

# Module 8

## VIDEO CODING STANDARDS

Version 2 ECE IIT, Kharagpur

# Lesson 24

## MPEG-2 Standards

## Lesson Objectives

At the end of this lesson, the students should be able to:

1. State the basic objectives of MPEG-2 standard.
2. Enlist the profiles and the levels supported by MPEG-2
3. Define field picture and frame picture for interlaced video
4. Illustrate how the field and the frame predictions are made
5. Define the chrominance format for MPEG-2.
6. Explain the basic philosophy of scalable coding
7. Define SNR scalability, spatial scalability and temporal scalability.
8. State the objectives of data partitioning.

## 24.0 Introduction

In lesson –23 we studied the first ISO/IEC MPEG coding standard MPEG-1. It was a generic standard that supported a broad range of applications and application-specific parameters. MPEG continued its standardization efforts and the next standard, MPEG-2 was given the charter to provide video quality not lower than NTSC/PAL and up to CCIR 601 quality. MPEG-2 addresses the emerging applications like digital cable television distribution, high definitions televisions (HDTV), satellite digital video broadcasts, networked multimedia through ATM etc.

In this lesson, we shall focus on the major features of MPEG-2. The MPEG-2 standard supports several profiles and levels for different applications and the lesson will first provide a familiarity with these profiles and levels. It is the first standard to support interlaced video, allowing both frame and field predictions and these will be discussed. The strength of MPEG-2 lies in its scalability support, with will be presented in details.

## 24.1 Basic Objectives of MPEG-2 standard

MPEG-2 standard was designed with the following objectives –

- Compression, coding and transmission of high quality multi-channel, multimedia signals for terrestrial broadcast, digital, cable TV distribution, broadband networks etc.
- Defining *profiles* and *levels* as the subset of syntax to suit wide range of applications.

- Scalable bit stream
- Error-correction capabilities
- Backward compatibility with MPEG-1, so that every MPEG-2 compatible decoder can decode a valid MPEG-1 bit stream

## 24.2 Profiles and levels of MPEG-2

Since MPEG-2 standard encompasses diverse applications requirements, a single syntax was defined by integrating many video coding algorithms. However, implementation of the full syntax was not very practical and some subsets of the syntax were defined with some *profiles* and *levels*. Accordingly a decoder's capabilities to decode a particular bit stream get defined.

MPEG-2 supports following five profiles in decreasing order of hierarchy.

- High
- Spatial scalable
- SNR scalable
- Main
- Simple

Each profile adds a new set of algorithms and acts as a superset of the algorithms supported in the profile below. A level specifies the range of parameters that are supported by the implementation, i.e., image size, frame rates and bit-rates. MPEG-2 supports following four levels-

- High
- High-1440
- Main
- Low

Table 24.1 lists the algorithms and functionalities supported by the different profiles and table 24.2 lists the upper bound of parameters at each level of a profile.

**Table 24.1 Algorithms and functionalities under each profile**

Profile	Algorithms and functionalities
High	All functionalities provided by spatial scalable profile plus <ul style="list-style-type: none"> <li>• 3-layers of SNR and spatial scalable coding</li> <li>• 4 :2 :2 YUV representation</li> </ul>
Spatial scalable	All functionalities provided by SNR scalable profile plus <ul style="list-style-type: none"> <li>• 2- layers of spatial scalable coding</li> <li>• 4 : 1 :1 YUV representation</li> </ul>
SNR scalable	All functionalities provided by Main profile plus <ul style="list-style-type: none"> <li>• 2-layers of SNR scalable coding</li> <li>• 4 :2:0 YUV representation</li> </ul>
Main	All functionalities provided by simple profile plus : <ul style="list-style-type: none"> <li>• Coding interlaced video</li> <li>• Random access</li> <li>• B-picture prediction modes</li> <li>• 4 :2 :0 YUV representation</li> </ul>
Simple	Does not support B-picture prediction

**Table 24.2 Upper bound of parameters at each level**

Level	Parameters constraints
High	1920 pixels/line, 1152 lines/frame, 60 frame/sec
High-1440	1440 pixels/line, 1152 lines/frame, 60 frames/sec
Main	720 pixels/line, 576 lines/frame, 30 frames/sec
Low	352 pixels/line, 288 lines/frame, 30 frames/sec.

## 24.3 Interlaced Video: Frame picture and field picture

Broadcast television applications follows interlaced scanning in which a frame is partitioned into a set of odd-numbered scan lines (referred to as *odd field*) and a set of even numbered scan line (referred to as *even field*). If the input is interlaced, the output of the encoder consists of a sequence of fields that are separated by one field period.

MPEG-2 supports two new picture formats – *frame pictures*, and *field pictures*. In field picture, every field is coded separately. Every field is separated into non

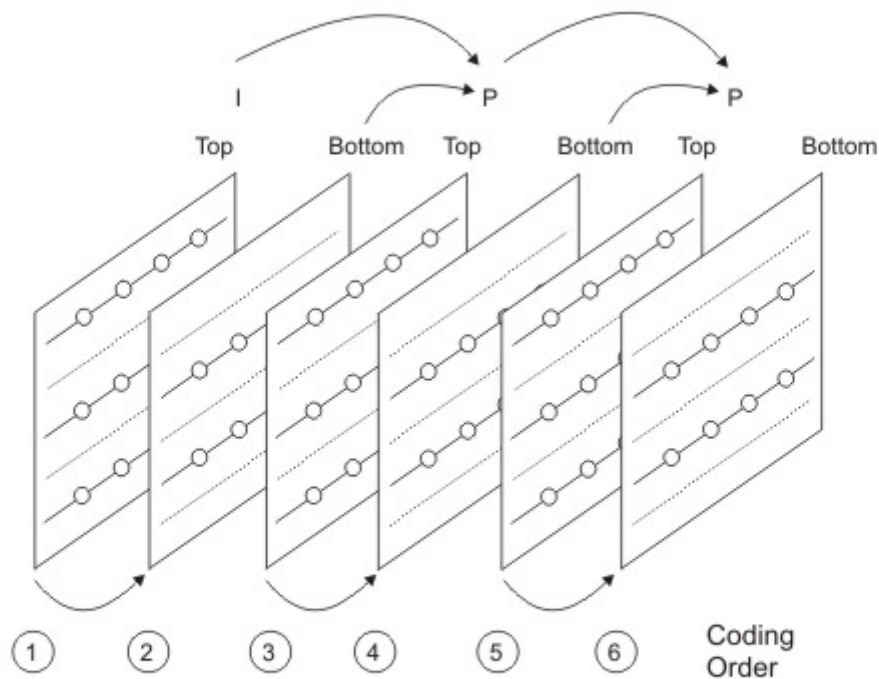
overlapping macroblock and DCT is applied on a field basis. In *frame pictures*, the two fields are coded together as a frame, similar to the conventional coding of progressive video sequence.

Frame pictures are preferred for relatively still images and field pictures give better results in presence of significant motion. It is possible to switch between the frame picture and the field picture on a frame-by-frame basis.

Each frame picture or a field picture may be I-type, P-type or B-type.

## 24.4 Field and frame prediction

It is possible to predict a *field picture* from previously decoded field pictures. Each *odd field* (top field) is coded using motion compensated inter-field prediction based on the previously coded *even field* (bottom field). Each *even field* may either be predicted through motion compensation on a previously coded *even field* or from previously coded *odd field* belonging to the same picture. Within a field picture, all predictions are field predictions. Fig. 24.1 illustrates the field picture prediction mechanism.



24.1 Field prediction in MPEG-2

*Frame pictures* can either have a frame prediction or field prediction and the prediction mode may be selected on a macroblock to macroblock basis. MPEG-2 also supports a *dual prime* prediction in which two independent predictions are made - one for the 8-lines which correspond to the odd (top)field, another for the 8 even (bottom)field lines.

## 24.5 Chrominance format for MPEG-2

In digital video encoding, chrominance format describes the ratio between the horizontal spatial sampling frequencies of the luminance and chrominance components. The chrominance format is expressed as three numbers - the first represents the luminance (Y) sampling frequency, the second and the third represent chrominance U and V sampling frequencies respectively. By convention, the first number is always taken as 4.

In MPEG-1, both U and V are sampled at half the sampling rate of Y in both horizontal and vertical directions (i.e., there is one sample each of U and V for every four Y samples). It should have been called as 4:1:1, but is referred to as 4:2:0 since the relative positions of luminance and chrominance in these two formats differ. In 4:2:0, the chrominance samples are located in between the grids for luminance samples, as shown in fig 24.2(b), whereas in 4:1:1 format, the U and V samples have same spatial locations as that of Y, as illustrated in Fig 24.2(a).

MPEG-2 not only supports the 4:2:0 format, but also the 4:2:2 format illustrated in fig 24.2(c), in which case, the chrominance sub-sampling is done in only one direction (horizontal), but in the vertical, the same sampling frequency as that of luminance is maintained.

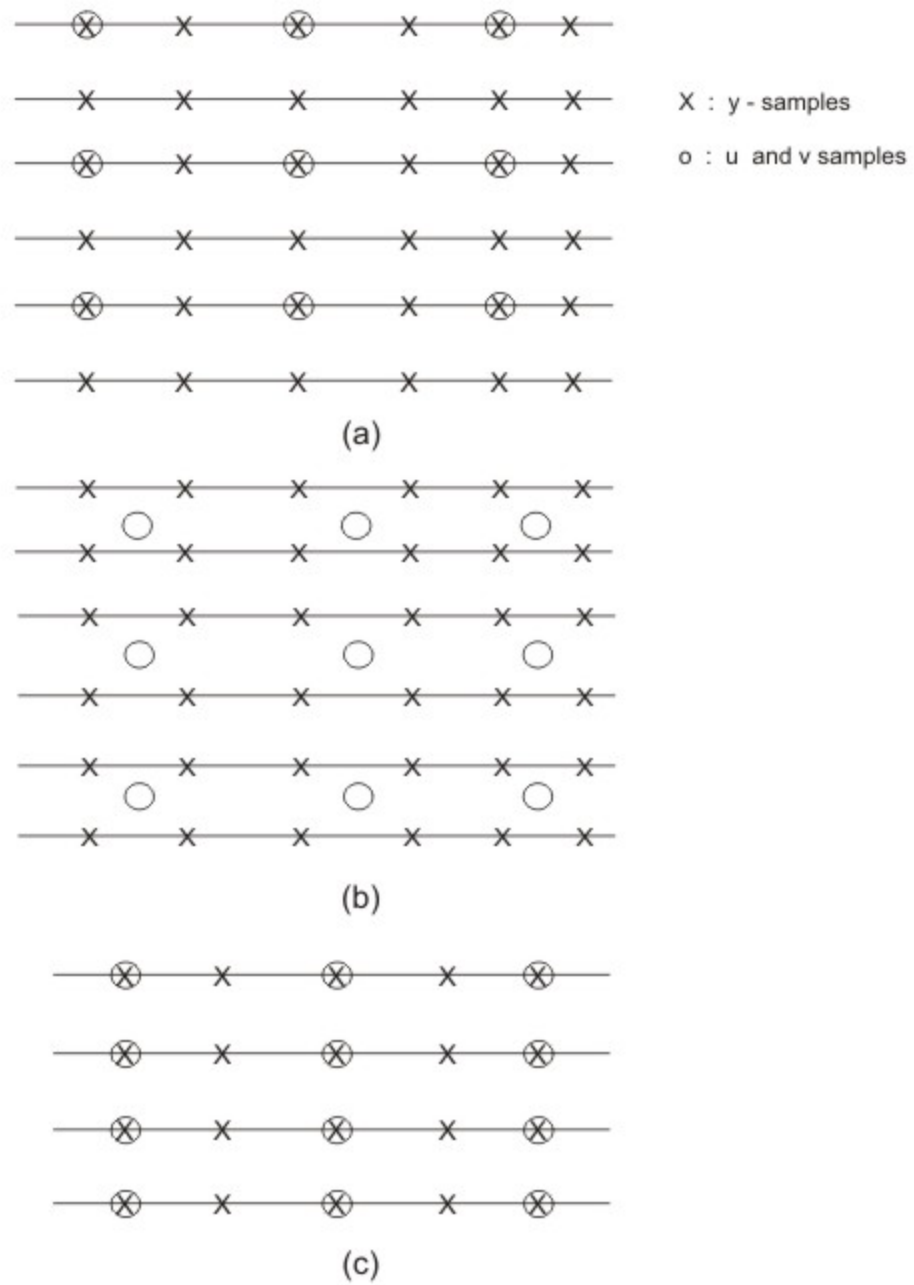


Figure 24.2 (a) 4:1:1 chrominance format  
 (b) 4:2:0 chrominance format  
 (c) 4:2:2 chrominance format



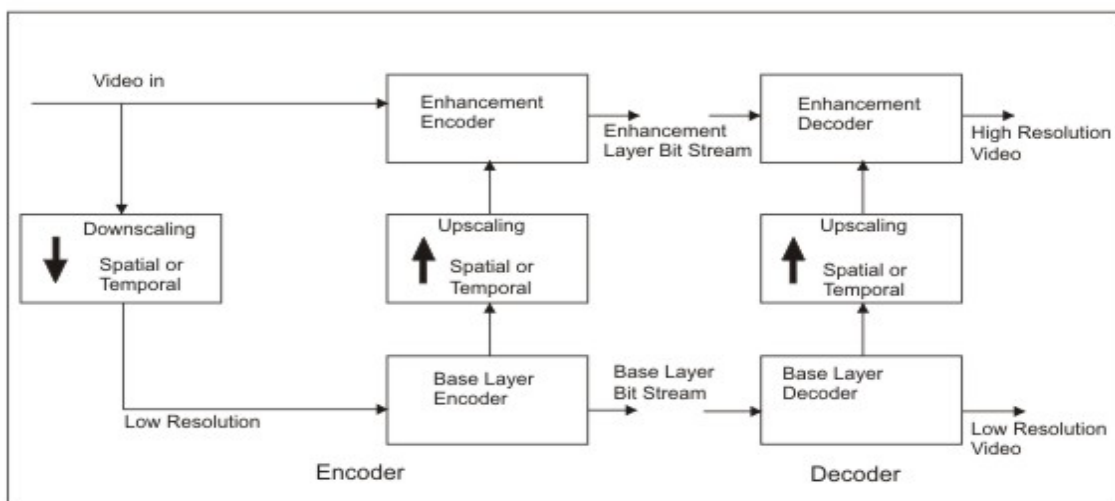
## 24.6 Scalability support of MPEG-2

MPEG-2 standard supports scalability to provide interoperability between different services and to support receivers with different display capabilities. Receivers not having the capability to reconstruct full resolution video can decode only a subset of the layered bitstream to reconstruct a reduced resolution video.

The bit-stream is organized into layers having two or three hierarchies. The bottom of the hierarchy contains base layer, which every receivers and every application must make use of. Above the base layer, enhancement layers exist, which will be used by high-end applications.

The scalability support is of particular interest for SDTV (Standard Definition Television) and HDTV applications. Instead of providing separate bitstreams for SDTV and HDTV, one common scalable bitstream is provided. The SDTV applications can be addressed by the base-layer and only a combination of base-layer and enhancement layers can address the HDTV applications.

Fig.24.3 illustrates the basic philosophy of a multi-scale video-coding scheme. Here, a downscaled version is encoded into a base-layer bitstream with reduced bit-rate. The reconstructed base-layer video is up-scaled spatially or temporally to predict the original input video. The prediction error is encoded into an enhancement layer bitstream. The scalable coding can be used to encode video with a suitable bit-rate allocated to each layer in order to meet the specific bandwidth requirement of the transmission channels or the storage media. Browsing through video databases or transmission of video over heterogeneous networks can benefit from the scalability.



24.3 Multi-scale video encoding and decoding

MPEG-2 has standardized three scalable coding schemes: (a) signal-to-noise ratio (SNR) scalability, (b) spatial scalability and (c) temporal scalability – each of which are targeted to specific requirements.

## 24.7 Scalable Coding Schemes

In this section, we are going to discuss each of the three scalable coding schemes just mentioned.

### 24.7.1 SNR Scalability

SNR scalability is intended for use in video applications involving telecommunications, video services with multiple qualities. The SNR scalable algorithms use a frequency (DCT-domain) scalability technique in which both base-layer and the enhancement layers are encoded at the same spatial scale but using different quantization for DCT coefficients. At the base-layer, the DCT coefficients are coarsely quantized to achieve moderate image quality at reduced bit rate. The enhancement layer encodes the difference between the non-quantized DCT coefficients and the coarsely quantized coefficients from the base-layer with fine quantization step-sizes. The SNR scalability is obtained as a straight forward extension to the main profile and obtains good coding efficiency.

### 24.7.2 Spatial scalability:

Spatial scalability is designed to support displays having different spatial resolution using one common layered bit-stream. This scheme best suits SDTV/HDTV applications. The base-layer encodes a spatially down-sampled video sequence and the enhancement layer encodes the extra information that would be necessary to support higher spatial resolution displays. The spatial scalability algorithm is based on the classical pyramidal approach for progressive image coding.

### 24.7.3 Temporal Scalability:

Temporal scalability is intended for use in systems where a migration into the higher temporal resolution from a lower one may be necessary.

Temporal scalability is achieved by skipping certain fields/ frames at the base-layer. The skipped frames are then encoded at the enhancement layer. The enhancement layer forms its predictions from either the decoded picture at the base layer or from previous temporal prediction at the enhancement layer.

Temporal scalability can be used to accommodate both interlaced and progressive video. The base layer can be interlaced and the enhancement layer can be a progressive HDTV video sequence.

## 24.8 Data partitioning in MPEG-2 bit-stream

MPEG-2 bit-stream has a provision for data partitioning according to the priorities to support error concealment in presence of transmission or channel errors. Similar to the SNR scalability, the algorithm is based upon the separation of DCT coefficients in two layers with different error likelihood. This scheme is implemented with a very low complexity as compared to the scalable coding schemes.