MPEG Video Coding I

Lecturer：Jun Xiao
(肖俊）
College of Software and Technology

Overview

- MPEG
  - Moving Pictures Experts Group, established in 1998 to delivery digital video and audio
- Membership
  - From 25 experts in 1988, to more than 350 experts form about 200 companies

Content

- **MPEG-1**
  - Motion Compensation in MPEG-1
  - Other Major Differences from H.261
  - MPEG-1 Video Bitstream
- **MPEG-2**
  - Introduction
  - Interlaced Video
  - MPEG-2 Scalabilities
  - Other Major Differences from MPEG-1

1. MPEG-1

Introduction of MPEG-1

- 1991 ISO/IEC
  - Coding of moving pictures and associated audio for Digital storage media at up to about 1.5Mbit/s
  - CDs and VCDs; 1.2M for Video and 256K for audio
  - 5 parts: Systems, Video, Audio, Conformance, Software
- Adopt the CCIR601 digital TV format -- SIF（Source Input Format）
  - Support only non-interlaced Video
  - 352*240 for NTSC at 30 fps
  - 352*288 for PAL at 25 fps
  - 4:2:0 chroma subsampling

1.1 Motion Compensation in MPEG-1

- Motion Compensation (MC) based video encoding in H.261 works as follows:
  - **In Motion Estimation (ME), each macroblock (MB) of the Target P-frame is assigned a best matching MB from the previously coded I or P frame - prediction.**
  - **prediction error: The difference between the MB and its matching MB, sent to DCT and its subsequent encoding steps.**
  - **The prediction is from a previous frame — forward prediction.**

1.1 Motion Compensation in MPEG-1

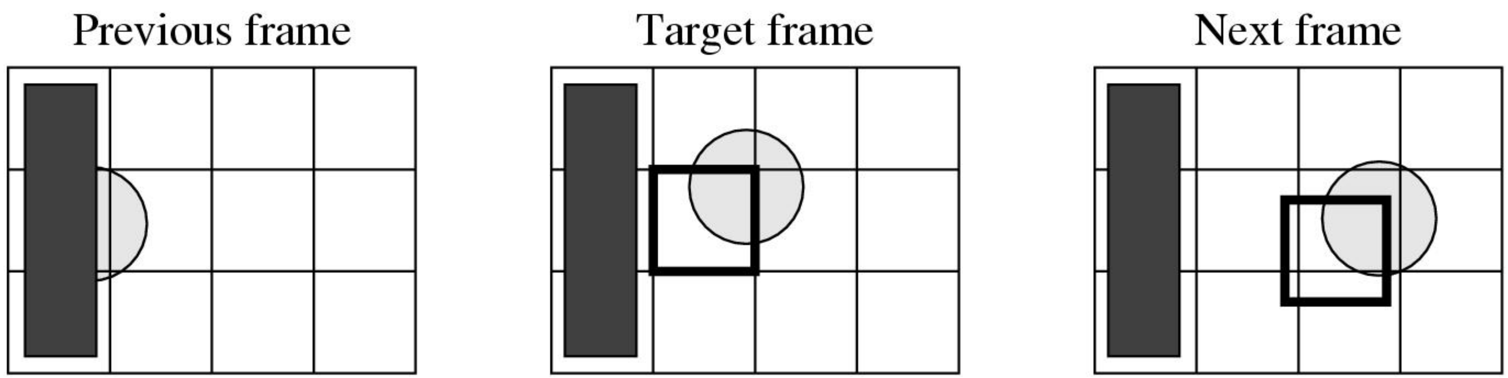
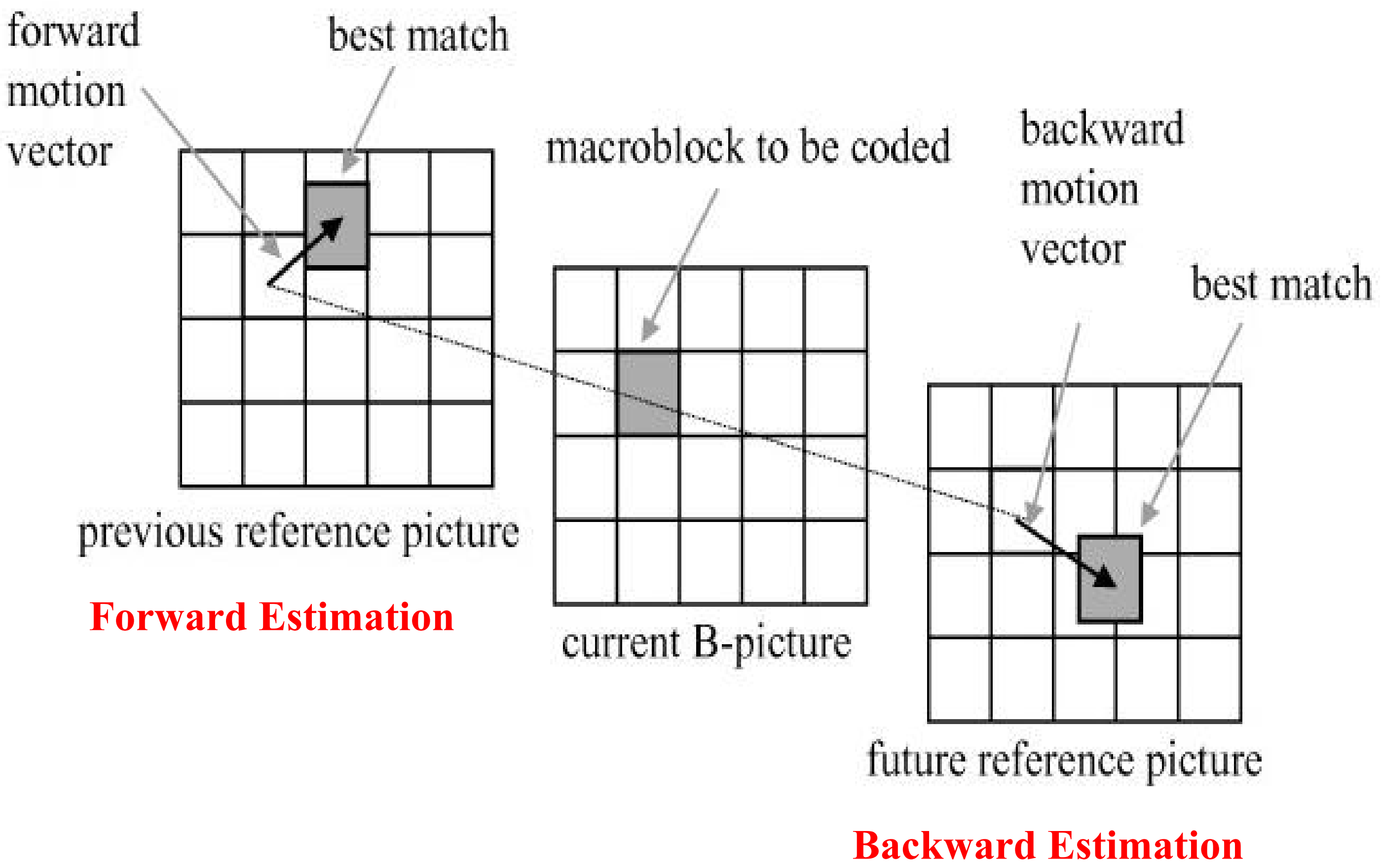


Fig 11.1: The Need for Bidirectional Search.

The MB containing part of a ball in the Target frame cannot find a good matching MB in the previous frame because half of the ball was occluded by another object. A match however can readily be obtained from the next frame.

1.1 Motion Compensation in MPEG-1

- MPEG introduces a third frame type — *B-frames*, and its accompanying bi-directional motion compensation.



1.1 Motion Compensation in MPEG-1

- The MC-based B-frame coding idea is illustrated in Fig. 11.2:
  - Each MB from a B-frame will have up to *two* motion vectors (MVs) (one from the forward and one from the backward prediction).
  - If matching in both directions is successful, then two MVs will be sent and the two corresponding matching MBs are averaged (indicated by '%' in the figure) before comparing to the Target MB for generating the prediction error.
  - If an acceptable match can be found in only one of the reference frames, then only one MV and its corresponding MB will be used from either the forward or backward prediction.

1.1 Motion Compensation in MPEG-1

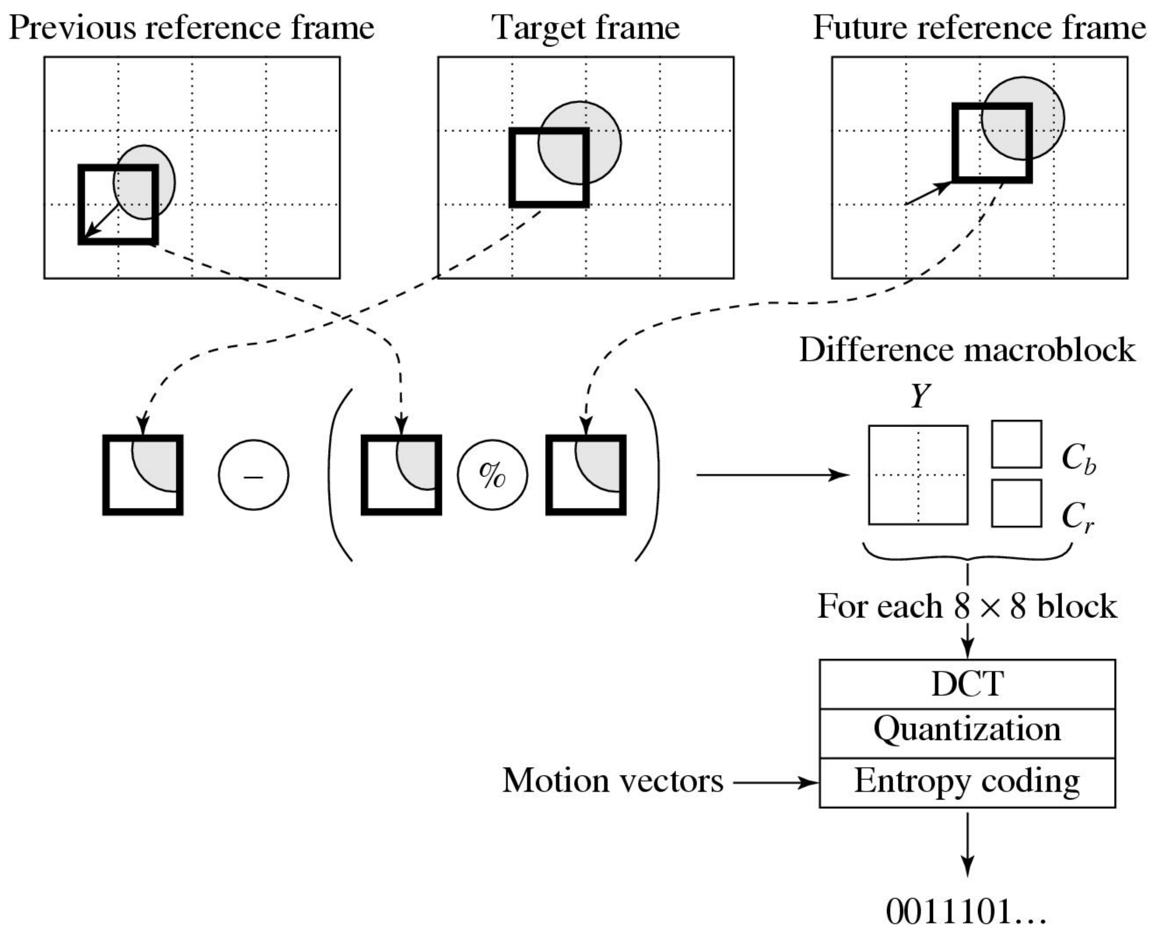


Fig 11.2: B-frame Coding Based on Bidirectional Motion Compensation.

1.1 Motion Compensation in MPEG-1

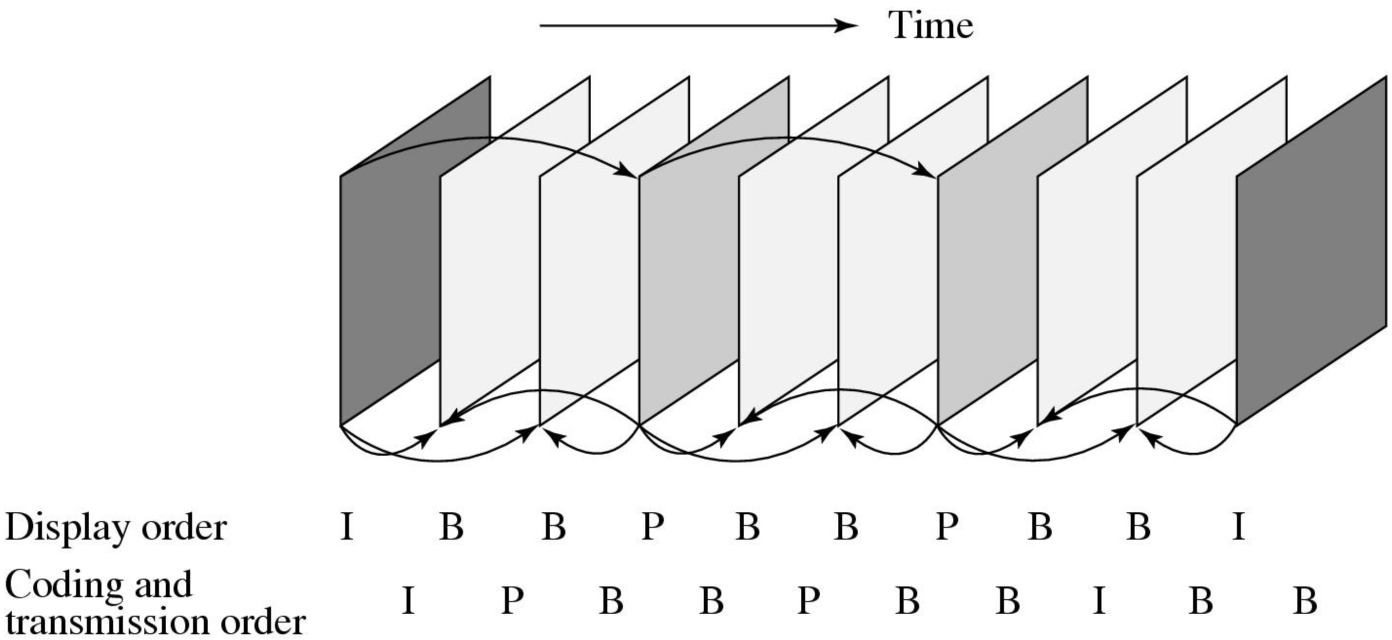
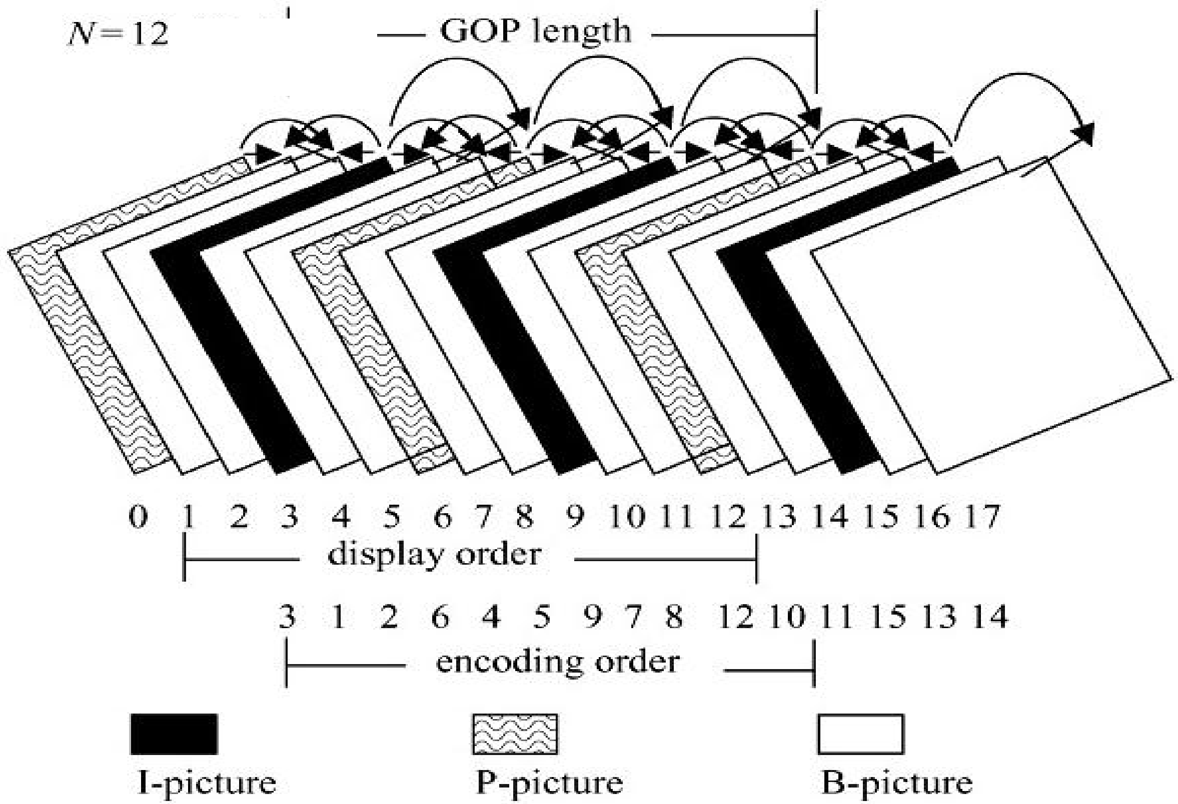


Fig 11.3: MPEG Frame Sequence.

1.1 Motion Compensation in MPEG-1

- B frame dependents on its succeeding P or I frame, so the play order is different from the coding order.



1.2 Other Major Differences from H.261

- H.261
  - Supports only CIF(352×288), QCIF (176×144)
- MPEG-1
  - Supports SIF(352×240 for NTSC ,352×288 for PAL),
  - Allows specification of other formats

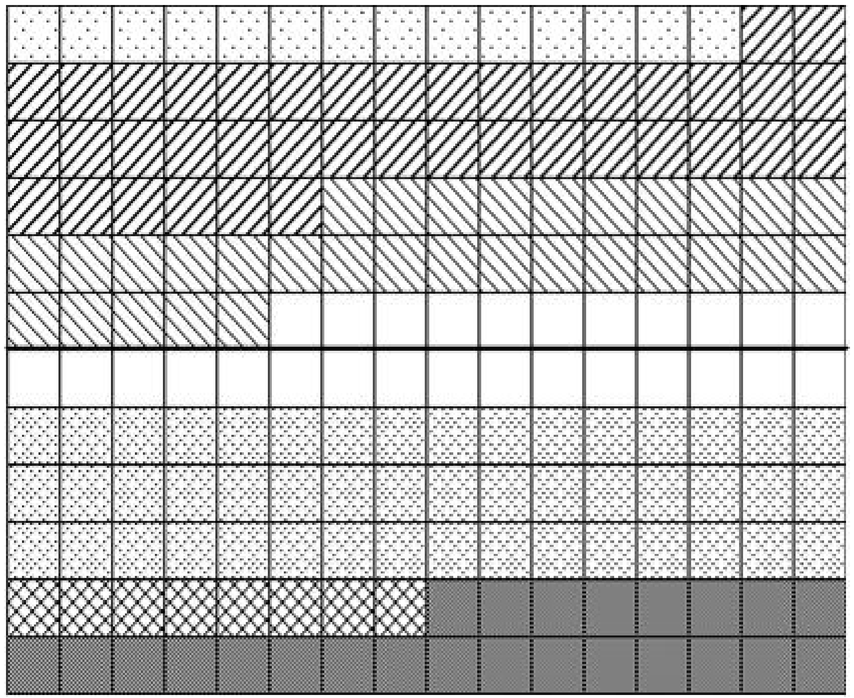
Parameter	Value
Horizontal size of picture	≤ 768
Vertical size of picture	≤ 576
No. of MBs / picture	≤ 396
No. of MBs / second	≤ 9,900
Frame rate	≤ 30 fps
Bit-rate	≤ 1,856 kbps

1.2 Other Major Differences from H.261

- Instead of GOBs as in H.261, an MPEG-1 picture can be divided into one or more slices (Fig. 11.4):
  – May contain variable numbers of macroblocks in a single picture.
  – May also start and end anywhere as long as they fill the whole picture.
  – Each slice is coded independently — additional flexibility in bit-rate control.
  – Slice concept is important for error recovery.

1.2 Other Major Differences from H.261

- MPEG-1 picture can be divided into one or more slices, each slice in coded independently.



Slices in an MPEG-1 picture

1.2 Other Major Differences from H.261

- Quantization:
  - **MPEG-1 quantization uses different quantization tables for its Intra and Inter coding (Table 11.2 and 11.3).**

For DCT coefficients in Intra mode:

$$QDCT[i,j] = round\left(\frac{8 \times DCT[i,j]}{step\_size[i,j]}\right) = round\left(\frac{8 \times DCT[i,j]}{Q_1[i,j] * scale}\right) \quad (11.1)$$

For DCT coefficients in Inter mode:

$$QDCT[i,j] = \left\lfloor \frac{8 \times DCT[i,j]}{step\_size[i,j]} \right\rfloor = \left\lfloor \frac{8 \times DCT[i,j]}{Q_2[i,j] * scale} \right\rfloor \quad (11.2)$$

1.2 Other Major Differences from H.261

- **The quantization numbers for intra-coding vary within a macroblock, which is different from H.261**

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

Intra Quantization Table

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

Inter Quantization Table

1.2 Other Major Differences from H.261

- MPEG-1 allows motion vectors to be of sub-pixel precision (1/2 pixel). The technique of “bilinear interpolation” for H.263 can be used to generate the needed values at half-pixel locations.

- Compared to the maximum range of ± 15 pixels for motion vectors in H.261, MPEG-1 supports a range of $[-512, 511.5]$ for half-pixel precision and $[-1{,}024, 1{,}023]$ for full-pixel precision motion vectors.

- The MPEG-1 bitstream allows random access — accomplished by GOP layer in which each GOP is time coded.

1.2 Other Major Differences from H.261

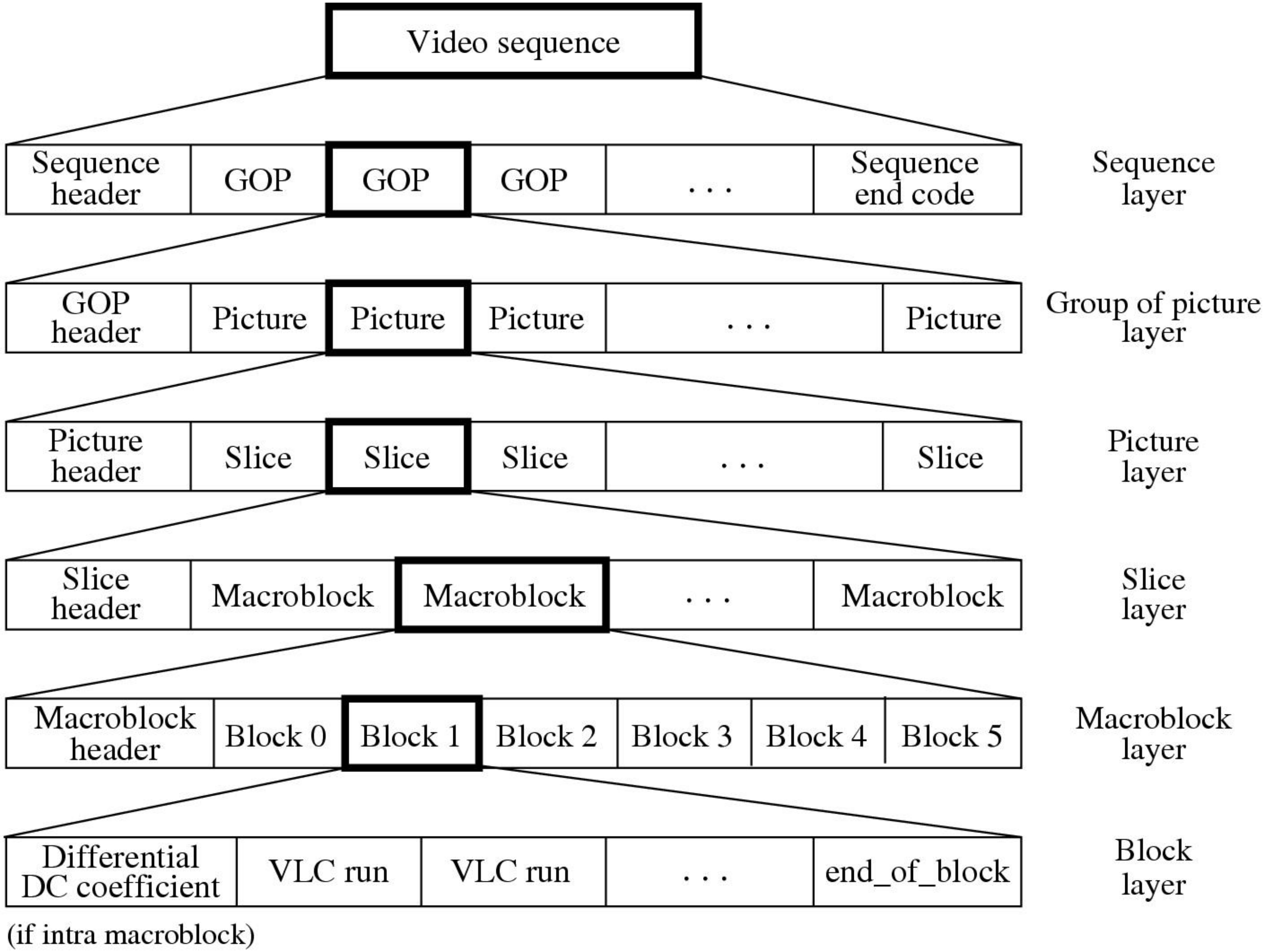
- The typical size of compressed P-frames is significantly smaller than that of I-frames — because temporal redundancy is exploited in inter-frame compression.

- B-frames are even smaller than P-frames — because of (a) the advantage of bi-directional prediction and (b) the lowest priority given to B-frames.

Table 11.4: Typical Compression Performance of MPEG-1 Frames

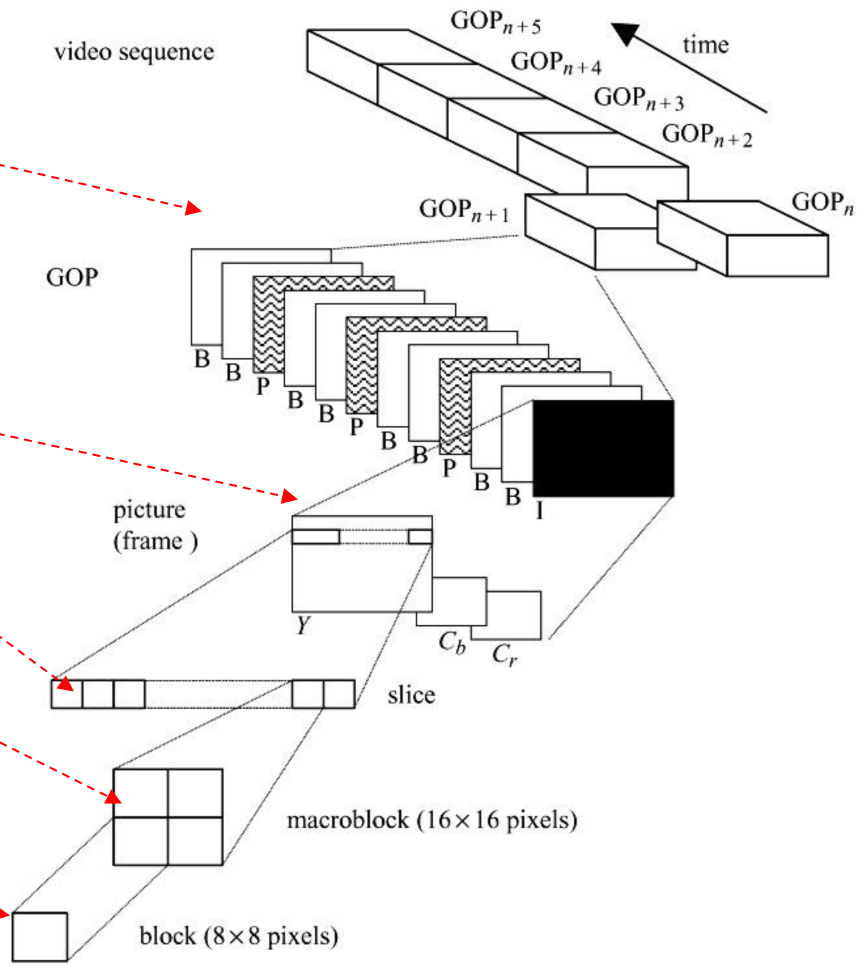
Type	Size	Compression
I	18kB	7:1
P	6kB	20:1
B	2.5kB	50:1
Avg	4.8kB	27:1

1.3 MPEG-1 Video Bitstream



1.3 MPEG-1 Video Bitstream

- GOP Layer : one or more pictures, one of which must be an I-picture
- Frame: can be I、P、or B type
- Slice: GOB (Group of Macroblock)
- Macroblock: 16*16 macroblock, subsample at 4:2:0, 4 Y blocks, 1 U block, 1 V block
- Block： unit for DCT and quantization



2. MPEG-2

2.1 Overview

- Started in 1990, for higher-quality video at a bitrate of more than 4 Mbps.
- Meets the compression and bitrate requirements of digital TV/HDTV
- Different resolutions, different compression complexities.
- Gained wide acceptance： terrestrial, satellite, cable network
- Other applications： Interactive TV, DVDs (digital video discs or digital versatile discs)

2.1 Overview

- MPEG-2 defined 7 profiles, aimed at different applications, up to 4 levels defined in each profile.
  - Simple, Main, SNR Scalable, Spatially Scalable, High, 4:2:2 and Multiview

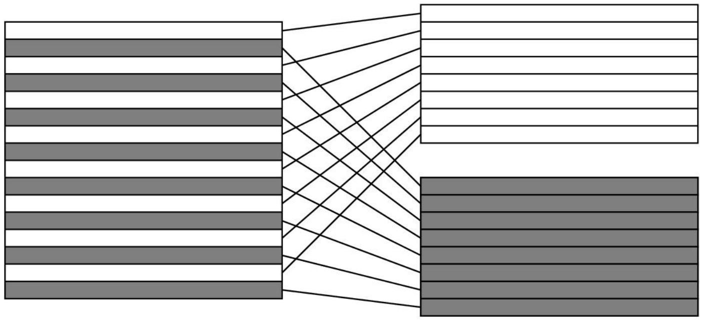
Level	Maximum	Maximum fps	Maximum Pixels$\times 10^6$/sec	Maximum coded Data rate(Mbps)	Application
High	19201152	60	62.7	80	Film production
High1440	1440$\times$1152	60	47.0	60	Consumer HDTV
Main	720$\times$576	30	10.4	15	Studio TV
Low	352$\times$288	30	3.0	4	Consumer tape equivalent

Four levels in the main profile of MPEG-2

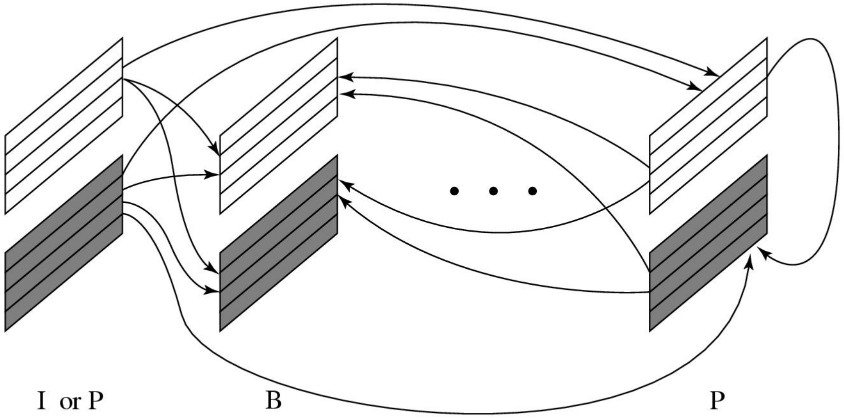
2.2 Supporting Interlaced Video

- MPEG-2 must support interlaced video as well since this is one of the options for digital broadcast TV and HDTV.
- In interlaced video each frame consists of two fields, referred to as the *top-field* and the *bottom-field*.
  - **In a *Frame-picture*, all scanlines from both fields are interleaved to form a single frame, then divided into 16×16 macroblocks and coded using MC.**
  - **If each field is treated as a separate picture, then it is called *Field-picture*.**

2.2 Supporting Interlaced Video



(a)



(b)

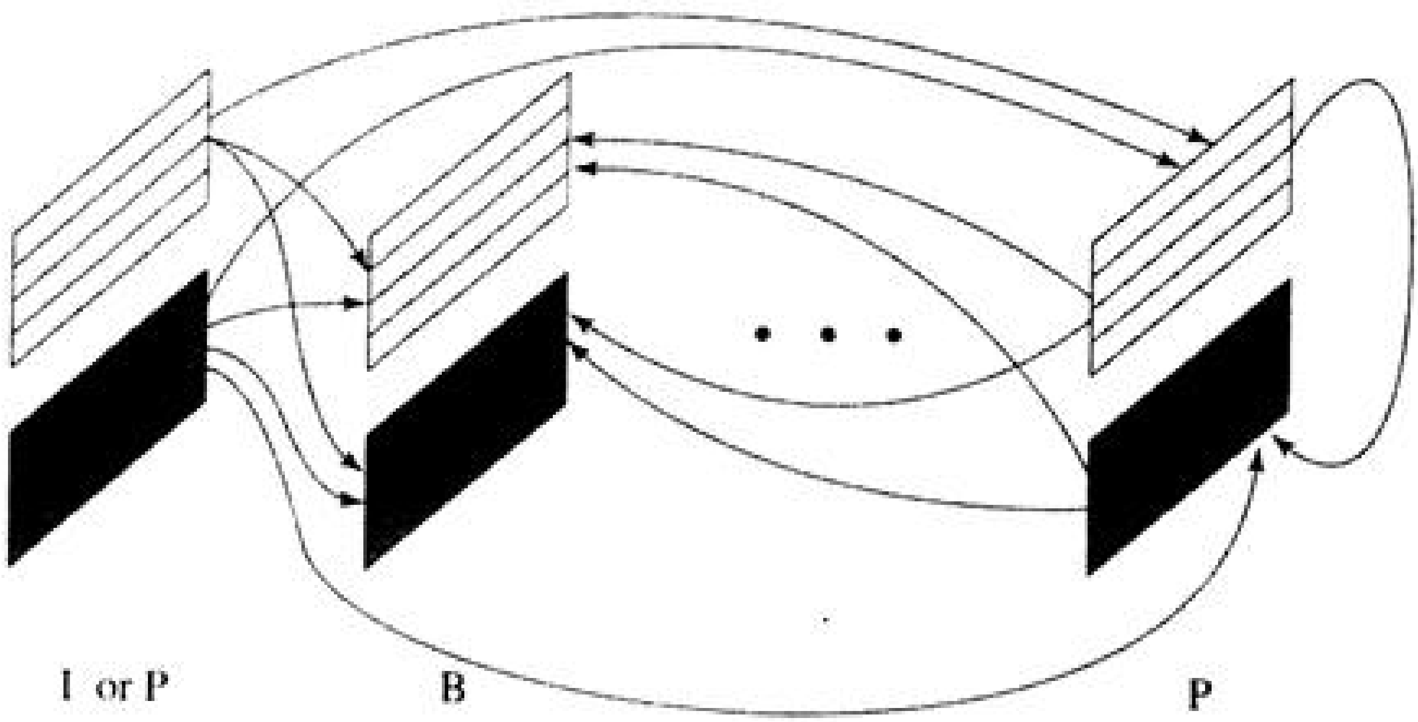
Fig. 11.6: Field pictures and Field-prediction for Field-pictures in MPEG-2.
(a) Frame−picture vs. Field−pictures, (b) Field Prediction for Field−pictures

2.2 Supporting Interlaced Video

- Five Modes of predictions
  - Frame prediction for frame-pictures
    - Identical to Mpeg1 motion compensation
    - Works well for videos containing slow and moderate object
  - Field prediction for field-pictures
    - See figure in the next page
  - Field prediction for frame-pictures
    - Treat top field and bottom field separately
  - 16×8 MC for field-pictures
    - Good for motion is rapid and irregular
  - Dual-prime for P-pictures
    - MV is used to derive a calculated motion vector CV

2.2 Supporting Interlaced Video

- **Field prediction for field-pictures**



2.2 Supporting Interlaced Video

Alternate Scan and Field_DCT

- Techniques aimed at improving the effectiveness of DCT on prediction errors, only applicable to Frame-pictures in interlaced videos:
  - **Due to the nature of interlaced video the consecutive rows in the 8×8 blocks are from different fields, there exists less correlation between them than between the alternate rows.**
  - **Alternate scan recognizes the fact that in interlaced video the vertically higher spatial frequency components may have larger magnitudes and thus allows them to be scanned earlier in the sequence.**
- In MPEG-2, **Field_DCT** can also be used to address the same issue.

2.2 Supporting Interlaced Video

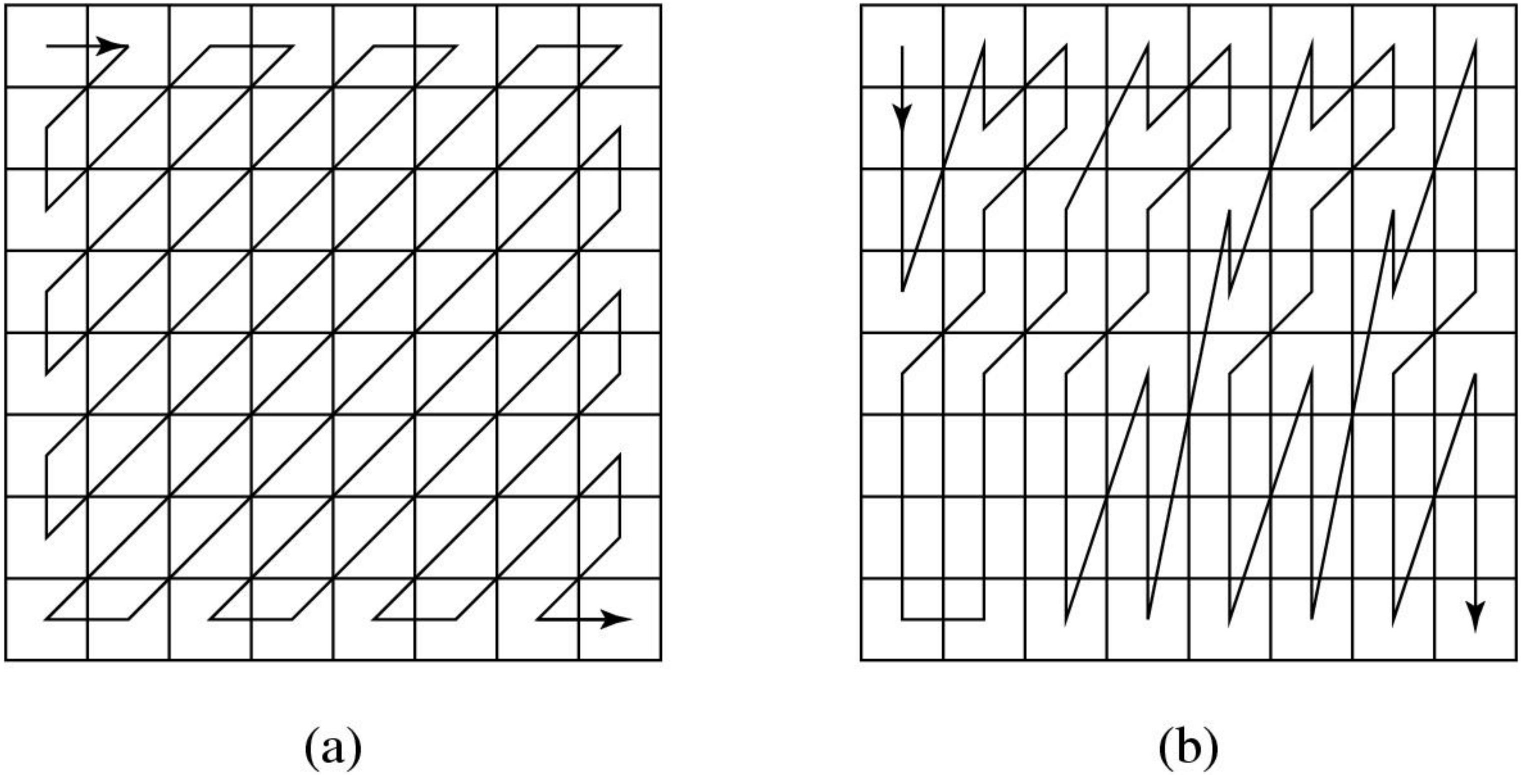


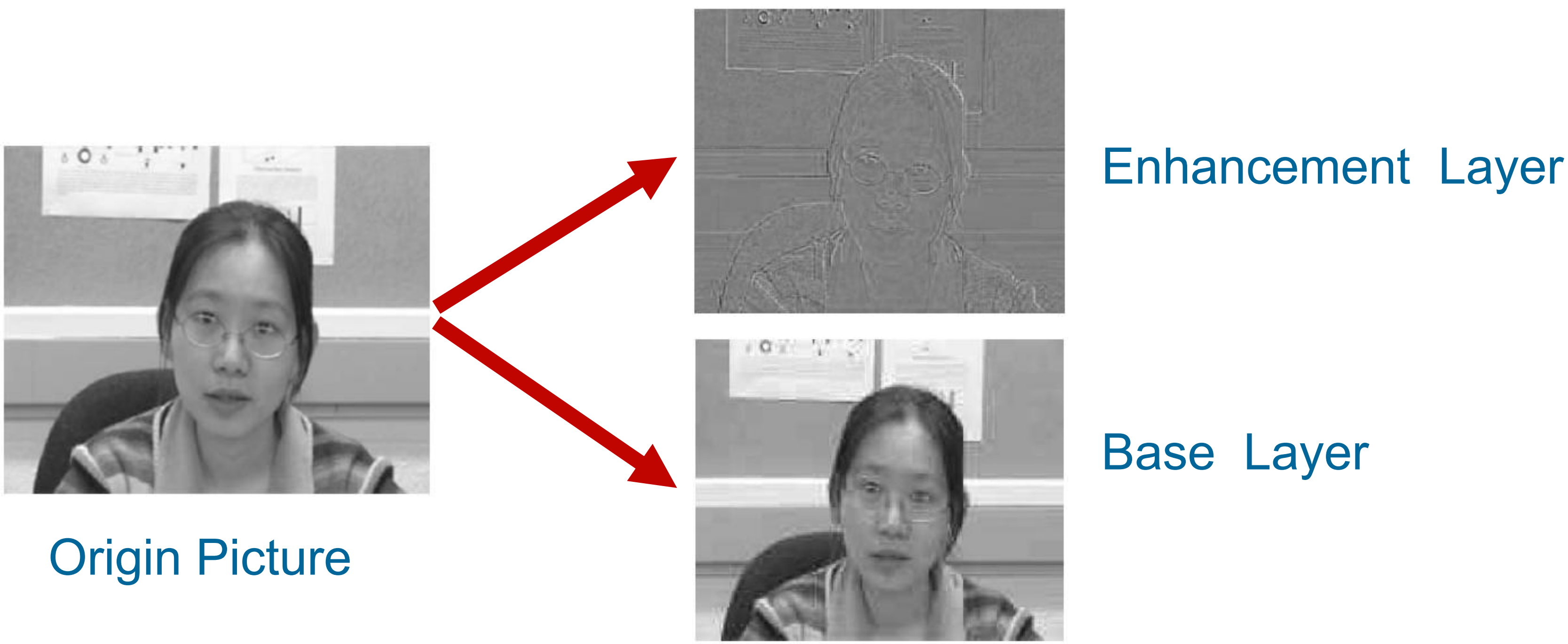
Fig 11.7: Zigzag and Alternate Scans of DCT Coefficients for Progressive and Interlaced Videos in MPEG-2.

2.3 MPEG-2 Scalabilities

- The MPEG-2 **scalable coding**: A base layer and one or more enhancement layers can be defined — also known **as layered coding**.
  - **The base layer can be independently encoded, transmitted and de-coded to obtain basic video quality.**
  - **The encoding and decoding of the enhancement layer is dependent on the base layer or the previous enhancement layer.**
- Scalable coding is especially useful for MPEG-2 video transmitted over networks with following characteristics:
  - **Networks with very different bit-rates.**
  - **Networks with variable bit rate (VBR) channels.**
  - **Networks with noisy connections.**

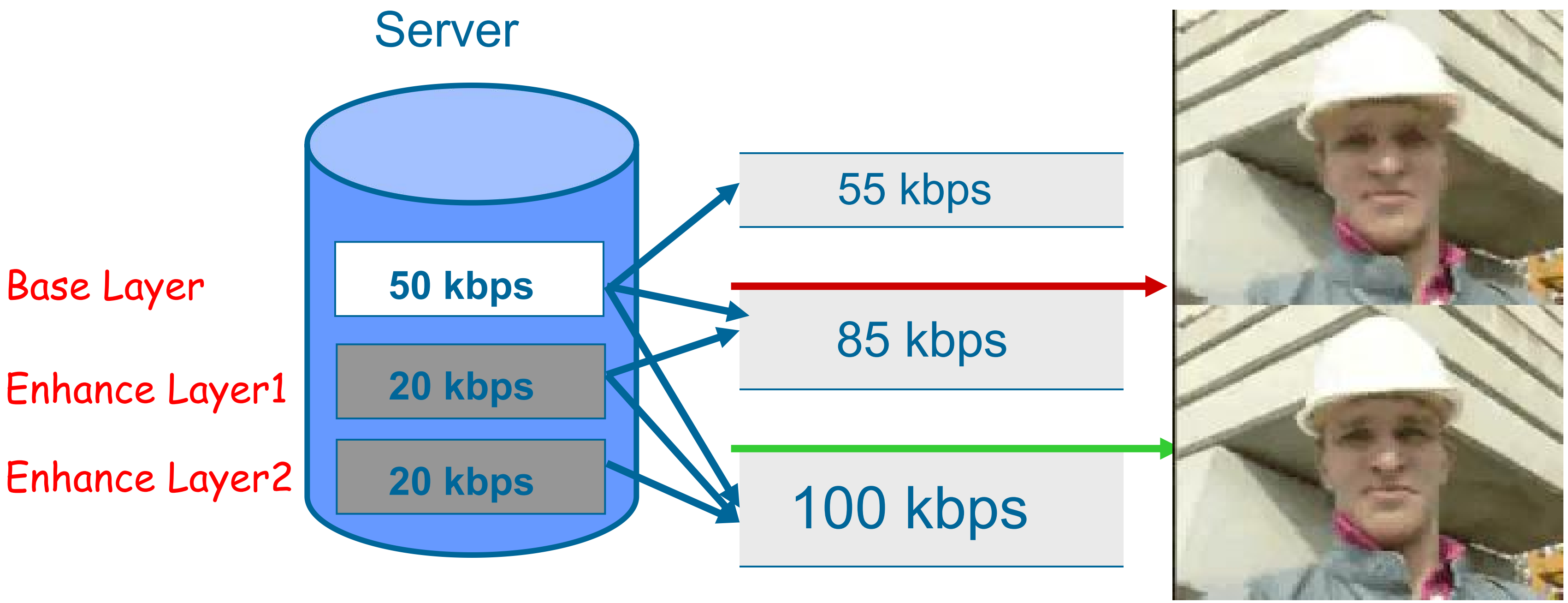
2.3 MPEG-2 Scalabilities

- Scalable Coding： The basic layer provide basic video quality, enhancement layers provide better qualities.
- The base layer is encoded independently, enhancement layer depends on the base layer or the previous enhancement layer.



2.3 MPEG-2 Scalabilities

- Send basic layer only over insufficient band
- Basic and enhancement layers for broad-band so that receivers can obtain better quality.



2.3 MPEG-2 Scalabilities

- MPEG-2 supports the following scalabilities:
  1. **SNR Scalability—enhancement layer provides higher SNR.**
  2. **Spatial Scalability — enhancement layer provides higher spatial resolution.**
  3. **Temporal Scalability—enhancement layer facilitates higher frame rate.**
  4. **Hybrid Scalability — combination of any two of the above three scalabilities.**
  5. **Data Partitioning — quantized DCT coefficients are split into partitions.**

2.3 MPEG-2 Scalabilities

- **SNR scalability**: Refers to the enhencement/refinement over the base layer to improve the Signal-Noise-Ratio (SNR).
- The MPEG-2 SNR scalable encoder will generate output bitstreams *Bits_base* and *Bits_enhance* at two layers:

  1. **At the Base Layer, a coarse quantization of the DCT coefficients is employed which results in fewer bits and a relatively low quality video.**
  2. **The coarsely quantized DCT coefficients are then inversely quantized (Q^{-1}) and fed to the Enhancement Layer to be compared with the original DCT coefficient.**
  3. **Their difference is finely quantized to generate a DCT coefficient refinement, which, after VLC, becomes the bitstream called Bits_enhance.**

2.3 MPEG-2 Scalabilities

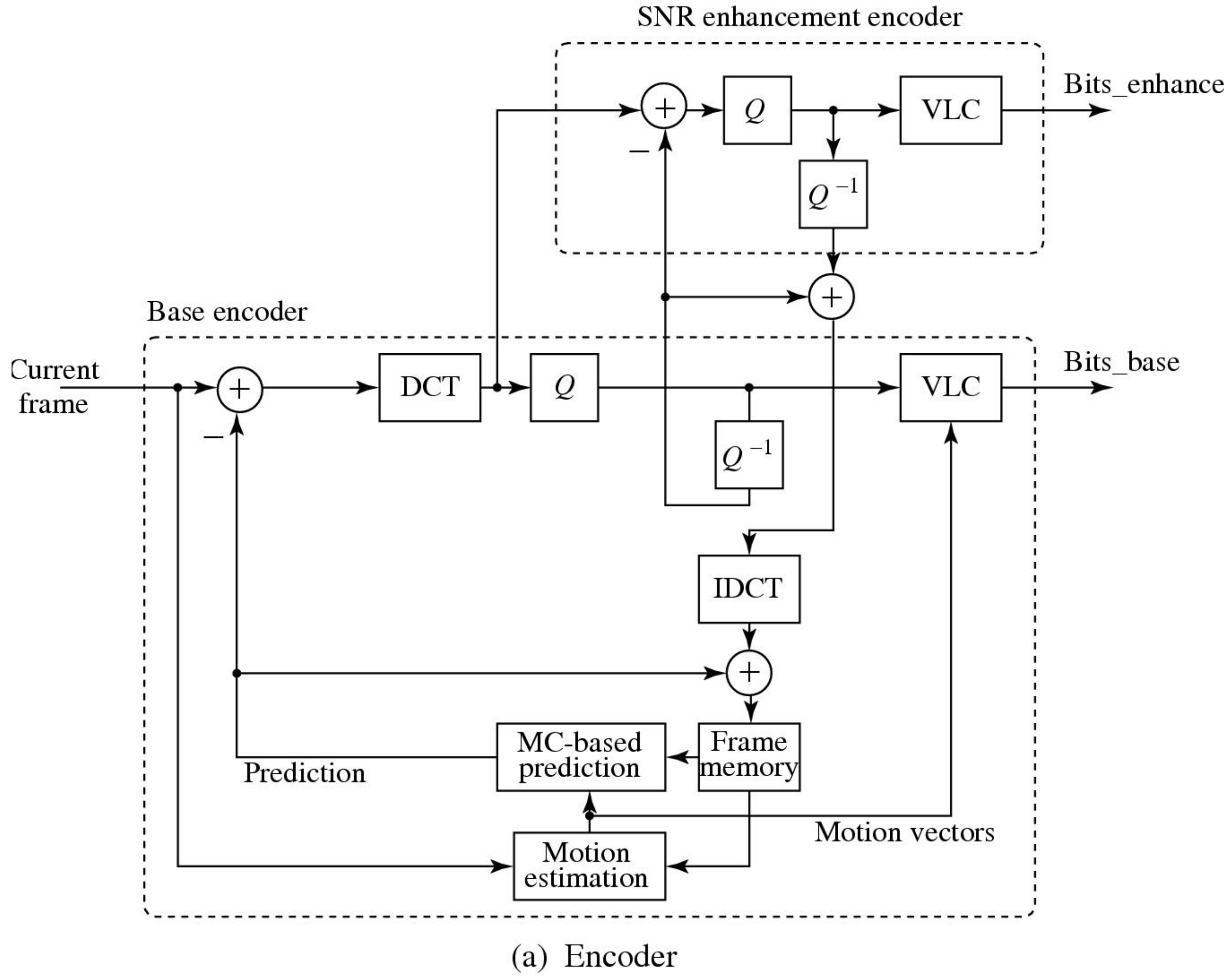


Fig 11.8 (a): MPEG-2 SNR Scalability (Encoder).

2.3 MPEG-2 Scalabilities

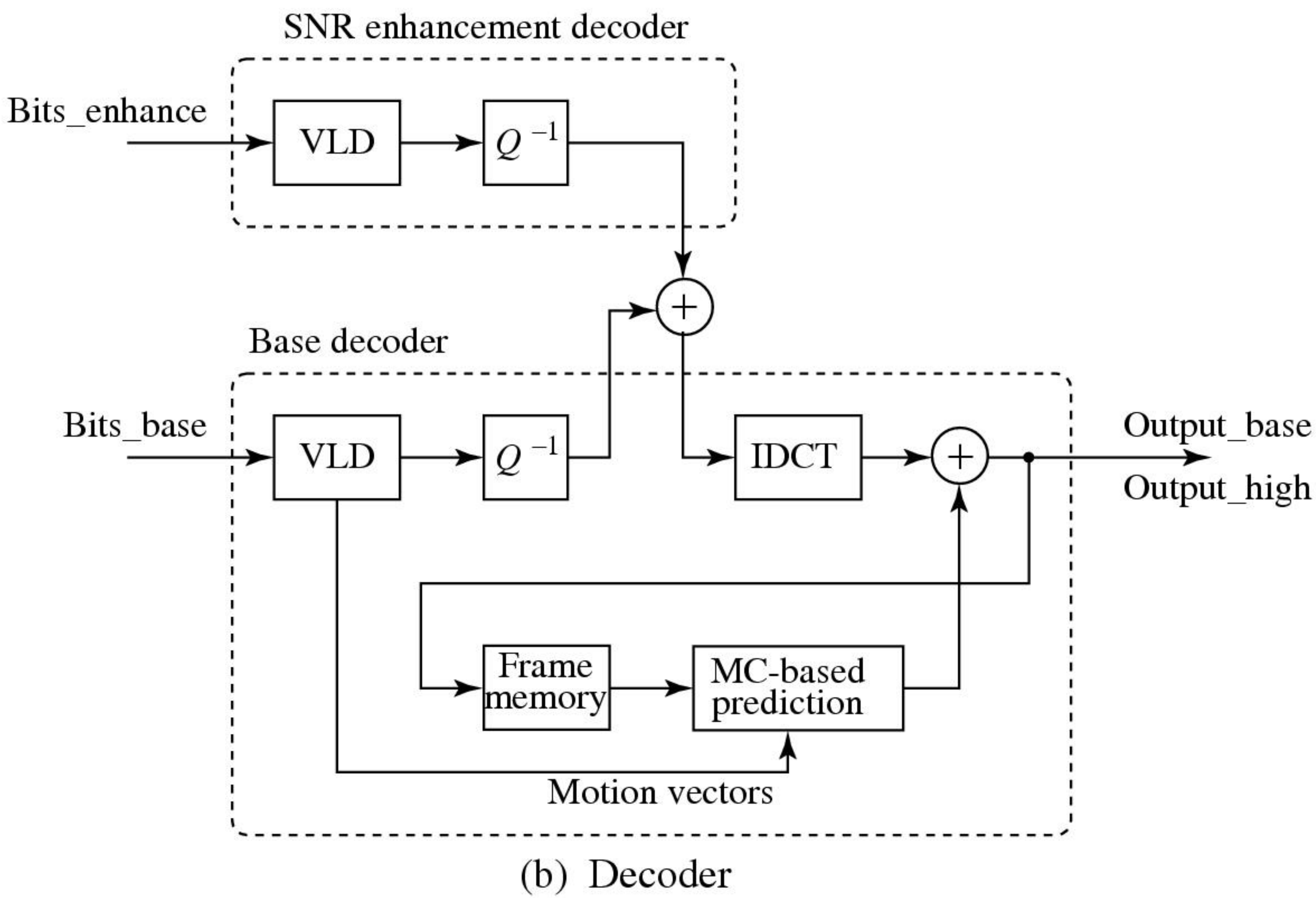
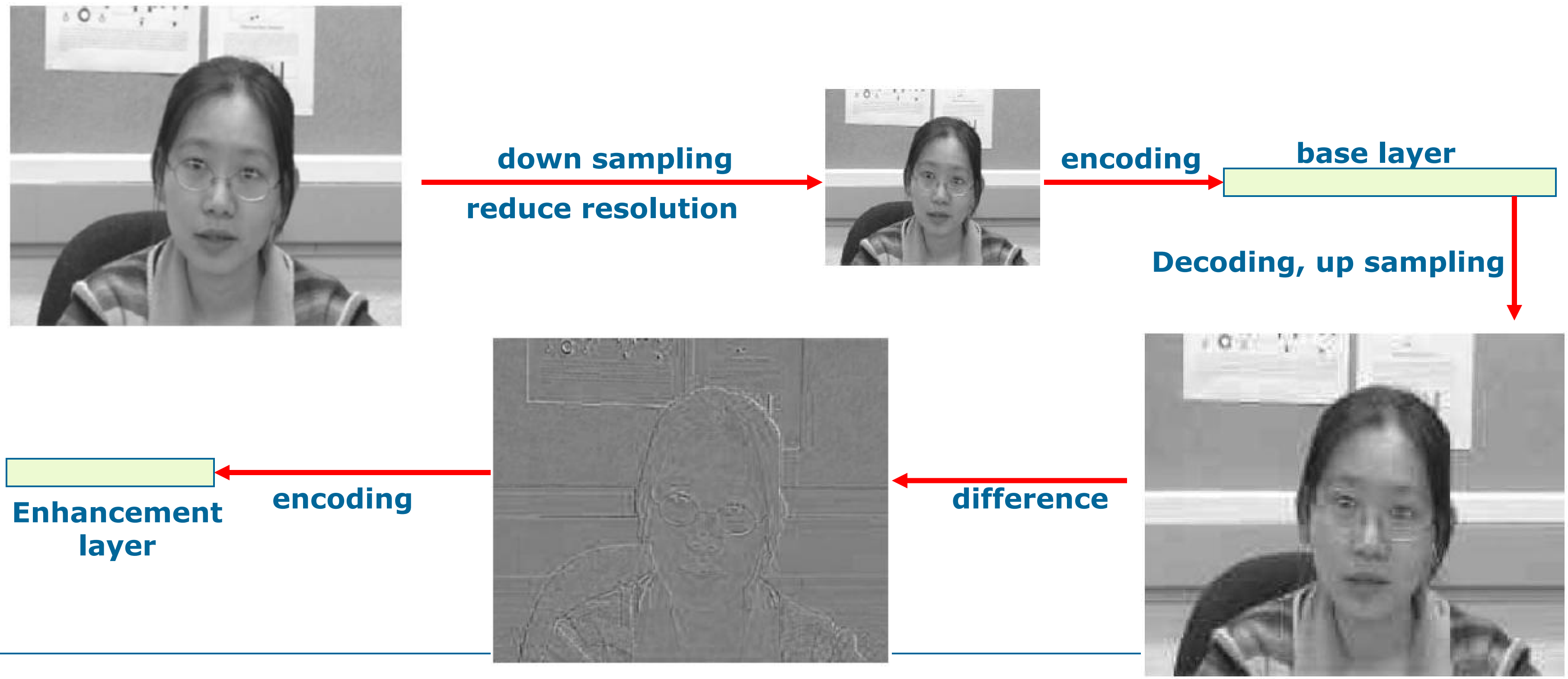


Fig 11.8 (b): MPEG-2 SNR Scalability (Decoder).

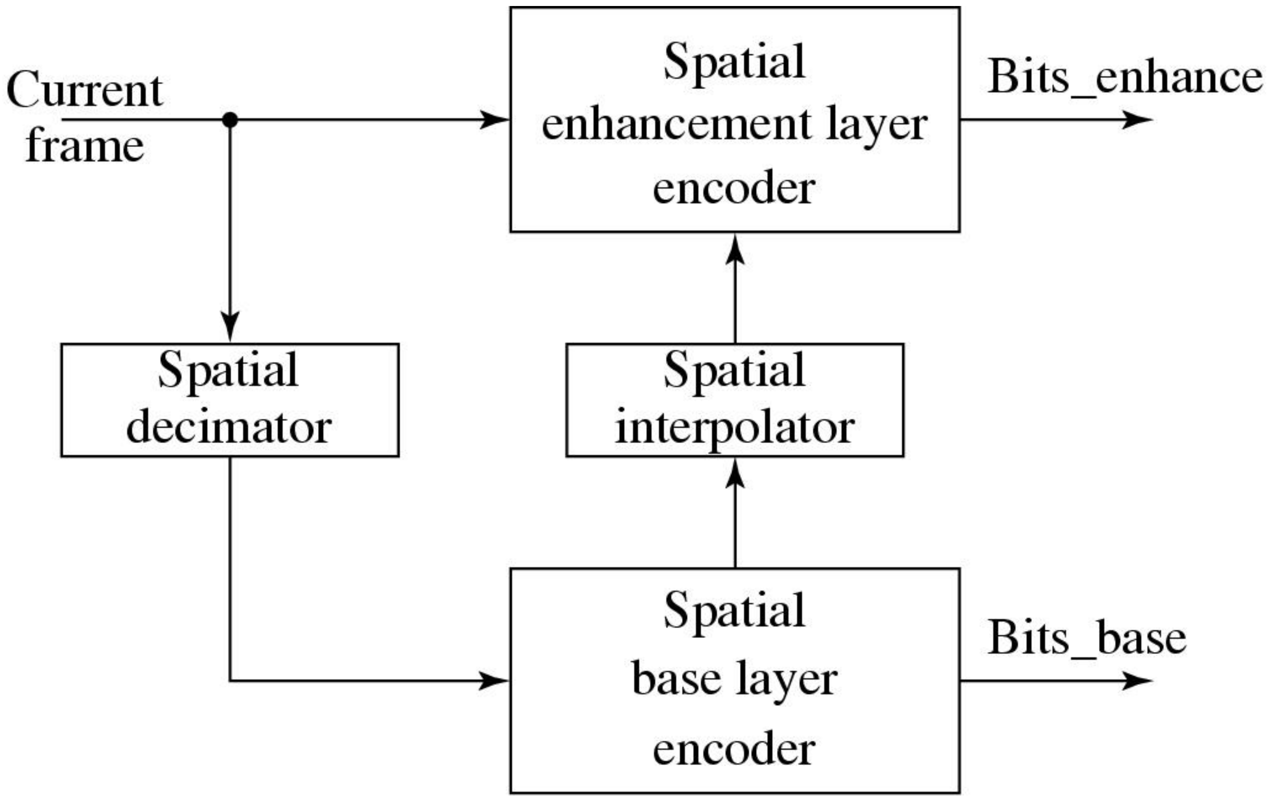
2.3 MPEG-2 Scalabilities

Spatial Scalability

The base layer is designed to generate bitstream of reduced resolution pictures. When combined with the enhancement layer, pictures at the original resolution are produced.



2.3 MPEG-2 Scalabilities



(a)

Encoder for MPEG-2 Spatial Scalability

2.3 MPEG-2 Scalabilities

Temporal Scalability

- The input video is temporally demultiplexed into two pieces, each carrying half of the original frame rate.
- Base Layer Encoder carries out the normal single-layer coding procedures for its own input video and yields the output bitstream Bits_base.
- The prediction of matching MBs at the Enhancement Layer can be obtained in two ways:
  - **Interlayer MC (Motion-Compensated) Prediction (Fig. 11.10(b))**
  - **Combined MC Prediction and Interlayer MC Prediction (Fig. 11.10(c))**

2.3 MPEG-2 Scalabilities

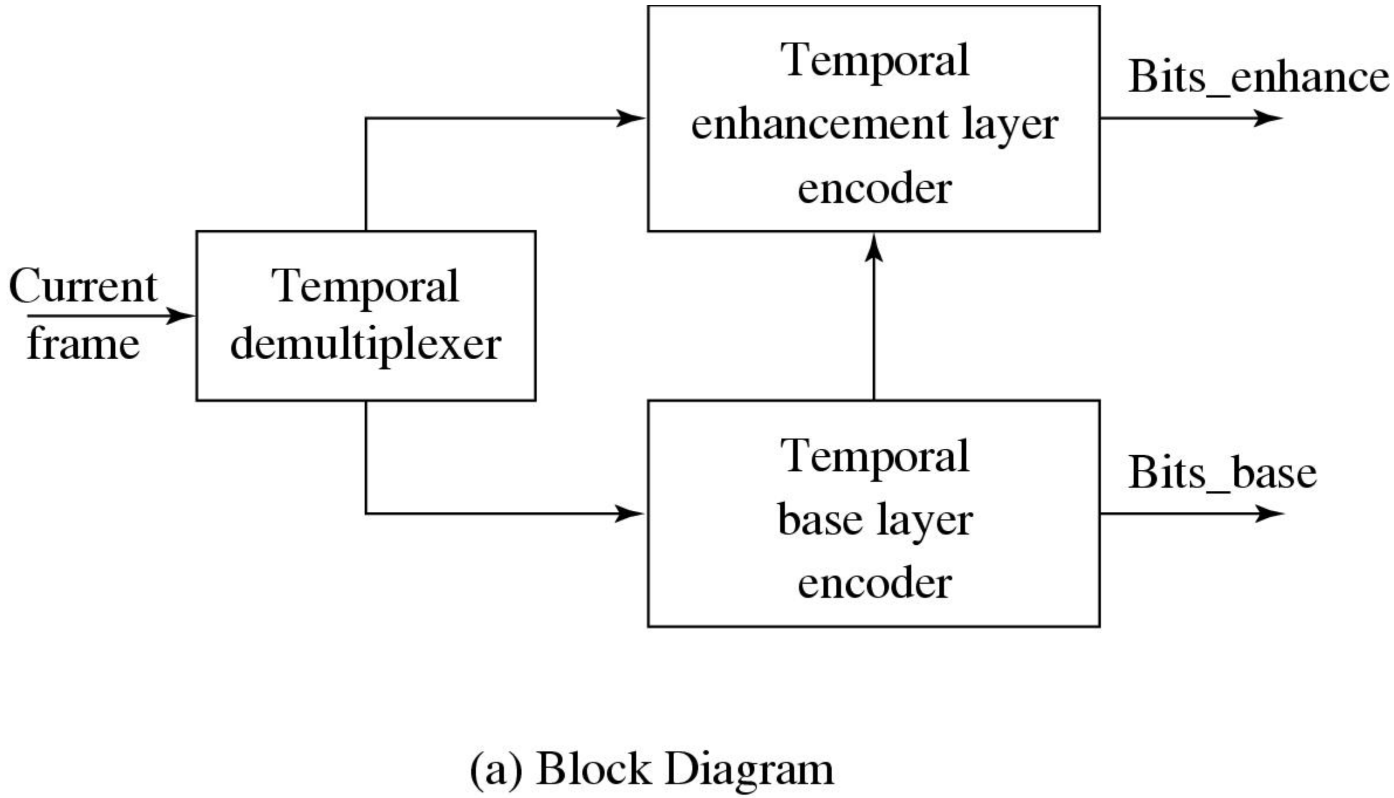


Fig 11.10: Encoder for MPEG-2 Temporal Scalability.

2.3 MPEG-2 Scalabilities

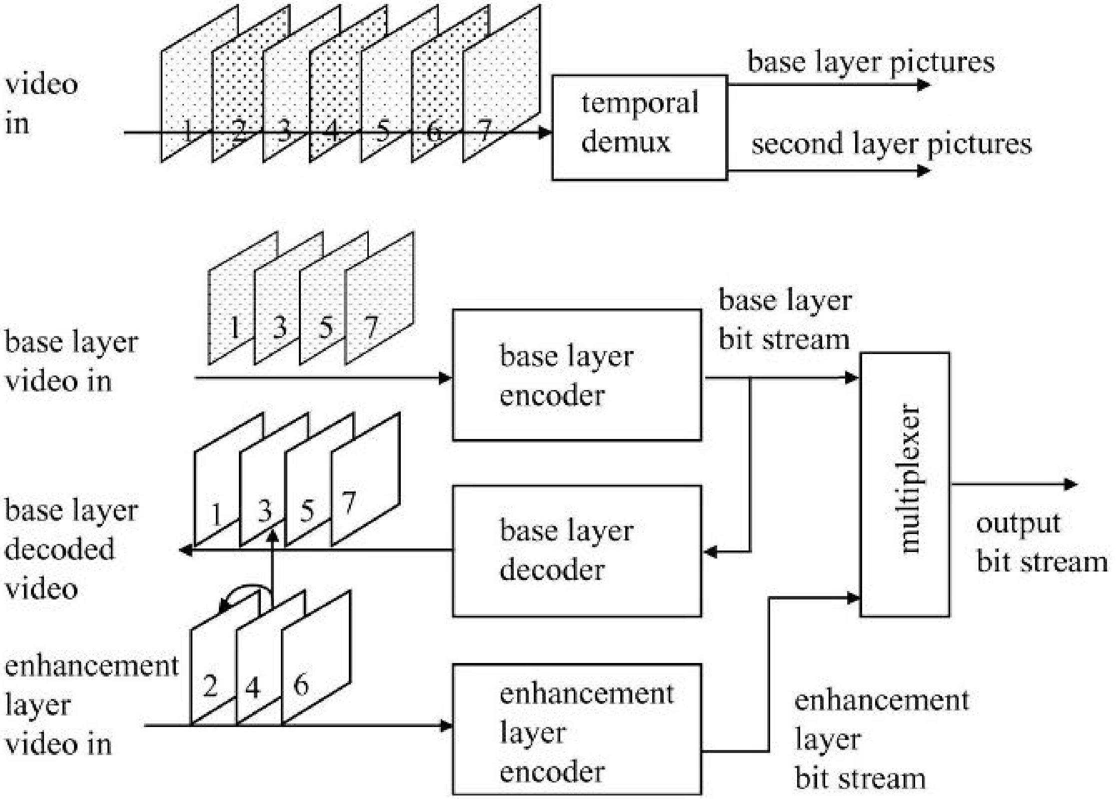
- Pictures from base layer and enhancement layers have the same spatial resolution as the input.

Temporal scalability

1、3、5、7 etc. odd frames as base layer, even frames as enhancement layer.

Base layer encoded as ordinary video, usually has I and P frame types.

Frames in enhance layer can take frames at base layer or its own layer as references



2.3 MPEG-2 Scalabilities

Hybrid Scalability

- Any two of the above three scalabilities can be combined to form hybrid scalability:

  1. Spatial and Temporal Hybrid Scalability.
  2. SNR and Spatial Hybrid Scalability.
  3. SNR and Temporal Hybrid Scalability.

- Usually, a three-layer hybrid coder will be adopted which consists of Base Layer, Enhancement Layer 1, and Enhancement Layer 2.

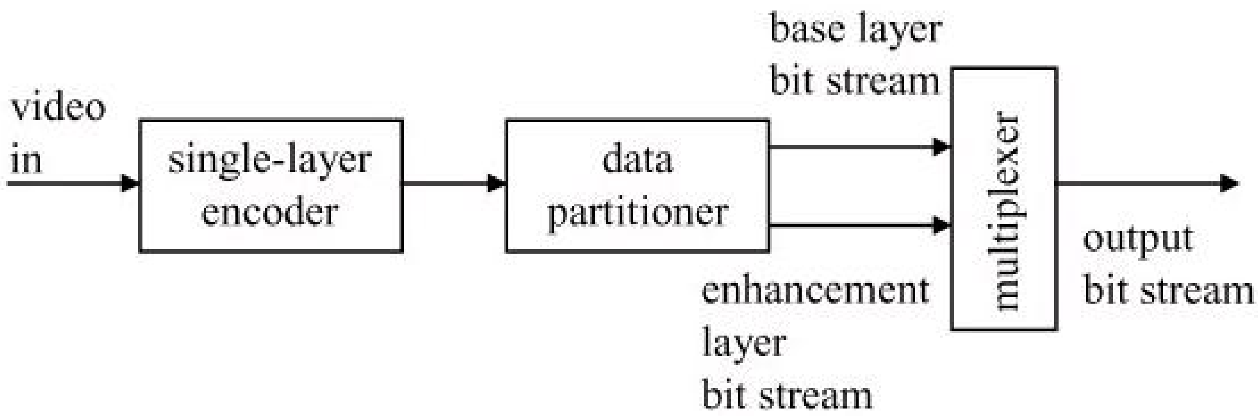
2.3 MPEG-2 Scalabilities

Data Partitioning

- The *Base partition* contains lower-frequency DCT coefficients, enhancement partition contains high-frequency DCT coefficients.
- Strictly speaking, data partitioning is not layered coding, since a single stream of video data is simply divided up and there is no further dependence on the base partition in generating the enhancement partition.
- Useful for transmission over noisy channels and for progressive transmission.

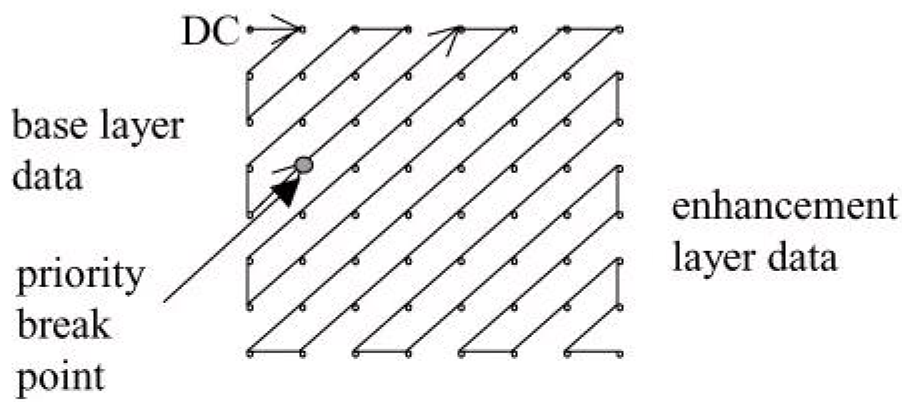
2.3 MPEG-2 Scalabilities

- it divided video into two partitions according to their importance.



Base layer: lower-frequency DCT coefficients, motion vectors, data head and other important information.

Enhancement layer: high-frequency DCT coefficients



2.3 MPEG-2 Scalabilities

- Illustration of data partitioning

Base Layer

Base Layer + Enhancement Layer

2.4 Other Major difference from MPEG-1

- **Better resilience to bit-errors**: In addition to *Program Stream*, a *Transport Stream* is added to MPEG-2 bit streams.

- **Support of 4:2:2 and 4:4:4 chroma subsampling**.

- **More restricted slice structure**: MPEG-2 slices must start and end in the same macroblock row. In other words, the left edge of a picture always starts a new slice and the longest slice in MPEG-2 can have only one row of macroblocks.

- **More flexible video formats**: It supports various picture resolutions as defined by DVD, ATV and HDTV.

2.4 Other Major difference from MPEG-1

- **Nonlinear quantization** — two types of scales are allowed:

  1. **For the first type, scale is the same as in MPEG-1 in which it is an integer in the range of [1, 31] and $scale_i = i$.**

  2. **For the second type, a nonlinear relationship exists, i.e., $scale_i \neq i$. The ith scale value can be looked up from Table 11.7.**

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$scale_i$	1	2	3	4	5	6	7	8	10	12	14	16	18	20	22	24
i	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
$scale_i$	28	32	36	40	44	48	52	56	64	72	80	88	96	104	112	

The End

Thanks！

Email: junx@cs.zju.edu.cn