
  - **Temporal prediction and motion compensation causes errors to spread, not just spatially but also in frequency**
  - **Use motion-compensated concealment**
  - **Use more Intra-blocks; predict only from known-good frames**
- **Prediction method affects degree of propagation**

Optimizations for one-layer video

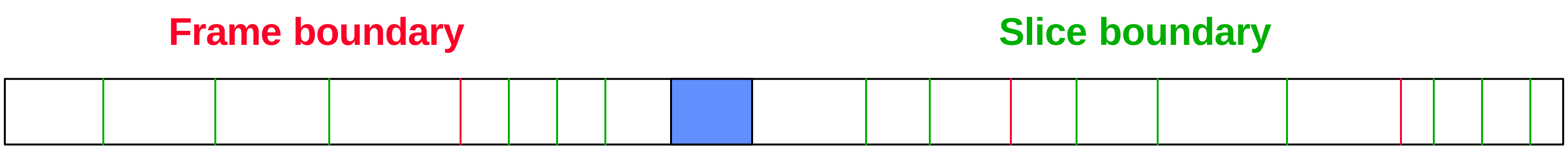
- Maximize efficiency: best quality for BW
  - Choose best possible predictor for each MB
  - Choose best quantizer given total available rate

- Minimize impact of losses: best quality for LR
  - Minimize temporal propagation: choose best location of intra blocks
  - Minimize spatial propagation: choose best slice length
  - Work with decoder’s known method of concealment

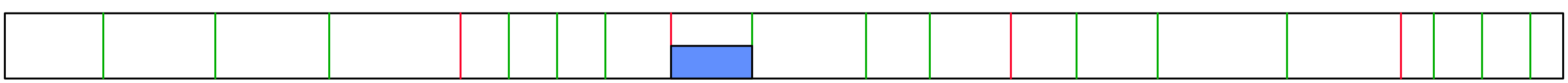
- Methods assume (BW, LR) known exactly

Transport for one-layer video

- **Fit entire slice into packet**
- **Rate adaptation on existing bitstream difficult**
- **Cannot prioritize within a slice!**
  - **Ex: Low frequency information is more critical**
  - **Loss of the slice loses “monolithically”**



- **Prioritize within slice**



Scalable (two-layer) video coding

- **Encode such that there is an easy partition**
  - **Base layer contains important information**
  - **Enhancement layer contains supplemental information**

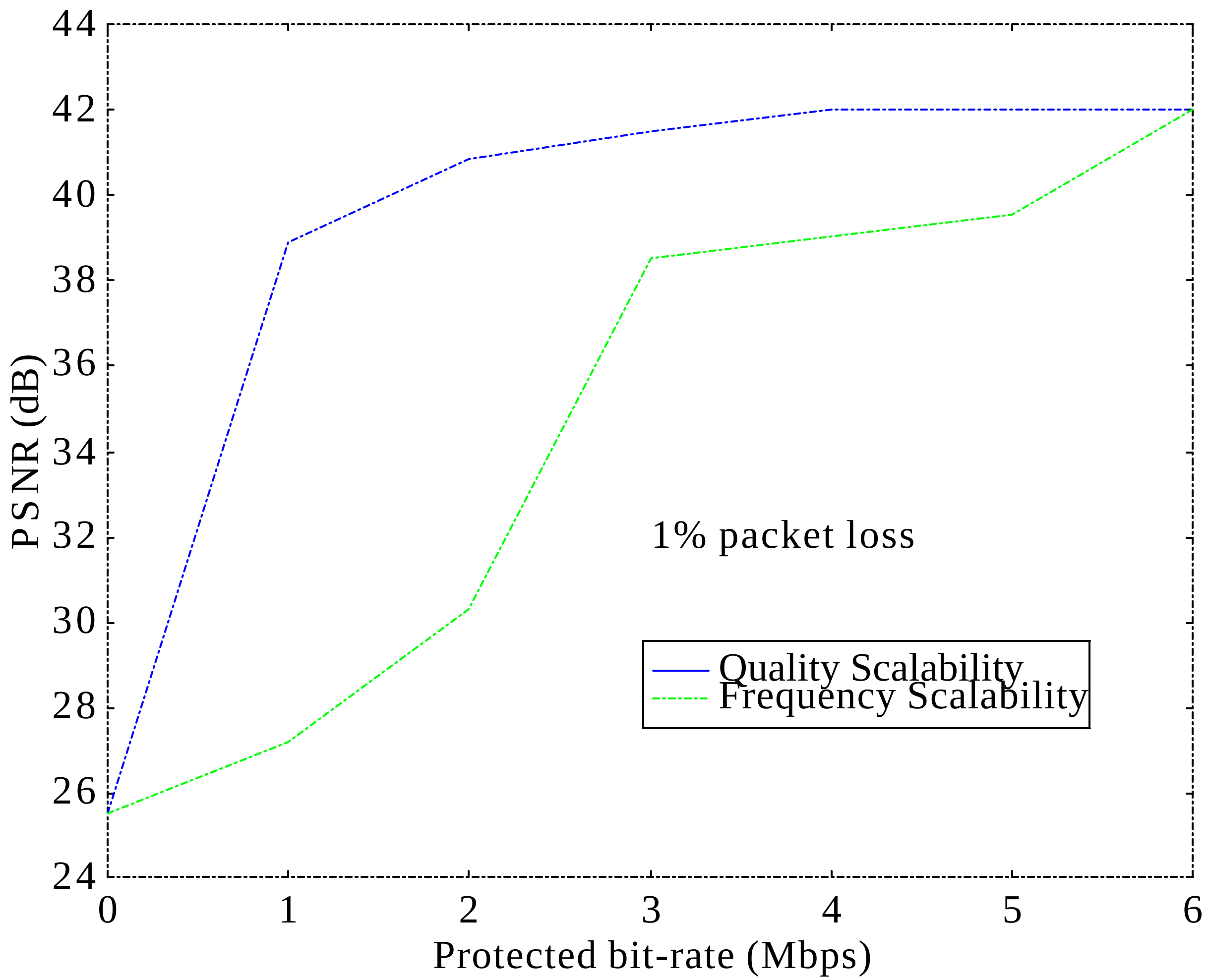
- **Flexible representation and easy prioritization are alluring for network transport**
  - **Unequal loss protection**
    - **Protect base layer more heavily in transport**
  - **Rate adaptation**
    - **Code once, decode to any bit-rate**
  - **Decouples network/video design**
    - **Performance is less sensitive to imprecision in BW**

Scalable encoding questions

- How to partition?
- What is impact on efficiency?
- Robustness?
- How many layers to use?
- How much BW in each layer?
- Which prediction strategy?

How to create partitioning?

Frequency, spatial, SNR/quality/bit-plane, temporal



How many layers?

- **Partitioning itself may introduce only minor inefficiencies *within a frame***
  - **Temporal: none**
  - **Frequency: marginal 1%**
  - **Quality/SNR: some, depends**
  - **Spatial: most for DCT, little for wavelets**
- **"Metadata" increases with number of partitions**
  - **Identifying where partitions start and what each contains**

So why do most standardized scalable video coders lose 1-2+ dB compared to a one-layer coder, and have only 2 layers???

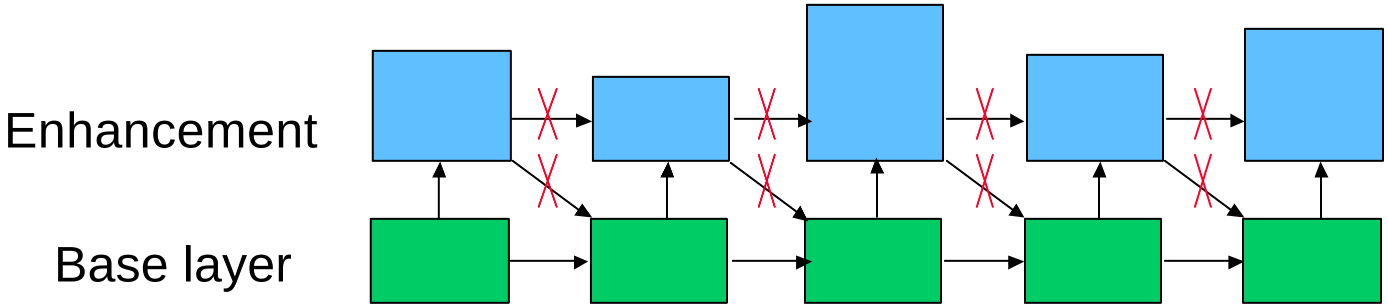
A simple scalable video encoder

- Add bit-plane encoding to one-layer video coder
- Change quantizer to be scalable quantizer
  - Loses about .1 dB quality
- Apply bit-plane arithmetic coding
  - Gains about 5-8% bit-rate
- Drop two bit planes and view decoder output

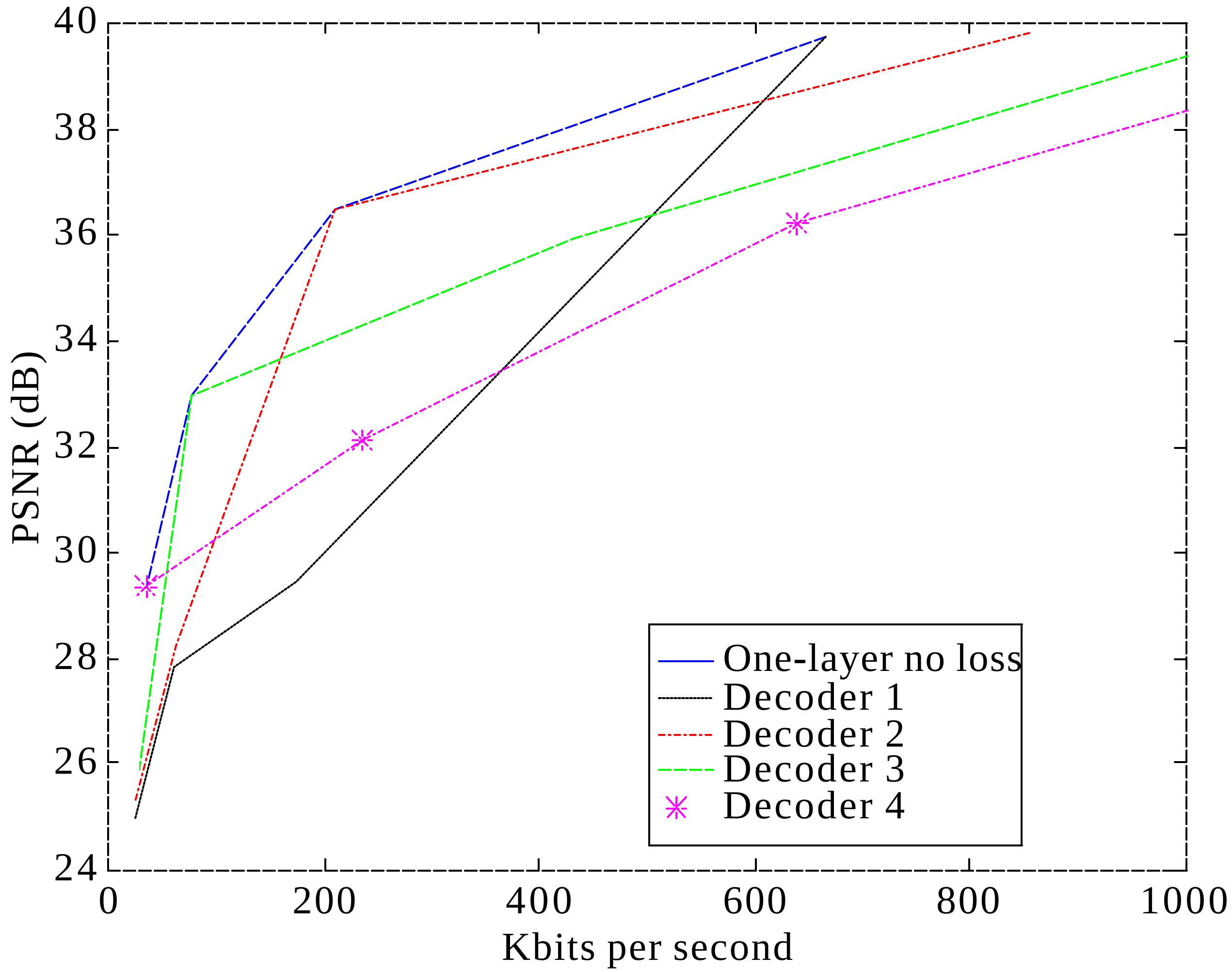
- *Mismatch*: encoder uses a predictor whose corresponding prediction error is not received by decoder
  - Ex: encoder uses a predictor with all bit-planes, but only subset of bit-planes is received by decoder

Prediction for scalable video

- What information is used for prediction?
  - Base layer can use base and/or enhancement information
  - Enhancement layer can use base and/or enhancement information
- Basis for decision
  - Efficiency
  - Robustness; error propagation; mismatch
  - Decoder complexity



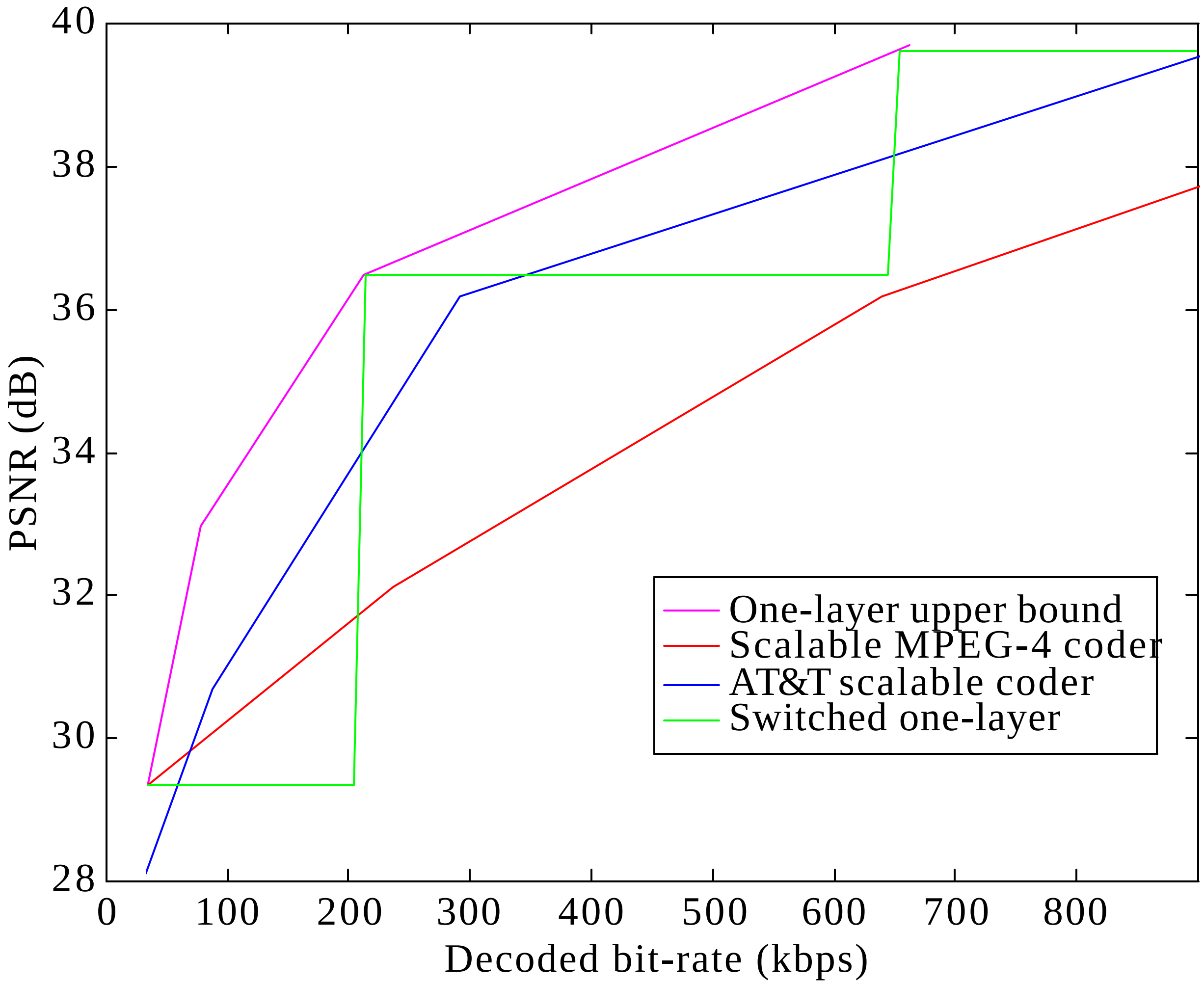
Influence of prediction, base-layer rate



Comparison: scalable vs. one-layers

- Download video when uncertain average thruput
  - Statistical knowledge about average throughput
  - Minimum sustainable throughput: 32 kbps
  - Maximum storage capacity: 900 kbps
  - Assume equally probable transmission rates in between
- Compare scenarios:
  - One scalable bitstream for all rates
  - Three one-layer bitstreams, switch depending on rate
  - Rate of all 3 one-layer streams must sum to 900 kbps
- Video quality (PSNR) averaged across range of rates

One-layer(s) vs. Scalability



Average PSNR for switched one-layer is more than 1.5 dB better than average PSNR for MPEG-4 FGS

More efficient scalable coders beat switched 1-layers by more than 0.3dB

Scalable video conclusions

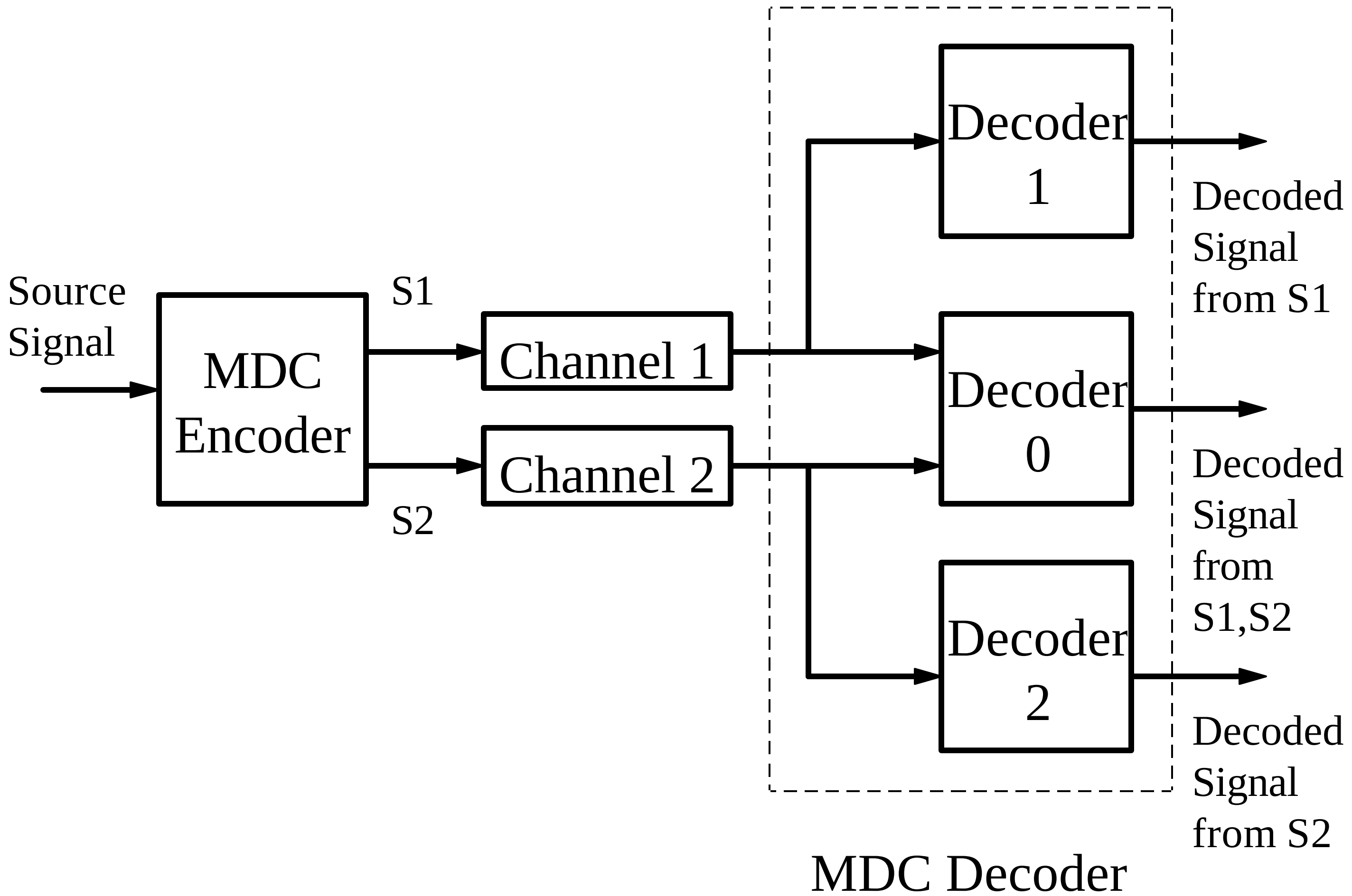
- Flexibility of representation is promising for network transport
- Partitioning of bitstream into two unequal pieces is helpful
  - If BW is not exactly known, easy rate adaptation
  - If LR is known, Unequal Error Protection effective
  - Enhancement-layer losses are not a problem
  - Base-layer losses are a problem!
- What about partitioning the video bitstream into two “equal” pieces?

The Multiple Description (MD) problem

- **Assumptions:**
  - Two channels between source and destination
  - Independent error/failure events
  - Probability that both channels fail simultaneously is low
- **An abstraction of an erasure channel**

- **Naive approaches**
  - Duplicate all information on each channel
  - Alternate all information between channels
- **Design considerations**
  - Trade-off between coding efficiency and robustness

The multiple description framework



Multiple Description video coding

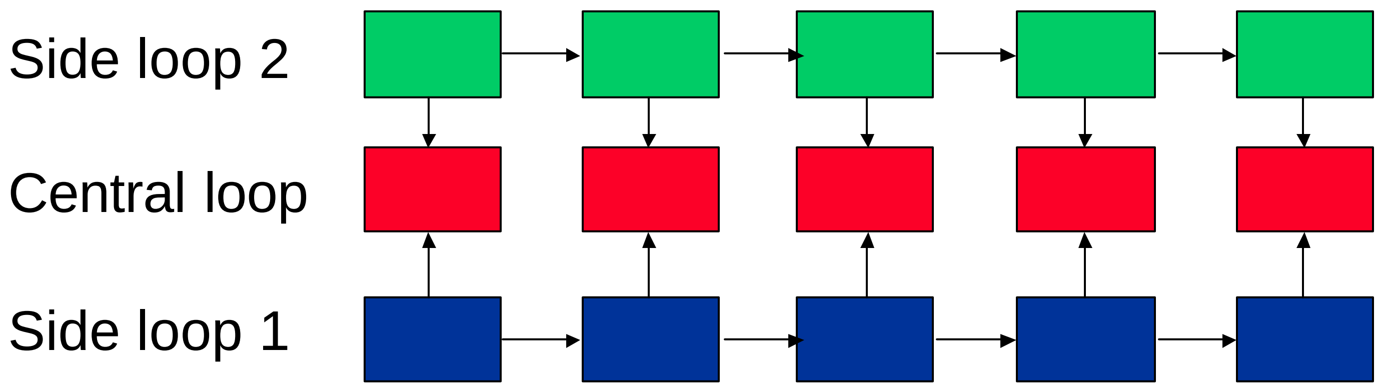
- Send two descriptions of the video
  - If one description is received, OK video quality
  - If both descriptions are received, even better video quality
- Simple approach: layered encoder
  - Duplicate base layer, alternate enhancement layer
- Other approaches:
  - Reintroduce correlation
  - Judicious alternation and duplication
- For same BW
  - Loss-free quality goes down
  - Robustness to packet losses goes up

Reintroducing correlation

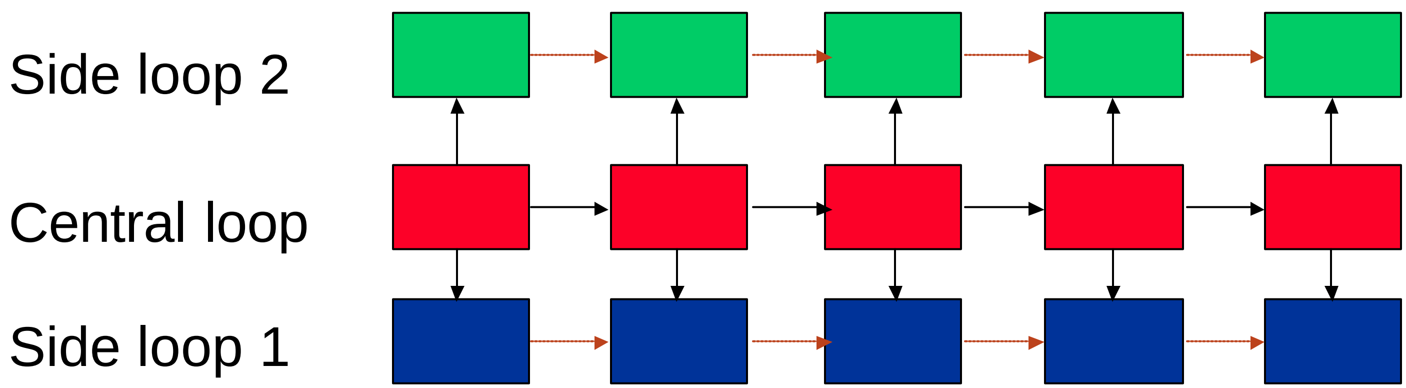
- Spatially
  - Subsample pixels in each frame
  - Code each sub-frame independently
- Temporally
  - Subsample temporally; code each substream separately
- In Quantizer
  - Design quantizer to have overlapping bins
  - One description received, like quantizing with 2*Qstep
  - Two descriptions received, like quantizing with Qstep
- In Transform
  - Apply correlating transform to pairs of coefficients
  - Use linear prediction to estimate a missing description

Prediction for MD video

- **Side loops have no mismatch**



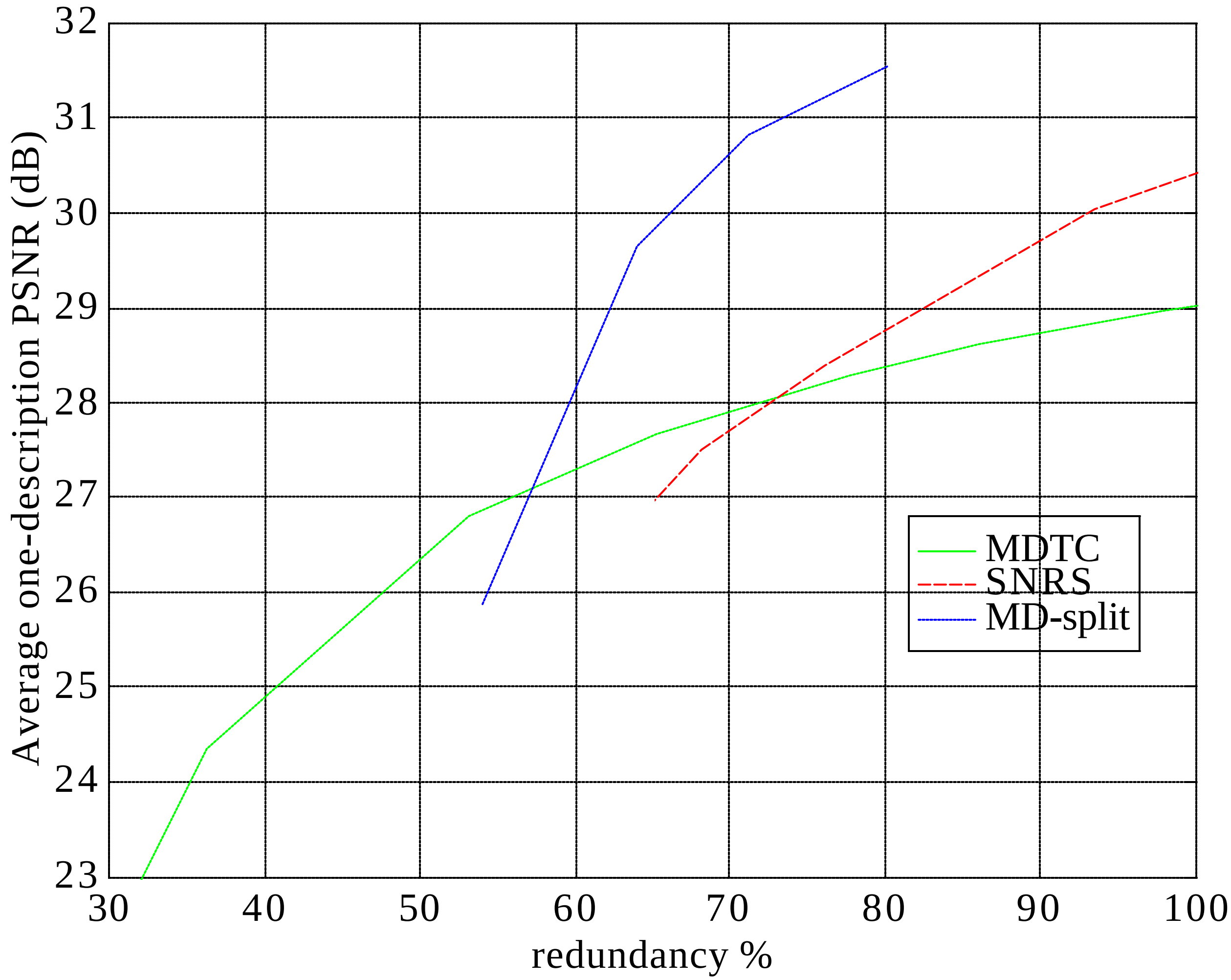
- **Central loop has no mismatch**



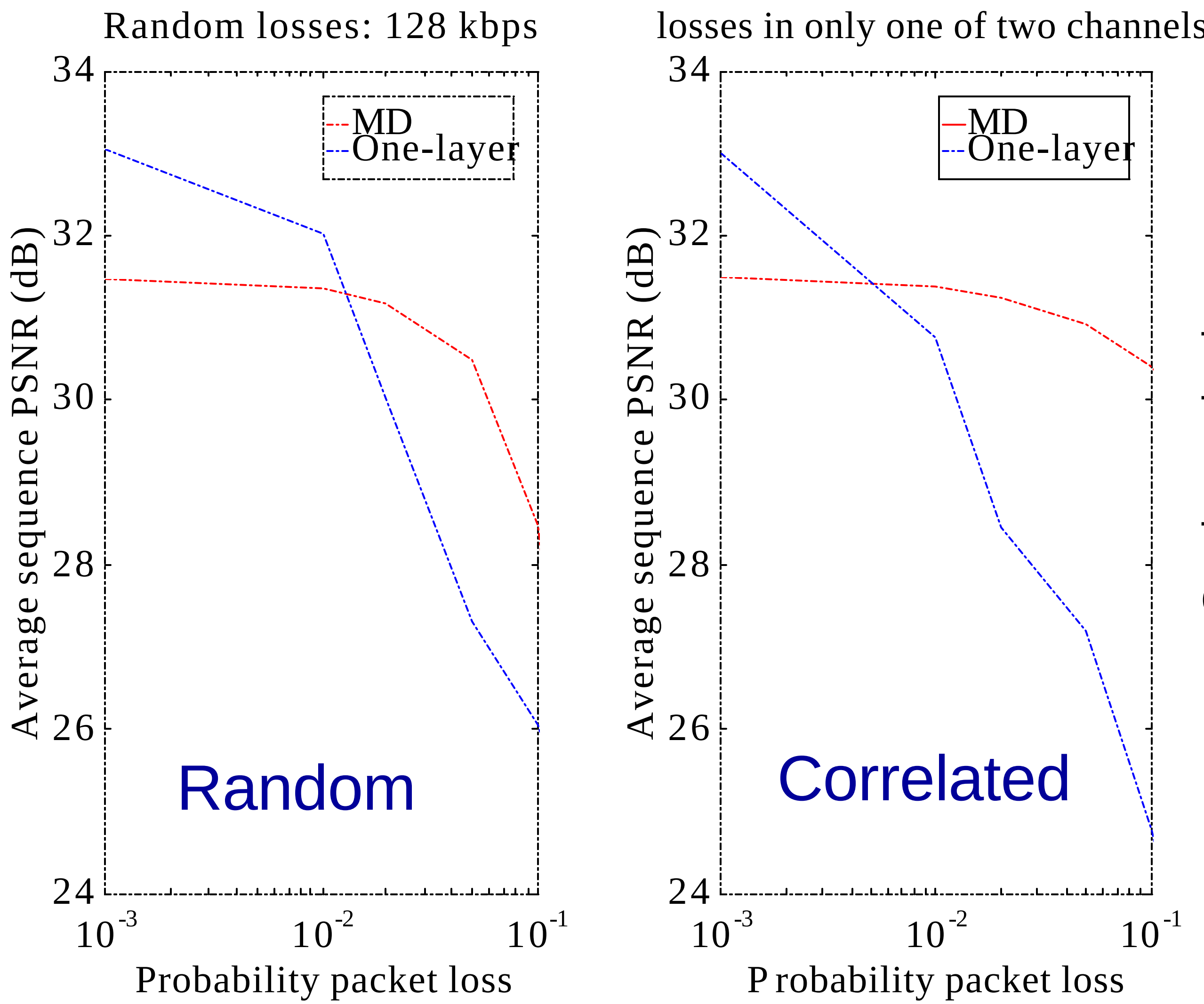
Issues for Multiple Description video

- Redundancy-Rate vs. Distortion
  - More redundancy improves quality with losses, but degrades loss-free quality
- How well do MD coders do in non-MD channels?
- Prediction structures
  - Using central-loop reconstruction introduces mismatch
  - Not using central-loop reconstruction loses efficiency
- Motion!!!
- As yet unsolved problems:
  - Achieving low (less than 40-50%) redundancy
  - More than two descriptions

Redundancy-Rate Distortion for video

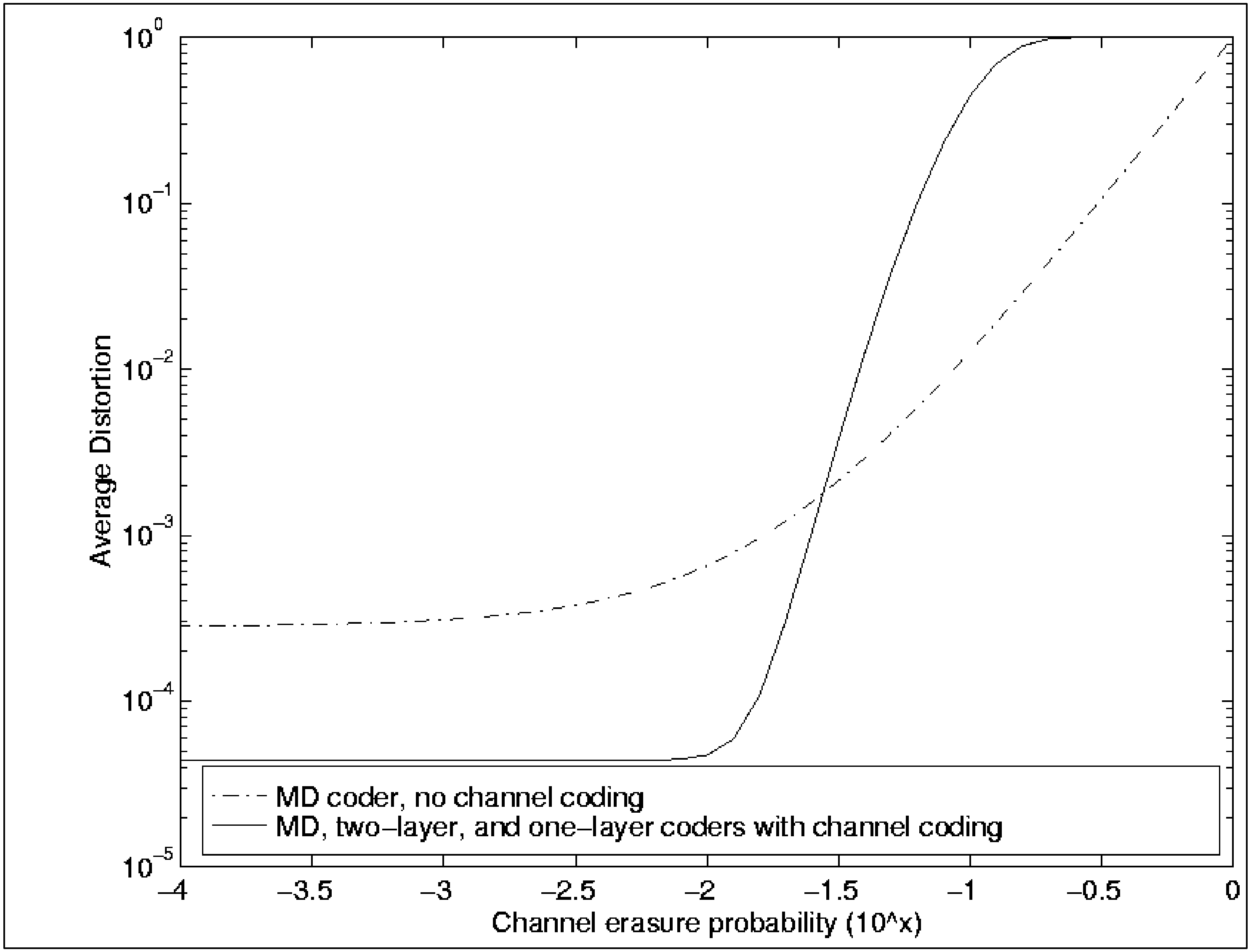


Packet loss: MD vs. one-layer



Better relative improvement for MD when the losses are more similar to ideal MD channels

Sensitivity of design to LR



All systems designed to be optimal with 10^{-2} loss rate, operating in range of loss rates

MD without channel coding is less sensitive to designed LR

MD video demo

- **10% packet loss rate**
- **MD vs. Single Description (SD)**
- **SD uses Intra-update, 10% of blocks Intra**

Conclusions

- **Representation matters for network transport!**
- **One-layer video**
  - **Best choice if exact knowledge of BW, LR small**
- **Multiple description video coding**
  - **Advantageous if lack of knowledge about current LR**
  - **More robust to 3-20% random losses**
  - **Cannot help with lack of knowledge in BW**
- **Scalable video coding**
  - **Less sensitive to imprecision in BW (rate adaptation)**
  - **Useful if known LR (unequal error protection)**

Acknowledgements

- **System level architecture**
  - **Reza Rejaie, AT&T**
- **Scalable video coding**
  - **Leon Bottou, AT&T**
- **Multiple description video coding**
  - **Hamid Jafarkhani, Michael Orchard, Yao Wang**