The Latest Development on MPEG Video Coding Standards

National Chiao Tung University
蔡淳仁 Chun-Jen Tsai
03/8/2004

Outline

- Introduction
- MPEG-4 Part 10: AVC (a.k.a. H.264)
  - Design Concept
  - What Makes the Differences
  - Licensing Issues
- MPEG-21 Part 13: Scalable Video Coding
  - MPEG SVC Call-for-Proposal
  - Potential Technologies
- Conclusion

Introduction

- MPEG-2 is an extremely successful video codec
- MPEG-4 SP and ASP is gaining momentum, but does it fulfill the demands of new applications?
- New tricks in the MPEG toolbox:
  - Back to the basics (well, sort of): MPEG-4 AVC
  - Jump into the future: MPEG-21 SVC

MPEG-4 AVC/ITU-T H.264

- Developed by the Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG
- Became an International Standard in May, 2003
- Goal: 50% coding efficiency gain over MPEG-4 ASP
- Three layer design:
  - Video Coding Layer (VCL):
    - How to compress source video
  - Network Abstraction Layer (NAL):
    - How to pack bitstream data error-resiliently
  - Transport Encapsulation Layer (TEL):
    - How to carry the NAL units over real transport systems
    - This layer was a “spin-off” from the original NAL design

Video Coding Layer Concept

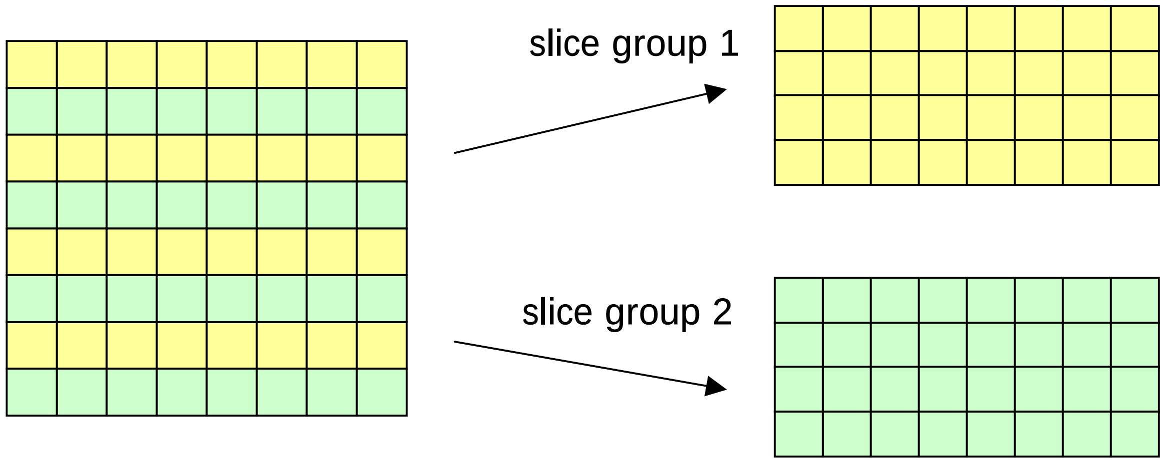
- Compression efficiency was the top priority
  - Bit-error resilience is no longer an important issue
- Slice-based operation (“picture-based” is obsolete?!)
  - Packet-based transport systems dominate the future
- Information provided by the transport (system) shall not present in VCL (e.g. timestamps)

VCL Key Features

- Predictive coding
  - 13-mode luma, 4-mode chroma intra prediction
  - Gazillion modes for inter prediction
- 16-bit integer combined transform/quantization
  - Exact forward-inverse transform pair is used
  - Transform block size is 4x4
- Two types of entropy coding methods
  - Universal VLC and Context Adaptive VLC
  - Context Adaptive Binary Arithmetic Coding
- In-loop filter

Controversial VCL Features

- Flexible Macroblock Ordering (FMO), Arbitrary Slice Ordering (ASO), Data-Partitioning (DP) are controversial “VCL” features
- For example, FMO causes hardware design problem:
  - MBs in a frame can be assigned to different slice groups
  - Each slice group is further divided into slices

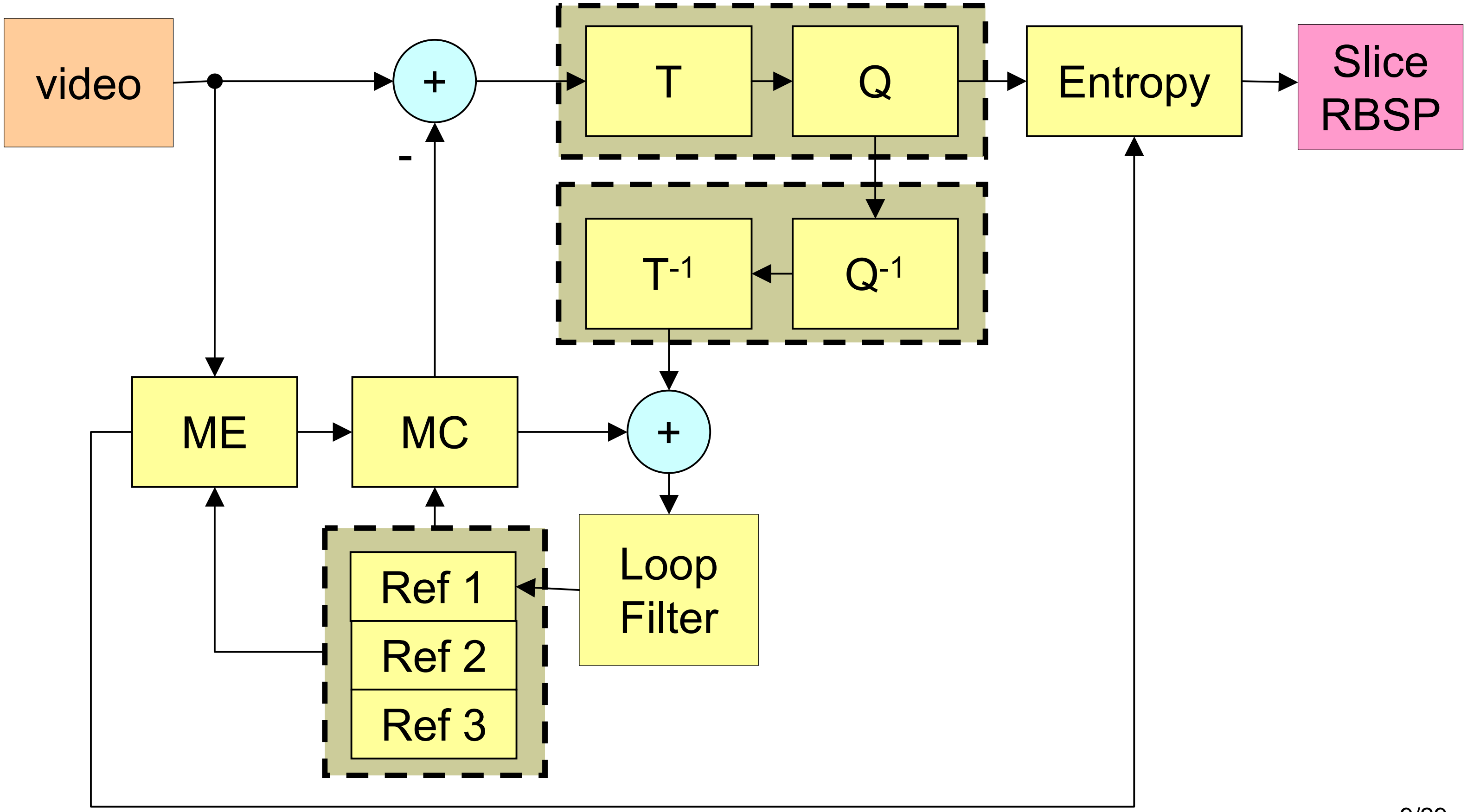


Network Abstraction Layer Concept

- NAL is an packet-based compressed data format mainly designed for error-resilience
- NAL is suitable for both packet-oriented and bitstream-oriented transports
- Each NAL units carries some video or system data
- In bitstream-oriented transports, each NAL can be preceded by a start code prefix $\rightarrow$ requires a *start code emulation prevention* mechanism

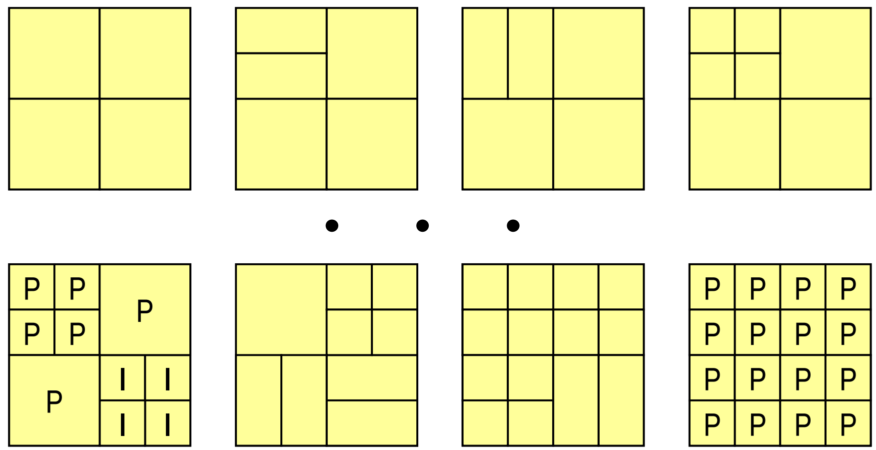
Not Much News in the Big Picture

- For example, the encoder diagram:

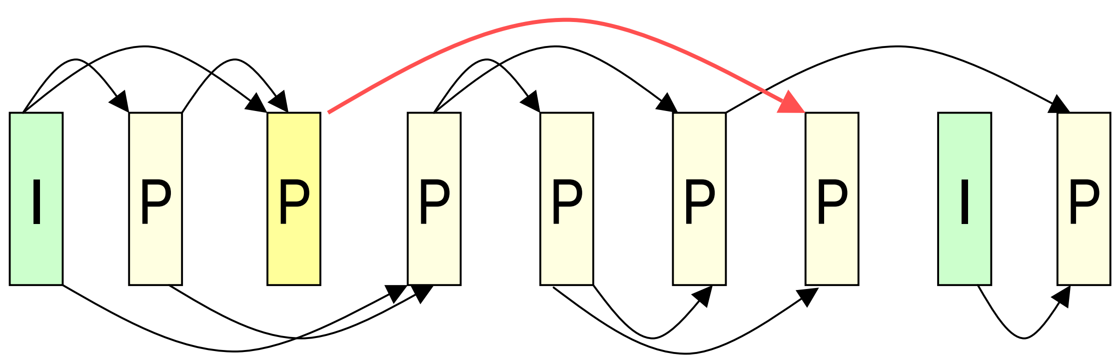


It’s All in the Details

- For example, the gazillions of coding modes:

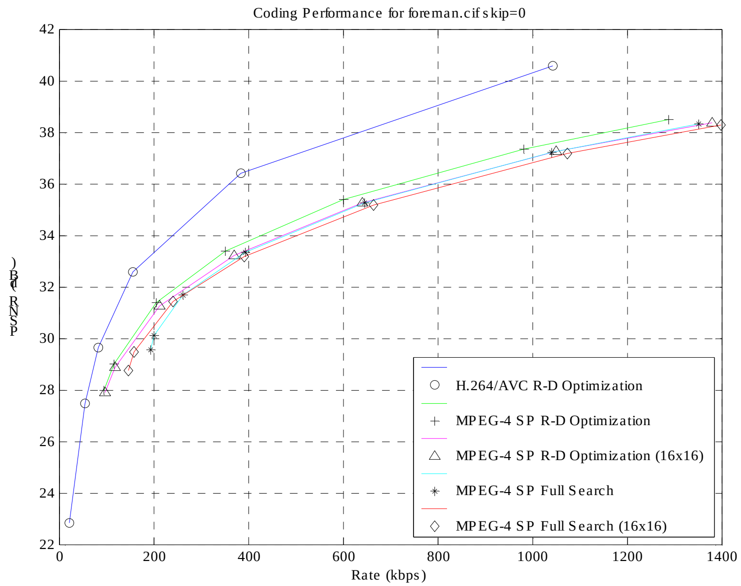


- Or, the complicated references frame patterns:



So, How Much Do We Gain?

- AVC Baseline vs. SP



- MPEG also conducted a verification test. AVC achieves 2x coding efficiency gain over competing codecs in roughly 77% of the test cases (from 48 kbps to 20Mbps)

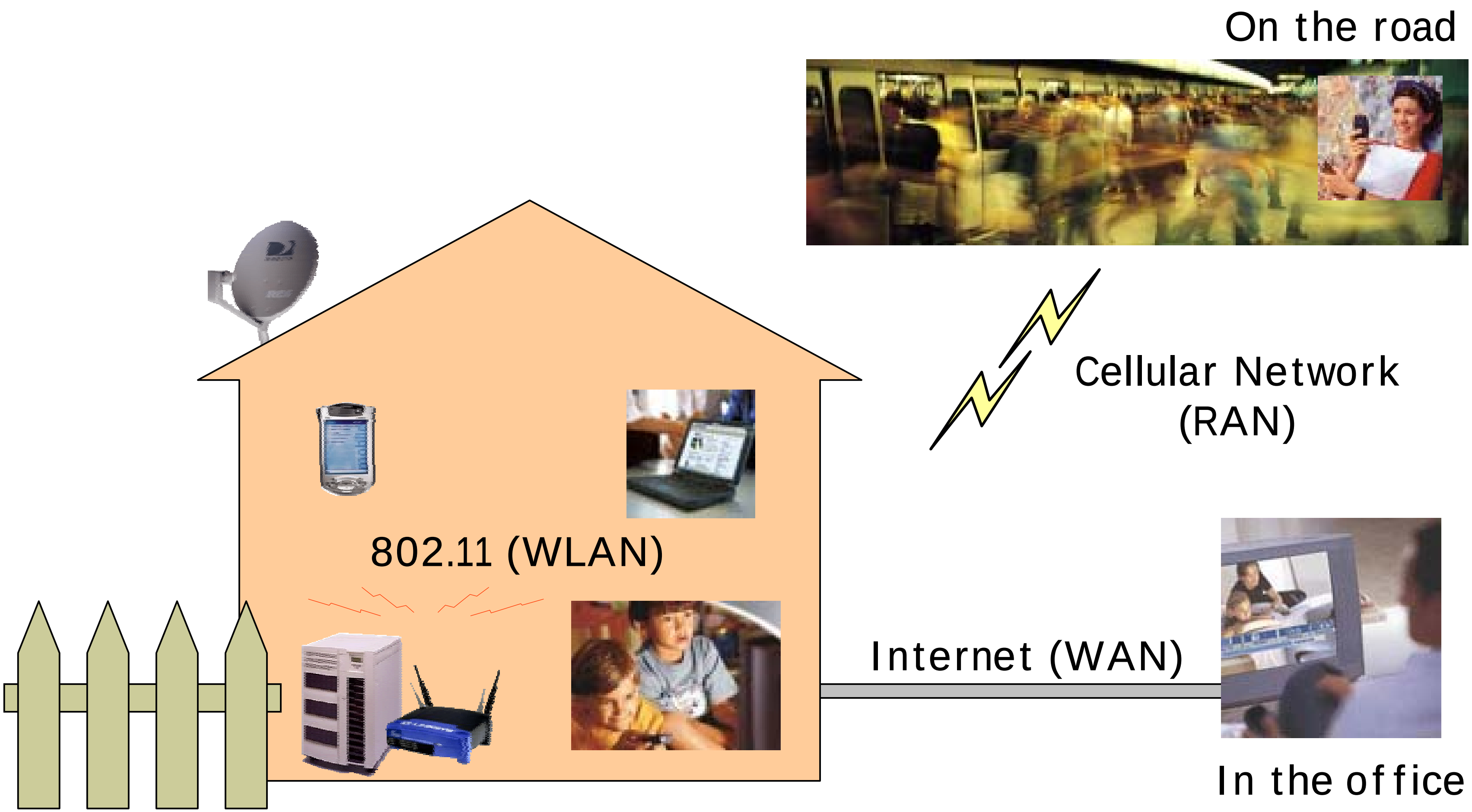
Potential AVC Killer: Licensing Issue

- JVT original goal was to make Baseline free
  - Some "better" technology was kicked out for this reason
  - Baseline tools used in Main Profile are not free
- MPEGLA does not allow different licensing terms for same technology
- Current term is especially harsh for the digital broadcasting community. EBU already published an open statement to boycott the new standard if licensing terms don't change

Scalable Video Coding

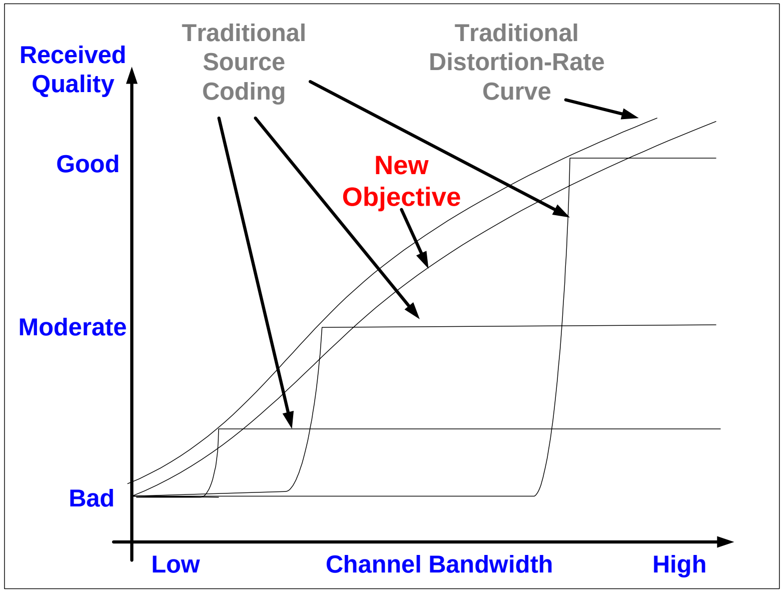
- There are three factors (dimensions) that determines the perceptual quality of a video presentation:
  - Picture Resolution
  - Frame rate
  - Bitrate
- Traditionally, these parameters was fixed once the coding is done
- For SVC, we want to be able change these parameters on-the-fly during the presentation

Application Scenario



Theoretical Goal of SVC

- Change quality smoothly in real time



- Million dollar question: how to measure “quality” in 3D scalable space?

MPEG SVC Call-for-Proposal

- Schedule

Dec. 1, 2003	Preliminary intention to participate
Dec. 31, 2003	Deadline for Pre-registration
Feb. 1, 2004	Formal registration (€ 1,500-2,000)
Feb. 16, 2004	Coded test material at the test site
Feb. 20, 2004	Subjective assessment starts
March 1, 2004	Registration/submission of documents
March 9, 2004	Report of the subjective test results
March 15-19, ‘04	MPEG 68th meeting, München, DE

Status of the CfP

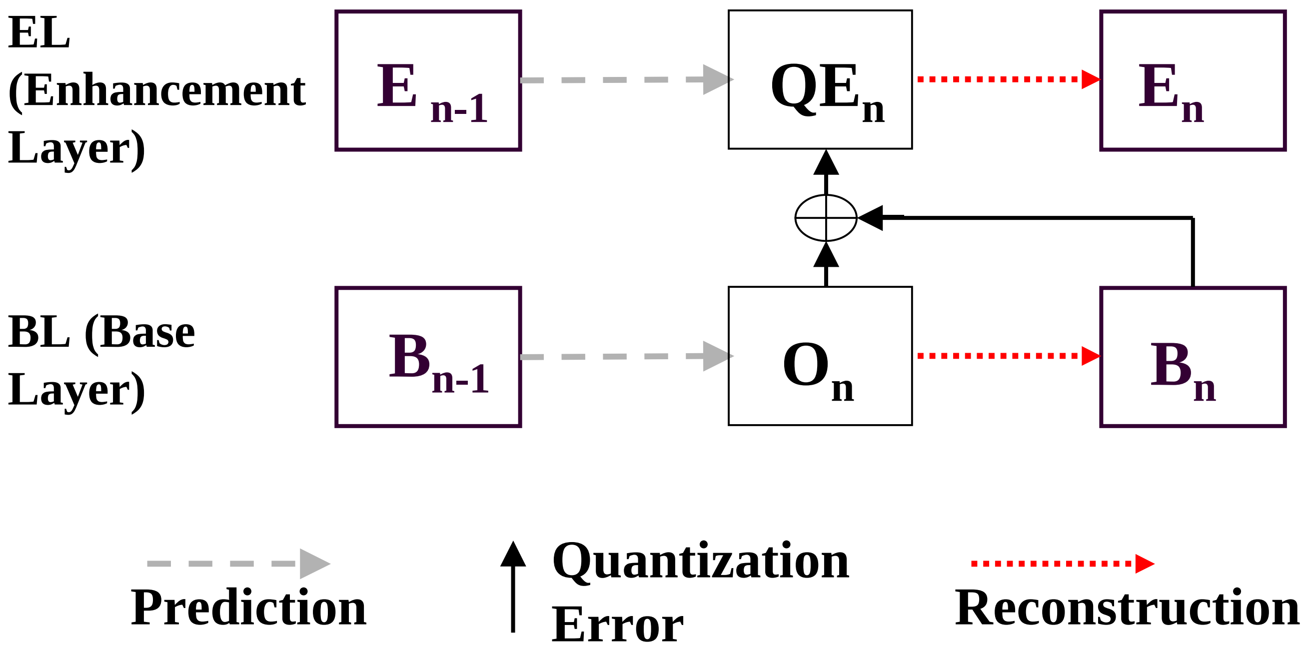
- 21 registrations of proposals (from 9 academic institutes/ 12 companies)
- 11 proposals are based on inter-frame wavelet,
  5 proposals are based on DCT, and
  5 proposals are based on undisclosed technologies.
- Two test scenarios:
  - Scenario 1: fully scalable codecs from 64kbps to 6Mbps; and from QCIF to 4CIF (13 proposals)
  - Scenario 2: limited scalable codecs from 48kbps to 1Mbps; and from OCIF to CIF (14 proposals)

Potential Technologies

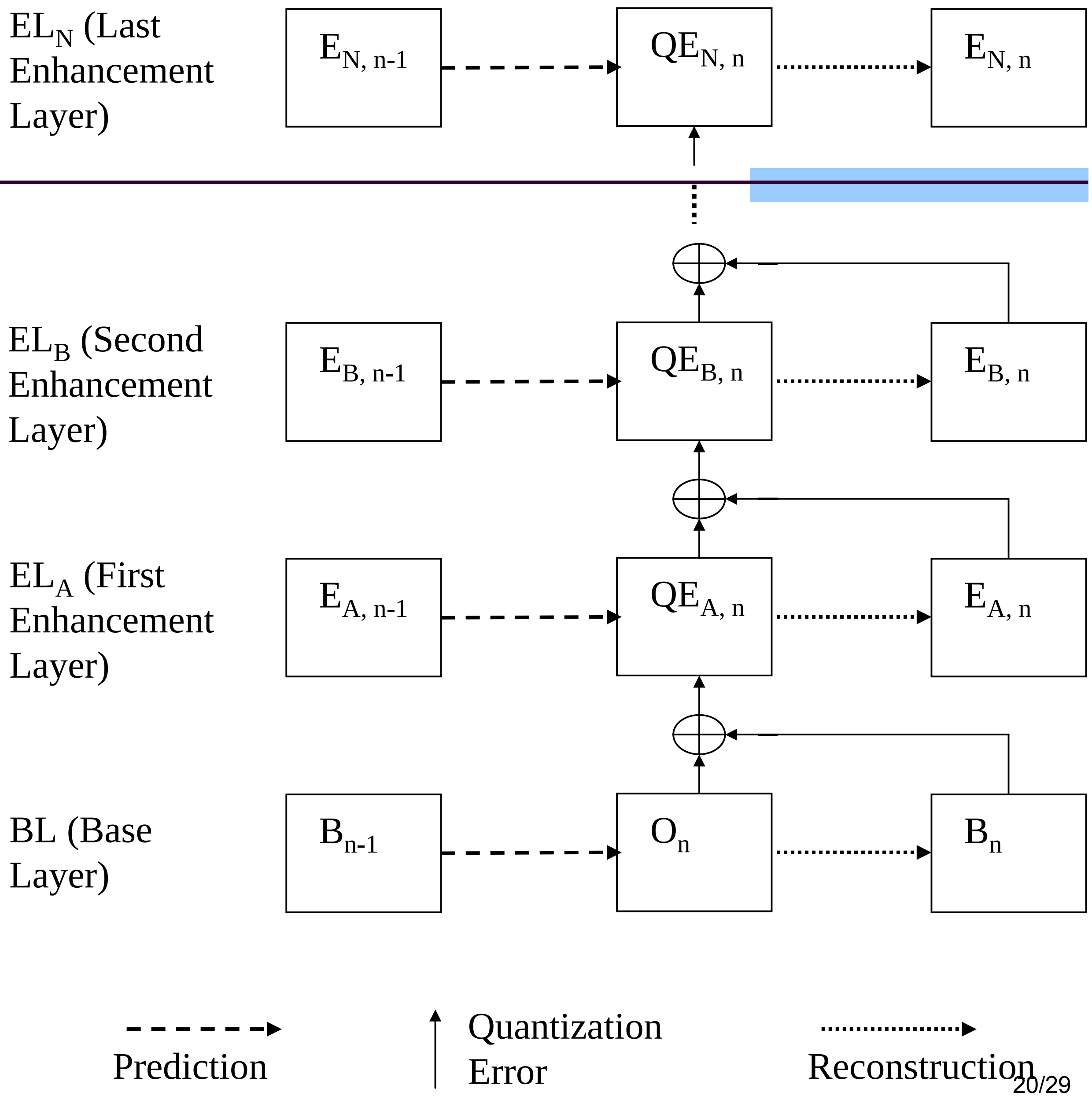
- DCT-based:
  - MPEG-4 SSP, FGS are well known, but lacks full-dimensional scalability and coding efficiency
  - New DCT-based techniques are based on AVC, use multiple prediction loops, and progressive layer approach to increase both coding efficiency and scalability
- Wavelet-based:
  - Most candidates are based on Prof. John Woods' 3D Subband Coding approach (and his reference code)
  - Full-dimensional scalability with decent high bitrate coding efficiency
  - The performance under low bitrate (below 512kbps) requires some improvements

DCT-based Example: SRFGS

- H.-C. Huang, W.-H. Peng, C.-N. Wang, T. Chiang, and H.-M. Hang, “**S**tack **R**obust **F**ine **G**ranularity **S**calability,” MPEG Document M9767
- RFGS Concept:

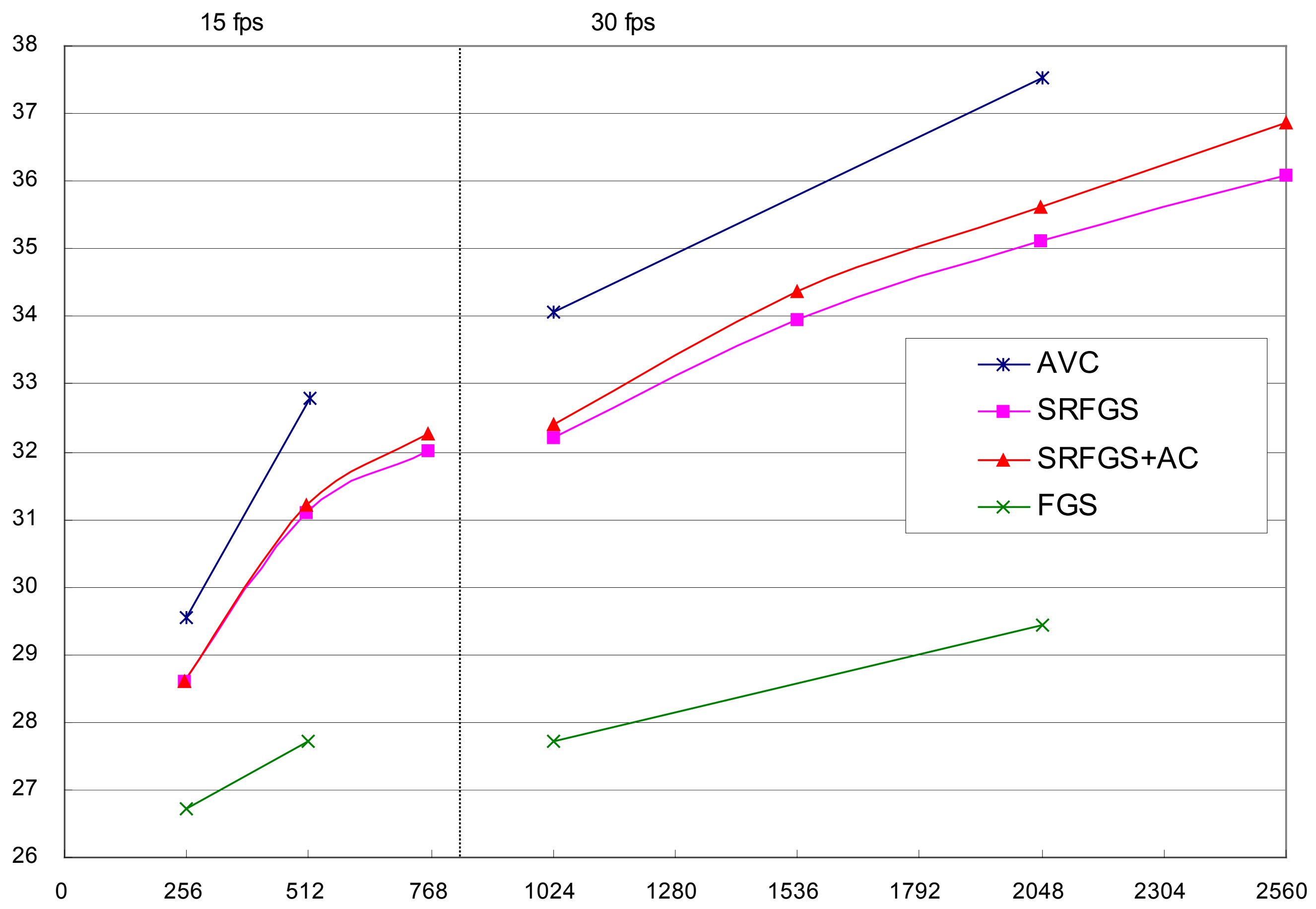


SRFGS

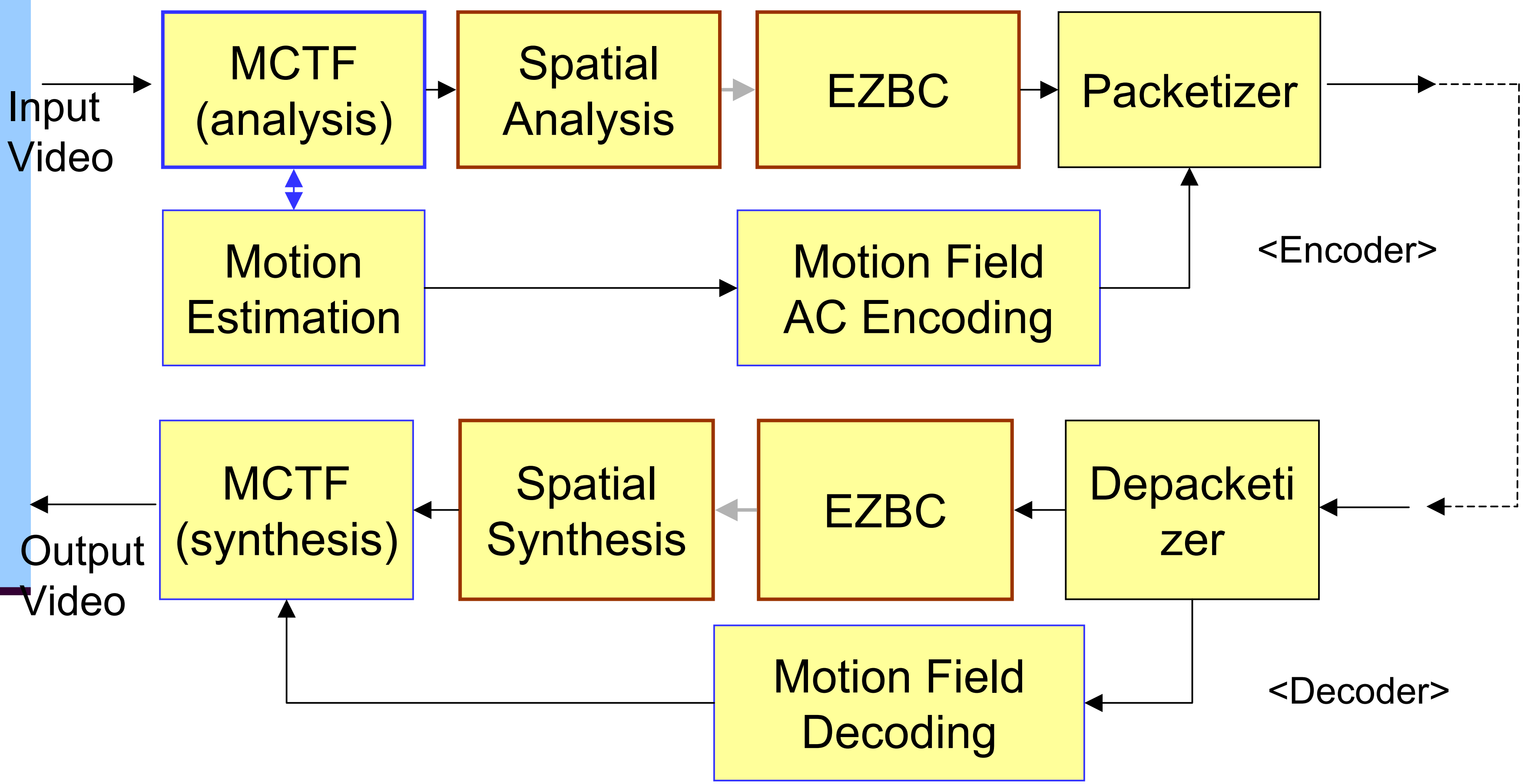


EL_N (Last Enhancement Layer)
$E_{N, n-1}$
$QE_{N, n}$
$E_{N, n}$
EL_B (Second Enhancement Layer)
$E_{B, n-1}$
$QE_{B, n}$
$E_{B, n}$
EL_A (First Enhancement Layer)
$E_{A, n-1}$
$QE_{A, n}$
$E_{A, n}$
BL (Base Layer)
B_{n-1}
O_n
B_n
Prediction
Quantization Error
Reconstruction

Rate-Distortion Plot for Mobile Seq.



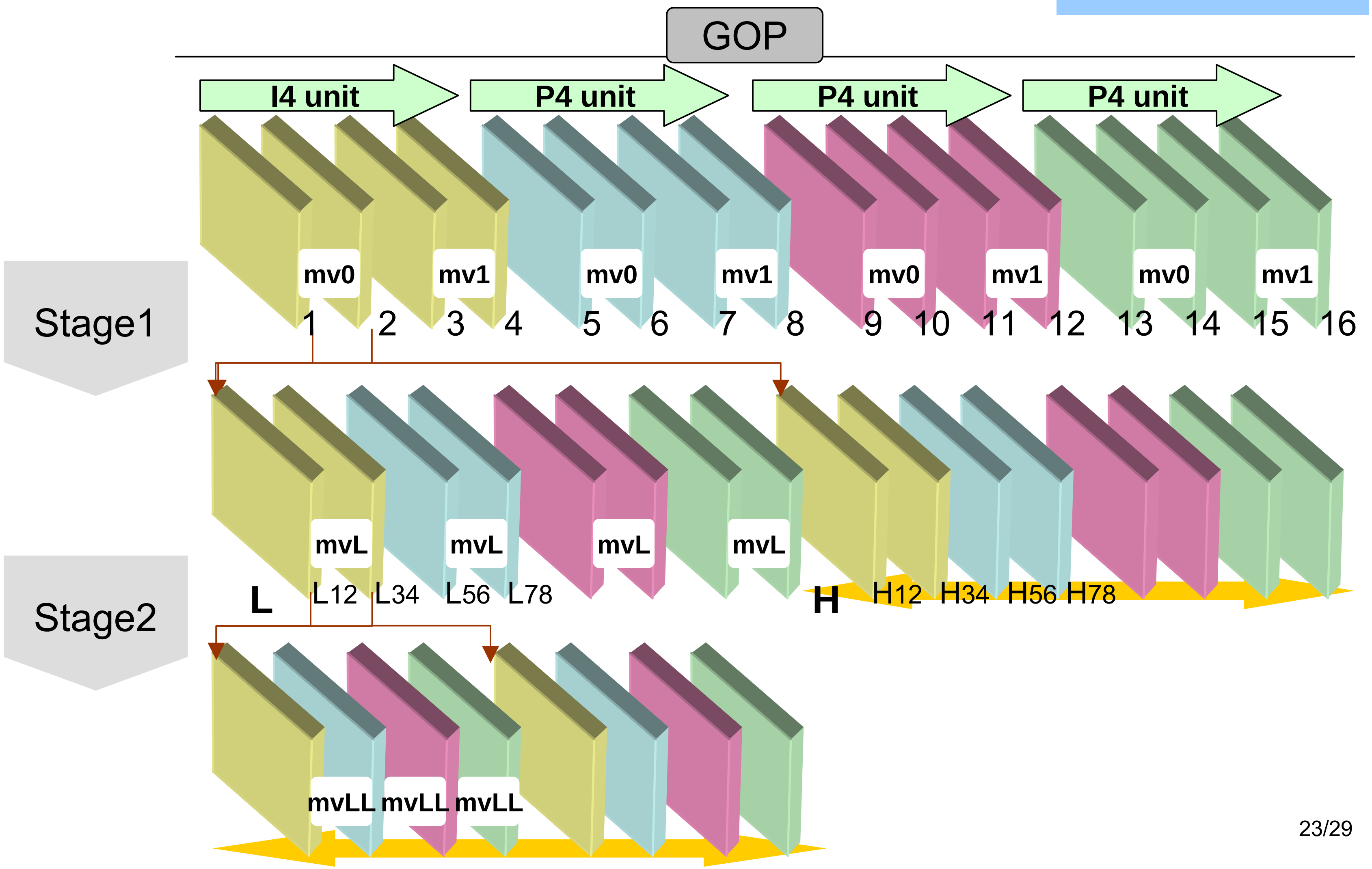
Wavelet-based Example: MC-EZBC



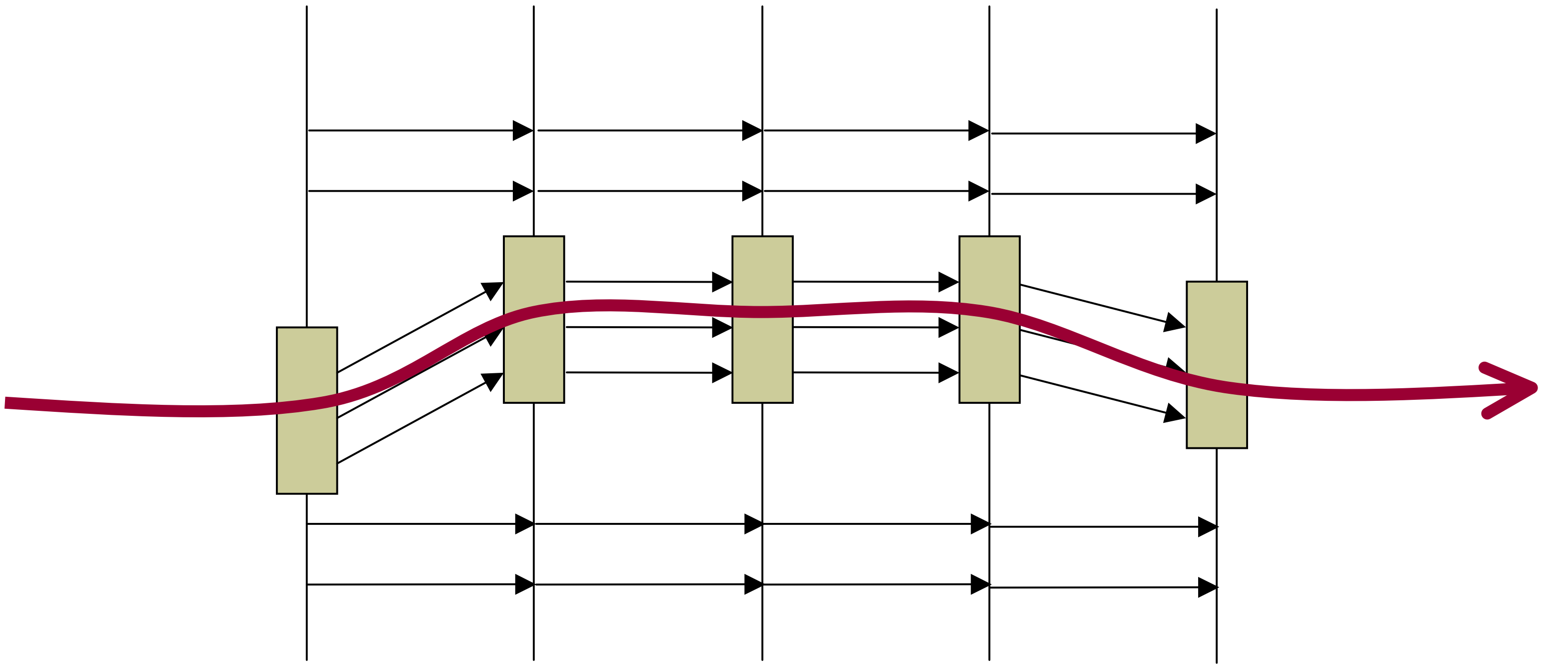
MCTF: Motion Compensated Temporal Filtering
EZBC: Embedded Zeroblock Coding

AC : Arithmetic Coding

Temporal Subband Decomposition

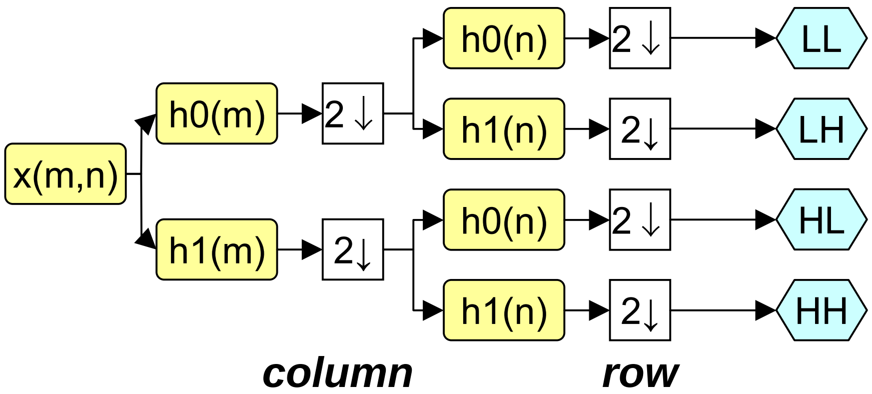


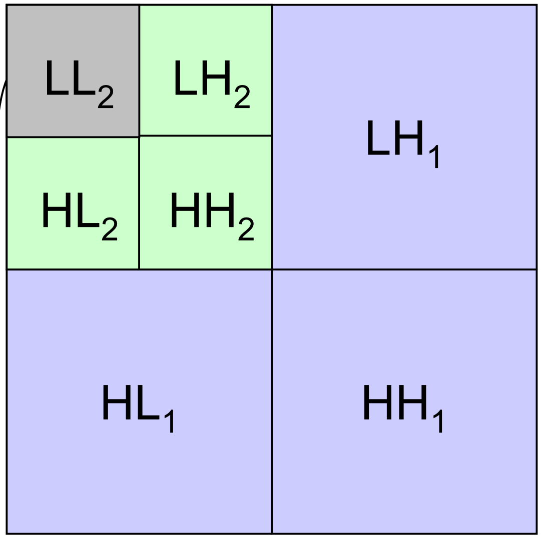
MCTF Concept



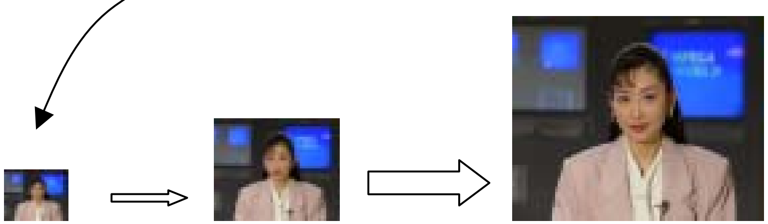
Spatial Subband Decomposition

- Decomposition





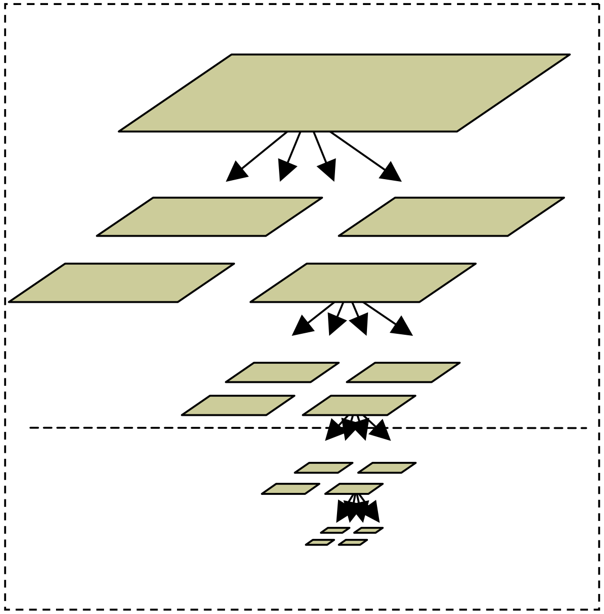
- Spatial Scalability



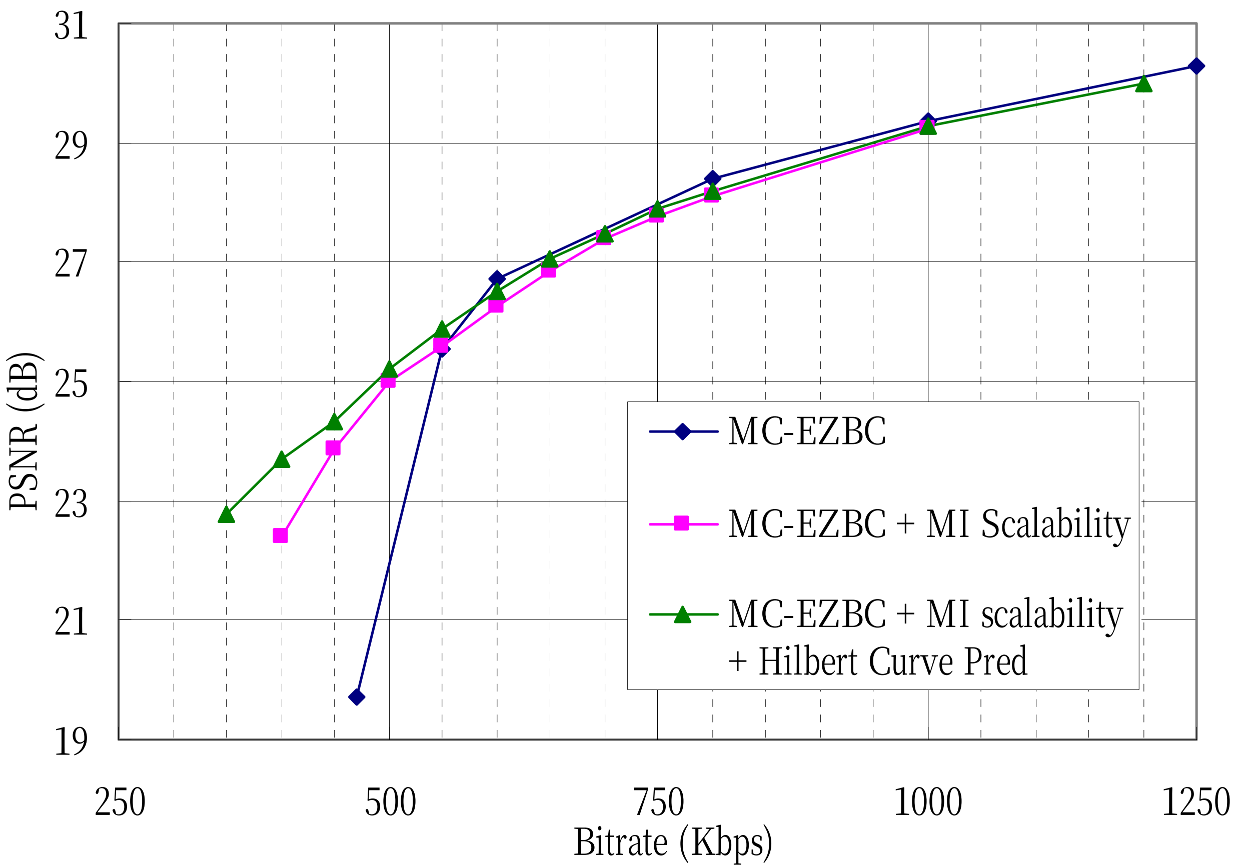
Rate-Distortion Plot for Mobile Seq.

Motion Information Scalability

- A Proposal from NCTU:
  S.S Tsai, H.-M. Hang, Tihao Chiang, “Motion Information Scalability for MC-EZBC,” MPEG Document M9723
- Compressed data = motion info. + residual wavelet-transformed image data
- Partition of motion vector by size
  - Base: 64-by-64 to 16-by-16
  - Enhancement: 8-by-8 to 4-by-4



Rate-Distortion Plot for Harbour Seq.



Conclusions

- It is still unclear whether AVC and/or SVC can be as successful as MPEG-2 due to various non-technical reasons
- In the long run (5 to 10 years), SVC may be more important than AVC since channel bandwidth/storage space is increasing rapidly. On the other hand, content authoring cost and device adaptability demand are also increasing rapidly