Optical Storage Technology

MPEG Data Compression
MPEG-1 Audio Standard

- Moving Pictures Expert Group (MPEG) was formed in 1988 to devise compression techniques for audio and video.
- It first devised the ISO/IEC International Standard 11172 “Coding of Moving Pictures and Associated Audio for Storage Media at up to about 1.5 Mbit/s” in 1992. It is known as MPEG-1.
- The standard has three major parts: system, video and audio.
- The max. audio bit rate is set at 1.856 Mbps. It supports coding of 32, 44.1 and 48 kHz PCM data at bit rates of approximately 32 to 224 kbps/channel.
- The ISO/MPEG-1 standard was developed to support audio and video coding for CD playback within the CD bandwidth of 1.41 Mbps.
MPEG-1 Audio Standard

- The 11172-3 standard describes three layers of coding, each with different applications.
- **Layer I** describes the least sophisticated method that requires relatively high data rates (~ 192 kbps/channel).
- **Layer II** is based on layer I but is more complex and operates at somewhat lower data rates (~ 96-128 kbps/channel).
- **Layer IIA** is a joint stereo version operating at 128 and 192 kbps per stereo pair.
- **Layer III** is somewhat conceptually different from I and II, is the most sophisticated, and operates at lowest data rates (~ 64 kbps/channel).
- In each case, the encoders are not defined by the ISO/MPEG-1 standard, only the decoders are specified.
MPEG-1 Audio Standard

- MPEG data is transmitted in **frames** with each frame being individually **decodable**.
- The **length of a frame** depends on the layer and MPEG algorithm used.
- A frame begins with a **32-bit ISO header** with a **12-bit synchronizing pattern** and **20 bits of general data** on layer, bit rate index, sampling frequency, type of emphasis, etc.
- This is followed by an **optional 16-bit CRCC check word**.
- Subsequent fields describe **bit allocation data**, **scale factor selection data**, and **scale factors** themselves.
- The largest part of the frame is occupied by **sub-band samples**. This varies among layers.
MPEG-1 Audio Standard

- Frames contain **384** samples in **layer I** and **1152** samples in **II** and **III** (8 or 24 ms respectively at 48 kHz).
- **Layer I** preserves **highest fidelity** for acquisition and production work at **high bit rates** where 6 or more codings can take place.
- Layer II distributes programs efficiently where two codings can occur. **Layer III** is **most efficient**, with **lowest rates**, with somewhat **lower fidelity**.
- At rates as low as **128 kbps**, **layer II** and **III** can convey **stereo** material that is subjectively **very close** to **16-bit fidelity**.
- **MPEG-2** incorporates the three audio layers of MPEG-1, and adds additional features, principally **surround sound**.
ISO/MPEG/AUDIO layer I frame structure: valid for 384 PCM audio input samples
Duration: 8 ms with a sampling rate of 48 kHz

<table>
<thead>
<tr>
<th>Header</th>
<th>Bit allocation</th>
<th>Scalefactors</th>
<th>Subband</th>
<th>Samples</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) 12 bit sync</td>
<td>16 bit</td>
<td>6 bit linear</td>
<td>1 subband sample corresponds to 32 PCM audio input samples.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B) 20 bit system info</td>
<td>4 bit linear</td>
<td>384 x 4 x 32 = 1536</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Auxiliary data field length not specified

2 channels = 3072
ISO/MPEG/AUDIO layer II frame structure: valid for 1152 PCM audio input samples
Duration: 24 ms with a sampling rate of 48 kHz

<table>
<thead>
<tr>
<th>Header</th>
<th>Bit allocation</th>
<th>Scalefactors</th>
<th>Subband</th>
<th>Samples</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) 12 bit sync</td>
<td>4 bit linear</td>
<td>6 bit linear</td>
<td>Gr0</td>
<td>12 granules [Gr] of 3 subband samples each.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3 bit linear</td>
<td></td>
<td>Gr11</td>
<td>Auxiliary data field length not specified</td>
<td></td>
</tr>
<tr>
<td>B) 20 bit system info</td>
<td>2 bit linear</td>
<td>3 subband samples correspond to 96 PCM audio input samples.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 10.16 Structure of the ISO/MPEG-1 audio Layer I, II, and III bit streams. The header and some other fields are common, but other fields differ. Higher-level coders might transcode lower-level bit streams. A. Layer I bit stream format. B. Layer II bit stream format. C. Layer III bit stream format.
Layer I

- Its aim is to provide **high fidelity** at **low cost**, at a somewhat **high data rate**.
- A **poly-phase filter** is used to split the wideband signal into **32 subbands** of equal width. The filter and its inverse are **not lossless**; however, the error is small.
- In layer I, **12** subband samples from each of the **32** subbands are grouped to form a **frame**; this represents **384** wideband samples.
- Based on the calculated **masking threshold** (**just audible noise**), the **bit allocation** determines the **number of bits** used to quantize those samples.
- The normalized samples are **quantized** by the **bit allocator** to achieve **data reduction**.
- Quantizer provides **$2^n - 1$** steps where $2 \leq n \leq 15$. 
Figure 10.17 ISO/MPEG-1 Layer I or II audio encoder and decoder. The 32-subband filter bank is common to all three layers. A. Layer I or II encoder (single-channel mode). B. Layer I or II two-channel decoder.
Layer I

Figure 10.18 Example of an ISO/MPEG-1 Layer I encoder; the FFT side chain is omitted. (Philips)
Figure 10.19  Example of an ISO/MPEG-1 Layer I decoder. (Philips)
Layer II

- Layer II is similar to Layer I, but is more sophisticated in design. It strives to provide high fidelity at moderate data rate, with somewhat higher cost.
- Generally, Layer I operating at 384 kbps achieves the same quality as a Layer II coder operating at 256 kbps.
- The filter creates 32 equal-width subbands, but frame size is tripled to 3 x 12 x 32, corresponding to 1152 wideband samples per channel.
- In other words, data is coded in three groups of 12 samples for each subband (Layer I uses one group).
- Quantization covers a range from 3 ($2^2-1$) to 65535 ($2^{16}-1$). Lower subbands can receive as many as 15 bits, middle subbands can receive 7 bits, and higher subbands are limited to 3 bits.
Layer II

- For greater efficiency, **three successive samples** (for all 32 subbands) are grouped to form a **granule** and quantized together.

*Figure 10.20* ISO/MPEG-1 Layer II audio encoder (single-channel mode) showing scale factor selection and coding of side information.
### TABLE 10.4 Comparison of parameters in ISO/MPEG-1 Layer I and Layer II.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MPEG layer-I</th>
<th>MPEG layer-II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame length (samples)</td>
<td>384</td>
<td>1152</td>
</tr>
<tr>
<td>Subbands</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Subband samples</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>FFT (samples)</td>
<td>512</td>
<td>1024</td>
</tr>
<tr>
<td>Bit allocation (bits)</td>
<td>4 per</td>
<td>2 to 4 depending on subband</td>
</tr>
<tr>
<td>Scalefactor select information (bits)</td>
<td>None</td>
<td>2 per subband</td>
</tr>
<tr>
<td>Scalefactors (bits)</td>
<td>6 per subband</td>
<td>6 to 18 per subband (selectable)</td>
</tr>
<tr>
<td>Sample grouping</td>
<td>None</td>
<td>3 per subband (granule)</td>
</tr>
</tbody>
</table>
Layer I & II encoding algorithm

Figure 10.22 Flow chart of the entire ISO/MPEG-1 audio Layer I and II encoding algorithm.
Layer III

- Layer III is more complex than Layers I and II; its strength is **moderate fidelity** even at **very low data rates**.
- Layer III files is also known as **MP3** files.
- As in Layers I and II, the wideband signal is first split into 32 subbands with a polyphase filter.
- Additionally, each subband is transformed into **18 spectral coefficients** by **MDCT** (modified discrete cosine transform) for a maximum of 576 coefficients each representing a bandwidth of 41.67 Hz at a 48-kHz sampling rate; time resolution is 24 ms.
- There are **three block modes**: in two modes, the outputs of all 32 filter banks are processed through the MDCT with equal block lengths; in a mixed mode, the two lower bands use long blocks and the upper 30 bands use short block.
The allocation control algorithm uses *dynamic quantization*.

**Huffman** and **run length entropy** coding exploit the **statistical properties** of the audio signal to achieve data compression.

The data rate from frame to frame can vary in Layer III; this can be used for **variable rate recording**.

The number of bits per frame is variable, but has a **constant long-term average**.

Layer III claims transparency (compared to a 1.4 Mbps CD recording) at **192 kbps**.
Figure 10.23  ISO/MPEG-1 Layer III audio encoder and decoder. A. Layer III encoder (single-channel mode). B. Layer III two-channel decoder.
MPEG-2 Audio Standard

- The **MPEG-2** audio standard was designed for applications ranging from digital HDTV television transmission, to Internet downloading.
- The MPEG-2 audio standard uses the **same** encoding and decoding **principles** as MPEG-1. It is **backward compatible** to MPEG-1.
- MPEG-2 provides **multichannel** sound at sampling frequencies of 32, 44.1 and 48 kHz. Overall, **5.1 channels** can be successfully coded at rates from **384 to 640 kbps**.
- The MPEG-2 audio standard was approved by the MPEG committee in November 1994 and is specified in **ISO/IEC 13818-3**.
MPEG-2 Audio Standard

**Figure 10.25** The MPEG-2 audio standard adds mono/stereo coding at low sampling frequencies, and multichannel coding. The three MPEG-1 layers are supported.
MPEG-2 Audio Standard

- The multiple channels of MPEG-2 are matrixed to form compatible MPEG-1 left/right channels as well as other MPEG-2 channels.
- MPEG-2 uses intensity coding, crosstalk reduction, inter-channel prediction coding, and center channel phantom image coding to achieve a combined bit rate of **384 kbps**.
- MPEG-2 allows for audio bit rates **up to 1066 kbps**.
- MPEG-2 also specifies Layer I, II and III at **low sampling frequencies (LSF)** of 16, 22.05 and 24 kHz; this extension is **not backward compatible** to MPEG-1 coders.
- The video portion of MPEG-2 specifies **704 pixels by 480 lines** and **704 pixels by 576 lines** resolutions at data rates of **4 Mbps to 8 Mbps**, **16:9** aspect ratio and **interlaced fields**.
Although it retains compatibility, MPEG-4 is a departure from previous MPEG coding methods.

It is a family of tools and algorithms for interactive audio-visual coding, with an emphasis on very low bit rates that allow operation over the internet and other networks.

MPEG-4 specifies how to represent both natural and synthetic (computer-generated) audio and video material as objects, and defines how those objects are transmitted or stored and then composed to form complete scenes.

MPEG-4 uses a language called Binary Format for Scenes (BIFS) to describe and dynamically change scenes.

BIFS allows a great degree of interactivity between objects in a scene and the user.
MPEG-4

- The user may interact with the presentation, either by using local processing or by sending information back to the sender.
- MPEG-4 uses an object-oriented approach to code multimedia information for both mobile and stationary users.
- Version I of the MPEG-4 standard (ISO/IEC-14496) was finalized in October 1998.
- MPEG-1 and -2 describe ways to compress, transmit and store frame-based video and audio.
- MPEG-4 provides for control over individual data objects and the way they relate to each other, and a transport mechanism.
MPEG-4

- MPEG-4 audio consolidates high quality **music coding**, **speech coding**, **synthesized speech** and **computer music** in a common framework.
- It supports high-quality mono, stereo and multichannel signal. In particular, MPEG-4 aims at **very low bit rates**; it codes natural audio at bit rates from **2 to 64 kbps**.
- **Four** audio profiles are defined: the **Speech** Profile, the **Synthesis** Profile, the **Scalable** Profile, the **Main** Profile.
- The **Speech** Profile provides a **HVXC (Harmonic Vector eXcitation Coding)** very-low bit-rate parametric speech coder, a **CELP (Code Excited Linear Predictive)** narrowband/wideband speech coder, and a **Text-To-Speech interface**.
MPEG-4

- The **Synthesis** Profile provides score driven synthesis using **SAOL** (Structured Orchestra Language) and **wavetables** and a **Text-To-Speech Interface** to generate sound and speech at very low bit rates.

- The **Scalable** Profile provides scalable coding of speech and music for **networks**. The bit rates range from **6 kbps** to **24 kbps**, with bandwidth between 3.5 and 9 kHz.

- The **Main** Profile is a superset of all the other profiles, providing tools for natural and synthesized audio.

- The **MPEG-7** standard is entitled “Multimedia Content Description Interface”. Its aim is to allow efficient searches for multimedia content using standardized descriptions.

- Many search engines are designed to find textual information on the World Wide Web. **MPEG-7** would allow searches using **audiovisual description**.