Wireless Communication Systems
@CS.NCTU

Lecture 8: Video – H261

Instructor: Kate Ching-Ju Lin (林靖茹)

Chap. 10.4 of "Fundamentals of Multimedia"

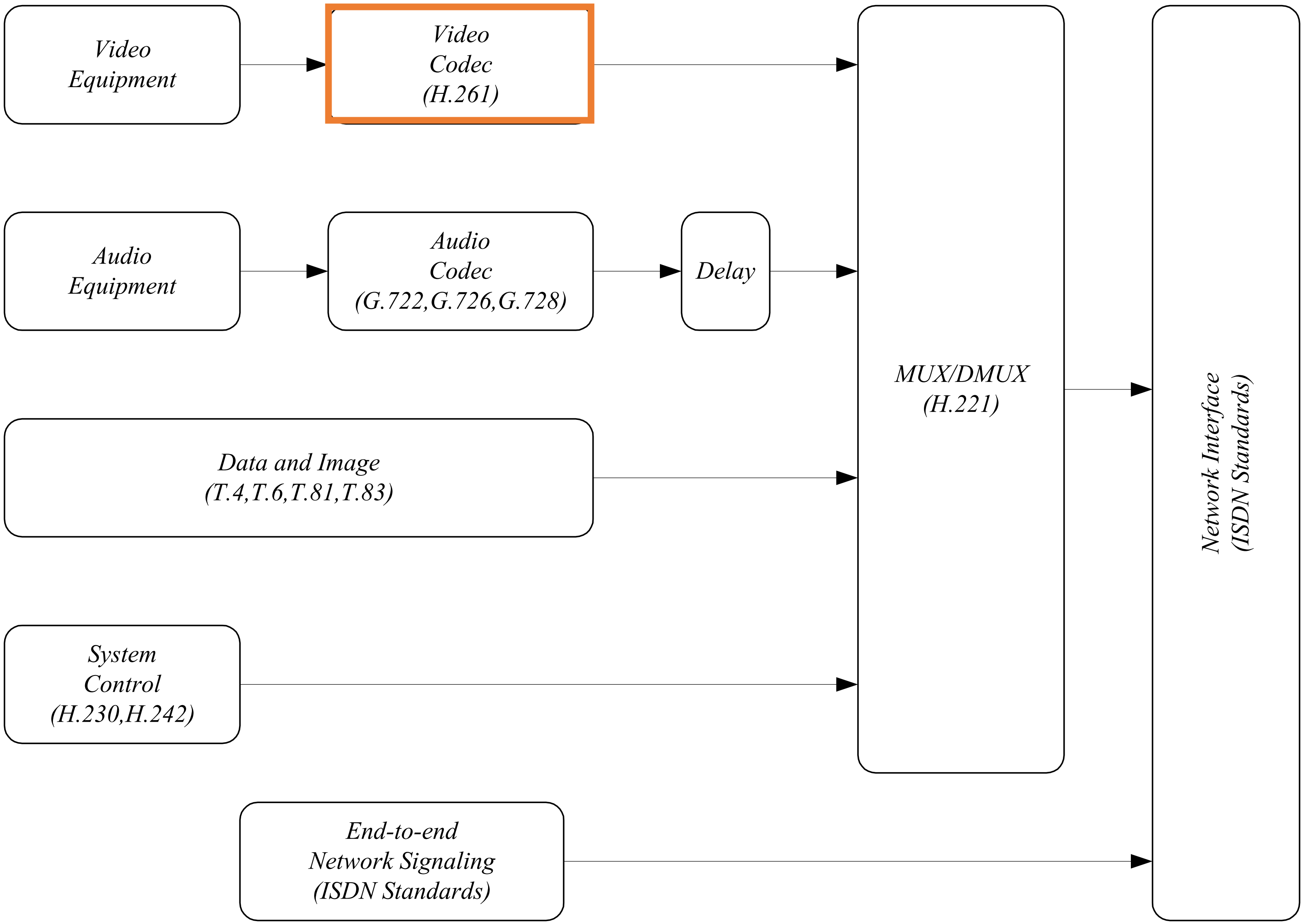
Outline

- **Introduction**
- Frame sequence
- Frame coding
- Quantization
- Encoder and decoder
- H.261 syntax

Introduction to H.261

- Video coding and decoding for the moving picture component of audiovisual services at the rates of p × 64 (kbps), where p ranges form 1 to 30
- Improve storage and transmission efficiency in ISDN (Integrated Service Digital Network)
  - considering a relatively low bitrate
- Belong to the following set of ITU recommendations
  - H.221, H230, H.242, H.261, H.320

Example: Recommendation H.320



Video Formats in H.261

Video format	Luminance image resolution	Chroma image resolution	Bitrate (Mbps) for 30fps	H.261 support
QCIF	176 x 144	88 x 72	9,1	required
CIF	352 x 288	176 x 144	36,5	optional

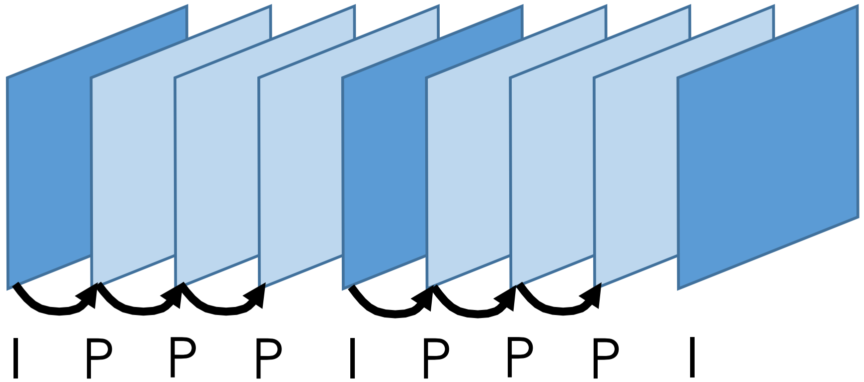
- Chroma subsampling: 4:2:0
- Compression ratio: about 100 to 300
- Designed for lowrate
  → QCIF is specified as required, while CIF is optional

Outline

- Introduction
- **Frame sequence**
- Frame coding
- Quantization
- Encoder and decoder
- H.261 syntax

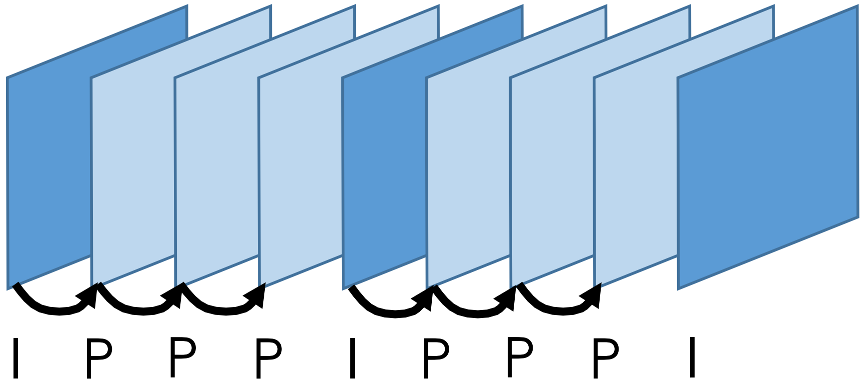
H.261 Frame Sequence

- **I-frame**
  - Independently coded
  - Coded using a transform coding method, e.g., DCT (similar to JPEG), hence called intra-frame
  - Spatial redundancy removal only
- **P-frame**
  - Not independent, coded using a forward predictive coding method
  - Difference between frames are coded
  - Both spatial and temporal redundancy removal



H.261 Frame Sequence (Cont.)

- Interval between pairs of I-frames is a variable
- Motion vectors are measured within a range of ±15 pixels, i.e., p=15
- A P-frame can be predicted by the preceding I- or P-frame



Outline

- Introduction
- Frame sequence
- **Frame coding**
- Quantization
- Encoder and decoder
- H.261 syntax

Stage 1 - Reducing Temporal Redundancy

- Segment a frame into macroblocks
- Compensate motion and remove temporal redundancy
- Output energy is related to the degree of temporal redundancy
- This stage is inter-frame coding

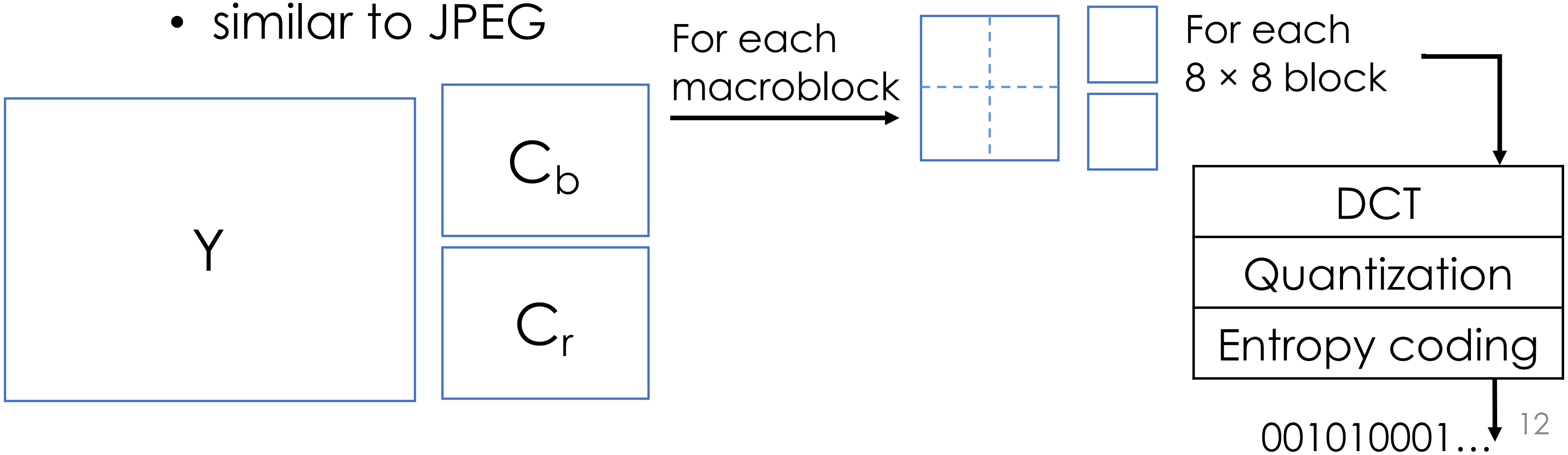
Stage 2 - Reducing Spatial Redundancy

- Processing the difference frame (spatially correlated) from stage 1
- Usually using DCT coding
- This stage is **intra-frame** coder

These two stages together are **hybrid coding**

I-frame Coding

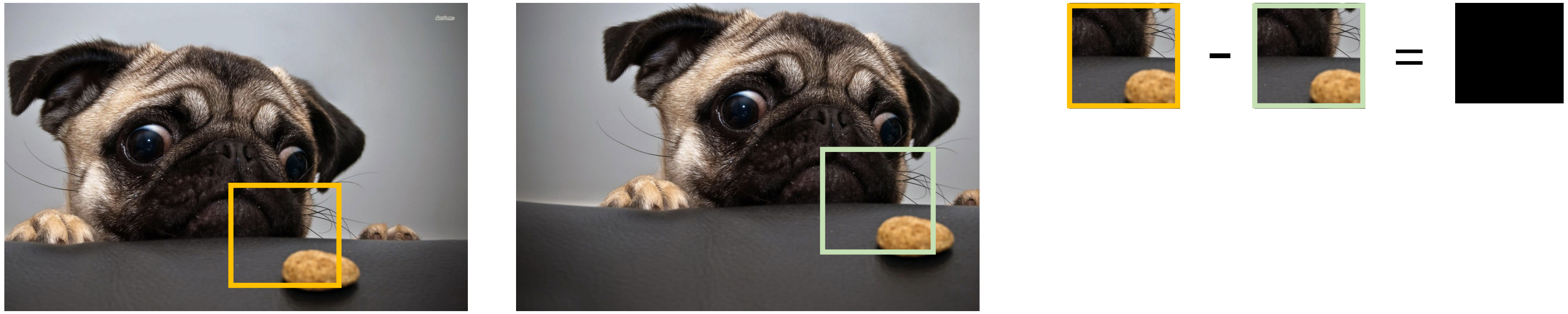
- Partition a frame into macroblocks
- Macroblocks is further partitioned into 8 × 8 blocks
- Adopt 4:2:0 chroma sampling
  - each macroblock is of 16 ×16 pixels for Y frames
  - each macroblock is of 8 × 8 pixels for Cb and Cr frames
- Apply DCT to code each 8 × 8 block
  - then go through quantization, zigzag scanning and entropy coding
  - similar to JPEG



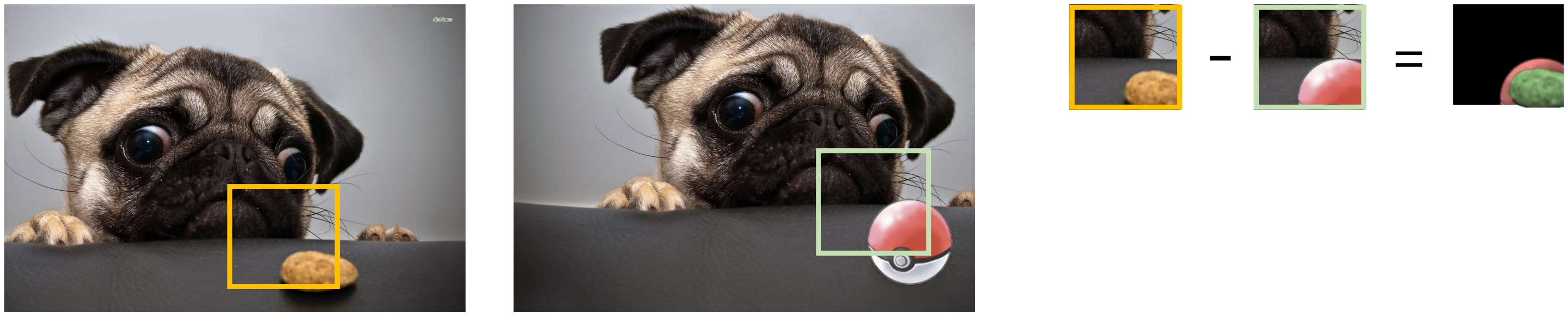
P-frame Coding

- Coded based on motion compensation
  - For each macroblock, find the motion vector
  - Find the difference between the target MB and reference MB
    → prediction error = difference macroblock

Example 1

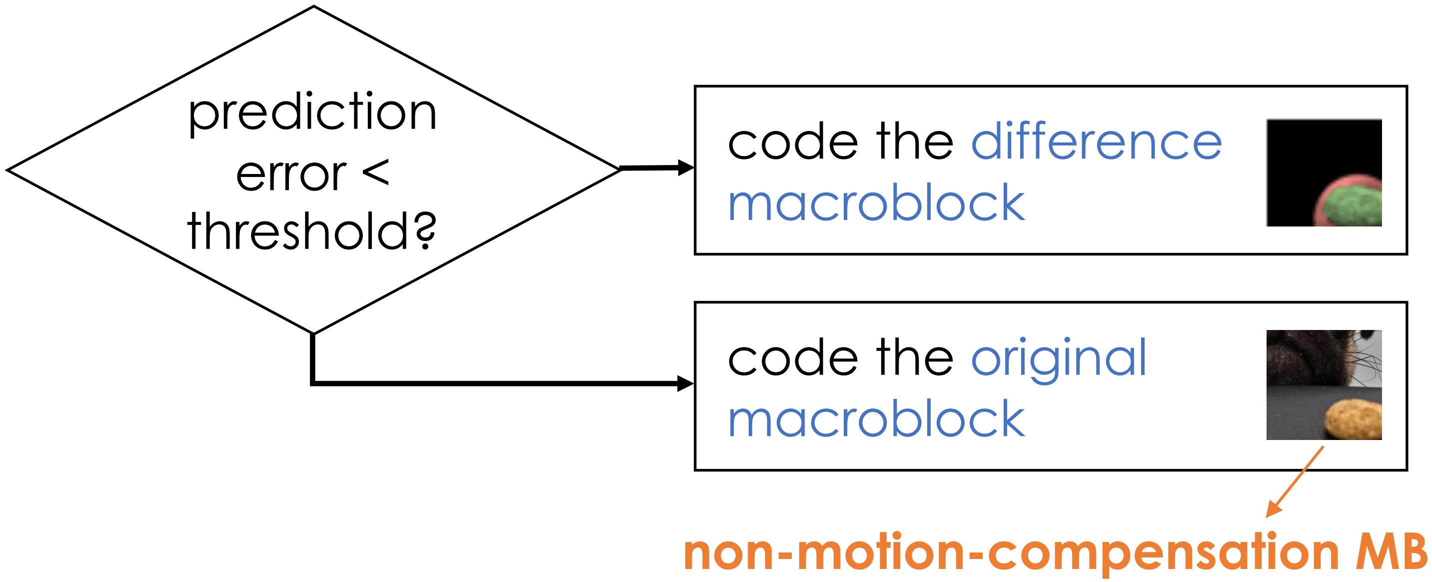


Example 2



P-frame Coding

- Coded based on motion compensation
  - For each macroblock, find the motion vector
  - Find the difference between the target MB and reference MB
    → prediction error = difference macroblock

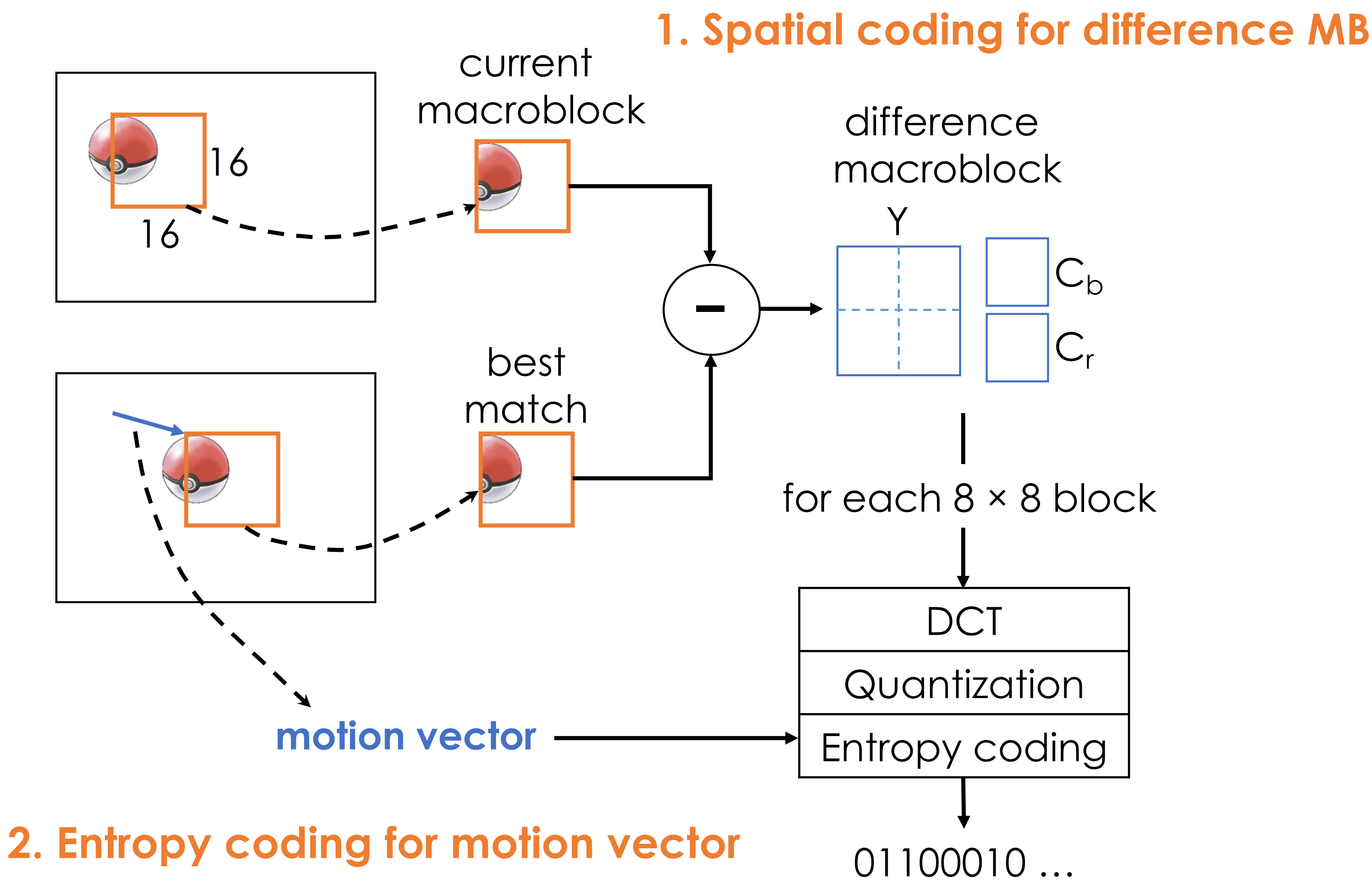


P-frame Coding

- Coded based on motion compensation
  - For each macroblock, find the motion vector
  - Find the difference between the target MB and reference MB
    → prediction error = difference macroblock

- Code the difference macroblock if the error is small enough
  - Difference MV usually has a much smaller entropy
- In fact, motion vector (MV) is not coded directly
- Instead, the difference between motion vectors (MVD) is coded
  - **$MVD = MV_{preceding} - MV_{current}$**

P-frame Coding



Summary of Hybrid Coding

- Temporal redundancy
  - Removed by Motion Estimation/Compensation
- Spatial redundancy
  - Removed by Transform coding
- Statistic redundancy
  - Removed by Entropy coding (VLC)
  - Applied for both 8x8 blocks and motion vectors

Outline

- Introduction
- Frame sequence
- Frame coding
- **Quantization**
- Encoder and decoder
- H.261 syntax

Quantization

- Does not use 8 × 8 quantization matrices (as in JPEG/MPEG)
- Instead, use a constant, called **`step_size`**, for all DCT coefficients
- **`step_size`** can be any one of the 31 even values from [2, 62]

$$\text{QDCT} = \left\lfloor \frac{\text{DCT}}{step\_size} \right\rfloor$$

- However, for the DC coefficient, **`step_size`** is always set to 8

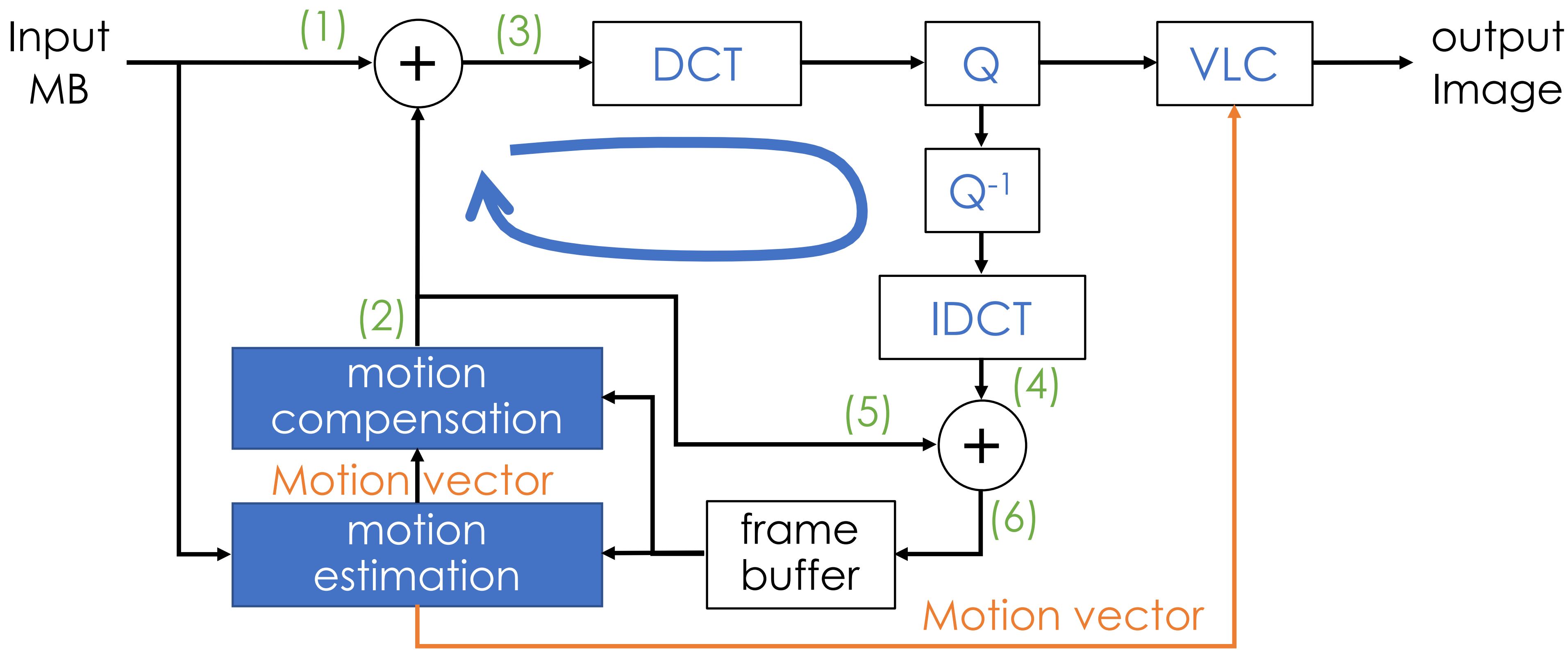
$$\text{QDCT} = \text{round}\left(\frac{\text{DCT}}{step\_size} \right)$$

Outline

- Introduction
- Frame sequence
- Frame coding
- Quantization
- **Encoder and decoder**
- H.261 syntax

Coding Loop

- Encoder and decoder should maintain the same reference frames (quantized one!)
  - To avoid error propagation (error drift)



Encoding I-frame

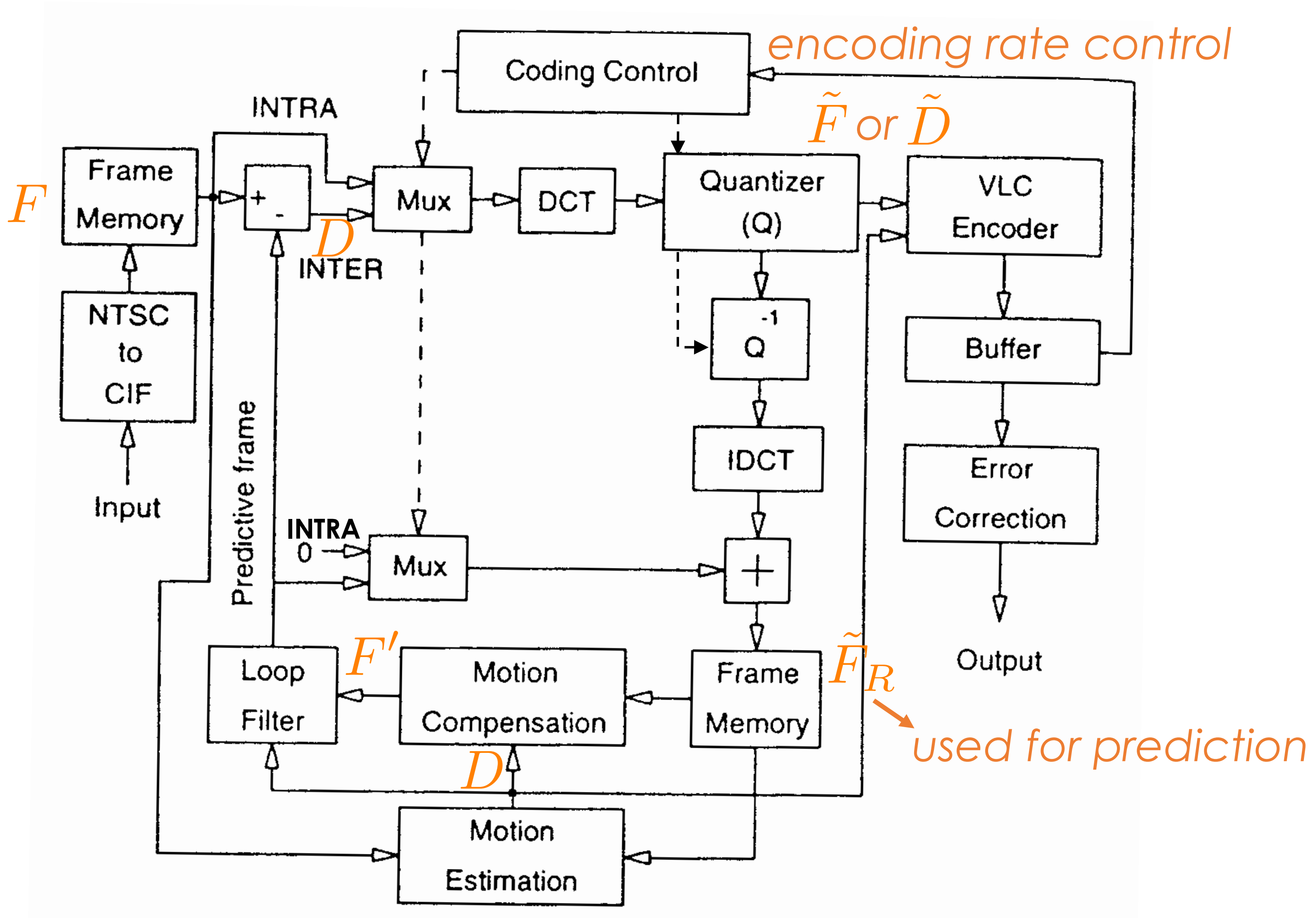
1. Receive macroblocks from the I-frame
   - Go to DCT, Q, VLC and be output
4. Use Q^{-1} and IDCT to get to get the reconstructed frame $\tilde{I}$
5. Combine with a zero input
6. Remain as $\tilde{I}$, stored in the frame buffer as the reference frame for the following P-frame

Why store the reconstructed $\tilde{\mathbf{I}}$ as the reference frame, instead of the original **I**?

Encoding P-frame

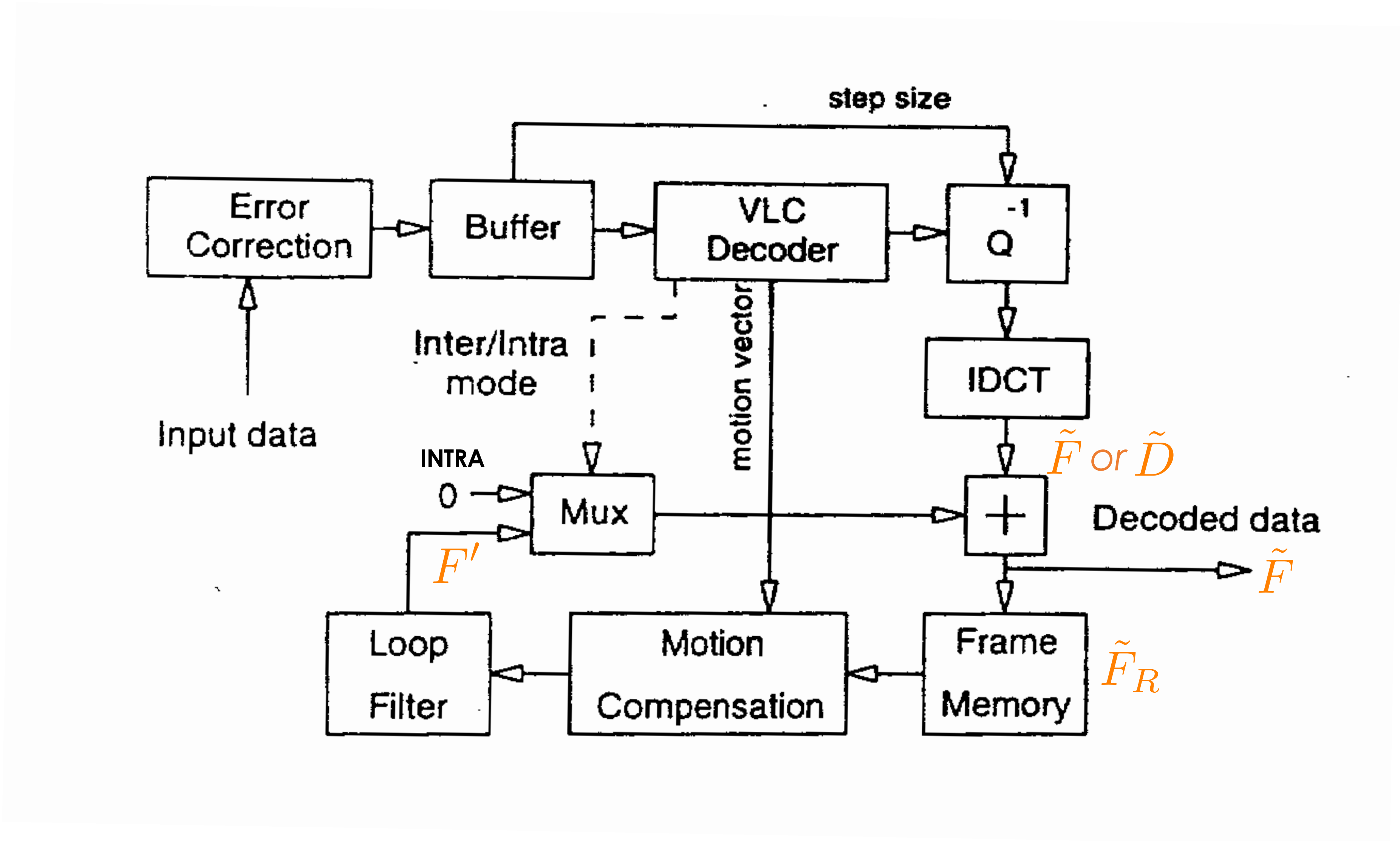
1. Receive macroblocks from the P-frame
   - Go to motion estimation
   - Find the motion vector best matching $\tilde{I}$
   - Send the motion vector to VLC
2. Yield the best matching MB **P'**
3. Find the difference (prediction error)
   D = P – P'
   - Send the error D to DCT, Q and VLC
4. Also send **D** to Q and IDCT to reconstruct $\tilde{D}$
5. Find the reconstructed P-frame $\tilde{P} = P' + \tilde{D}$
6. Store $\tilde{P}$ in the frame buffer as the reference frame of the next P-frame

Encoding System



- Switching between intra- and inter-frame modes by a multiplexer
- "Coding control" controls **`step_size`** according to the buffer level

Decoding System



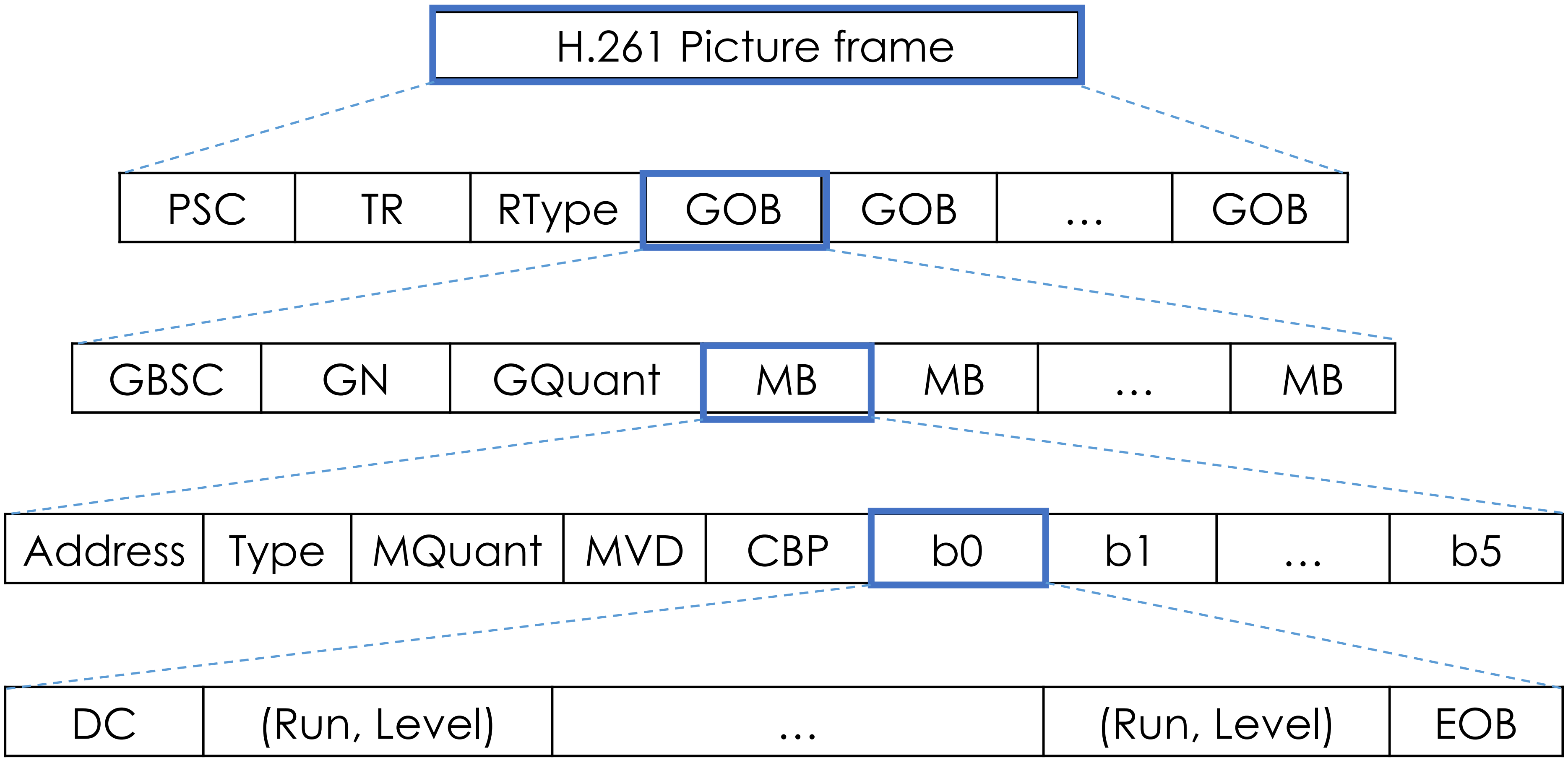
Outline

- Introduction
- Frame sequence
- Frame coding
- Quantization
- Encoder and decoder
- **H.261 syntax**

Coding Hierarchy

- Defined in the standard to ensure that the system can be interpreted universally
- Four-layer hierarchy
  - Picture layer: corresponding to one video frame (CIF or QCIF)
  - Group of block (GOB) layer: each is of 11×3 macroblocks (i.e., 176×48 pixels in luminance images), corresponding to 1/12 of CIF or 1/3 of QCIF
  - Macroblock layer: corresponding to 16×16 pixels of luminance (Y) and 8x8 of chrominance (Cb, Cr)
  - Block layer: corresponding to 8×8 pixels, coded by *DCT and run-length coding*

H.261 Video Bitstream



Blocks Arrangement

1	2
3	4
5	6
7	8
9	10
11	12

CIF

1
2
3

QCIF

(a) GOB arrangement in a picture

1	2
3	4

Y

6	2
Cb	Cr

(c) block arrangement in a MB

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

(b) macroblock arrangement in a GOB

Macroblock Type

Prediction	MQUANT	MVD	CBP	TCOEFF	VLC
Intra				x	0001
Intra	x			x	0000001
Inter			x	x	1
Inter	x		x	x	00001
Inter + MC		x			000000001
Inter + MC		x	x	x	00000001
Inter + MC	x	x	x	x	0000000001
Inter + MC + FIL		x			001
Inter + MC + FIL		x	x	x	01
Inter + MC + FIL	x	x	x	x	000001

H.261

- Does not specify:
  - Preprocessing and post processing
  - The criteria for choosing the mode for coding a macroblock
  - The use of BCH (511,493) in the decoder
  - Motion estimation in the encoder
  - The quantizer decision levels
  - Rate-control algorithm
  - Frame-rate
- Specifies:
  - Bit-stream syntax and decoding
  - A macroblock should be forcibly updated at least once per every 132 times it is transmitted
  - For CIF, the number of bits created by coding any single picture must not exceed 256 kb; for QCIF, 64 kb
  - Hypothetical Reference Decoder (HRD)

Summary

- **Frame sequence**
  - Consist of I-frames and P-frames
- **Frame coding**
  - I-frames are coded similar to JPEG
  - P-frames are predicted by I-frames
- **Quantization**
  - Slightly different from JPEG
- **Encoder and decoder**
  - Code loop for implementing motion estimation and prediction
- **H.261 syntax**