11.1 MPEG 1 and 2

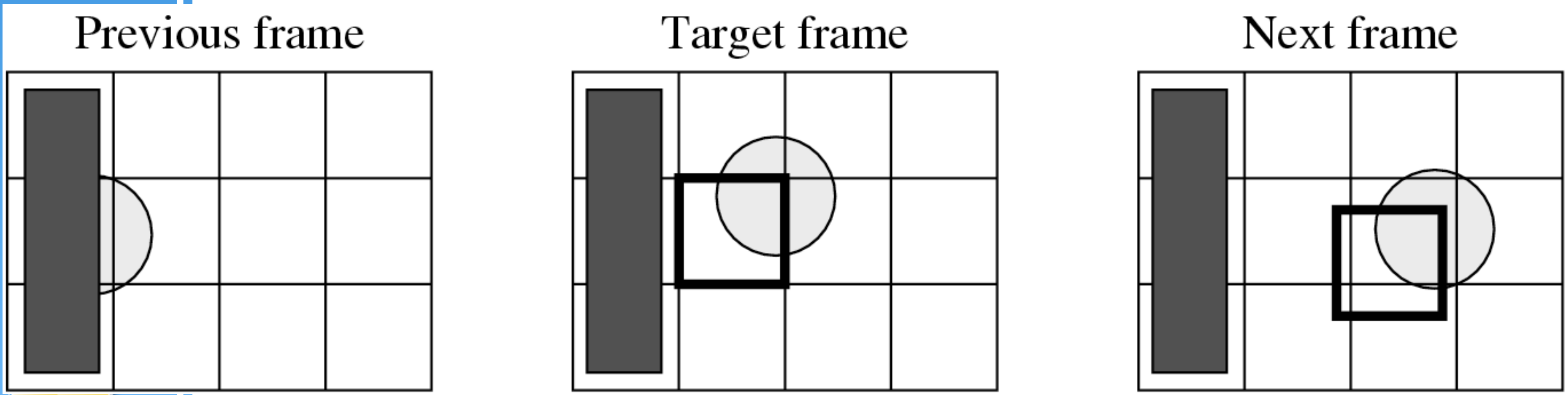
- **MPEG**: *Moving Pictures Experts Group* for the development of digital video
- It is appropriately recognized that proprietary interests need to be maintained within the family of MPEG standards:
  - Accomplished by defining only a compressed bitstream that implicitly defines the decoder. The compression algorithms, and thus the encoders, are completely up to the manufacturers
- From Wikipedia: approximately 640 patents worldwide make up the "essential" patents surrounding MPEG-2. These are held by over 20 corporations and one university. Where software patentability is upheld, the use of MPEG-2 requires the payment of licensing fees to the patent holders via the MPEG Licensing Association. The development of the standard itself took less time than the patent negotiations

11.2 MPEG-1 – used in VCD

- 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 sub-sampling
- **MPEG-1 Audio Layer 3 is mp3**

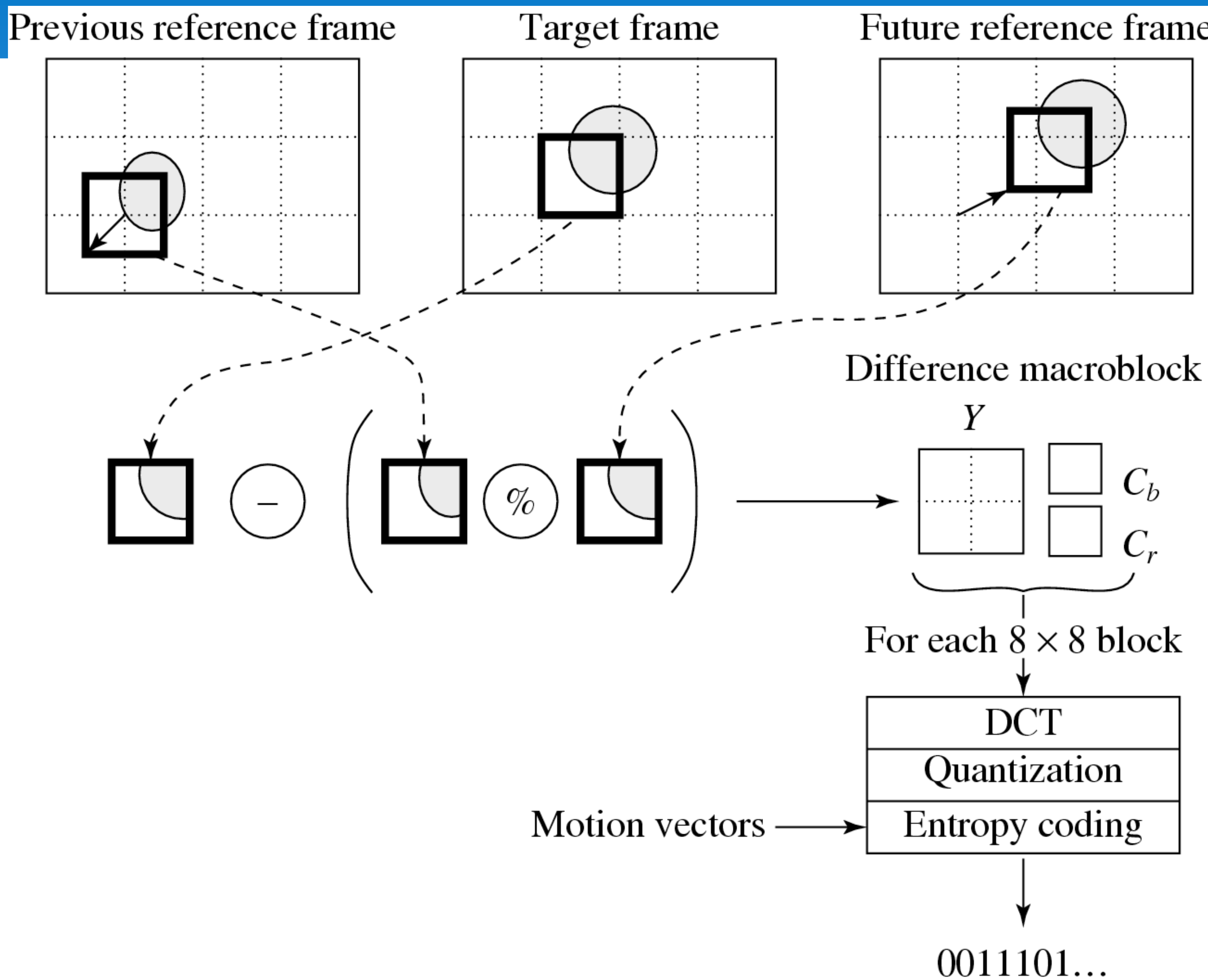
Bidirectional motion compensation

- 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

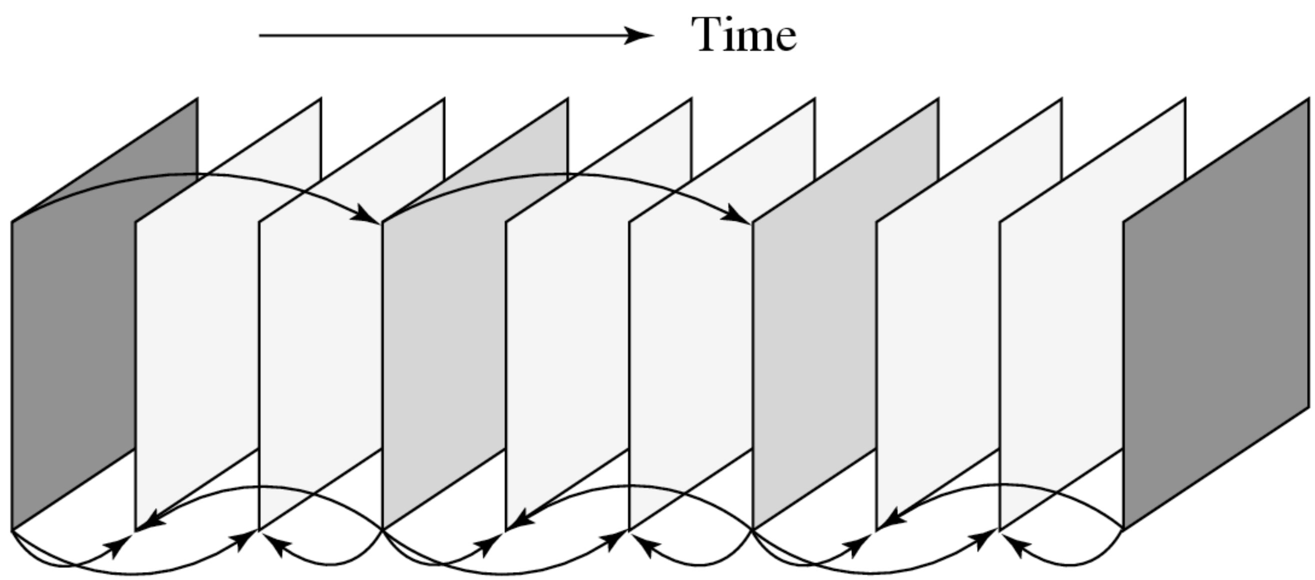


Motion Compensation in MPEG-1 (Cont'd)

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



Previous reference frame
Target frame
Future reference frame
Difference macroblock
Y
C_b
C_r
For each 8 × 8 block
DCT
Quantization
Entropy coding
Motion vectors
0011101…



Display order	I	B	B	P	B	B	P	B	B	I
Coding and transmission order	I	P	B	B	P	B	B	I	B	B

Major Differences from H.261

- Quantization:
  - MPEG-1 quantization uses different quantization tables for its Intra and Inter coding
- For DCT coefficients in Intra mode:

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

- For DCT coefficients in Inter mode:

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

Typical Sizes of MPEG-1 Frames

- 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

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

Problems with B frames (from BMRC Faq)

- A. Computational complexity, bandwidth, end-to-end delay, and picture buffer size
  - Computational complexity in the decoder is increased since some macroblock modes require averaging between two block predictions (macroblock_motion_forward==1 && macroblock_motion_backward==1). Worst case, memory bandwidth is increased an extra 15.2 MByte
  - an extra picture buffer is needed to store the future reference picture (backwards prediction frame)
  - an extra picture delay is introduced in the decoder since the frame used for backwards prediction needs to be transmitted to the decoder and reconstructed before the intermediate B-pictures in display order can be decoded

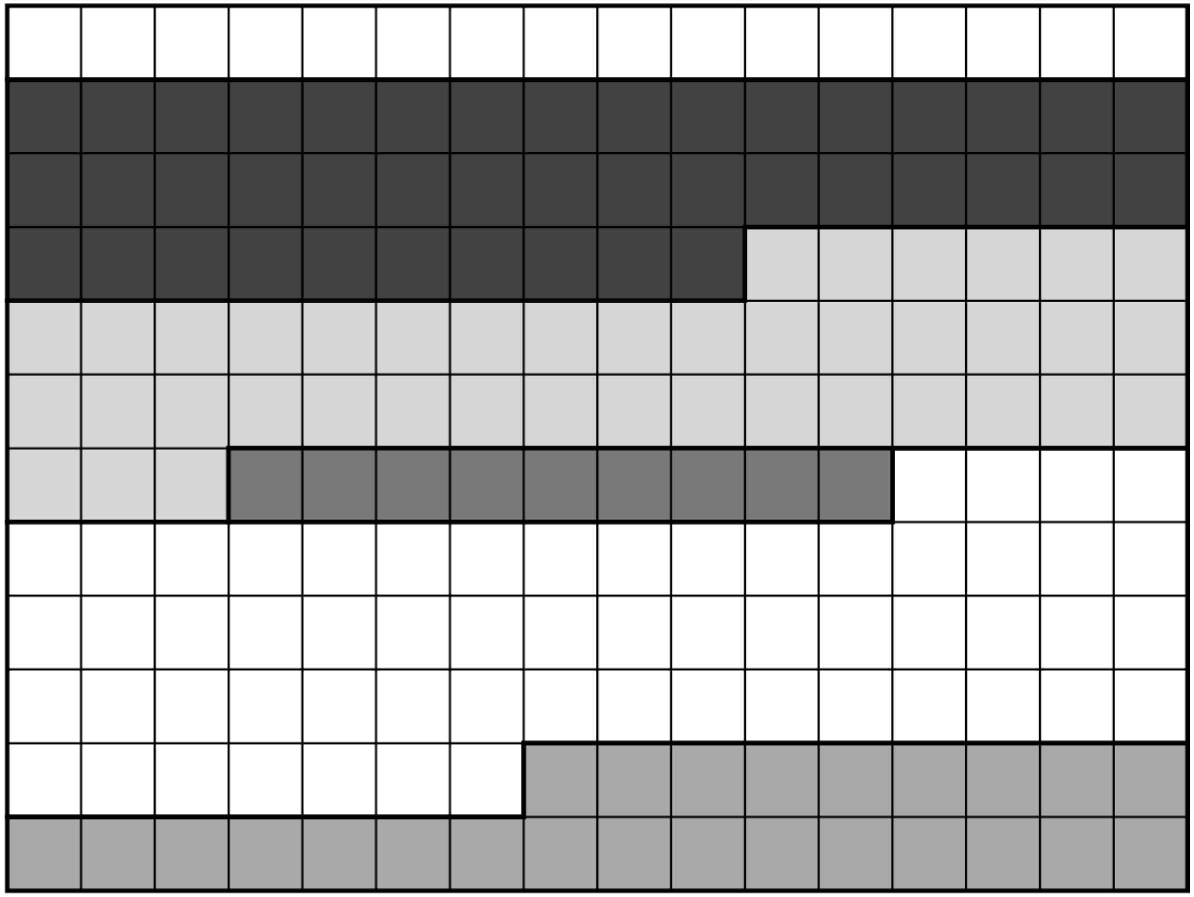
Other Major Differences from H.261

- 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) is satisfied

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

Other Major Differences from H.261 (Cont'd)

- Instead of GOBs as in H.261, an MPEG-1 picture can be divided into one or more **slices**
  - 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



- Fig 11.4: Slices in an MPEG-1 Picture.

Other Major Differences from H.261 (Cont'd)

- 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 [−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.

11.3 MPEG-2

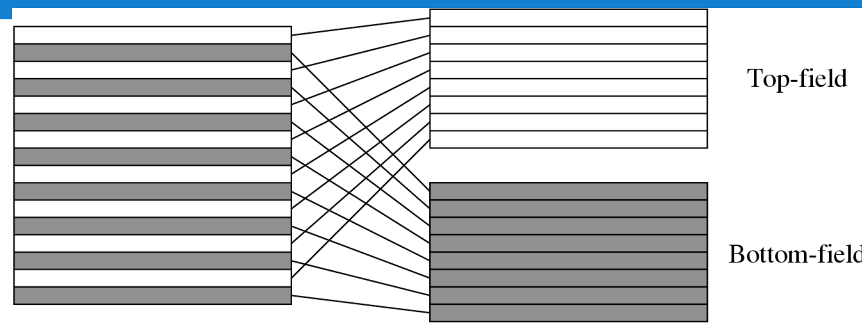
- MPEG-2: For higher quality video at a bit-rate of more than 4 Mbps
- Defined seven profiles aimed at different applications:
  - Simple, Main, SNR scalable, Spatially scalable, High, 4:2:2, Multiview
  - Within each profile, up to four levels are defined
  - The DVD video specification allows only 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
    - Video peak 9.8 Mbit/s
    - Total peak 10.08 Mbit/s
    - Minimum 300 kbit/s

Level	Simple profile	Main profile	SNR Scalable profile	Spatially Scalable profile	High Profile	4:2:2 Profile	Multiview Profile
High		*			*		
High 1440		*		*	*		
Main	*	*	*		*	*	*
Low		*	*				

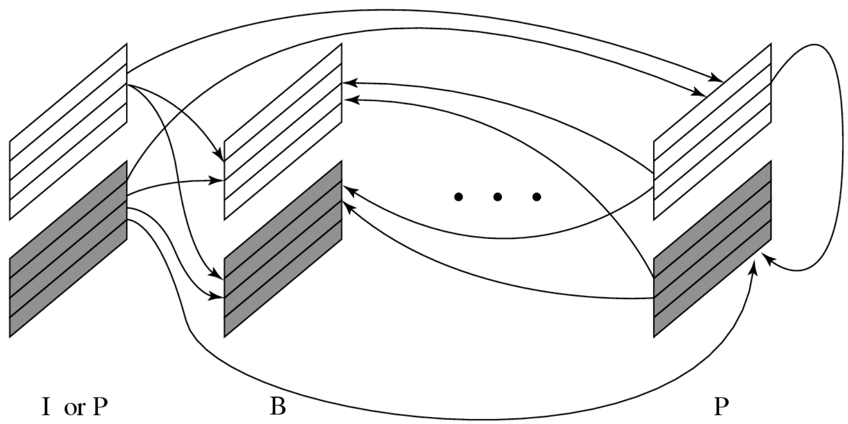
Level	Max. Resolution	Max fps	Max pixels/sec	Max coded Data Rate (Mbps)	Application
High	1,920 × 1,152	60	62.7×10^6	80	film production
High 1440	1,440 × 1,152	60	47.0×10^6	60	consumer HDTV
Main	720 × 576	30	10.4×10^6	15	studio TV
Low	352 × 288	30	3.0×10^6	4	consumer tape equiv.

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

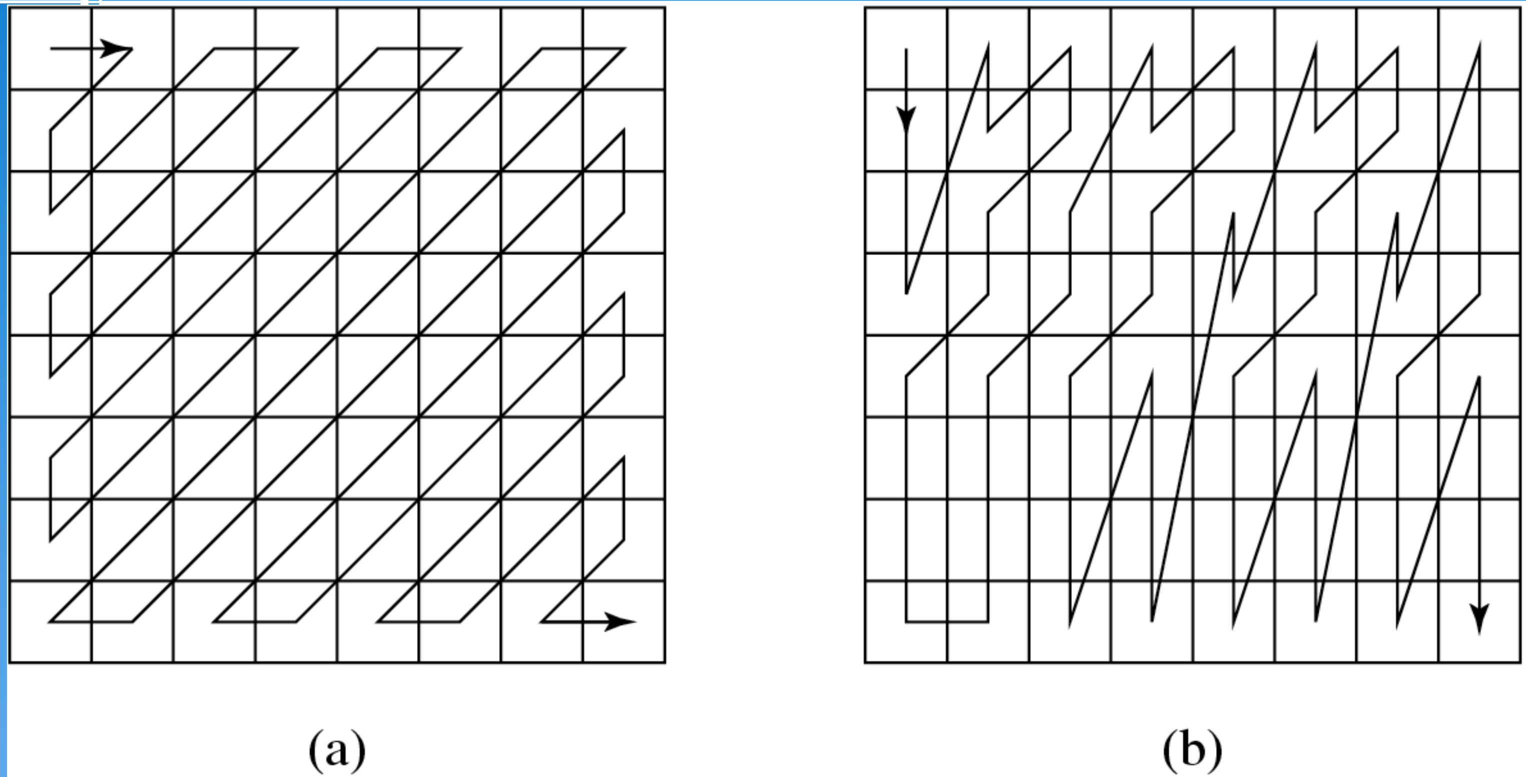


(a)



(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



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

MPEG-2 layered coding

- The MPEG-2 scalable coding: A base layer and one or more enhancement layers can be defined
  - The base layer can be independently encoded, transmitted and decoded 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

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

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 macro block 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 macro blocks
- 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:
  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

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	