CMPN206: Multimedia

Lecture 10: Video Compression II

Mohamed Alaa El-Dien Aly
Computer Engineering Department
Cairo University
Spring 2014

Agenda

- MPEG-1
- MPEG-2
- MPEG-4
- H.264

Acknowledgments: Most slides are adapted from Richard Ladner, from Li and Drew, from Khaled Sayood, and from David Marshall and Kirill Sidorov.

MPEG

- Moving Pictures Experts Group, established in 1988 to create standards for digital audio and video
- It defines a *compressed bitstream* that implicitly define the decoder
- The *compression method*, and thus the encoder, are completely up to the manufacturer
- Each standard contains several *parts* e.g. systems, video, audio, software, …
- Usually:
  - Part2: video standard
  - Part3: audio standard

MPEG-1

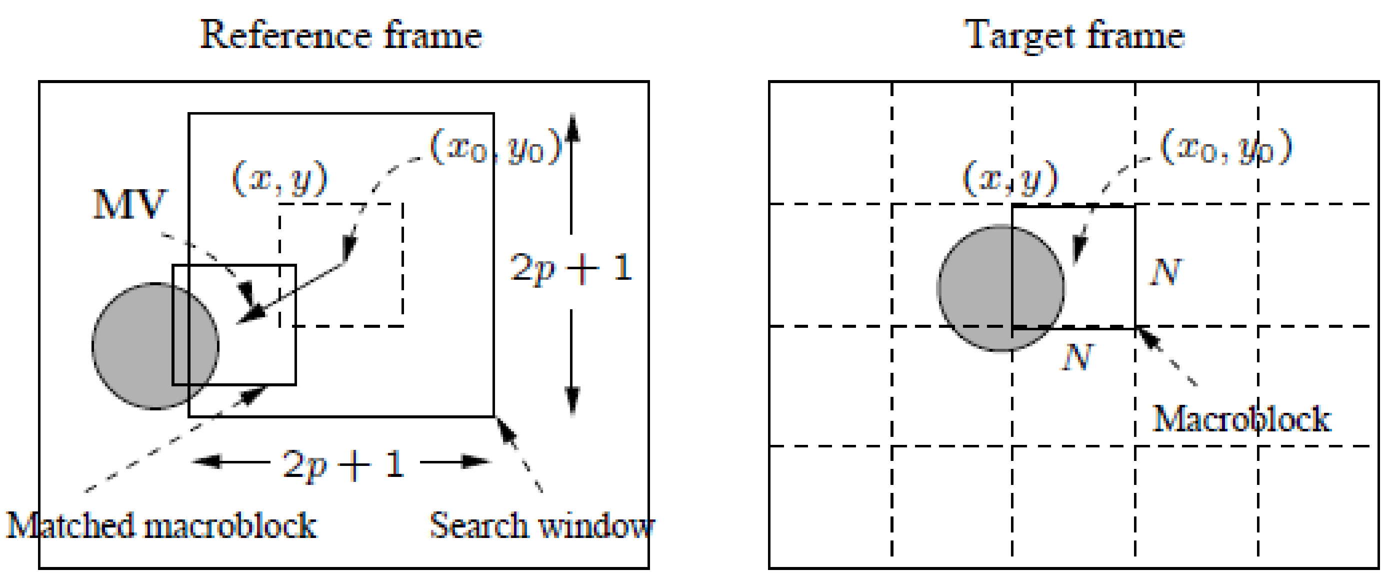
- Approved in 1991 by ISO MPEG group
- Designed for digital video for storage media (CDs and VCDs) at bitrates up to 1.5 Mbps
- Adopts the CCIR601 digital TV format (also known as Source Input Format SIF)
- Supports only non-interlaced (*progressive*) video at resolutions:
  - 352×240 for NTSC video at 30 fps
  - 352×288 for PAL video at 25 fps
  - Uses *only* 4:2:0 chroma subsampling

MPEG-1

- MPEG-1 standard ISO/IEC 11172 contains *five* parts:
  - *systems:* dividing output into packets, multiplexing of audio and video, …
  - *video:* encoding/decoding of video signals
  - *audio:* encoding/decoding of audio signals
  - *conformance:* or compliance, specifies tests to ensure a bitstream conforms to the standard
  - *software*: implementation of a standard decoder, and sample encoder

Recall: H.261 Motion Compensation

- Motion Estimation: each macroblock MB of the target frame is assigned a best matching MB from the *previously decoded* I- or P-frame
- Prediction Error: the difference between the target MB and its best matching MB, and is sent to DCT, quantization, and entropy coding
- The prediction is from a *previous* frame → *forward prediction*



MPEG-1 Motion Compensation

- Sometimes *backward prediction*, i.e. taking the reference frame as the *next* frame, is required
- For example, when the ball is coming out from behind the bar, it better matches a block in the next frame *not* the previous frame
- This is called *bidirectional* search

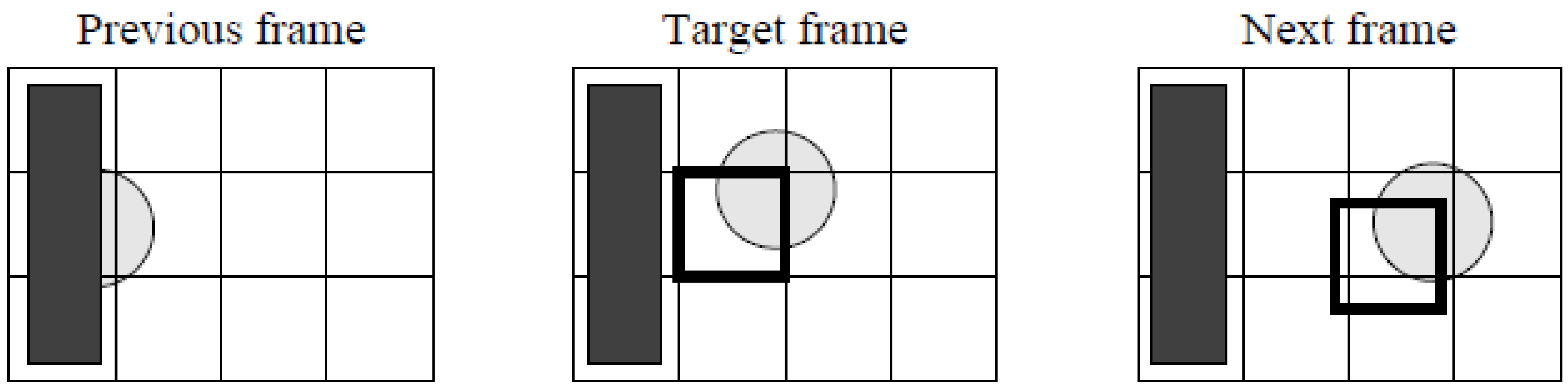


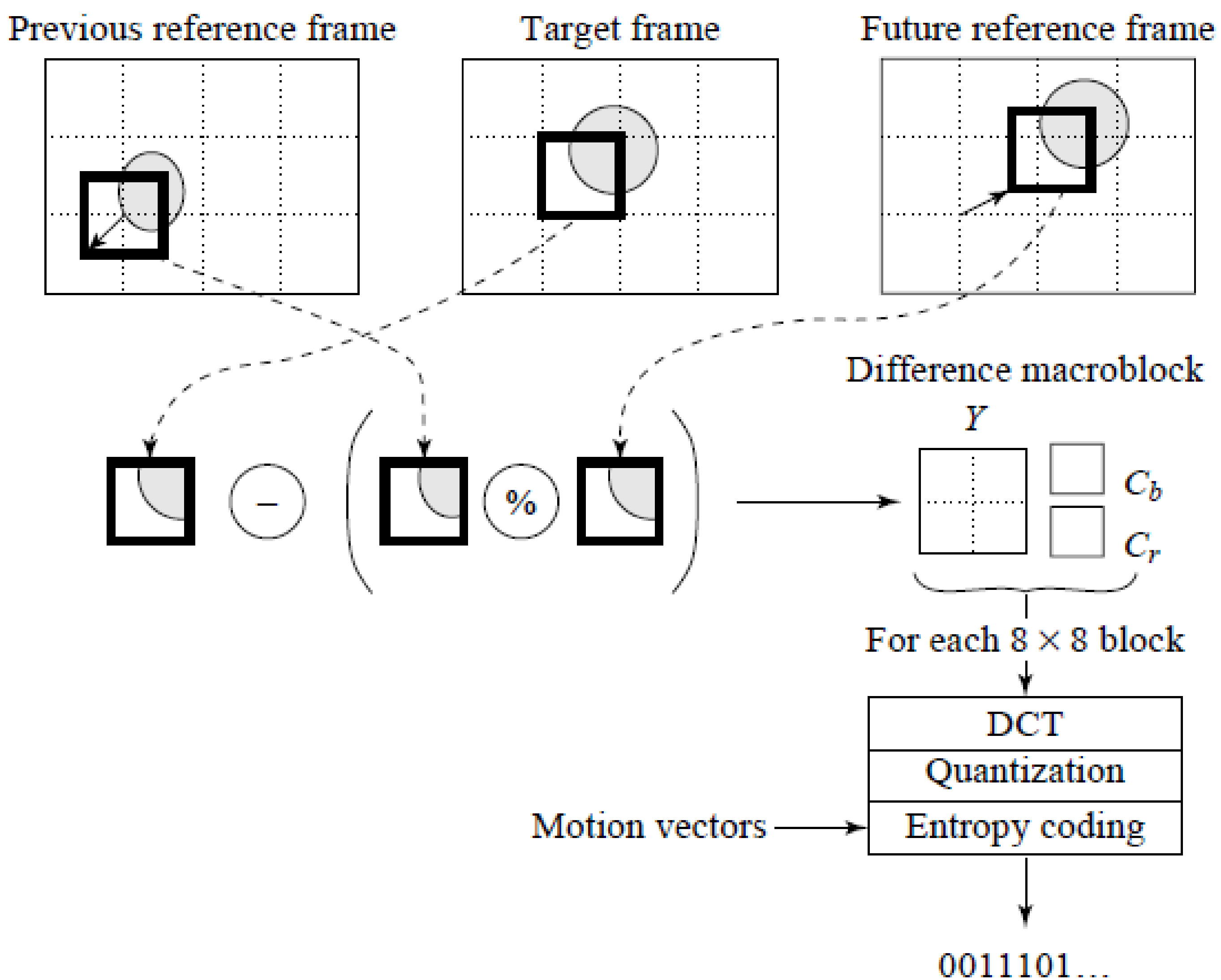
Fig 11.1: The Need for Bidirectional Search.

MPEG-1 Motion Compensation

- MPEG-1 introduces a third frame type: B-frames (bi-directional frames) which used *bi-directional* motion compensation
- The coding of the B-frame uses both *previous* and *next* frames for the motion compensation

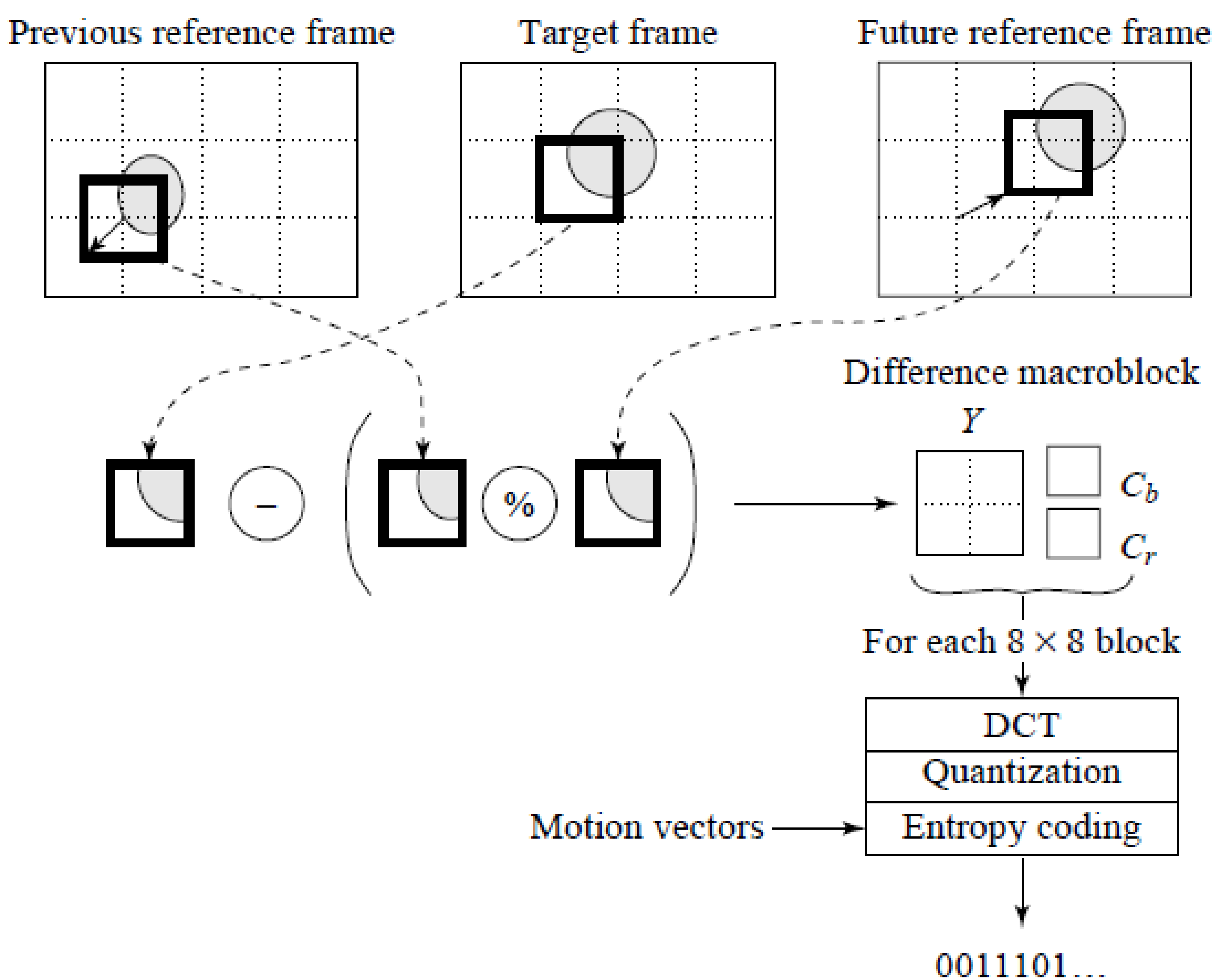
B-frame Coding

- Each MB in the B-frame will have up to *two* MV (one from the forward and one from the backward prediction)



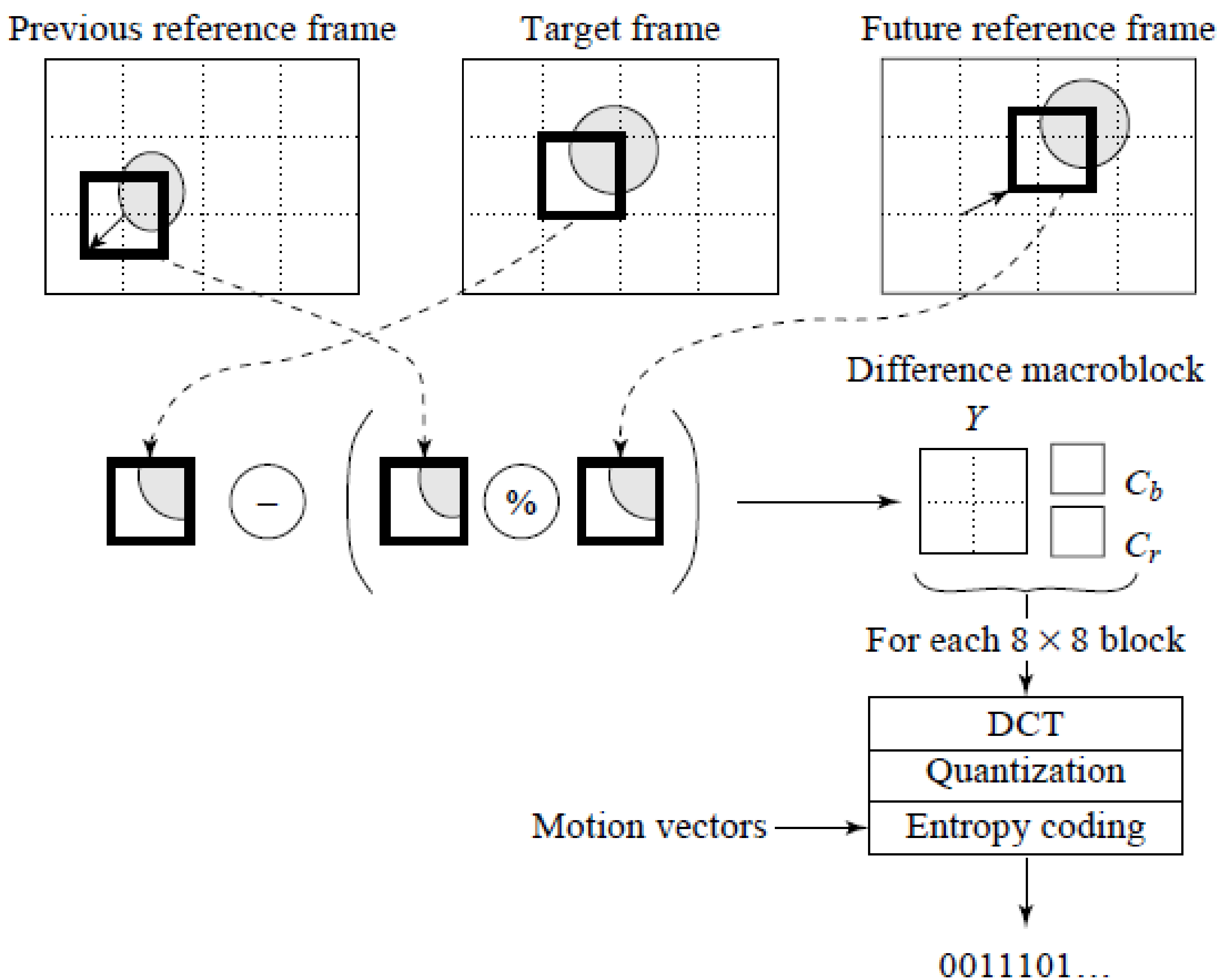
B-frame Coding

- If matching in *both* directions is successful, then two MVs are sent and the corresponding matching MBs are *averaged* before computing the *difference*



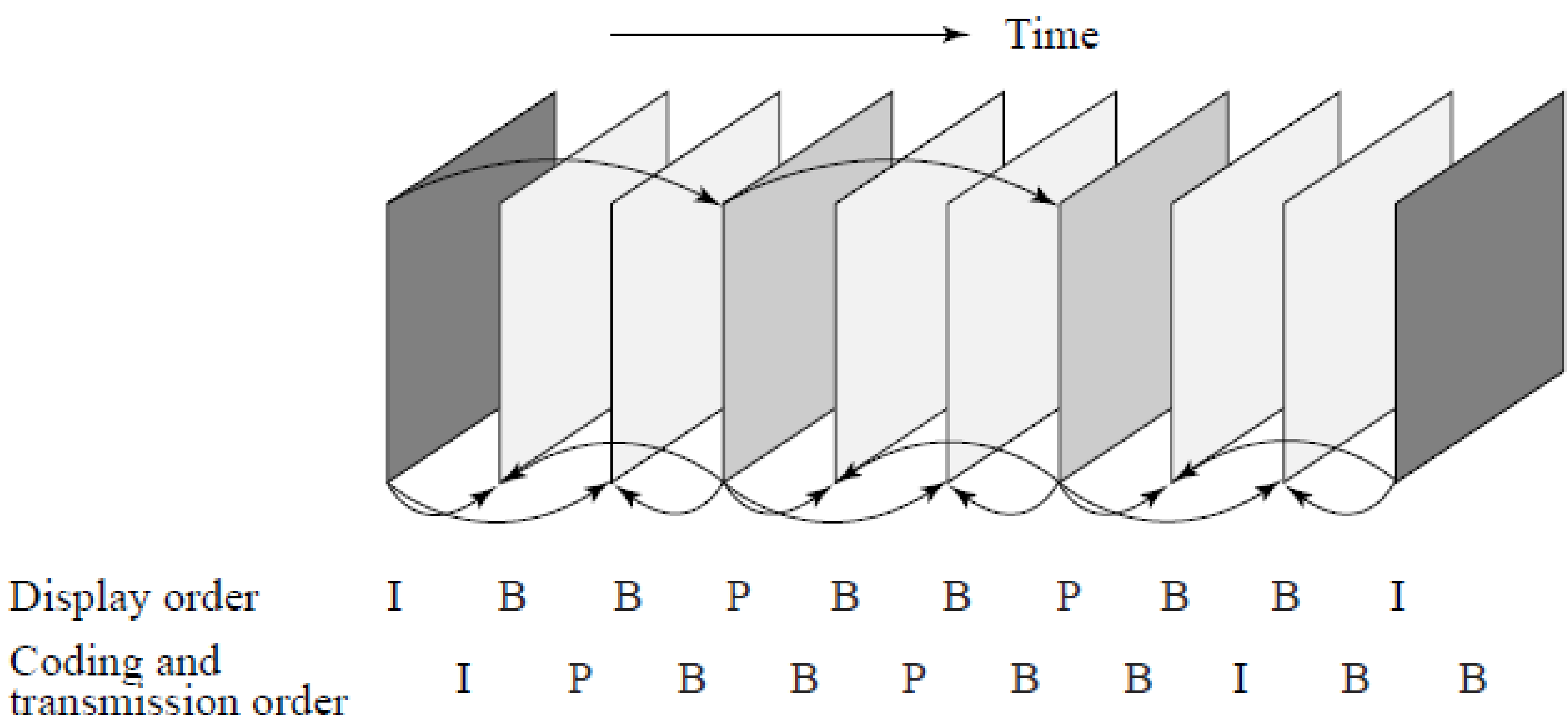
B-frame Coding

- If matching is successful in only one direction, then that MV is sent and the corresponding MB is used for computing the *prediction error*



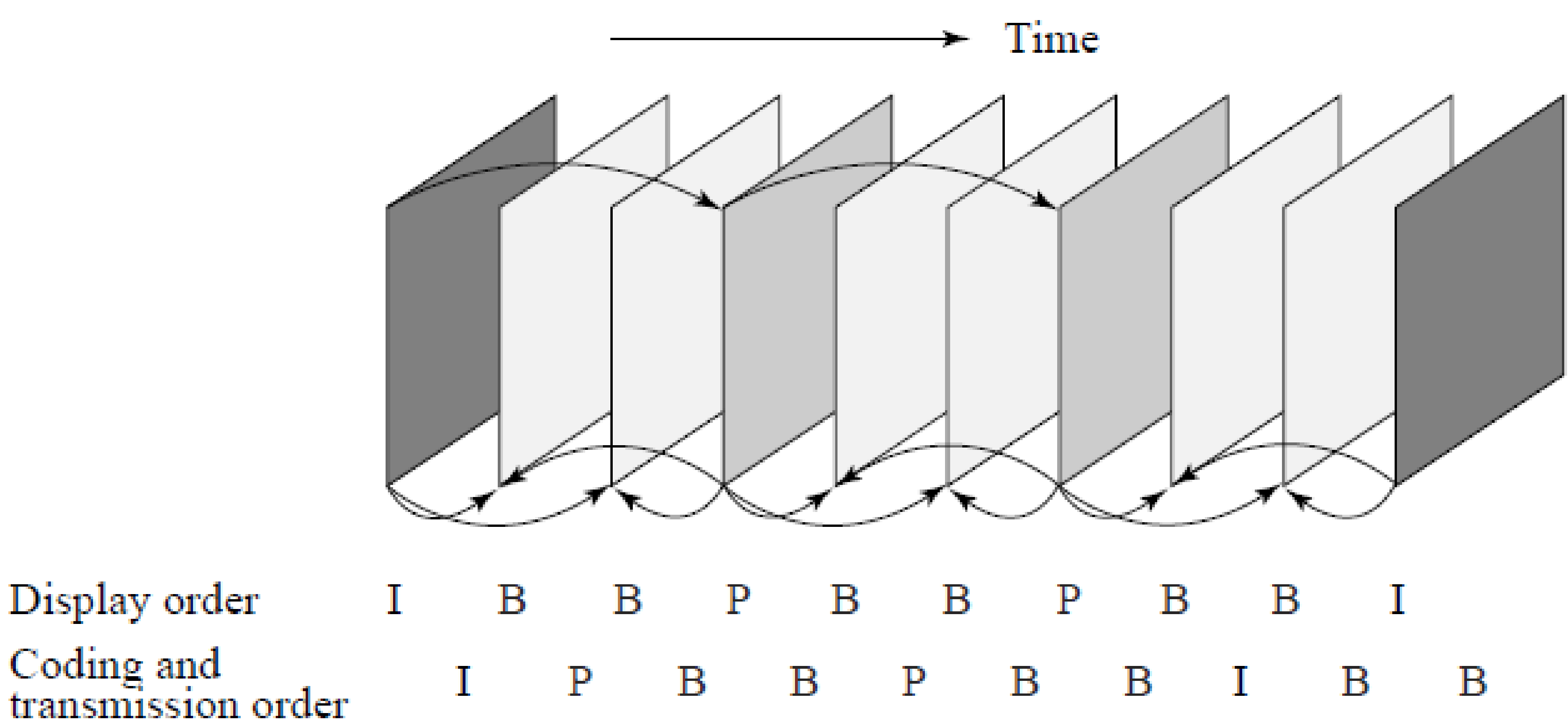
MPEG-1 Frame Sequence

- B-frames are predicted from previous and next I- or P-frames
- P-frames are predicted from previous I- or P-frames
- I- and P-frames have to be sent and decoded *before* being able to decode their dependent B-frames, that's why the *display* order is different from the *coding* order



MPEG-1 Frame Sequence

- The MPEG standard defines:
  - *M*: the interval between I- or P-frames
  - *N*: the interval between I-frames
- Below, $M = 3$ and $N = 9$



Differences from H.261

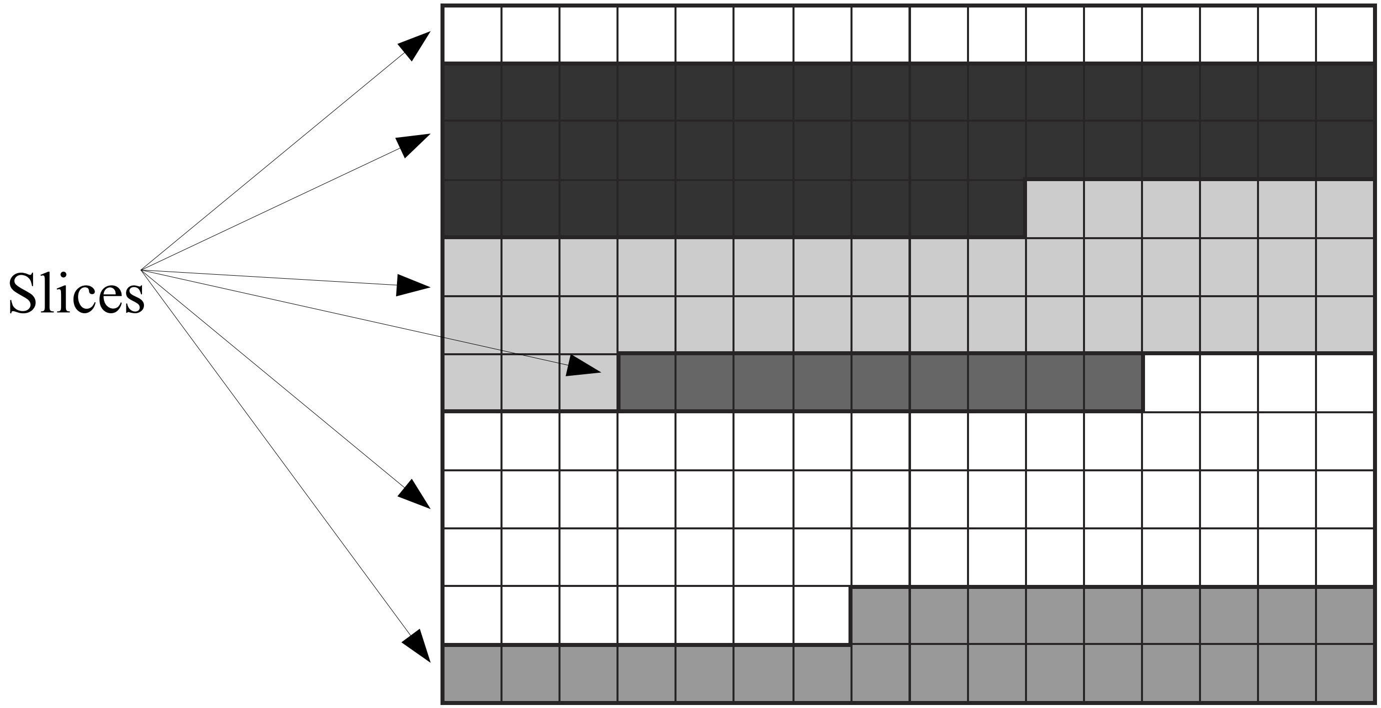
- Bi-directional Motion Compensation
- Source Formats:
  - H.261 only supports CIF (352×288) and QCIF (176×144)
  - MPEG-1 supports SIF (352×240 for NTSC and 352×288 for PAL)
- MPEG-1 also allows specification of other formats subject to these constraints:

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

Differences from H.261

- Instead of GOBs in H.261, MPEG-1 pictures can be divided into one or more *slices* that:
  - can contain variable number of MBs in a single picture (not fixed in size like GOBs)
  - can start and end anywhere in the picture
  - are coded *independently* (with different quantization scales)
  - important for *error recovery* (like GOBs in H.261 and H.263)



Differences from H.261

- MPEG-1 *quantization* uses different quantization tables for its intra- and inter-frame coding
  - DCT coefficients in intra-frame mode:

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

  - DCT coefficients in inter-frame mode:

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

  where *scale* is an integer in the range [1, 31]

Differences from H.261

- MPEG-1 *quantization* uses different quantization tables for its intra- and inter-frame coding

Table 11.2: Default Quantization Table (Q_1) for Intra-Coding

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

Table 11.3: Default Quantization Table (Q_2) for Inter-Coding

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

Differences from H.261

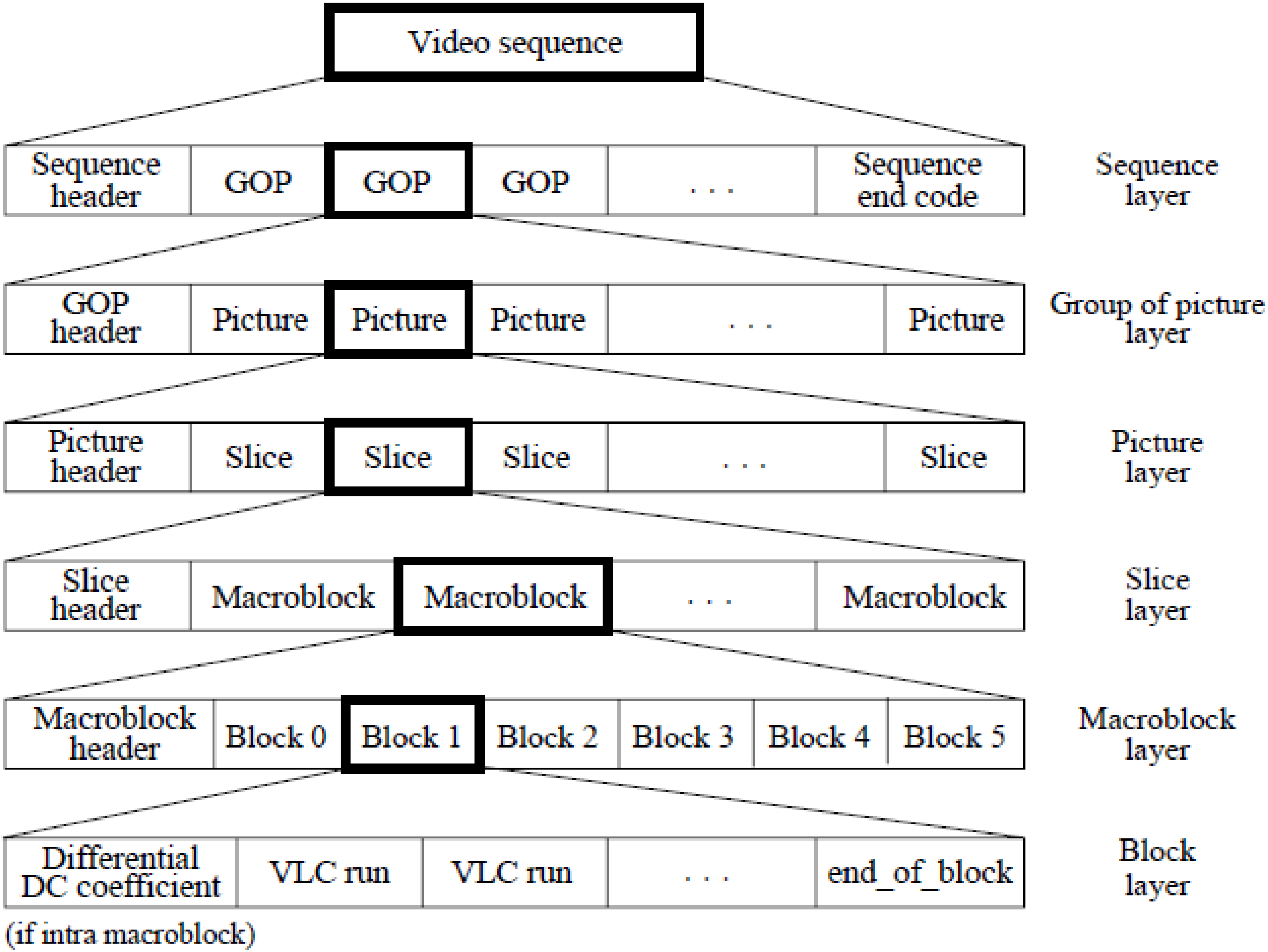
- MPEG-1 allows *half-pixel* precision motion compensation
- MPEG-1 allows a range of [–512, 511.5] for the motion vectors in half-pixel precision and [–1024, 1023] for full-pixel precision, unlike the range of [–15, 15] in H.261.
  - This is needed because of the introduction of B-frames which have to depend on previous I- or P-frames, and can't be used as *reference* frames
  - This increases the difference in time between the B-frame and the reference I- or P-frames, and thus a larger search window is required
- Allows *random access* in the bitstream using the Group of Pictures GOP layer that contains a number of frames, is time-coded, and always starts with an I-frame

Typical MPEG-1 Frame Sizes

- P-frames are significantly smaller than I-frames because of exploiting the *temporal redundancy*
- B-frames are even smaller than P-frames because of the bi-directional motion compensation

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

MPEG-1 Bitstream



MPEG-1 Bitstream

- Sequence Layer: contains one or more Groups of Pictures (GOPs) and starts with a header containing information about the video
- GOP Layer: contains one or more pictures with at least one I-frame, and contains a time_code that specifies the time hour:minute:second:frame from the video start
- Picture Layer: contains an I-, P-, or B-frame
- Slice Layer: contains one or more MBs
- Macroblock Layer: contains 4 *Y* blocks and 1 *Cb* and 1 *Cr*
- Block Layer: contains quantized entropy-coded DCT coefficients for the block

MPEG-2

- Also known as H.262
- Started in 1990, approved by ISO in 1994
- Aims at videos with higher quality and bitrates of over 4 Mbps
- Initially developed as a standard for digit TV broadcast
- Used for digital TV and satellite broadcasts and Digital Video Disks (DVDs)

MPEG-2 Profiles

- Defines *seven profiles* (video *quality*) and different *levels* (video *resolutions*) *per profile,* suitable for different applications
  - Low latency applications like video conferencing
  - Non-interactive applications like DVDs
  - …

Table 11.5: Profiles and Levels in MPEG-2

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

Main Profile

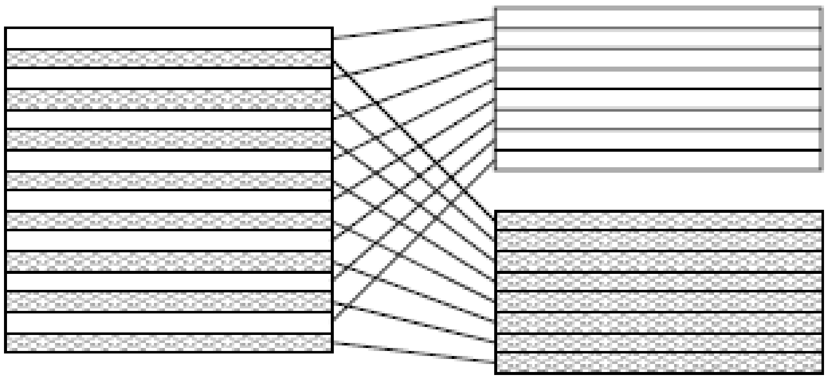
- The four *levels* in the *main* profile:
  - Low: SIF video, backward compatible with MPEG-1
  - Main: CCIR601 video
  - High 1440: European HDTV
  - High: North American HDTV

Table 11.6: Four Levels in the Main Profile of MPEG-2

Level	Max Resolution	Max fps	Max Pixels/sec	Max coded Data Rate (Mbps)	Application
High	$1,920 \times 1,152$	60	62.7×10^6	80	film production
High 1440	$1,440 \times 1,152$	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.

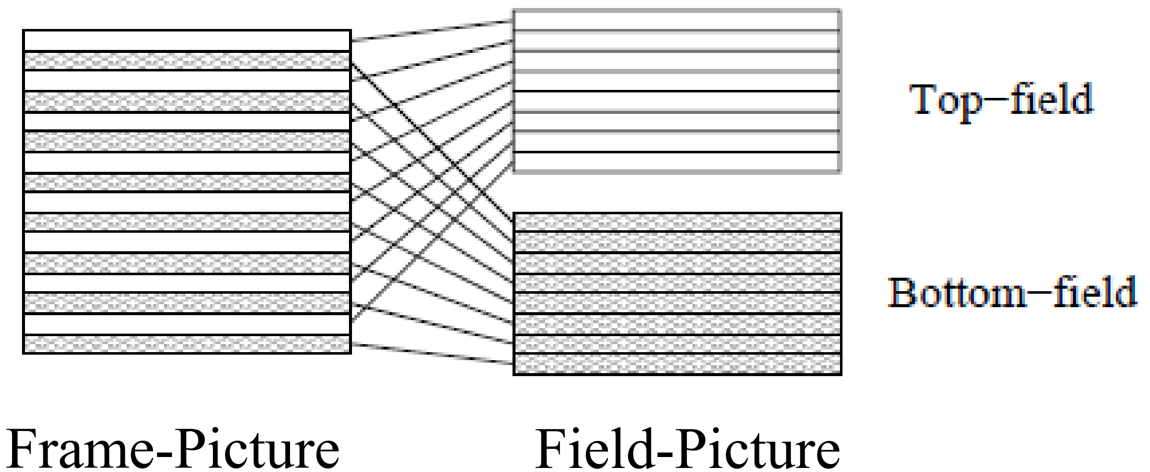
Supporting Interlaced Video

- MPEG-1 supports *only* non-interlaced (*progressive*) video
- MPEG-2 supports interlaced video because it is used in HDTV
- In interlaced video, each *frame* consists of two *fields*: the *top* field and the *bottom* filed



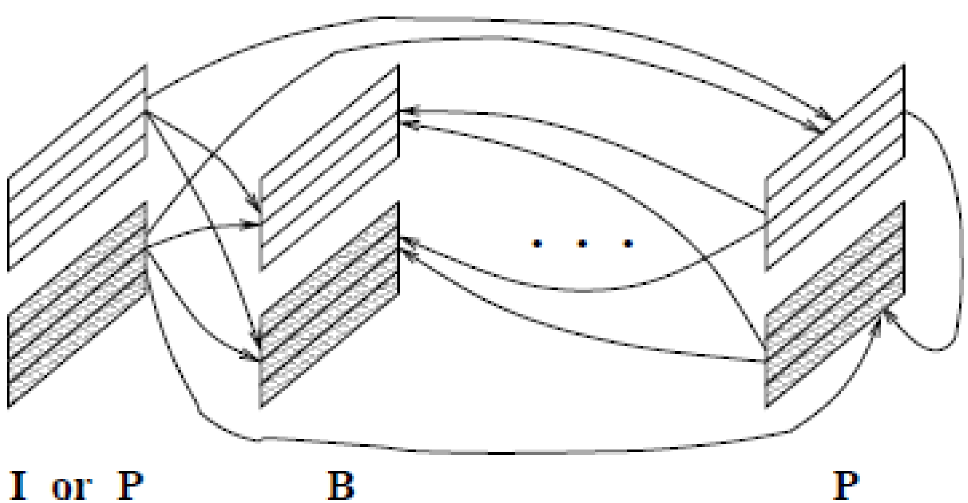
Supporting Interlaced Video

- Two kinds of *pictures*:
  - Frame-Picture: scan lines from *both* fields are interleaved to form a single frame, then divided into 16×16 macroblocks and coded with motion compensation
  - Field-Picture: each field is treated as a separate picture
- Each 16×16 macroblock in the *field-picture* corresponds to a 16×32 block in the *frame-picture*, and each 16×16 block in the *frame-picture* corresponds to a 16×8 block in the *field-picture*



MPEG-2 Motion Compensation

- *Five* modes for motion estimation and prediction:
  - Frame prediction for frame-pictures: similar to MPEG-1 motion compensation in both P- and B-frames
  - Field prediction for field-pictures: macroblocks of 16×16 pixels in the field-pictures are used for MC:
    - P-frames: MBs in top-field are predicted from both fields of previous I- or P-frames, and MBs in bottom-field are predicted from previous frame *and* top-field
    - B-frames: MBs predicted from both fields in previous/next frames



MPEG-2 Motion Compensation

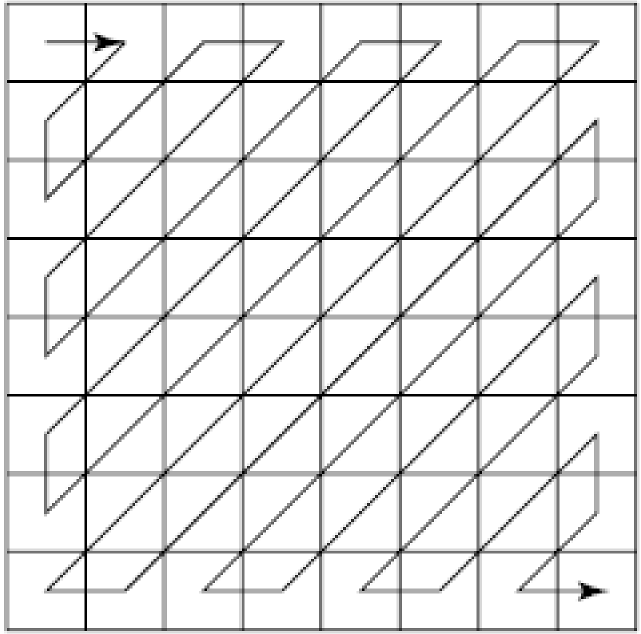
- *Five* modes for motion estimation and prediction:
  - Field prediction for frame-pictures: Divide each 16×16 MB in the target frame into two 16×8 MB one for each field, and do MC for each smaller MB like the *field prediction* before
  - 16×8 MB prediction for field-pictures: Divide each 16×16 from each target *field-picture* into two 16×8 Mbs (top 8 rows separate from bottom 8 rows). Perform field prediction on each smaller MB alone. This produces two MVs for each P-field-picture and up to four MVs for each B-field-picture. Suitable for when there is rapid motion.

MPEG-2 Motion Compensation

- *Five* modes for motion estimation and prediction:
  - Dual-prime for P-pictures:
    - Can be used for either frame-pictures or field-pictures
    - Perform field-prediction on each previous field with the same parity (top or bottom field)
    - Use this MV to derive another calculated MV from the field with opposite parity
    - Average the prediction error from both MVs to arrive at the final prediction error

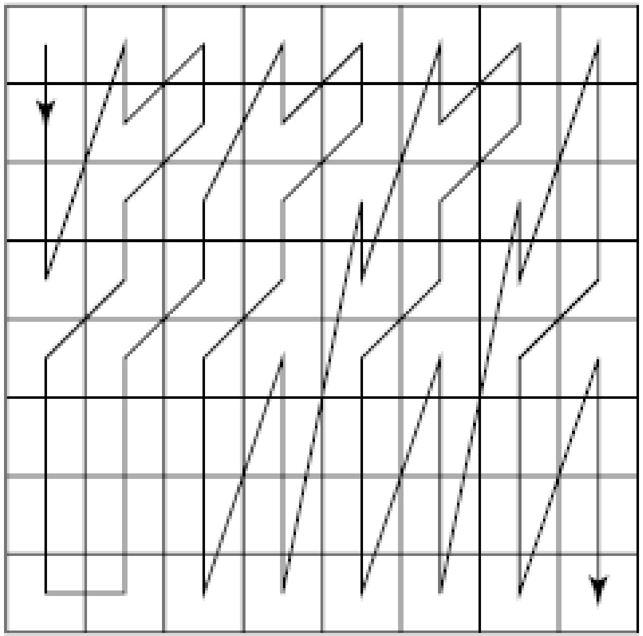
MPEG-2 Alternate Scan

- Recall the zig-zag scanning order in JPEG and MPEG-1 was used to scan the coefficients with *larger* magnitudes *first* followed by coefficients that are close to zero to improve the efficacy of *run length coding* .
- This works well for MPEG-2 *progressive* frames



MPEG-2 Alternate Scan

- For MPEG-2 *interlaced* frames, *consecutive* rows come from *different* fields, and thus have less correlation than *alternate* rows
- Therefore, MPEG-2 uses an *alternate scan* order to scan coefficients with *larger* magnitudes first, and this has been shown by experiments to improve the performance



MPEG-2 Scalability

- MPEG-2 provides different *scalability* modes where the video is sent in *layers* and each layer adds more *quality* to the video
- This is important because MPEG-2 is application-independent and has many different use cases
- The main information is sent in a *base layer* that can be independently encoded, transmitted, and decoded to provide basic video quality
- This is followed by one or more *enhancement layers* that enhance the quality of the video and depend on the previous base or enhancement layers

MPEG-2 Scalability

- MPEG-2 provides the following scalabilities:
  - *SNR Scalability*: enhancement layer provides higher SNR by using coarse quantization (larger steps) in the base layer and finer quantization (smaller steps) in the enhancement layers
  - *Spatial Scalability*: enhancement layer provides higher spatial resolution by sending a lower resolution version in the base layer and higher resolution versions in the enhancement layers
  - *Temporal Scalability*: enhancement layer facilitates higher frame rate by dividing the frames between two layers each with *half* the frame rate

MPEG-2 Scalability

- MPEG-2 provides the following scalabilities:
  - *Hybrid Scalability*: combination of any two of the above three scalabilities
  - *Data Partitioning*: quantized DCT coefficients are split into partitions by sending the DC and lower AC coefficients in the base layer, and higher AC coefficients in the enhancement layers

Other differences from MPEG-1

- 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.

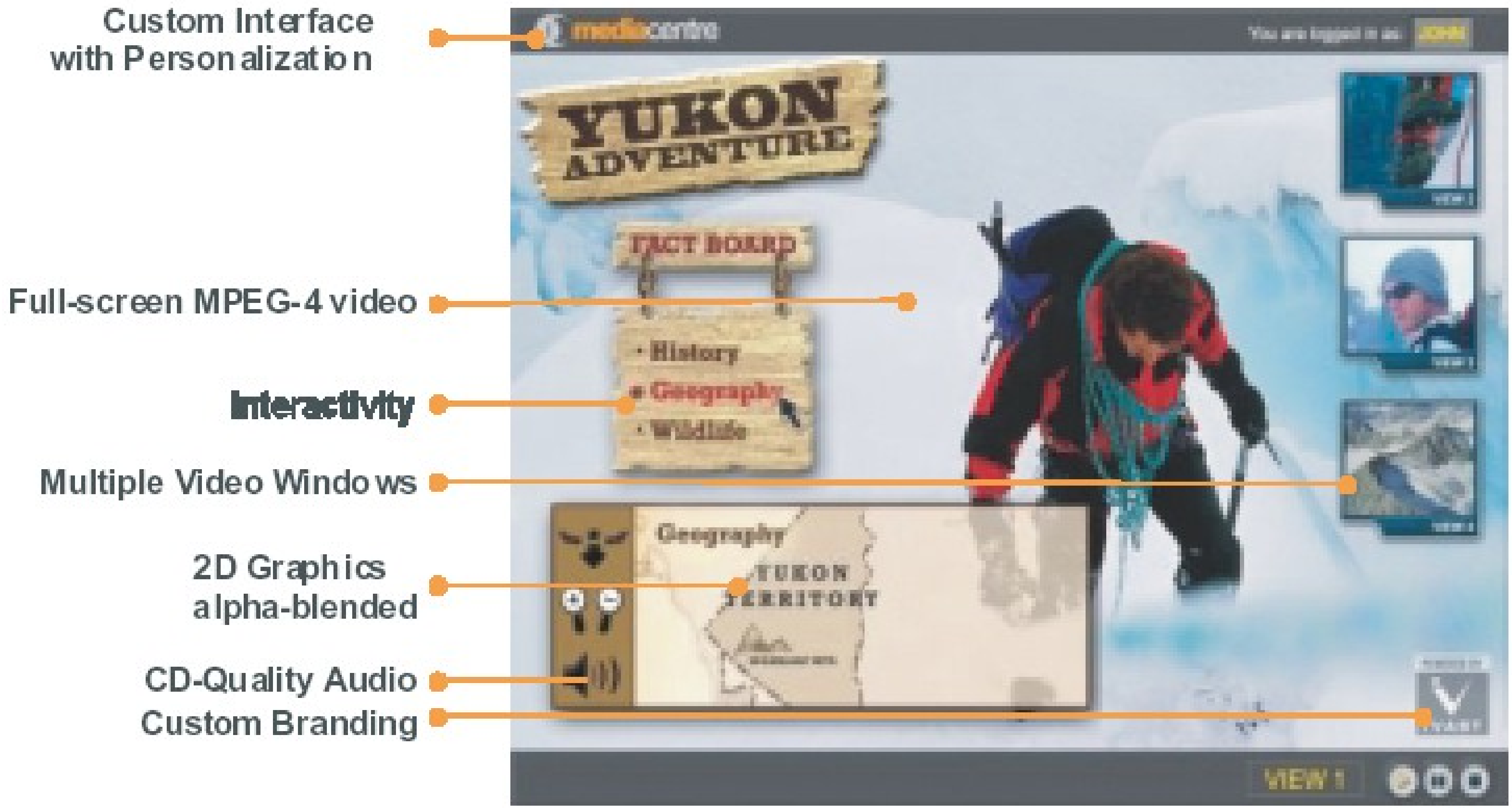
MPEG-3

- Originally started while MPEG-2 was underway
- Was supposed to handle HDTV signals
- Was folded into MPEG-2 as a profile
- So, there is *no* MPEG-3 standard

MPEG-4

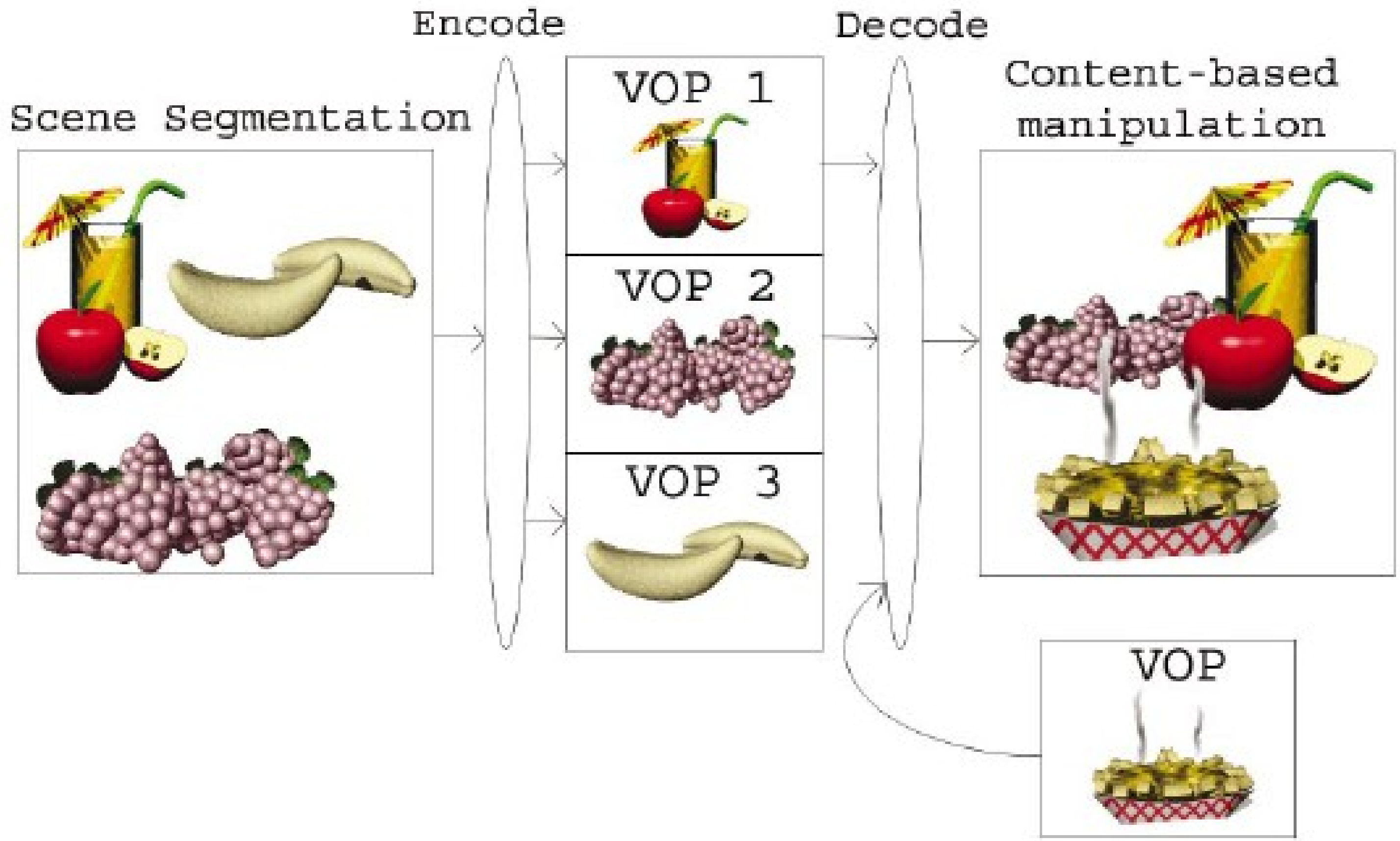
- MPEG-4 (version 1) adopted in 1999 by ISO
- MPEG-4 (version 2) adopted in 2000 by ISO
- Covers a large range of bitrates from 5 kbps up to 10 Mbps
- Targets various applications e.g. digital TV, interactive graphics on WWW, computer games, …
- Focuses more on user *interactivity* than previous MPEG standards:
  - MPEG-1 does not allow any interactivity
  - MPEG-2 allows limited interactivity in e.g. VoD
  - MPEG-4 allows much more interactivity by the user with the audio/visual objects

MPEG-4 Interactivity



Object-Based Coding

- Employs *object-based coding* unlike *frame-based coding* used by MPEG-1 and MPEG-2
  - Offers higher compression rates
  - Video can be composed and manipulated by simple operations on *video objects*



Object-Based Coding

- MPEG-4 is quite different from MPEG-2 and provides:
  - *Composing* media objects to create desirable audiovisual scenes
  - *Multiplexing* and *synchronizing* the bitstreams for these media objects to reach the end-user
  - *Interacting* with the audiovisual scene at the end user

Object-Based Coding

- *Video-object Sequence* (VS): delivers the complete MPEG-4 visual scene, which may contain 2-D or 3-D natural or synthetic objects.
- *Video Object* (VO): a particular object in the scene, which can be of arbitrary (non-rectangular) shape corresponding to an object or background of the scene.

Video-object Sequence (VS)
Video Object (VO)
Video Object Layer (VOL)
Group of VOPs (GOV)
Video Object Plane (VOP)

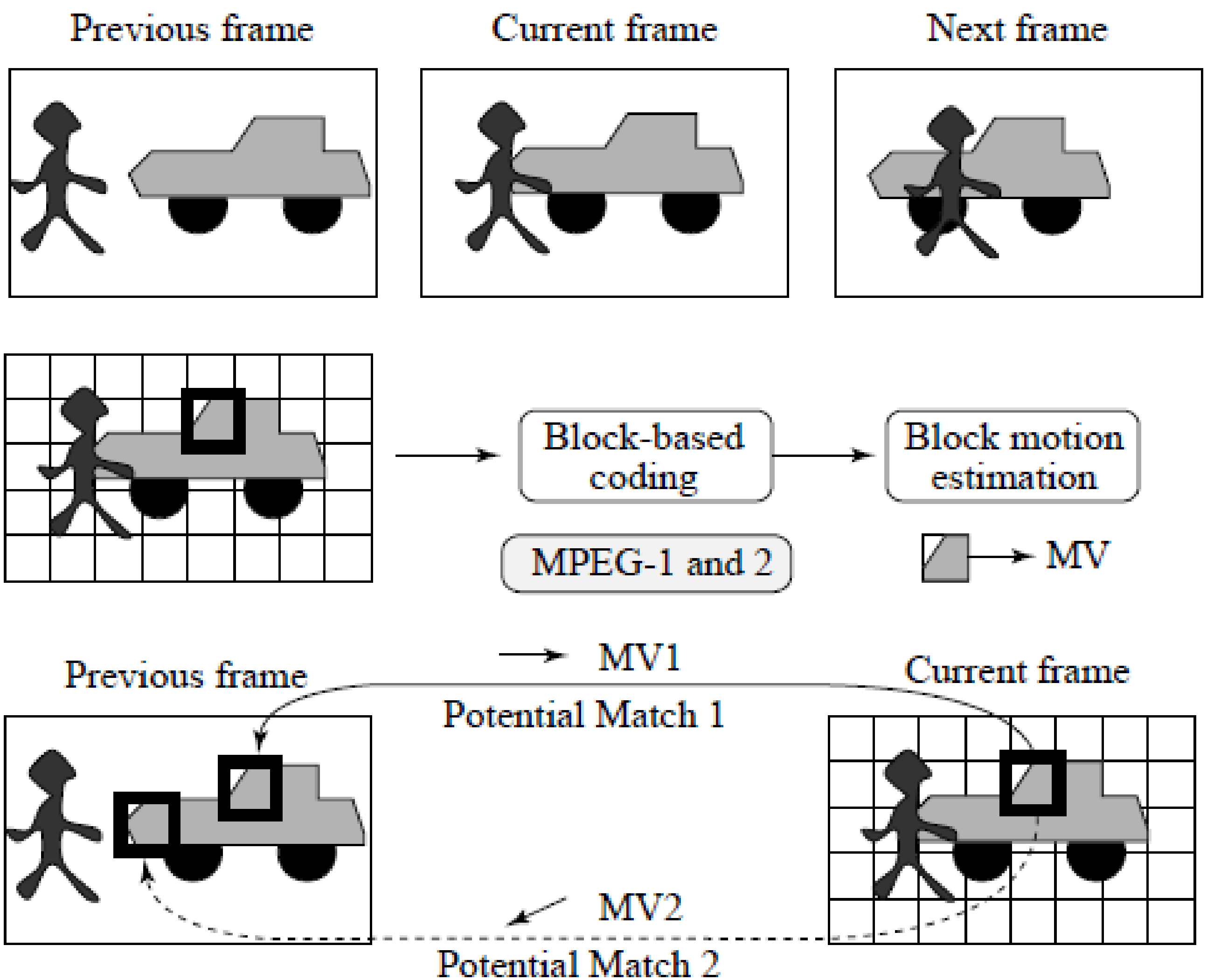
Object-Based Coding

- *Video Object Layer* (VOL): facilitates a way to support (multi-layered) scalable coding. A VO can have multiple VOLs under scalable coding, or have a single VOL under non-scalable coding.
- *Group of Video Object Planes* (GOV): groups Video Object Planes together (optional level).
- *Video Object Plane* (VOP): a snapshot of a VO at a particular moment.

Video-object Sequence (VS)
Video Object (VO)
Video Object Layer (VOL)
Group of VOPs (GOV)
Video Object Plane (VOP)

VOP-based vs Frame-based Coding

- MPEG-1 and -2 do not support the VOP concept, and hence their coding method is referred to as *frame-based* (also known as *block-based*) coding



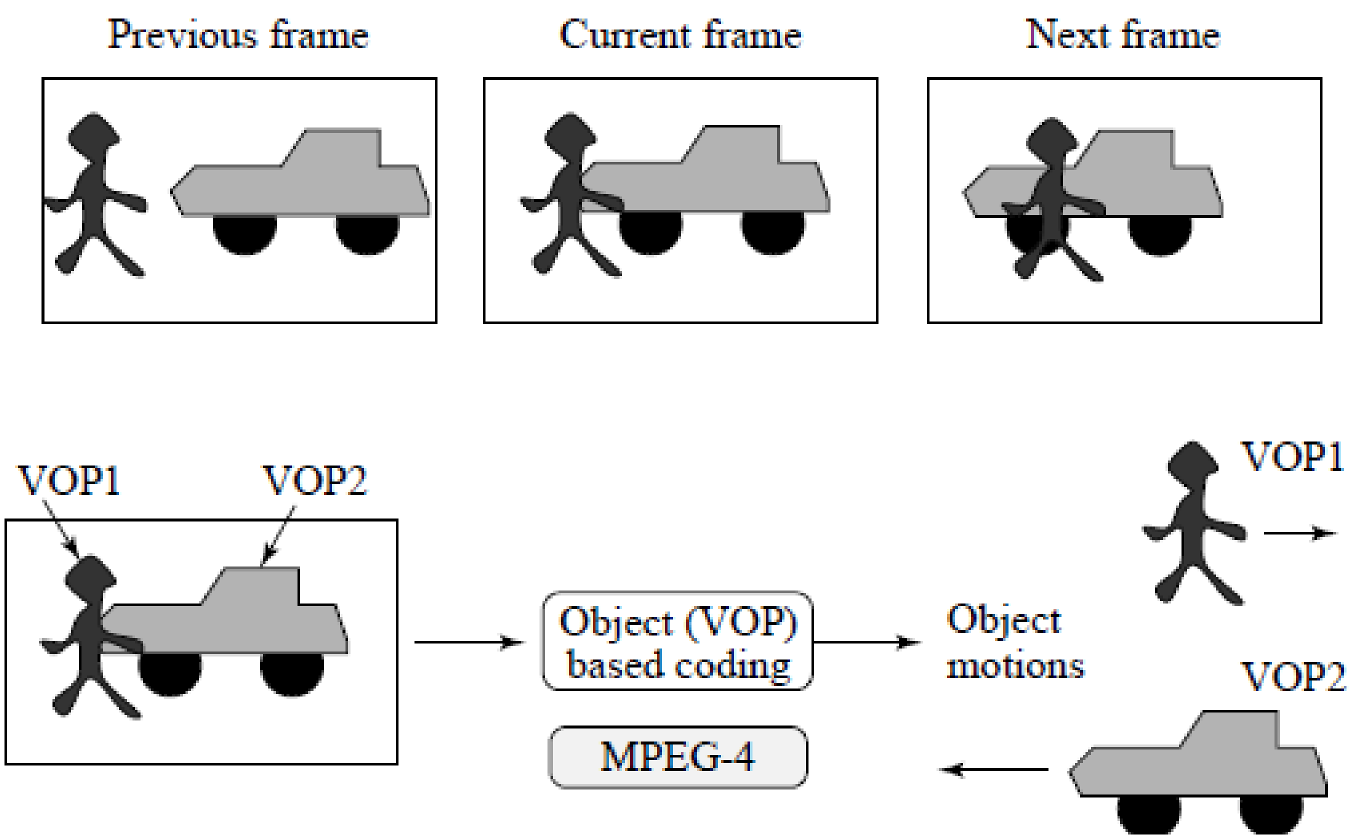
Each *block* gets its own *motion vector*

These *motion vector* might not be consistent with the object's motion

VOP-based vs Frame-based Coding

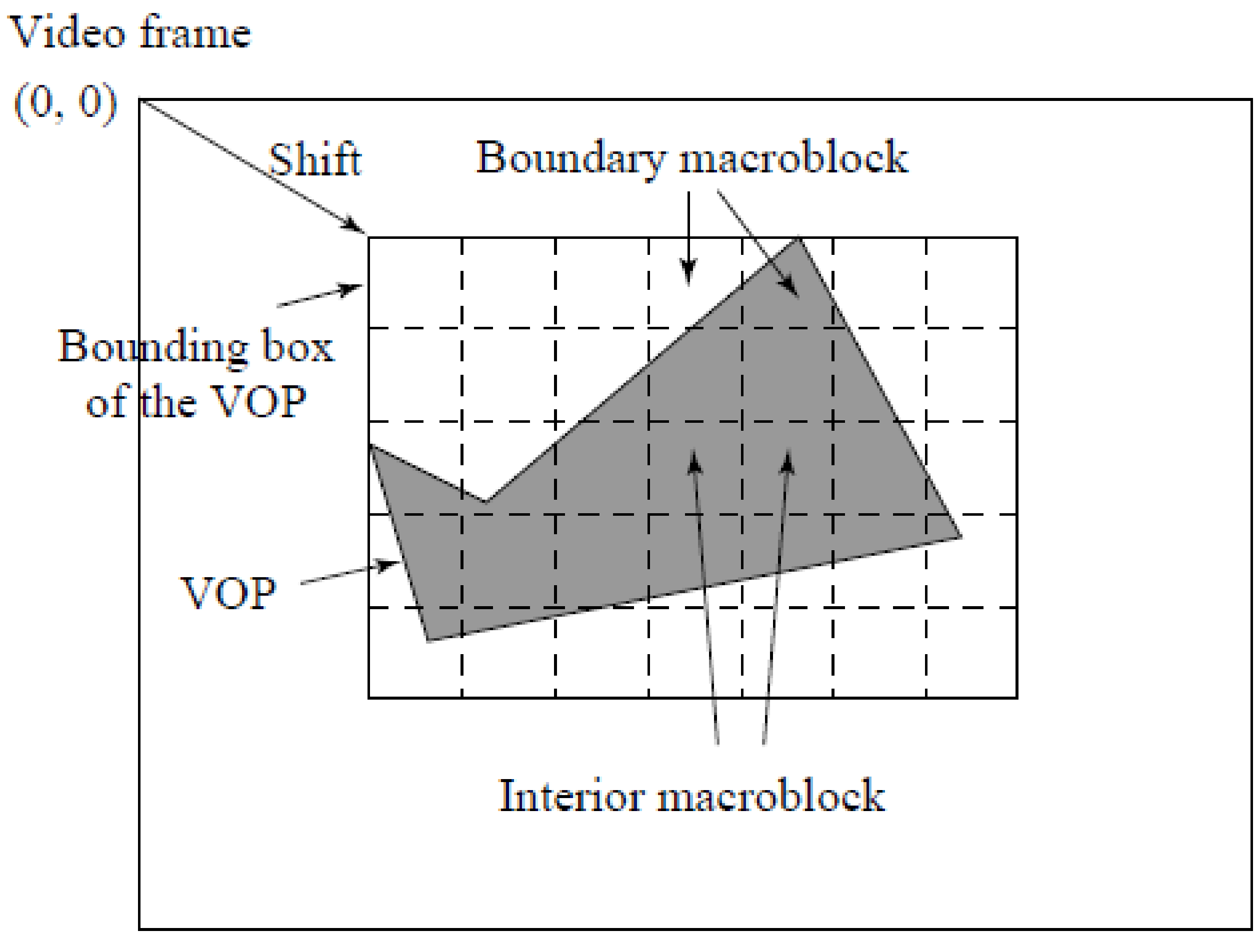
- MPEG-4 supports VOP-based motion compensation, and each VOP can be of arbitrary shape

Each *VOP* gets its own *motion vector*



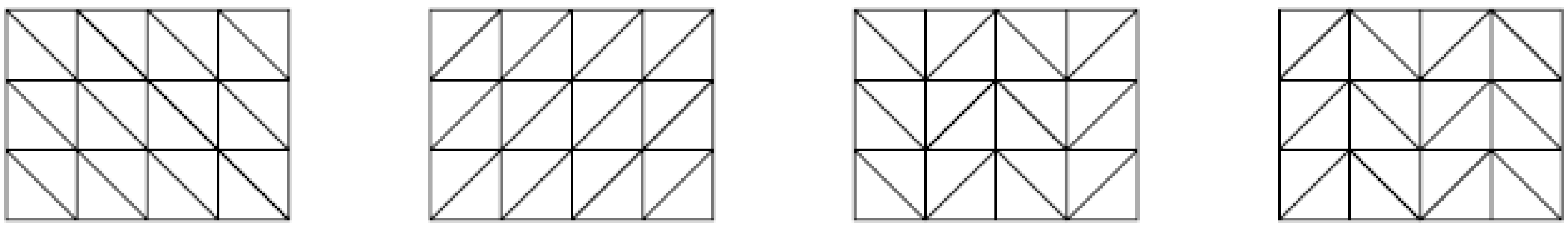
VOP-based Coding

- Each VOP is divided into many 16×16 macroblocks
- MBs from the target VOP are matched with MBs from reference VOPs in previous/next frames
- Boundary MBs need to be *padded* to perform the matching

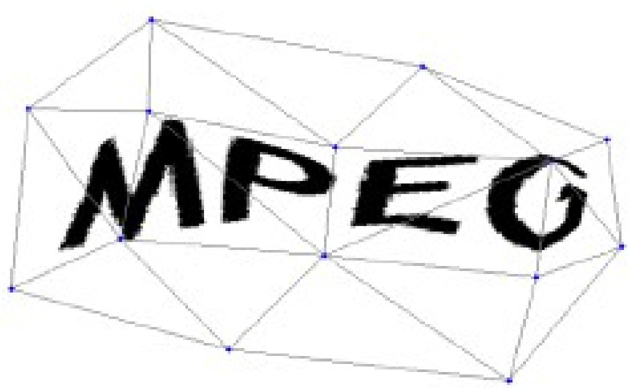


Synthetic Object Coding

- MPEG-4 allows coding of several *synthetic* objects frequently used in videos:
  - *2D Mesh*: a tesellation of a 2D planar region with *triangles*

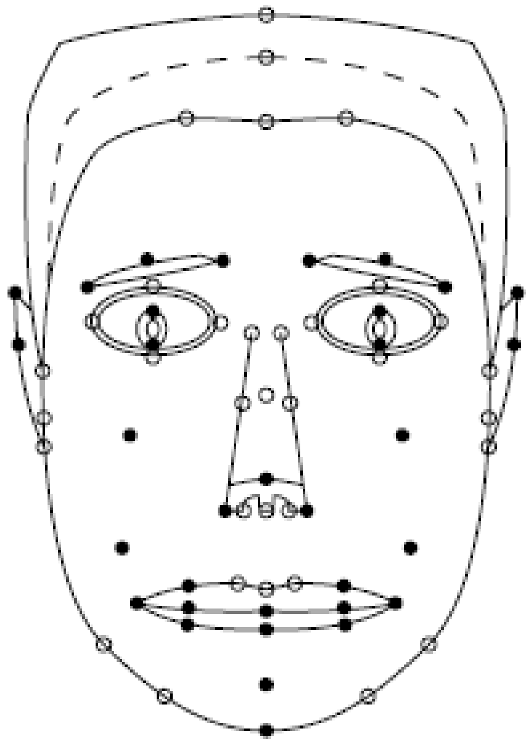


(a) Type 0 (b) Type 1 (c) Type 2 (d) Type 3



Synthetic Object Coding

- MPEG-4 allows coding of several *synthetic* objects frequently used in videos:
  - *3D Models*: 3D face or body objects. Useful in applications such as video conferencing, human-computer interaction, computer games, ...



MPEG-4 Levels and Profiles

- MPEG-4 defines several profiles and levels

Table 12.3: MPEG-4 Levels in Simple, Core, and Main Visual Profiles

Profile	Level	Typical picture size	Bit-rate (bits/sec)	Max number of objects
Simple	1	176×144 (QCIF)	64 k	4
	2	352×288 (CIF)	128 k	4
	3	352×288 (CIF)	384 k	4
Core	1	176×144 (QCIF)	384 k	4
	2	352×288 (CIF)	2 M	16
Main	1	352×288 (CIF)	2 M	16
	2	720×576 (CCIR601)	15 M	32
	3	1920×1080 (HDTV)	38.4 M	32

Famous MPEG-4 Codecs

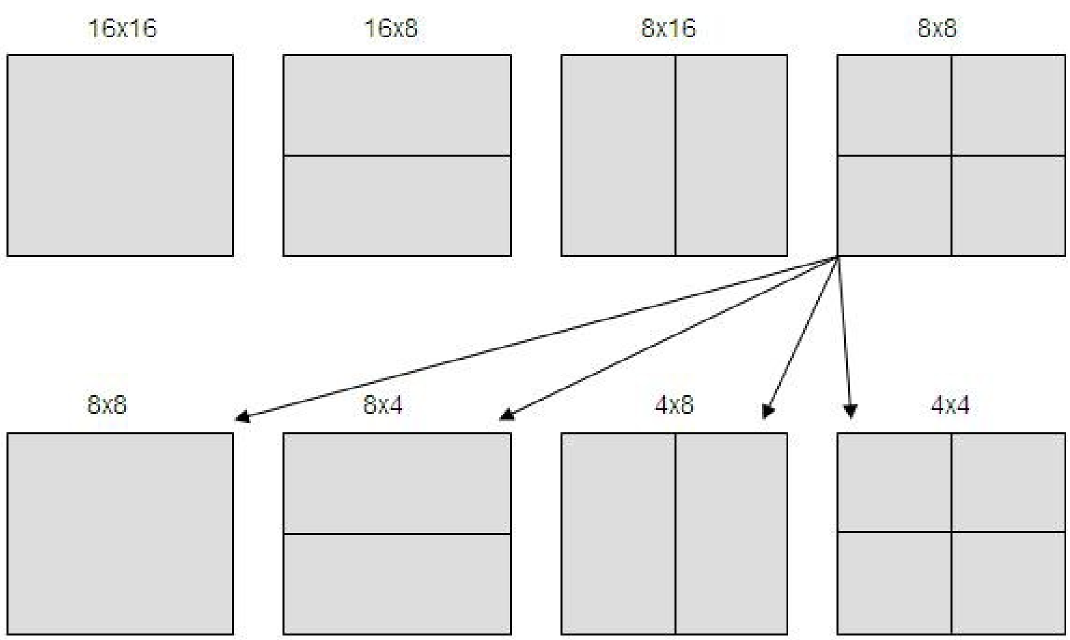
- DivX:
  - A proprietary MPEG-4 Part 2 (MPEG-4 video) encoder/decoder software
  - Started around 2000 as an open source project, but was later withdrawn and used to found what would become DivX, Inc.
- XviD:
  - An open source MPEG4 encoder/decoder
  - Initially based on the then-open source DivX code base, was forked and maintained separately as a free and open source project
- libavcodec:
  - An open source video/audio manipulation library containing many implementations for audio/video formats and standards
  - Used in many open source software programs e.g. ffmpeg, VLC Player, …

MPEG-4 Part10/H.264

- Improved video compression standard adopted in 2002
- ISO MPEG-4 Part 10 Advanced Video Coding AVC
- ITU-T H.264
- Provides 30-50% better compression than MPEG-2 and up to 30% over H.263+ and MPEG-4

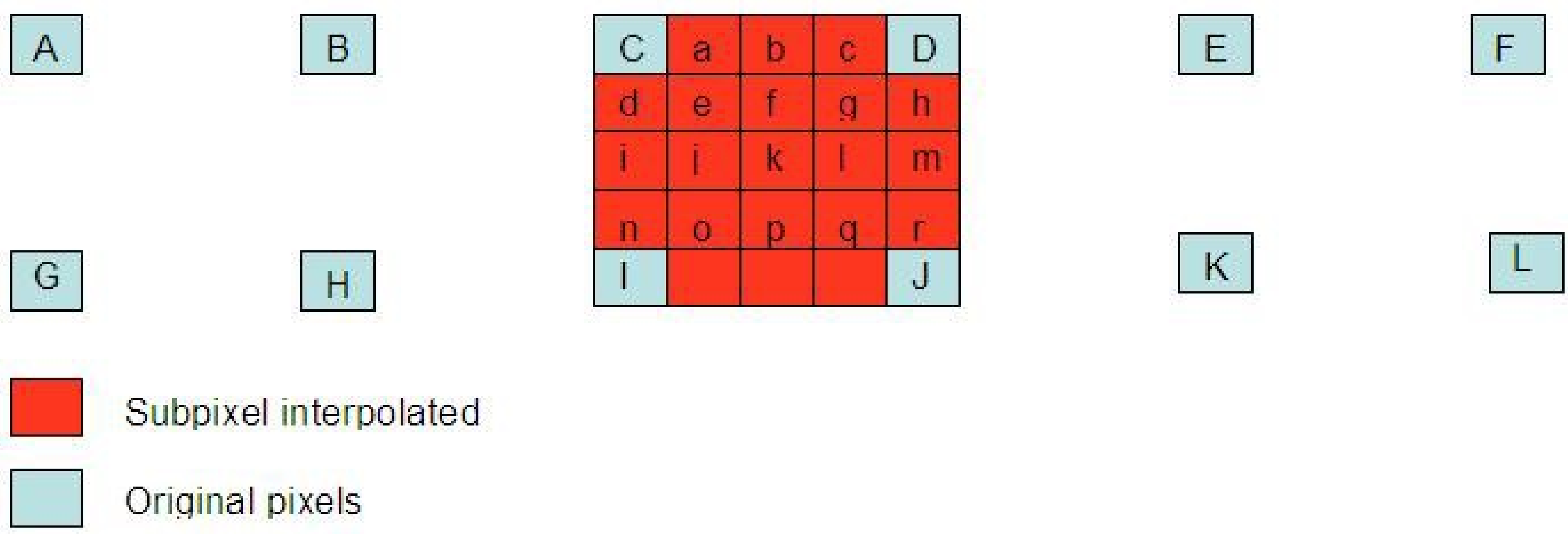
H.264 Main Features

- Flexible Block Partitioning in MC:
  - Macroblock in MPEG-2 uses 16×16 luminance values. MPEG-4 AVC uses a tree-structured motion segmentation down to 4×4 block sizes (16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4). This allows much more accurate motion compensation of moving objects.



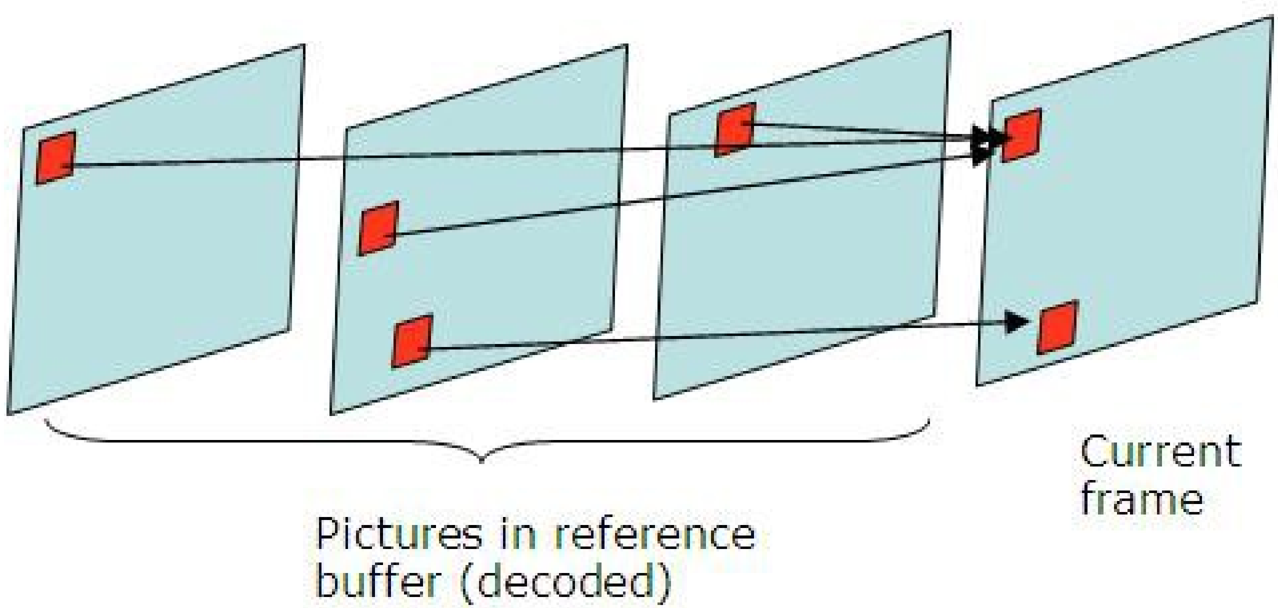
H.264 Main Features

- Sub-pixel Precision in MC:
  - Motion vectors can have up to quarter-pixel precision. This provides more accuracy in the motion compensation process.



H.264 Main Features

- Multiple References:
  - Allows searching *multiple* (up to 32) previous and next frames for the motion vectors



H.264 Main Features

- Intra Prediction:
  - Adds *spatial* prediction in the I-frame coding to reduce their bit rates

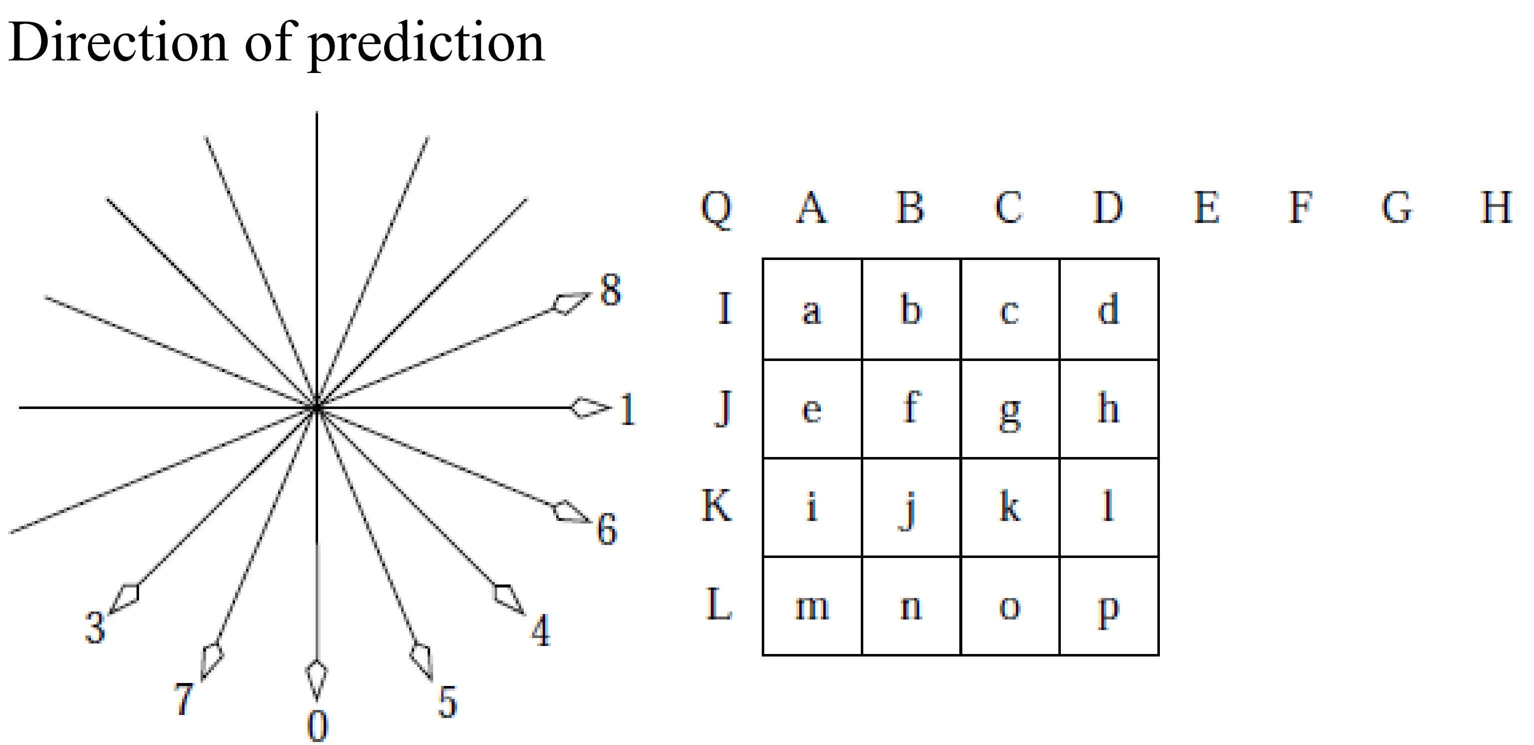


FIGURE 18.17 **Prediction modes for 4 x 4 intra prediction.**

Famous H.264 Codecs

- DivX:
  - A proprietary H.264 codec
- x264:
  - A open source H.264 codec
  - Used by other frontends e.g. ffmpeg, mencoder, …

Video Container Formats

- A format that provides the specification for including different streaming contents in one file e.g. audio and/or video, and can support video/audio encoded in different formats
- The most famous ones are:
  - Audio Video Interleave (.avi):
    - Video container format
    - Can include MPEG-1, MPEG-2, MPEG-4, H.264, …
  - MPEG (.mpeg, .mpg):
    - Can include MPEG-1, MPEG-2, MPEG-4, H.264
  - MPEG-4 (.mp4):
    - MPEG-4 Part2, H.264/MPEG-4 Part10
  - Matroska (.mkv):
    - MPEG-1, MPEG-2, MPEG-4, H.264, …

[https://en.wikipedia.org/wiki/Comparison_of_container_formats]

Recap

- MPEG-1
- MPEG-2
- MPEG-4
- H.264

- Next: Audio Compression

- More information: **FM** Ch. 11.