The MLP Lossless Compression System for PCM Audio*

M. A. GERZON (deceased), **P. G. CRAVEN,** *AES Member*,[2] **J. R. STUART,** *AES Fellow*,[1]
M. J. LAW, *AES Member*,[2] **AND R. J. WILSON,** *AES Fellow*[1]

[1]*Meridian Audio Ltd., Huntingdon, Cambs., PE29 6EX, UK*

[2]*Algol Applications Ltd., Steyning, West Sussex, BN44 3RE, UK*

Lossless compression provides bit-exact delivery of the original signal and is ideal where the highest possible confidence in the final sound quality is required. Meridian lossless packing (MLP) was adopted in 1999 as the lossless coding method used on DVD-Audio. MLP uses four principal strategies to reduce both the total quantity and the peak rate of encoded data. MLP can invert a matrix transformation losslessly, this allows a two-channel representation to be transmitted alongside a multichannel signal, with a minimal increase in the data rate. It is illustrated how the characteristics of the incoming audio affect the coding performance, and MLP's versatility, achieved by the use of substreams and an open-ended metadata specification, is demonstrated.

0 INTRODUCTION

Meridian lossless packing (MLP)[3] is a lossless coding system for use on high-quality digital audio data originally represented as linear pulse-code modulation (PCM). High-quality audio nowadays implies high sample rates, large word sizes, and multichannel. This paper describes the MLP system while presenting insights into lossless coding in general.

1 OVERVIEW

MLP performs lossless compression of up to 63 audio channels at any bit depth up to 24. There is no inherent limitation on the sample rate, although on DVD-A this is limited to 192 kHz.

Lossless compression has many applications in the recording and distribution of audio. In designing MLP we have paid particular attention to the application of lossless compression to data-rate-limited transmission (such as storage on DVD), to the option of a constant data rate in the compressed domain, and to aspects that impact on mastering and authoring. MLP was targeted to provide:

- Good compression of both peak and average data rates
- Use of both fixed- and variable-rate data streams
- Automatic savings on bass-effects channels
- Automatic savings on signals that do not use all of the available bandwidth (for example, sampled at 96 kHz)
- Automatic savings when channels are correlated
- Comprehensive metadata
- Hierarchical access to multichannel information
- Modest decoding requirements.

Reduction of the peak data rate is equivalent to reducing the word width of 48 kHz sampled signals by 4 bit or more. At least 8 bit is removed from signals sampled at 96 kHz, and so 24-bit audio can be compressed into a 16-bit channel. MLP provides for up to 63 channels, but applications tend to be limited by the available data rate. To aid compatibility, MLP uses a hierarchical stream structure containing multiple substreams and hierarchical additional data. With this stream structure decoders need to access only part of the stream to play back subsets of the audio. Suitable use of the substreams also allows two-channel compatibility. A low-complexity decoder can recover a stereo mix from a multichannel stream. Fig. 1 gives an overview of the process of compressing a stream

*Presented at the AES 17th International Conference on High-Quality Audio Coding, Florence, Italy, 1999 September; revised 2003 December 18.

[3]Meridian, Meridian Lossless Packing, and MLP are registered trademarks of Dolby Laboratories, Inc.

containing multiple audio channels and auxiliary data onto a disk.

2 LOSSLESS COMPRESSION

Unlike perceptual or lossy data reduction, lossless coding does not alter the final decoded transmitted 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 reflects closely 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. Fig. 2 illustrates such a variation through 30 s of a six-channel recording of baroque chamber music at 96 kHz 24-bit precision (original data rate 13.824 Mbit/s).

While a music example can show this kind of compression, we reasonably expect (and see) wider variations in the 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 the average data rate (that is minimizing compressed file size). The ARA proposal [1] describes the important requirement of reducing the instantaneous peak data rate for optimum results at high sampling rates such as 96 or 192 kHz and for data-rate-limited disk-based applications such as DVD-Audio. MLP was developed by the authors as a simple-to-decode method optimized for these special requirements of high-rate high-quality audio combined with an unbreakable requirement to reduce the peak data rate at all times. MLP tackles this by attempting to maximize the compression at all times using the following set of techniques:

- Looking for "dead air"—channels that do not exercise all the available word size
- Looking for "dead air"—channels that do not use all the

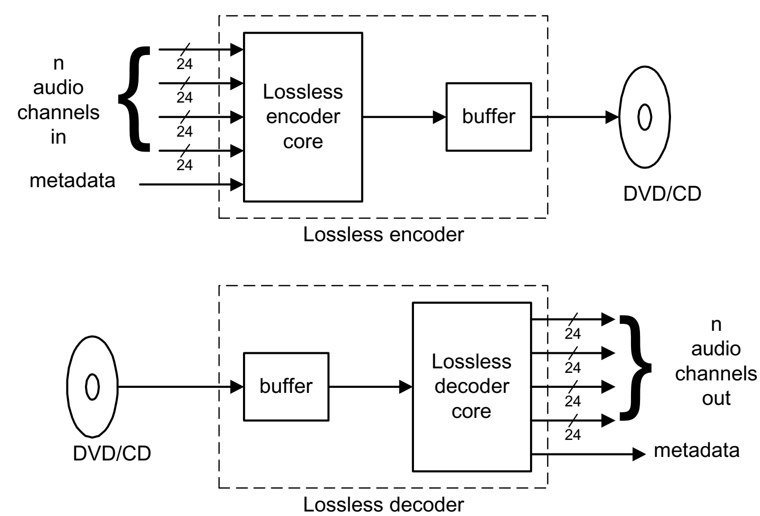

Fig. 1. Overview of MLP used on disk.

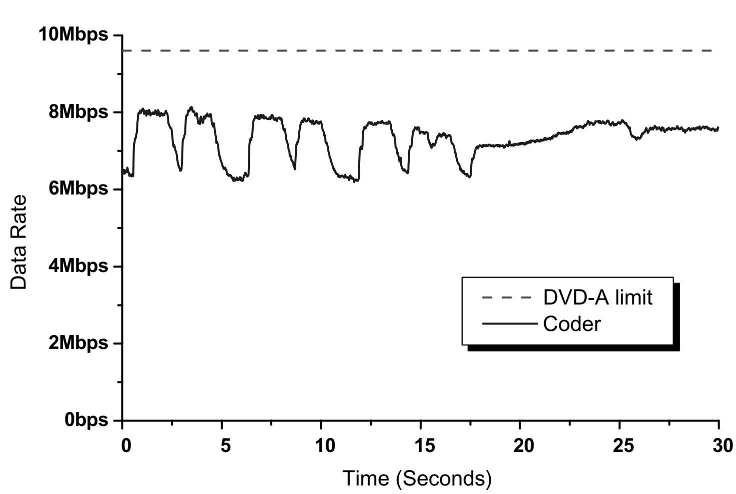

Fig. 2. Data rate over a six-channel excerpt of chamber music recorded at 96 kHz 24 bit.

available bandwidth
- Removing interchannel correlations
- Coding the residual information efficiently
- Smoothing coded information by buffering.

2.1 Application Factors

A lossless compression system must guarantee lossless (that is bit-for-bit) recovery over an encoding–decoding pass. If this is achieved, then the system will remain lossless over multiple cascades of encoding–decoding; there will be no generation loss. A significant requirement of a versatile coding system is that the process remain lossless, regardless of the encoder or decoder computing platform.

The average data rate after compression (coding ratio) affects playing time and hard-disk storage applications the most. MLP allows the compressed data to be packed to a variable data rate on the disk, which maximizes playing time. However, as explained earlier, the peak data rate can be very important in two cases:

- When there is a need to fit the compressed data into a channel that has a lower rate capacity than the incoming audio.
- When, for a particular application, the compressed data are packed to a constant data rate, then this rate cannot be less than the peak rate of the item. Examples include packetizing MLP in Sony/Philips Digital Interconnect Format (S/PDIF) or in a constant-rate stream to accompany motion video.

2.2 Integrity

A lossless encoding–decoding system displays an inherent integrity. Once audio has been "wrapped up" in the MLP stream, it will remain intact through any intermediate storage or transmission process. An MLP decoder can continuously test against checks inserted by the encoder that the overall transmission has been lossless. This makes the audio more secure than transmitting LPCM, since in that case the receiver cannot tell whether intermediate processes have occurred on the data. However, any coded stream is subject to random media transmission errors. To minimize the impact of these, MLP has several error-detection crosschecks in the stream. Another important consideration for a practical system is to be able to start and stop decoding quickly and to avoid unnecessary latency.

3 HOW DOES IT WORK?

MLP coding is based on established concepts. However there are some important novel techniques used in this system, including the following:

- Lossless processing
- Lossless matrixing
- Lossless use of IIR filters
- Managed first-in, first-out (FIFO) buffering across transmission
- Decoder lossless self-check
- Operation on heterogeneous channel sample rates.

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

4 MLP ENCODER

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

1) Incoming channels may be remapped to optimize the use of substreams (described later).

2) Each channel is shifted to recover unused capacity (such as less than 24-bit precision or less than full scale).

3) A lossless matrix technique optimizes 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.

4.1 Lossless Matrix

A multichannel audio mix will usually share some common information between channels. On occasion, such as when widely spaced microphones are used, the correlations will be weak. However, there are other cases where the correlations can be high. Examples include multitrack recordings where a mixdown to the delivered channels may pan signals between channels and thus place common information in some channels. There are also

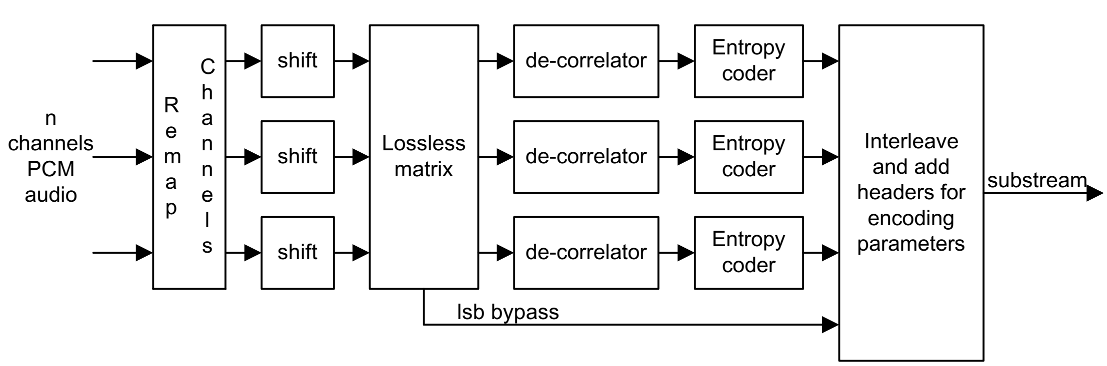

Fig. 3. Block diagram of lossless encoder core.

specific examples where high inter-channel correlations occur, including the following:

- 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 loudspeaker feeds resulting from a hierarchical upscale
- Multichannel loudspeaker 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 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 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 (Fig. 4). For example, if the encoder subtracts a particular linear combination, then the decoder must add it back. The quantizers Q in Fig. 4 ensure constant input–output word width and lossless operation on different computing platforms.

4.2 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 (so called because when optimally adapted there is 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. Fig. 5 shows the short-term spectrum of a music excerpt. If this spectrum were 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 [2] 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 Fig. 5, this deci-

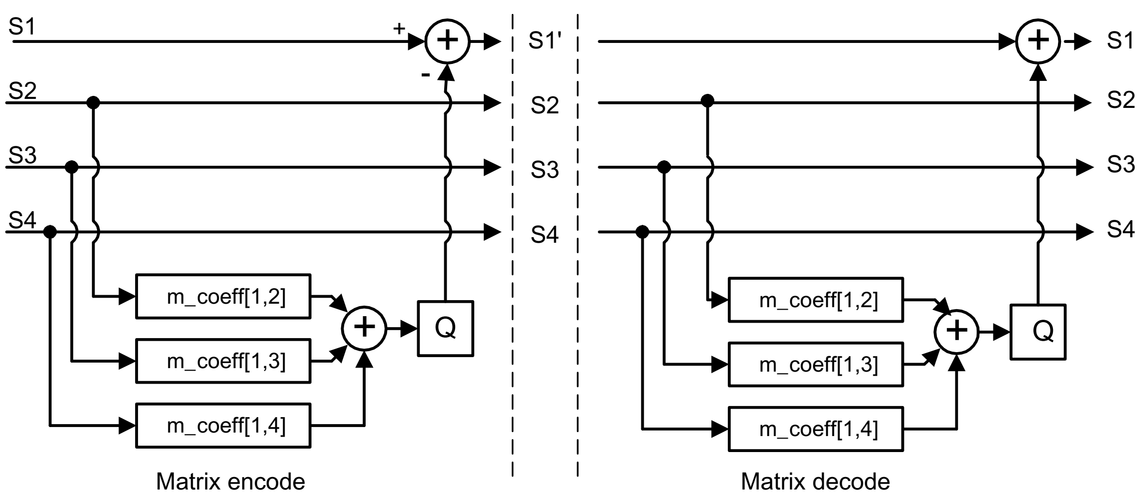

Fig. 4. Single affine transformation used in lossless matrix encode and decode.

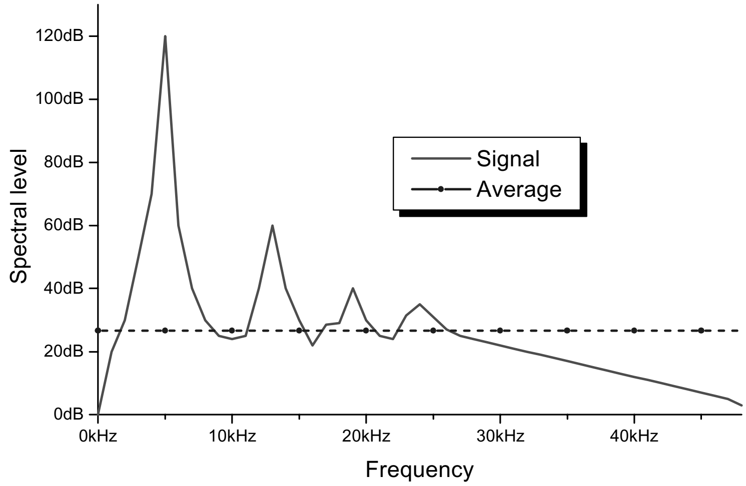

Fig. 5. Spectra of a signal and its average level.

bel average can have significantly less power than the original signal, hence the reduction in data rate. In fact this reduced power represents the information content of the signal as defined by Shannon [3].

In practice, the degree to which any section of music data can be "whitened" depends on its content and on the complexity allowed in the prediction filter. Infinite complexity could theoretically achieve a prediction at the entropy level shown in Fig. 5. However, all the coefficients that 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.

4.3 FIR and IIR Prediction

Most previous lossless compression schemes use FIR prediction filters and can achieve a creditable reduction in data rate on conventional CD-type material [4]–[6]. However, it is pointed out in [7]–[9] that IIR filters have advantages in some situations, in particular:

- Cases where control of the peak data rate is important
- Cases where the input spectrum exhibits an extremely wide dynamic range.

The ARA proposal [1] pointed out the particularly increased likelihood of a wide dynamic range in the spectrum of audio sampled at higher rates such as 96 or 192 kHz. The 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 have the freedom to control the lossless data rate by adjusting supersonic filtering during mastering. A powerful lossless compression system will require the use of FIR and IIR prediction.

Fig. 6 shows the spectrum of a 3.6-ms frame taken from the ending of the "William Tell Overture." This section is high level, contains a cymbal crash, and has a spectrum that is easily flattened by a low-order filter. Fig. 6 also shows the residual spectrum after decorrelation by a fourth-order FIR filter.

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 Fig. 7 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 underused spectrum between the 20-kHz cutoff 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 is placed at, say, 30 or 35 kHz (see [9]).

4.4 Lossless IIR Filtering

IIR predictors are used widely in lossy compression, but a conventional prediction architecture such as that in Fig. 8 does not adapt straightforwardly to lossless compression. To see this, consider that the output of the prediction filter in Fig. 8 generally has a longer word length than the input signal because of the multiplication by fractional coefficients. As the transmitted data rate depends on the total word length at this point, extending the word size would be counterproductive.

Fig. 9 shows a conventional way of dealing with this. Here the output of the prediction filter is quantized so that the transmitted prediction error has the same word length as the input signal. By symmetry, the decoder (also in Fig. 9) can recover the original signal despite the fact that the side chain is now nonlinear.

However, this assumes that the predictors of the decoder and the encoder produce outputs which, when requantized, agree exactly. This is not difficult to achieve with an FIR filter, as its output can be computed exactly using finite-word length arithmetic.

In contrast, IIR filters with fractional coefficients can-

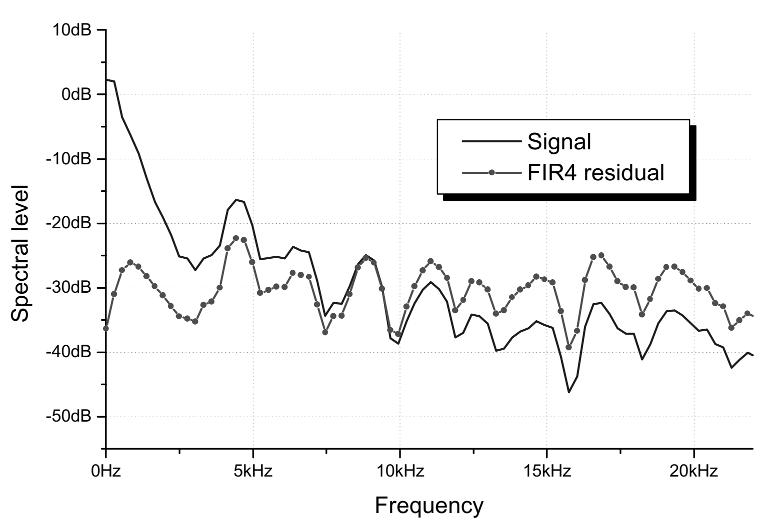

Fig. 6. Spectra of a signal excerpt and residual using a fourth-order FIR predictor.

not be exactly implemented since representation of the recirculating signal requires an ever-increasing word length. The IIR output is thus dependent on the rounding behavior of the underlying arithmetic, and it is difficult to ensure that this will not sometimes affect the quantized output also.

Thus we have the possibility that a decoder implemented on different hardware (such as a computer or a DSP chip) from the encoder will not reproduce exactly the same bits and the compression will not be lossless.

In [8] the encoding architecture of Fig. 10 with the corresponding decoder of Fig. 11 was proposed. As the input and output signals are both quantized and filters A and B are both FIR, the input to the quantizer Q is a finite-precision signal, and the quantization can therefore be specified precisely. On the other hand, because of the recirculation through filter B, the total response is IIR. We have achieved the aim of constructing an IIR predictor that is portable across hardware platforms.

4.5 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 Fig. 12, which graphs the data rate through a 30-s 96 kHz 24-bit six-channel orchestral excerpt.

The lowest curve in Fig. 12 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 the coding ratio was

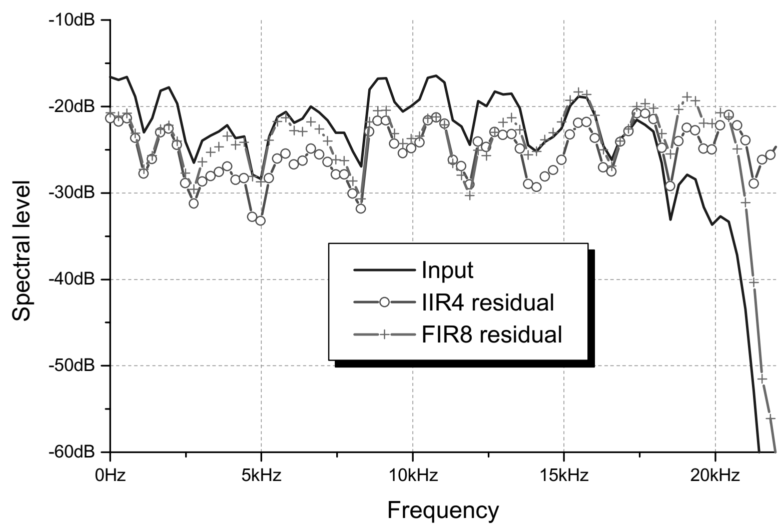

Fig. 7. Spectra of a signal excerpt and residuals using eighth-order FIR and fourth-order IIR predictors.

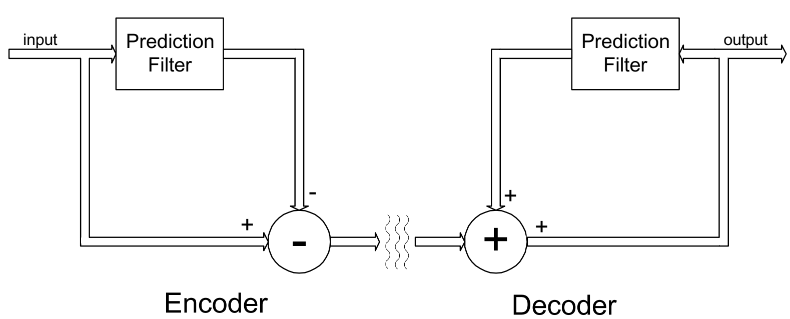

Fig. 8. General schematic of encode and decode using prediction filters.

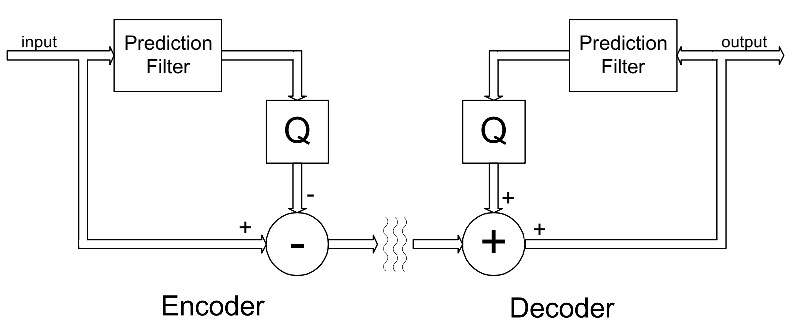

Fig. 9. Schematic of encode and decode using prediction filters with quantizers.

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

4.6 Entropy Coding

Once the crosschannel 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 term given to this process, 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 (see [4]–[6]), that is, the histogram is a two-sided decaying exponential. This appears to be true even after decorrelation. The Rice code (see [4], [5]) provides a simple and near optimal way of encoding such a signal to a binary stream and has the

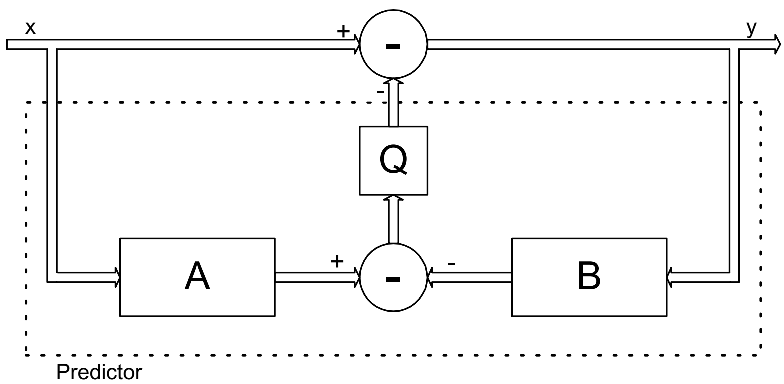

Fig. 10. Lossless IIR prediction structure (encoder).

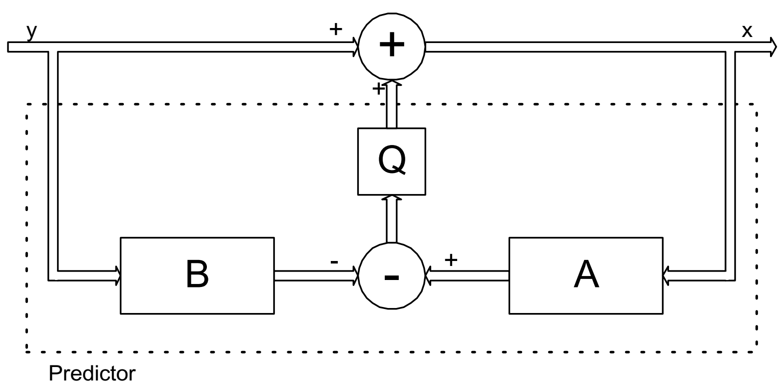

Fig. 11. Lossless IIR prediction structure (decoder).

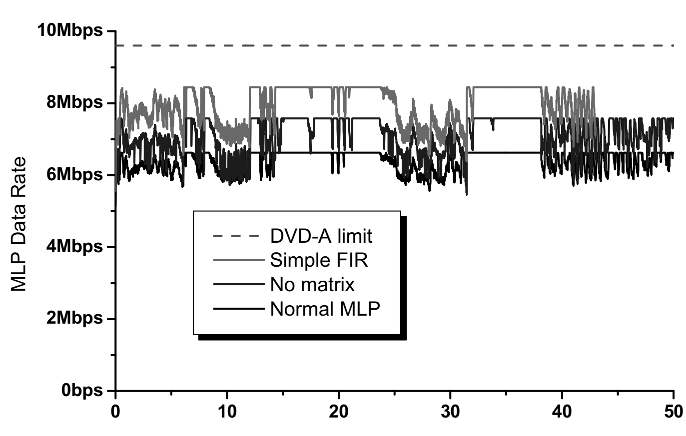

Fig. 12. Data rate for MLP encoder showing benefit of encoder stages.

advantage that encoding and decoding need not use tables.

The Rice code is not used unconditionally. The MLP encoder may choose from a number of entropy coding methods. Although MLP is designed principally for music or speech signals, it is always possible that it may be asked to encode peak-level rectangular probability density function (all values equally probable) white noise. In fact ordinary PCM (which would be optimal for this rogue case) is one of the coding options available to the MLP encoder.

4.7 Buffering

We have explained that while normal audio signals can be well predicted, there will be occasional fragments such as sibilants, synthesized noise, or percussive events that have high entropy. MLP uses a particular form of stream buffering that can reduce the variations in the transmitted data rate, absorbing transients that are hard to compress.

FIFO memory buffers are used in the encoder and decoder as shown in Fig. 13. These buffers are configured to give a constant notional delay across encode and decode. This overall delay is small—typically on the order of 75 ms. To allow rapid start up or cuing, the FIFO management minimizes the part of the delay due to the buffer of the decoder. So the decoder buffer is normally almost empty and fills only when the encoder (which incorporates look-ahead) sees that a section with a high instantaneous data rate lies ahead.

During these sections, the decoder 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 disk, 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, such as:

- Keeping the data rate below a preset (format) limit
- Minimizing the peak data rate over an encoded section.

Fig. 14 shows an example of the latter. The entropy-coded data rate from the encoder core is shown along with the buffered result. The buffered data have a characteristic flat-topped curve. This is not due to clipping or overload, but to rate absorption in the encoder–decoder FIFOs.

Another illustration of data-rate minimization is shown in Figs. 15 and 16. Again the encoded data rate is plotted through a 30-s 96 kHz 24-bit six-channel excerpt featuring a close recording of a jazz saxophone. Fig. 15 indicates the underlying compression when the encoder does not limit the data rate. The minimum-rate encode shown in Fig. 16 makes long-term use of the decoder buffer. It should be obvious that the situation in Fig. 16 is preferable if the transmission channel (maybe a DVD disk) has other calls on the bandwidth—for example, the bandwidth to transmit an associated picture or text.

Fig. 17 illustrates how hard-to-compress signals can be squeezed below a preset format limit. This 30-s 96 kHz 24-bit 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 Mbit/s. Buffering allows the MLP encoder to hold the transmitted data rate below 9.2 Mbit/s by filling the decoder

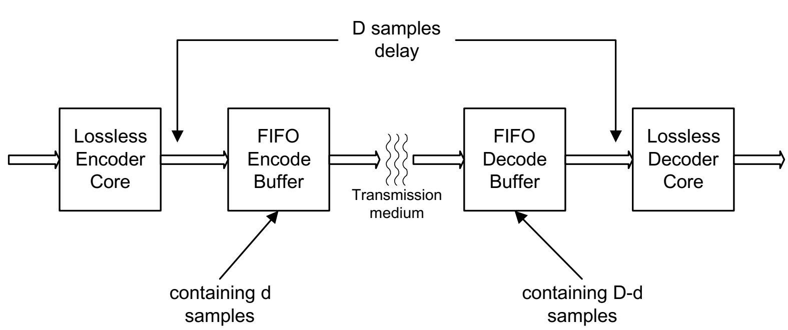

Fig. 13. Buffering used in MLP.

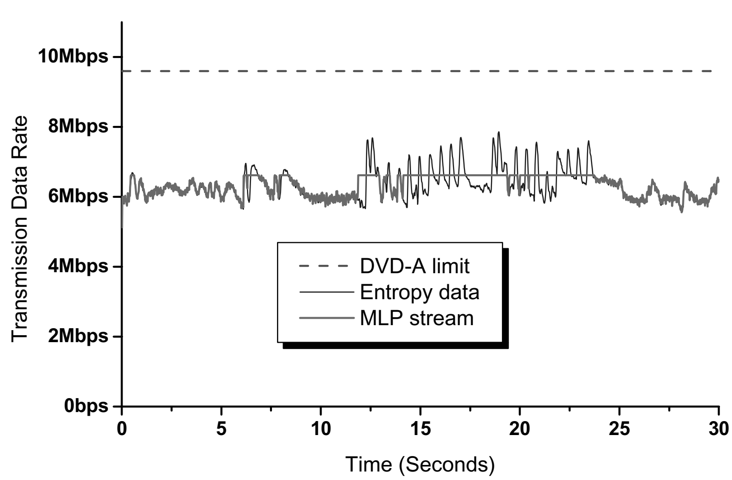

Fig. 14. Illustration of how buffering can minimize data rate in transmission channel.

buffer to a short-term maximum of 86 kbyte (bottom curve). Fig. 18 shows the potential for peak data-rate reduction on this item with different amounts of available FIFO memory.

5 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 in Figs. 19 and 20 for the encoder and the decoder respectively. Note that each substream is buffered separately. We see in Fig. 20 that the output of decoder 0 is (losslessly) matrixed into the output of decoder 1 to build up the overall signal.

6 MLP DECODER

The MLP decoder core is shown in Fig. 21. The decoder unwinds each encoder process in reverse order. The decoder is of relatively low complexity.

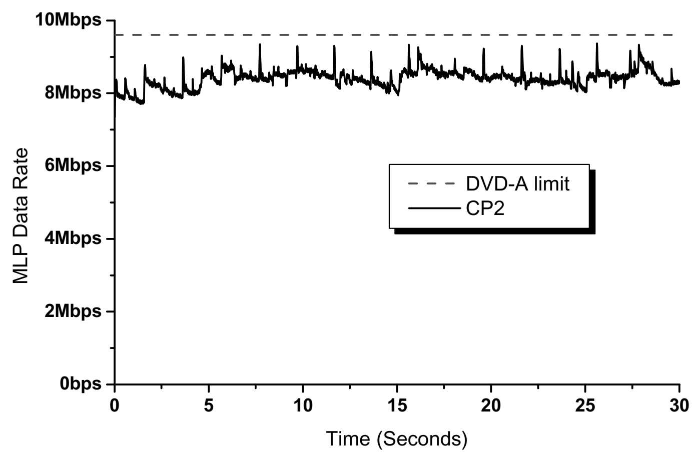

Fig. 15. Unlimited data-rate encoding.

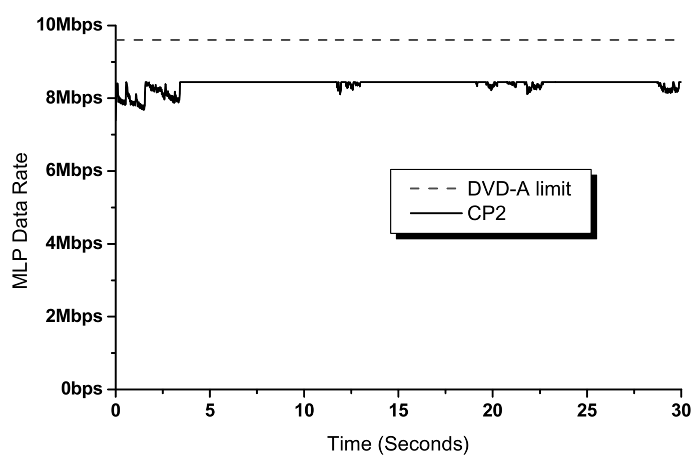

Fig. 16. Minimum data-rate encoding.

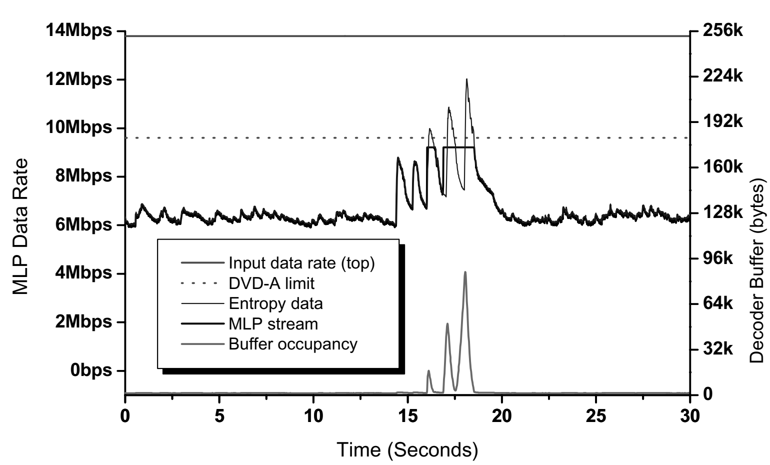

Fig. 17. Buffering allows a difficult passage to remain below a hard format limit.

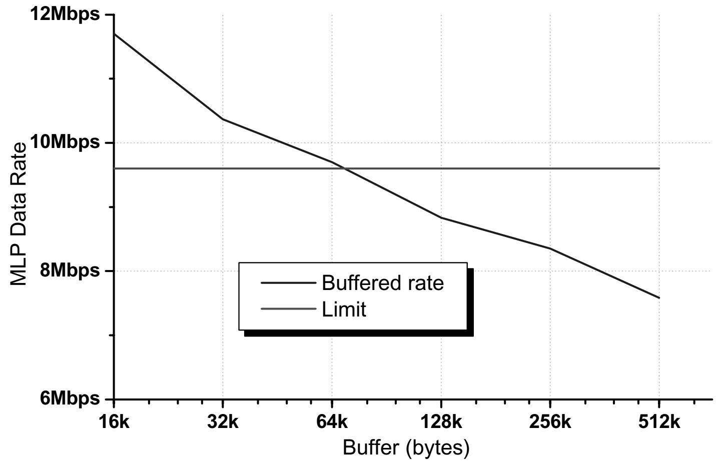

Fig. 18. Effect of buffering on data rate.

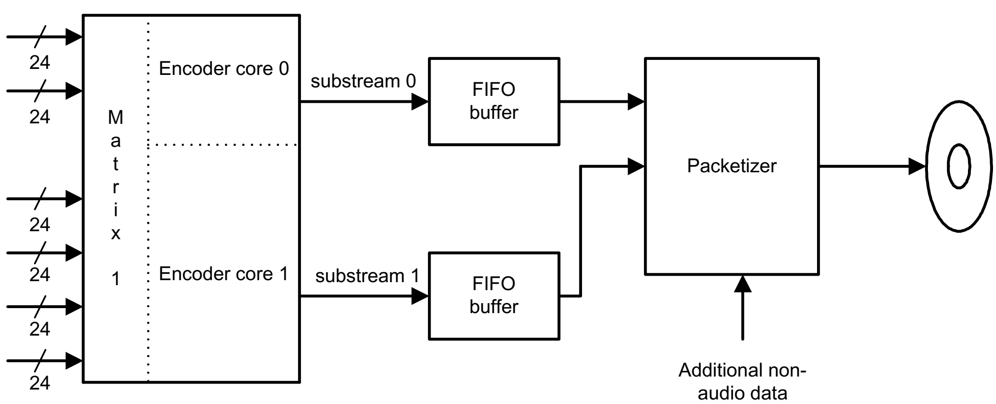

Fig. 19. Encoding two substreams.

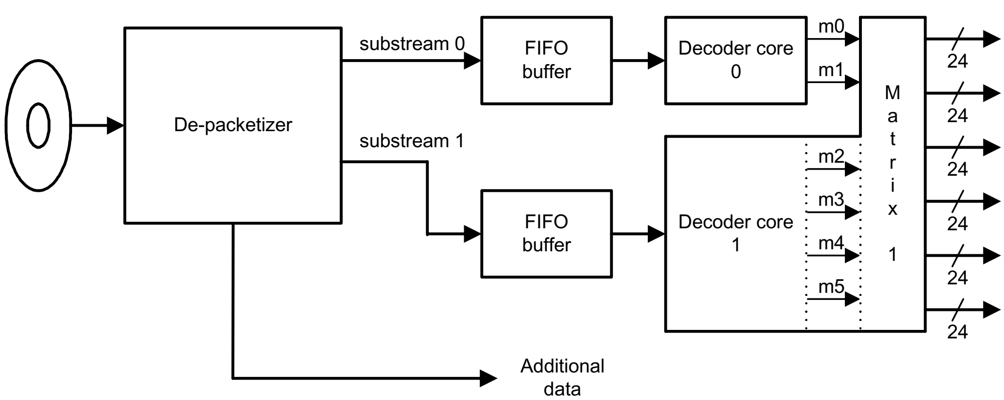

Fig. 20. Decoding two substreams.

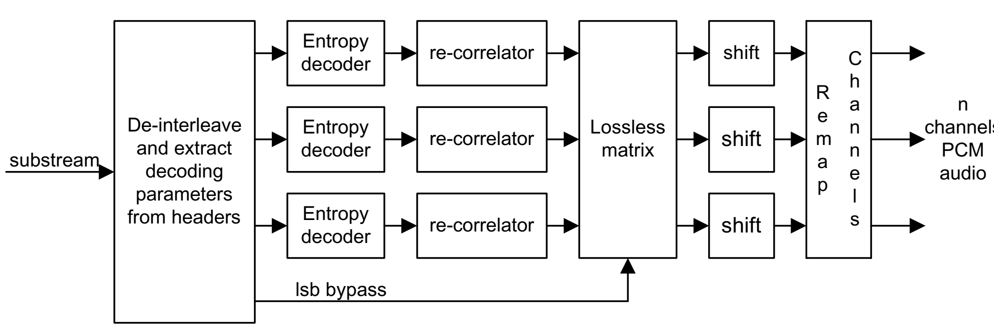

Fig. 21. Block diagram of lossless decoder core.

7 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 disk. However, to do this requires separate mixing, mastering, and authoring processes and uses disk capacity.

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

7.1 Performing Mixdown 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 two-channel downmix and the multichannel version. This method is illustrated in Fig. 22.

Downmix instructions are used to determine some coefficients for the lossless matrices. The matrices then perform a transformation 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 Fig. 22 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 the overall data rate (typically 1 bit per sample, such as 96 kbit/s at 96 kHz). Fig. 23 shows an example where a downmix is added to the six-channel segment from Fig. 16.

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 mixdown. Instead, the lossless decoder need only decode substream 0.
- A more complex decoder may access both the two-channel and the multichannel versions losslessly.
- The downmix coefficients do not have to be constant for a whole track, but can be varied under artistic control.

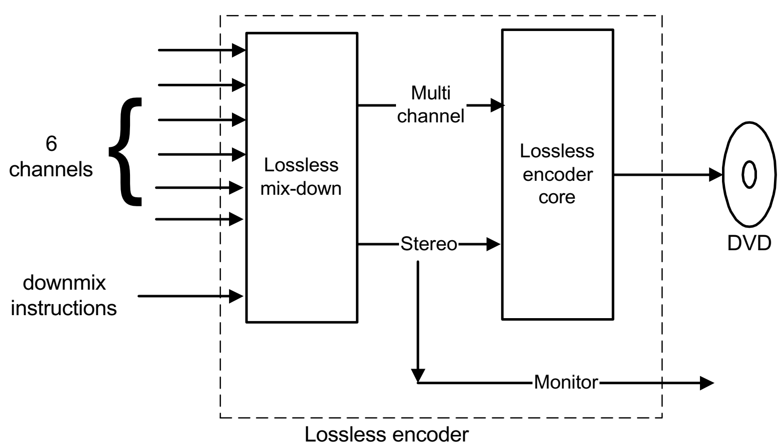

Fig. 22. Encoder downmix.

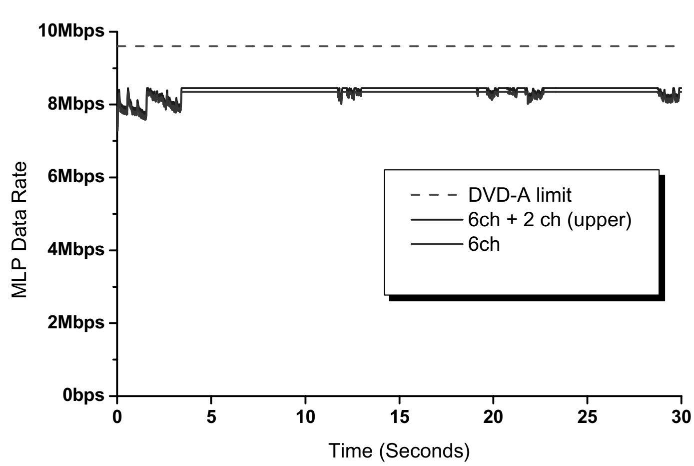

Fig. 23. Impact on data rate of adding a two-channel downmix to six-channel content.

8 MLP BIT-STREAM FORMATS

The encoded stream carries all the information necessary to decode the stream. This information includes the following:

- Instructions to the decoder
- Compressed data
- Auxiliary data (content provider's information)
- CRC check information
- Lossless testing information.

Incoming audio is encoded in segments and the bit stream uses a packet structure as follows:

- Data are encoded in blocks that typically contain between 40 and 160 samples.
- Blocks are assembled into packets. The user and/or the encoder can adjust the length of packets. A typical range is between 640 and 2560 samples.
- Each packet contains full initialization and restart information. Therefore the decoder can recover from severe transmission errors, or start up losslessly in midstream typically, within 7 ms.

8.1 Error Handling

MLP has powerful built-in detection that allows rapid recovery from bit stream errors. In addition,

- Errors cannot propagate beyond a packet boundary.
- Recovery from 1-bit errors generally occurs within 1.6 ms.
- Multiple checks in the stream prevent erroneous noises, "clicks" or "bangs."

8.2 Variable-Rate Bit Stream

A variable-rate MLP stream is packetized to minimize file size. The packetizing method can ensure that the short-term peak data rate is kept as low as possible. Several examples of variable-rate streams have been given in this paper.

8.3 Fixed-Rate Bit Stream

The fixed-rate stream is packetized to provide losslessly compressed audio at a constant data rate. Encoding for a fixed rate can be a single-pass process if the target data rate is always attainable. At times when the compressed data rate is less than the target, the encoder will fold in padding data or transmit a pending payload of additional data (see Section 12).

8.4 MLP Stream Transcoding

An MLP bit stream contains sufficient data to allow transcoding between fixed- and variable-rate streams. Fig. 24 shows circumstances in disk production and playback where transcoding may be useful. Transcoding is a lightweight operation, not requiring a full decode and reencode.

9 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 word width. With lossless compression, increases in incoming precision, that is additional least significant bits (LSBs) on the input, must be reproduced losslessly. However, these LSBs typically contain little redundancy that can be removed by an encoder and thus they contribute directly to the transmitted data rate. Therefore we tend to quote the *saving* in data rate, as this measure is relatively independent of incoming precision. (see Table 1). In Table 1 peak savings are for "difficult" signals while average savings reflect the uncertainty introduced by quiet passages and other variables. Table 1 gives the compression for two-channel material. Compression generally increases as more channels are added or if any channels are correlated, or have low-noise bandwidth (like a subwoofer channel) or low occupancy

Table 1. Peak and average data-rate reduction on two-channel material.

Sampling (kHz)	Data-Rate Reduction (bits/sample/channel) Peak	Average
48	4	5–11
96	8	9–13
192	9	9–14

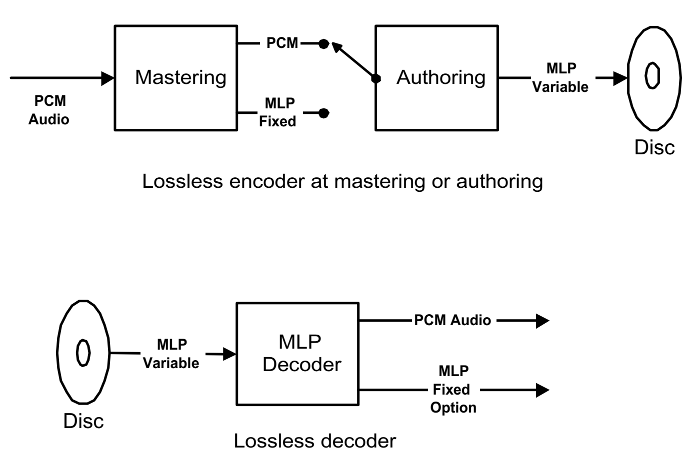

Fig. 24. In a mastering or replay environment it may be desirable to transcode MLP streams.

(such as typical surround channels).

At 44.1 or 48 kHz the peak data rate can almost always be reduced by at least 4 bit per sample, that is, 16-bit audio can be losslessly compressed to fit into a 12-bit channel. At 96 kHz the peak data rate can similarly be reduced by 8 bit per sample, that is, 24-bit audio can be compressed to 16 bit and 16-bit 96-kHz audio can be losslessly compressed to fit into an 8-bit channel. The important parameter for transmission applications is the reduction of the peak rate. In the case of DVD-Audio, the peak rate is a key parameter because the encoded stream must always operate below the audio buffer data-rate limit of 9.6 Mbit/s.

The average number in Table 1 indicates the degree of compression that could be obtained when using MLP in an archive, mastering, or editing environment. For example, a peak-data-rate reduction of 8 bit per sample means that a 96-kHz 24-bit channel can be carried on the disk with a rate equal to that of a $24 - 8 = 16$-bit LPCM channel. However, the space used on the disk is estimated by the average saving, in this case the residual will be $24 - 11 = 13$ bit per channel.

Consider that an 11-bit saving represents a compression ratio of 1.85:1 with 24-bit material, whereas the same saving compresses 16-bit audio by 3.2:1. Of course the amount of lossless compression attainable is limited by the noise floor of the recording itself.

Fig. 25 shows a typical progression through two-channel 192-kHz 24-bit material (original data rate 9.216 Mbit/s). Figs. 26 and 27 show compression examples at CD quality. The two-channel example in Fig. 26 shows an average 2:1 compression. Note that the three-channel horizontal ambisonic B-format (WXY) stream in Fig. 27 (opening of Rachmaninov's Second Piano Concerto) shows sufficient peak-rate compression to allow the stream to fit on a CD.

9.1 Compression Adjustment

A producer may wish to save space used by a recording, or to reduce the data rate. Lossless compression extends the number of options. With MLP, data are saved automatically if the incoming precision is reduced. So reducing, for example, a few or all channels in a mix from 24 to 22 bit will provide an automatic data saving. The concept is illustrated in Fig. 28. The authors have previously described appropriate quantizing strategies. [2], [10]–[12].

In an overall sense the process of Fig. 28 could be viewed as lossy. However, this is not the case if the producer makes the adjustment. A conventional lossy system

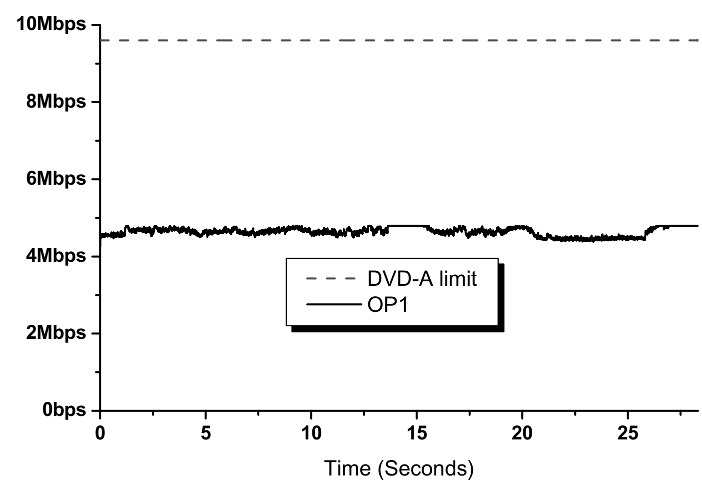

Fig. 25. Compressed data rate for 24-bit two-channel item sampled at 192 kHz.

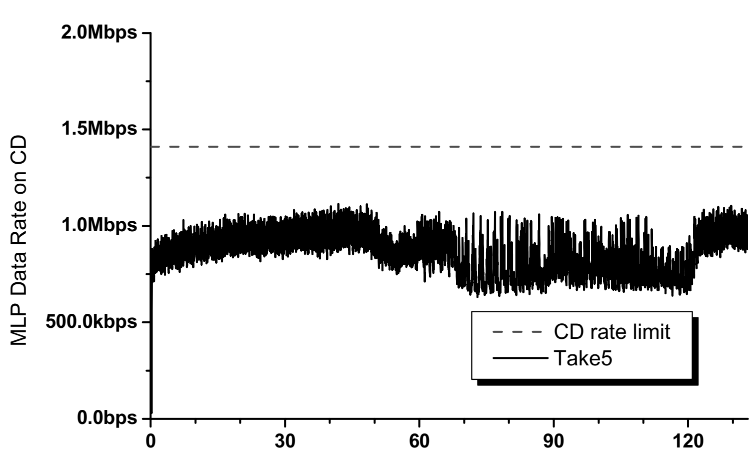

Fig. 26. Compressed data rate for "Take Five," a 16-bit two-channel 44.1-kHz item from CD.

provides no choice about how the signal is modified in order to fit the desired data rate, whereas in Fig. 28 the producer can use artistic judgment to select both the method and the amount of word-width reduction. The output of the quantizer can be monitored, and that signal will be delivered losslessly by the MLP decoder.

This does not exclude the possibility that a quantizer for use with an MLP encoder could adapt incoming precision automatically, a circumstance envisaged by the authors in [13]. An intriguing property of a lossy encode made in this way is that it can be losslessly cascadable, that is, it would be lossless over subsequent encode–decode passes. Another option for reducing encoded data is to low-pass filter some of the incoming channels. Low-pass filtering reduces the entropy in the signal and the lossless coder generally provides a lower data rate. A typical 96-kHz 24-bit six-channel program would encode to an average of 7.2 Mbit/s. Reducing the audio bandwidth with simple filtering from 48 to 24 kHz will generally reduce the rate to below 5 Mbit/s.

A less drastic alternative is to use an “apodizing” filter [14], which will reduce the data rate to about 6 Mbit/s. The apodizing filter potentially provides an improved transient response as well as reducing the data rate.

10 FEATURES FOR CONTENT PROVIDERS

MLP allows the record producer to make a personal tradeoff between playing time, frequency range, number of active channels, and precision. The packed channel conveys this choice implicitly in its control data, and the system operation is transparent to the user. This method has the following example benefits:

- A producer mastering at 48 kHz can control the incoming precision of each channel—and trade playing time or channels for noise floor.
- A producer mastering at 96 or 192 kHz can in addition trade bandwidth for playing time, active channels, and precision.

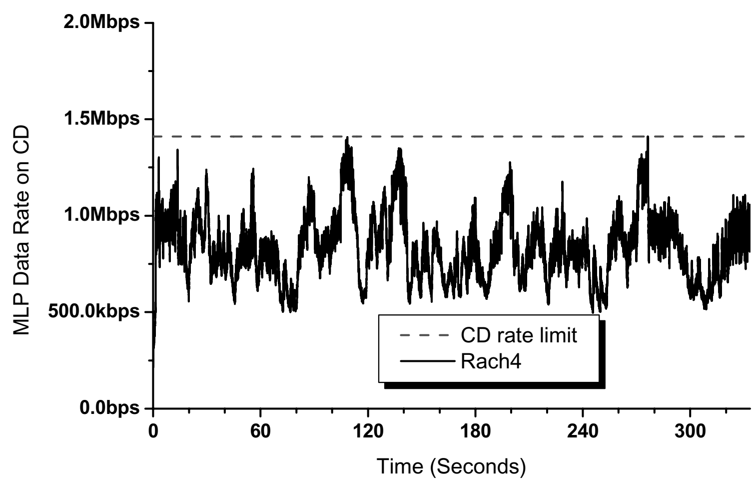

Fig. 27. Compressed data rate for a horizontal ambisonic WXY 16-bit three-channel 44.1-kHz fragment compressed for delivery on CD.

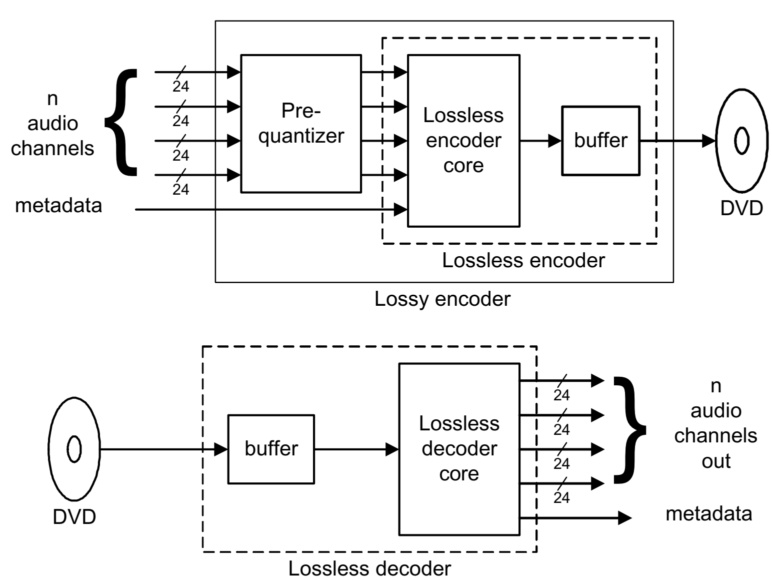

Fig. 28. Generalized schematic of prequantization showing a lossy–lossless encode and lossless decode.

For example:

- Playing time or precision may be extended by prefiltering information above some arbitrary frequency (such as 30 kHz), thereby allowing more compression.
- Playing time or precision may be extended by only supplying a two, three, or four-channel mix.
- Feeding smaller word sizes to the encoder will extend playing time. For example, when reducing from 24 to 23 or 22 bit, each bit removed will increase playing time by around 8%.

MLP always returns the streams bit for bit intact once any mastering adjustments have been made.

10.1 DVD-Audio Content

MLP has some features that assist content providers in providing material for issue on DVD-Audio, including the following:

- Longer playing time than allowed by LPCM
- Higher quality by delivering more channels or bits for the same playing time
- Guaranteed quality; the lossless decoder delivers bit-accurate data
- High-quality mixdown options; longer playing time with multichannel material
- Fine control over delivered quality and playing time
- The large reduction in the audio data rate means that many more options for audio with pictures are possible
- Additional data channel in the stream to carry copyright information
- Additional data channel in the stream to carry signature information
- Bit-stream definition allows more than six channels for recording and archive.

10.2 Playing Time on DVD-Audio

DVD-Audio holds approximately 4.7 Gbyte of data and has a maximum data transfer rate of 9.6 Mbit/s for an audio stream. Six channels of 96-kHz 24-bit LPCM audio has a data rate of 13.824 Mbit/s which is well in excess of 9.6 Mbit/s. Also, at 13.824 Mbit/s, the data capacity of the disk would be used up in approximately 45 min. So lossless compression is needed to reduce the data on the disk to extend the playing time to the industry norm of 74 min and to guarantee a minimum reduction of 31% in the 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 [15]. Here are some examples of playing times that can be obtained:

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

11 SYSTEM DEFINITION AND FLEXIBILITY

MLP was conceived as a general-purpose lossless compression system. However, a high-density replacement for the consumer CD was foreseen as an early application, and this has driven the system design in two directions:

- Any complexity must be in the encoder rather than the decoder.
- The system is defined in terms of the bit stream and the required decoder behavior.

As a result of the second point, encoder developments may continue (for example, for increased compression) without outdating the installed base of decoders. Current decoders are required to decode any legal bit stream, so there will be no question of “old” decoders being unable to decode “new” software.

The bit stream has been designed to keep open as many options as possible for future encoder developments, while not impacting decoder complexity and data rate more than necessary. While the highest compression requires sophisticated encoders, near optimal encoding of most music signals can be obtained with much simpler encoders that have modest data-rate requirements and can run in real time on cheaply available DSP devices. Thus future use in consumer record–playback systems is entirely feasible.

Neither encoding nor decoding mandate the use of fast Fourier transforms or other block processing, so it is also possible to construct encoders and decoders with very low latency, for use in radio microphones or other real-time applications.

12 SIGNAL AND METADATA

A design aim of MLP was to provide a simple external connectivity. An encoder has (conceptually) n identical input sockets, and the corresponding decoder has n output sockets. Externally the system is just like an n-channel 24-bit PCM link. Thus, there is no concept of a 5.1-channel or a 7.1-channel encoder or decoder. If a 5.1-channel signal is presented to a six-channel encoder, the .1 channel will be recognized by the encoder as being highly predictable (on account of its low bandwidth) and should be encoded to an extremely low data rate, ideally about 2 bits per sample. If someone were to invent a 4.2 multichannel format having two low-frequency channels, this too would be automatically and optimally handled.

Likewise there is no need for the word width to be flagged to the encoder. If a 20-bit signal were presented to some or all channels of an MLP encoder, the 4 unused bits would be evident to the encoder and the appropriate economies made.

Channel meaning and word width are examples of metadata. MLP regards metadata as important and provides intact delivery of any metadata that are supplied along with the audio. However, the feeding of metadata to the encoder is entirely optional (unless mandated by a particular application such as DVD-Audio) and in no way affects the handling of the audio signals.

12.1 MLP Metadata Specification

The MLP metadata specification is deliberately open-ended. Items that have been discussed include the following:

- Dynamic-range control data (should there be a wish to compress after a lossless decode)
- Ownership and copy protection fields
- SPL reference
- SMPTE time code
- Content signature
- Provenance information for decoders
- A Rosetta stone text field.

In a system in which the number of signal channels may be up to 63 in the future, it is hard to predict exactly what variations of "channel meaning" data may be needed. Therefore in designing the MLP metadata format:

- Fixed-length bit fields have been avoided.
- Hierarchical data structures are supported.

The significance of the latter item may become apparent from the next section.

12.2 Future Audio Possibilities

In [16] the ARA highlighted the desirability of recording and storing multichannel signals in "hierarchical" (MSTBF) or "ambisonic" (WXYZ) format, while recognizing the need to matrix these signals to feeds for the standard "3 + 2" loudspeaker layout before issuing on a consumer disk. If the matrix coefficients are recorded within the MLP metadata, a suitable decoder can apply the inverse matrix and recover the original hierarchical or ambisonic signals. This is advantageous if it is desirable to use a different layout or number of loudspeakers from the standard 3 + 2.

Moreover, using the same technique of lossless matrixing discussed in Section 4.1, it is possible to recreate the original MSTBF or WXYZ signals in a bit-exact fashion. Methods for using the lossless matrix in this manner are described in [17]. Having done this, the enthusiast will then wish to have a metadata description of the original signal, hence the need for a hierarchical capability on the description language.

Further possibilities include lossless equalization. Having established in Section 4.4 the possibility of a losslessly invertible IIR filter architecture, it is natural to apply this to equalization. Thus a mastering engineer may choose to apply such equalization to "sweeten" a track for consumer release, but if the coefficients are recorded as MLP metadata, it will be possible for the original signal to be recovered losslessly by an appropriate decoder. Clearly a sophisticated audiophile decoder could unravel several nested layers of metadata description, and thus undo several cascaded stages of studio processing, to recover an original signal with bit-exact precision.

13 SUMMARY

The authors set out to develop a lossless coding system whose options enabled the highest possible audio quality in a hierarchical architecture that would allow future extensions. First and foremost MLP is truly lossless and guarantees delivery of the original audio data. The decoder can confirm true end-to-end lossless operation.

Great attention has been paid to the audio compression strategies. A four-level approach incorporating novel lossless use of matrices, processing, and IIR filters allows a high degree of compression at all times. Because MLP will be used on carriers such as DVD-Audio, which have a limited data rate, particular attention was also paid to methods that control the peak rate of the encoded bit stream. The bit stream itself has been defined to allow robust operation, fast error recovery, and rapid cuing (typically recovering in 7 ms).

An unusual feature is the ability to use fixed- or variable-rate streams according to the application.

Following the sensible paradigm that as much system complexity as possible should be embodied in the encoder rather than the decoder, the MLP decoder is relatively simple. The decoder is also hierarchical, has a low computational complexity, is portable, and is lossless over different hardware platforms.

Flexible encoding options include automatic adaptation to the bandwidth of incoming audio and to the incoming word size in 1-bit steps.

In addition to audio, the MLP stream carries additional information of benefit to the decoder, to the content provider, and to the end user. A flexible extensible hierarchical metadata option also allows very effective use of MLP in advanced surround applications.

14 ACKNOWLEDGEMENTS

The authors are grateful to Tony Faulkner (Green Room Productions), Warner Music, JVC, and Nimbus Records for providing experimental recordings and material used in the compression examples. Several aspects of the MLP encode, decode, packetizing and bit stream are covered by patents.

15 REFERENCES

[1] Acoustic Renaissance for Audio, "A Proposal for High-Quality Application of High-Density CD Carriers," private publ. (1995 Feb.), www.meridian-audio.com/ara.

[2] P. G. Craven and M. A. Gerzon, "Optimal Noise Shaping and Dither of Digital Signals," presented at the 87th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 37, p. 1072 (1989 Dec.), preprint 2822.

[3] C. E. Shannon, "A Mathematical Theory of Communication," Bell Sys. Tech. J., vol. 27, pp. 379–423, 623–656 (1948 July, Oct.).

[4] A. Robinson, "Shorten: Simple Lossless and Near-Lossless Waveform Compression," Tech. Rep. CUED/F-INFENG/TR.156, Cambridge University, Cambridge, UK

(1994 Dec.).

[5] A. A. M. L. Bruekers, A. W. J. Oomen, R. J. van der Vleuten, and L. M. vande Kerkhof, "Lossless Coding for DVD Audio," presented at the 101st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 1166 (1996 Dec.), preprint 4358.

[6] C. Cellier, P. Chenes, and M. Rossi, "Lossless Audio Bit Rate Reduction," in *Proc. AES UK Conf. Managing the Bit Budget*, pp. 107–122 (May 1994).

[7] P. G. Craven and M. A. Gerzon, "Lossless Coding for Audio Discs," *J. Audio Eng. Soc.*, vol. 44, pp. 706–720 (1996 Sept.).

[8] P. G. Craven and M. A. Gerzon, "Lossless Coding Method for Waveform Data," Int. patent application PCT/GB96/01164 (1996 May).

[9] P. G. Craven, M. J. Law, and J. R. Stuart, "Lossless Compression Using IIR Prediction Filters," presented at the 102nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 45, p. 404 (1997 Mar.), preprint 4415.

[10] M. A. Gerzon, P. G. Craven, J. R. Stuart, and R. J. Wilson, "Psychoacoustic Noise-Shaped Improvements in CD and Other Linear Digital Media," presented at the 94th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 41, p. 394 (1993 May), preprint 3501.

[11] J. R. Stuart, and R. J. Wilson, "Dynamic Range Enhancement Using Noise-Shaped Dither Applied to Signals with and without Preemphasis," presented at the 96th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 42, p. 400 (1994 May), preprint 3871.

[12] J. R. Stuart and R. J. Wilson, "Dynamic Range Enhancement Using Noise-Shaped Dither at 44.1, 48, and 96 kHz," presented at the 100th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 44, p. 646 (1996 July/August), preprint 4236.

[13] P. G. Craven, and J. R. Stuart, "Cascadable Lossy Data Compression Using a Lossless Kernel," presented at the 102nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 45, p. 404 (1997 Mar.), preprint 4416.

[14] P. G. Craven, "Antialias Filters and System Transient Response at High Sample Rates," *J. Audio Eng. Soc.*, this issue, pp. 216–242.

[15] DVD Format/Logo Licensing Corp. (DVD FLL), "DVD Specifications for Read-Only Disc, Part 4: Audio Specification," version 1.0 (1999 Mar.)

[16] Acoustic Renaissance for Audio, "DVD: Application of Hierarchically Encoded Surround Sound—Including Ambisonics," private publ. (1996 Oct.), www.meridian-audio.com/ara.

[17] P. G. Craven, M. J. Law, J. R. Stuart, and R. J. Wilson, "Hierarchical Lossless Transmission of Surround Sound Using MLP," presented at the AES 24th Int. Conf. on Multichannel Audio (Banff Canada, 2003 June).

THE AUTHORS

M. J. Law

R. J. Wilson

Michael Gerzon's untimely death in 1996 precluded his participation in the MLP project, but he was certainly its progenitor. Mr. Gerzon identified lossless compression as a key component in allowing multichannel audio of the highest resolution to be conveyed in a high-density optical disc, and he made these ideas public through the Acoustic Rennaisance for Audio (see ARA Web site www.meridian-audio.com/ara).

Mr. Gerzon invented the IIR lossless predictor and the lossless matrixing described in this paper. The other authors of the paper are also indebted to Mr. Gerzon for a wealth of inspiration, techniques, and ways of thinking that are still being worked through. Mr. Gerzon combined a passion for audio with deep intuition, a sound knowledge of information theory, and an ability to cope with difficult mathematics. (Mr. Gerzon's ability with difficult mathematics was also exemplified in his other research interest, Quantum Field Theory.)

Further information about Mr. Gerzon's life and contributions to audio can be found in his obituary published in *JAES*, vol. 44, pp. 669–670 (1996 July/Aug.).

●

Malcolm J. Law studied mathematics and computation at Oxford University, UK, where he graduated from with First Class Honours in 1991 and was awarded the Junior Mathematical Prize.

He then joined B&W Loudspeakers and worked on its digital room and loudspeaker equalization project. In 1996 he started working for Algol Applications and is now co-principal. Consultancy projects through Algol have included low-cost reverberation for consumer products and lossy compression algorithms. He is involved with the development of the MLP lossless compression system and has written much of the enabling software. He recently designed the convolution algorithms in the Waves IR1 reverberator.

Mr. Law is married with two children. In his spare time he is a retained firefighter with West Sussex Fire Brigade.

●

Rhonda J. Wilson was born in Sydney, Australia. She studied at the University of Sydney, from which she received a B.Sc. degree in pure mathematics and computer science in 1985 and a B.E. (electrical) degree with First Class Honours in 1987.

Dr. Wilson has worked in the audio industry since 1989 when she joined KEF as a research engineer on the Archimedes project. Late 1990 she commenced work for Meridian Audio, where she is now the engineering research manager. She continues to do research and to develop audio products, with a particular emphasis on high-resolution digital signal processing for loudspeakers, surround sound decoders, and DVD-Audio. With Meridian's support, she also earned a Ph.D. degree, with a thesis on "Noise Source Cancellation in Audio Recordings," from the Imperial College, UK, in 1997.

Dr. Wilson has been a member of the *Journal* Review Board since 1993. She chaired the AES British Section's DSP Conference in 1992 September and many papers sessions at AES conventions and conferences. She has presented several papers at AES conventions and has papers published in *JAES*. She has also served the AES British Section as committee (1991-1994), vice chair (1992-1993), and chair 1993-1994; and has served the AES as governor (1993-1995) and vice president, International Region (1999-2001).

Dr. Wilson appreciates live and recorded music. She earned an A.Mus.A performance diploma for the bassoon in 1982 and enjoys playing in orchestras and smaller ensembles.

The biography of J. Robert Stuart appears in this issue, p. 144.

The biography of Peter G. Craven appears in this issue, p. 242.