Chapter 11
MPEG Video Coding I — MPEG-1 and 2

11.1 Overview

- **MPEG**: *Moving Pictures Experts Group*, established in 1988 for the development of digital video.
- It is appropriately recognized that proprietary interests need to be maintained within the family of MPEG standards:
  - Accomplished by defining only a compressed bitstream that implicitly defines the decoder.
  - The compression algorithms, and thus the encoders, are completely up to the manufacturers.

11.2 MPEG-1

- MPEG-1 adopts the CCIR601 digital TV format also known as SIF (*Source Input Format*).
- MPEG-1 supports only non-interlaced video. Normally, its picture resolution is:
  - 352×240 for NTSC video at 30 fps
  - 352×288 for PAL video at 25 fps
  - It uses 4:2:0 chroma subsampling
- The MPEG-1 standard is also referred to as ISO/IEC 11172. It has five parts: 11172-1 Systems, 11172-2 Video, 11172-3 Audio, 11172-4 Conformance, and 11172-5 Software.

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**.

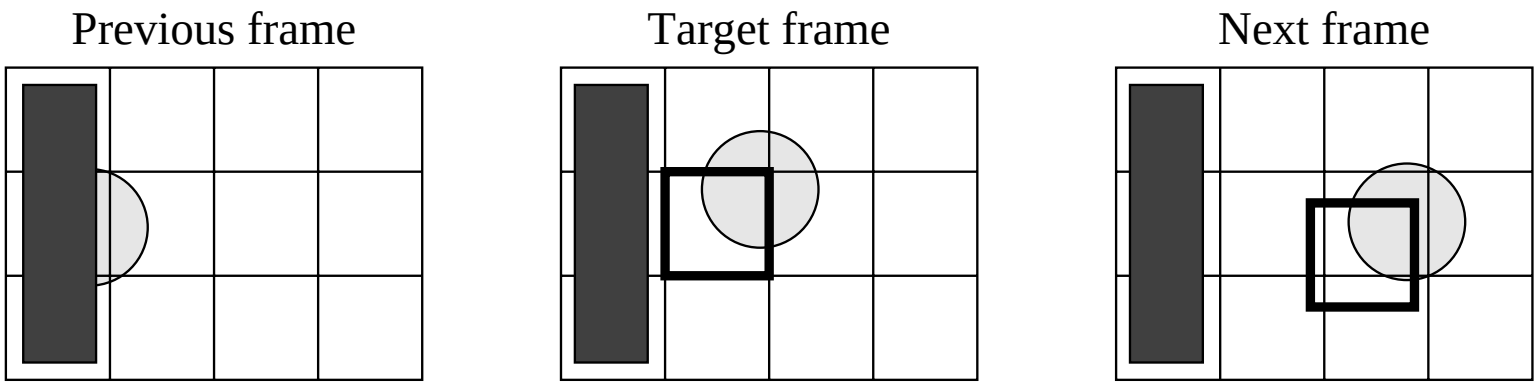

Fig 11.1: The Need for Bidirectional Search.

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.

Motion Compensation in MPEG-1 (Cont'd)

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

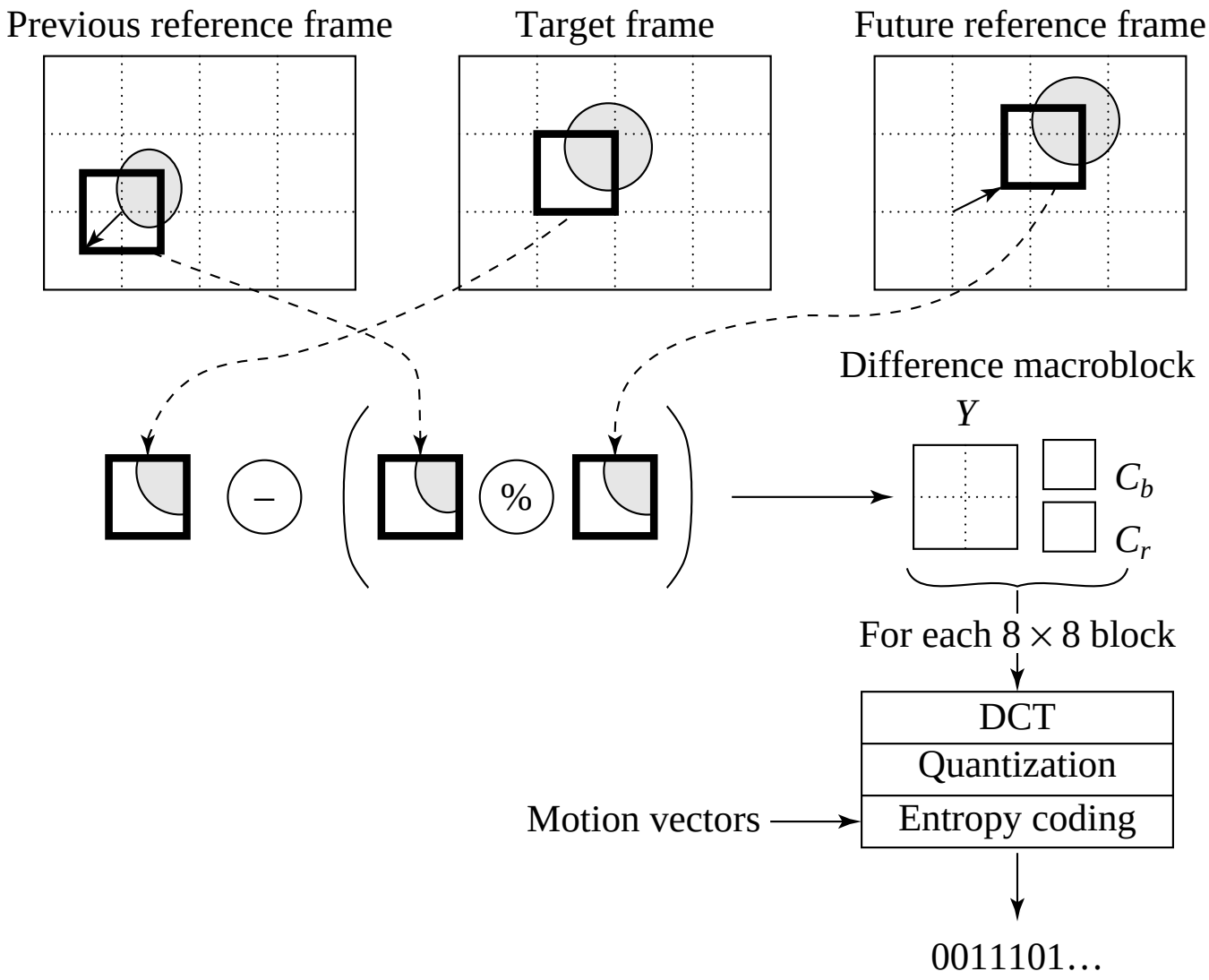

Fig 11.2: B-frame Coding Based on Bidirectional Motion Compensation.

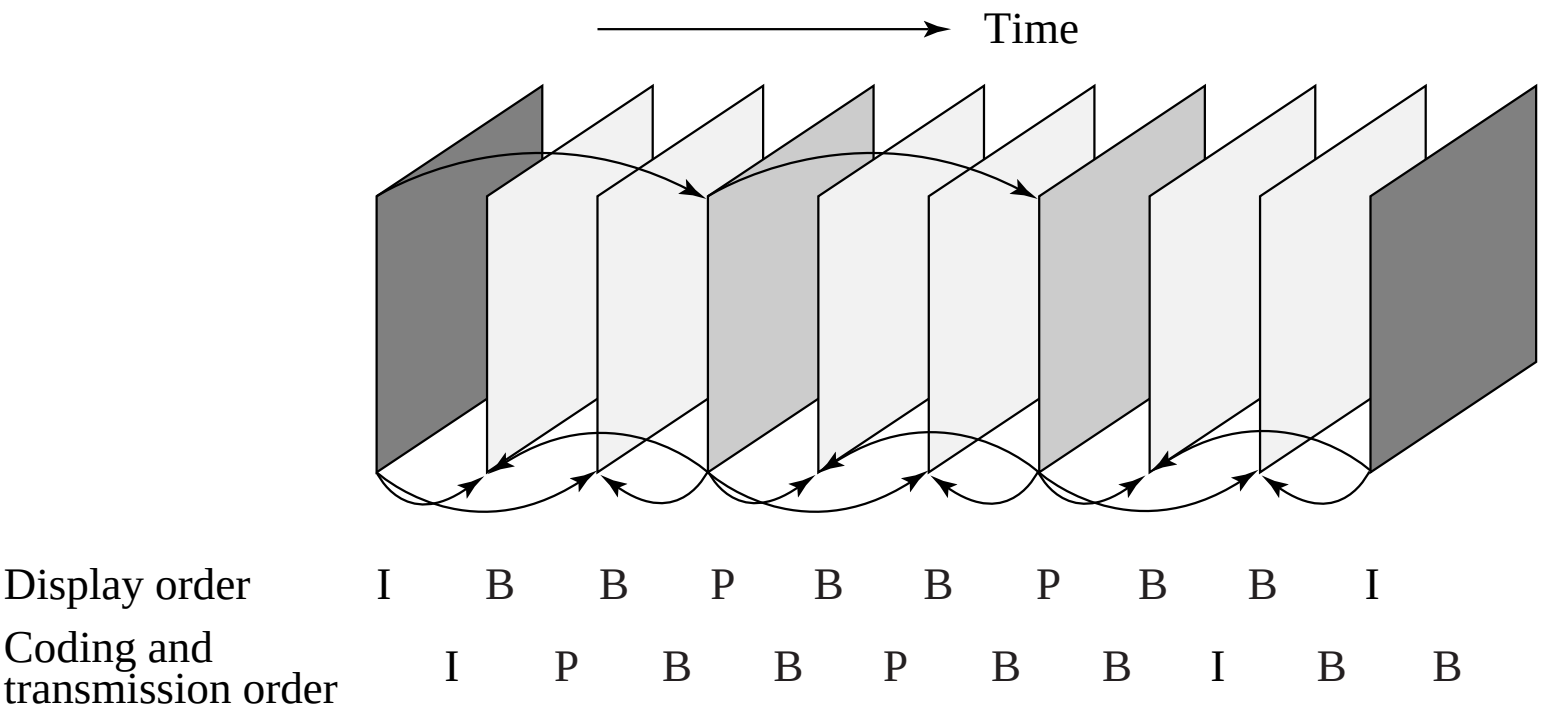

Fig 11.3: MPEG Frame Sequence.

Other Major Differences from H.261

- Source formats supported:
  - H.261 only supports CIF (352×288) and QCIF (176×144) source formats, MPEG-1 supports SIF (352×240 for NTSC, 352×288 for PAL).
  - MPEG-1 also allows specification of other formats as long as the Constrained Parameter Set (CPS) as shown in Table 11.1 is satisfied:

Table 11.1: The MPEG-1 Constrained Parameter Set

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

11.3 MPEG-2

- **MPEG-2**: For higher quality video at a bit-rate of more than 4 Mbps.
- Defined seven **profiles** aimed at different applications:
  - **Simple, Main, SNR scalable, Spatially scalable, High, 4:2:2, Multiview**.
  - Within each profile, up to four *levels* are defined (Table 11.5).
  - The DVD video specification allows only four display resolutions: 720×480, 704×480, 352×480, and 352×240 — a restricted form of the MPEG-2 Main profile at the Main and Low levels.

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*.

Five Modes of Predictions

- MPEG-2 defines **Frame Prediction** and **Field Prediction** as well as five prediction modes:

  1. **Frame Prediction for Frame-pictures:** Identical to MPEG-1 MC-based prediction methods in both P-frames and B-frames.

  2. **Field Prediction for Field-pictures:** A macroblock size of 16×16 from Field-pictures is used. For details, see Fig. 11.6(b).

3. **Field Prediction for Frame-pictures:** The top-field and bottom-field of a Frame-picture are treated separately. Each 16×16 macroblock (MB) from the target Frame-picture is split into two 16×8 parts, each coming from one field. Field prediction is carried out for these 16×8 parts in a manner similar to that shown in Fig. 11.6(b).

4. 16×8 **MC for Field-pictures:** Each 16×16 macroblock (MB) from the target Field-picture is split into top and bottom 16×8 halves. Field prediction is performed on each half. This generates two motion vectors for each 16×16 MB in the P-Field-picture, and up to four motion vectors for each MB in the B-Field-picture.

   This mode is good for a finer MC when motion is rapid and irregular.

5. **Dual-Prime for P-pictures:** First, Field prediction from each previous field with the same parity (top or bottom) is made. Each motion vector $\mathbf{mv}$ is then used to derive a calculated motion vector $\mathbf{cv}$ in the field with the opposite parity taking into account the temporal scaling and vertical shift between lines in the top and bottom fields. For each MB the pair $\mathbf{mv}$ and $\mathbf{cv}$ yields two preliminary predictions. Their prediction errors are averaged and used as the final prediction error.

   This mode mimics B-picture prediction for P-pictures without adopting backward prediction (and hence with less encoding delay).

   This is the only mode that can be used for either Frame-pictures or Field-pictures.

Other Major Differences 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.

Other Major Differences from MPEG-1 (Cont'd)

- **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.

Table 11.7: Possible Nonlinear Scale in MPEG-2

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	

11.4 Further Exploration

- **Text books:**
  - *Video Compression Standard* by J.L. Mitchell et al
  - *Digital Video: An Introduction to MPEG-2* by B.G. Haskell et al

- **Web sites:** $\longrightarrow$ Link to Further Exploration for Chapter 11.. including:
  - The MPEG home page.
  - MPEG FAQ page.
  - Overviews and working documents of the MPEG-1 and MPEG-2 standards.