H.264/AVC and Its Extensions:

How Close is this Family?

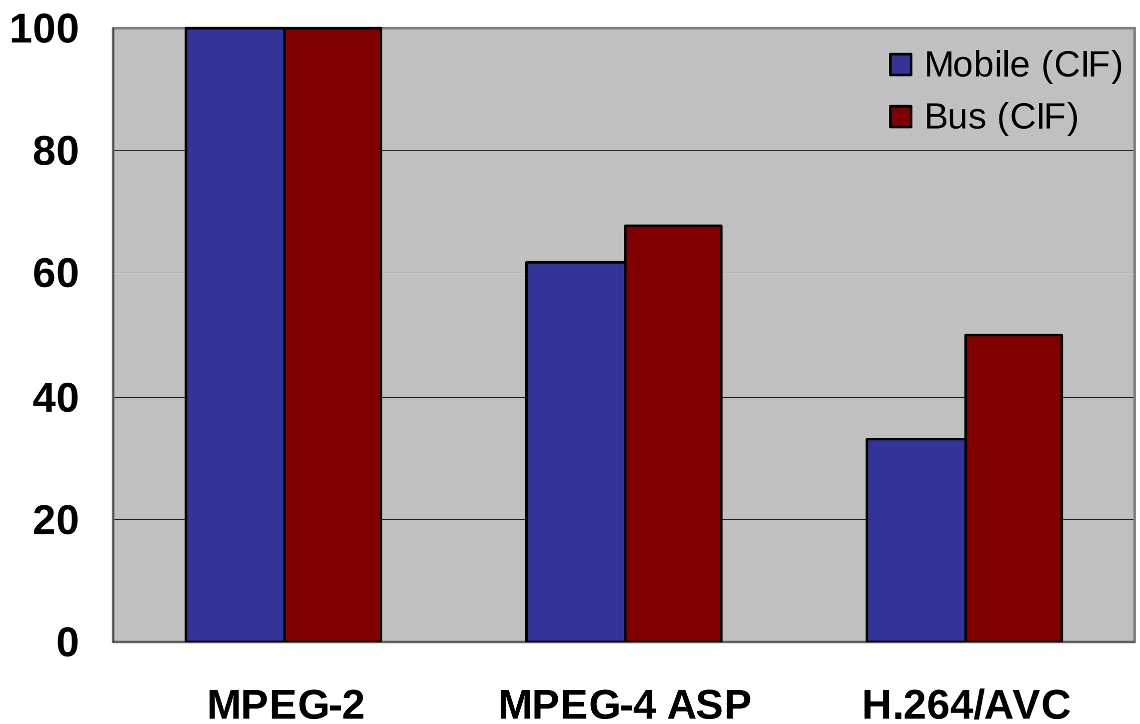
Anthony Vetro
Mitsubishi Electric Research Labs
Cambridge, USA

(avetro@merl.com)

Coding Efficiency Is King

- H.264/AVC satisfies a fundamental need in digital video systems

High coding efficiency



Percentage bit-rate relative to MPEG-2 (at 32 dB)

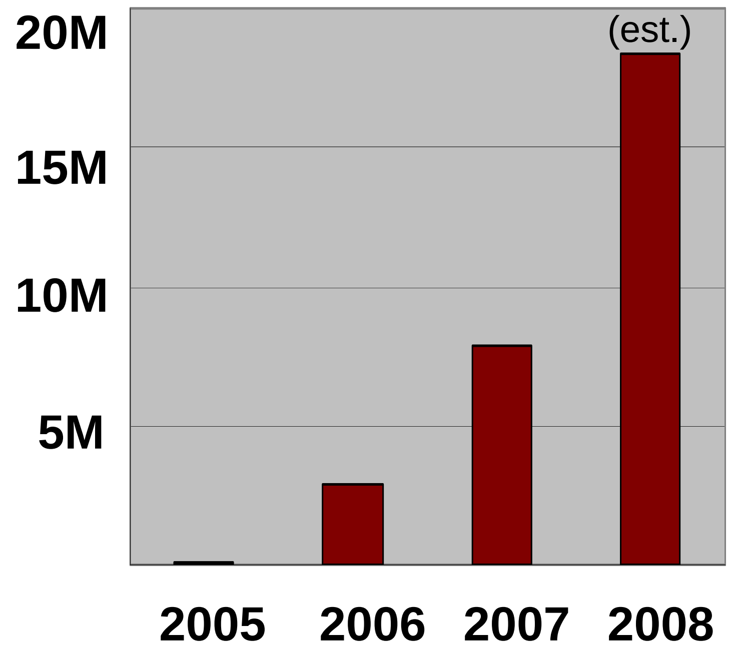
Additional Design Features

- Network friendliness
  - Enabled through NAL design
- Error resilience
  - Data partitioning, FMO, slices, redundant slices, resynchronization markers, multiple reference pictures, parameter sets, etc.
- Temporal scalability
  - Enabled by flexible reference picture management via hierarchical predictions

Commercial Deployments

- Optical Storage
  - HD-DVD, Blu-Ray Disc
- Television Broadcast
  - Cable, Satellite, Terrestrial
  - Telco IPTV
  - Mobile TV
- Streaming
- Video Conferencing
- Portable Media Players
- Video Game Consoles

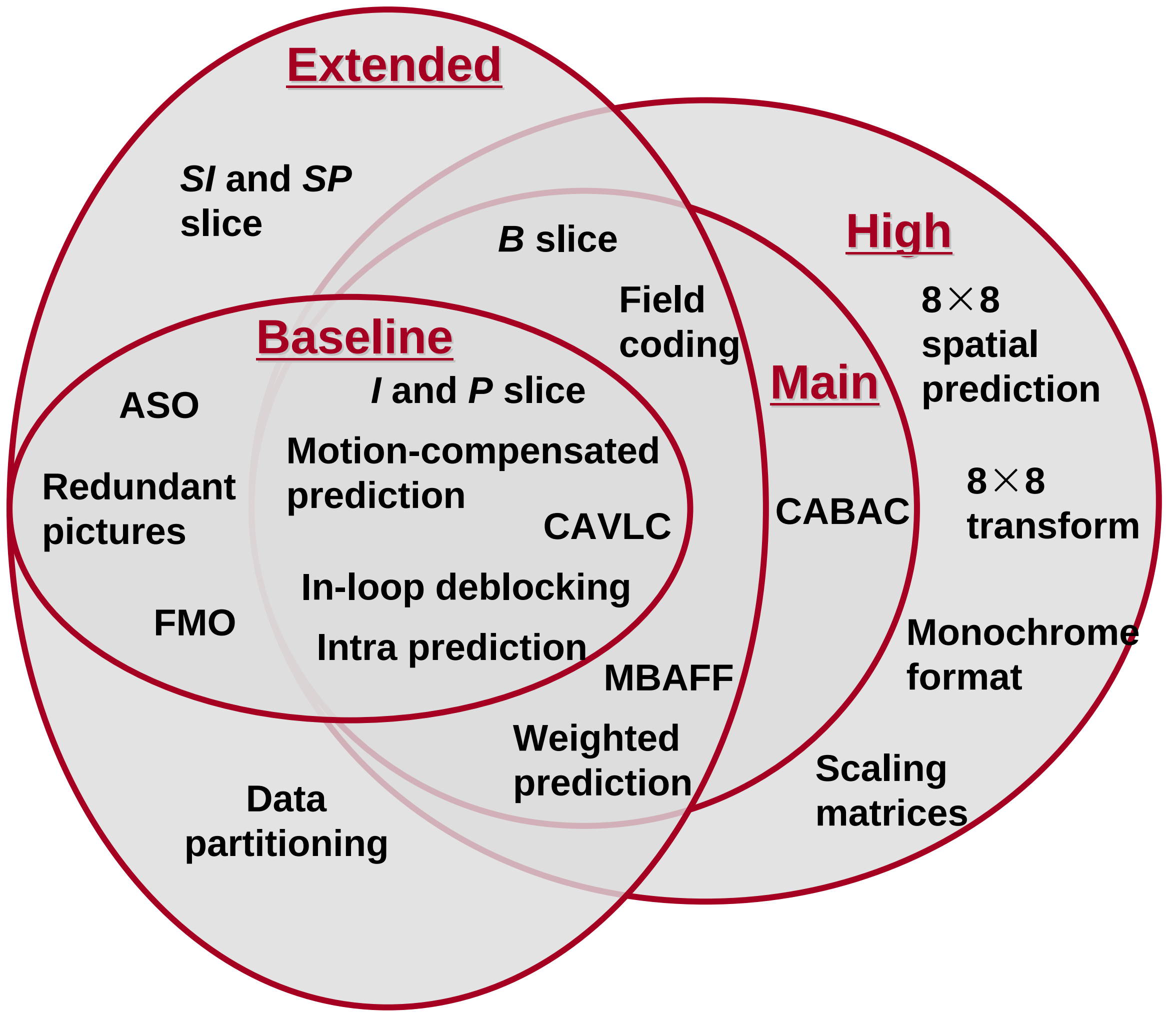
Number of One-Seg Devices (Japan)



[Source: Yano Research]

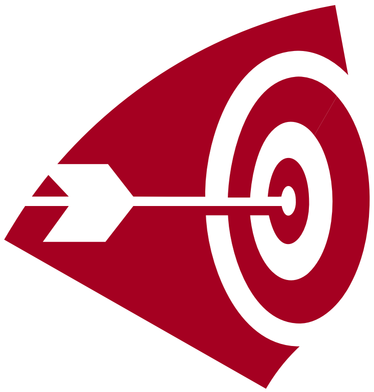
AVC Tools & Profiles

*** 3rd Edition, 2005**



What Role do Extensions Play?

- Extensions broaden the functionality and capabilities of the base spec
  - Wide range of potential
  - Targeted application domains
  - Enable new services



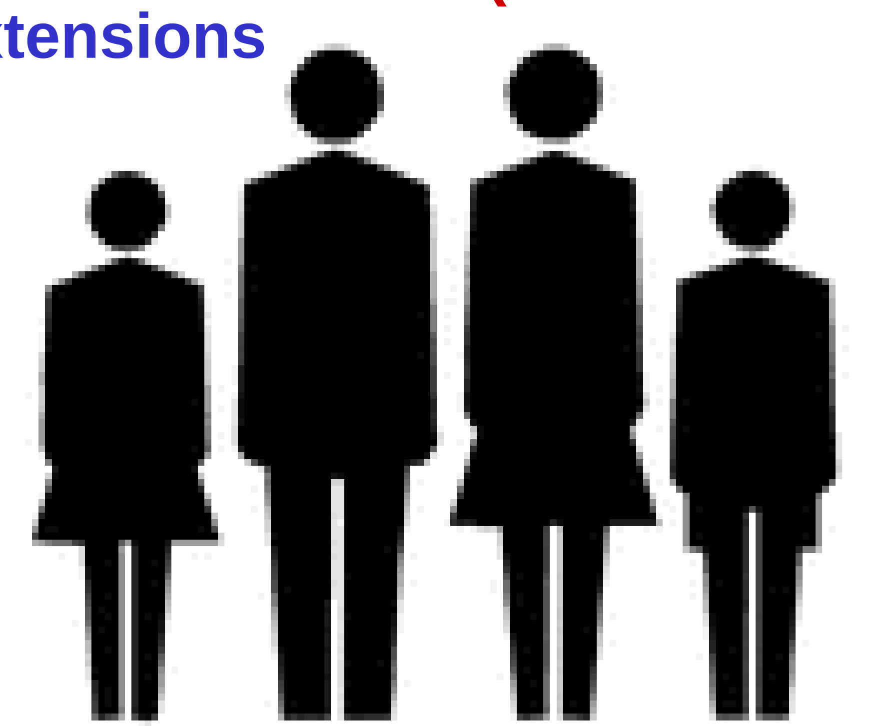
- Obstacles
  - Complexity, performance
  - System constraints, business factors
  - Sometimes difficult to get traction

    (need extensive follow up in industry consortia)

The Current Family

Professional extensions

Base H.264/AVC (the mother spec)



Scalable video coding (SVC)

Multiview video coding (MVC)

Objectives This Morning

- Cursory overview of H.264/AVC extensions
  – Professional, Scalable, Multiview

- Analysis of coding tools, concepts and performance across various extensions

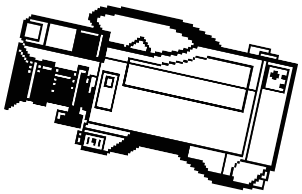
- Do these extensions have what it takes to succeed? What other factors exist?

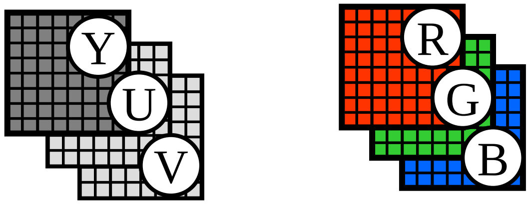
- What’s on the horizon…

Professional Extensions

Professional Quality Video

- Key applications
  - Professional studio
  - Digital cinema
  - Super HD imaging systems
  - Medical applications





- Technical requirements
  - High coding efficiency (HD bit rates > 150Mbps)
  - High dynamic range (typically up to 14 bits/pixel)
    - Mainly to keep the signal fidelity during various encoding/decoding iterations
  - Random access (essential for content editing)
  - Low complexity (algorithm and implementation)
  - Cover 4:2:2 and 4:4:4 color sampling structures
    - 4:4:4 can represent different color spaces (RGB, YUV, ...)

Common & Independent Modes

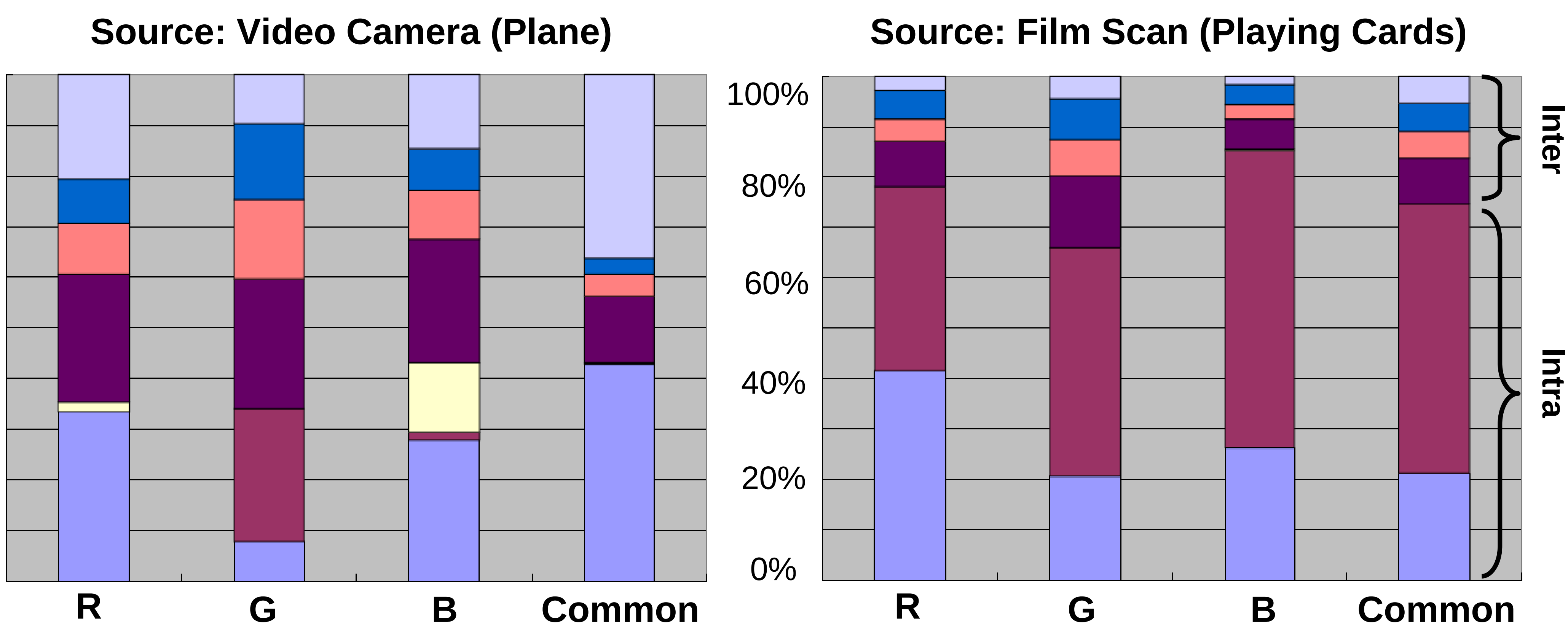
Common Mode

- Use existing coding architecture and tools
  - Single set of MB-level coding parameters, e.g., MB types, prediction modes, MVs, etc.
  - Minimizes syntax changes and algorithm complexity
- Residual signals for each color plane coded using same “luma” coding

Independent Mode

- Separate, independent coding of each color plane
  - Each color plane coded with monochrome coding tools
  - Enhanced parallel processing capability
- Adaptive MB-level coding parameters for each color
  - Flexible prediction by adapting to local signal statistics of each color

Analysis of Coding Modes



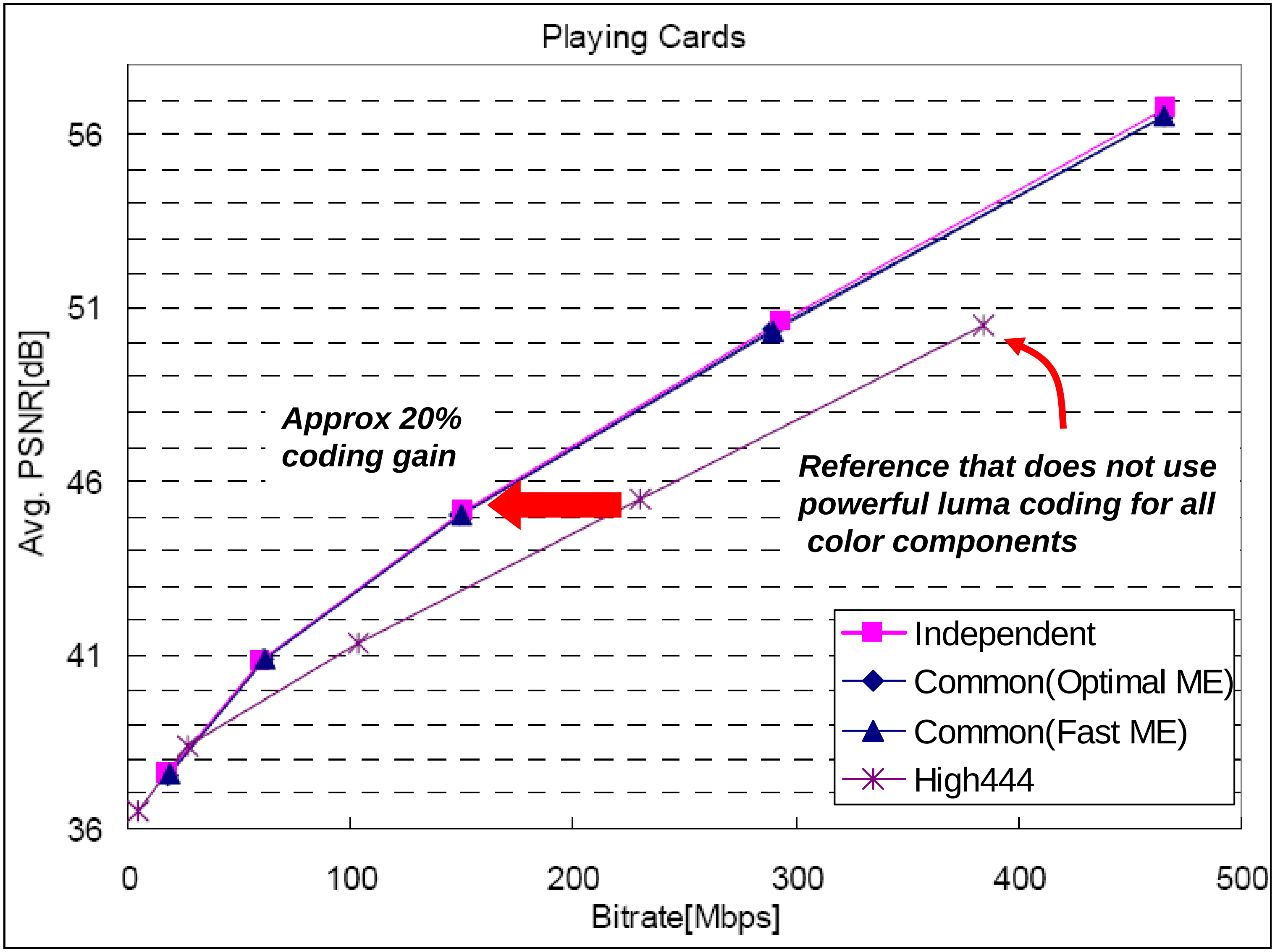
Independent mode adapts to local statistics in each color plane

Intra is dominant for film scan
Adaptivity is less critical

[Sekiguchi, et al., ICIP 2006]

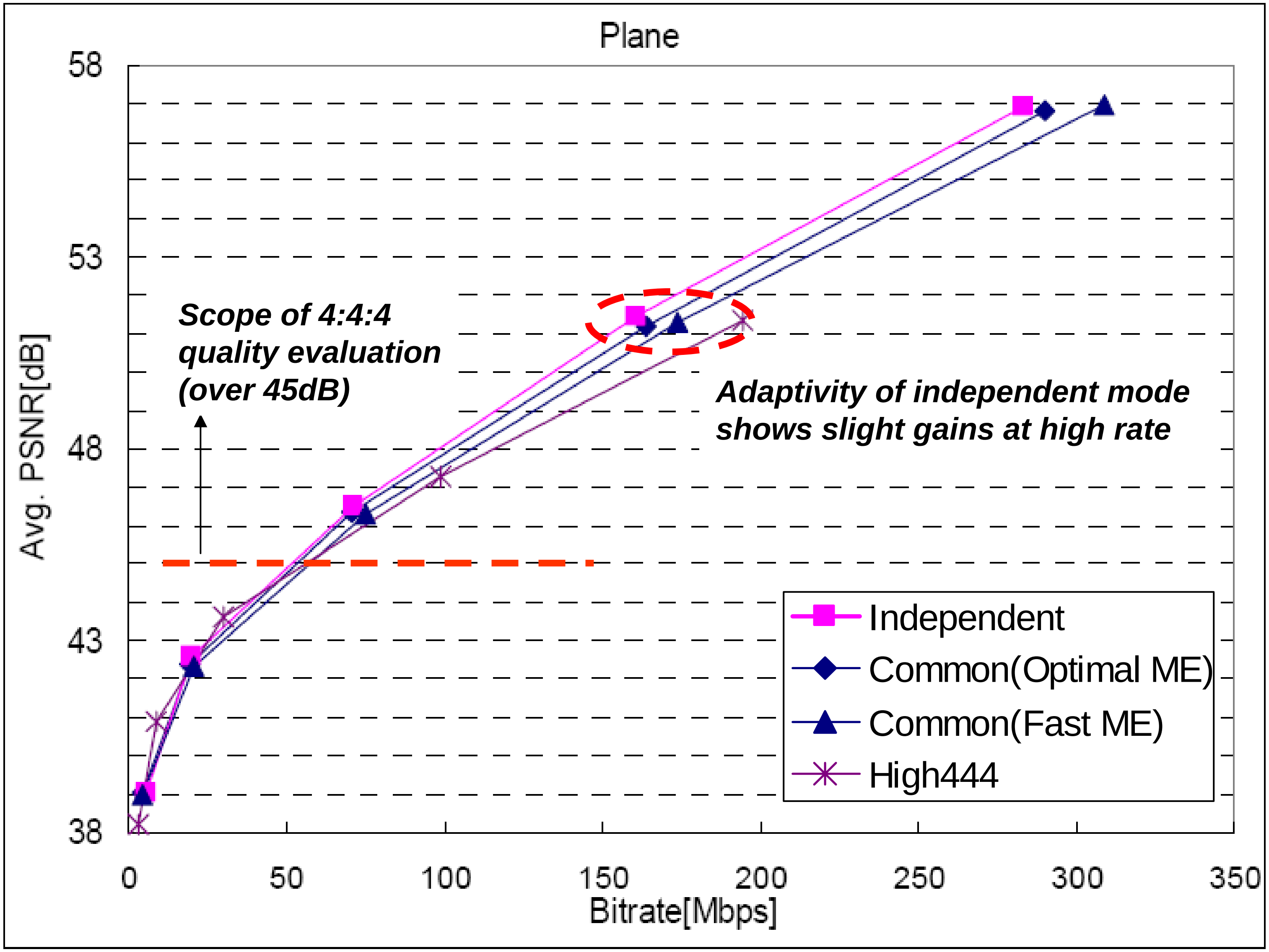
4:4:4 Coding Performance

(Film Scan)

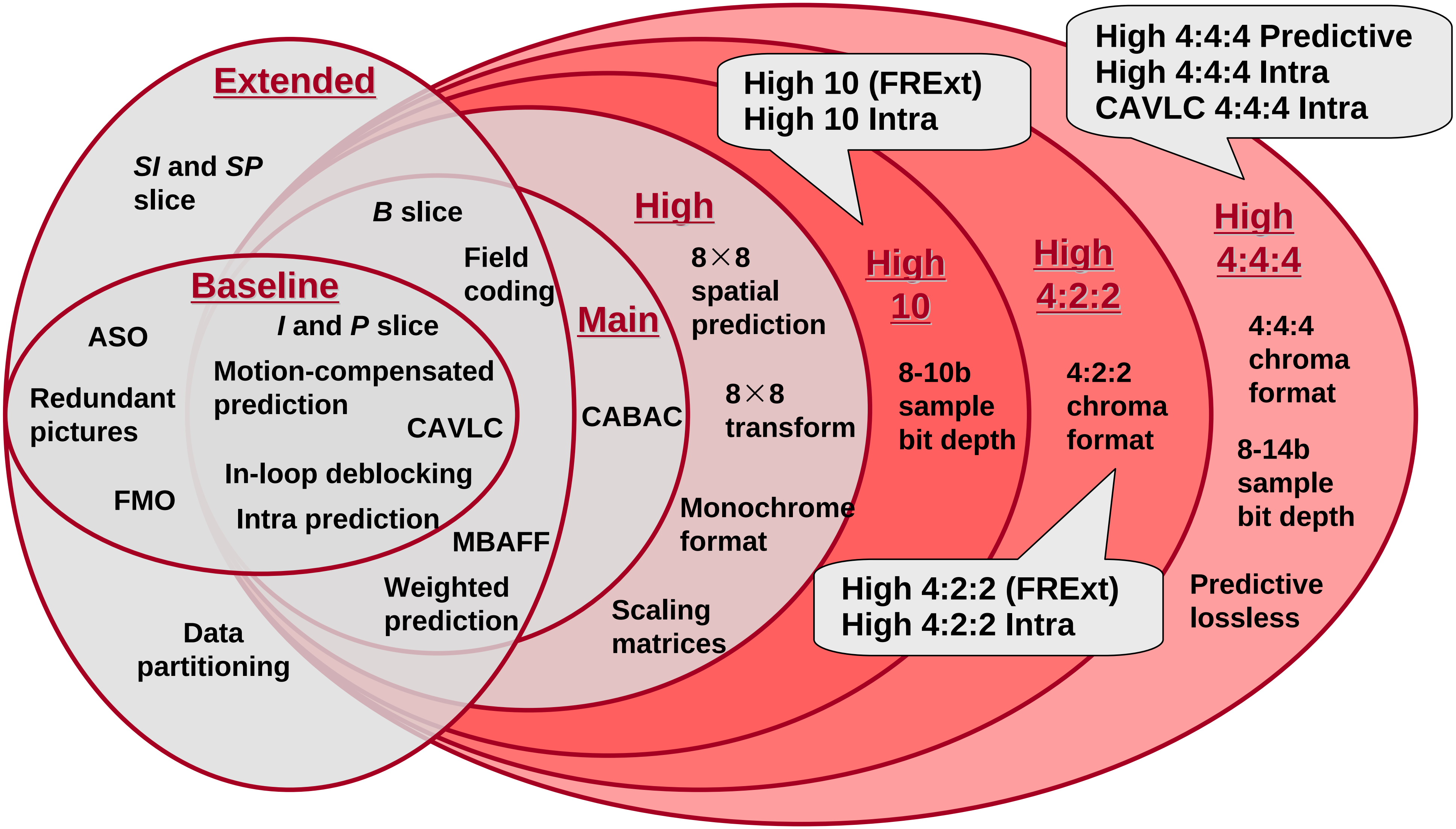


4:4:4 Coding Performance

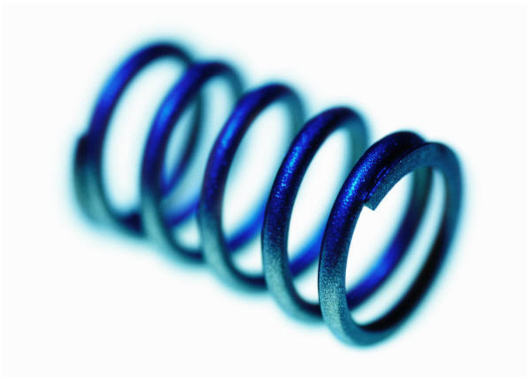
(Video Camera)



Professional Profiles

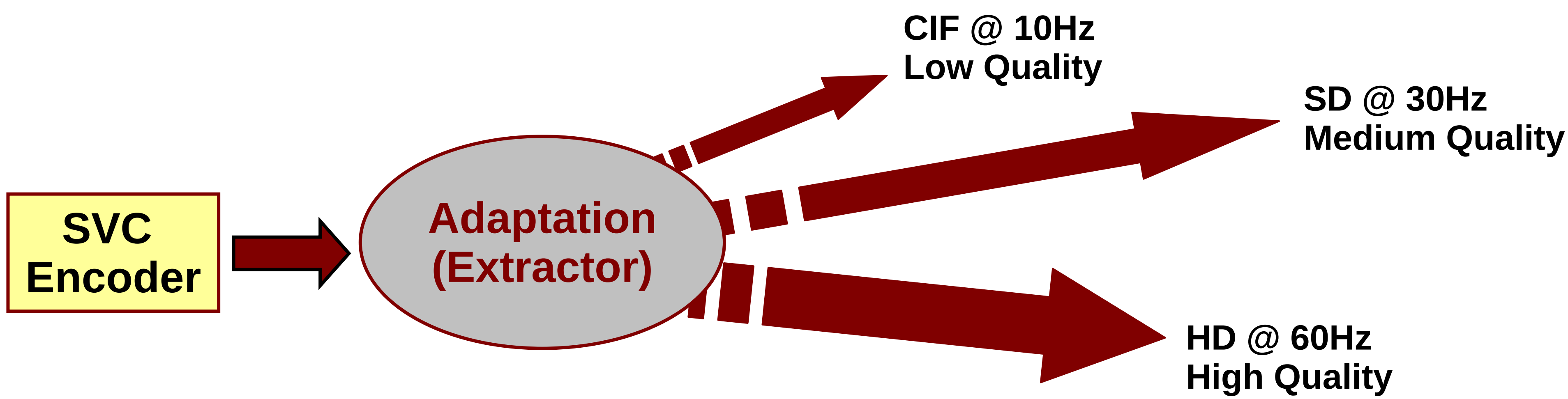


Scalable Video Coding (SVC)



Scalable Video Coding

- Traditional Dimensions of Scalability
  - Temporal, Spatial, SNR

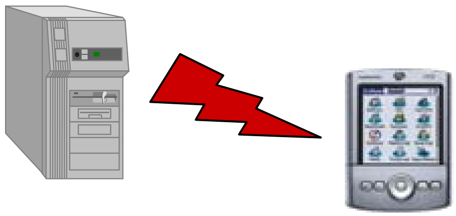


Encode once, decoding many

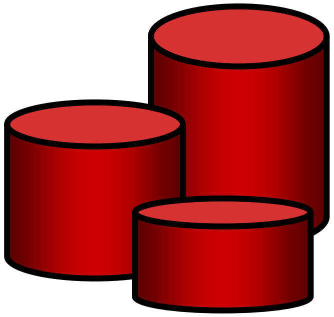
- Key technical challenges
  - <u>Coding Efficiency</u>: min loss wrt single layer coding
  - <u>Complexity</u>: solution must not incur substantial increase

SVC Applications

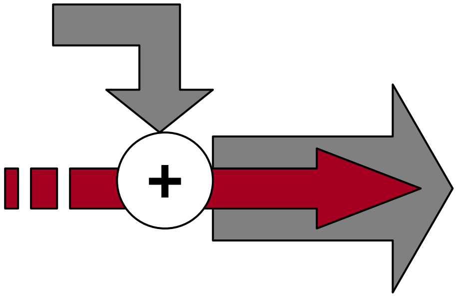
- Robust Video Delivery
  - Adaptive delivery over error-prone networks and to devices with varying capability
    - Combine with unequal error protection
    - Guarantee base layer delivery
  - Internet/mobile transmission



- Scalable Storage
  - Scalable export of video content
  - Graceful expiration or deletion
  - Surveillance DVR's and Home PVR's

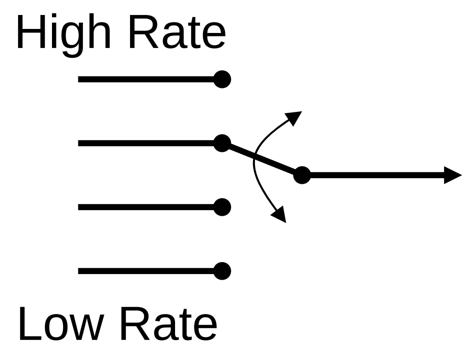


- Enhancement Services
  - Upgrade delivery from 1080i/720p to 1080p
  - DTV broadcasting, optical storage devices

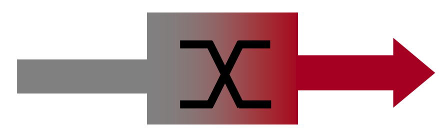


SVC Competition

- Simulcast
  - Simplest solution
  - Code each layer as an independent stream
  - Incurs increase of rate



- Stream Switching
  - Viable for some application scenarios
  - Lacks flexibility within the network
  - Requires more storage/complexity at server



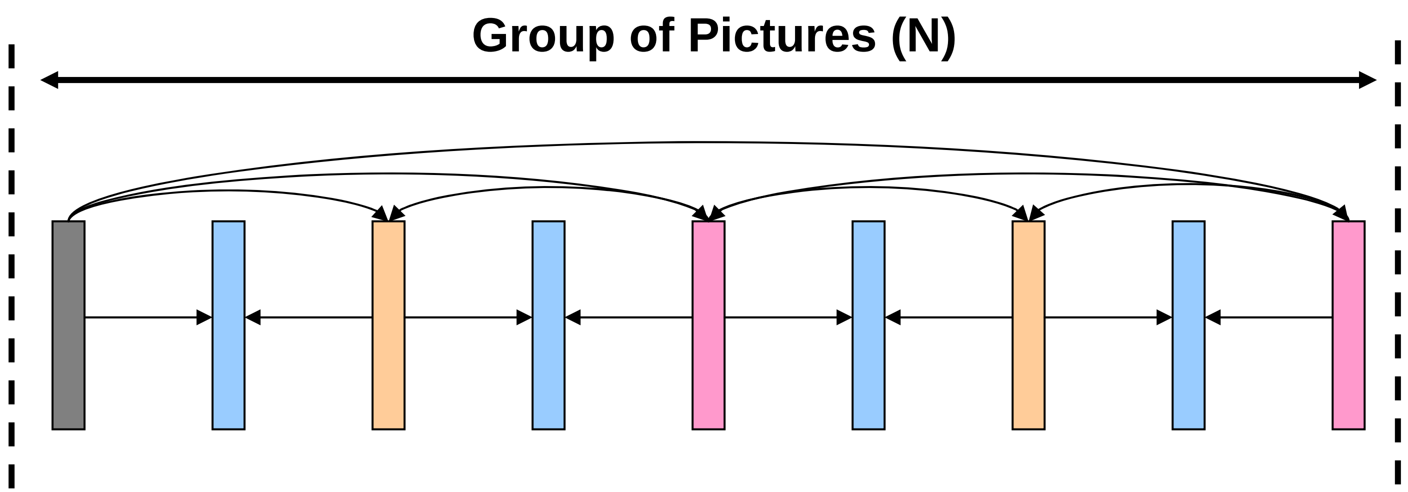
- Transcoding
  - Low cost, designed for specific application needs
  - Already deployed in many application domains

Relevant SVC Technologies

- Wavelet-based
  - Typically, a critically sampled decomposition (but not always)
  - Filters designed for perfect reconstruction, high coding efficiency
  - Aliasing often included in decimated image - problems for video:
    - Motion compensated prediction could become more difficult
    - Temporal artifacts may result from truncation of higher subbands
  - Solution for JPEG 2000 and adopted for Digital Cinema applications

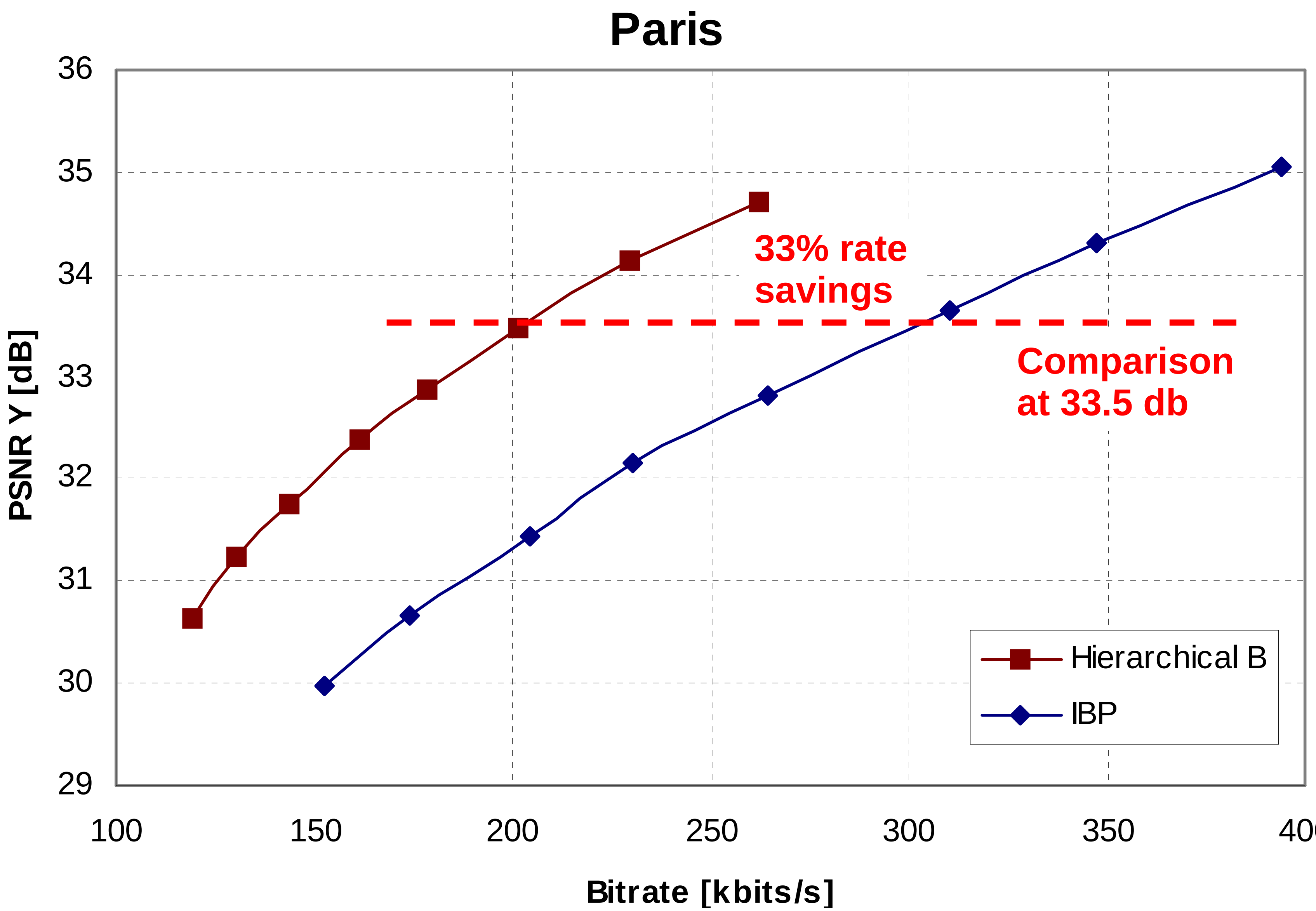
- Image pyramid
  - An oversampled representation, amenable to the traditional block-based coding schemes
  - Down-sampling filters designed for high visual quality of decimated pictures
  - Scalable extension of H.264/AVC employs layered block-based (image pyramid) approach

Temporal Scalability



- Achieved using hierarchical prediction structures
  - Possible due to flexible reference picture management in H.264/AVC design
  - Many configurations possible with varying delay, picture buffer requirements, non-dyadic structures
- No loss in coding efficiency
  - Hierarchical prediction structures improve coding performance!
  - Quantizer selection for each temporal layer is important

Improved Coding with Hierarchical B



Visual Comparison at 33.5 dB

Hierarchical B (N=16)

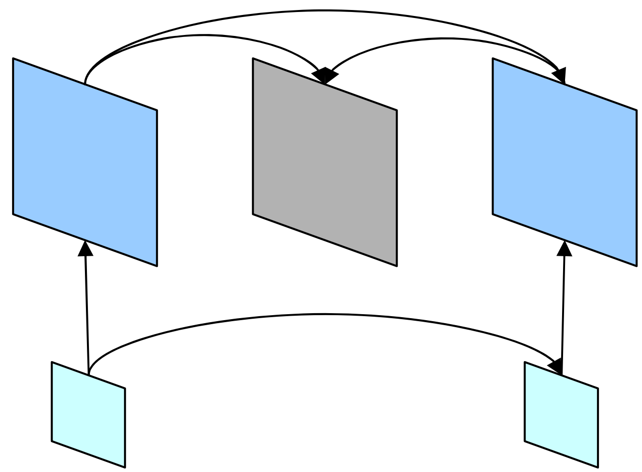
201 kbps

Conventional IBP

310 kbps

Spatial Scalability

- Similar approach as in past standards
  – Multi-layered coding of each spatial scale



- Each spatial layer
  – MC prediction
  – Inter-layer prediction
- Combined spatial/temporal scalability
  – Possible since lower layer pictures need not be present in every access unit (time instant)

Spatial Scalability

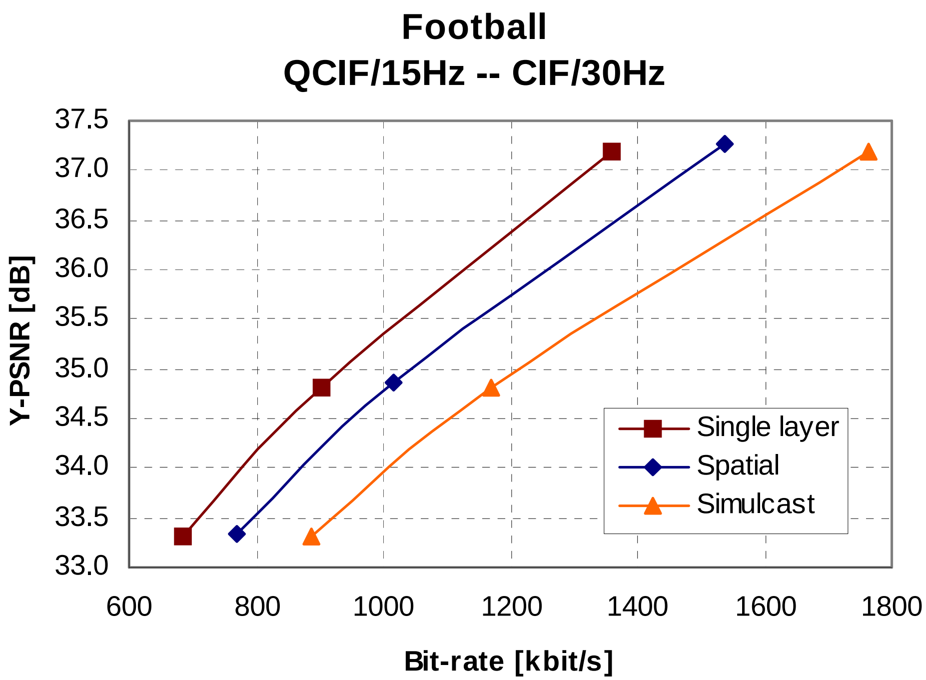
- Options for inter-layer prediction
  - Prediction of intra MBs (upsample intra MBs in ref layer)
  - Inter-layer motion prediction (infer MB info from ref layer)
  - Residual prediction (upsample residual from ref layer)

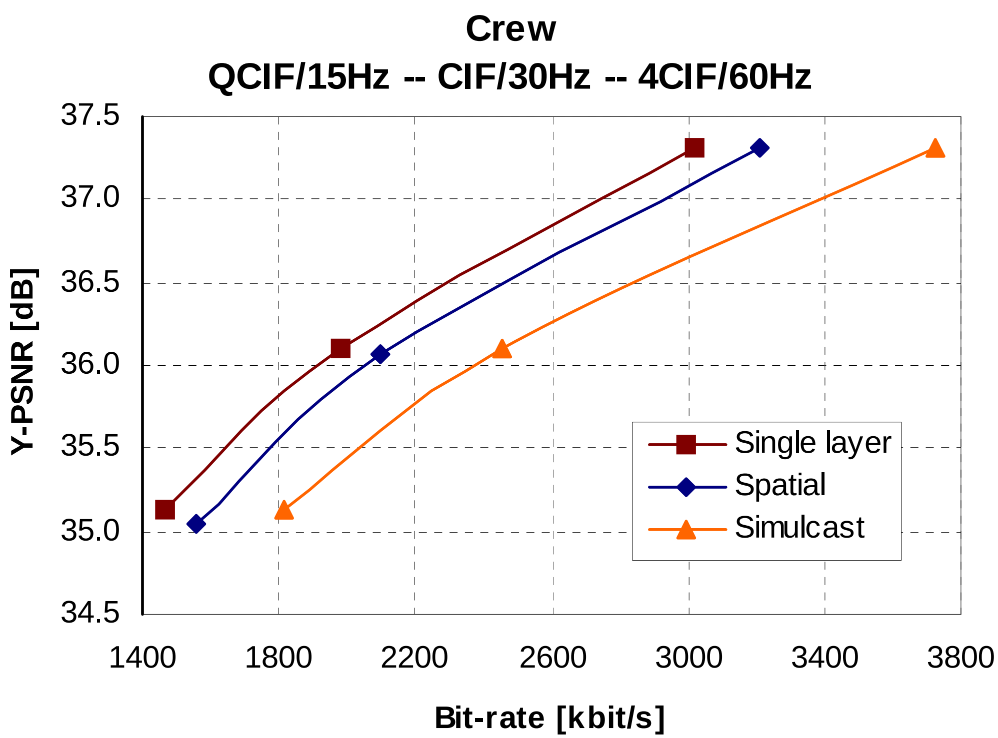
- Single-loop decoding
  - A central concept in SVC
  - Constrained inter-layer prediction
    - Do not allow inter-layer prediction based on MC reconstruction in reference layer
    - MC only needed for MBs at target decode/display resolution (hence only need to store pictures from target layer)
    - Substantial reduction in complexity compared to multi-loop decoding (as required in past standards)
    - Relatively small increase in decoder complexity over single layer decoding

Coding Performance

Sample Dyadic Test Results for Spatial Scalability

[Segall & Sullivan, T-CSVT, Sept'07]





- 10~15% gains over simulcast
- Performs within 10% of single layer coding

SNR Scalability

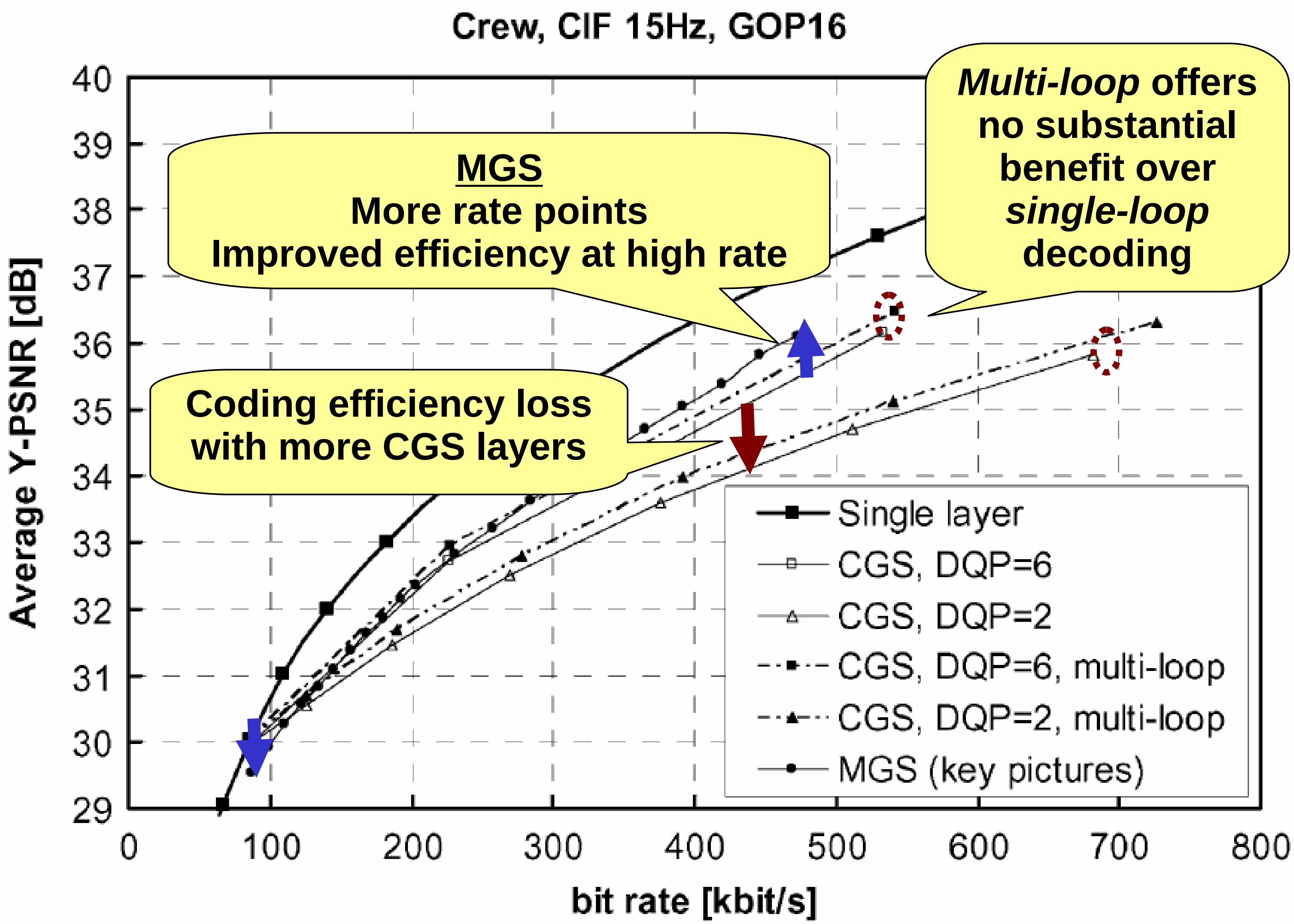
CGS (coarse grain)

- Discrete set of rate points for each layer
  - Same as spatial scalability, but w/out upsampling
  - Use inter-layer prediction
- Texture refinement
  - Requantization of residual
  - Finer quantization step with increasing layer
- Suitable for select number of rate points
  - Coding efficiency reduces with too many CGS layers

MGS (medium grain)

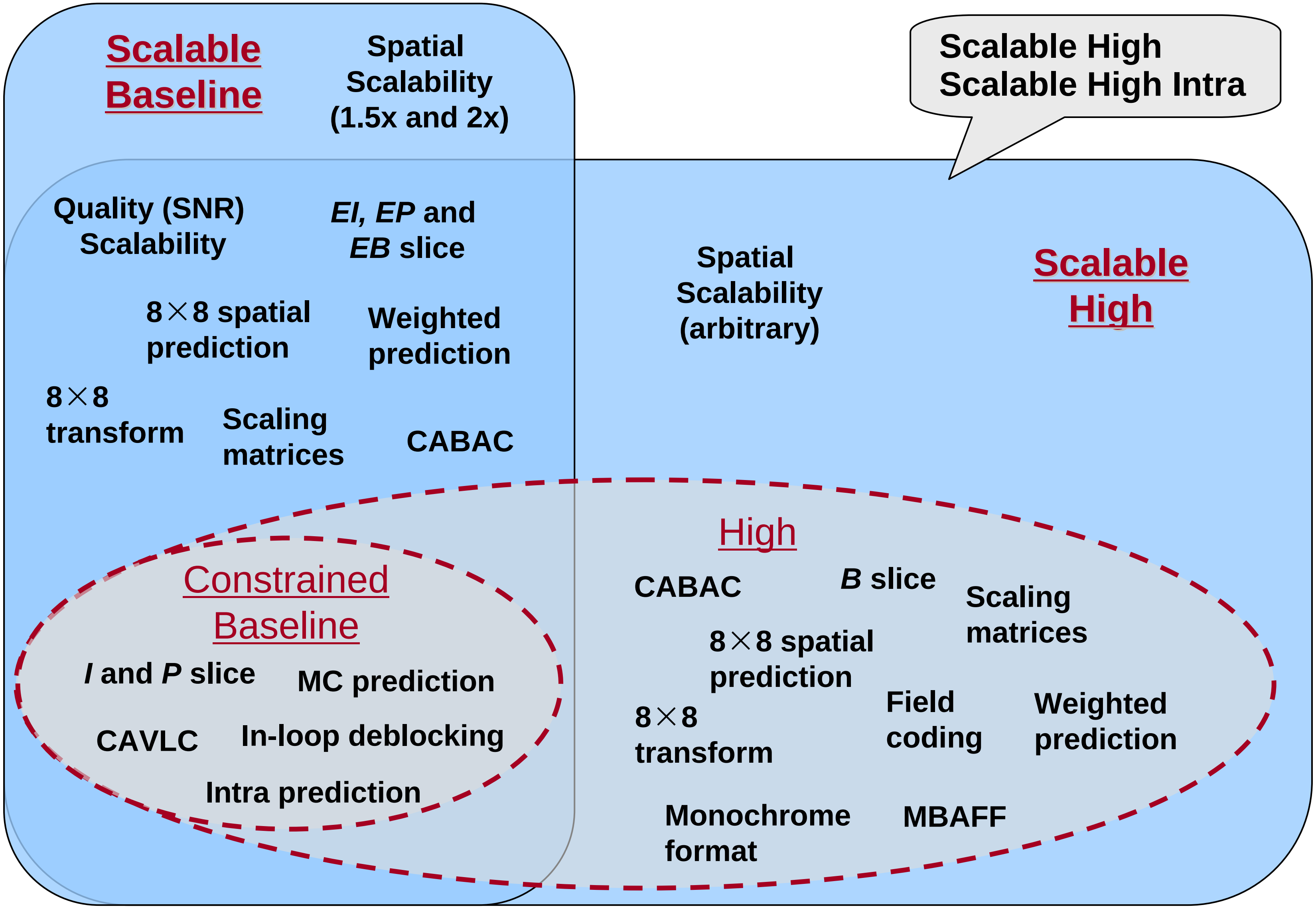
- Successive refinement of quality within a layer
  - Code fragments of transform coefficients
  - Allows graceful degradation
- Switching between layers
  - Possible in any access unit
  - Predict from base: control drift due to packet loss
  - Predict from enhancement: improve coding efficiency
  - Key picture concept to signal whether base or enhancement used for prediction

CGS/MGS Coding Performance

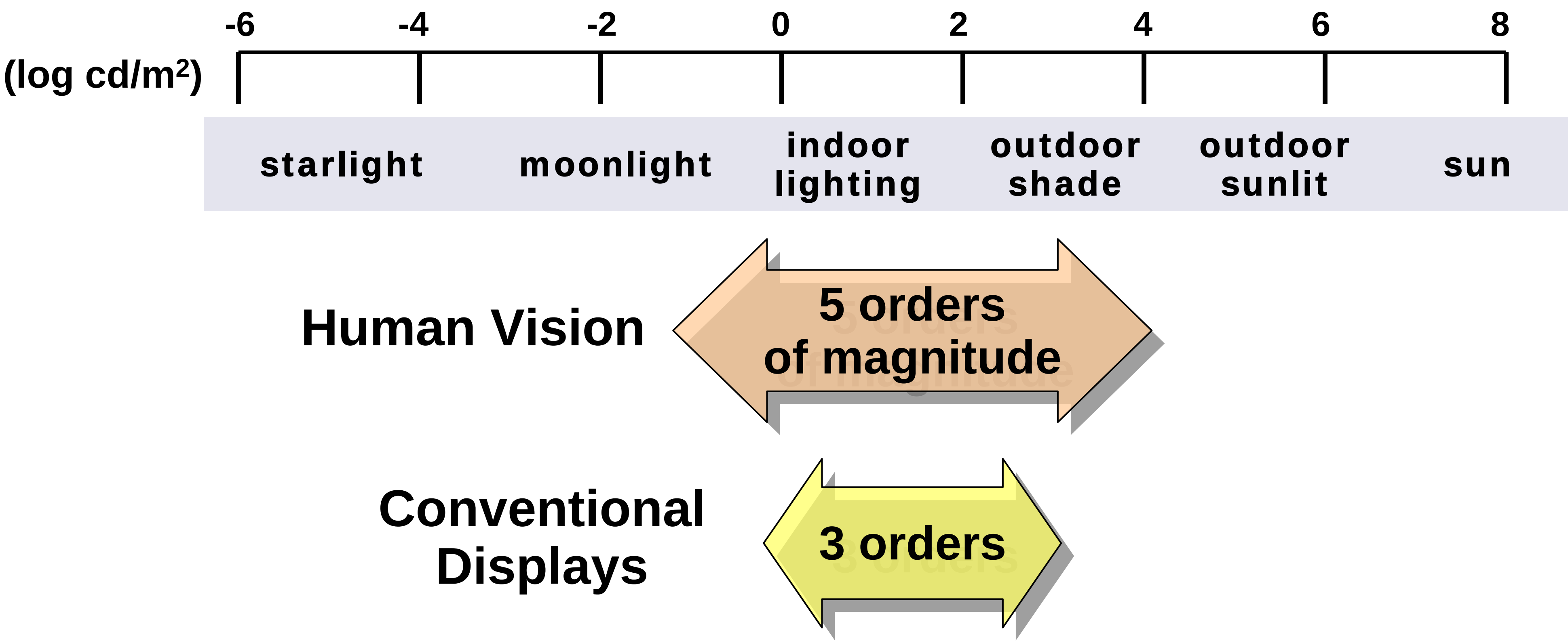


[Schwarz, et al. T-CSVT, Sept'07]

SVC Profiles



New Scalability Dimension



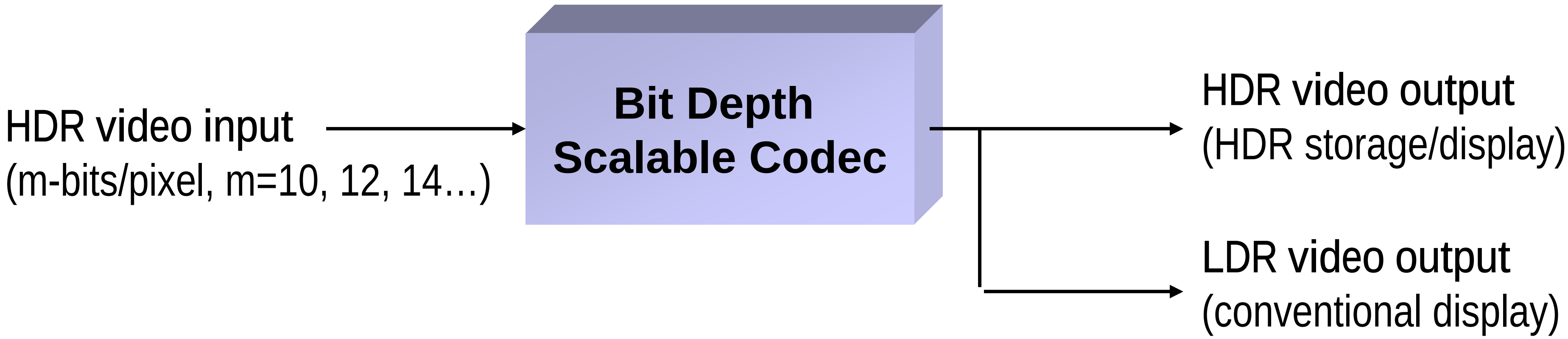
High Dynamic Range Scenes + Mixed Display Environment → **Bit Depth Scalable Compression**

Potential Applications

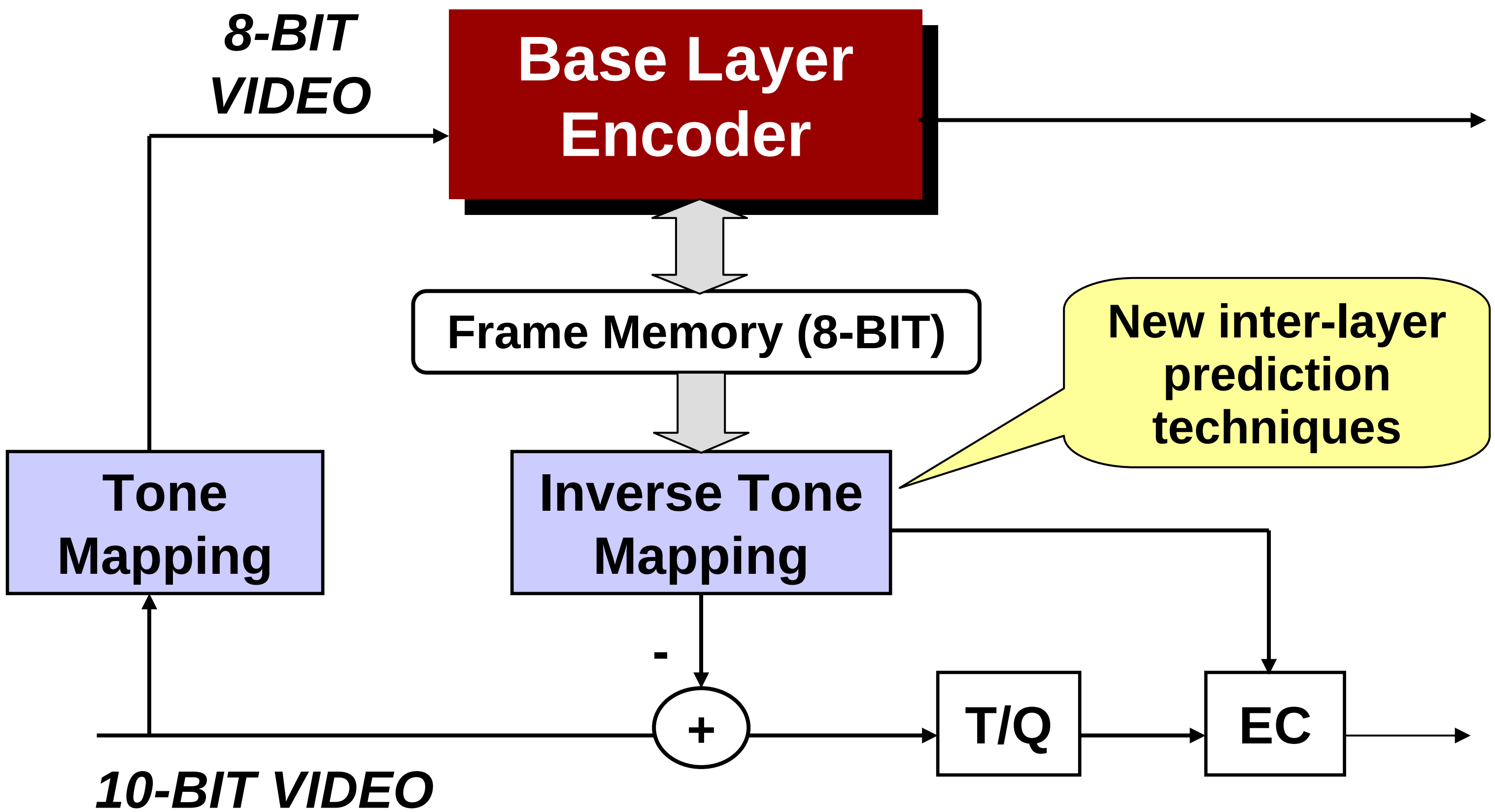
- Next Generation Optical Disc
  - Consumer displays will become more capable, e.g., 10 bits/pixel output
  - HD-DVD and Blu-Ray may consider supporting both legacy displays as well as advanced displays
- Post-Production
  - Already use 10/12/14 bit video
  - Provide compatibility with broadcast/DVD stream
- Other (Medical, Satellite, etc.)
  - Most content is captured using 16 bits sensors
  - Need efficient way to store and process HDR video, but display conventional 8-bit video

Technical Objectives

- High compression efficiency for video with bit-depths greater than 8 bits/pixel
- Scalable compression architecture
  - Retains HDR for storage, processing and advanced display
  - Enables simple access to 8-bit format for conventional display



Reference Architecture



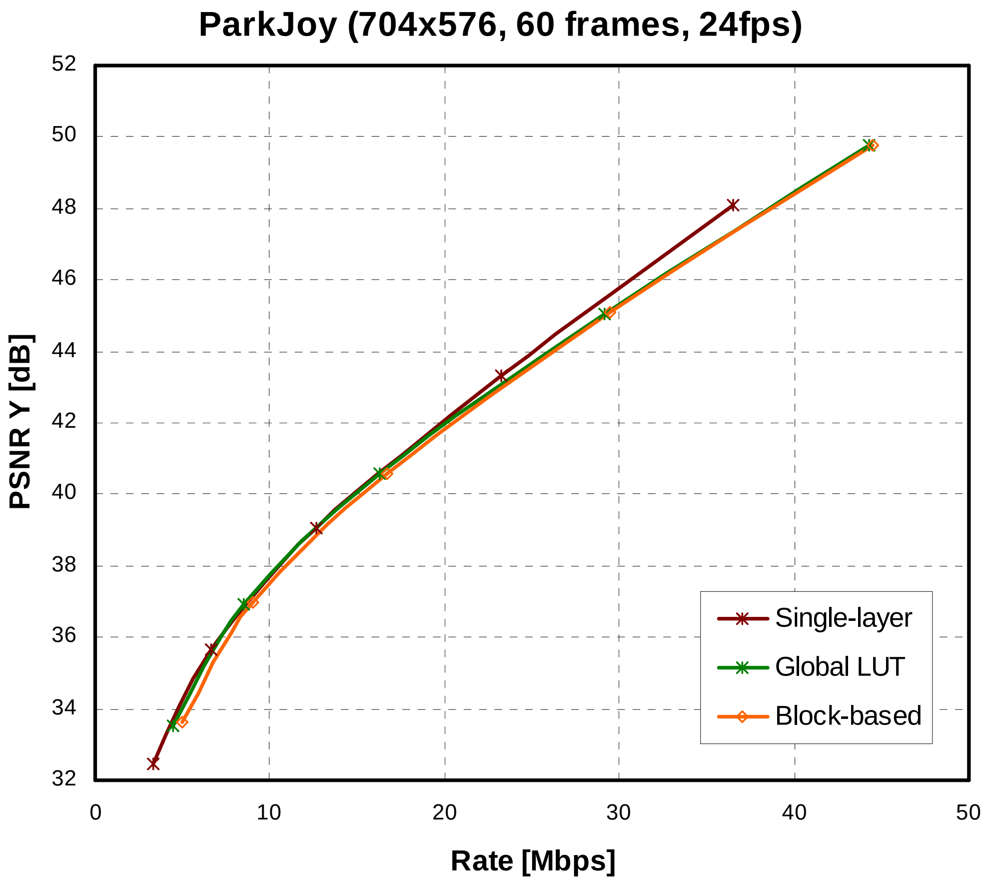
Motion compensation in base layer only

Technical Issues

- Architecture
  - E.g., determine if motion compensated prediction in enhancement layer useful - follow CGS architecture
- Combined scalability
  - Efficiently achieving bit-depth and spatial scalability
  - A three-layer scalable hierarchy is one possibility

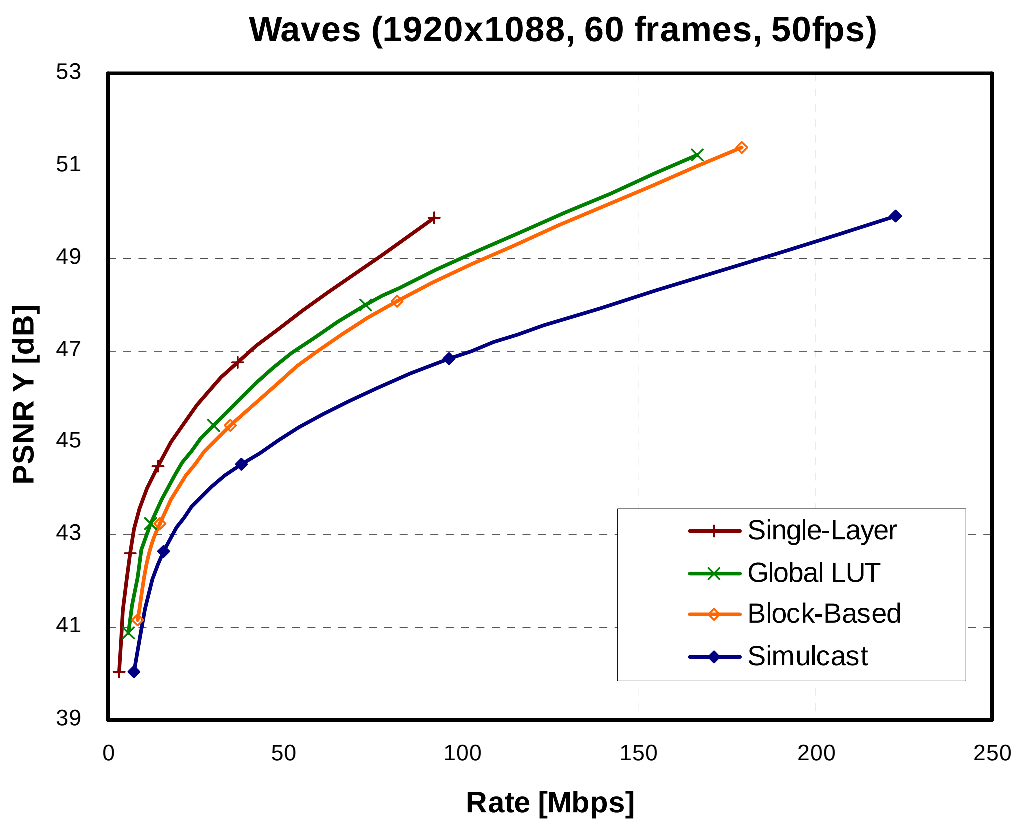
    (base → spatial enhancement → bit-depth enhancement)
- Inter-layer prediction
  - Inverse tone mapping: predict original 10-bit video pictures from 8-bit tone mapped pictures
  - Global vs local operators

Coding Performance (10/8-bit)



Linear tone mapped input sequence

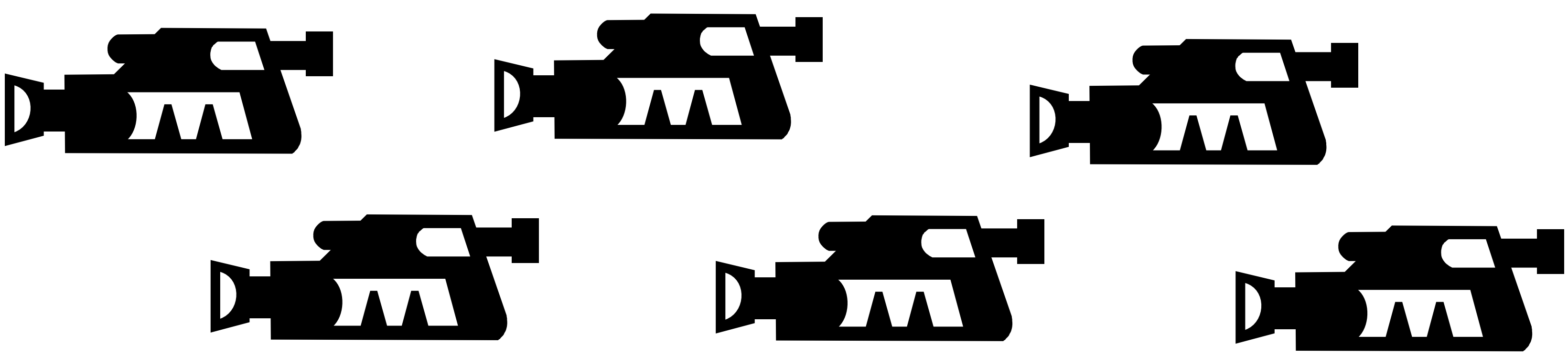
Non-linear tone mapped input sequence



Remarks on Bit-Depth Scalability

- Still very much a work in progress...
  - Beginning to look at 12 bit videos
  - Base layer sequences with spatially varying tone mappings applied
  - Applications and requirements need further clarification
- Coding Tools
  - Fits neatly within SVC framework
- Compatibility with professional profiles
  - Chroma format scalability (e.g., 4:4:4 to 4:2:0)
  - Relation with common and independent modes

Multiview Video Coding (MVC)



Multiview Video

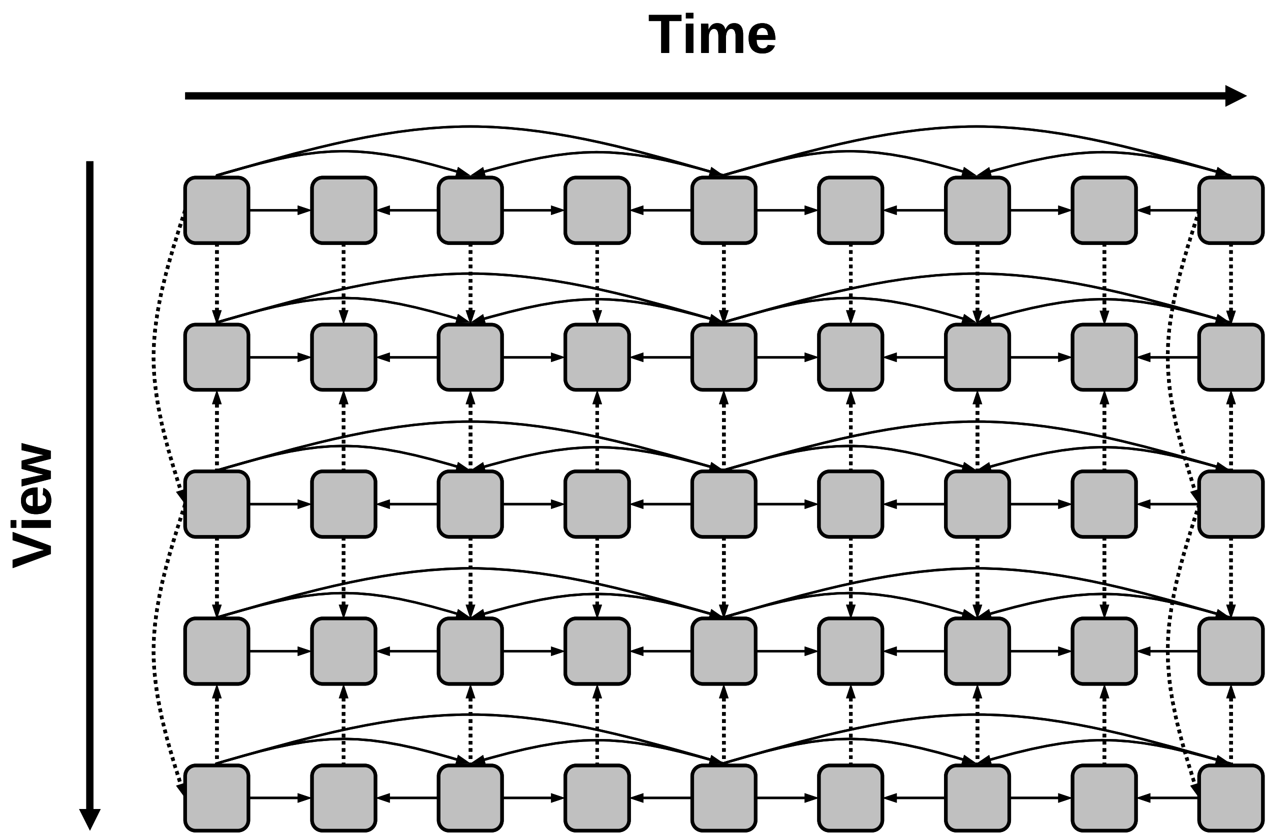
Scope of MVC
Coding of N views
(N in / N out)

- Key Applications
  - Free Viewpoint Video
  - 3D Video

- Substantial redundancy between camera views
  - Need to cope with color/illumination mismatch problems
  - Alignment may not always be perfect either

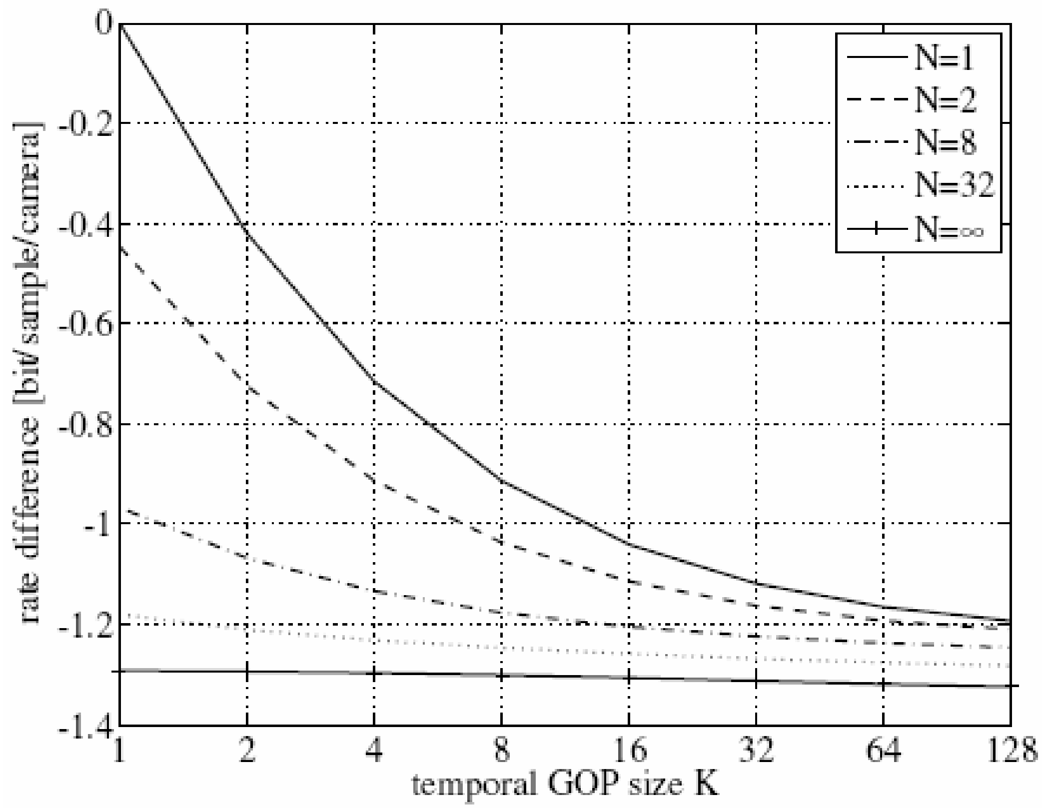
Prediction Structures

- Prediction structures exploit inter-camera redundancy
  - Trade-off in memory, delay, computation and coding efficiency



Performance Bounds & Coding Schemes

- Theoretical performance bounds on compression of multiview video have been derived
  - Function of temporal GOP size and number of cameras jointly considered by encoder
- Coding schemes to exploit inter-view similarities
  - Predictive Coding: use disparity compensated prediction
  - Subband Coding: rely on adaptive subband decomposition (e.g., disparity-compensated lifted wavelets)
- Standardization efforts focused on predictive coding scheme

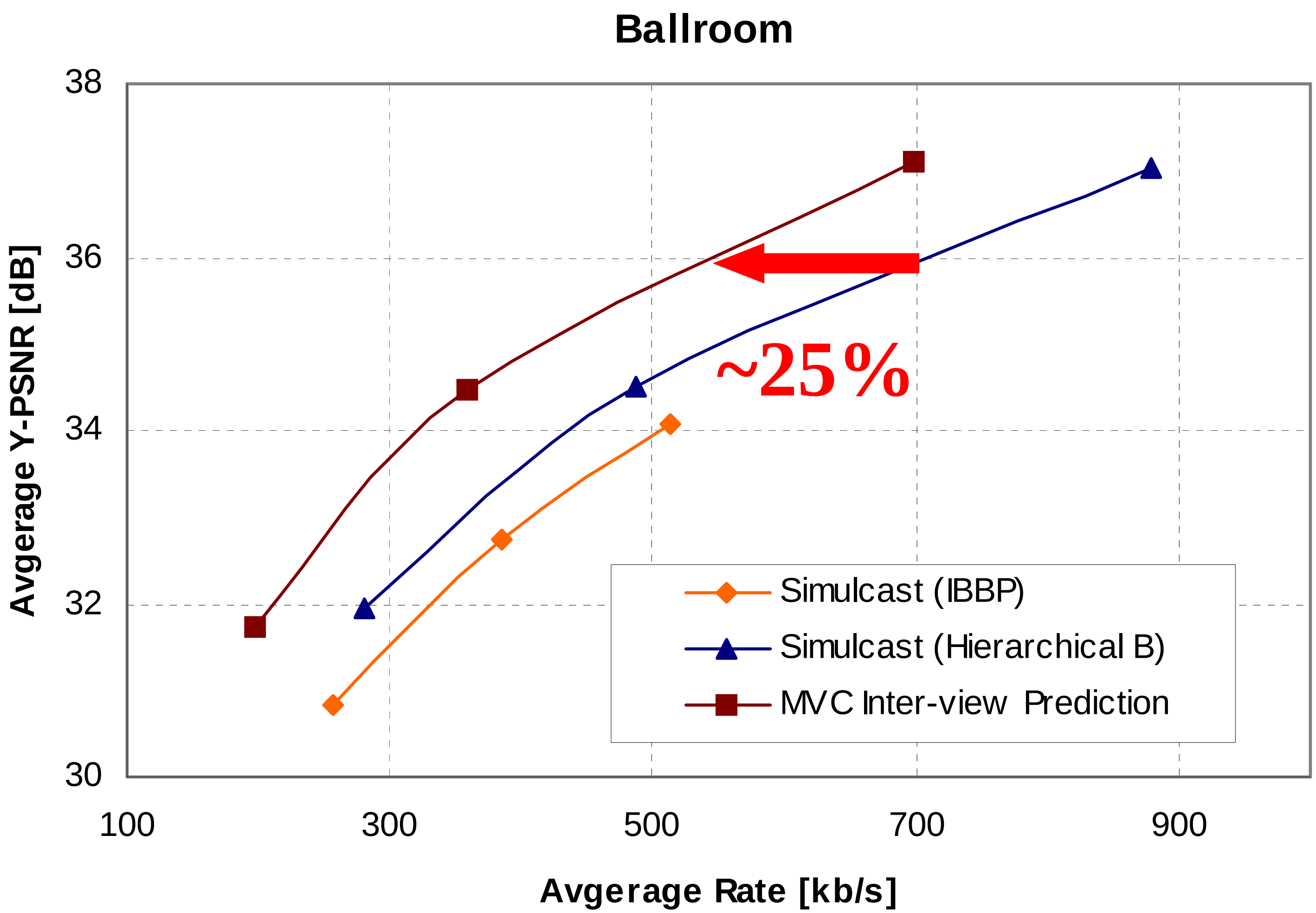


[Flierl & Girod, IEEE SPM, Nov'07]

Coding Efficiency: Multiview Extension of H.264/AVC

Sample comparison of simulcast with inter-view prediction

(majority of gains due to inter-view prediction at I-picture locations)



Compatibility of MVC Extension

- Current multiview extension of H.264/AVC does not require *any* changes to lower-level syntax
  - Very compatible with single-layer AVC hardware
- Inter-view prediction
  - Enabled through flexible design of decoded reference picture management
  - Allow decoded pictures from other views to be inserted and removed from reference picture buffer
- Small changes to high-level syntax
  - E.g., specify view dependency

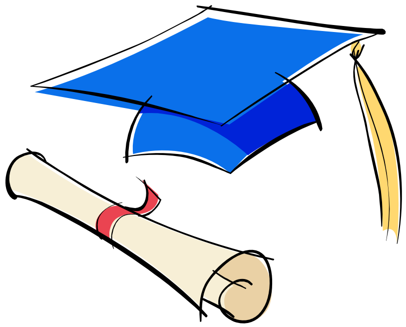
Proposals for New Coding Tools

- Illumination Compensation
  - Incorporates illumination change into MC process
- Adaptive Reference Filtering
  - Compensate for focus mismatches between views
- Motion skip mode
  - Infer motion info from corresponding block in neighboring view
  - Similar in concept to inter-layer motion prediction of SVC
  - Concept of single-loop decoding being studied
- View Synthesis Prediction
  - Generate synthesized view from neighboring views using estimated depth, then use synthesized view for prediction
- All would require changes to slice/MB level
  - Additional 10-15% gains could be expected

The Road Ahead

Definitions of Success

- Technical success
  - Were the design goals achieved?
  - Substantial advance forward over the state-of-the-art?



- Business success
  - Were products deployed?
  - New markets, large profits?

Professional Extensions

- History of use in professional domains (broadcast studio)
  - AVC products employing 4:2:2 profiles already available
  - 4:4:4 products will be available and deployed within a few years
- Trend towards consumer applications for 4:4:4 (near future)
  - High quality video for Blu-Ray and HD-DVD
  - Future DTV broadcasting services

What is happening with SVC?

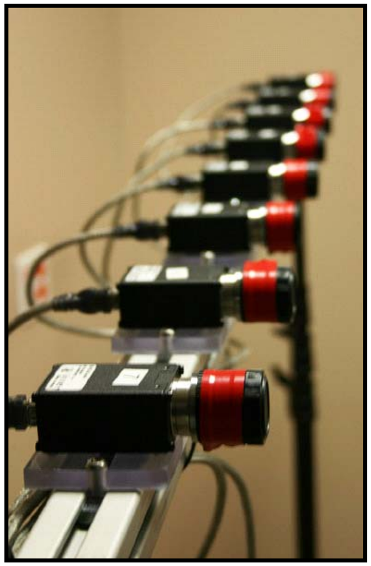
- Technically, the standard is a great success
- Industry appears to be open towards embracing SVC for DTV broadcast services
  - Specifically, enhancement of 720p to 1080p
- Others might be less certain, but still possible...
  - SVC for video conferencing equipment
  - Talk of using SVC for surveillance recorders
  - Lots of discussion on Scalable Baseline in ATSC-M/H
- Time will tell...

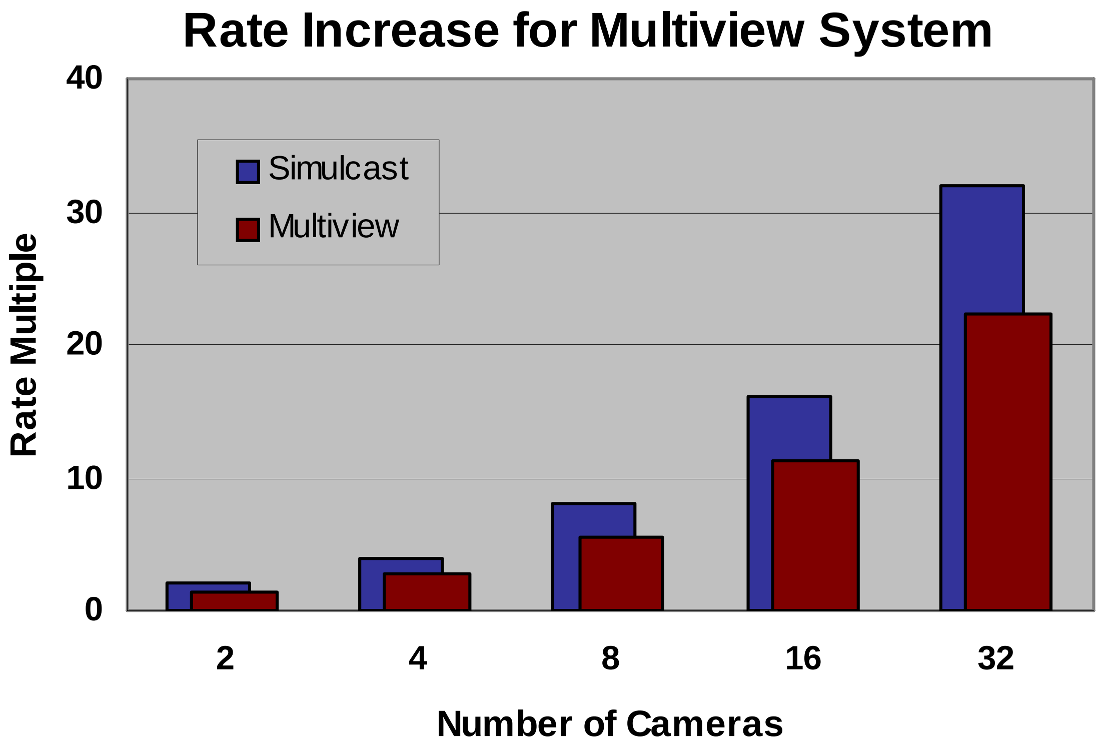
Potential for Bit-Depth Scalability

- Consumer displays with higher bit depth
  - Expected in next 2-3 years
  - Should create demand for better content
- Optical disc storage (e.g., Blu-Ray Disc) might be a promising route
  - Satisfy both legacy and higher-bit depth displays
  - However, scalability might just be a temporary solution for transition period in this case
- Use in other environments less certain
  - Broadcast studio, post-production environments
  - Medical, satellite, HDR imaging, etc.

Deployment of Multiview?

- Key issues for MVC [current scope, near-term]
  - Acquisition/production with large camera arrays difficult
  - Require rate of multiview video proportional to number of views



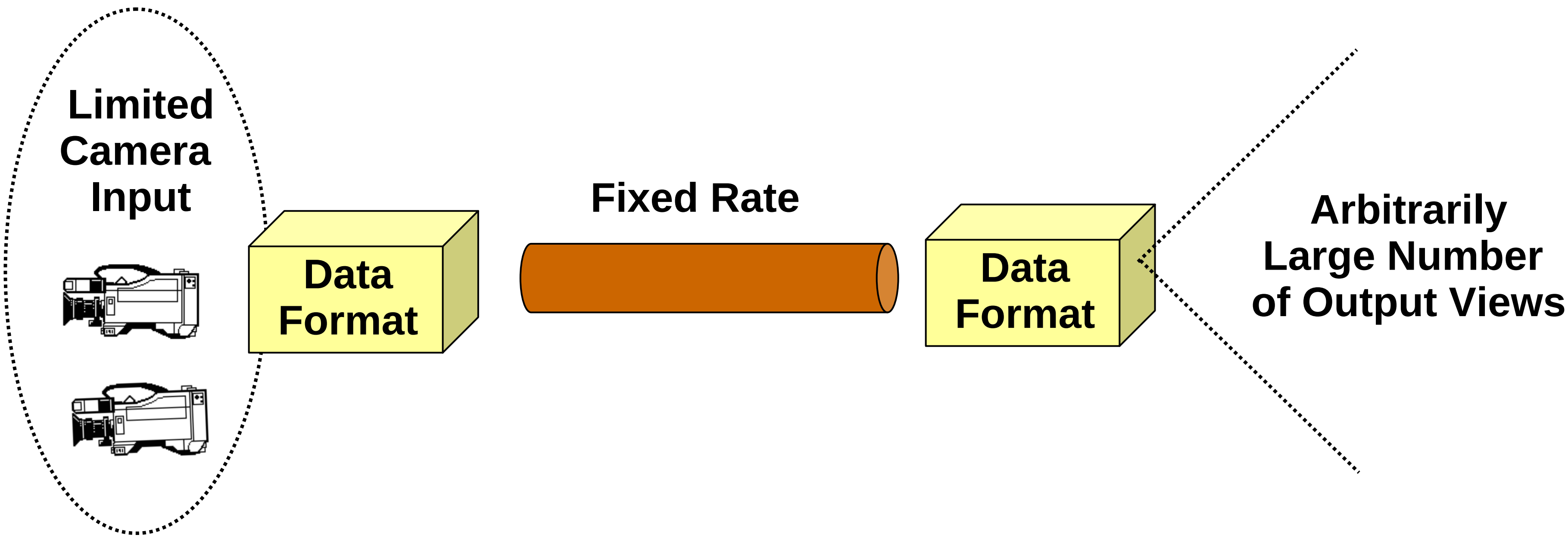


Note: 3D TV is here…

- 3D-ready sets on the market
  - Samsung & Mitsubishi
  - Use 3D DLP from TI
  - Shuttered glasses
- Content sources
  - 3D-enabled Blu-ray Disc players, game consoles
  - Software conversion, e.g., via Media Center PC
- *Where can we go from here?*

One Option: Get Rid of the Glasses

- More views needed to drive **auto-multiscopic** displays
- Technology challenge
  - Synthesize a continuum of views based on a limited set of decoded views
  - Specify a format that fixes the rate, but allows an arbitrarily large number of views to be rendered



Other Benefits

- Improved support for free-viewpoint video
  - Interactive navigation of scene
- Improved support for 3D displays
  - Benefits for advanced signal processing, e.g., upsample along view dimension for anti-aliasing

Plans are underway to standardize a multiview video + depth format

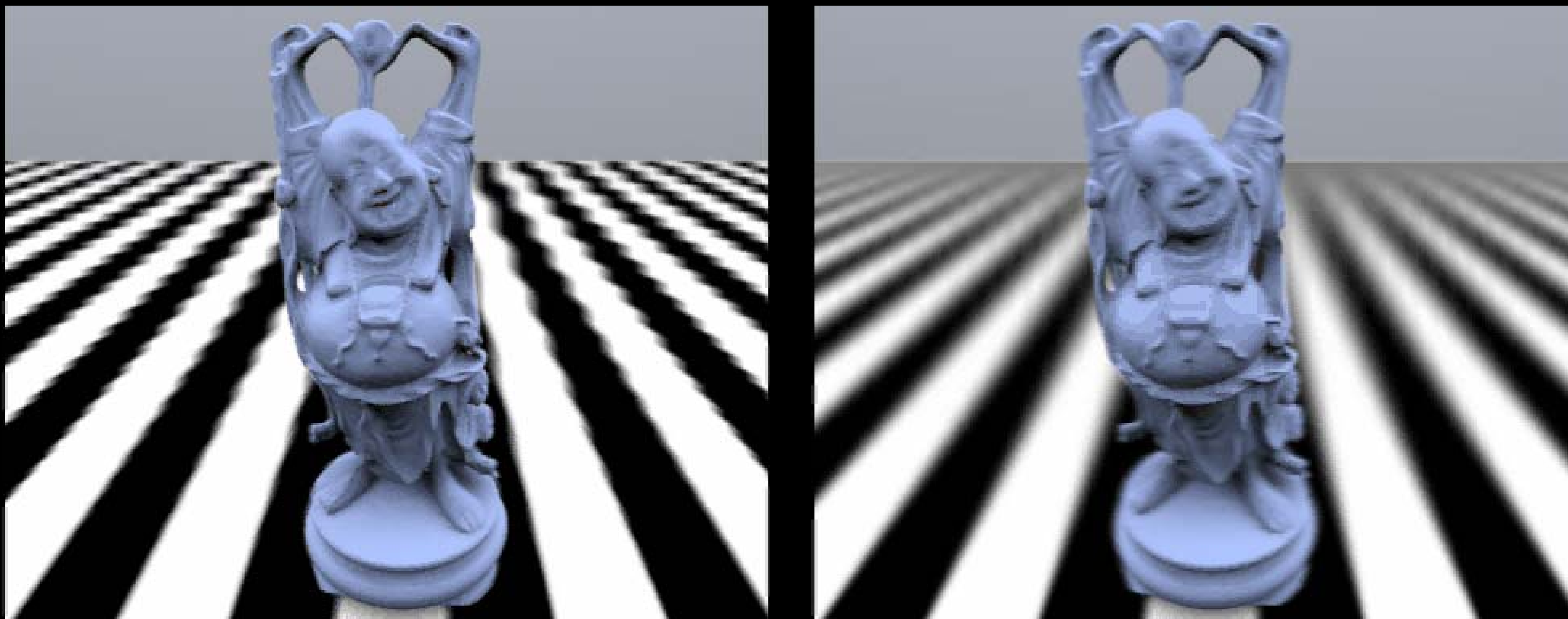
(but still at an early stage)

Free Viewpoint Video Using Depth

Microsoft Research - SIGGRAPH 2004

High-Quality Video View Interpolation Using a Layered Representation
C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, R. Szeliski

Anti-aliasing for 3D Displays Using Depth



Mitsubishi Electric Research Labs - Eurographics 2006

Antialiasing for Automultiscopic Displays
M. Zwicker, W. Matusik, F. Durand, H. Pfister

Interesting Technical Issues

- What is the correct data representation format?
  - Depth map associated with each camera view
  - Global depth map of the scene / sampling problem
  - Occlusion maps
- What is the best way to code depth?
  - Apply conventional (AVC / MVC) coding techniques
  - New methods: better preserve edge, exploit spatial characteristics
- How should depth be integrated into the framework?
  - Integral part of video payload, e.g., share motion vectors
  - As an independent and auxiliary stream
- Measuring the effectiveness of depth data
  - Quality of the depth: both estimation and coding perspectives
  - Quality of intermediate view (outside traditional coding domain)

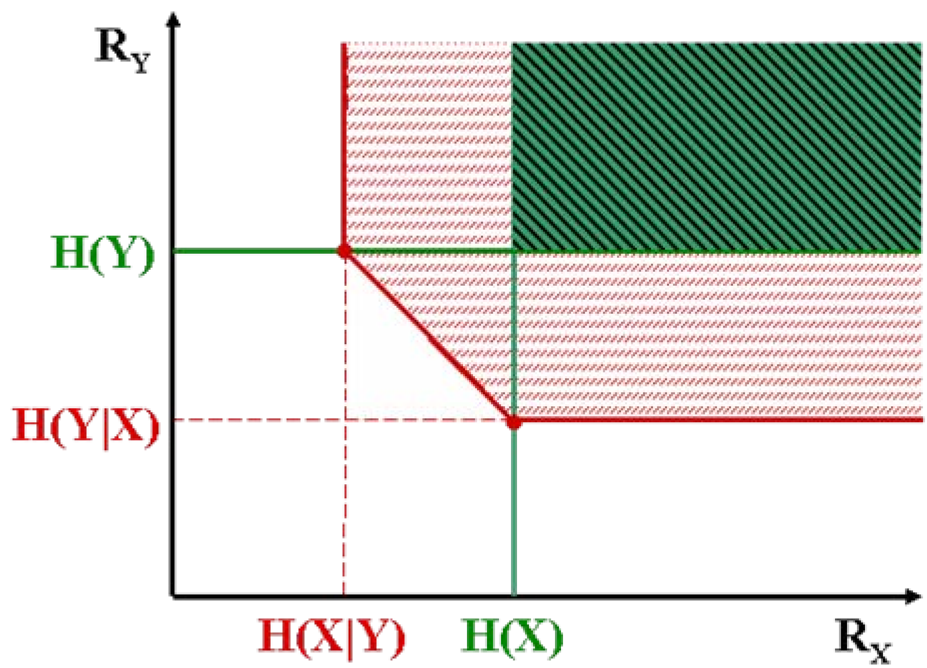
Additional Directions & Challenges

Future Directions

- Increasing single-layer coding efficiency
  - Better prediction and motion modeling
  - Better entropy coding and reduction of side info
  - Better transform and decomposition of source signal (talk by Martin Vetterli on DCT, Wavelets and X-lets)
- Perceptual Coding
  - How can we measure - subjective assessment only?
  - Geometric modeling, e.g., of textures, regions
  - Computer vision, analysis/synthesis techniques (talk by Tsuhan Chen on image understanding)

Future Directions

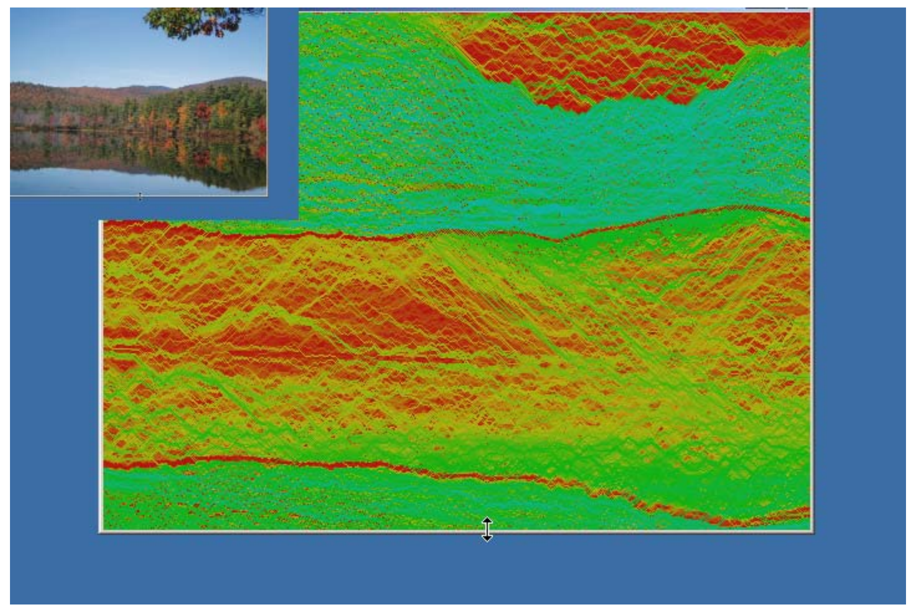
- Distributed Source Coding
  - Doesn't yet seem ready for standardization
  - Let's talk more about this in the panel...



Future Directions

- Other Dimensions of Scalability
  - Seam Carving : Content-aware resizing
    : Fine grain "spatial" scalability

[Avidan and Shamir, SIGGRAPH 2007]



Concluding Remarks

- **Unified design across all AVC extensions**
  - Substantial re-use of primary coding engines, layered/modular design
  - Borrowing of coding concepts across SVC and MVC extensions (e.g., inter-layer vs inter-view motion prediction)
  - Starting to see some relation between professional extensions and SVC
- **Breaking this mold will not be easy**
  - AVC is a set of very mature and well engineered technology
  - Sets the current benchmarks in conventional rate-distortion video coding performance - good for the community to have this!
  - Academia could learn from the technology selection process of standards (common conditions, accessible s/w, continually ratchet performance)
- **Moving forward**
  - Expect further improvements within AVC framework, but it would be more exciting to see alternative frameworks reach the same level of maturity
  - Requirements for new coding/transport paradigms should be considered
  - Potential shift away from maximizing pixel fidelity (as measured by MSE) towards alternative measures for visual quality assessment

Obrigado!