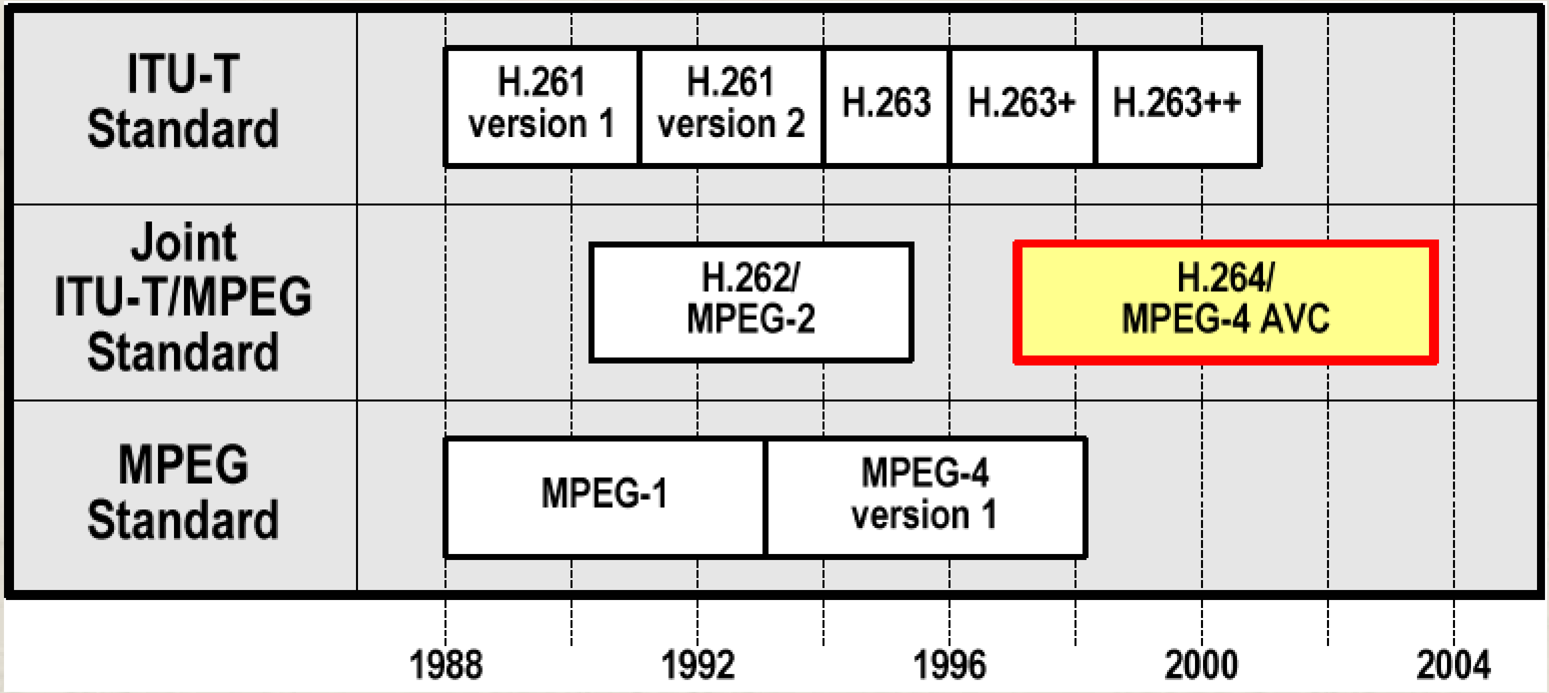
Overview of H.264 Video Coding

Trac D. Tran
ECE Department
The Johns Hopkins University
Baltimore, MD 21218

Outline

- Video coding standards
  - History
  - Generic framework
- H.264/ MPEG-4 AVC
  - Main features
  - Key technical innovations
  - Coding performance
  - Profiles: basic, main and ***high*** profiles
- Challenging problems
- Applications and markets

History of Video Standards



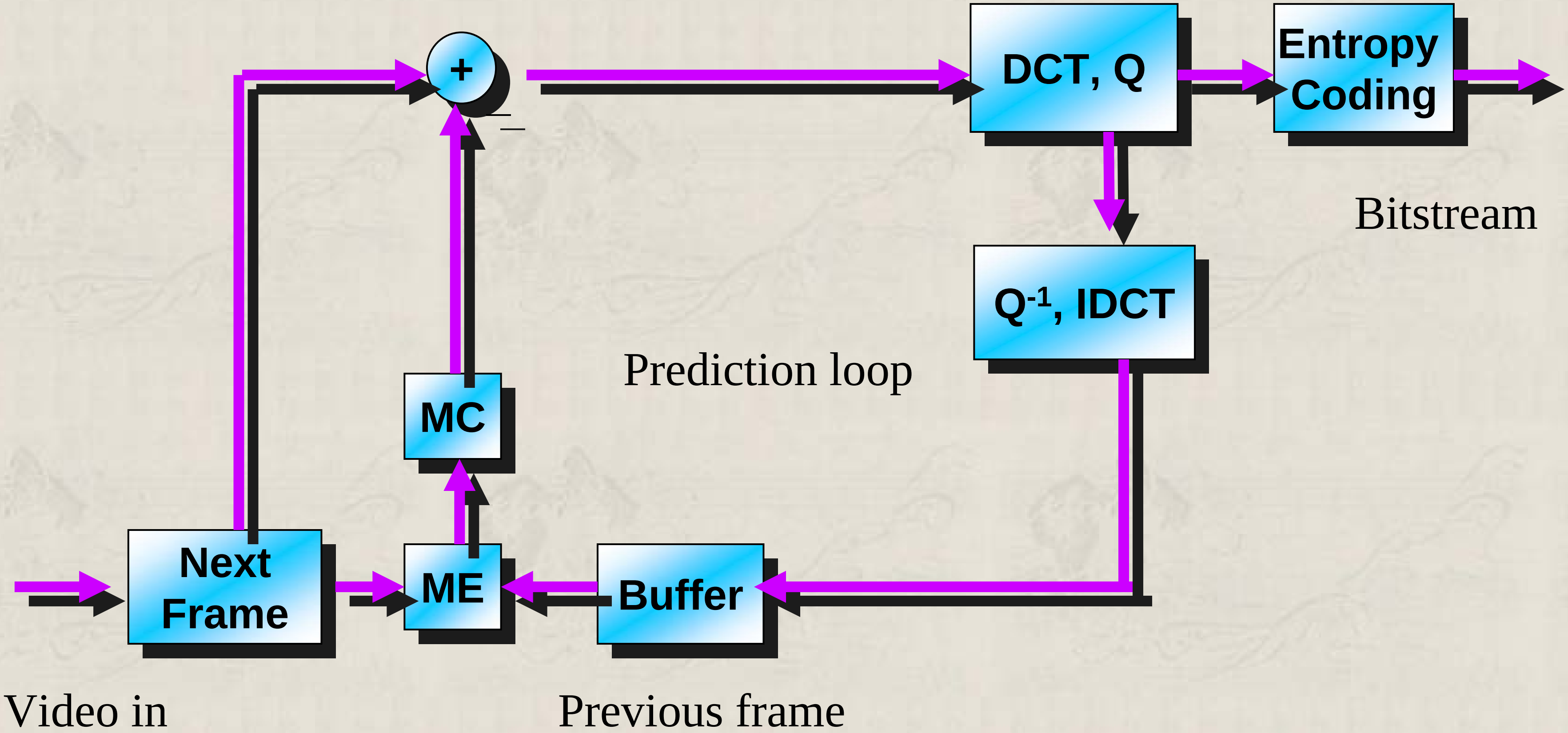
ITU H.26x History

- ITU H.26L: “long-term” solution for low bit-rate video coding for communication apps
- Predecessors include
  - H.261 (1990): “px64”, video conf. solution
  - H.263 (1995): next conf. solution, used in H.323
  - H.263+, H.263++, follow-on solutions
- H.26L project dates back to early ’90s
- Call for formal proposals in January 1998
- First draft in August 1999
- Joining forces with MPEG: Dec. 2001
- H.264 (H.26L) completed in May 2003

MPEG History

- MPEG-1 (1993)
  - Video on CD (VCD)
- MPEG-2 (1994)
  - DTV Broadcast, DVD, HD
- MPEG-4 (1999 -)
  - Cell phone, interactive, high rate communication
  - Object-oriented
  - Over-ambitious?
- AVC (2003)
  - Conventional to HD
  - Emphasis on compression performance and loss resilience

Generic Framework*



* H.261, 263, 263+, MPEG-1/2/4

H.264 Video Coding

- Development history
- Main features
- Key compression techniques
  - Tools
  - Framework
- Performance
- Profiles
  - Basic and main profiles
  - High profile
  - Other new profiles

Development History

- Dec 2001 – Start
  - Joint Video Team (JVT) formed between ITU/MPEG
- Dec 2002 – Tech freeze
- May 2003 – ITU-T Rec. H.264
- June 2003 – ISO/IEC final draft (FDIS)
- July 2003 – Launch of FRExt (Fidelity Range) extension project
- Oct 2003 – ISO/IEC (14496-10) AVC
- Dec 2003 – Verification tests by MPEG
- Jun 2004 – FRExt project is finalized
- Jan 2005 – Scalable Video Coding (SVC) project starts
- Jul 2006 – Multi-View Video Coding (MVC) project starts

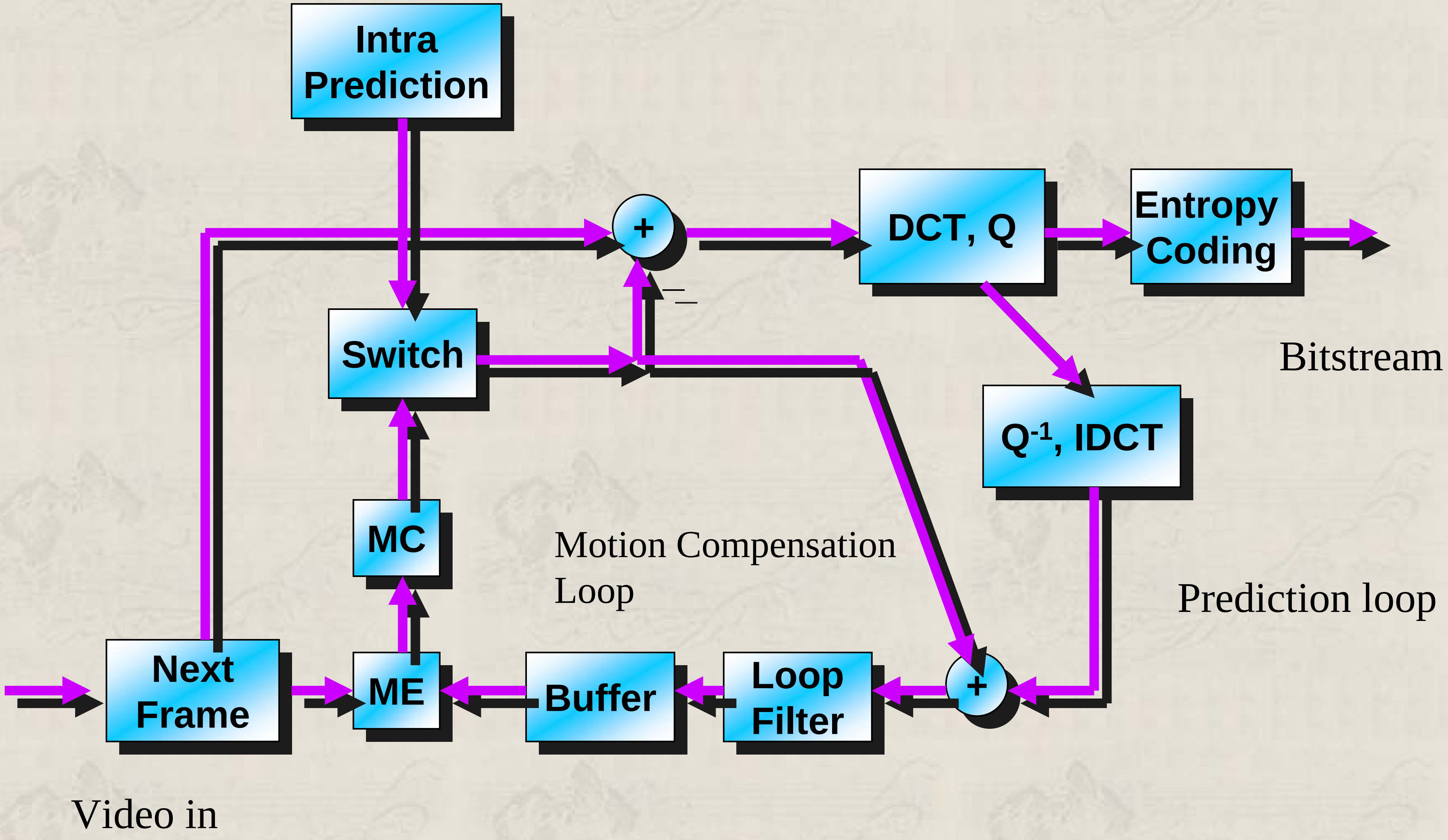
Main Features

- High compression performance
  - Advanced compression tools
  - Average 50% bit rate reduction given fixed fidelity compared to other standards
- Exact match decoding
  - Integer transform
- Improved perceptual quality
  - In-loop deblocking filter
- Network friendliness
  - NAL (network abstraction layer)
  - Enhanced error resilience

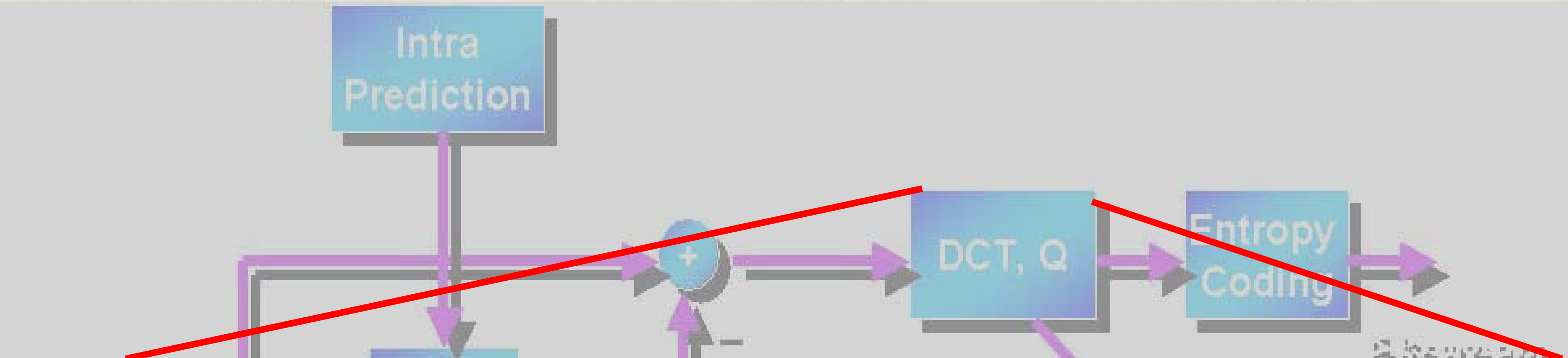
H.264 Technical Tools

- Structure
  - Sequence ->GOP->Picture->Slice->MB->Block
- Picture type: I, P, B, SI, SP
- Frame structure: interlaced, progressive
- Adaptive frame/field: per picture, per MB
- Deblocking filter – in loop
- MV resolution – ¼ pixel
- Tree-like motion segmentation – 16x16 to 4x4
- Entropy coding – CAVLC/CABAC
- Data partition – NAL unit, priority
- ASO (arbitrary slice order) – independently decodable
- FMO (flexible macroblock order) – map
- ABP (adaptive bi-prediction) – adaptive weighting

Block Diagram: H.264 Encoder



Innovation 1: Transform



$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

Quantization step size control is nonlinear: step size increases gradually by about 12% (double after 6 steps)

16 bit 4x4 DCT

- **EXACT MATCH** simplified transform
  - 4x4 transform

4x4 DCT

$$H = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ a & -a & -a & a \\ c & -b & b & -c \end{bmatrix}$$

where $a = 1/2$, $b = \sqrt{1/2} \times \cos(\pi / 8)$,

and $c = \sqrt{1/2} \times \cos(3\pi / 8)$.

H.264

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix}$$

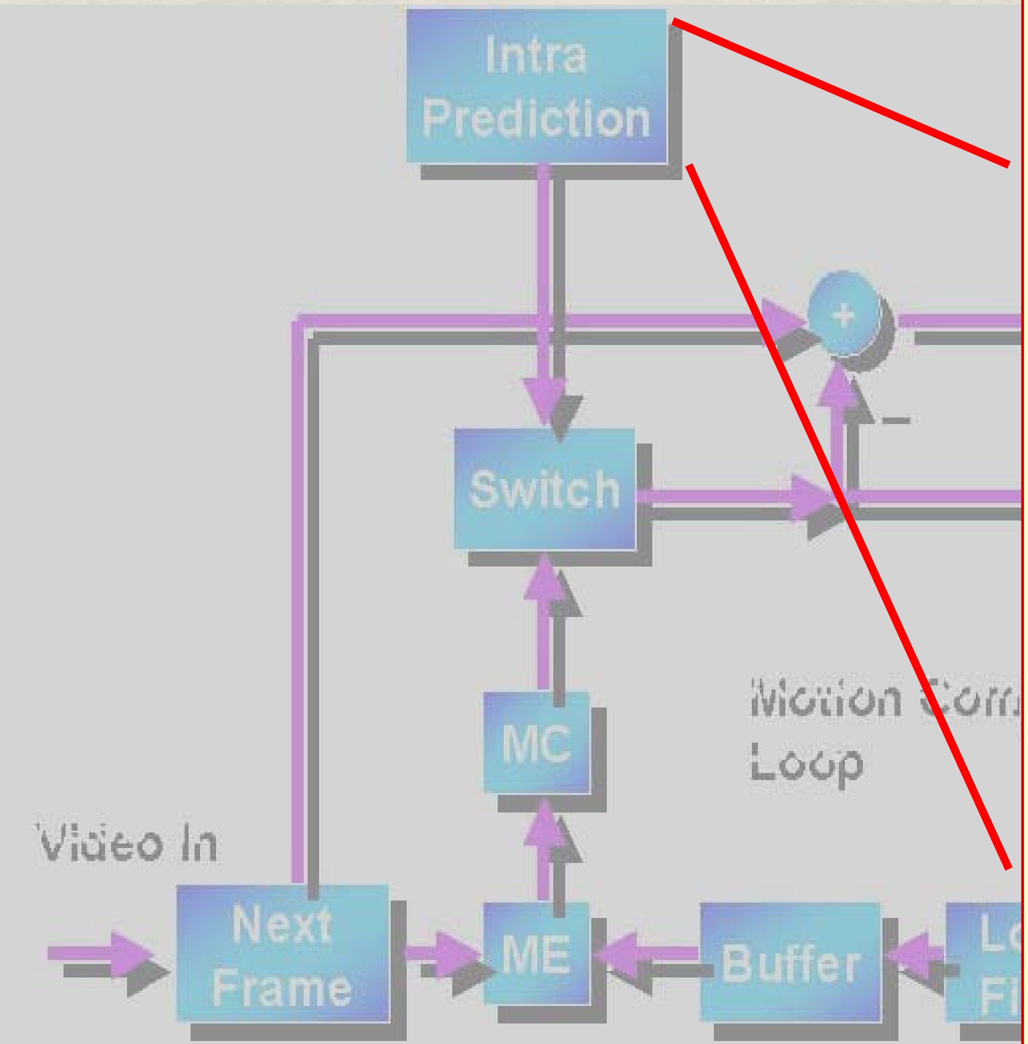
  - Non-orthonormality of the integer transform, i.e., position dependent scaling
  - Requires only 16 bit arithmetic (including intermediate values)
  - Expanded to 8x8 for Chroma by 2x2 transform of the DC values

Quantization

- Quantization of transform coefficients
  - Logarithmic step size control
  - Extended range of step sizes
  - Smaller step size for chroma
  - 16-bit multiply, add and shift
  - Table-driven: 2 times in Qstep for every 6th increment in Qp

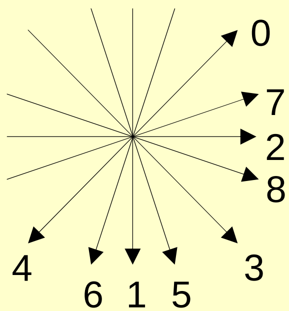
QP	0	1	2	3	4	5	6	7	8	9	10	11	12	...
qStep	0.625	0.6875	0.8125	0.875	1	1.125	1.25	1.675	1.625	1.75	2	2.25	2.5	...
QP	...	18	...	24	...	30	...	36	...	42	...	48	...	51
qStep	...	5	...	10	...	20	...	40	...	80	...	160	...	224

Innovation 2: Intra Prediction



- Directional spatial prediction (9 types for luma, 1 for 4x4 chroma)

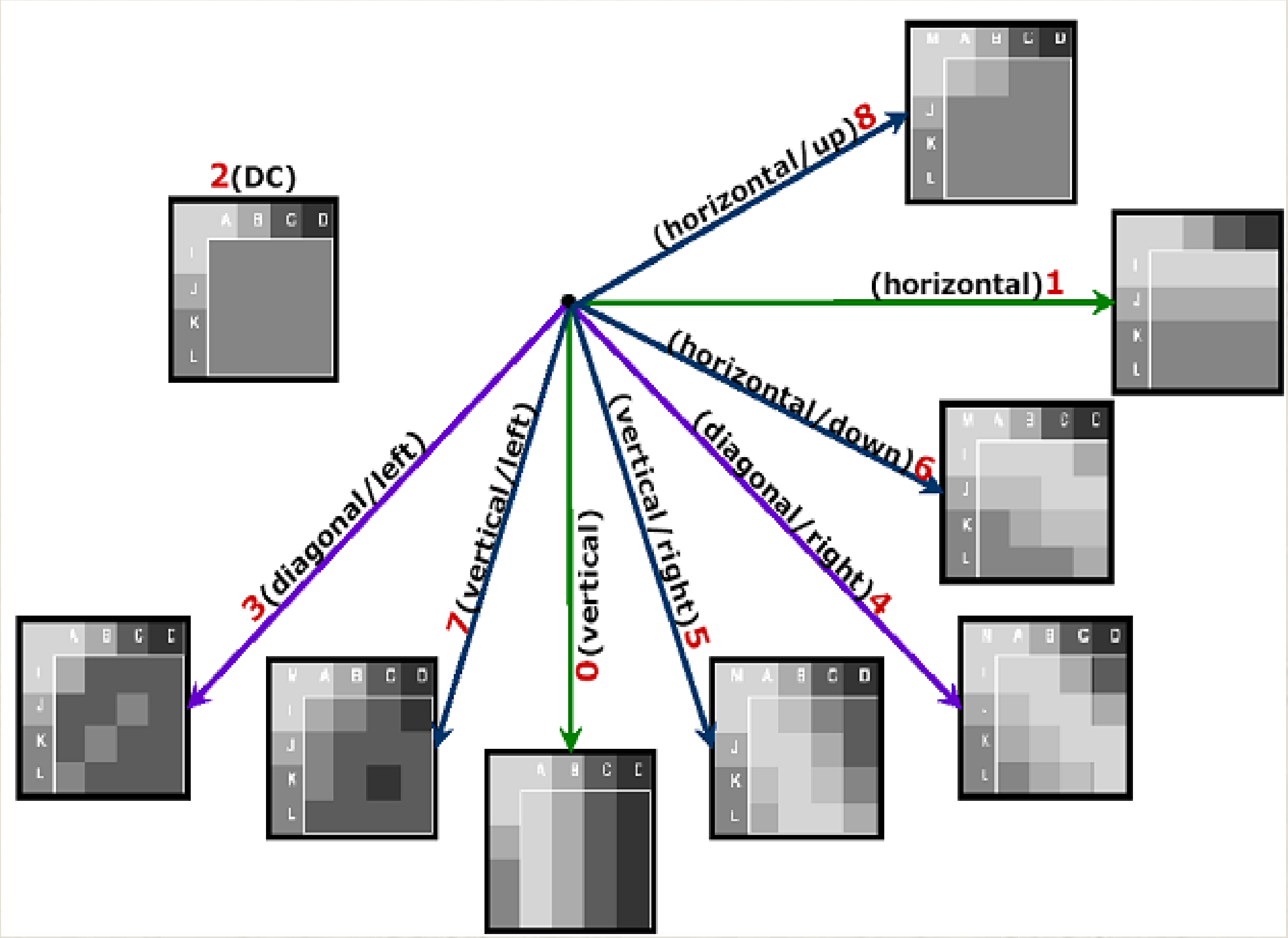
```
Q A B C D E F G H
I a b c d
J e f g h
K i j k l
L m n o p
M
N
O
P
```



- e.g., Mode 3:
  diagonal down/right prediction
  a, f, k, p are predicted by
  (A + 2Q + I + 2) >> 2

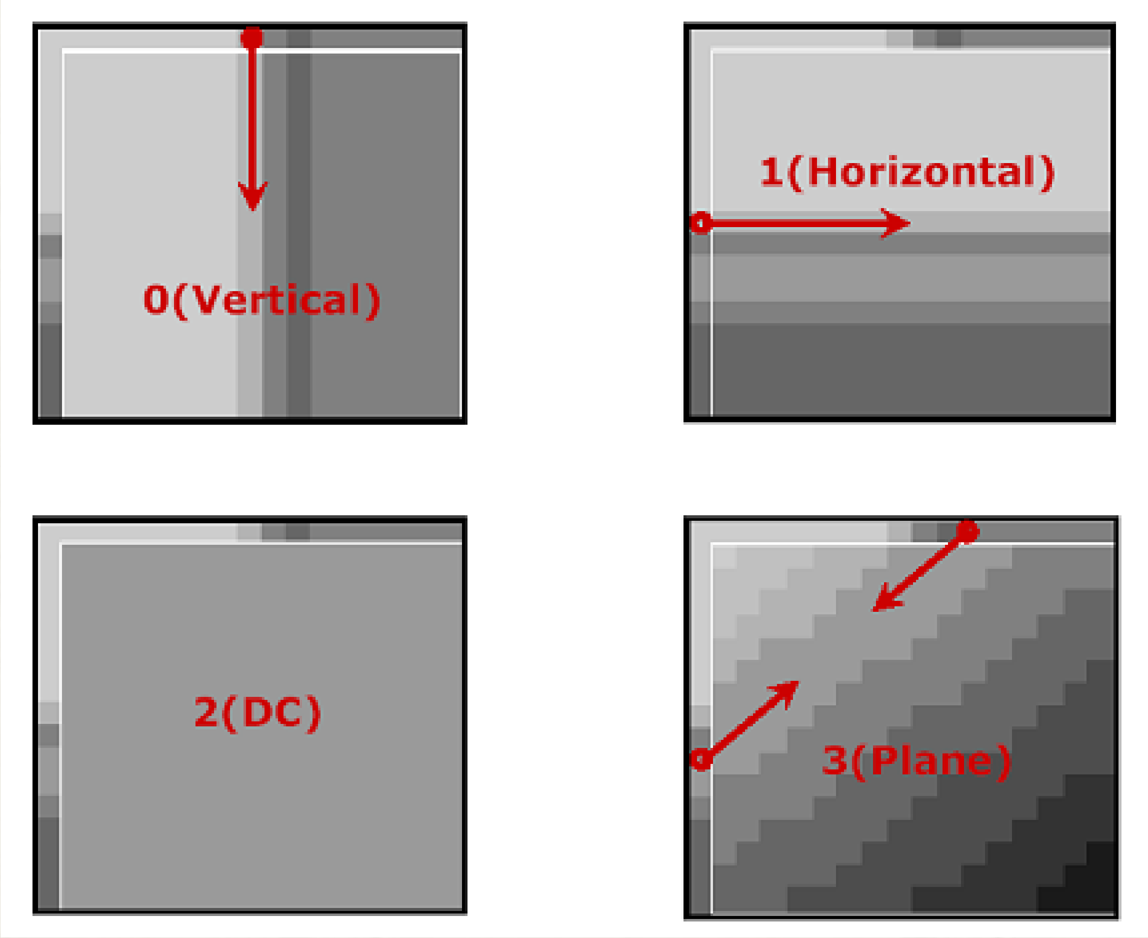
4x4 Intra Block Prediction Modes

- Nine 4x4 block prediction modes

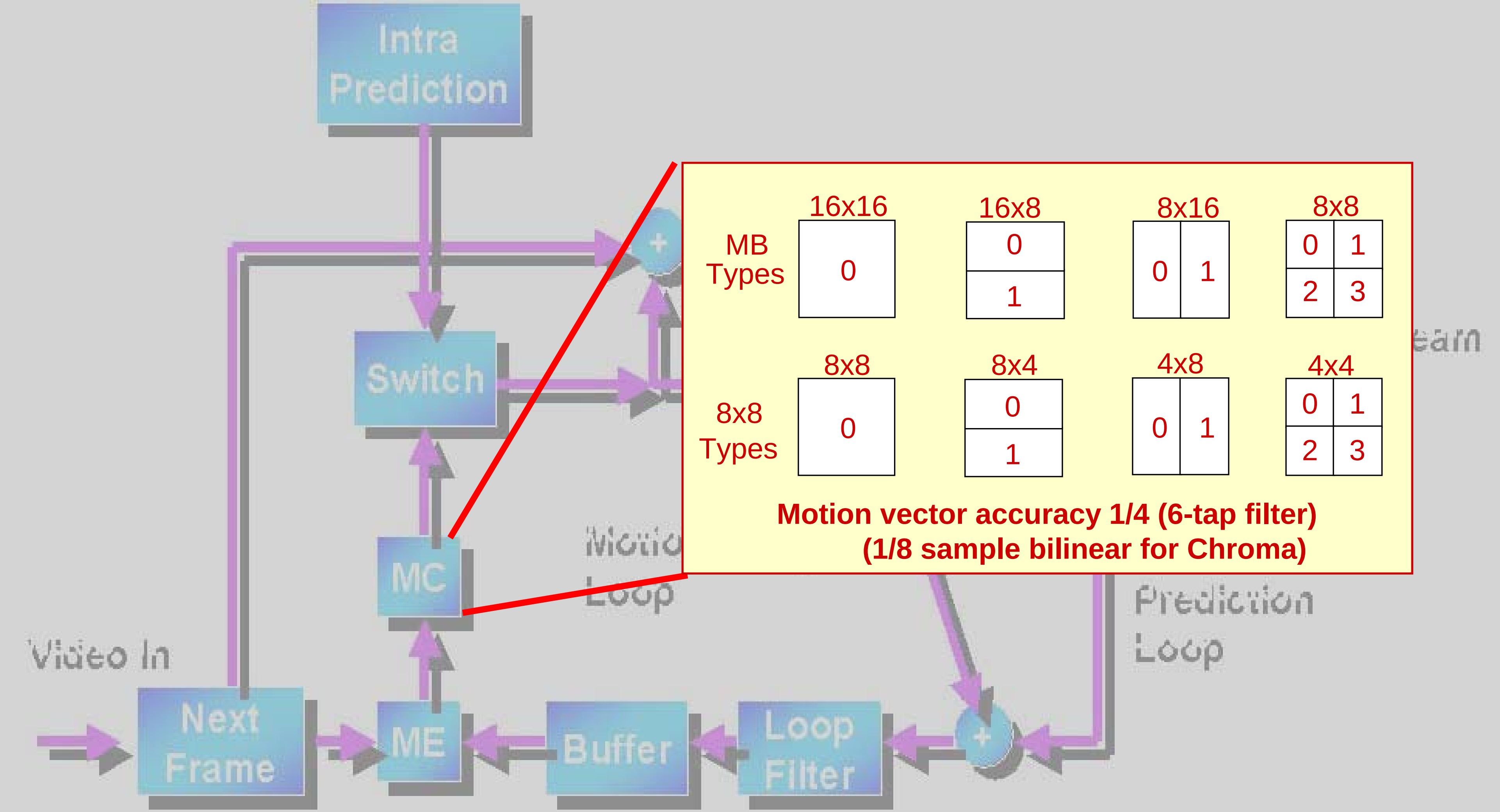


16x16 Luma (8x8 Chroma) Intra Prediction

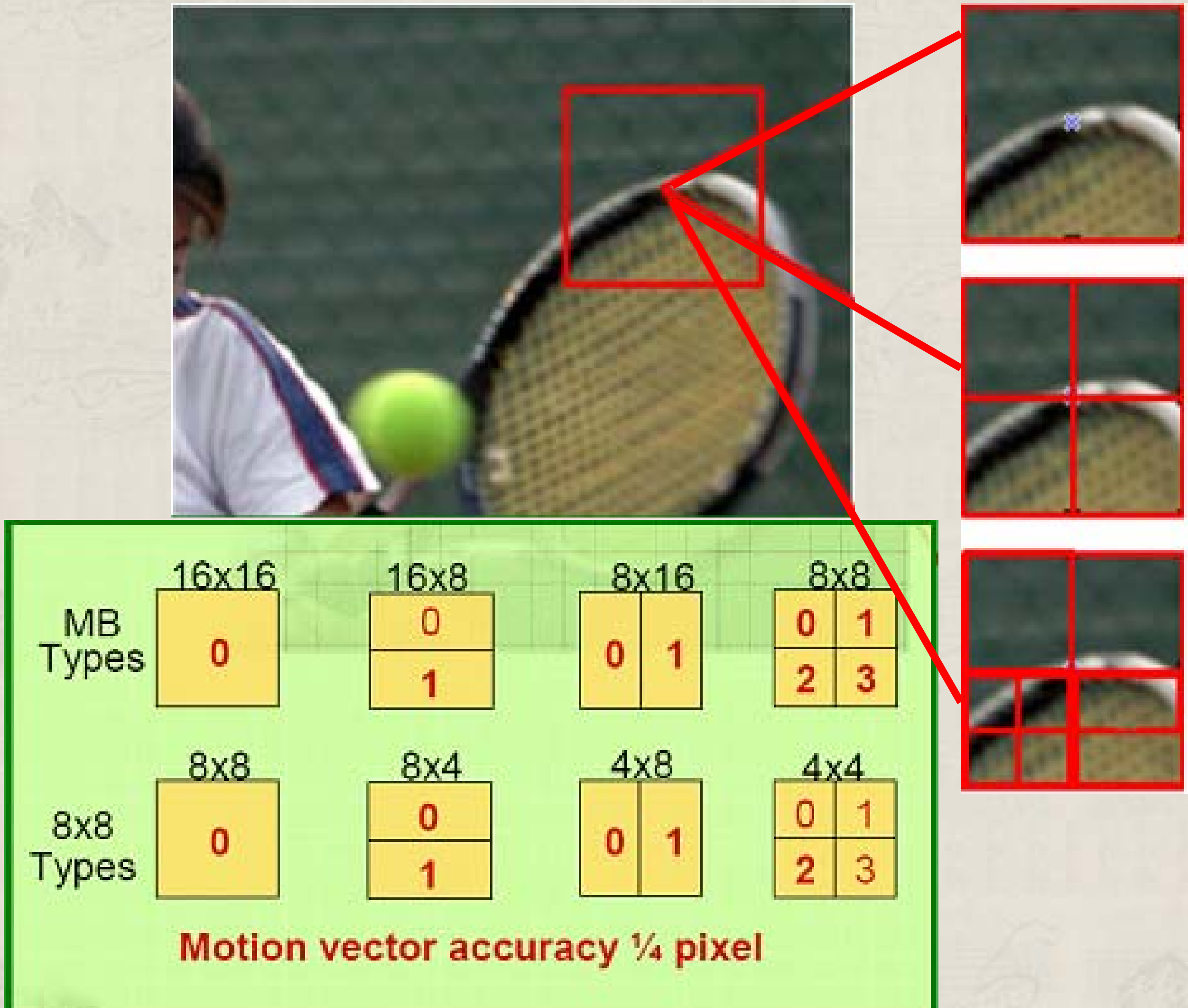
- Four 16x16 Luma (8x8 chrominance) intra predication modes



Innovation 3: Flexible Block MC



Example: H.264 MC

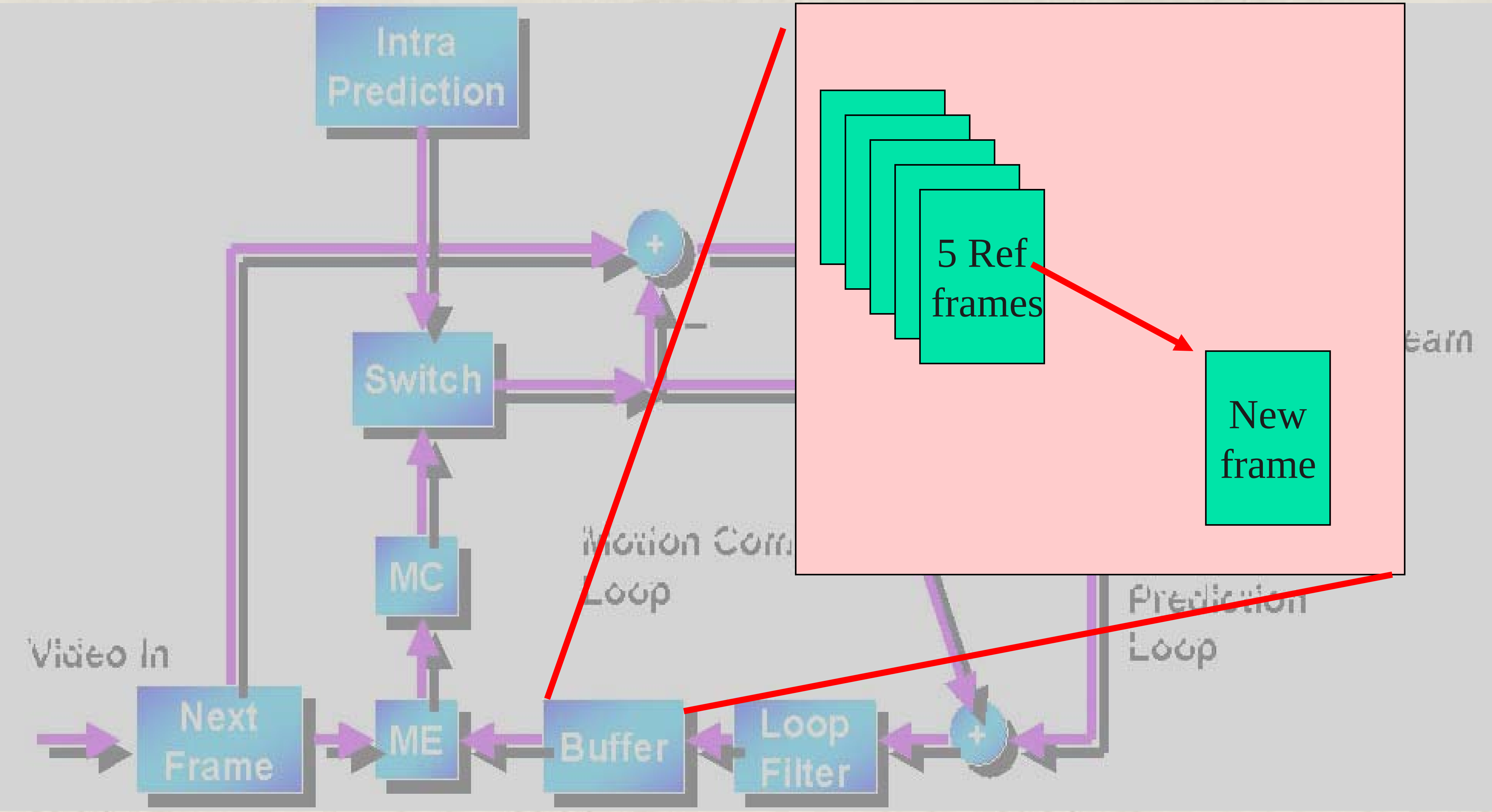


MPEG-2
- 16x16 block size
- Square shape
- ½ pel motion vector
- → Weak Motion Isolation !

MPEG-4
- 8x8 block size
- Square shapes
- ½ pel motion vector
- → Moderate Motion Isolation !!

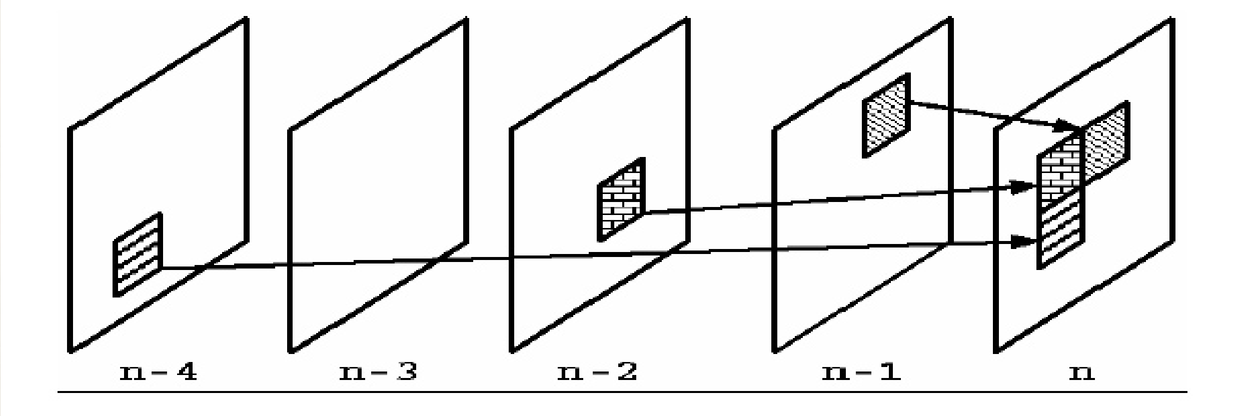
H.264
- 4x4 block size
- Arbitrary shapes
- ¼ pel motion vector
- → Strong Motion Isolation !!!

Innovation 4: Multiple Reference Frames

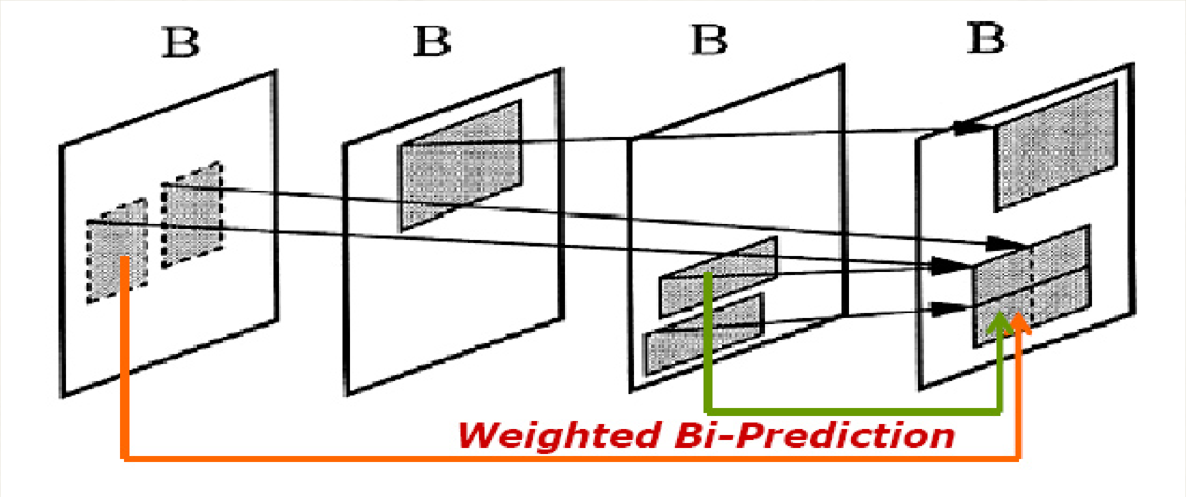


Multiple Reference Frames

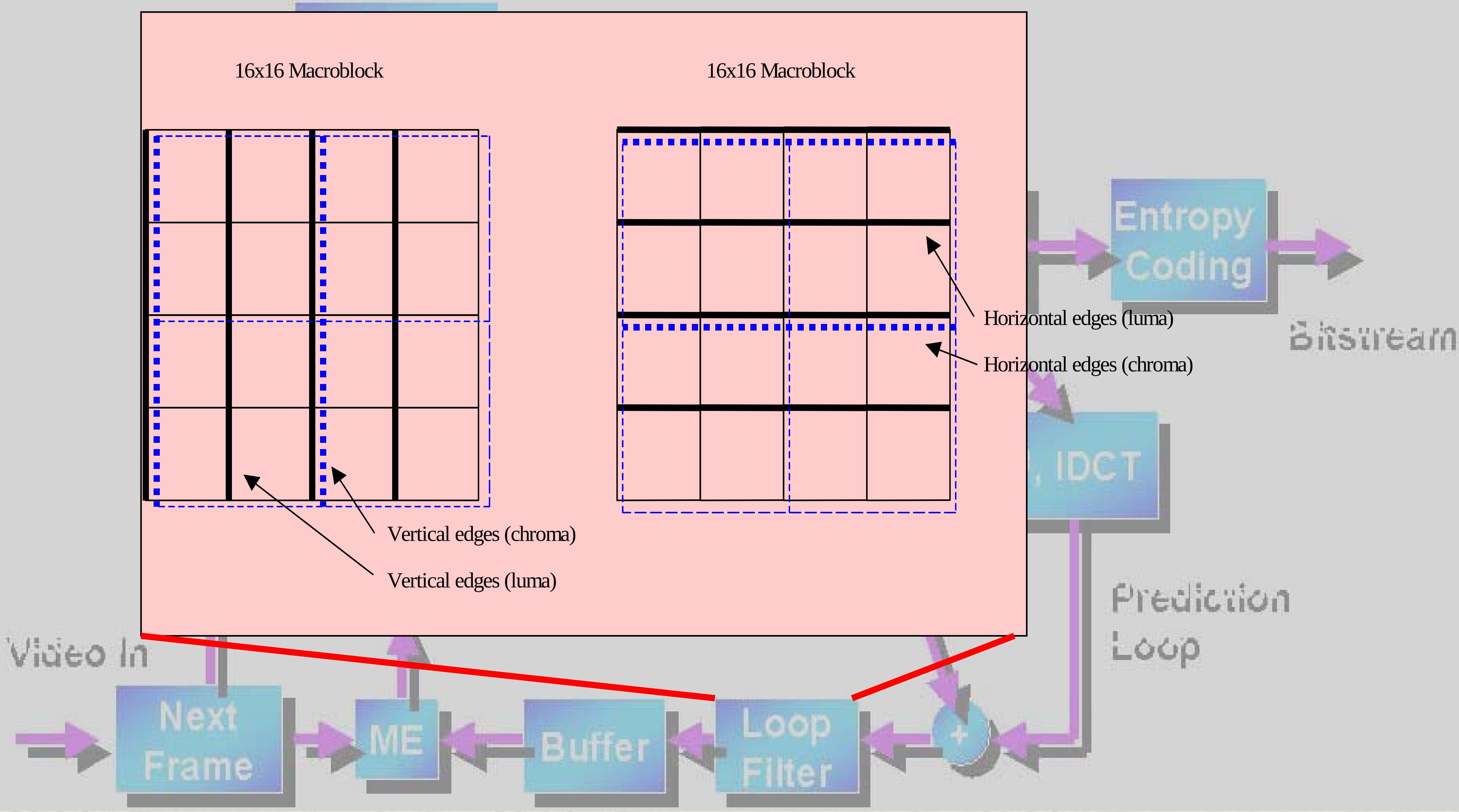
- Reference blocks



- Weighted bi-prediction



Innovation 5: In-Loop Deblocking



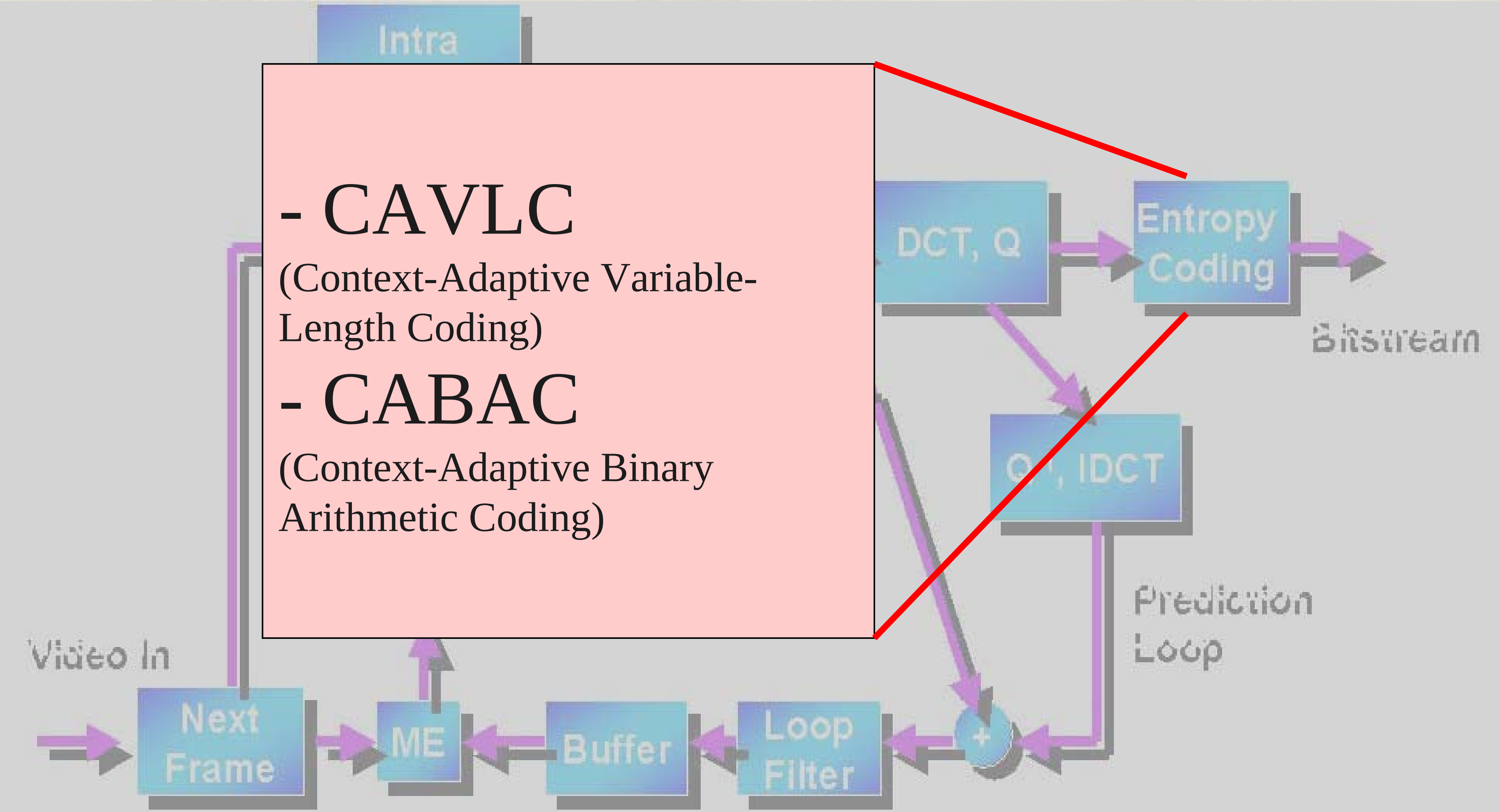
In-Loop Deblocking Filter

- Improves subjective visual quality
- Much better than out-of-loop post-filtering
- Highly context adaptive

Without loop filter

With H.264/AVC loop filter

Innovation 6: Two Entropy Coding Methods



- CAVLC

(Context-Adaptive Variable-Length Coding)

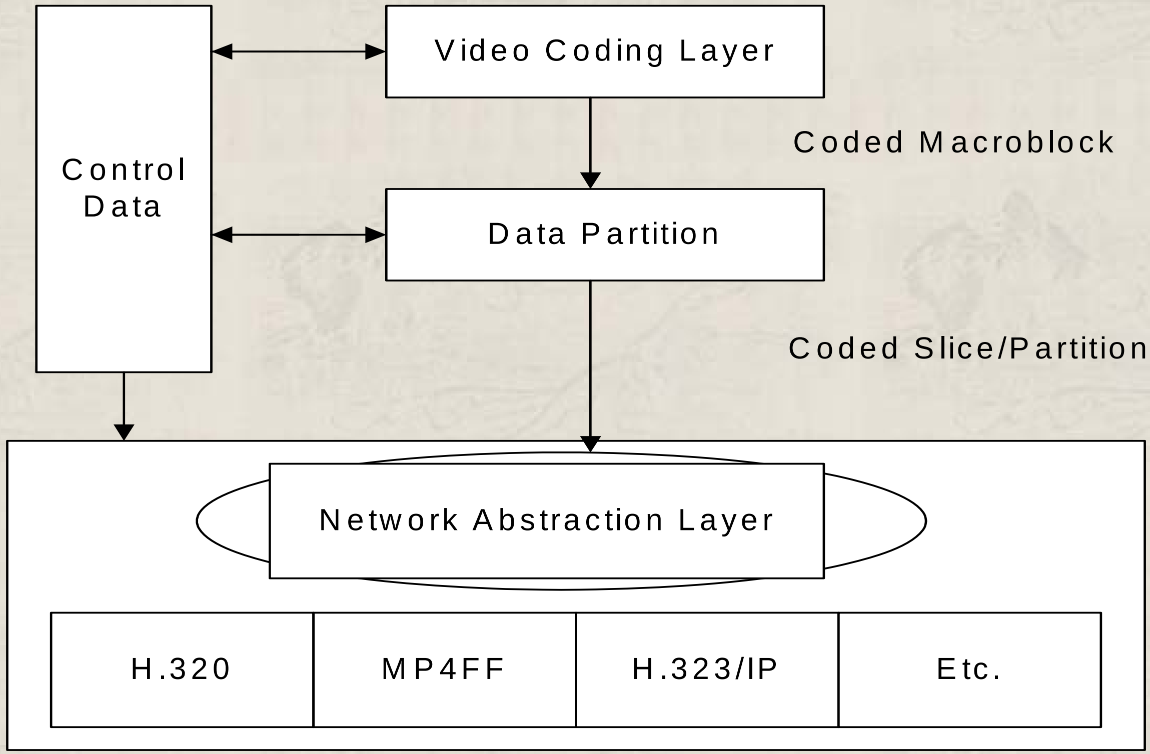
- CABAC

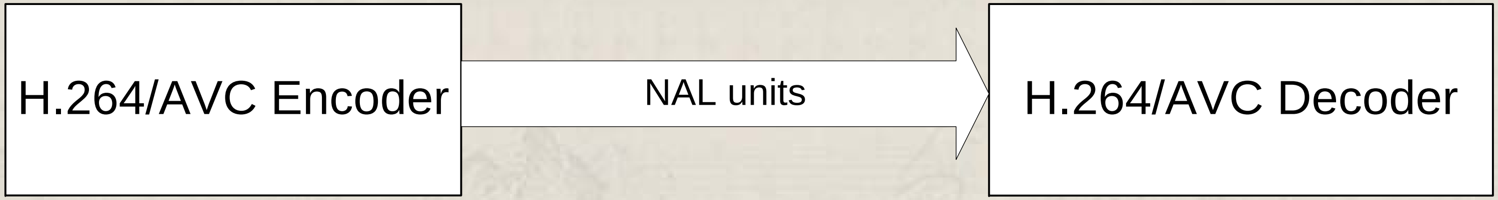
(Context-Adaptive Binary Arithmetic Coding)

H.264 Entropy Coding

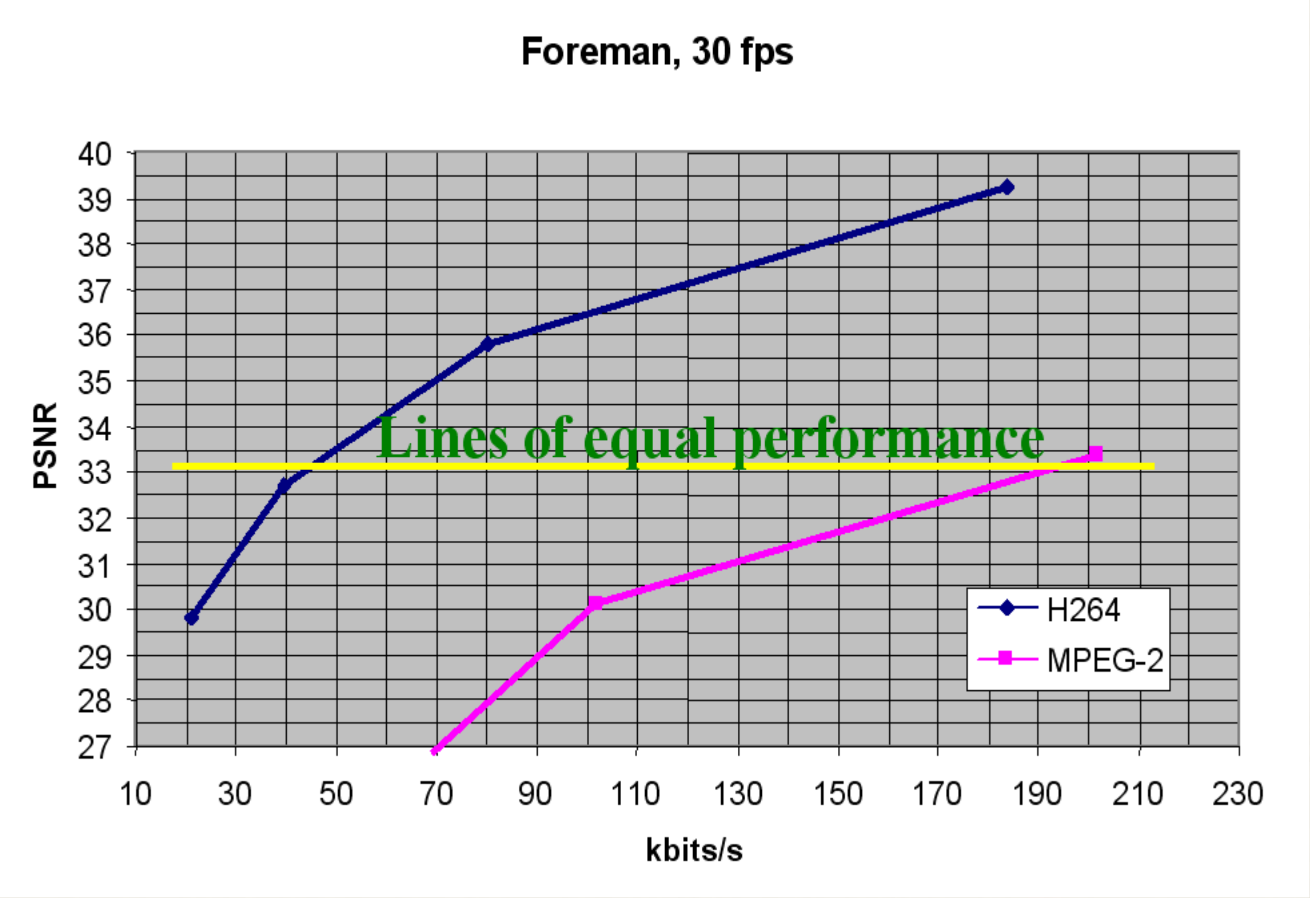
- Exp-Golomb Code
  - For all symbols except transform coefficients
  - Variable length codes with a regular construction, e.g.,
    0 -> 1; 1-> 010; 2 -> 011; 3 -> 00100; 4 -> 00101; 5 -> 00110
    6 -> 00111; 7 -> 0001000; 8 -> 0001001 …
- CAVLC (Context adaptive VLC)
  - For transform coefficients
  - No end-of-block, but the number of coefficients is encoded
  - Coefficients are scanned backwards
  - Contexts are built dependent on transform coefficients
- CABAC (Context-based binary arithmetic coding)
  - For transform coefficients
  - Uses adaptive probability models for most symbols
  - Exploiting symbol correlations by using contexts
  - Average bi-rate saving over CAVLC 10-15%

Innovation 7: Network Abstraction Layer





H.264 vs. MPEG-2: Low bit-rate (1)

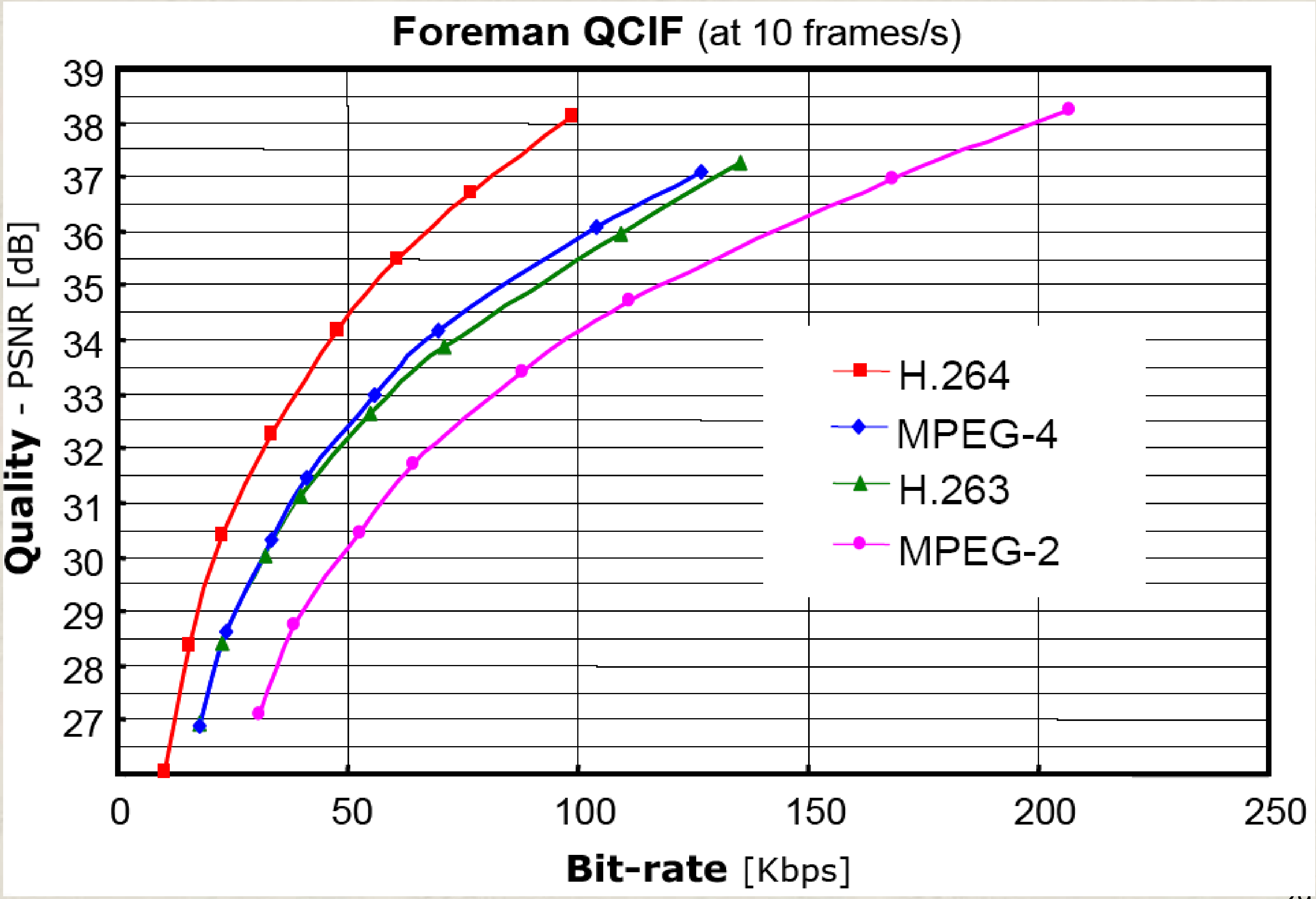


H.264 vs. MPEG-2: Low bit-rate (2)

MPEG-2 203kbps

H.264 39 kbps

Comparison to Other Standards



Basic H.264 Profiles

- Baseline (Video-conferencing & Wireless)
  - I and P frames (no B frame)
  - Interlace
  - Adaptive frame/field
  - In-loop deblocking filter
  - ¼ -sample motion compensation
  - Variable block motion estimation
  - CAVLC
  - Some error resilience features, e.g., ASO, FMO
- Main profile (Broadcast)
  - All baseline features except enhanced error resilience features
  - B frame
  - CABAC
  - MB-level frame/field switching
  - Adaptive weighting for B and P picture prediction

Enhanced H.264 Profiles

- Extended Profiles (Streaming)
  - Main profiles + Error resilience - CABAC
  - More error resilience: data partition
  - SP/SI switching pictures
- High profile
  - Old name: *Fidelity-Range Extensions* (**FRExt**)
  - Main profile
  - Switchable 8x8 transform
  - Scaling matrix for subjective quality optimization
  - Implementation beyond Main Profile affects Intra prediction, transform, deblocking filter control, CABAC decoding

High Profile

- H.264/AVC standard finished 2003
  - ITU-T/H.264 finalized May, 2003
  - MPEG-4 AVC finalized July, 2003
- High profile
  - Initiated in July 2003 and finished in July 2004
  - Motivation: higher quality and higher rates
  - Consider more than 8 bits sequences, and various color spaces
  - Improved coding efficiency (bit-rate reduction): e.g., 12% for HD films and progressive HD video
  - Complexity issues:
    - No increase in computational requirements
    - Slight increase in memory requirements (CABAC, transform)
  - *No reason not to move to High profile !*

New Features in High Profile

- Larger transforms
  - 8x8 transform
  - Drop 4x8, 8x4, and larger transforms
- Quantization matrix
  - 4x4, 8x8, intra, inter trans. coefficients weighted differently
  - Full capabilities not yet explored (visual weighting)
- Coding in various space
  - 4:4:4, 4:2:2, 4:2:0, and monochrome
  - New integer color transform
- Efficient lossless interframe coding
- Film grain characterization for analysis/synthesis representation
- Stereo-view video support
- De-blocking filter display preference

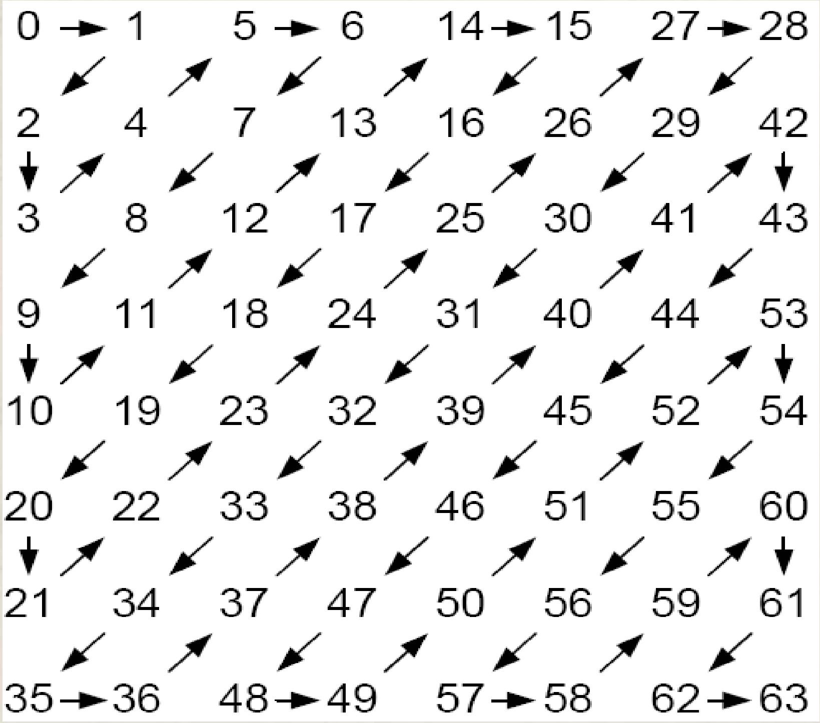
8x8 16-bit Transform

$$\begin{bmatrix} 8 & 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 12 & 10 & 6 & 3 & -3 & -6 & -10 & -12 \\ 8 & 4 & -4 & -8 & -8 & -4 & 4 & 8 \\ 10 & -3 & -12 & -6 & 6 & 12 & 3 & -10 \\ 8 & -8 & -8 & 8 & 8 & -8 & -8 & 8 \\ 6 & -12 & 3 & 10 & -10 & -3 & 12 & -6 \\ 4 & -8 & 8 & -4 & -4 & 8 & -8 & 4 \\ 3 & -6 & 10 & -12 & 12 & -10 & 6 & -3 \end{bmatrix} \cdot \frac{1}{8}$$

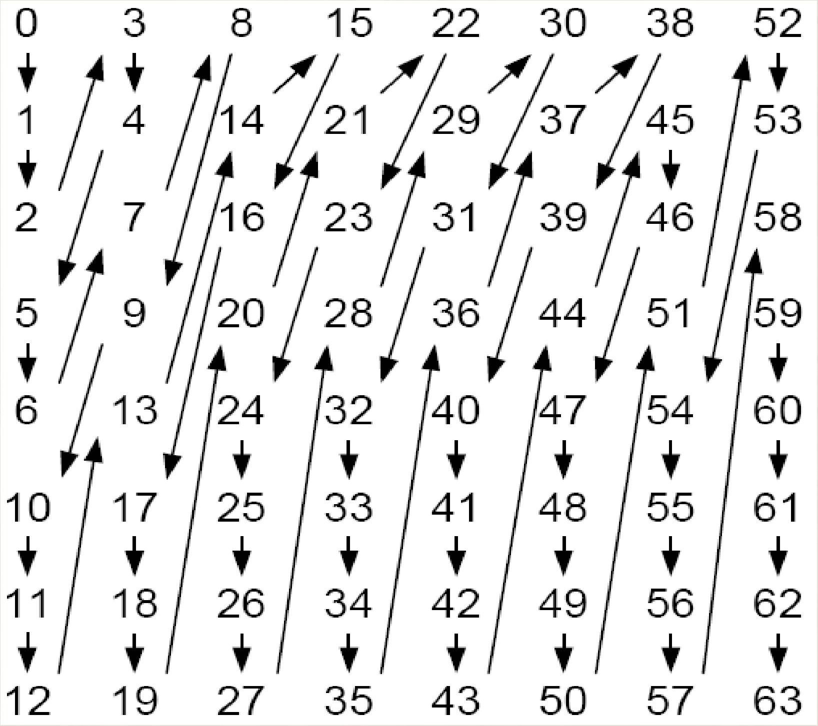
- Computational complexity
  - One 8x8 block has the same number of adds (64) and 4 extra shifts (20 vs. 16) compared with four 4x4 transform

8x8 Transform Coefficients Scan

- Two Scans
  - Different scan for frame/field coding



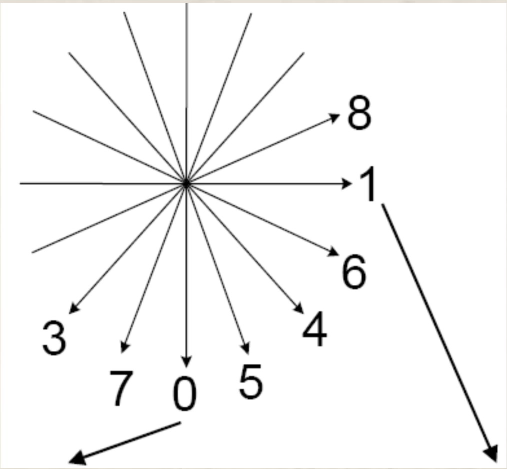
Frame scan



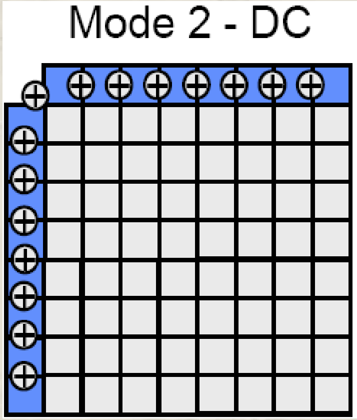
Field scan

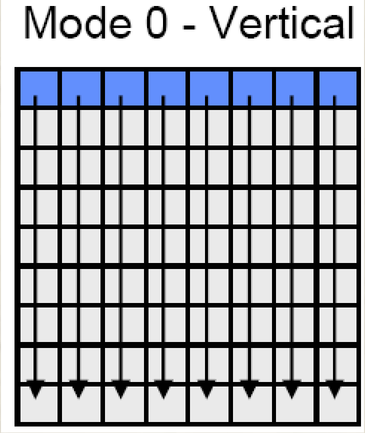
8x8 Intra Block Prediction

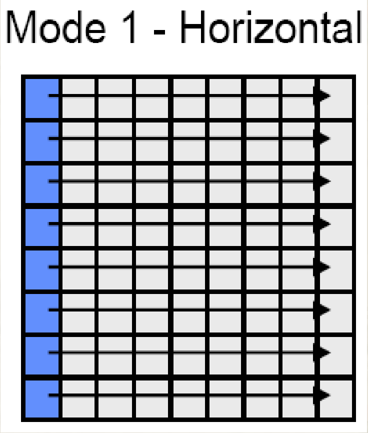
- Nine intra-prediction modes similar to the nine modes for 4x4 block prediction



&







Quantization Matrix

- Similar concept to MPEG-2 design
- Vary step size based on frequency
- Adapted to modified transform structure
- More efficient representation of weights
- Separate matrix for inter and intra
- Matrix can be included in picture/slice head information
- Eight downloadable matrices (at least for 4:2:0)
  - Intra 4x4 Y, Cb, Cr
  - Intra 8x8 Y
  - Inter 4x4 Y, Cb, Cr
  - Inter 8x8 Y

Reversible Integer Color Transform

- Color transform for YUV

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} .299 & .587 & .114 \\ -.147 & -.289 & .436 \\ .615 & -.515 & .100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

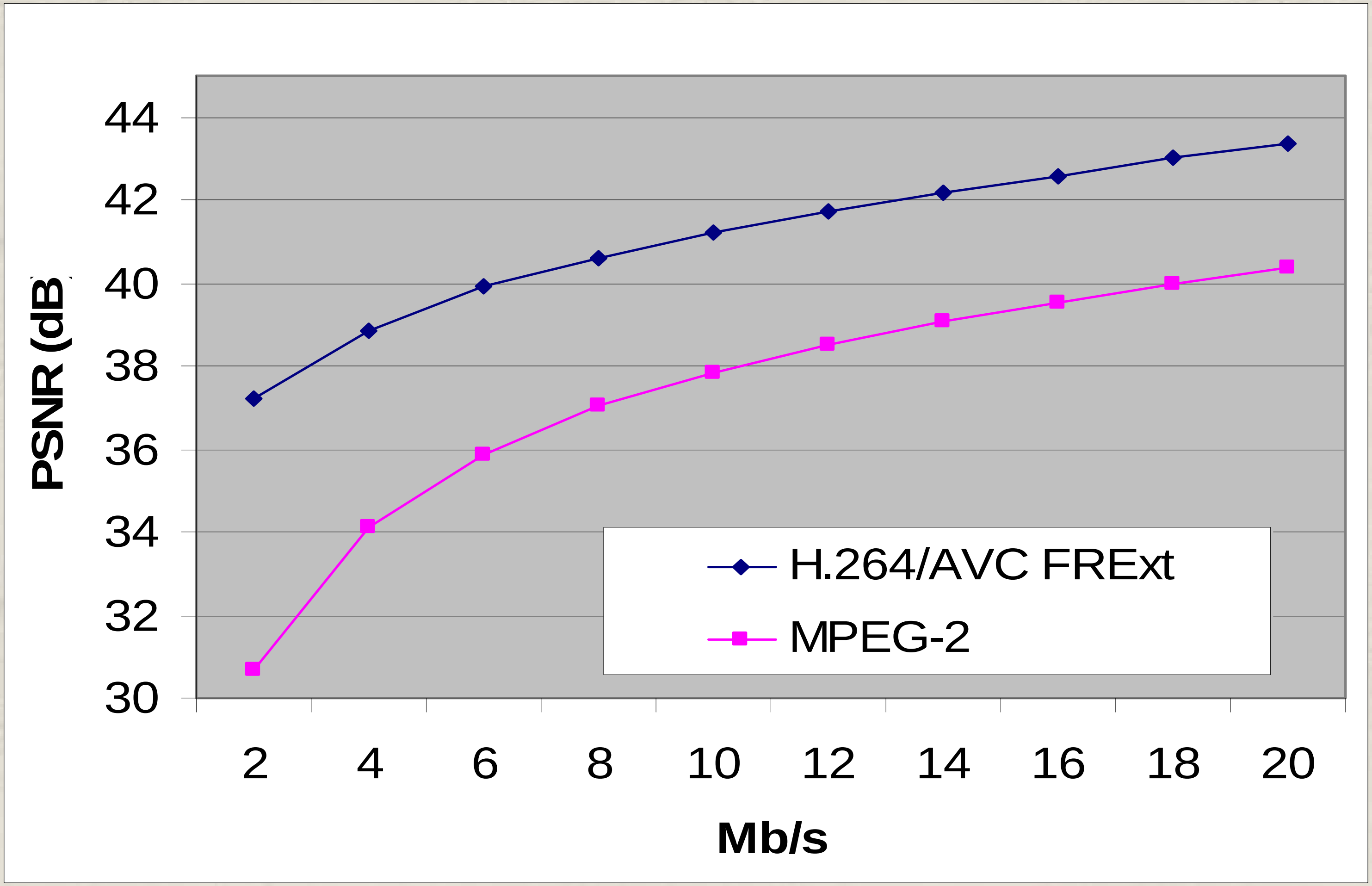
- Integer color transform (YCoCg)

$$\begin{bmatrix} Y \\ Co \\ Cg \end{bmatrix} = \begin{bmatrix} 1/4 & 1/2 & 1/4 \\ 1 & 0 & -1 \\ -1/2 & 1 & -1/2 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Other High Profile Details

- Deblocking filters:
  - Only control of filter is adjusted: do no filter for 4x4 blocks
  - Filter operation itself does not change
- CABAC
  - 61 contexts and their corresponding initial values
  - No change to CABAC engine
- Information signaling
  - 8x8 transform on/off flag at the picture head information
  - 8x8 transform on/off flag at per macroblock allows adaptive use

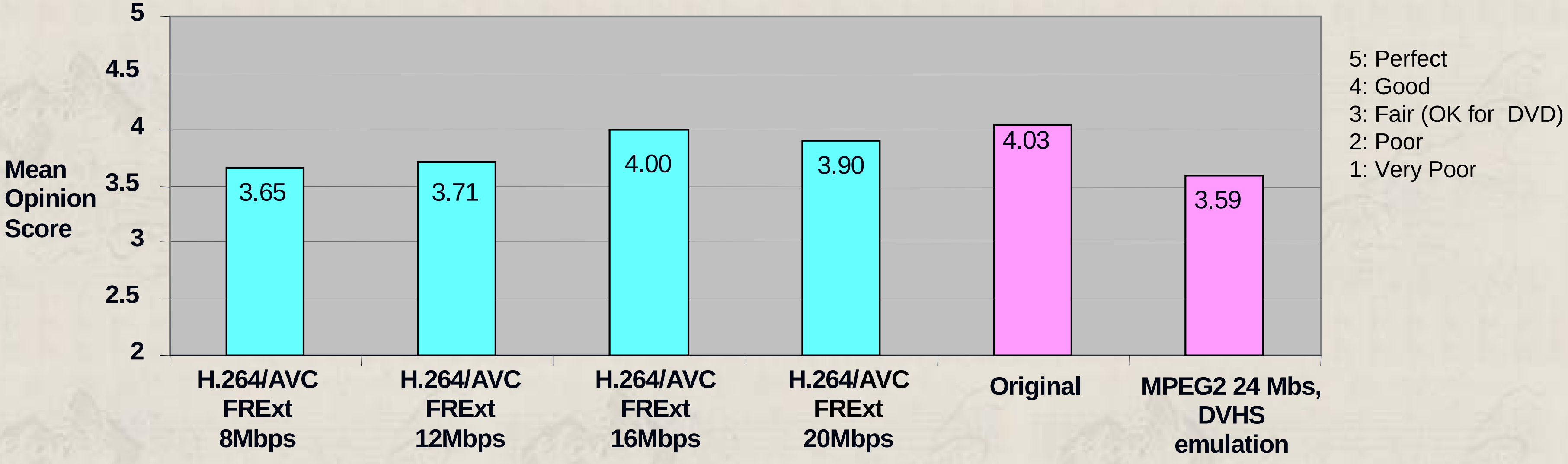
H.264 High Profile vs. MPEG-2



BigShip HD sequence (1280x720, 720p)

Subjective Performance *

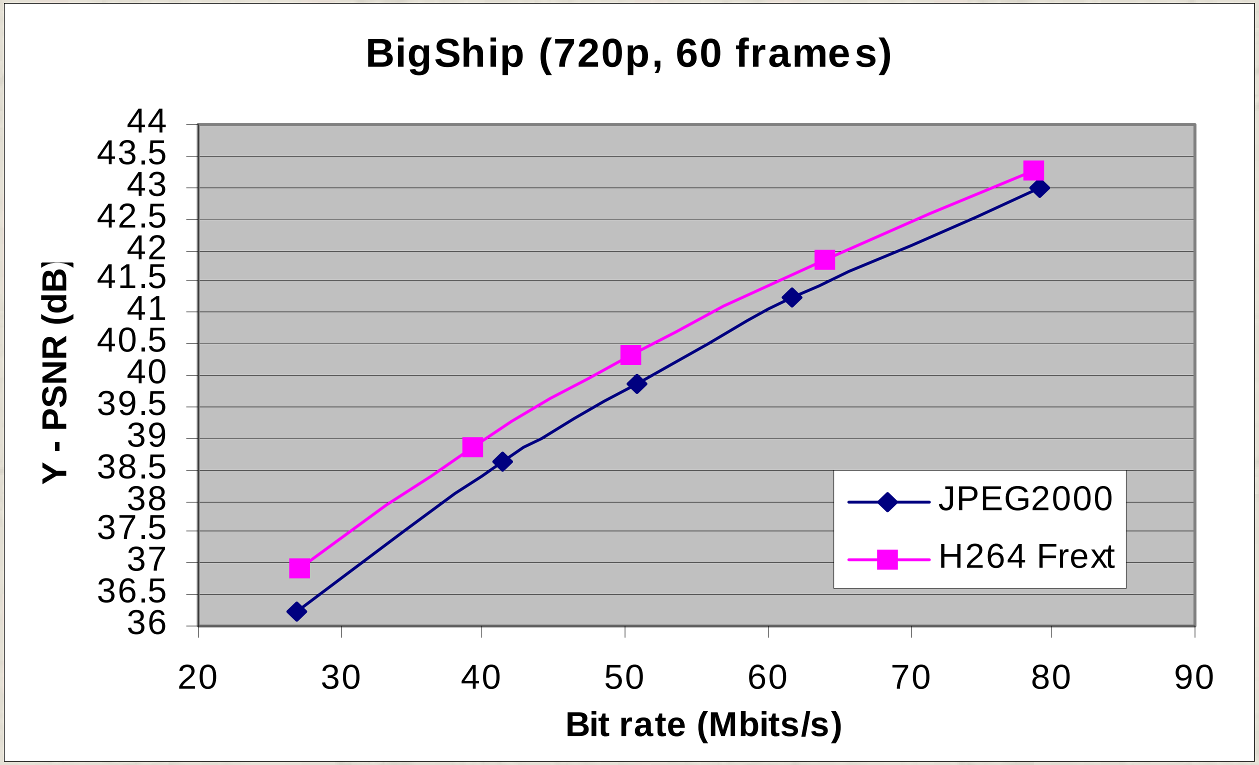
- Subjective tests by Blu-Ray Disk Founders of FRExt HP
  - 4:2:0/8 (HP) 1920x1080x24p (1080p), 3 clips.
  - Notional 3:1 advantage to MPEG-2
    - 8 Mbps HP scored better than 24 Mbps MPEG-2!
  - Apparent transparency at 16 Mbps!



*JVT-L033, M1116, Draft JVT Redmond report

High Profile I-Frame Coding vs. JPEG2000

- High profile I frame coding with RD-optimization model selection
- RD-optimized JPEG2000 coder used



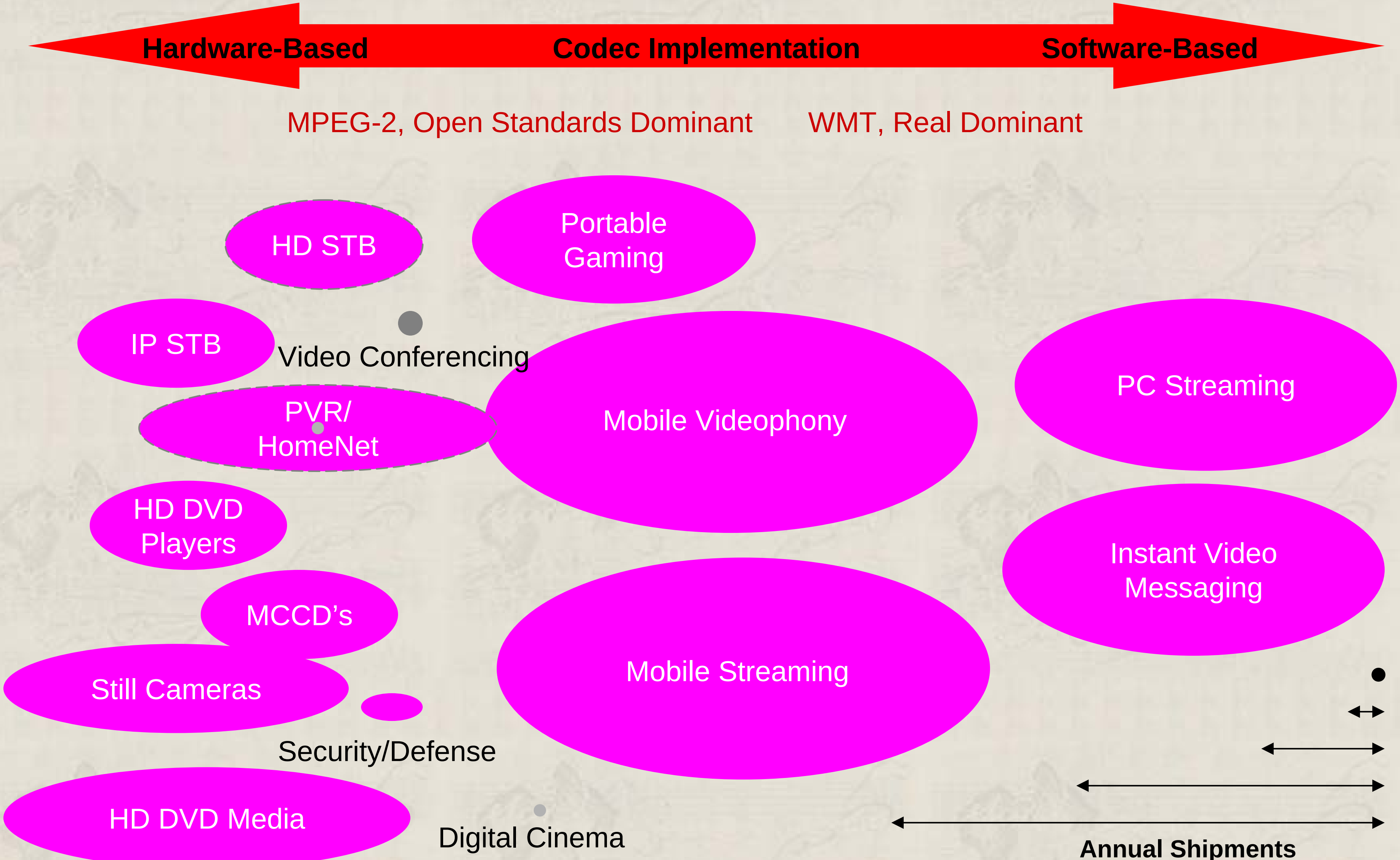
Challenging Problems

- Major problem: reduce the computational complexity without sacrificing the performance
  - Motion estimation
    - Fast motion search
    - Reference frames selection
  - Macroblock mode decisions
    - Seven inter modes, intra mode with prediction
    - Try all and select the best?
    - Mode decision criterion needed
  - Etc.
- Implementation issues
  - Read time H.264 encoding and decoding
  - Hardware implementations
  - Etc.

Applications and Markets

- Storage
  - Video CD, DVD, Hard Disk, Web publishing
- Broadcast
  - Satellite, Cable, Terrestrial
- Conversational
  - Video-conferencing, Cell phones, PDAs
- Streaming
  - Video-on-demand, music video, streaming ads
- Future Applications! – unknown

H.264 Opportunities Map



Example: HD DVD Multimedia

- With H.264, put 2 hours of HD on DVD-9
  - Note: a 100-min HD movie fits in 8.25 GB @ 11 Mb/s
- Keep MPEG-2 skin
  - Systems, audio… minor change to DVD player
  - Small cost, big quality jump
- Even better with blue-ray when ready
  - Tech is “laser-agnostic”
- Studios can recycle catalog in HD
  - Double the money!!

Source: DVD-FAQ (Jim Taylor)

Format	Laser	Density	Data Rate	Encoding	Supporters
HD-DVD9	Red	9 GB	6-11 Mb/s*	H264/WM9	Warner
AOD	Blue	15 GB	10-20 Mb/s	MPEG2, …	Toshiba/NEC
Blue 1	Blue	17 GB	25 Mb/s	H264?	ITRI (Korea)
Blue-Ray	Blue	27 GB	10-30 Mb/s	MPEG2, …	LG, Philips, Pan., Sony, Sharp…
Blue 2	Blue	17 GB	10-30 Mb/s	H264?	Matsushita(Panosonic)?

H.264/AVC Organization Adoptions

- ITU-T systems adoption completed
- MPEG-2 and MPEG-4 systems & file format adoption completed
- IETF WG last call for RTP payload
- 3GPP2
  - Adopted Baseline (restricted) for streaming and MMS
- HD DVD in DVD Forum: Mandatory player support
- Blue-Ray Disc Founders (BDF)
  - High Profile (HP) is their first choice beyond MPEG-2
- Digital Multimedia Broadcast in Rep. of Korea
- Mobile broadcast announcement in Japan
- France Terrestrial Broadcast announcement
  - H.264/AVC HD instead of MPEG-2
- Etc.

Companies Publicly Known to Implement H.264 Standard

- Ahead Software / ATEME
- Amphion
- Apple Computer
- British Telecom
- Broadcom / Sand Video (chips)
- Conexant (chipset for STB)
- Cradle
- Deutsche Telekom
- DG2L
- Dicas
- DSP Research / W&W Communications
- Emblaze Group
- Envivio
- Equator
- FastVDO
- France Telecom
- Hantro
- Harmonic (filtering and motion estimation)
- HHI (PC & DSP encode & decode; demos)
- i3 Micro Technology
- iVast
- Intel
- KDDI R&D Labs
- Ligos
- LSI Logic / Videolocus
- Mainconcept
- Mcubeworks
- Media Excel
- Mobile Video Imaging
- Mobilygen
- Modulus Video (main profile levels 3 & 4 b'cast encoders & professional-use decoders)
- Moonlight Cordless
- Motorola
- Neomagic
- Nokia
- Oki Electric
- Optibase
- Packetvideo
- PixelTools
- PixSil Technology
- Polycom (videoconferencing & MCUs)
- Prodys
- Radvision (videoconferencing)
- Richcore
- Samsung (Terrestrial DMB receiver)
- Scientific Atlanta
- Setabox
- SkyStream Networks
- Sony (encode & decode, software & hardware, including PlayStation Portable 2004 & videoconferencing systems)
- ST Micro (decoder chip in '03)
- Tandberg (shipping with all videoconferencing endpoints since July '03, GW and MCU since Oct.)
- TandbergTV
- Tektronix
- Techno Mathematical
- Telesuite
- thin multimedia
- Thomson
- TI (DSP partner with UBV for one of two UBV real-time implementations)
- Toshiba
- Tuxia
- UB Video (demoed real-time encode and decode, software and DSP implementations)
- Videosoft / Vanguard Software Solutions (s/w, enc/dec)
- VideoTele.com (a division of Tut Systems)
- VCON
- Vqual
- W&W Communications / DSP Research

CAUTION: This information should be considered preliminary and should not be considered to be product announcements – only preliminary implementation work. It may be a while before robust interoperable implementations are well-established.

References

- IEEE Transactions on Circuits and Systems for Video Technology, July 2003.
- http://www.vcodex.com/h264.html
- ftp site: http://bs.hhi.de/~suehring/tml/
- P. Topiwala, H.264/AVC: Overview and Introduction to Fidelity-Range Extensions, http://www.fastvdo.com
- T. Wiegand, S. Gordon, A. Luthra: H.264/AVC High Profile, Presented to DVB, Sept 2004
- H.264 Overview, AddPac Tech. Co. Ltd.
- JVT-L033, M1116, Draft JVT Redmond report
- G. Sullivan, P. Topiwala, and A. Luthra, The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions, SPIE Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard: H.264/AVC, August, 2004
- L. Liu, P. Topiwala, P. Rault and T. D. Tran, Comparison of JPEG2000 with H.264/AVC FRExt I - Frame Coding on 720p Video Sequences, JVT-N010, Jan. 2005
- Google H.264