MPEG-2

ISO/IEC 13818-2 (or ITU-T H.262)
- High quality encoding of interlaced video at 4-15 Mbps for digital video broadcast TV and digital storage media
- Applications
  - Broadcast TV, Satellite TV, CATV, HDTV, video services on networks (e.g., ATM)
  - Killer applications: DVD
MPEG-2 (cont.)

- Parts of MPEG-2:
  - ISO/IEC 13818-1: Systems
  - ISO/IEC 13818-2: Video
  - ISO/IEC 13818-3: Audio
  - ISO/IEC 13818-4: Compliance Testing
  - ISO/IEC 13818-5: Software
  - ISO/IEC 13818-6: DSM-CC
  - ISO/IEC 13818-7: NBC Audio
  - ISO/IEC 13818-8: 10-Bit Video (dropped!)
  - ISO/IEC 13818-9: Real-Time Interface
  - ISO/IEC 13818-10: DSM-CC Conformance

MPEG-2 Requirements

- Coding of **interlaced video** with high quality at 4-15 Mbps
- Random access/channel switching in limited time
- Fast forward/reverse (FF/FR) using access points
- **Scalable video coding** for multi-quality video applications
- System supporting audio-visual synchronized play/access for multiple streams
- A practical/implementable decoder
MPEG-2 Main New Feature

- Frame/field picture structure
- Frame/field/dual prime adaptive motion compensation
- Frame/field adaptive DCT
- **Alternate scan** for DCT coefficients
- Chrominance formats: 4:2:0, 4:2:2, 4:4:4
- Nonlinear quantization table
  - increased accuracy for small values

Positions of Samples

```
4:2:0

4:2:2

: Y samples
: Cr, Cb samples
```
Positions of Samples (cont.)

Interlaced 4:2:0
- `top_field_first` = 1
- `top` = top field pixels
- `bottom` = bottom field pixels

Interlaced 4:2:0
- `top_field_first` = 0
- `bottom` = bottom field pixels
- `top` = top field pixels

Interlaced 4:2:2/4:4:4
- `top_field_first` = 1
- `top` = top field pixels
- `bottom` = bottom field pixels

Progressive
- `time`
Group of Pictures

- Encoder input: 1 2 3 4 5 6 7 8 9 10 11 12 13
  1 B B P B B P B B I B B P
- Encoder output: 1 4 2 3 7 5 6 10 8 9 13 11 11
  I P B B P B B I B B P B B
- Decoder output: 1 2 3 4 5 6 7 8 9 10 11 12 13

MPEG-2 Slice

- Slice – a series of an arbitrary number of consecutive macroblocks
  - The first and last macroblocks of a slice shall not be skipped macroblocks
  - Every slice shall contain at least one macroblock
  - Slices shall not overlap
  - The position of slices may change from picture to picture
  - The first and last macroblock of a slice shall be in the same horizontal row of macroblocks
MPEG-2 Slice (cont.)

- Two slice structure
  - **General slice structure** – the slices do not cover the entire picture
  - **Restricted slice structure** – every macroblock shall be enclosed in a slice
MPEG-2 Macroblocks

- Three different chrominance format for a macroblock

  4:2:0
  Y: 0 1 2 3
  Cb: 4
  Cr: 5

  4:2:2
  Y: 0 1 2 3
  Cb: 4 6
  Cr: 5 7

  4:4:4
  Y: 0 1 2 3
  Cb: 4 6 10
  Cr: 5 7 11

MPEG-2 Streams

- Program streams
  - for error-free environments (such as a disk)
  - use long and variable-length packets for software-based processing

- Transport streams
  - offer robustness necessary for noisy channels
  - use fixed-length packets of 188 bytes
  - well suited for delivering compressed video and audio over error-prone channels such as CATV and satellite transponders
MPEG-2 Scalability

- **Scalability** – allows decoder of various complexities to be able to decode video of resolution/quality commensurate with their complexity from the same bit stream.

MPEG-2 Scalability (cont.)

- MPEG-2 **non-scalable** video codec
MPEG-2 Scalability (cont.)

- MPEG-2 scalable video codec

MPEG-2 Scalability

- Scalability (cont.)
  - Data Partitioning
  - SNR scalability
  - Spatial scalability
  - Temporal scalability
MPEG-2 Scalability (cont.)

- **Data Partitioning**
  - All header, MVs, first few DCT coefficients in the base layer
  - Can be implemented at the bit stream level
  - Simple

MPEG-2 Scalability (cont.)

- **SNR Scalability**
  - Base layer includes coarsely quantized DCT coefficients
  - Enhancement layer further quantizes the base layer quantization error
MPEG-2 Scalability (cont.)

Spatial Scalability

- Spatial Scalability
MPEG-2 Scalability (cont.)

- **Temporal Scalability – option 1**

![Temporal Scalability Option 1 Diagram](image1)

MPEG-2 Scalability (cont.)

- **Temporal Scalability – option 2**

![Temporal Scalability Option 2 Diagram](image2)
MPEG-2 Levels and Profiles

- **Levels** – define the *resolution* of the picture
  - Low level – SIF (360×288)
  - Main level – standard 4:2:0 resolution (720×576)
  - High-1440 level – HDTV (1440×1152)
  - High level – wide-screen HDTV (1920×1152)

MPEG-2 Levels and Profiles (cont.)

- **Profiles** – determine the set of compression tools, compromise between compression rate and the cost of the decoder
  - Simple profile – higher bit-rate, no bidirectional prediction (B pictures)
  - Main profile – the best compromise between rate and cost, use all three image types (I, P and B)
  - SNR scalable profile – enhance quantization accuracy
  - Spatially scalable profile – enhance spatial resolution
  - High profile – for HDTV broadcast applications
MPEG-2 Levels and Profiles (cont.)

- **New Profiles – 4:2:2 and Multiview**
  - **4:2:2** profile – similar to main profile but higher chrominance resolution
  - **Multiview** profile – stereoscopic video for two views

<table>
<thead>
<tr>
<th>Level</th>
<th>Profile</th>
<th>SNR Scalable</th>
<th>Spatial Scalable</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>High 1920×1152 x60</td>
<td>Main</td>
<td>MP@HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-1440 1440×1152 x60</td>
<td>Main</td>
<td>MP@H1440</td>
<td>SSP@H1440</td>
<td>HP@H1440</td>
</tr>
<tr>
<td>Main 720×576 x30</td>
<td>SP@ML</td>
<td>MP@ML</td>
<td>SNP@ML</td>
<td></td>
</tr>
<tr>
<td>Low 352×288 x30</td>
<td>MP@LL</td>
<td>SNP@LL</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
MPEG-2 Levels and Profiles (cont.)

- **MP@ML**
  - Digital TV
  - DVD

- **SP@ML**
  - Digital CATV and VCR
  - 1/2 buffer needed

- **MP@HL**
  - HDTV

---

Motion Estimation/Compensation

- Performed on luminance macroblock (16×16)
- Supporting half-pixel motion compensation
- Chrominance motion vectors are half of luminance MB’s
- -2048 to +2047.5 for half-pixel motion vector
- Depending on motion types:
  - Frame motion vector
  - Field motion vector
  - Motion vector in forward direction
  - Motion vector in backward direction
Motion Estimation/Compensation (cont.)

- MPEG-2 provides two types of picture structures
  - Field picture
  - Frame picture
- Five motion compensation modes
  - Frame prediction for frame pictures
  - Field prediction for field pictures
  - Field prediction for frame pictures
  - Dual-prime prediction for p-pictures
  - 16x8 MC for field pictures

Motion Estimation/Compensation (cont.)

- Mode 1- frame prediction for frame pictures
  - Works well for videos with slow and moderate object and camera motions

Reference frame → Possible interleaving B-picture (Not yet decoded) → Frame-prediction for P-pictures
Motion Estimation/Compensation (cont.)

- Mode 1: frame prediction for frame pictures (cont.)

Reference frame

Possible interleaving B-picture
(Already decoded)

Possible interleaving B-picture
(Not yet decoded)

Frame-prediction for B-pictures

Motion Estimation/Compensation (cont.)

- Mode 2: field prediction for field pictures

Top reference field

Bottom reference field

Possible interleaving B-picture
(Not yet decoded)

Field-prediction for the first field of P-field pictures
Motion Estimation/Compensation (cont.)

- Mode 2: field prediction for field pictures (cont.)

Field-prediction for the 2nd field of P-field pictures when it is bottom field

Possible interleaving
B-picture
(Not yet decoded)

Field-prediction for the second field of P-field pictures when it is top field
Mode 3: field prediction for frame pictures

- The target MB in a frame picture is split into top field pixels and bottom field pixels
- Field prediction is carried out independently for each 16×8 field
- For P-frames, two motion vectors are assigned to each target MB, and two or four motion vectors are assigned to each target MB for B-frames
Motion Estimation/Compensation (cont.)

- **Mode 3: field prediction for frame pictures (cont.)**

  - **Top reference field**
  - **Bottom reference field**
  - Possible interleaving B-picture (Already decoded)
  - Possible interleaving B-picture (Not yet decoded)

  Field-prediction for B-frame pictures

---

Motion Estimation/Compensation (cont.)

- **Mode 4: dual-prime for P-pictures**
  - Only one motion vector is transmitted per MB together with a small differential motion vector
    - Field prediction from each previous field with the same parity is made
    - Each motion vector, $\text{MV}$, is used to derive a calculated motion vector, $\text{CV}$, in the field with opposite parity, taking into account the temporal scaling and vertical shift between lines in the top and bottom fields
    - The pair $\text{MC}$ and $\text{CV}$ yields two preliminary predictions for each MB
    - The prediction errors are averaged and used as the final prediction error
Motion Estimation/Compensation (cont.)
- Mode 4: dual-prime for P-pictures (cont.)
  - For field pictures two motion vectors are used to form predictions from two reference fields (one top, one bottom)
  - For frame pictures, a total of four field predictions are made

Field prediction in field picture
Motion Estimation/Compensation (cont.)

Mode 4: dual-prime for P-pictures (cont.)

Field prediction in frame picture

Field Vector from bitstream

Transmitted field MV

For mvy12: e = -1

mvy12 = mvy11/2 + e

For mvy21: e = +1

mvy21 = 3mvy22/2 + e

Derived Vectors

Motion vector from prediction field 'p' to reference field 'r': mv

Vertical shift correction: e

Transmitted field MV

mv11 = (mvx11, mvy11)

dmv = (dmvx, dmvy)

Derived field MV

mvx22 = mvx11

mvy22 = mvy11

For mvy12: e = -1

mvy12 = mvy11/2 + e

For mvy21: e = +1

mvy21 = 3mvy22/2 + e
Motion Estimation/Compensation (cont.)

- Mode 5: 16×8 MC for field pictures
  - The target MB in a field picture is split into upper half region and lower half region
  - Field prediction is carried out independently for each 16×8 half region
  - For p-frames, two motion vectors are assigned to each target MB, and two or four motion vectors are assigned to each target MB for B-frames
  - Good for finer motion compensation when motion is rapid and irregular
Motion Estimation/Compensation (cont.)

<table>
<thead>
<tr>
<th>Motion Compensation Mode</th>
<th>Use in Field Pictures</th>
<th>Use in Frame Pictures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame Prediction for Frame Pictures</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Field Prediction for Field Pictures</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>Field Prediction for Frame Pictures</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td>Dual-Prime for P-Pictures</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>16×8 MC for Field Pictures</td>
<td>YES</td>
<td>NO</td>
</tr>
</tbody>
</table>

Motion Mode Decision

- For P-Pictures
  - Compute mean square error (MSE) between block and zero motion prediction
  - Compute MSE between block and its MC frame prediction block
  - Compute MSE between block and its MC field prediction block
  - Compute MSE between block and its MC dual-prime prediction block
  - Choose the prediction mode with the least MSE
  - A better strategy may be to weight MSE before mode selection
Motion Mode Decision (cont.)

- For B-Pictures
  - Compute MSE between block and its forward MC frame prediction block
  - Compute MSE between block and its forward MC field prediction block
  - Compute MSE between block and its backward MC frame prediction block
  - Compute MSE between block and its interpolated MC frame prediction block
  - Compute MSE between block and its interpolated MC field prediction block

DCT Coding

- Two types of luminance macroblock structure for DCT coding
  - **Frame DCT coding** - each block shall be composed of lines from the two fields alternately
  - **Field DCT coding** - each block shall be composed of lines from only one of the two fields, applicable only to frame-picture in interlaced videos
DCT Coding (cont.)

- frame DCT coding
- field DCT coding

DCT Coefficients Scan

- Scan order should depend on frequency energy distribution
  - Zigzag scan
  - Alternate scan
Nonlinear Quantization

- The quantization step size, \( \text{step}_\text{size} \), is determined by the product of \( Q[i, j] \) and \( \text{scale} \), where \( Q \) is the default quantization tables for inter- or intra-coding.

- Two types of scales are allowed
  - Linear scale – scale is the same as MPEG-1
    - an integer in the range of \([1, 31]\)
    - \( \text{scale}_i = i \)
  - Nonlinear scale – \( \text{scale}_i \neq i \)

Nonlinear Quantization (cont.)

- Nonlinear scale in MPEG-2

| \( i \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| \( \text{scale}_i \) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 14 | 16 | 18 | 20 | 22 | 24 |
| \( i \) | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| \( \text{scale}_i \) | 28 | 32 | 36 | 40 | 44 | 48 | 52 | 56 | 64 | 72 | 80 | 88 | 96 | 104 | 112 |