Konstantinos Konstantinides

An Introduction to Super Audio CD and DVD-Audio

For many, the audio CD is considered the most successful new format in consumer electronics. Since its introduction in 1982, the CD has become the distribution medium of choice for both music and data. Using a sampling frequency of 44.1 kHz, analog audio is sampled and converted to pulse code modulated (PCM) samples. Each sample has 16 b of resolution and provides a signal-to-noise ratio in excess of 96 dB. Audio CDs use no compression; however, a cross-interleaved Reed-Solomon code (CIRC) allows audio CDs to satisfy the need for error detection and correction during playback.

In 1997, the introduction of DVD-Video provided not only superior video quality but also allowed consumers to experience and enjoy multichannel surround sound in the home. DVD movies do provide multichannel audio, but it is stored compressed using the lossy Dolby digital or MPEG-audio formats. The sampling rate for compressed audio in the DVD-Video format is 48 kHz.

The DVD-Audio specification allows for up to 24-b PCM data and uses the Meridian lossless packing (MLP) algorithm to provide up to six channels of high-quality, multichannel audio at sampling rates of up to 96 kHz for six channels or 192 kHz for two channels.

Super-audio CD (SACD), introduced in March 1999, integrates a variety of new technologies, such as the hybrid disc, direct stream digital (DSD), and direct stream transfer coding. Unlike DVD-Audio, SACD does not use PCM encoding. The idea behind DSD is to remove the front-end and back-end decimation and interpolation filters and directly record 1-b data, sampled at approximately 2,822.4 kHz. At the end, DSD signals yield a dynamic range greater than 120 dB. Table 1 summarizes the key characteristics of audio CDs, SACD, and DVD-Audio.

Table 1. Main parameters in DVD-Audio, SACD, and audio CD recordings.

Parameter	DVD-Audio	SACD	CD
Audio coding	16-, 20-, or 24-b LPCM	1-b DSD	16-b LPCM
Sampling rate	44.1, 48, 88.2, 96, 176.4, or 192 kHz	2,822.4 kHz	44.1 kHz
Channels	1-6	2-6	2
Compression	Yes (MLP)	Yes (DST)	None
Content protection	Yes	Yes	No
Playback time	62-843 min*	70-80 min	74 min
Frequency response	DC-96 kHz	DC-100 kHz	DC-20 kHz
Dynamic range	Up to 144 dB	Over 120 dB	96 dB

*For 62 min, we assume f_s = 96 kHz, 20-b samples, and five channels. For 843 min, we assume f_s = 44.1 kHz, 16-b samples, and one channel.

The purpose of the next two articles is to present and highlight the latest developments in consumer audio and specifically in DVD-Audio and SACD.

The first article by B.H. Suzuki, N. Fuchigami, and J.R. Stuart, from JVC and the Meridian Group, presents an overview of the DVD-Audio specification. The authors place special emphasis into the design and implementation of the MLP coder, especially as it relates to the coding of multichannel audio.

The second article by E. Janssen and D. Reefman, from Philips Research Laboratories, introduces SACD. Here, the authors place special emphasis on the new challenges in digital signal processing using DSD streams.

I hope that readers will enjoy reading both articles and that this Forum will stimulate a healthy and interesting exchange of ideas in future columns.

1053-5888/03/$17.00©2003IEEE

Bike H. Kuzuki, Norihiko Fuchigami, and J. Robert Stuart

DVD-Audio Specifications

Since 1983 audio CD (CD-DA) has been the primary carrier for high-quality stereo music. In the intervening years optical media and signal processing technologies have greatly advanced and enabled the development of higher-density discs such as DVD (digital versatile disc). DVD offers more than six times the capacity of CD.

The DVD Forum [1] has developed a family of specifications to accommodate both prerecorded (DVD-ROM) and recordable (DVD-R, RW, and RAM) applications. DVD-Video [2] and DVD-Audio [3] are both applications of DVD-ROM.

DVD-Video provides a carrier for high-quality video assets, and its specification reflects a balance between both data capacity and data rate for the sound and picture components of a movie. By contrast, DVD-Audio uses the capacity of DVD-ROM to provide a music carrier with supreme audio quality and better sound-field reproduction. The DVD-Audio Specification was drawn up in cooperation with the music industry, which demanded very high-quality audio in stereo and multichannel, multimedia functions and security.

The lossless coding scheme adopted in DVD-Audio is a key technology that allows high-quality multichannel music to be available on the huge, but still limited, capacity DVD disc. This article mainly focuses on MLP Lossless [4] since it will be of most interest for readers of this magazine. (MLP stands for Meridian Lossless Packing and is a trademark of Dolby Laboratories Inc.)

Features of DVD-Audio

Working Group 4 (WG4) of the DVD Forum developed the DVD-Audio Specification, and during this effort they received 15 key requirements from the international music industry, represented by ISC (International Steering Committee). The main features of DVD-Audio, which closely reflect these requirements, are summarized below.

High-Quality Audio

The first requirement from artists, the music industry, and listeners is to provide sound quality that is superior to CD and as close to the original master as possible. DVD-Audio meets this requirement flexibly by offering several options based on linear pulse code modulation (LPCM). By including options that increase both the sampling frequency and quantization word length, DVD-Audio offers four times the frequency range and 256 times the resolution of CD. DVD-Audio supports:

- up to 192 kHz/24-b for two-channel stereo
- up to 96 kHz/24-b for multichannel (up to six channels).

In some combinations the audio can be stored on the disc as raw LPCM. However, for efficient recording of the huge data resulting from high-resolution multichannel, the lossless compression must be used. As the name implies, MLP compresses data without any loss and decompresses it into data that is exactly identical to the original LPCM master. MLP enables any of the audio formats permitted on DVD-Audio to play for longer than 74 min (on a single layer 4.7 GB disc).

DVD-Audio also supports the sampling frequencies 44.1, 88.2, and 176.4 kHz for making good use of assets that may be issued on CD.

DVD-Audio offers great flexibly to producers by supporting not only the high-bit/high-sampling parameters but also more conventional coding such as 44.1 kHz/16-b/two channel. Titles can be made in a wide variety of formats.

Attractive Multichannel Surround

Although DVD-Audio provides for super high-quality, two-channel audio, perhaps its most attractive feature is the support for high-quality multichannel sound. The superb quality of sound field that can be achieved through exerting the whole bit rate of DVD for audio data is clearly superior to that obtained using the lossy-compressed surround sound options of DVD-Video.

DVD-Audio also offers methods to enjoy multichannel music in two-channel listening environments by providing backward compatibility for the listener who has a two-channel speaker system or who may be using headphones.

A Variety of Value-Added Contents

Another important requirement for this new generation medium is the provision of a variety of value-added content to supplement the high-quality audio experience. DVD-Audio offers attractive value-added content: still pictures that can be enjoyed in conjunction with the high-quality audio, DVD-video compatible content suitable for video clips such as interviews, and visual menu for selecting songs or additional information. These features were developed by transplanting or expanding on the multimedia capability of DVD-Video.

Compatibility with DVD-Video

As will be explained later, DVD-Video compatible content is directly compatible with the DVD-Video Specification, which means that such content can be played on existing DVD-Video players.

Additionally, since functions or operations for visual content such as still pictures and menus are almost the same as those of DVD-Video, users of DVD-Video easily become familiar with DVD-Audio. Thus, DVD-Audio turns the compatibility with DVD-Video to good account for the propagation of the format.

Supporting Copyright Management System

Illegal copying damages the rights of content owners and so new generation carriers must provide a credible protection system against illegal copying. DVD-Video compatible content makes use of the content scramble system (CSS) encryption method [5] that is used to protect DVD-Video movies. For the high-quality audio, DVD-Audio adopts a newer copy protection system: content protection for prerecorded media (CPPM) [6]. Additionally, audio data on a DVD-Audio disc may contain an "audio watermark" that enables downstream control of illegal copying.

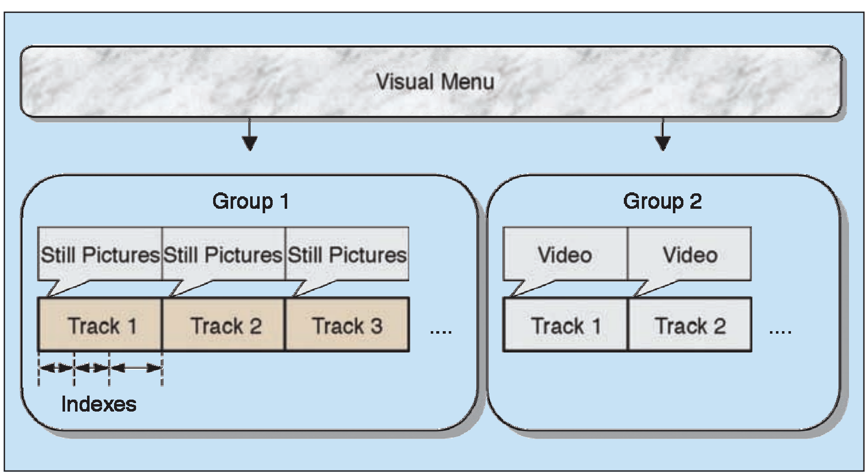
A User's View of DVD-Audio Content

As with CD-DA, DVD-Audio defines the "track" as the unit corresponding to one tune. So as to efficiently manage the large number of tracks that are possible on this large capacity disc, a higher layer, "group," is introduced. A disc can contain 1 to 9 groups, each of which in turn can contain up to 99 tracks. Users can access a target tune by selecting a group and a track.

Furthermore, a track may be segmented into multiple "index" portions. As with CD-DA, index 0, and 1 to 99 can be specified.

Besides audio data, a DVD-Audio disc can include still pictures, DVD-Video compatible content, and so on. DVD-Video compatible contents are also represented as tracks at the user level. A "visual menu" may also be included if required. Figure 1 shows the user's view of DVD-Audio.

For distinction, we call tracks for high-quality audio contents audio tracks and those for DVD-Video compatible contents video tracks.



▲ *1. The user's view of DVD-Audio.*

Internal Data Structure of DVD-Audio

The internal data structure of DVD-Audio is somewhat different from that implied by the user's view. Because it can store a variety of content, DVD-Audio has a fairly complex data structure, especially when compared to the rather simpler structures of CD-DA. Figure 2 shows how data is allocated on a disc from inside to outside. In the figure, the dotted-line boxes indicate optional content.

Since DVD-Audio is an application format of DVD-ROM (DVD Specifications for Read-Only disc), the DVD-Audio disc complies with Part 1: Physical Specifications of DVD-ROM [7] and the data is stored as files in compliance with Part 2: File System Specifications [8].

Data for DVD-Audio is recorded as several files in the AUDIO_TS directory, in compliance with Part 4: Audio Specifications, the body of DVD-Audio format.

Audio manager (AMG) manages the total presentation. The audio data is recorded in audio objects (AOB) in an audio title set (ATS). An AOB is a program stream (PS) complying with MPEG-2 System (ISO/IEC 13818-1) [9] and contains either a packetized LPCM stream or a packetized MLP stream. Each pack of the MPEG PS on a DVD disc is stored in a logical sector of 2,048 B, where roughly 2,000 B are used for the main data (audio data in the case of AOB) and the rest is for headers.

Still pictures to be presented with audio data are recorded in an audio still video set (ASVS) separately from the audio data. Each still picture, called audio still video (ASV), is also a MPEG PS containing one MPEG-intra picture.

DVD-Video compatible content, if any, is recorded in the "VIDEO_TS" directory in compliance with Part 3: Video Specifications. The video-with-audio data for the video track is recorded as a video object (VOB) in a video title set (VTS). This video data is also managed (i.e., referred to) by the AMG for playback on DVD-Audio players. VMG is the video manager information used by DVD-Video players.

Audio Specifications

This section describes the general audio specifications for audio data in the ATS.

Audio Coding Schemes and Audio Parameters

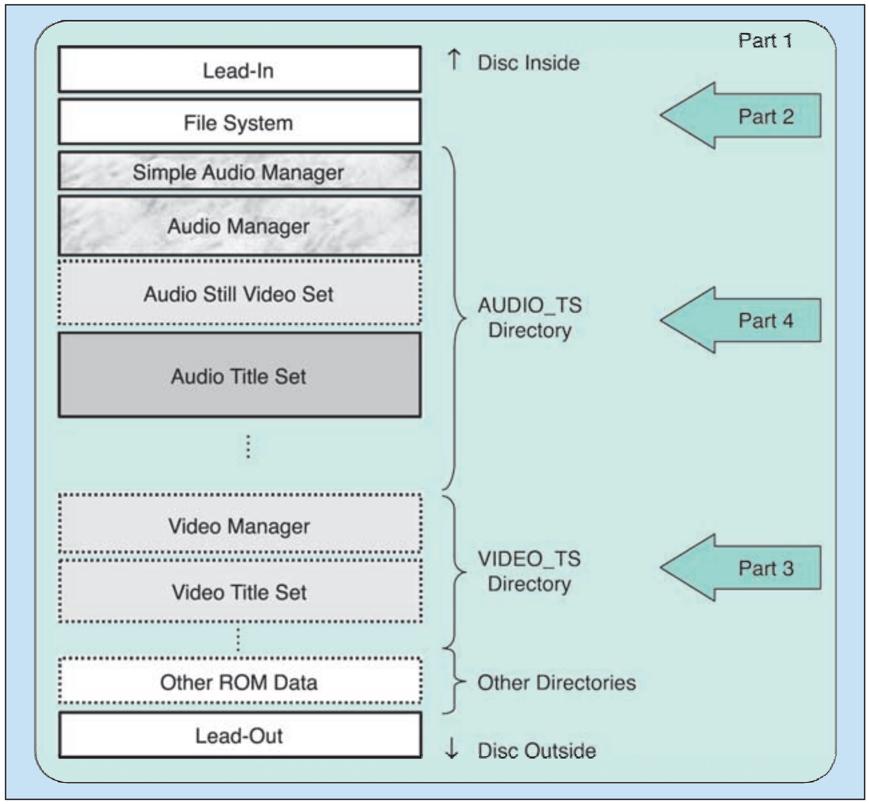
DVD-Audio supports two audio coding schemes: linear PCM (LPCM) and MLP (MLP is also called "packed PCM" in the format book).

Audio data in an audio track is coded in one of these two schemes and all DVD-Audio players support their playback. MLP, which is described later, is a lossless compression system for LPCM audio data, and it is used for two key purposes:

▲ to extend playing time: lossless compression extends playing time, by typically reducing the data on the disc by a factor of 2

▲ to enable the highest quality multichannel data: the maximum bit rate allowed for audio data in DVD-Audio is 9.6 Mb/s. Without compression, the highest quality audio that could be accommodated is 96 kHz/20-b/five-channel LPCM, which runs at 9.6 Mb/s. MLP also reduces the maximum data rate on the disc and meets the highest quality requirement of 96 kHz/24-b/six channel, which would otherwise require a carrier capable of 13.824 Mb/s in native LPCM format.

Table 1 shows the combinations of sampling frequency, quantization word length, and number of channels (one to six channels) supported on DVD-Audio.



▲ 2. Data structure of a DVD-Audio disc.

Table 1. Supported parameter configurations for audio tracks.

For Both LPCM, MLP — For Only MLP

Sampling Frequency	Quantization Word Length	1ch	2ch	3ch	4ch	5ch	6ch
48 kHz or 44.1 kHz	16 bit						
	20 bit						
	24 bit						
96 kHz or 88.2 kHz	16 bit						
	20 bit						
	24 bit						
192 kHz or 176.4 kHz	16 bit						
	20 bit						
	24 bit						

DVD-Audio supports up to quadruple sampling (192 kHz, 176.4 kHz) for two-channel stereo and up to double sampling (96 kHz, 88.2 kHz) for multichannel.

As 44.1 kHz sampling and its multiples are also supported, master materials for CD can be transferred to DVD-Audio discs without a sampling conversion process.

Channel Assignment

DVD-Audio supports up to six full bandwidth audio channels. The channel assignment follows the layout for movie (video), i.e., front three channels, rear-surround two channels (or one channel) plus LFE (low frequency effect). This generally means that users are not required to change the speaker layout for playback of DVD-Video and DVD-Audio.

This layout is basically assumed to comply with the ITU-R BS.775-1 recommendation [10]; however, it is important to note that any exact layout or relation of playback speakers (distance, azimuth, etc.) are not defined in the DVD-Audio Specification.

Multichannel and Stereo Compatible Presentation Methods

Multichannel audio is a primary feature of DVD-Audio; however, a number of users may have two-channel stereo equipment at home. DVD-Audio provides two methods for multichannel and stereo compatible presentation.

Down-Mix Method

If a track contains only multichannel audio data then two-channel stereo is derived from the multichannel data by a downmix process. The downmix coefficients are specified during the process of authoring.

The actual processing is slightly different depending on the coding scheme used.

▲ In the case of linear PCM, the player performs the downmix to two channel according to down-mix coefficients stored on the disc. Different tracks basically can have different coefficients.

▲ In the case of MLP, the downmix is performed during the authoring process (MLP encoding) according to specified coefficients, and the downmixed substream (called "L_0, R_0") is stored on the disc together with another substream containing the rest of channels (e.g., C, LFE, Ls, Rs). For two-channel stereo reproduction, players directly decode and play back the L_0, R_0 substream.

Audio Selection

In case of audio selection, a track logically contains separate multichannel and two-channel stereo data: for example it could be as "audio #1: multichannel" and "audio #2: two-channel stereo."

The coding schemes for audio #1 and #2 can be any combination of LPCM or MLP. Obviously, in terms of data efficiency, using lossless compression (MLP) for both selections would be the best.

When playing back a track with audio selection, players automatically select the appropriate version (i.e., multichannel for multichannel players or two channel for two-channel stereo players), and it is also possible for users to switch between audio #1 and #2 with the "audio" key on a remote control unit (RCU).

Using this method, content providers can include dedicated two-channel mixes which may have used sound effects processing such as reverberation, equalization, and delay in addition to downmix in their production. However, providing two-channel compatibility by audio selection does require more data space, since the disc must hold both the multichannel and two-channel mixes in separate locations.

In the case of DVD-Video, multiple audio streams (if any) are multiplexed with the video stream. In that case, the sum of bit rates for the video stream and all audio streams must not exceed 9.8 Mb/s (the maximum bit rate of DVD-Video). By contrast, in the case of DVD-Audio, audio selections are put on the disc as separate audio data (AOB) without multiplexing. Therefore, each AOB can adopt a bit-rate up to 9.6 Mb/s. From the viewpoint of audio data recording, this scheme is equivalent to storing audio #1 and #2 as two separate tracks.

MLP Lossless

MLP for DVD-Audio is a tailored subset of the MLP algorithm that complies with the MPEG program stream system and DVD specific features. This section describes the MLP system [11].

Lossless Compression

Unlike perceptual or lossy data reduction, lossless coding does not alter the final decoded signal in any way but merely "packs" the audio data more efficiently into a smaller data rate.

Audio information that is of interest to the human listener contains some redundancy. On music signals, the information content varies with time and the input channel information capacity is rarely fully exercised.

The aim of lossless compression is to reduce incoming audio to a data rate that closely reflects the inherent information content plus a minimum overhead.

An important insight then is that the coded output of a lossless compressor will have a variable data rate on normal audio content.

While normal music may be compressed by a factor of 2, we reasonably experience wider variations in compressed rate. There are also pathological signals: for example, silence or near-silence will compress greatly and signals that are nearly random will not. Indeed, should a section of channel data appear to be truly random, then no compression is possible. Fortunately it turns out that real acoustic signals tend not to provide full-scale white noise in all channels for any significant duration!

Previously lossless audio data compression systems have been optimized for reducing average data rate (i.e., minimizing compressed file size).

Additionally, it is important to reduce the instantaneous peak data rate for optimum results at high sampling rates such as 96 kHz or 192 kHz for data-limited, disc-based applications like DVD-Audio.

MLP tackles this by attempting to maximize the compression at all times using this set of techniques:

▲ looking for "dead air," channels that do not exercise all the available word length
▲ channels that do not use the available bandwidth
▲ removing interchannel correlations
▲ efficiently coding the residual information
▲ smoothing coded information by buffering.

How Does It Work?

MLP coding is based on established concepts; however, there are some important novel techniques used in this system, including:

▲ lossless processing
▲ lossless matrixing
▲ lossless use of infinite impulse response (IIR) filters
▲ managed first-in, first-out (FIFO) buffering across transmission
▲ decoder lossless self-check
▲ operation on heterogeneous channel sampling frequency.

These methods are described next, in the context of the encoder.

MLP Encoder

The MLP encoder core is illustrated in Figure 3. The steps for encoding blocks of data are the following:

▲ 1) Incoming channels may be remapped to optimize the use of substreams (described later).
▲ 2) Each channel is shifted to recover unused capacity (e.g., less than 24-b precision or less than full scale).
▲ 3) A lossless matrix technique optimizes the channel use by reducing interchannel correlations.
▲ 4) The signal in each channel is decorrelated using a separate predictor for each channel.
▲ 5) The decorrelated audio is further optimized using entropy coding.
▲ 6) Each substream is buffered using a FIFO memory system to smooth the encoded data rate.
▲ 7) Multiple data substreams are interleaved.
▲ 8) The stream is packetized for fixed or variable data rate and for the target carrier.

Lossless Matrix

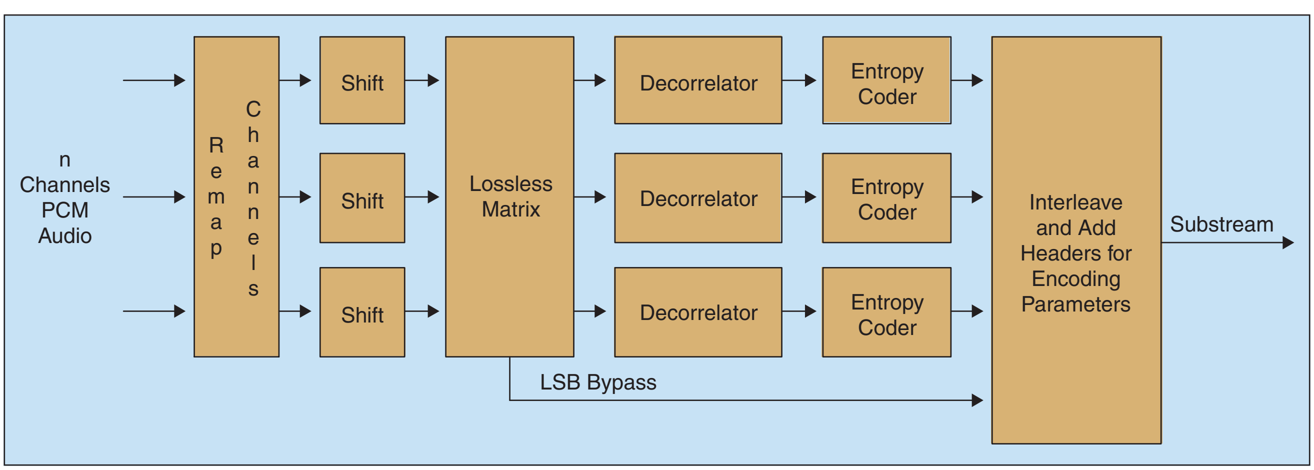
A multichannel audio mix will usually share some common information between channels.

On occasion the correlations will be weak, but there are other cases where the correlations can be high. Examples include multitrack recordings where a mix-down to the delivered channels may pan signals between channels and thus place common information in some channels.

There are also specific examples where high interchannel correlations occur, including:

▲ mono presented as dual-mono with identical left and right (common in "talking book" or archive recordings)
▲ derived surround signals based on left minus right
▲ multichannel speaker feeds resulting from a hierarchical upscale
▲ multichannel speaker feeds resulting from an ambisonic decode from B-format WXYZ.

The MLP encoder uses a matrix that allows the encoder to reduce correlations, thereby concentrating larger amplitude signals in fewer



▲ 3. Block diagram of the lossless encoder core.

channels. A trivial (though important) example would be the tendency of the matrix process to rotate a stereo mix from left/right to sum/difference. In general the encoded data rate is minimized by reducing the commonality between channels.

However, conventional matrixing is not lossless: a conventional inverse matrix reconstructs the original signals but with rounding errors.

The MLP encoder decomposes the general matrix into a cascade of affine transformations. Each affine transformation modifies just one channel by adding a quantized linear combination of the other channels; see Figure 4. For example, if the encoder subtracts a particular linear combination, then the decoder must add it back. The quantizers Q in Figure 4 ensure constant input-output wordwidth and lossless operation on different computing platforms.

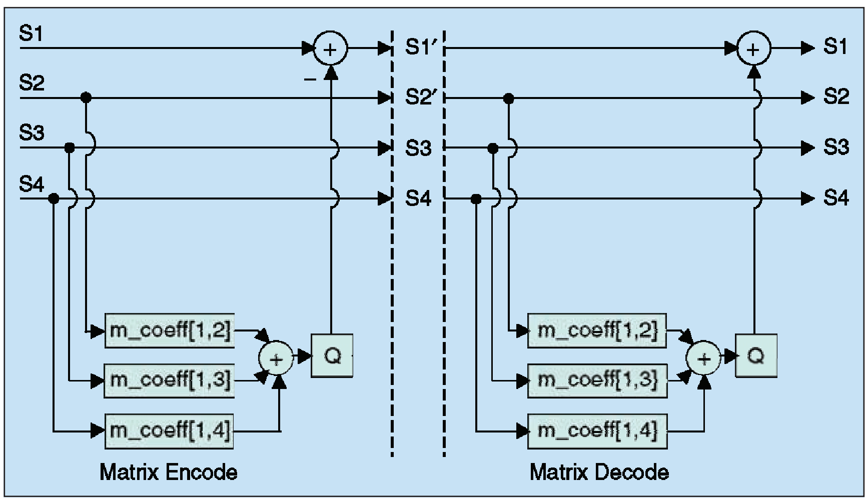
Prediction

If the values of future audio samples can be estimated, then it is only necessary to transmit the rules of prediction along with the difference between the estimated and actual signals. This is the function of the decorrelator (optimal coding shows no correlation between the currently transmitted difference signal and its previous values). It is useful to consider how prediction operates in the frequency (Shannon) domain.

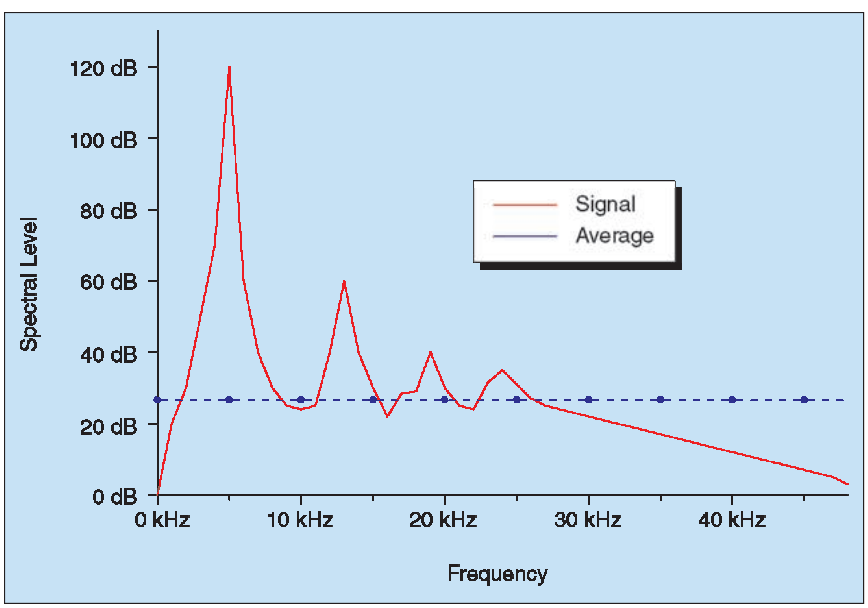
Figure 5 shows the short-term spectrum of a music excerpt. If this spectrum was flat, a linear prediction filter could make no gains. However, it is far from flat, so a decorrelator can make significant gains by flattening it, ideally leaving a transmitted difference signal with a flat spectrum, essentially being white noise. The Gerzon/Craven theorems [20] show that the level of the optimally decorrelated signal is given by the average of the original signal spectrum when plotted as decibels versus linear frequency. As illustrated in Figure 5, this decibel average can have significantly less power than the original signal, hence the reduction in data rate. In fact this power reduction represents the information content of the signal as defined by Shannon [13].

In practice, the degree to which any section of music data can be "whitened" depends on the content and the complexity allowed in the prediction filter.

Infinite complexity could theoretically achieve a prediction at the entropy level shown in Figure 5; however, all the coefficients which define this decorrelator would then need to be transmitted to the decoder (as well as the residual signal) to reconstruct (recorrelate) the signal. There is therefore a need to obtain a good balance between predictor complexity and performance.



▲ *4. A single lossless matrix encode and decode.*



▲ *5. Spectra of a signal and its average level.*

FIR and IIR Prediction

Most previous lossless compression schemes use finite impulse response (FIR) prediction filters and can achieve creditable reduction of data rate on conventional CD-type material [14]-[16]. However, we pointed out in [17]-[19] that IIR filters have advantages in some situations, particularly:

▲ cases where control of peak data rate is important

▲ cases where the input spectrum exhibits an extremely wide dynamic range.

The ARA proposal [12] pointed out the particularly increased likelihood of wide dynamic range in the spectrum of audio sampled at higher rates such as 96 or 192 kHz. Spectral energy at high frequencies is normally quite low and may be further attenuated by microphone response or air absorption.

The ARA also indicated the desirability that a music provider should have the freedom to control lossless data rate by adjusting supersonic filtering during mastering. A powerful lossless compression system will require the use of FIR and IIR prediction.

Track 6 of the CD "Hello, I Must be Going!" by Phil Collins shows an example that is quite difficult to compress. The original signal spectrum in Figure 6 includes a percussion instrument with an unusually extended treble response. An eighth-order FIR filter is able to flatten the major portion of the spectrum. However, it is completely unable to deal with the drop above 20 kHz caused by the anti-alias filter. A fourth-order denominator IIR filter is able to do this very effectively as shown.

In this case the improvement in compression is small, as there is only 2 kHz of underutilized spectrum between the 20 kHz cut-off and the Nyquist frequency of 22.05 kHz. IIR filtering gives a bigger improvement if filtering leaves a larger region of the spectrum unoccupied, for example if audio is sampled at 96 kHz but a filter or other roll-off is in place at say 30 or 35 kHz (see [19]) (as may happen with a microphone).

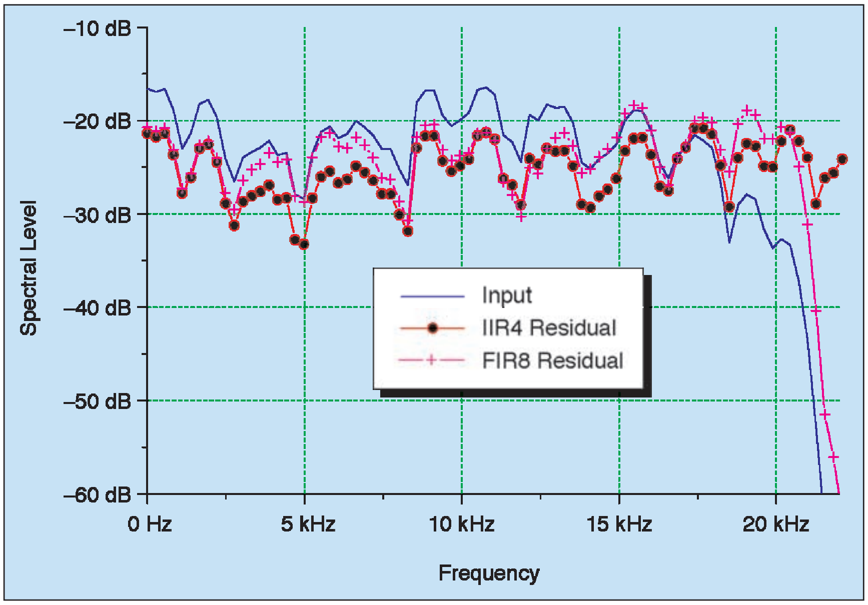
Lossless Prediction in MLP

The MLP encoder uses a separate predictor for each encoded channel. The encoder is free to select IIR or FIR filters up to eighth order from a wide palette. These extensive options ensure that good data reduction can be provided on as many types of audio as possible.

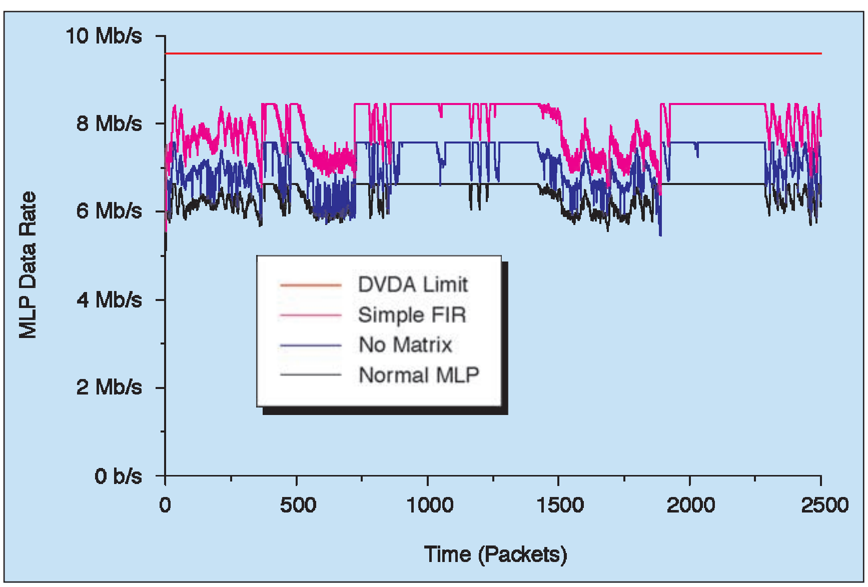
The effectiveness of the encoder tactics described so far can be seen in Figure 7, which graphs the data rate through a 30-s, 96 kHz/24-b/six-channel orchestral excerpt.

The lowest curve in Figure 7 is the data rate for the normal MLP encoder; the flat-topped sections will be explained later.

The middle curve shows the impact of switching off the lossless matrix and illustrates that in this case a significant improvement in coding



▲ *6. Spectra for a signal excerpt and the residuals using eighth-order FIR and fourth-order IIR predictors.*



▲ *7. Resulting data rate in MLP encodes showing the benefit of the encoder stages.*

ratio was obtained by removing interchannel correlations. The upper curve shows the further reduced effectiveness by constraining the predictor choices to a simple FIR. The top line shows the 9.6 Mb/s data-rate limit for DVD-Audio.

The input data rate is 13.824 Mb/s, so in this example the options of IIR and lossless matrixing improved the coding ratio from 1.64:1 to 2.08:1.

Entropy Coding

Once the cross-channel and intersample correlations have been removed, it remains to encode the individual samples of the decorrelated signal as efficiently as possible. "Entropy coding" is the general name given to this process, with its aim being to match the coding of each value to the probability that it occurs. Infrequent values are coded to a large number of bits, but this is more than compensated by coding frequent values to a small number of bits.

Audio signals tend to be peaky, and so linear coding is inefficient. For example, in PCM one has to allocate enough bits to describe the highest peak, and the most significant bits (MSBs) will be used infrequently. Audio signals often have a Laplacian distribution (c.f., [14]-[16]), that is, the histogram is a two-sided decaying exponential; this appears to be true even after decorrelation. The MLP encoder may choose from a number of entropy coding methods.

Buffering

We have explained that while normal audio signals can be well predicted there will be occasional fragments like sibilants, synthesized noise, or percussive events that have high entropy.

MLP uses a particular form of stream buffering that can reduce the variations in transmitted data rate, absorbing transients that are hard to compress.

FIFO memory buffers are used in the encoder and decoder (see Figures 9 and 10). These buffers are configured to give a constant notional delay across encode and decode. This overall delay is small, typically of the order of 75 ms. To allow rapid startup or cueing, the FIFO management minimizes the part of the delay due to the decoder buffer. So, this buffer is normally empty and fills only ahead of sections with high instantaneous data rate.

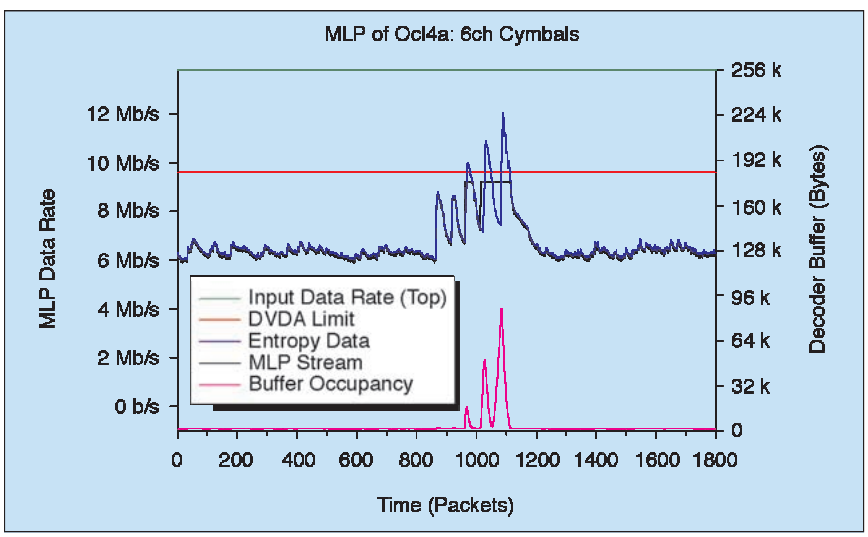
During these sections, the decoder's buffer empties and is thus able to deliver data to the decoder core at a higher rate than the transmission channel is able to provide. In the context of a disc, this strategy has the effect of moving excess data away from the stress peaks, to a preceding quieter passage.

The encoder can use the buffering for a number of purposes, e.g.,

▲ keeping the data-rate below a preset (format) limit

▲ minimizing the peak data rate over an encoded section.

Figure 8 shows how hard-to-compress signals can be squeezed below a preset format limit. This 30-s, 96 kHz/24-b recording features closely recorded cymbals in six channels. At the crescendo this signal is virtually random and the underlying compressed data rate is 12.03 Mb/s. Buffering allows the MLP encoder to hold the transmitted data rate below 9.2 Mb/s by filling the decoder buffer to a short-term maximum of 86 KB (bottom curve).



▲ *8. Buffering allows a difficult passage to remain below a hard format limit.*

Use of Substreams

The MLP stream contains a hierarchical structure of substreams. Incoming channels can be matrixed into two (or more) substreams. This method allows simpler decoders to access a subset of the overall signal.

This substream principle is illustrated for the encoder in Figure 9 and the decoder in Figure 10; note that each substream is separately buffered.

MLP Decoder

The MLP decoder unwinds each encoder process in reverse order.

The decoder is relatively low complexity. A decoder capable of extracting a two-channel stream at 192 kHz requires approximately 27 MIPS (mega instruction per second), while 40 MIPS will be required to decode six channels at 96 kHz.

Two-Channel Downmix

It is often useful to provide a means for accessing high-resolution multichannel audio streams on two-channel playback devices. In an application such as DVD-Audio, the content provider can place separate multi- and two-channel streams on the disc as described earlier. However, to do this requires separate mix, mastering, and authoring processes and uses disc capacity.

In cases where only one multichannel stream is available, then there are very few options at replay; one is to use either a fixed or guided downmix. However, to create such a downmix it is first necessary to decode the full multichannel signal; this contravenes the desirable principle that decoder complexity should decrease with functionality.

Performing Mix-Down in the Lossless Encoder

MLP provides an elegant and unique solution. The encoder combines lossless matrixing with the use of two substreams in such a way as to optimally encode *both* the L_0, R_0 downmix *and* the multichannel version. This method can be explained by referring to Figures 9 and 10.

Downmix instructions are fed to the matrix 1 to determine some coefficients for the lossless matrices. The matrices then perform a rotation such that the two channels on substream 0 decode to the desired stereo mix and combine with substream 1 to provide full multichannel.

Because the two-channel downmix is a linear combination of the multichannel mix, then strictly, no new information has been added. In the example shown in Figure 9 there are still only six independent channels in the encoded stream. So, theoretically, the addition of the two-channel version should require only a modest increase in overall data rate (typically 1 b per sample, e.g., 96 kb/s at 96 kHz).

The advantages of this method are considerable.

▲ The quality of the mix-down is guaranteed. The producer can listen to it at the encoding stage and the lossless method delivers it bit-accurate to the end user.

▲ A two-channel-only playback device does not need to decode the multichannel stream and then perform mix-down. Instead, the lossless decoder only need decode substream 0.

▲ A more complex decoder may access both the two-channel and multichannel versions losslessly.

▲ The downmix coefficients do not have to be constant for a whole track, but can be varied under artistic control.

How Much Compression?

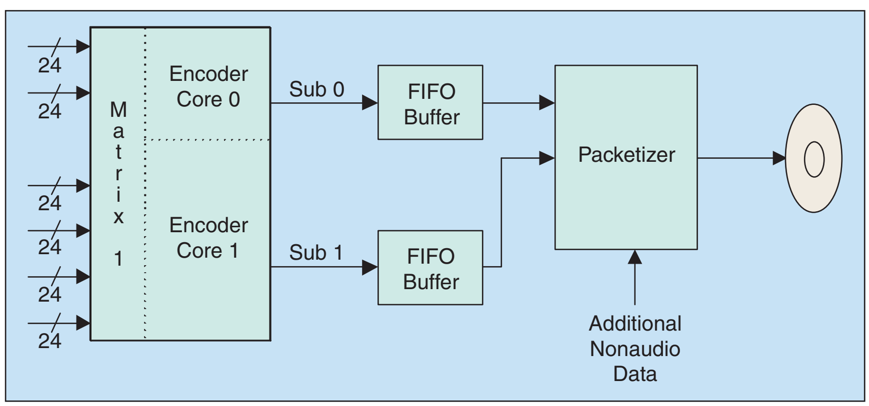
In specifying a lossy system, the critical compression measure is the final bit rate for a given perceptual quality, and this is independent of the input wordwidth. With lossless compression, increases in incoming precision (i.e., additional LSBs on the input) must be losslessly reproduced. However, these LSBs typically contain little redundancy and so contribute directly to the transmitted data rate. Therefore we tend to quote the *saving* of data rate, as this measure is relatively independent of incoming precision. The saving of data rate (in bits per original sample) is indicated in Table 2.

At 44.1 or 48 kHz, the peak data rate can almost always be reduced by at least 4 b/sample, i.e., 16-b audio can be losslessly compressed to fit into a 12-b channel.

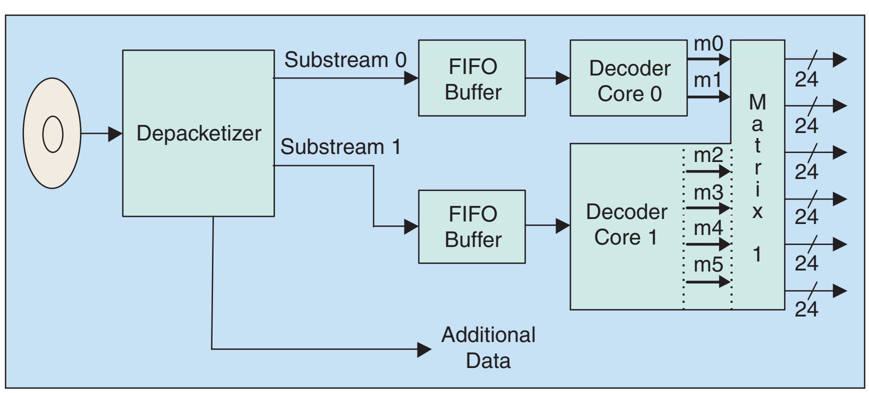
At 96 kHz, the peak data rate can similarly be reduced by 8 b/sample, i.e., 24-b audio can be compressed to 16 b and 16-b/96 kHz audio can be losslessly compressed to fit into an 8-b channel.

The important parameter for transmission applications is the reduction of the peak rate and in the case of DVD-Audio peak rate is a key parameter, because the encoded stream must always operate below the audio buffer data-rate limit of 9.6 Mb/s.

The average number in Table 2 indicates the degree of compression that could be obtained when using



▲ 9. Illustrating two substreams in encoding.



▲ 10. Decoding two substreams.

Table 2. Data-rate reduction.		
Sampling kHz	Data-Rate Reduction: b/sample/channel	
	Peak	Average
48	4	5 to 11
96	8	9 to 13
192	9	9 to 14

MLP in an archive, mastering, or editing environment. For example, a peak data rate reduction of 8 b/sample means that a 96 kHz/24-b channel can be carried on the disc with a rate equal to that of a 24 – 8 = 16 b LPCM channel. However the space used on the disc is estimated by the average saving, in this case the residual will be 24 – 11 = 13 b/channel. Consider that an 11-b saving represents a compression ratio of 1.85:1 with 24-b material, whereas the same saving compresses 16-b audio by 3.2:1!

Playing Time on DVD-Audio

DVD audio holds approximately 4.7 GB of data and has a maximum data transfer rate of 9.6 Mb/s for an audio stream. Six channels of 96 kHz/24-b LPCM audio have a data rate of 13.824 Mb/s, which is well in excess of 9.6 Mb/s.

Also, at 13.824 Mb/s, the data capacity of the disc would be used up in approximately 45 min.

So, lossless compression is needed to reduce the data on the disc to extend playing time to the industry norm of 74 min *and* to guarantee a *minimum* reduction of 31% in instantaneous data rate.

MLP meets this requirement with a sophisticated encoder, a simple decoder, and a specific subset of features limited to two substreams and six channels.

Here are some examples of playing times that can be obtained:

- ▲ 5.1 channels, 96 kHz/24-b: 100 min
- ▲ 6 channels, 96 kHz/24-b: 86 min
- ▲ 2 channels, 96 kHz/24-b: 4 h
- ▲ 2 channels, 192 kHz/24-b: 2 h
- ▲ 2 channels, 44.1 kHz/16-b: 12 h
- ▲ 1 channel, 44.1 kHz/16-b: 25 h (talking book).

Other Features

Still Picture

The audio track can be accompanied by multiple still pictures (ASVs) and these can be presented in "slideshow" or "browsable" modes. In slideshow mode ASVs are presented at predetermined timings. In browsable mode users can change (or browse) the pictures freely. Browsable pictures are a unique feature of DVD-Audio which is not supported in DVD-Video.

A set of ASV(s) to be presented together with one or more audio track(s) is called an ASV unit (ASVU). Because the audio may use the full data-rate capability of DVD-Audio, ASVUs are preloaded into a memory buffer in the player in advance of the audio presentation. Since each picture (ASV) in the ASVU is read from the memory, changing the ASVs does not affect the audio presentation at all. Moreover, since the display timing of each ASV can be flexibly controlled, the browsable functionality becomes possible.

The maximum size of one ASVU is 2 MB (2097152 B), which would allow for storing typically 15 to 20 MPEG-intra pictures at 720*480 pixels. The lower the resolution of the stored images, the more pictures can be contained in an ASVU.

Video Track

A video track or DVD-Video compatible content is used mainly for a music video clip accompanied by moving pictures. It is optionally recorded in the VIDEO_TS directory in compliance with a restricted subset of Part 3, Video Specifications. The restrictions forbid the use of some DVD-Video functions such as multistory (story branching), region control, and parental control, which are not necessary for an audio application.

Video tracks can also be used to contain the same tunes as the audio tracks (in lower quality) so that the disc can be compatible with DVD-Video players.

Visual Menu

The visual menu allows content in a track or group of a DVD-Audio disc to be accessed through a menu picture on the screen. The menu can also give access to other information such as liner notes, discography, and lyrics. The visual menu of DVD-Audio provides capabilities which are quite like those of the title menu in DVD-Video.

Copyright Management System

DVD-Audio accommodates a copyright management system called CPPM. CPPM technology together with copyright management information on the disc protects the recorded assets and aims to control any copying.

CPPM and CSS

CPPM uses encryption to protect content in the AUDIO_TS directory. CSS, the copy protection system for DVD-Video, is used for contents in the VIDEO_TS directory. Note that CPPM and CSS are systems developed by organizations outside of the DVD Forum. Further information may be found in [5] and [6].

Encryption of Contents

A primary function of CPPM or CSS is to encrypt the content. Since encrypted data cannot be decrypted without the correct key, a direct copy of the encrypted data (very likely on a PC) has no meaning. As the decryption key is obtained only through official contract and managed securely, the chance of key hacking is extremely low.

Furthermore, as a countermeasure against the theft or hacking of the keys, CPPM provides a method to disable any stolen key; that is, discs manufactured after hacking can be protected from being decrypted illegally with that key.

Audio Watermark (Watermark)

An audio track can optionally include an audio watermark [22], [6]. The watermark can carry copy control information (WM-CCI) and other information about the copyright holder. The watermark is primarily intended for the following:

- ▲ Tracing illegal copies. Even if audio data is captured from a player's analog output and illegally distributed in some way, the watermark can identify the original rights holder, etc.
- ▲ Stopping playback of some pirated discs. Under the CPPM contract, audio data must be CPPM-encrypted when a WM-CCI is included that restricts copying. Therefore, a non-CPPM disc which contains such a watermark will be an illegal disc. Players supporting CPPM detect the watermark and stop playing back such discs.
- ▲ Control of copying via an analog interface. The watermark is the only way to confirm the copyright of data when it is transmitted via an analog interface. If the destination recording device complies with the contract for the same watermark, the device will detect the watermark to control the copying.

Conclusions

DVD-Audio offers a broad and blank canvas for music producers and artists with flexible high-quality audio configurations, multimedia features, and compatibility with DVD-Video platforms. One can produce multichannel music sampled at 96 kHz/24-b with the aid of the MLP Lossless and super-wide-range stereo music sampled at 192 kHz. At CD quality, the large capacity of DVD can contain complete works, such as all Beethoven's symphonies, on one disc. Visual images that would otherwise be distributed in a paper booklet can be included, and as with a DVD-Video disc, moving pictures can offer value-added contents such as music clips or interviews.

Recently, security is a topic of great importance for commercial media. DVD-Audio supports CPPM, a state-of-the-art content protection system and an audio watermark to take care of the analog based copying. Both measures cooperate to prevent piracy, unbounded casual copying, or clone-copying. In addition, DVD-Audio allows for administrable copying to a secure destination media such as DVD audio recording, which has been under development by the DVD Forum. The DVD Audio Recording Specification supports the same audio specifications as DVD-Audio (including the MLP Lossless) and has options for some lossy coding schemes for longer time recording.

DVD-Audio in collaboration with DVD audio recording will provide high-quality and secure platforms for music assets, and the authors hope that music lovers will recognize the value of such storage media.

Acknowledgments

We wish to thank everyone concerned in the establishment of the DVD-Audio specifications, its promotion, and business. We are also grateful to the developers of MLP.

References

[1] DVD Forum secretariat. Tokyo, Japan. Available: http://www.dvdforum.org/

[2] "DVD Specifications for Read-Only Disc Part 3: VIDEO SPECIFICATIONS Version 1.1_3," in *DVD Forum*, Tokyo Japan, Mar. 2002.

[3] "DVD Specifications for Read-Only Disc Part 4: AUDIO SPECIFICATIONS Version 1.2," in *DVD Forum*, Tokyo, Japan, Mar. 2001.

[4] "DVD Specifications for Read-Only Disc Part 4: AUDIO SPECIFICATIONS, Packed PCM: MLP Reference Information Version 1.0," in *DVD Forum*, Tokyo, Japan, Mar. 1999.

[5] DVD Copy Control Association (DVD CCA). Available: http://www.dvdcca.org/

[6] 4C Entity, LLC. Available: http://www.4centity.com/

[7] "DVD Specifications for Read-Only Disc Part 1: PHYSICAL SPECIFICATIONS Version 1.0_4," in *DVD Forum*, Tokyo, Japan, June 2002.

[8] "DVD Specifications for Read-Only Disc Part 2: FILE SYSTEM SPECIFICATIONS Version 1.0_4," in *DVD Forum*, Tokyo, Japan, June 2002.

[9] *Information Technology—Generic Coding of Moving Pictures and Associated Audio*, ISO/IEC Standard 13818, 1994.

[10] *Multi-Channel Stereophonic Sound System with and Without Accompanying Picture*, ITU-R Recommendation BS.775-1, 1994.

[11] M.A. Gerzon, J.R. Stuart, R.J. Wilson, P.G. Craven, and M.J. Law, "The MLP Lossless compression system," in *AES 17th Int. Conf. High-Quality Audio Coding*, Florence, Italy, Sept. 1999, pp. 61-75.

[12] Acoustic Renaissance for Audio. (1995, Feb.). A Proposal for High-Quality Application of High-Density CD Carriers. [Private publication]. Available: http://www.meridian-audio.com/ara

[13] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27 pp. 379-423, 623-656, July 1948, Oct. 1948.

[14] A. Robinson, "Shorten: Simple lossless and near-lossless waveform compression," Cambridge Univ., Cambridge, U.K., Tech Rep CUED/F-INFENG/TR.156, Dec 1994.

[15] A.A.M.L. Bruekers, A.W.J. Oomen, and R.J. van der Vleuten, "Lossless coding for DVD audio," in *Proc. AES 101st Conv.*, Los Angeles, CA, Preprint 4358, Nov. 1996.

[16] C. Cellier, P. Chenes, and M. Rossi, "Lossless audio bit rate reduction," in *Proc. AES UK Conf. Managing the Bit Budget,* May 1994, pp. 107-122.

[17] P.G. Craven and M.A. Gerzon, "Lossless coding for audio discs," *J. Audio Eng. Soc.*, vol. 44, pp. 706-720, Sept. 1996.

[18] P.G. Craven and M.A. Gerzon, "Lossless coding method for waveform data," International Patent Application PCT/GB96/01164, May 15, 1996.

[19] P.G. Craven, M.J. Law, and J.R. Stuart, "Lossless compression using IIR prediction filters," *J. Audio Eng. Soc. (Abstracts)*, vol. 45, p. 404, Mar. 1997.

[20] P.G. Craven and M.A. Gerzon, "Optimal noise shaping and dither of digital signals," presented at the AES 87th Convention, New York, Preprint 2822, Oct. 1989.

[21] *Documentation—International Standard Recording Code (ISRC)*, ISO Standard 3901, 1986.

[22] Verance Corporate Headquarters. CA. Available: http://www.verance.com/index.html