Lecture 3: Image & Video Coding Techniques (II) & Standards (I)

A/Prof. Jian Zhang

NICTA & CSE UNSW
COMP9519 Multimedia Systems
S2 2009
jzhang@cse.unsw.edu.au

THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY · AUSTRALIA

NICTA

3.1 Subband Coding

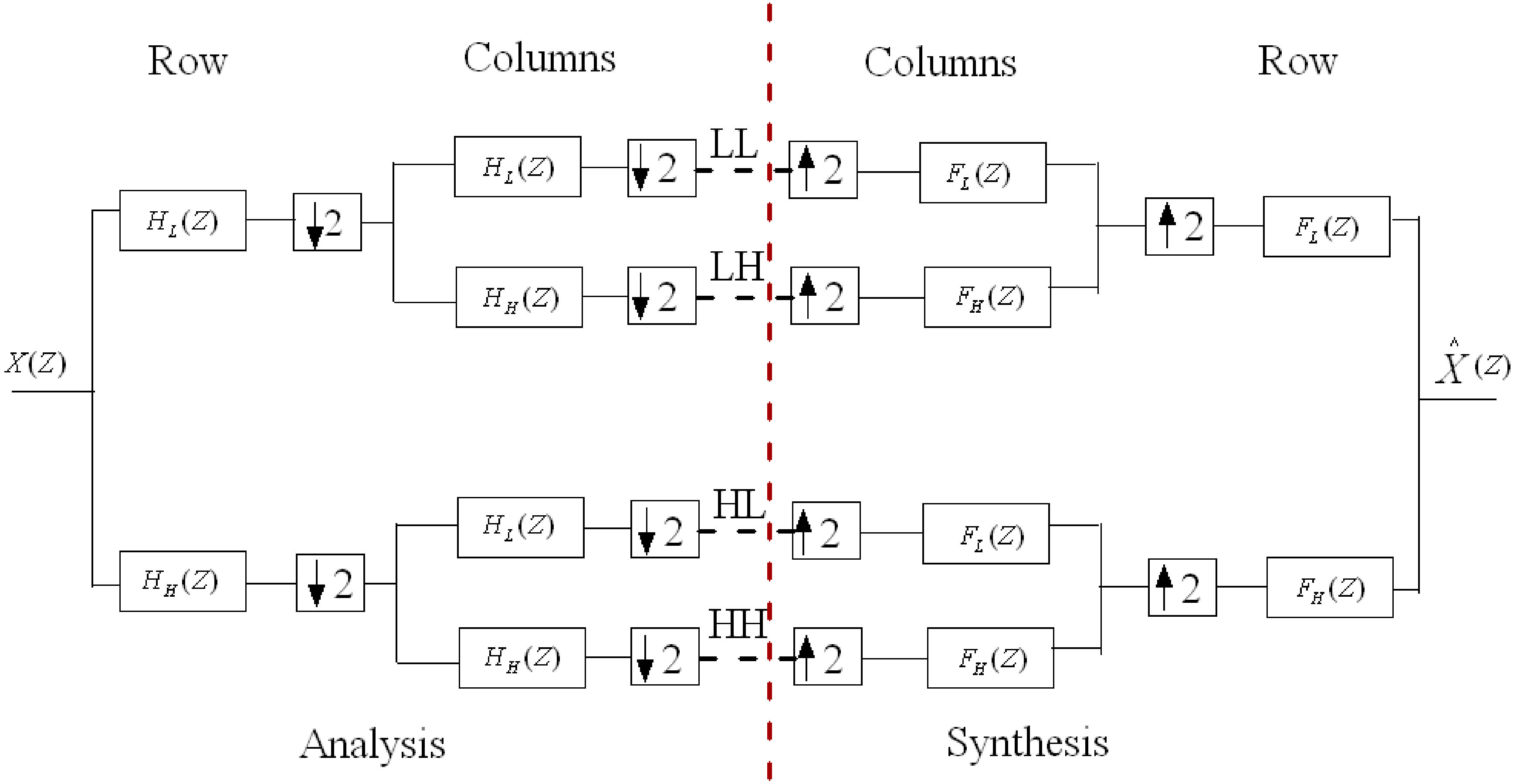
- The fundamental concept behind Subband Coding is to split up the frequency band of a signal and then to code each subband using a coder and bit rate accurately matched to the statistics of the band
- This makes each subband at the lower sampling rate
- At the receiver, the subband are resampled, fed through interpolation filters and added to reconstruct the image
- With an appropriate choice of filters, perfect reconstructions can be achieved

3.3.1 Analysis/Synthesis Stages

- Analysis
  - A signal is first filtered to create a set of signals, each of which contains a limited range of frequencies. These signals are called subbands.
  - Since each subband has a reduced bandwidth compared to the original fullband signal, they may be downsampled. That is, a reduced number of samples may be taken of the signal without causing aliasing
- Synthesis-- Reconstruction is achieved by:
  - Upsampling the decoded subbands.
  - Applying appropriate filters to reverse the subbanding process.
  - Adding the reconstructed subbands together.

3.1 Analysis/Synthesis Stage

- **2D dimensional decomposition structure**

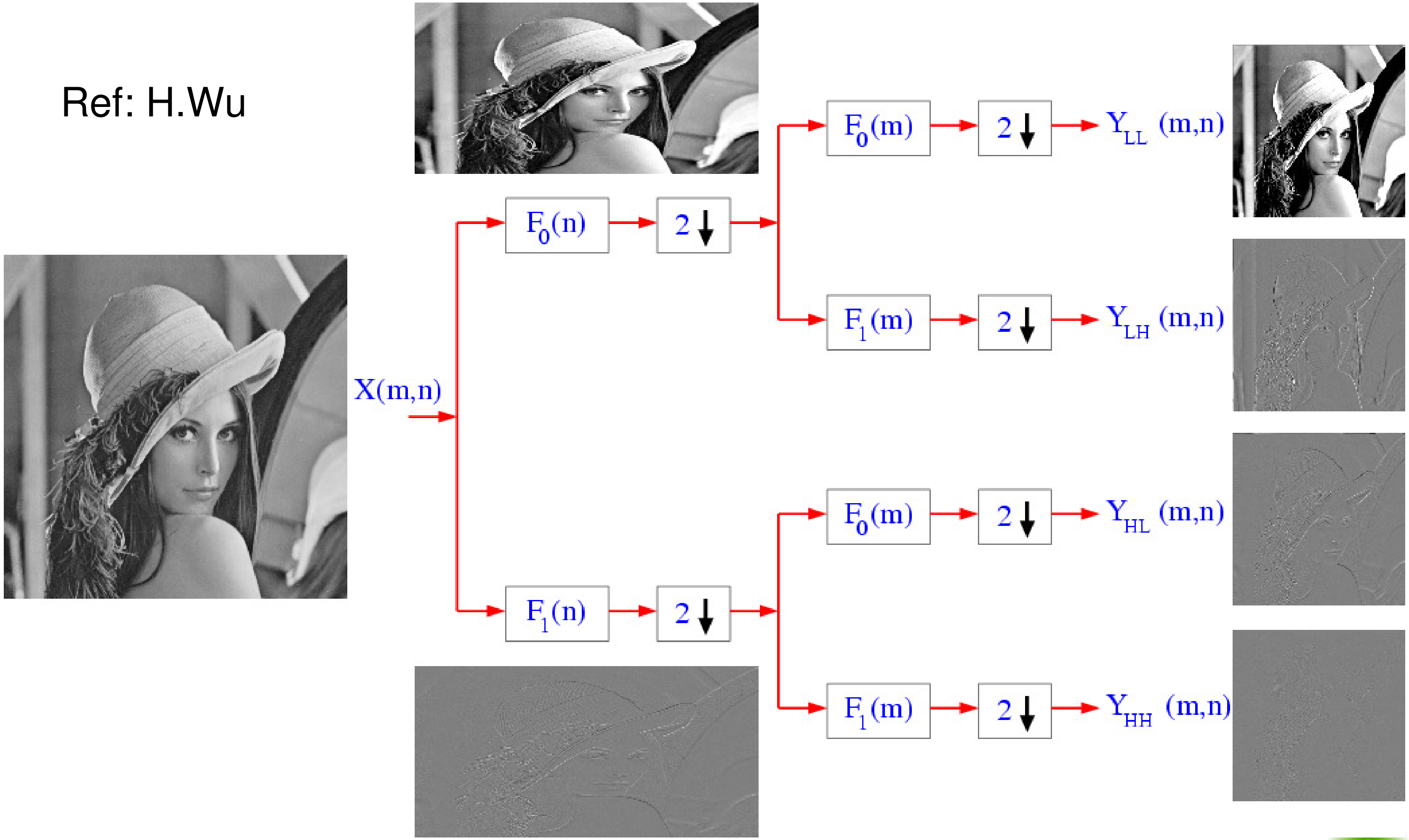


3.1 Analysis/Synthesis Stage

- The formation of subbands does not create any compression in itself.
  - The same total number of samples is required to represent the subbands as is required to represent the original signal.
- The subbands can be encoded efficiently:
  - The significance of the different spatial frequencies are not uniform, and this fact may be exploited by different bit allocations to the various subbands
  - The subbands are then encoded using one or more coders. Different bit rates or even different coding techniques may be used for each subband.

3.1 Analysis/Synthesis Stage

Ref: H.Wu



$X(m,n)$

$F_0(n)$ → $2\downarrow$

$F_0(m)$ → $2\downarrow$ → $Y_{LL}\ (m,n)$

$F_1(m)$ → $2\downarrow$ → $Y_{LH}\ (m,n)$

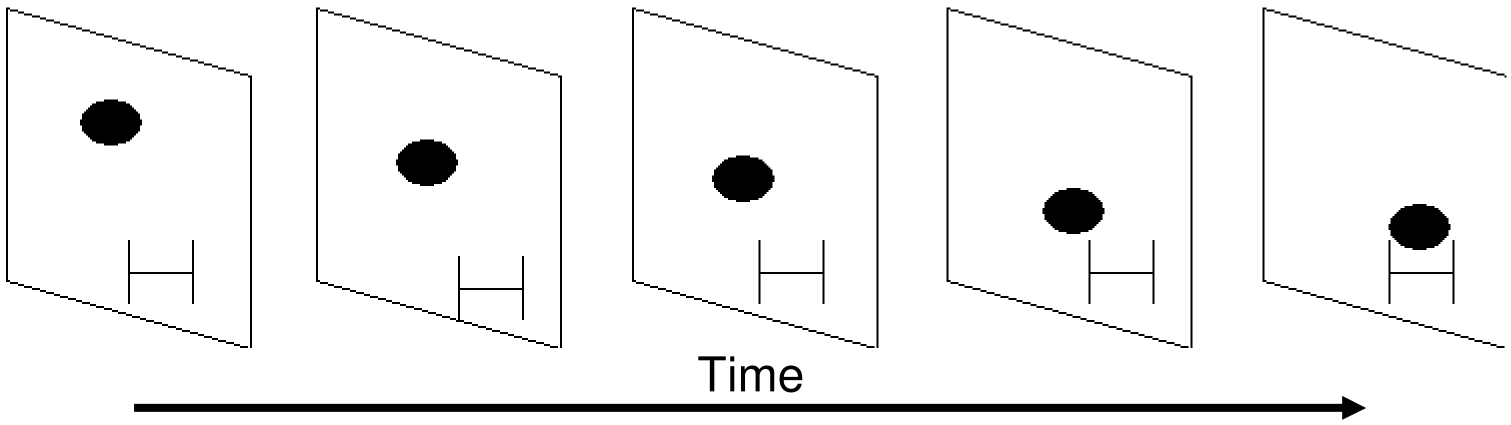
$F_1(n)$ → $2\downarrow$

$F_0(m)$ → $2\downarrow$ → $Y_{HL}\ (m,n)$

$F_1(m)$ → $2\downarrow$ → $Y_{HH}\ (m,n)$

3.2 Temporal Redundancy and Prediction for Video Coding

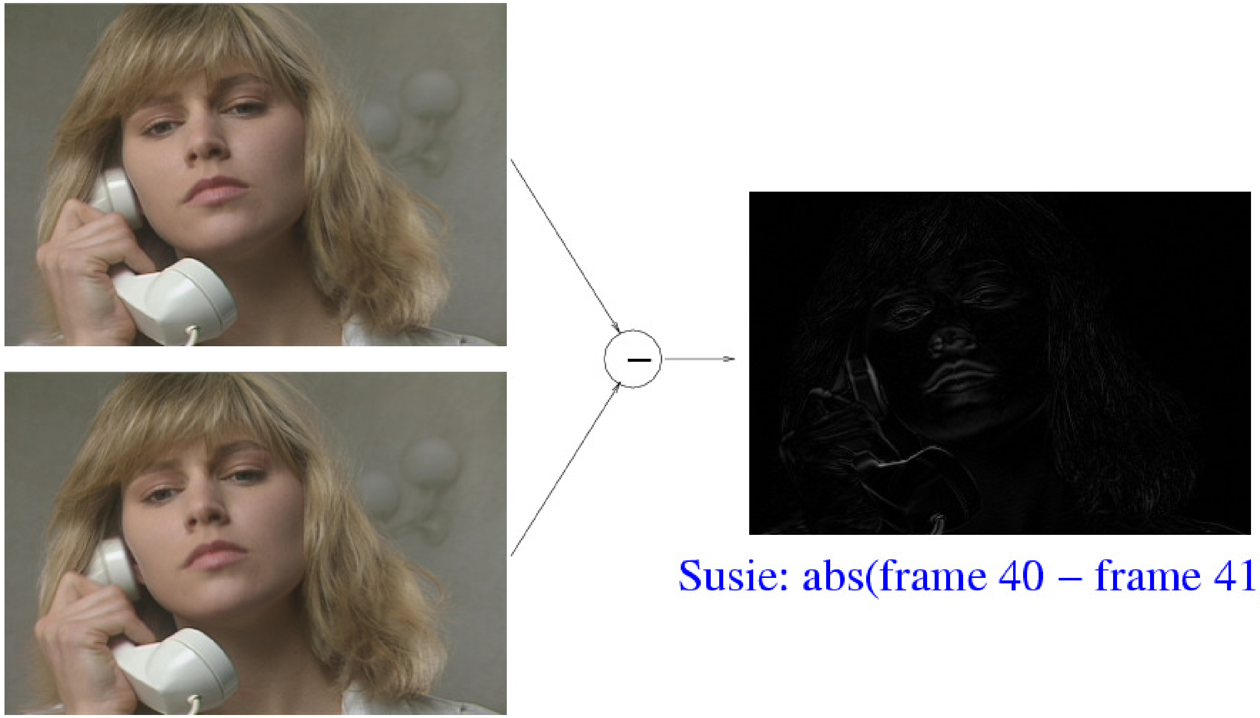
- In video coding, it is necessary to explore the compression in the spatial domain (spatial redundancy in intra-picture) and in time domain (temporal redundancy in inter-picture)



- A video sequence is a series of images with limited motion between adjacent images in most of the time

3.2 Temporal Statistical Redundancy

- Temporal redundancy between pixels of adjacent frames of video sequence
  - The pixel differences at the same spatial location between consecutive frames are typically small.



Susie frame 40

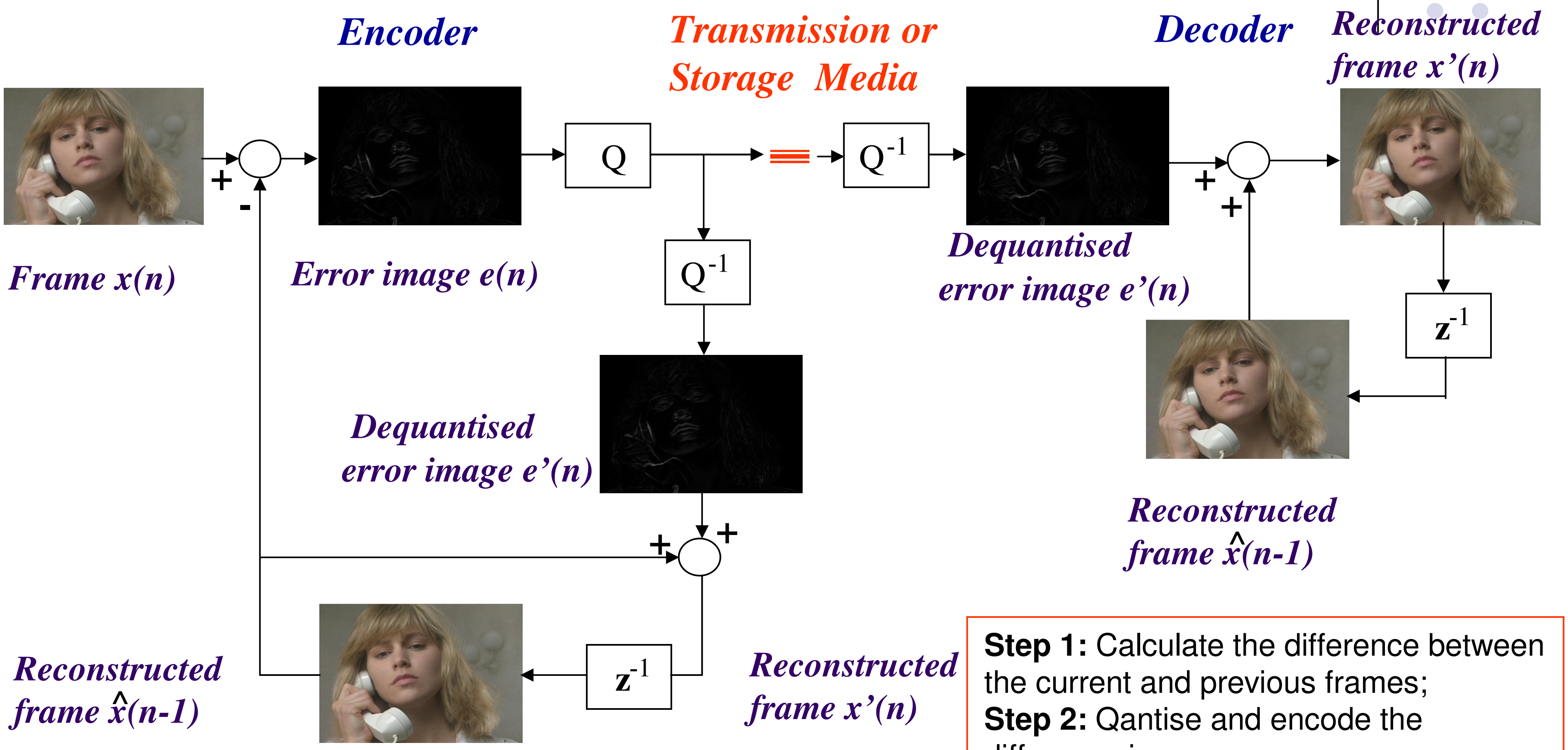
Susie frame 41

Susie: abs(frame 40 – frame 41)

3.2 Inter-frame Encoder

- Conditional Replacement
    - Inter-frame coders produce differential signals that are effectively zero in the non-changing parts of the picture and non-zero only in the moving areas
    - It is needed to transmit differential signal values only for the moving areas of the picture

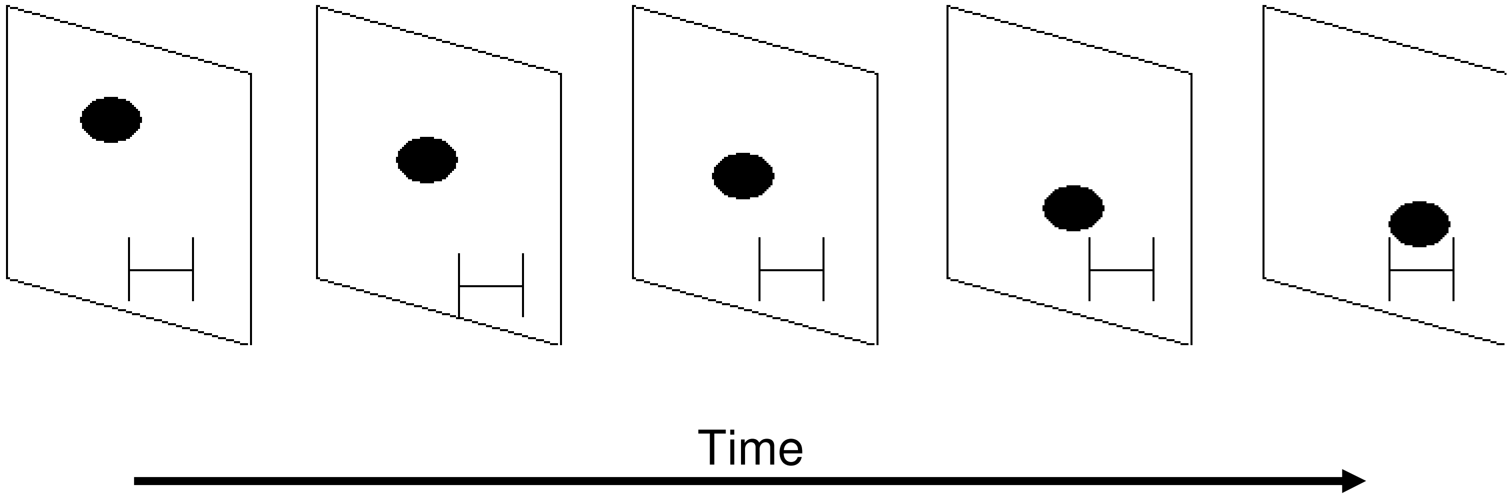
3.2 Inter-frame Encoder



Step 1: Calculate the difference between the current and previous frames;
Step 2: Qantise and encode the difference image.
Step 3: Add the dequantised (residual) image to the previous frame to reconstruct the current frame of image.

3.3 Motion Estimation & Compensation Algorithms

- The football has moved a limited distance but its shape remains almost constant between two adjacent frames.
- To reduce these temporal redundancies, several motion compensation methods can be applied



3.4 Block Based Motion Est.

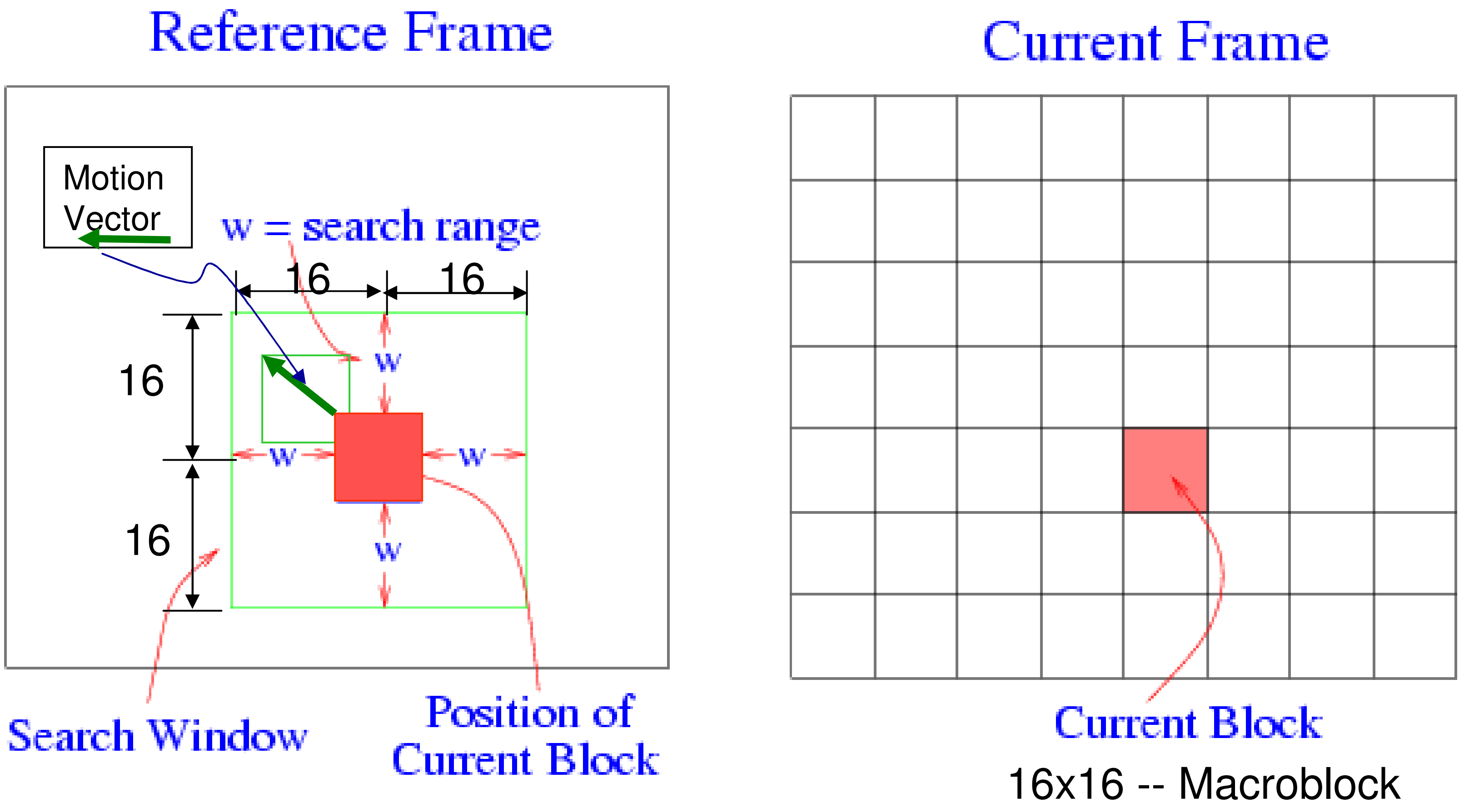
- A common approach in predictive encoder is to attempt to compensate for motion which occurs between frames using block matching
- Using this approach, the reconstruction of the previous frame is searched for the best match to the current frame on a block by block basis. The location of this matching block (called the motion vector) and the prediction error (residual) is then transmitted

3.4 Block Based Motion Est.

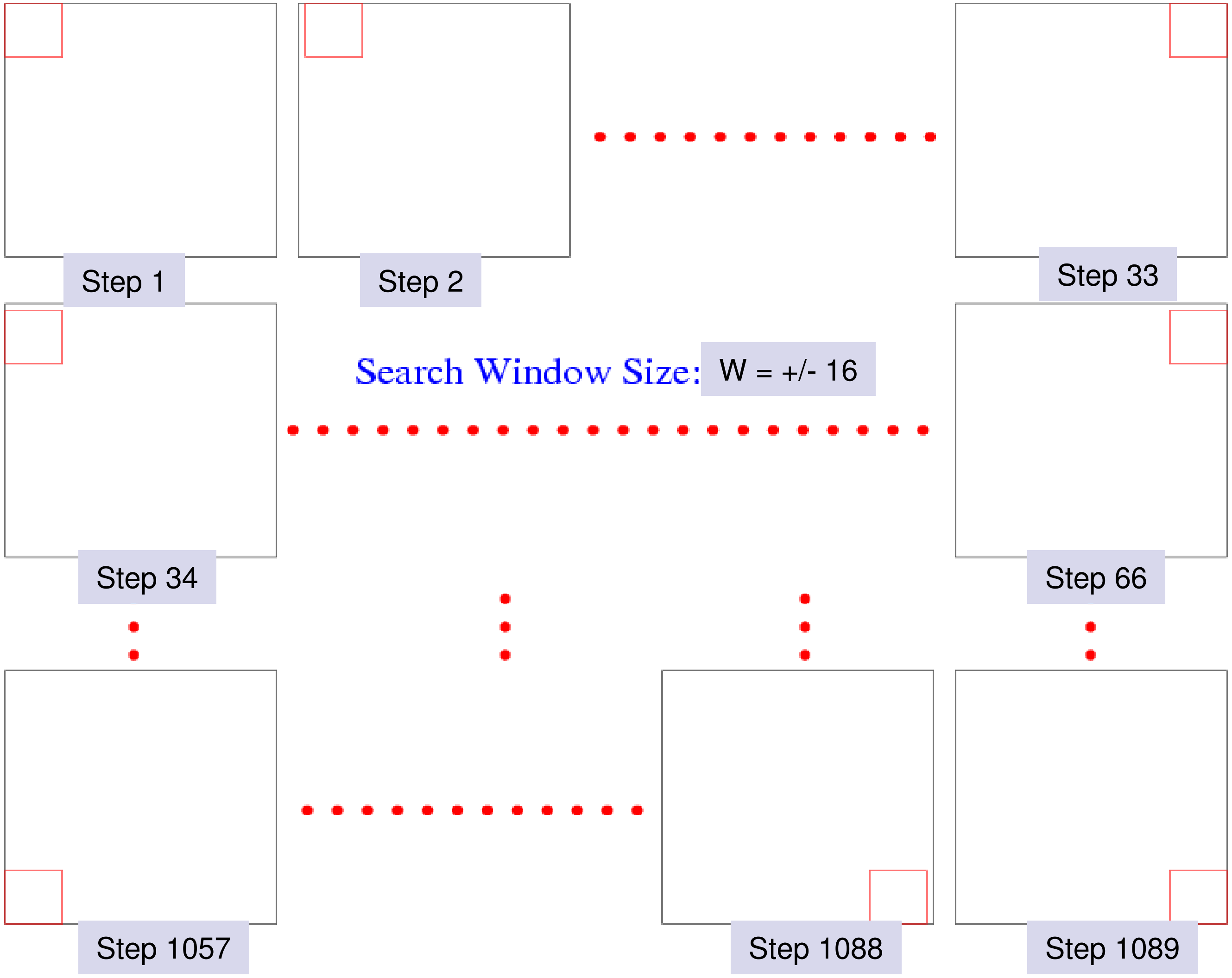
- Each picture is divided into block based subimage for motion estimation rather than using each pixel as a unit.
- To reduce computational and storage requirements, a limited search area is defined to the position around the current block
- The range of possible displacement in the X and Y directions is +/- 16.
  - 5 bits to specify the Hor/Ver displacement
- The total motion vector overhead is 10 bits/pixel (more than the entropy of the orig. image)
- 16x16 Macroblock is defined for ME.

3.4 Block Based Motion Est.

- Block base search



3.5 Full Search Algorithm



3.5 Full Search Algorithm

- For each Macroblock (MB)
  - Total search steps = 33x33 = 1089
  - Total pixel operations = 1089x256=278784
  - For CCIR Rec.601 video (704 pixels x 576 pixels)
  - Pixel operation/frame = 441.6 million/frame
  - Pixel operation/second=11.0 billion/second

    (25 frame/second)
- Conclusion:

  Motion Estimation is Highly Computationally Intensive !!!!!!!

3.6 Motion Compensation

“Previous frame (No. 1)”

3.6 Motion Compensation
“Current original frame (No. 2)”

3.6 Motion Compensation

“Motion compensated prediction (No. 2)”

3.6 Motion Compensation

“Motion compensated prediction error (No. x)”

3.7 Fast Motion Estimation Techniques

- Full search motion estimation is computationally complex.
- Several sub-optimum fast search techniques have been developed.
- Many work on the assumption that block matching will improve monotonically as the search moves closer to the optimum point.
- Since they do not examine all of the candidate blocks, the choice of matching block might not be as good as that chosen by a full search. However, the quality-cost trade-off is usually worthwhile

3.7 Fast Motion Estimation Techniques

Two-dimensional logarithmic search (TDL)

- This and the following techniques are quadrant monotonic searches.
- Quadrant monotonic assumes that the value of the distortion function increases as the distance from the point of minimum distortion increases.
- Special case of the principle of locality: not only locality to optimal block, but also distance from optimal block.
- TDL is a multi-stage search which successively reduces the search area during each stage until the search area is trivially small.
- Search positions: $2 + 7 \log_2 w$. Example: for w=16, 30 positions are searched, compared with 1089 for the full search.
- The processing of a search step relies on the previous search steps; therefore, the search steps cannot be preformed in parallel.

3.7 Fast Motion Estimation Techniques

Two-dimensional logarithmic search (TDL)

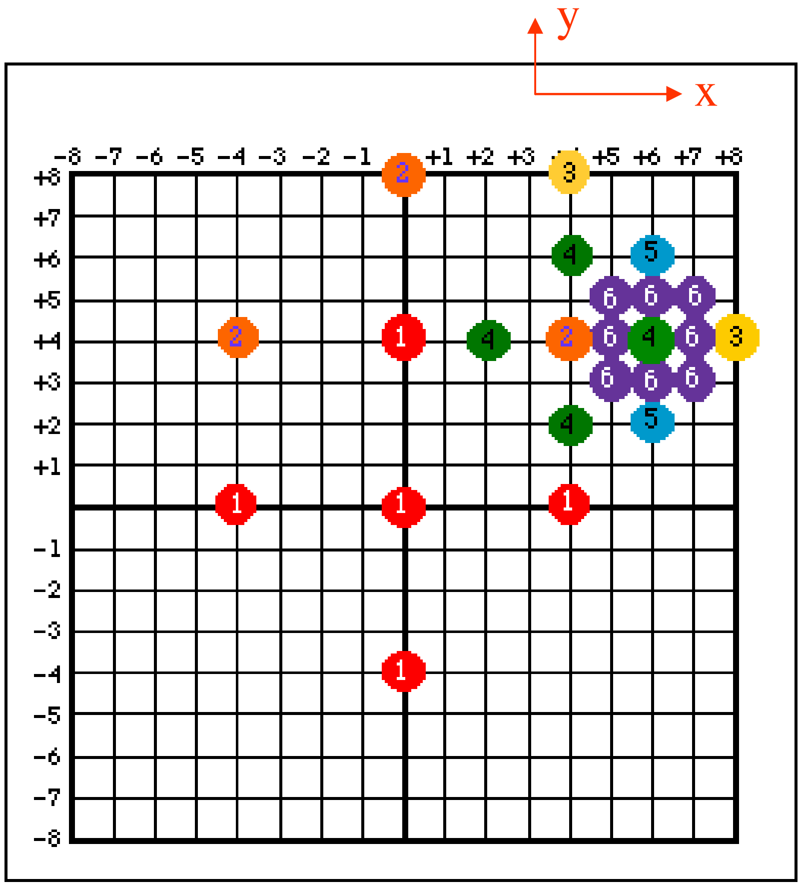
Step 1: The block at the centre of the search area and four blocks at distance ***s*** from the centre on the X and Y axes are searched for a best match.

Step 2: If the position of best match is the centre, halve the step size (***s***/2). Else, if the best match is in one of the four outer positions, then it becomes the new centre point ([cx,cy]) for the next stage.

Step 3: If the step size ***s*** is 1, then all nine blocks around the centre are examined, and the best match chosen for the target block. Otherwise, blocks at positions ([cx,cy], [cx+s,cy], [cx-s,cy], [cx,cy+s], and [cx,cy-s]) are searched, and the algorithm goes to stage 2.

Note: The points [0,+4], [+4,+4], [+6,+4] are the minima at each stage, and finally [+7,+4] is chosen as the matching block.

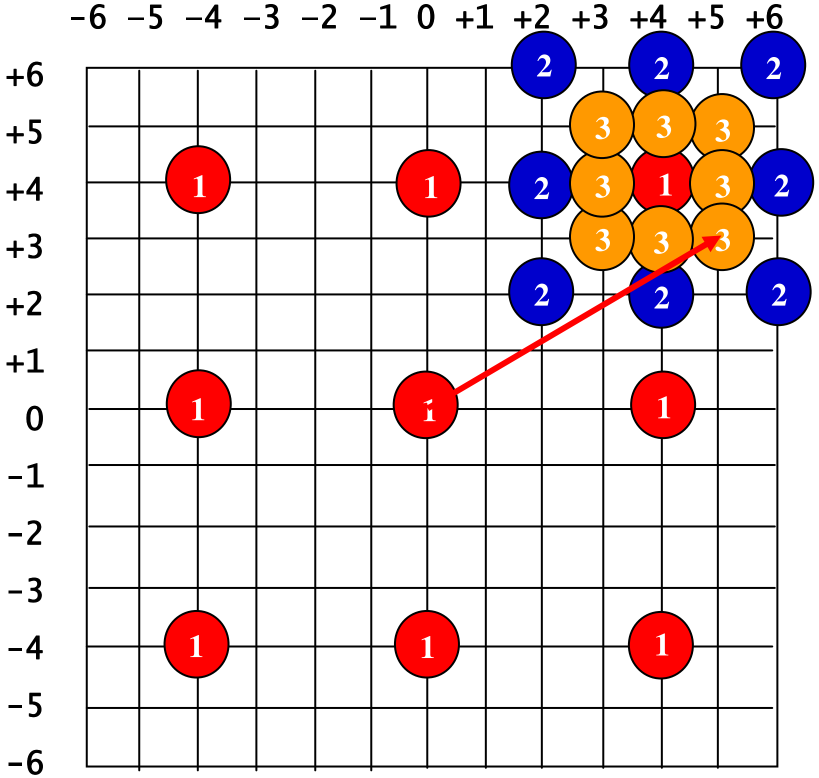
For step size ***s*** and search window size **w**, the step update algorithm is given by



$$s_0 = 2^{\lfloor \log_2 w \rfloor - 1}, \quad s_n = s_{n-1}/2$$

3.7 Fast Motion Estimation Techniques "Three Step Search (TSS)"

- Very similar to TDL search, and developed around the same time.
- The three step search tests eight points around the centre instead of four, with the position of minimum distortion becoming the new centre.
- After each stage the step size is reduced.
- For step size ***s*** and search window size **w**, the step update algorithm is given by



$$s_0 = 2^{\lceil \log_2 w \rceil - 1}, \quad s_n = 2^{\lceil \log_2 s_{n-1} \rceil - 1}.$$

3.7 Fast Motion Estimation Techniques

Ghanbari: M. Ghanbari, “The cross-search algorithm for motion estimation”, IEEE Trans. Comm., vol.38, pp.950-953, July 1990.
PHS: A. Puri, H.-M. Hang and D.L. Schilling, “An efficient block-matching algorithm for motion-compensated coding”, in Proc. IEEE ICASSP, pp.25.4.1-25.4.4, 1987.
SR: R.Srinivasan and K.R. Rao, “Predictive coding based on the efficient motion estimation”, IEEE Trans. Comm., vol.COM-33, pp.888-896, Aug. 1985.
KR: S Kappagantula and K.R. Rao, “Motion compensated interframe image prediction”, IEEE Trans. Comm., vol.COM-33, pp.1011-1015, Sep. 1985

Algorithm	Maximum number of search points	w		
		7	15	31
Full search	$(2w+1)^2$	225	961	3969
Ghanbari	$2+7\log_2 w$	22	30	37
PHS	$1+4\log_2 w$	13	17	21
SR	$3+2w$	17	33	65
KR	$1+6\log_2 w$	18	25	31

3.7 Fast Motion Estimation Techniques

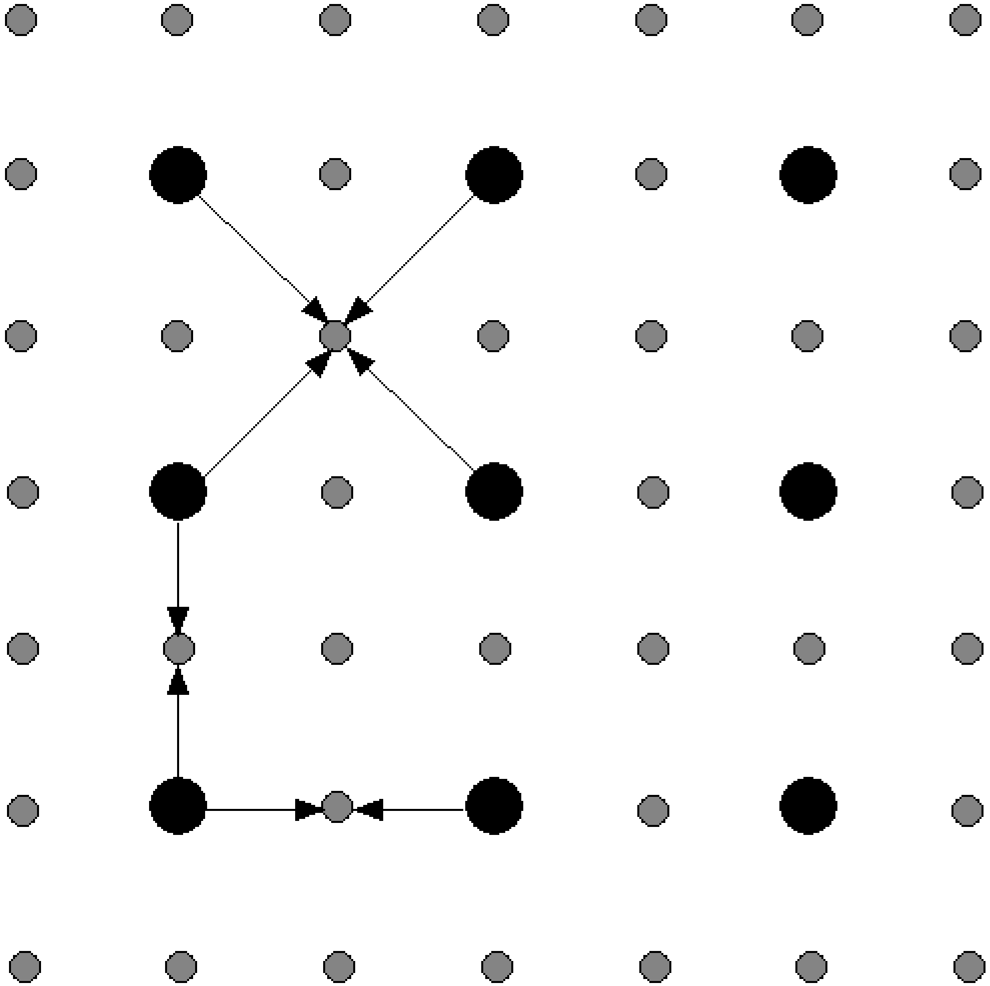
- Issues with fast motion estimation algorithms
  - Full search motion estimation is computationally complex.
  - Fast search algorithms only work based on the assumption that block matching will improve monotonically as the search moves closer to the optimum point, which may not be the case in general.
  - All block matching based motion estimation algorithms assume pixel motion within the block is uniform, which will not work if the block involves object deformation or varying motion speeds.
  - Overlapped block matching algorithm and hexagonal matching algorithms have been introduced to address these two issues.

3.8 Motion Estimation to Sub-pixel Accuracy

- After a single pixel accuracy of motion estimation, it can be extended to sub-pixel accuracy by performing bi-linear interpolation between pixels in the search area.

- This can lead to a worthwhile improvement in the performance of motion compensated prediction

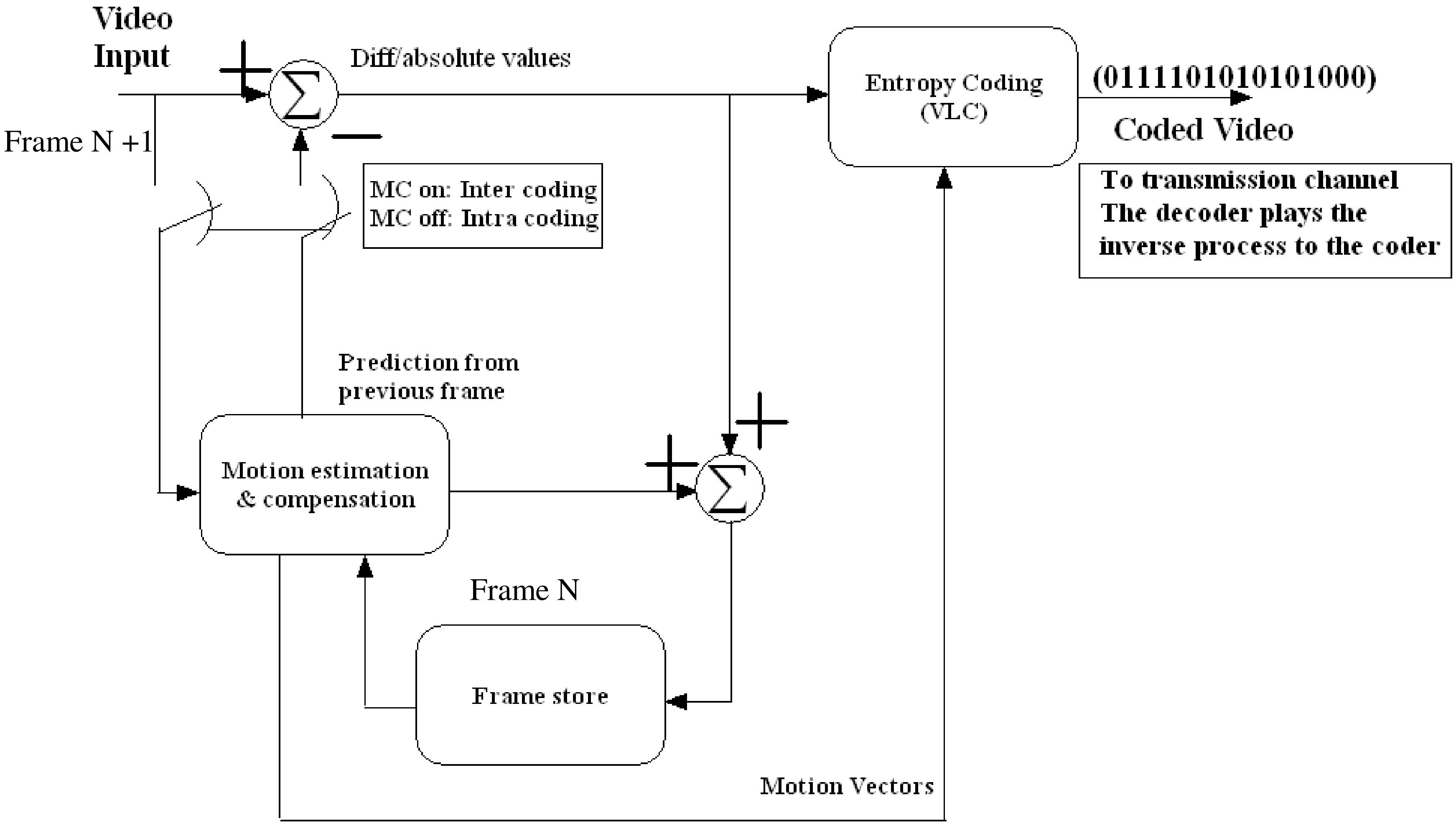
3.8 Motion Estimation to Sub-pixel Accuracy



3.9 Introduction to Video Coders –
Motion Compensated Coder

- Motion Compensated Coder is a lossless coder
- Central to the operation of the coder is the frame store
- It contains one or more previously transmitted frames.
- A Marcoblock (MB) can be transmitted directly to the decoder is called “intra” mode transmission
- A different block between the current block and a corresponding block in a transmitted frame in the frame store is called “inter” mode transmission
- This is a hybrid MC/DPCM

3.9 Introduction to Video Coders –
Motion Compensated Coder



To transmission channel
The decoder plays the inverse process to the coder

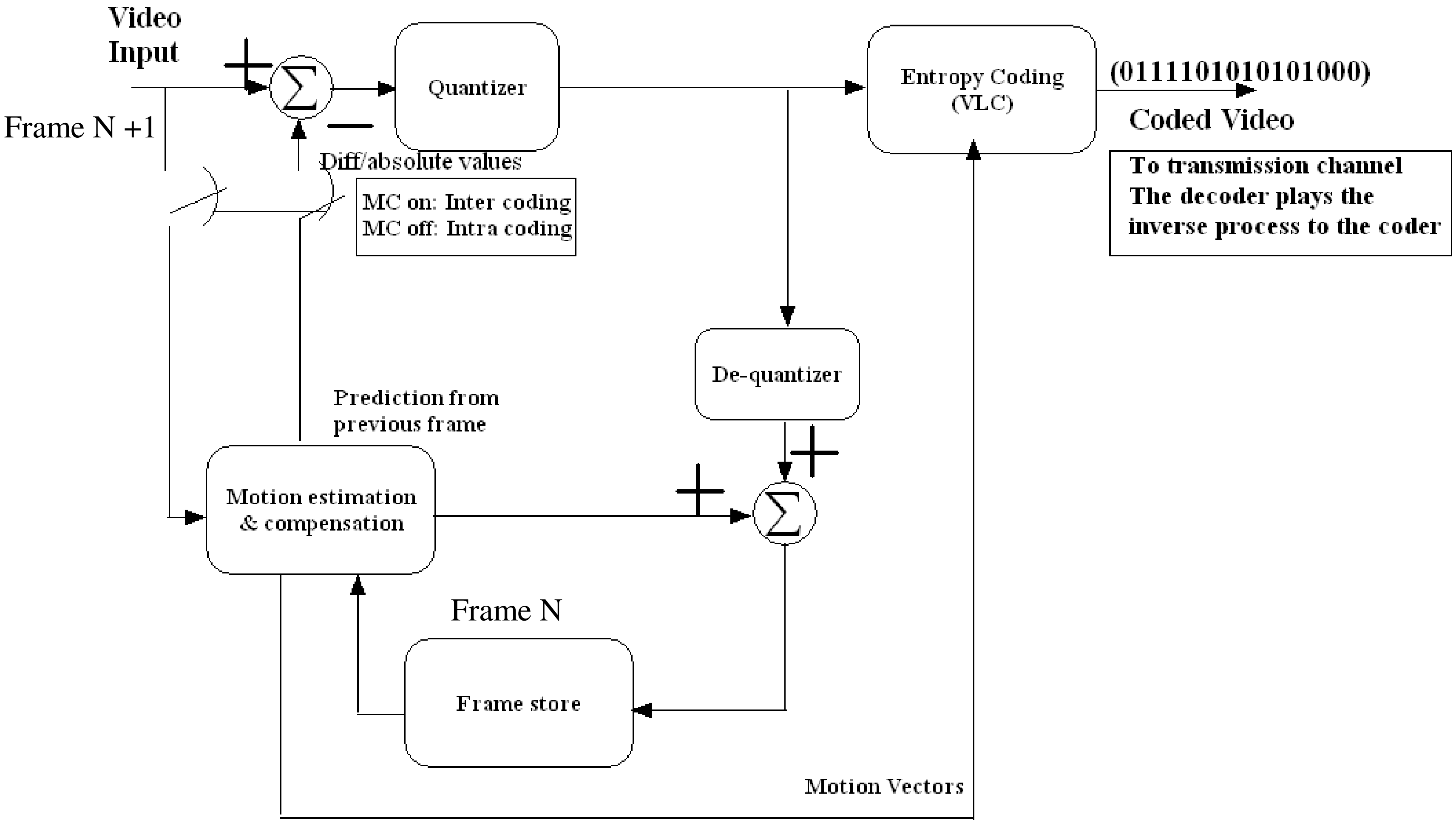
3.9 Introduction to Video Coders –

Motion Compensated Coder with Quantization

- This is a lossy coder since a quantization scheme is introduced
- The key part for the coder is the quantizer to reduce the residue values after the motion compensation for inter-mode.
- This is a hybrid MC/DPCM with quantization coder

3.9 Introduction to Video Coders –

Motion Compensated Coder with Quantization

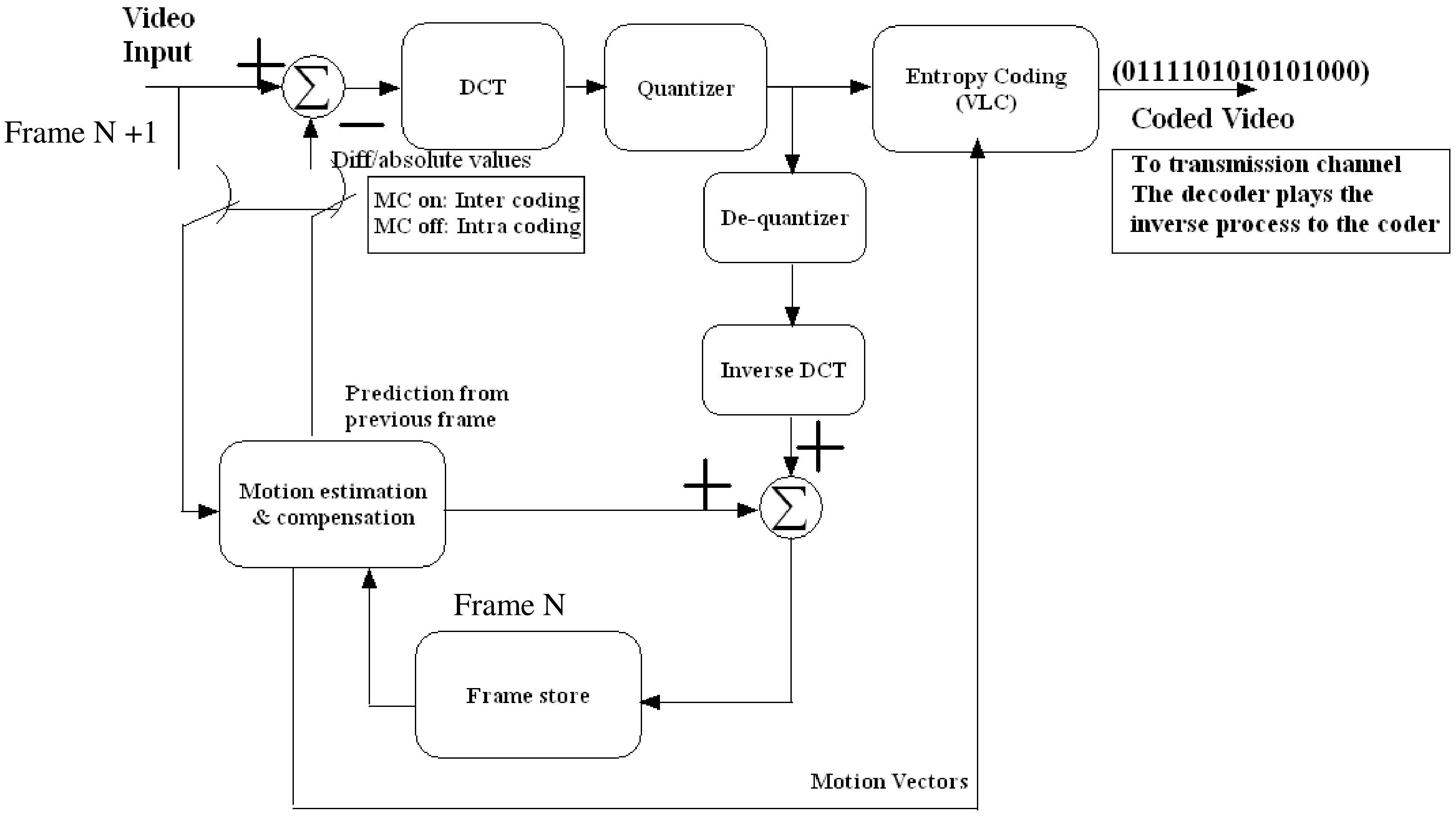


3.9 Introduction to Video Coders –

Motion Compensated DCT/Quantization Coder

- The pixel data for a motion compensated MB is DCT transformed before quantization.
- The 2-D DCT coef. of each 8x8 block is then re-arranged by zig-zag into a 1-D arrangement
- A quantization scheme is applied to differently quantize each DCT coeff. according to its frequency spectrum
- This will achieve a better compression result based on perceptual weight to the HVS
  - The HVS is less sensitive to high spatial frequencies than lower frequencies
  - The low DCT coeff. the fine quantization
- This is a hybrid MC/DPCM/DCTcoder

4.1 Introduction to Video Coders –
Motion Compensated DCT/Quantization Coder



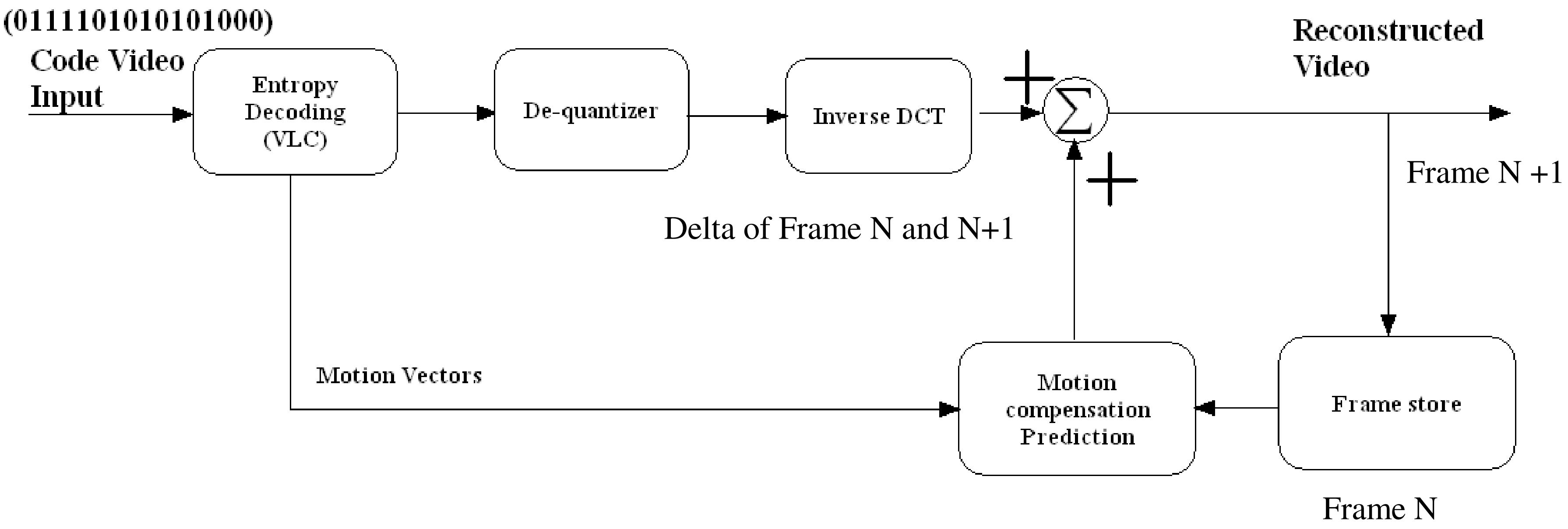
MC on: Inter coding
MC off: Intra coding

To transmission channel
The decoder plays the
inverse process to the coder

3.9 Introduction to Video Coders –

Motion Compensated DCT/Quantization Decoder

- The procedure followed by the decoder does the inverse of the process performed at encoder



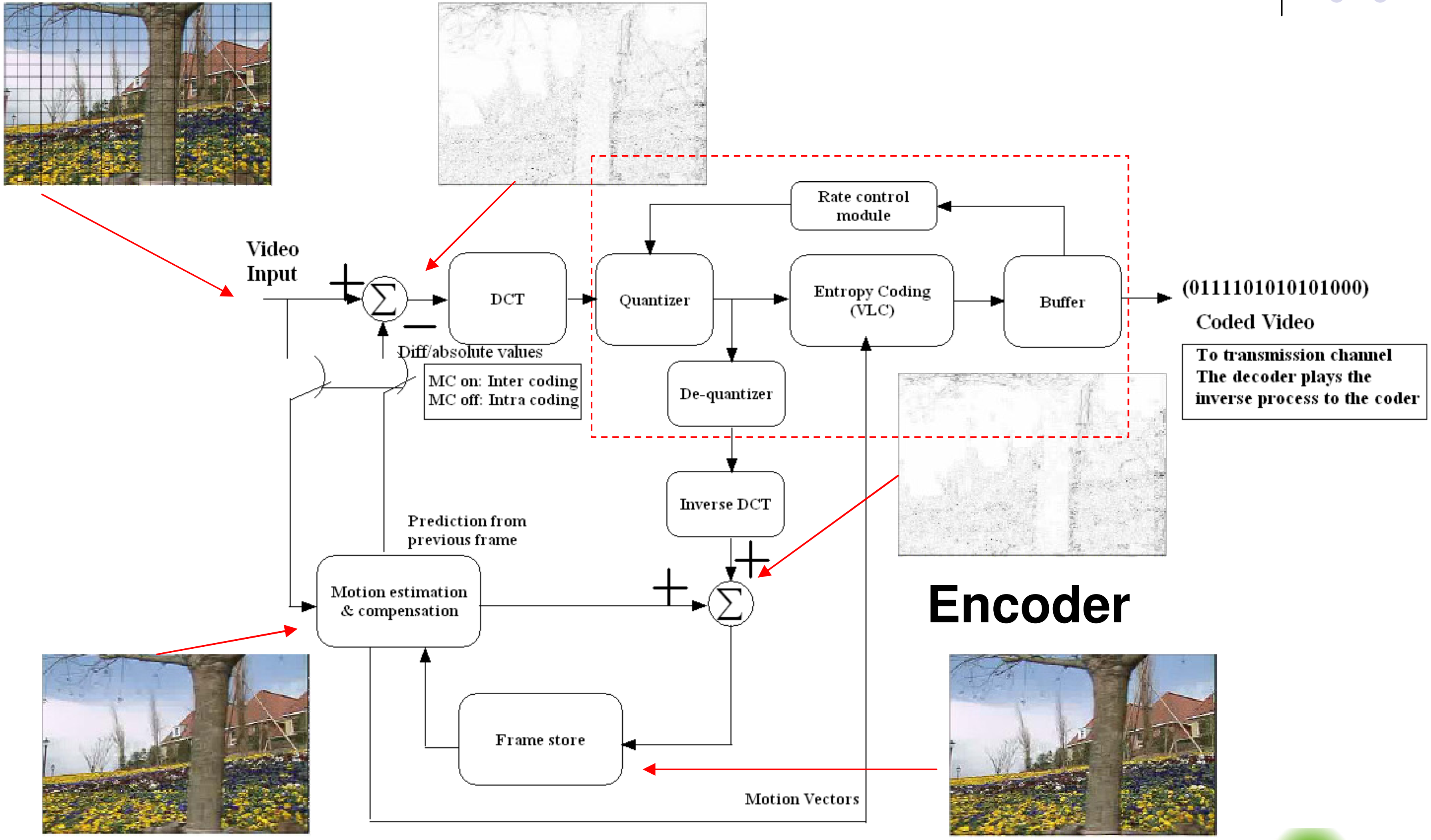
3.10 Digital Video Coding (DVC) Structure

- All the DVC standards are based on the Hybrid MC/DPCM/DCT video coding structure

- Since the constant rate constrain for most of current video codec applications, the quantization scheme should be considered to achieve the maximum of rate/distortion (R-D) ratio

- Standards defined the decoder process while provide verification model for industry to develop their encoder. Therefore, there are many challenges about how to develop advanced algorithms to realize encoder at low complexity, low power and high performance (e.g: R-D ratio)

3.10 Digital Video Coding (DVC) Structure – Hybrid MC/DPCM/DCT



Video Input

DCT

Rate control module

Quantizer

Entropy Coding (VLC)

Buffer

(0111101010101000)

Coded Video

To transmission channel
The decoder plays the inverse process to the coder

Diff/absolute values

MC on: Inter coding
MC off: Intra coding

De-quantizer

Inverse DCT

Prediction from previous frame

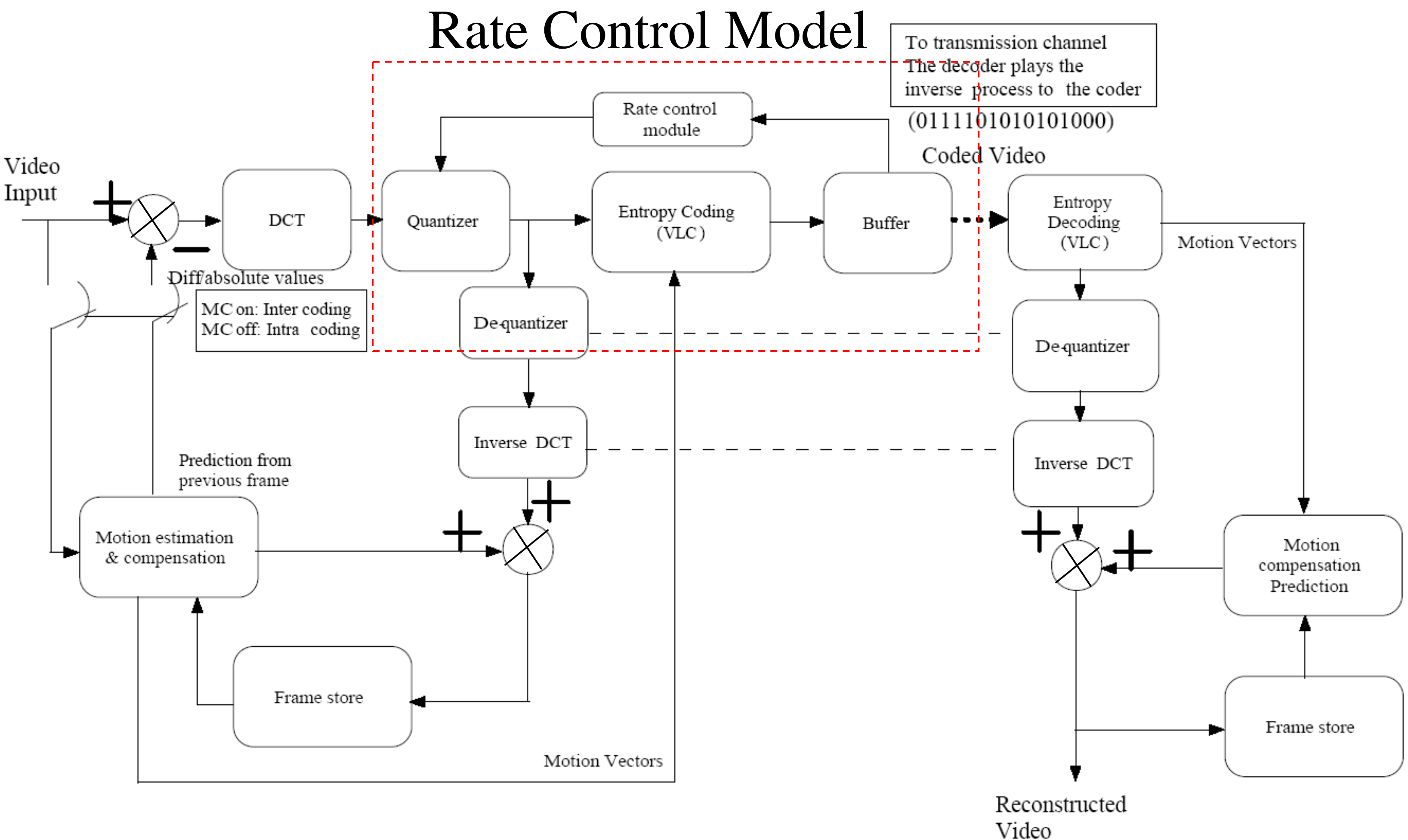
Motion estimation & compensation

Frame store

Motion Vectors

Encoder

3.10 Digital Video Coding (DVC) Structure – Hybrid MC/DPCM/DCT

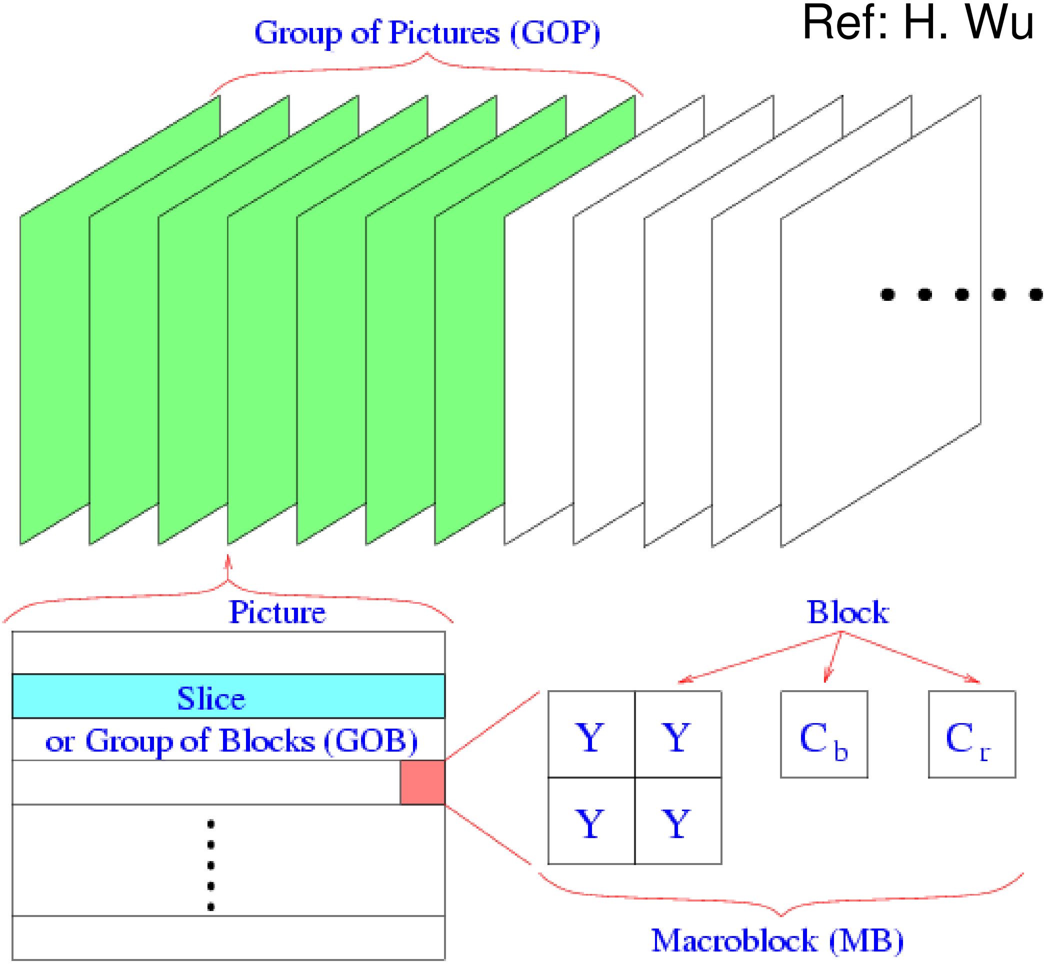


Codec = encoder/decoder

3.10 Digital Video Coding (DVC) Structure – Video Block Data Structure

- Typical MC/DPCM/DCT video coding architectures use the 4:2:0 (YCbCr) format as block data structure
- The architecture consists of the following layered data structure for each picture:
  - Picture layer
  - Slice or group of blocks (GOB) layer
  - Macroblock (MB) layer: 16x61 pixels 4Y,Cb & Cr
  - Block size layer: 8x8 pixels.
- . MC is applied to the MBs.

Ref: H. Wu



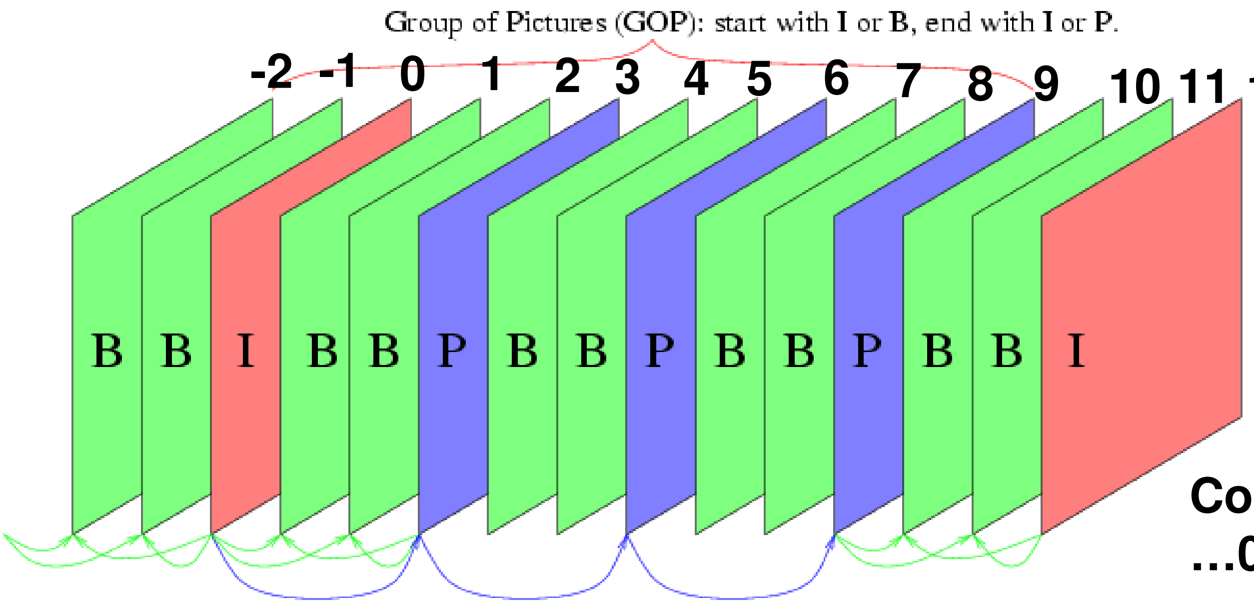
3.11 Digital Video Coding (DVC) Standards -- Overview

- MPEG: Motion Pictures Experts Group
- ISO/IEC JTC1/SC 29/WG 11: (JPEG is WG 1). Formed in Jan. 1988.
- MPEG-1 (ISO/IEC 11172, Nov 92).
  - Audio and video storage media such as CD-ROM 1x CD-ROM: 150 KB/s = 1.2 Mbps).
  - Targeted at 1 to 1.5 Mbps (~1.2 Mbps for video, and ~250 kbps for audio).
- MPEG-2 (ISO/IEC 13818, Nov 94).
  - Digital TV: SDTV, HDTV, DVD, etc.
  - Wider range of bitrates: 4 to 80 Mbps (optimised for 4 Mbps).
  - Supports interlaced video and scalable coding.
- MPEG-4 (ISO/IEC 14496, Oct 98).

3.11 Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)

- Moving Picture Experts Group - ISO/IEC JTC1/SC29/WG11
- Coded representation of moving pictures and associated audio stored on digital storage media
- Basic Requirements:
  - Generic video coding at 1 to 1.5 Mbps (~VHS and 1.2 Mbps for video, and ~250 kbps for audio)
  - Fast forward/reverse: seek and play in FF/FR using access points.
  - Random access to a frame in limited time: frequent access points
  - System supporting audio-visual synchronized play and access
- Typical features and parameters:
  - Bi-directional in temporal processing (I,P,B frame)
  - Larger motion compensation range with half pxiel MC (no loop filters)
  - Quantization table
  - 4:2:0 format and SIF (~CIF) resolution 352x240@30 or 352x288@25

3.11 Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)



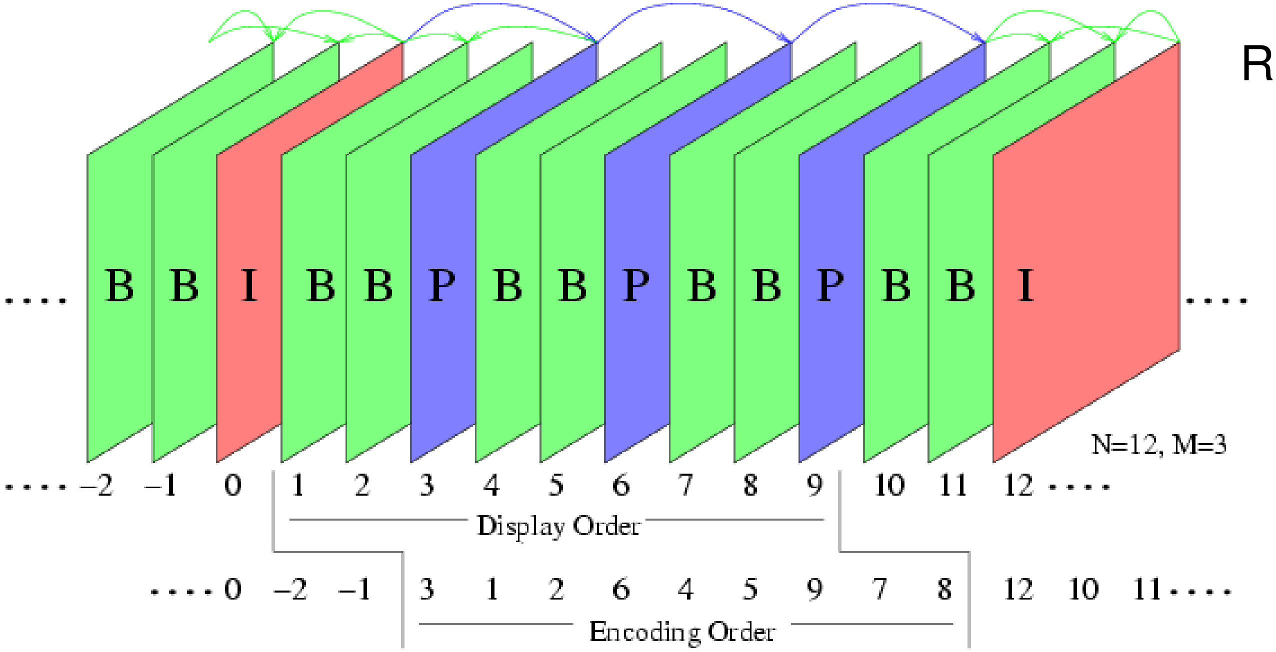
N = number of pictures in GOP
M = prediction distance
(M–1 in–between B–pics)

Example: N=12, M=3

Coding order:
…0,3,1,2,6,4,5,9,7,8…

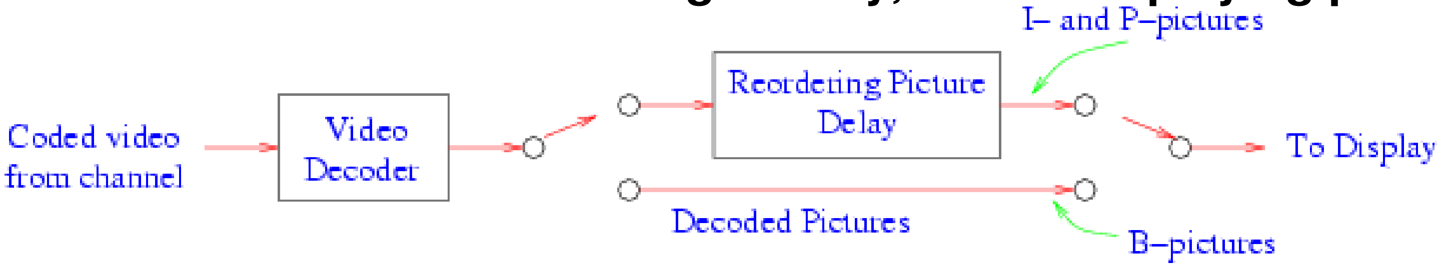
- **Intra coded picture (I-Picture):**
  - Coded on their own (all MBs are intra) and server as random access.
- **Predicted picture (P-Picture):**
  - Coded with reference (MC predictions) to the previous anchor I or P picture.
- **Bi-directionally predicted picture (B-Picture):**
  - Coded with reference to the previous and/or future anchor I or P pictures (forward or backward MC prediction and/or linear interpolation).

3.11 Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)

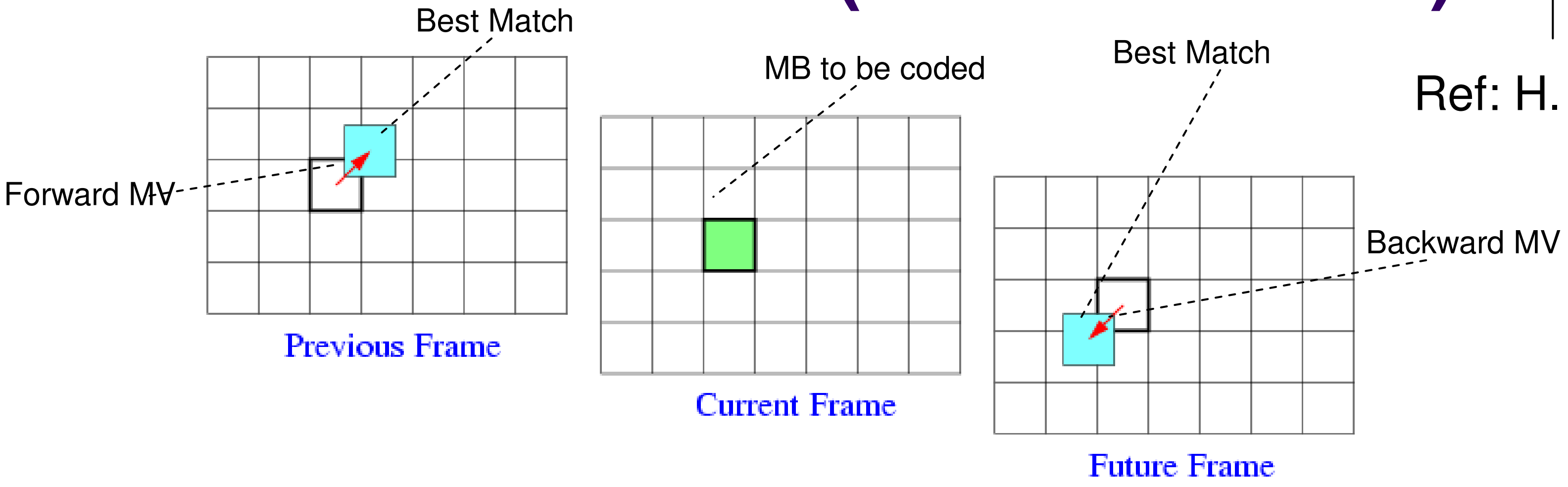


Ref: H. Wu

The decoder needs to reorder using a delay, before displaying pictures.



3.11 Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)



Ref: H. Wu

Forward prediction: Predict where the pixels in a current frame were in a past frame.

Backward prediction: Predict where the pixels in a current frame will go in a future frame.

Prediction for a macroblock may be backward, forward, or an average of both.

Advantages

Main Advantage:

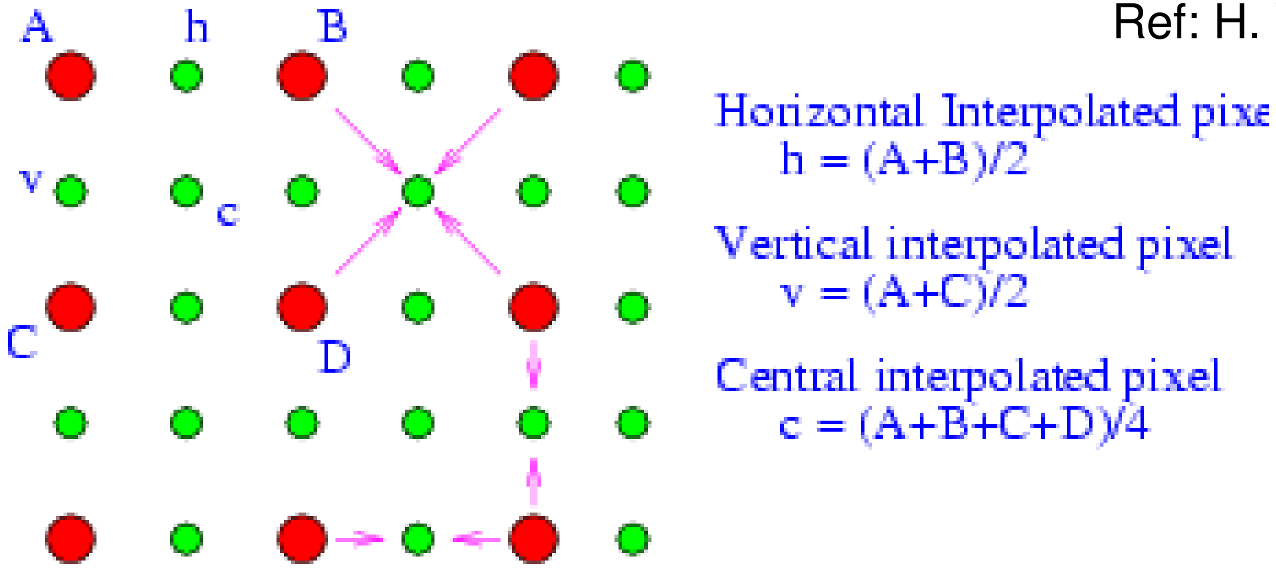
- High coding efficiency (gain/cost is significant)
- No uncovered background problem

Main disadvantage: long delay and more memory to store two anchor frames

3.11 Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)

- Half-pel refinement of motion vectors using simple linear interpolation.
- Half-pel causes filtering of prediction image: loop-filter not required.

Ref: H. Wu



Horizontal Interpolated pixe
$h = (A+B)/2$

Vertical interpolated pixel
$v = (A+C)/2$

Central interpolated pixel
$c = (A+B+C+D)/4$

- The use of bi-directional prediction in MPEG-1 can also lead to sub-pixel accuracy motion compensation and top of half pixel MC to give better-than-half-pixel accuracy

3.11 Digital Video Coding (DVC) Standards– MPEG-1 (ISO/IEC 11172)

- Quantization weighting matrix (QWM)
  - Diff quantizers for diff. DCT Coef.
  - Default QWM

Quantized Setpsize

$$\hat{C}_{ij} = 16 * C_{ij} // (Q * W_{ij})$$

Coef.ij

weighting

Recommended MPEG quantization matrix

$$\begin{bmatrix} 8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\ 16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\ 19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\ 22 & 22 & 26 & 27 & 29 & 34 & 37 & 40 \\ 22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\ 26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\ 26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\ 27 & 29 & 35 & 38 & 46 & 56 & 69 & 83 \end{bmatrix} \qquad \begin{bmatrix} 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \\ 16 & 16 & 16 & 16 & 16 & 16 & 16 & 16 \end{bmatrix}$$

3.12 Digital Video Coding (DVC) Standards– MPEG-2 (ISO/IEC 13818)

- ISO/IEC 13818-2 (or ITU-T H.262) -- An extension of MPEG-1 standard
- Broadcast TV, cable/satellite TV, HDTV, DVD, etc
  - To achieve PAL quality at rates between 4 and 9 Mbits/s -- SDTV
  - To cover HDTV around 20 Mbits/s
- To cope with fully interlaced content such as ITU-R 601 resolution (i.e. 720x576@50 Hz interlaced field rate).
- Extremely flexible to support adaptive filed/frame processing (ME and DCT) and higher chroma resolution (4:2:2 and 4:4:4)
- Downloadable quantization matrix
- Scalable video for multi-quality video applications

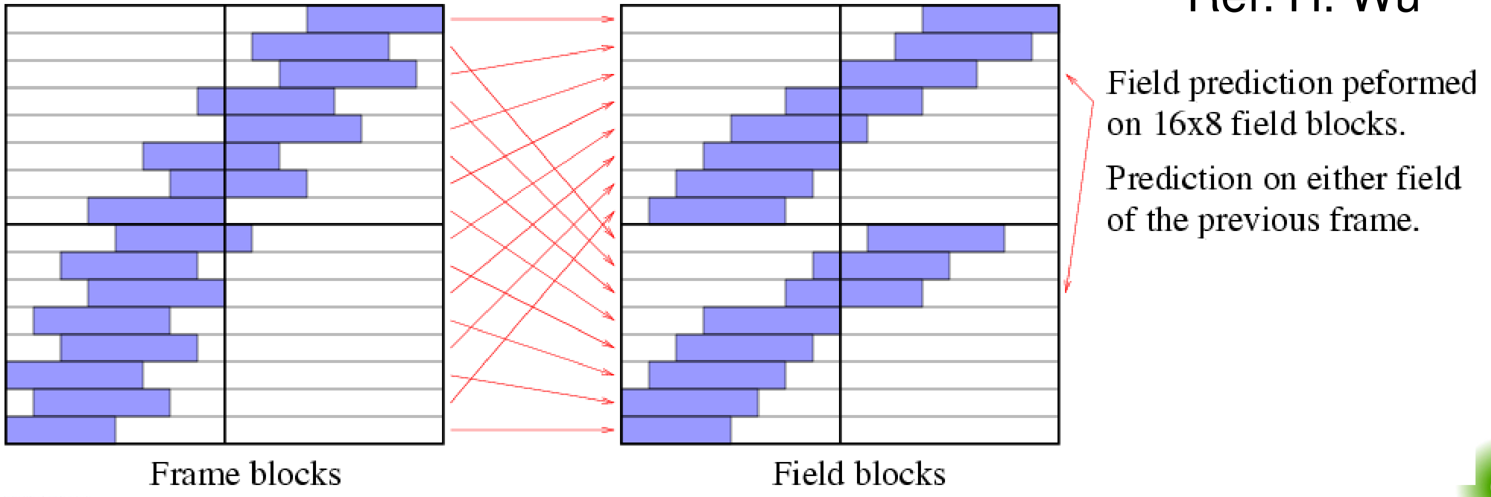
3.12 Digital Video Coding (DVC) Standards– MPEG-2 (ISO/IEC 13818)

- Profile and level
  - DVC standard subset to satisfy the functional requirements of different applications
  - Each profile "levels" are defined to restrict the values of various parameters to build flexible encoder/decoder
- Support for interlaced content (selected on frame-by-frame basis):
  - Frame-pictures: frame encoded as a single picture (as in MPEG1).
  - Field-pictures: two fields encoded as two picutres (can use prediction).
- Motion estimation from either alternate field, or previous frame
- For Intra_coding
  - Alternative Intra VLC table, Zig-zag scan order ….

3.12 Digital Video Coding (DVC) Standards– MPEG-2 (ISO/IEC 13818)

- MPEG-2 field and frame pictures
  - Two interlaced fields make up one frame
  - If first field is P/B, then second field will also be P/B
  - If first field is I, then second field can be I or P
  - Independent predictions for each field from one or more previous fields
  - The two files of the frame are interleaved
  - Each macroblock may be adaptively frame or field encoded and predicted to achieve high coding efficiency!

Ref: H. Wu



3.12 Digital Video Coding (DVC) Standards– MPEG-2 Profile/Levels

- Profiles and levels provide a means of defining subsets of the standard syntax and thereby the decoder capabilities
- A profile defines a subset of constraints upon the allowed values of parameters within the full syntax
- Conformance tests will be carried out against defined profiles at defined levels
- Three profiles have been defined at this stage
  - Main Profile
  - Next Profile
  - Simple Profile

3.12 Digital Video Coding (DVC) Standards– MPEG-2 Profile/Levels

- Level\Profile parameters

Level\Profile		Simple 4:2:0 Single Layer	Main 4:2:0 Single Layer	Nextg 4:2:2 Scalable
High	Pixels/line	1920	1920	1920
	Lines/frame	1152	1152	1152
	Frame/s	60	60	60
	Pixels/s	62.7 million	62.7 million	62.7 million
High-1440	Pixels/line	1440	1440	1440
	Lines/frame	1152	1152	1152
	Frame/s	60	60	60
	Pixels/s	44.9 million	44.9 million	44.9 million
Main	Pixels/line	720	720	720
	Lines/frame	576	576	576
	Frame/s	30	30	30
	Pixels/s	10.4 million	10.4 million	10.4 million
Low	Pixels/line	352	352	Not defined
	Lines/frame	288	288	
	Frame/s	30	30	
	Pixels/s	2.53 million	2.53 million	

3.12 Digital Video Coding (DVC) Standards– MPEG-2 Profile/Levels

- MP@ML
  - Chroma format – 4:2:0
  - Bit rate flexibility: Yes, CBR and VBR operation
  - Random access: Yes, access point at I frames
  - Editability: Yes, but not necessarily at every frame
  - Error resilience: Yes, details in late slides
  - Video windowing: Yes, for display of 16:9 service on a 4:3 receiver need to signal part to be displayed
  - Low Delay: Yes
  - Trick Modes: Yes, basic fast forward/fast reverse supported in main syntax
  - Scalability: No
  - Compatibility: Full compatibility with MPEG-1
  - Quality: Able to trade picture quality against bit rate
  - Flexibility in implementation: Yes, a high degree of encoder flexibility provided