PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number:
H04B 14/04, 1/66	A1	WO 94/18762
		(43) International Publication Date: 18 August 1994 (18.08.94)

(21) International Application Number: PCT/GB94/00297

(22) International Filing Date: 15 February 1994 (15.02.94)

(30) Priority Data:
9302982.5 15 February 1993 (15.02.93) GB

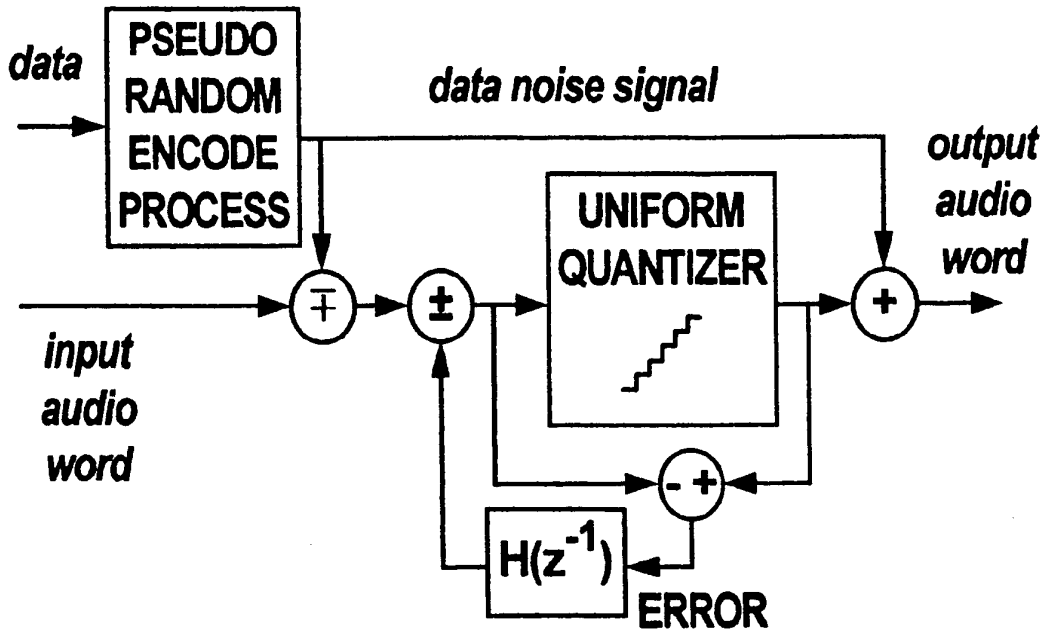
(71)(72) Applicants and Inventors: GERZON, Michael, Anthony [GB/GB]; Ground Floor Flat, 14 Bartlemas Road, Oxford OX4 1XX (GB). CRAVEN, Peter, Graham [GB/GB]; Old School House, Old School Lane, East Challow, Wantage, Oxon OX12 9SG (GB).

(74) Agent: GILL JENNINGS & EVERY; Broadgate House, 7 Eldon Street, London EC2M 7LH (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: TRANSMISSION OF DIGITAL DATA WORDS REPRESENTING A SIGNAL WAVEFORM



(57) **Abstract**

A method and apparatus for encoding digital data within digital words representing signal waveforms, includes the step of modifying least significant digits of the digital words representing signal waveforms in dependence upon the digital data. This is achieved by pseudo-randomising the digital data thereby forming data noise words having levels which are small relative to those of the waveform words, subtracting the pseudo-randomised data words from the waveform words to thereby produce a dithered waveform word, and then quantizing the dithered waveform word and adding the data noise word to the quantized word to thereby form an output of reduced noise carrying information representing the digital data in its least significant digits.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

Transmission of Digital Data Words Representing a Signal Waveform

BACKGROUND TO THE INVENTION

This invention relates to methods and apparatus for transmitting, encoding and decoding data signals using parts of low significance of the digital words representing signal waveforms, particularly in applications where the degradation of the waveform signal resulting from the data coding is desired to be of minimal or benign effect.

In many applications where a signal waveform is represented by a sequence of digital words, the accuracy of the digital word is greater than is strictly required for a satisfactory representation of the original waveform, For example, in compact disc audio, audio signal waveforms are represented by 16-bit wordlength data sampled at 44.1 kHz, and by use of techniques such as dither and noiseshaping described in references [1]-[4] and [16]-[19], this wordlength is capable of producing a perceived dynamic range exceeding 110 dB, whereas existing technologies and consumer requirements rarely require perceived dynamic ranges of more than 90 or 100 dB. Similarly, professional digital video standards often use 10 bit words to represent video signal waveforms, which has a significant quality margin above the quality found acceptable to most viewers.

In such situations, it may be desired to reallocate some of the information used to transmit the waveform to instead transmit and receive other data signals. For example, in compact disc audio, only two channels of stereo information are conveyed by the audio words, whereas systems of sound reproduction and recording using three or more related channels of audio information are found to be subjectively preferable to conventional two-channel stereo. It may therefore be desired to reallocate some of the data in the compact disc audio words to transmitting additional audio channel signals. Many other applications, some of which

are detailed in section 1 below, exist for using data reallocated from the existing waveform information to other uses.

However, such reallocation of data is required to be compatible with reproduction systems which are not designed specifically to take account of such re-allocation of data. For example, it is undesirable in compact disc audio to reallocate audio word data to other uses if this results in a markedly degraded sound when the disc is played on existing players.

Methods are already known in the prior art of reallocating waveform word data to other uses, in which the reallocated data produces a waveform error designed to be minimally perceivable. For example, in ref. [5] there is described a method whereby an audio signal is divided into many sub-bands.

This prior-art sub-band method has many disadvantages. Firstly, the process of both encoding and decoding is a complicated one requiring a high level of signal processing complexity. Secondly, there is an inherent time delay in the signal processing involved in splitting signals into subbands. Thirdly, the data rate that can be encoded by the sub-band method is reduced for small input waveform to a low level.

A particular disadvantage of the sub-band method is that it relies on models of auditory masking perceptually to hide the error caused by data coding in the signal waveform words, coding the data at levels within particular sub-bands that are determined adaptively by a model for auditory masking to be masked by the audio signal. Such masking models are still imperfect, and moreover, masking thresholds are not deterministic but probabilistic in nature, so that there is a finite probability of error signals below a masking threshold being detected by the

ears. Thus the waveform degradations produced by the sub-band method will in general produce audible effects that may not be acceptable for the highest quality uses.

Another approach to transmitting data in an audio waveform, for use with the NICAM system, has been described by Emmett [22], in which the shape of the error spectrum is adaptively changed to be masked by the audio signal. The present proposal does not require the use of such level-adaptive data rates.

A further disadvantage of the sub-band method is that in practice it is not efficient in an information-theoretic sense.

SUMMARY OF THE INVENTION

The invention allows coding of data within the digital words representing signal waveforms such that the coding is efficient in the sense of information theory, thereby minimising added error noise levels, involves only short or zero time delays in the signal processing, and in which nonlinear distortion and data-related error variation effects are avoided, and also allows if desired avoidance of all modulation noise effects as well. The invention also allows the spectral characteristics of the error noise to be modified so as to minimise its perceptual level, which in general depends on the spectral characteristics.

These advantages of the invention permit data to be encoded within the words of signal waveforms at a higher data rate and with less waveform degradation than was possible in the prior art.

According to the invention in a first aspect, there is provided a method of encoding digital data within digital words representing signal waveforms, including the step of modifying least significant digits of said digital words

representing signal waveforms in dependence upon said digital data,

characterised by pseudo-randomising said digital data thereby forming data noise words having levels small relative to those of said waveform words,

subtracting the pseudo-randomised data words from said waveform words thereby producing a dithered waveform word, and

quantizing said dithered waveform word and adding said data noise word to said quantized word thereby forming an output of reduced noise carrying information representing digital data in the least significant digits thereof.

The method may be implemented using:

means of receiving input digital waveform words representing input waveform signals,

means of receiving input data information,

means for outputting output digital waveform words representing an output waveform signal and incorporating data information,

means for pseudo-randomising said data information and for forming it into a word signal termed the data noise signal having a level or range of levels small relative to that of the waveform words,

means for subtracting said data noise signal from digital words representing said input waveform producing dithered waveform words,

means for uniformly quantizing said dithered waveform words, and

means for adding said data noise signal to the output of said uniform quantizing means to produce output digital words,

wherein least significant digits of the digital words representing signal waveforms are replaced in the output digital words by information representing said data information in a pseudo-randomised form.

The least significant digits of a digital word may be the digits in a binary representation of the word, or the least significant digits in representations of the word in any other integer base or bases.

In a preferred implementation of the invention in this aspect, there is provided noise shaping means around said uniform quantizing means adapted to modify the spectrum of the difference between output and input waveform signals in a desired predetermined manner.

In one preferred implementation of the invention in its first aspect, said uniform quantizer means may be a uniform vector quantizer for a plurality n of signal channels in the sense defined below, and said data noise signal may be a vector noise signal in said plurality n of signal channels.

In preferred implementations of the invention in its first aspect, the difference between output and input waveform signals has the form of a noise signal substantially free of nonlinear distortion products related to the input waveform signal, because the data noise signal has a probability distribution function adapted to subtractively dither the uniform quantizer with substantially no resulting nonlinear distortion.

Additionally preferred implementations of the invention in its first aspect provide for encoding data at a constant data rate and the difference between output and input waveform signals has the form of a noise signal substantially free of nonlinear distortion products related to the input waveform signal, and substantially free of variations in statistics dependent on the encoded data or on the input waveform signal.

According to the invention in a second aspect, there is provided a means for decoding data information encoded into the least significant digits of digital waveform words representing waveform signals, comprising

means for receiving said digital waveform words,

means of separating said least significant digits from said digital waveform words,

means for inverting pseudo-random encoding in said least significant digits to provide data information, and

means of outputting said data information.

According to the invention in a third aspect, there is provided a system for encoding and decoding data information within the least significant digits of digital waveform words representing waveform signals, comprising

encoding means according to the above first aspect,

decoding means according to the above second aspect,

and transmission means for conveying the output of said encoding means to the input of said decoding means.

The said transmission means may, by way of example, be a wire or optical link, or a link using radio, acoustic or infra red waves, or may be via a storage medium such as memory storage media, hard disc media, magnetic tape or optical disc recording, storage and playback media or any sequential combination of these.

When applied to audio CD (compact disc), the invention provides a new method for burying a high data rate data channel (with up to 360 kbit/s or more) compatibly within the data stream of an audio CD without significant impairment of existing CD performance. A proposal in this description is to replace a number (typically up to four per channel) of the least significant bits (LSBs) of the audio words by other data, and to use the psychoacoustic noise shaping techniques associated with noise shaped

subtractive dither to reduce the audibility of the resulting added noise down to a subjective perceived level equal to that of conventional CD.

Simply replacing the LSBs of existing audio data would, of course cause a drastic audible modification of the existing audio signal for two reasons :

1) the wordlength of existing signals would be truncated to (say) only 12 bits, which would not only reduce the basic quantization resolution by 24 dB, but also would introduce the problems of added distortion and modulation noise caused by truncation (e.g. see refs. [1-4]).

2) Additionally, the replaced last (say) 4 LSBs would themselves constitute an added noise signal, which itself may not have a perceptually desirable random-noise like quality, and will also add to the perceived noise level in the main audio signal, typically increasing the noise by a further 3 dB above that due to truncation alone, giving in this case as much as 27 dB degradation total in noise performance.

The invention incorporates methods of overcoming all these problems in replacing the last few LSBs of an audio signal by other data. The new method involves the following preferred steps:

A) Using a pseudo-random encode/decode process, operating only on the LSB data stream itself without extra synchronizing signals, to make the added LSB data effectively of random noise form, so that the added signal becomes truly noise-like.

B) Using this pseudo-random data signal as a subtractive dither signal (e.g. see [1-4]), so that simultaneously it does not add to the perceived noise and that it removes all nonlinear distortion and modulation noise effects caused by truncation. Remarkably, and unlike in the ordinary subtractive dither case [3], this does not require the use

of a special subtractive dither decoder, so that the process works on a standard off-the-shelf CD player, and

C) preferably additionally, at the encoding stage, incorporating psychoacoustically optimized noise shaping of the (subtractive) truncation error, thereby reducing the perceived truncation noise error by around 17 dB further.

The overall effect of combining these three processes is that if one incorporates data into the last few LSBs, then the effects of distortion, modulation noise and perceived audible patterns in the LSB data are completely removed, and the resulting perceived steady noise is reduced by around 23 dB below that of ordinary unshaped optimally dithered quantization to the same number of bits. For example, when the last 4 LSBs of the 16 bit CD wordlength is used for buried-channel data, the perceived S/N (signal-to-noise ratio) is around 91 dB - approximately the same as ordinary 16 bit CD quality when unshaped dither is used.

The result of this process is that as much as 2 x 4 = 8 bits of data per stereo sample is available for buried data without significant loss of audio quality on CD, giving a data rate of 8 x 44.1 = 352.8 kbit/s.

While the new process achieves potentially high data rates for the buried channel, it does of course reduce room for improvements in CD audio quality approaching 20 bits effective audio quality, such as described in refs. [3],[4]. However, there is no reason why the process should only be used with one fixed number of LSBs, and by reducing the data rate of the buried channel to a smaller number of LSBs, one correspondingly improves the resolution of the audio - for example achieving an effective perceived S/N of around 103 dB for a system using 2 LSBs of data per signal channel sample, with a data rate still of 176.4 kbit/s.

One can even make the number of LSBs used fractional, say, ¼ or ½ or 1½ LSBs per sample. This may be used either to precisely match the buried channel to a desired data rate, or to minimize the loss of audio quality, especially at very low data rates.

Additionally, by including in the LSB data channel itself low-rate data indicating the number of LSBs "stolen" from the main audio channels, it is possible to vary the number of LSBs stolen in a time-variant way, so that, for example, more LSBs can be taken by the buried channel when the resulting error is masked by a high-level main audio signal. The noise-shaping can also be varied adaptively at the encoding stage so that at high audio levels, the noise error is maximally masked by the audio signal, thereby increasing the data rate of the buried channel during loud passages to, in some cases, as much as 700 kbit/s.

It is also shown in this description that with stereo signals, it is possible to code data jointly in the least significant parts of the audio words of the two (or more) channels, using a multichannel version of the data encoding process involving the use of uniform vector quantizers and subtractive vector dithering by a multichannel pseudo random data signal for the dithering. The basic theory of vector dithering is described in section 5. It is shown that the vector multichannel version of the data coding process ensures left/right symmetry of any added noise in the audio reproduction, and an advantageous noise performance.

The approach in this invention is substantially different from an alternative method of burying data described in [5], which involved a process of splitting the audio signal into subbands, replacing the LSBs of the subbands with data based on auditory masking theory, and then reassembling the resulting data by recombining the subbands. Not only is

that process very complicated, with a considerable time-delay penalty in the subband encoding/decoding process, but it has to be done with extraordinary precision to prevent data errors in the band splitting and recombining process. By contrast, the present process involves little time delay, involves relatively simple signal processing, and further is such as to guarantee the lack of audible side-effects due to nonlinear distortion, modulation noise or data-related audible patterns.

1. Uses Of Buried Data

1.1 Additional audio channels

One application of a buried data channel particularly with an audio CD is to transmit alternative mixes of sound to that conveyed in the main channels. For example, a data-reduced audio signal may be conveyed using the buried data channel to convey an alternative sound mix of a piece of music particularly suited for special listening conditions, such as radio air play or use in exceptionally noisy environments such as in-car or background music use.

A further extension of such uses is in Library music, where functional music for use as backgrounds in radio, film, advertising, multi-media, audio visual or television productions is put onto CD. With existing library music, essentially only one mix can be conveyed on a track of a CD, but by incorporating in the buried data channels, additional data - compressed mixes or submixes in synchronism with the main channels, alternative mixes can be created by mixing together information from the main and buried data audio channels.

Typically, by way of example, one might have three basic stereo mixes A, B and C which may for example convey the respective rhythm, harmony and melody lines of a piece of music. The main stereo channels may contain a

pre-determined mix $a_1A+b_1B+c_1C$ for mixing coefficients a_1, b_1, c_1 for general use, and the data compressed channels may convey two further mixes $a_2A+b_2B+c_2C$ and $a_3A+b_3B+c_3C$ for mixing coefficients a_2, b_2, c_2 and a_3, b_3, c_3. After data recovery and data compression decoding of the additional audio channels, any mix of the form

$$d_1(a_1A+b_1B+c_1C) + d_2(a_2A+b_2B+c_2C) + d_3(a_3A+b_3B+c_3C)$$

may be recovered to obtain any desired mix $(a_0A+b_0B+c_0C)$ of A, B and C. This technically may be done by putting them in the 3 x 3 matrix.

$$M = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$$

multiplying the three signals $a_1A+b_1B+c_1C$, $a_2A+b_2B+c_2C$, $a_3A+b_3B+c_3C$ by its 3 x 3 matrix inverse M^{-1} to recover A, B and C and then to form whatever mix of these is required using conventional mixing methods. The encoding and decoding stages of such a proposal are shown schematically in Figure 16.

This mix down method can be used also for consumer music releases where it is desired to give to the public the ability to produce modified mixes other than the standard mix of the main audio stereo channels.

One application of this ability to provide alternative mixes is the possibility of providing a choice of languages for vocals in a music release aimed at a multi-lingual market, with the main channels conveying one language, and the subsidiary channels conveying, for example the difference between the vocals in the first and in a second language. Subtracting this "difference" vocal channel from the main channels will produce a track with vocals in

the second language, while still retaining the full quality of the main channel for all the backing musical lines.

1.2 Application to multichannel sound

One application of the new data channel is using the additional bits to add, using audio data compression, additional audio channels for three- or more-speaker frontal stereo or surround sound as shown in Figure 16, such as described for example in [6], [7], [8]. In using the buried channel to transmit additional directional audio channels, it is important to design the codec error signals so that they do not become audible through the mechanism of directional unmasking described in three of the inventor's references [9], [10], [11].

The data rate available is sufficient to transmit a Dolby AC-3 or MUSICAM surround 5-channel surround-sound signal, but these systems involve a quality compromise with the data rate, so that this is not a preferred procedure.

High-quality data compressed additional audio channels can, unlike existing data compression systems, minimize the risk of destruction of subtle auditory cues such as those for perceived distance, thereby maintaining CD digital audio as the preferred medium for high quality audio, while adding additional channels. For high quality (and especially musical) use, it may be preferred to use additional buried audio channels either for frontal-stage 3- or 4- speaker stereo or for 3-channel horizontal or 4-channel full-sphere with height [13] ambisonic surround sound (see refs. [7], [8], [15]), rather than for the rather cruder theatrical "surround-sound" effects considered appropriate for cinema or video-related surround-sound systems. However, systems have been proposed for intercompatible use of both kinds of system [7], [8].

Since the main audio channels in this proposal convey high-quality audio, it is possible to use the spectral envelope of the main audio channels to convey most or all of the dynamic ranging information used for the subbands in data reduction systems for related subsidiary channels conveyed in the buried data channel, especially if the main audio channels incorporate a mixture of all the transmitted channels so that no direction is canceled out. This saves the data overhead of conveying ranging data, which in high quality systems may save of the order of 60kbit/s, as compared to a stand-alone data compression system. This will allow a system conveying n related channels using 4 LSBs per main CD audio channel to give a performance equivalent to that of a stand-alone data compression system conveying n-2 channels in about 420 kbit/s. For 3-channel systems, such as horizontal B-format surround-sound or 3-channel UHJ [15] or frontal-stage 3-channel stereo, this quality is unlikely to be audibly distinguishable from an uncompressed data channel, and for 4-channel systems, the results will still subjectively approach that of critical studio-quality material, and even for 5-channel material, the results will be considerably less compromised than that for DAB or cinema surround-sound, using a data rate for the additional channels of well over twice that used in those applications.

1.3 Video and computer data

Alternatively, the buried data channel can be used for conveying related computer data, such as graphics, data files, computer games, multilingual text or track copyright information and a data rate of 350 kbit/s is even enough to convey a reasonable video image by using a video data reduction system such as MPEG.

1.4 Dynamic range data

Another use would be to convey dynamic-range reduction or enhancement data, e.g. a channel conveying the setting of

a gain moment by moment. This would allow the same CD automatically to be played with different degrees of dynamic compression according to environment, by choosing the gain adjustment channel appropriate to that environment. This would include the possibility of completely uncompressed quality for high-quality use. It is known in the prior art to add buried data to a CD to control the switching on or off of a dynamic expander to alter the dynamic range of sounds, or to control the characteristics of such a dynamic expander. The present proposal differs in that the buried data channel is used to convey a signal representing the actual gain to which the audio waveform signal is to be subjected to moment by moment. By this means, the gain need not be rigidly specified by any particular design of compressor or expander, but may be chosen freely from that derived by many different kinds of compressors or expanders, or even derived from manual gain adjustment by an artistically skilled operative.

The gain signal may be conveyed by any known method. For example, it might by conveyed using say 12 successive bits in a data signal to convey using PCM the value of the gain control signal to 12 bit resolution with a bandwidth limited to the Nyquist resolution of the sample rate of the 12 bit words. Preferably, the gain waveform will be coded in the data stream using Differential PCM techniqes rather than PCM techniques, since this will generally convey the gain control signal with a higher resolution at a given data rate. Well-known techniques of efficient data transmission such as Huffman coding may be used to maximise the gain control signal resolution within the available data rate.

The decoder will recover the buried data as described in the following, will recover from this, by DPCM and Huffman decoding as appropriate, the original gain control signal,

and will then alter the gain of the main audio channels by multiplying these channels by the value of the gain control signal as shown schematically in Figure 17.

Other continuous control signals can be conveyed digitally by the buried data channel in a similar manner, for example by transmitting the data in one or more channels of MIDI (Musical Instrument Digital Interface) control information. these MIDI control signals can be used to adjust the reporduciton parameters of the main audio channels by means of MIDI controlled gains, panpots and equalisers, reverberation units and similar effects devices in order to produce desired alterations for special reproduction purposes.

Such MIDI or similar control signals in the buried data channel can additionally or instead be used to cause the performance of MIDI-controlled synthesiser sound modules for the purposes of adding additional musical lines to those conveyed in the main audio waveform of the CD.

1.5 Frequency Range Extension

A further use related to the original audio is shown schematically in Figure 18 and is to add in the subchannel data-reduced information allowing information above 20 kHz to be reconstructed. It is widely noted that there is a significant loss of perceived quality cuased by the sharp bandlimiting to 20kHz when comparing high-quality digital signals sampled at say 44.1 kHz as compared to 88.2 kHz.

From a quality viewpoint, it may be more important to use an extended bandwidth to provide a more gentle roll-off rate than to provide a response flat to 40 kHz, since (unlike the brickwall filters used with ordinary CD), such gentler roll-offs are similar to those encountered in natural acoustical situations.

The extended bandwidth can be provided, for example, by using a high-order complementary mirror filter pair of the kind described in Regalia et al. [20] and in Crochiere and Rabiner [21] to split an 88.2 kHz-rate sampled digital signal into two bands sampled at 44.1 kHz. The filters involved will overlap, although using a high-order filter [20], the region of significant overlap can be reduced to of the order of a kHz. Within the overlap region there will be aliasing from the other frequency range, although the reconstruction of the full bandwidth [20,21] will cancel out this aliasing. The band below 22.05 kHz can then be transmitted as the conventional audio, and the band above 22.05 kHz can be transmitted in data reduced form in the buried data channel at a reduced data rate of, say, between 1 and 4 bits per sample per channel, using known sub-band or predictive coding methods. This arrangement is illustrated in Figure 18. Phase compensation inverse to the phase response of the low pass filter in the complementary filter pair may be employed to linearise the phase response of the main sub-22.05 kHz signal for improved results for standard listeners, with the use of an inverse phase compensating filter in the decoding process for reconstructing the wider bandwidth signal.

The potential quality problem caused by aliasing within the main audio waveform may be avoided by conveying a lower frequency range via a low pass filter that has substantially zero response above 22.05 kHz, and a higher frequency range in data-reduced form that includes some overlap of frequency range with the band below 22.05 kHz.

1.6 Combined applications

Any or all of these uses can, of course, be combined, subject only to the restrictions of the data rate, so that the buried data channel could be used for example to convey one additional audio channel, a dynamic range gain signal, extended bandwidth and additional graphics, text (possibly

in several languages), copyright and even insert video data as appropriate.

1.7 Other applications

Although for simplicity of exposition, much of the description of the invention is discussed in detail in the case of digital words representing audio signal waveforms as discrete time series and with respect to compact disc audio, it will be appreciated that the invention is not confined to this application, but may equally be applied to other cases such as video or image waveforms, or waveforms representing analog data such as seismographic or electroencephalographic waveforms, or to digital words representing waveforms in a transform domain such as a Fourier or cosine or sine transform or Hilbert transform domain, or the domains produced by the invertible sub-band or discrete transforms used in waveform coding applications.

Particularly in remote sensing applications where data waveforms have to be transmitted via a limited or expensive communications link, for example data sent from a space probe or satellite or a sensor in an oil well drill or a sensor used to gather meteorological information for use in weather forecasting, the invention may be used to convey other data in the least significant digits of the waveform data, while minimally affecting the noise performance in the waveform data.

Other aspects, embodiments, objects, uses and advantages of the invention will be apparent from the description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Embodiments of the invention will now be described by way of example, and the theoretical background to the invention

discussed, with reference to the accompanying drawings in which :-

Figure 1 shows pseudo random encoding and decoding of data transmitted via a digital channel to ensure noise-like behavior.

Figure 2 shows a binary pseudo-random sequence generator using shift-register logic, with input "exclusive or" gate for encoding and decoding of a binary data stream.

Figure 3 shows a schematic of processing of data to form an audio noise-like signal.

Figure 4 shows subtractive dither around a uniform quantizer.

Figure 5 shows subtractive dither using a combination of discrete and continuous RPDF dither.

Figure 6 shows a noise shaped subtractively dithered uniform quantizer.

Figure 7 shows an "outer" form equivalent to that of fig. 6 for a noise-shaped subtractively dithered uniform quantizer, where $H'(z^{-1}) = H(z^{-1})/(1-H(z^{-1}))$.

Figures 8a and 8b show noise shaping round pseudo random data noise signal encoding of data into an audio word using the standard noise shaper form and round a modified process.

Figures 9a and 9b show noise shaping round pseudo random data noise signal encoding of data into an audio word using an "outer" noise shaper form equivalent to fig 8 if $H'(z^{-1}) = H(z^{-1})/(1-H(z^{-1}))$, and round a modified process

Figure 10 shows a further implementation of noise shaping round pseudo random data noise signal encoding of data into an audio word.

Figure 11 shows the recovery of the data signal from the received coded audio word.

Figure 12 shows the 2-dimensional rhombic quantizer region (shaded square with sides tilted 45°) shown against a background (squares with horizontal and vertical sides) of conventional independent quantizers (whose square

quantizer region is darkly shaded) on each channel y_1 and y_2.

Figure 13 shows the use of extra subtractive dither to eliminate nonlinear distortion and modulation noise at LSB level, using noise shaped triangular PDF dither having ±1 LSB peaks to achieve good results in both nonsubtractive reproduction of output audio word and (shown) subtractive reproduction.

Figures 14 and 15 show the use of autodither to generate triangular dither in the encoder and audio decoder.

Figure 16 shows the encoding of three or more audio channels as a pair of normal audio channels on CD with the remaining information conveyed using audio data compression in the least significant digits, and the recovery of these channels and their mixing or combining to form output audio channels, either for mixdown use or for multi-channel directional sound reproduction.

Figure 17 shows the encoding and the decoding as least significant bit buried data a signal intended for optional gain alteration of the reproduced CD sound.

Figure 18 shows the encoding and the decoding of information coded in data-reduced form in the least significant bits for the increase of audio bandwidth beyond the 20 kHz limits of conventional CD.

DESCRIPTION OF EXAMPLES

A signal processing circuit for encoding data within a digitised signal waveform comprises a pseudo-random encoder 1 and a uniform quantizer 2. The pseudo-random encoder applies a reversible pseudo-random function to the input data so that the output is noise-like. The output from the pseudo-random encoder 1 is substracted from an input digital word representing a signal waveform. After subtraction, the modified waveform word is quantized by the quantizer which in this example has a uniform quantization

characteristic. A noise-shaping loop is provided around the quantizer 2 and the data noise subtraction node.

This circuit, and the alternative circuits dicsussed below, are conveniently implemented by means of signal processing algorithms programmed and implemented in ways well known to those skilled in the art on general purpose digital signal processing chips, such as those in the Motorola DSP 56000 family such as the DSP56001 or DSP56002, or chips of the Texas Instrument TMS320 family, although any digital signal processing hardware capable of performing the required arithmetic and logic operations may be used, including programmable logic chips and arithmetic logic units and general purpose central processors used in computers. Logic algorithms for pseudo-random encoding and decoding of data, such as described below in connection with Figures 2a and 2b, have in themselves been well-known for over twenty years in the prior art, and may be implemented by using standard digital logic elements to implement Boolean logic operations.

When used with general purpose digital signal processing chips, the data encoding algorithms are implemented as programs stored in program memory, operating on a time series of digital input words representing waveform signals and on data signals, and providing a time series of digital output signal words representing modified waveform words incorporating data signal information. In a similar way when used with general purpose digital signal processing chips, the data decoding algorithms are implemented as programs stored in program memory, operating on a time series of digital input words representing waveform signals incorporating data signals, and providing an output data signal information.

The digital waveform signal being processed will usually have been derived either by passage through an analogue-to-

digital converter from analogue waveform signals, or from signals directly synthesised in the digital domain. The output waveform words may be converted to analogue waveforms by use of a digital to analogue converter. Where in this description the data signal is a data-reduced signal for analogue waveforms, such as data-compressed audio or image video signals, any available hardware or software method may be used to encode the extra waveform information into a data-reduced form, and to decode the recovered buried data signal back into a waveform signal. Such hardware and software methods of encoding are available commercially for many data-reduction systems such as the MUSICAM, ISO/MPEG and Dolby AC2 and AC3 and APTX-100 and ATRAC systems for data-reduced audio signals, and for the JPEG method of data reduction for image signals and the MPEG method of data reduction for video signals.

2. Pseudo-Random Coding of Data

2.1 Pseudo-Randomized data

It is desirable, if the LSBs of an audio signal are to be replaced by data, that the replacing data should truly resemble a random noise signal (albeit perhaps one that may be spectrally shaped for psychoacoustic reasons). Most data signals, when listened to as though they were digital audio signals, have some degree of systematic pattern which may well prevent them from sounding or behaving truly like random noise. Such departures from random noise like behavior are generally much more perceptually disturbing or distracting than a simple steady noise.

Also, if we can ensure that added data behaves like a noise signal with known statistical properties, one can use all that is known in the literature on dither and noise shaping (see [1]-[4],[16]-[19]) to optimize the perceptual properties of the added data to minimize its audible effects.

The data signal is rendered pseudo-random with predictable statistics in our proposal by a data encode/decode process, the encode process having the effect of pseudo-randomizing the data signal, and the decode process having the effect of recovering the original data signal from the pseudo-randomized data signal, as in figure 1. From a practical point of view, it is highly desirable that the encode and decode process require no use of an external synchronizing signal, but that the decode process should work entirely from the pseudo-randomized data sequence itself.

The simplest way of constructing such an encode/decode pseudo-randomizing process for data is to use a cyclic pseudo-random logic sequence generator separately on each bit. For example, if its input is zero, fig.2 shows a well-known binary pseudo-random logic sequence generator using feedback around three logic elements and a total shift register delay of 16 samples (a 1-sample delay is denoted by the usual notation z^{-1}). Provided that the logic state in the 16 samples stored in the shift register is not all zero, this binary sequence generator has the 16 logic states cycle through all $2^{16}-1 = 65,535$ non-zero states in a pseudo-random manner.

If, instead of using a zero input, the pseudo-random sequence generator of fig. 2 is fed with a binary data stream s_n, then it has the effect of a pseudo-randomizer for the input data. This encoding scheme is based on the recursive logic

$$t_n = s_n \oplus t_{n-1} \oplus t_{n-3} \oplus t_{n-14} \oplus t_{n-16} , \qquad (2\text{-}1)$$

where t_n is the output binary logic value of the network at integer sample time n, s_n is the input binary logic value of the network at integer sample time n, and $\oplus$ represents the logic "exclusive or" or Boolean addition operator (with truth table $0 \oplus 0 = 1 \oplus 1 = 0$, $0 \oplus 1 = 1 \oplus 0 = 1$).

Conversely, if exactly the same arrangement of logic gates is fed with the pseudo-randomized data t_n, then the effect of the "exclusive or" gates on the input signal is to restore the original data stream. This is achieved by the inverse decoding logic process

$$s_n = t_n \oplus t_{n-1} \oplus t_{n-3} \oplus t_{n-14} \oplus t_{n-16} ,$$

(2-2)

illustrated in the second diagram in fig. 2.

Thus by using a logic network recursively with a total of L = 16 samples delay and only 4 "exclusive or" gates, a binary data stream can be pseudo-randomized, and the same network can decode the data stream back to its original form. For constant signals, there is a one in 65,536 chance that the undesirable non-random zero state will be encountered, but this low probability is probably acceptable, given that even a single binary digit change of input is likely to "jog" the system back into a pseudo-random output state.

Other well-known pseudo-random binary sequence logic generators with shift registers of longer length L than 16 samples can be used for encoding and decoding in the same way, with their fed-back output given by subjecting the delayed sequence output and the input to a "sum" logic gate. Such length L sequences will have, for a constant input, only one chance in 2^L-1 of giving an unrandomised output, and will have a sequence length of 2^L-1 samples.

Although the pseudorandom binary sequence generator described in (2-1) and fig. 2 is a maximum length sequence for a zero input, it has a shorter length for an all-one constant input, and in general, the precise behavior with, say periodic inputs is hard to predict. Partly for this reason, it is not absolutely essential to use a maximum-length sequence generator, provided that the length of the sequence is not too short for constant inputs

It will be noted that the network of fig. 2 only has L = 16 samples of memory, so that when used as a decoder, any data errors in the input will only propagate for L samples, and then the output will recover. This lack of long-term memory in the decoding process means that there are no special requirements on the error-rate of the transmission channel. Because of the small number of logic elements in fig. 2, a single sample error in the received data stream will only cause five sample errors in the decoded output.

As shown in fig. 3, typically, for use with CD, the data will first be arranged to form a number of bits of data per sample of each audio channel, for example 8 bits of data constituting bits 12 to 15 of the left and the right audio channels (where bit 0 is the most significant bit (MSB) of a 16 bit audio word and bit 15 the least significant bit).

Then each of these (say 8) bits will, separately, be encoded by a pseudo-random logic such as that of fig. 2 to form a pseudo random sequence, and the resulting pseudo-randomized bits used to replace the original bits in (say) bits 12 to 15 of the left and the right audio channels. The resulting noise signals in the left and right audio channels will be termed the (left and right) data noise signals.

Alternatively, instead of pseudo-randomizing individual bits of the audio words representing data separately, they can be pseudo-randomized jointly by regarding the successive data bits of a word as being ordered sequentially in time, and applying a pseudo-random encoder such as that of figure 2 to this sequence of bits. For example, eight bits of data per audio sample can be sequentially ordered before the next eight bits of data corresponding to the next audio sample, and the pseudo random-logic encoding can be applied to this time series of bits at eight times the audio sampling rate.

An advantage of this strategy is that errors in received audio samples propagate for (in this example) for only one eighth of the time as in the case where each word bit is separately pseudo-randomized.

M-level data signals, taking one of M possible values, conveying $\log_2 M$ bits per sample can also be pseudo-randomized by a direct process involving congruence techniques, whereby the coded version w'_n of the current sample M-level word w_n is given by

$$w_n' = w_n + \sum_{j=1}^{L} a_j w_{n-j}' \pmod{M}, \qquad (2\text{-}3)$$

where the a_j's are (modulo M) integer coefficients chosen (if necessary by empirical trial-and-error) to ensure that all M possible constant inputs result in a pseudo-randomized output with reasonably long sequence lengths. The inverse decoding of the pseudo-randomized M-level words is

$$w_n = w_n' - \sum_{j=1}^{L} a_j w_{n-j}' \pmod{M}. \qquad (2\text{-}4)$$

The logic techniques described with reference to figure 2 are just the special case when M = 2 of this more general congruence technique. The congruence technique can result in sequence lengths for constant inputs of length up to a maximum of M^L-1 samples, so that in general, the larger the value of M, the smaller need be L, with a consequent shortening of the time duration of propagation of errors.

A slightly more complex pseudo-randomization of data will provide an initial pseudo-randomization of M-level data by a method such one of those described here, and follow it by an additional one-to-one map between the M possible data values. The decoding will first subject the M levels to an

inverse map before applying the inverse of the above pseudo-random encodings.

There are many similar but more complicated methods of pseudo-randomization of data streams, and as we have seen, these need have no coding delay or increase in data rate after coding, and can limit the duration of any errors in received data in the inversely decoded output to not more than a few samples after the occurrence of an erroneous audio sample.

As audio signals, the resulting pseudo-randomized data noise signals have a steady white noise spectrum and a (discrete) uniform or rectangular PDF (probability distribution function), in the example case described above having 16 levels in each of the left and right channels. Such discrete noise does not have the ideal properties of rectangular dither noise, although Wannamaker *et al* [16] have shown that it approximates many of these desirable properties in a precise mathematical sense. However, adding to it an extra random or pseudo-random white rectangular PDF noise signal with peak level ± ½ LSB converts it into noise with a true rectangular PDF with peak levels (in this example) of ± 8 LSB. In this case the added noise to convert from a discrete to a continuous PDF is at a very low level, being 24 dB below the level of the data noise signal.

2.2 Stereo parity coding

Although in the above example, we have described data being conveyed on each audio word bit of the data signal separately, it will be realized that data can alternatively be conveyed by more complicated combinations of the least significant digits (in any numerical base M, not just the binary base 2) of audio words, for example on the Boolean sum of the corresponding bit in the left and right audio signal.

For example, consider the case that a data rate of only one bit per stereo audio sample is required. Such a signal can be conveyed as the Boolean sum of the LSB in the left and the right audio channels, leaving the values of the LSB in individual audio channels separately unconstrained. Conveying a data channel using the Boolean sum of the corresponding bits of the left and right audio signals is herein termed stereo parity coding.

It is of course desirable that the effect on the conventional audio of reallocating bits to a buried data channel should be left/right symmetrical. In particular, if a buried data channel is used with a data rate of just one BPSS (bit per stereo sample), then one does not wish to code the data in the LSBs of only one of the two stereo channels. If the value of the respective N'th bits of the respective left and right channel signals are denoted by L^N_n and R^N_n at time n, then one codes a pseudo-randomized one bit per sample data channel t^N_n as

$$t^N_n = L^N_n \oplus R^N_n \quad . \tag{2-5}$$

If desired, an additional second pseudo-randomized one bit per sample data channel u^N_n can be encoded in the N'th bits of the stereo audio signal say as

$$u^N_n = L^N_n \quad . \tag{2-6}$$

in which case the data can be encoded via $L^N_n = u^N_n$ $R^N_n = L^N_n \oplus t^N_n$. and decoded via $u^N_n = L^N_n$, $t^N_n = L^N_n \oplus R^N_n$. Alternatively u^N_n can be encoded as R^N_n . The use of stereo parity encoding allows the separate one BPSS data channels to be separately decoded while maintaining left/right symmetry in the audio when an odd number of one BPSS channels are used.

One could standardize a basic one BPSS data channel as being conveyed via the parity (Boolean sum) of the LSBs (i.e. bit 15) of the left and right audio channels. Information about the way other data channels conveying more BPSS are coded will, in such a standardization, be

conveyed by this basic data channel. By this means, a data decoder can read from the basic one BPSS stereo parity data channel how to decode any other data channels (if any) present. In particular, this allows if desired moment-by moment variation of the data rate, either adaptively to the amount of data needing transmission or adaptively to the audio signal according to its varying ability to mask the error signal caused by the hidden data channels.

For example, in loud passages in pop/rock music, the data rate allocated to say a video signal could be increased, allowing quite high quality video images in, say, heavy metal music.

2.3 Fractional bit rates

There is no reason why the buried data channels should be restricted to data rates of an integer number of BPSS, although this may be a convenient implementation. Several methods can be used to allocate less significant parts of audio words to data at fractional bit rates.

One method conveys $\log_2 M$ bits for integer M in the less significant parts of audio words by conveying data in the M possible values of the remainder of the integer audio word after division by M, whereas the rounding quantization process used for the audio involves rounding to the nearest multiple of M. For M a power of 2, this reduces to conventional quantization to $\log_2 M$ fewer bits.

In Eqs. (2-3) and (2-4) above, we described how such M-level data channels can be pseudo-randomized by pseudo-random congruence encoding and decoding.

Alternatively, if M can be expressed as nontrivial product of K = two or more integer factors

$$M = \prod_{j=1}^{K} M_j,$$

then one can uniquely expand the M-level data word w in the form

$$w = \sum_{k=0}^{K-1} w_{(k)} \prod_{j=1}^{k} M_j \qquad (2\text{-}7)$$

with $w_{(k)}$ an integer between 0 and $M_{k+1}-1$. Eq. (2-7) is the generalization of the expansion of a number to base M_0 in the case $M_j = M_0$ for all $j = 1,\ldots,K$. Each of the expansion coefficients $w_{(k)}$ can, if desired, be separately pseudo-randomized before the final length M word is formed. Again, this generalizes the binary case described above where the M_j's equaled 2.

A second method for fractional bit rates especially suitable for very low data rates of 1/q BPSS for integer q is to code data only in one out of every q audio samples. The encoding schemes are as before but with a data sampling rate divided by q, and decoding involves the decoder trying out and attempting to decode each of the q possible sub-sequences until it finds out (e.g. by confirming a parity check encoded into the data) which one carries data.

For integers $p < q$, a data rate of p/q BPSS can similarly be obtained by encoding data in the LSBs of p out of every q samples (for example, samples 1 and 3 out of every successive 5 samples for $p = 2$ and $q = 5$).

A third method for fractional bit rates also codes data in the LSBs of q successive samples, but codes the data into different logical combinations of all q bits. For example, a data rate of 1/q BPSS can be obtained by encoding data as the parity (Boolean sum) of the q LSBs. It turns out that

this option is often capable of significantly less audio noise degradation than the simpler scheme of the second method. A part of the advantage is that if one needs to modify the parity, then one can choose to modify that sample out of the q successive samples causing the least error in an original high-resolution audio signal, rather than being forced to alter a fixed sample.

We shall see in the following that, for all three kinds of fractional bit rate data encoding, it is possible to use a subtractive dithering technique by a data noise signal to eliminate unwanted modulation noise and distortion side effects on the modified waveform data. The advantages of the new process are not confined to integer bit rates per sample.

3. Subtractively dithered noise shaping

3.1 Subtractive dither

Here we briefly review the ideas of subtractively dithered noise shaping, detailed by the inventors in refs. [1], [3] and [4]. In this description, by a "quantizer" we mean a signal rounding operation that takes higher resolution audio words and rounds them off to the nearest available level at a lower resolution. We assume that the quantizer is uniform, i.e. that the available quantization levels are evenly spaced, with a spacing or step size denoted as STEP.

The quantizer rounding process introduces nonlinear distortion, but this distortion may be replaced by a benign white noise error at the same typical noise level by using the process of subtractive dither shown in figure 4. The process comprises adding a dither noise before the quantizer and subtracting the same dither noise afterwards. Provided that the statistics of the dither noise are suitable, it can be shown (see [1], [2]) that this results

in the elimination of all correlations between the error signal across the subtractively dithered quantizer and the input signal. One such suitable dither statistics is what we term RPDF dither, i.e. dither each of whose samples is statistically independent of other samples and with a rectangular probability distribution function (PDF) with peak levels ±½ STEP.

An audio word of B bits each of which is a pseudo-random binary sequence, is a 2^B-level approximation to a signal with RPDF statistics, so that the data noise signals considered above may be used as dither signals for dithering audio to eliminate nonlinear quantization distortions and modulation noise. Similarly, the M-level data noise signals described above in section 2.3 using the remainder modulo M for data, if made to be of a pseudo-random form by a pseudo-random data encoding/decoding process, can be used as an M-level approximation to RPDF noise.

Although data noise signals are discrete approximations to RPDF noise, they can be converted to continuous RPDF noise statistics by the simple process of adding to them an additional smaller RPDF noise with peak levels ±½ LSB, where LSB is the step size of the LSB's of the transmitted audio words (as distinct from the step size STEP = M LSBs of any rounding process used in encoding hidden data channels.) This is shown schematically in figure 5.

Conventionally, as described in refs. [1] and [3], the use of subtractive dither requires the use of a decoding process in which during playback, the original dither noise added before the quantizer is reconstructed before being subtracted; this requires either the use of synchronized pseudo-random dither generation algorithms, or an encode/decode process in which the dither noise is generated from the LSB's of previous samples of the audio

signal [3]. However, in the application of this invention, as will be seen, no special dither reconstruction process is required for the discrete dither, since this is already present in the transmitted LSBs.

3.2 Noise shaping

A white error spectrum is not subjectively optimum for audio signals, where it is preferred to weight the error spectrum to match the ears' sensitivity to different frequencies so as to minimize the audibility or perceptual nuisance of the error. The spectrum of the error signal may be modified to match any desired psychoacoustic criteria by the process of noise shaping, discussed for example in refs. [1], [4], [12], [17]-[19].

Noise shaping may be static (i.e. adjusting the spectrum in a time-invariant way) and made to minimize audibility or optimize perceptual quality at low noise levels, or alternatively it can be made adaptive to the audio signal spectrum so as to be optimally masked by the instantaneous masking thresholds of audio signals at a higher level. The latter option is particularly valuable in the present application, where loud audio signals may well allow an increased error energy to be masked, thereby allowing a higher data rate to be transmitted in the hidden data channels during loud audio passages.

The form of noise shaping with subtractive dither that may be used in this description is indicated in the schematic of figure 6. It will be noted that, while it is equivalent to some of the forms described in ref. [1], it is not the arrangement described previously by the inventors in ref. [3], in that here we put the noise shaping loop around the whole subtractive process. With the arrangement of figure 6, the output of the quantizer itself differs from the noise shaped output of the whole system by a spectrally white dither noise, so that in this arrangement, unlike

those suggested in ref. [3], the spectral shape of the quantizer output error and system output error is not identical.

With the noise-shaped subtractively dithered quantizer of fig. 6, the error feedback filter $H(z^{-1})$ must include a 1-sample delay factor z^{-1} in order to be implementable recursively, and the originally white spectrum of the subtractively dithered quantizer is filtered by the frequency response of the noise shaping filter

$$1 - H(z^{-1}) , \qquad (3\text{-}1)$$

which is preferably chosen to be minimum phase to minimize noise energy for a given spectral shape [1], and may be chosen to be of any desired spectral shape.

Other implementations of noise shaping around a dithered quantizer system are possible. Alternative implementations are reviewed in ref. [4]. By way of example, fig. 7 shows an alternative "outer" form of noise shaping architecture described in ref. [4], that is equivalent to fig. 6 if one puts

$$H'(z^{-1}) = H(z^{-1})/(1-H(z^{-1})). \qquad (3\text{-}2)$$

The application of noise shaping around a subtractively dithered quantizer will not result in any unwanted nonlinear distortion or modulation noise, provided that the dither noise added in figs. 6 or 7 is RPDF dither matched to the step size STEP of the quantizer.

4. Application to buried data channels

4.1 Noise-shaped subtractively dithered buried channel encoding

Either the arrangement of fig. 6 or fig. 7 can be applied to obtain subtractively dithered noise-shaped audio results when the last digits of an audio signal word (whether the

last N binary digits or the remainder after division by M) are replaced by buried data bits.

The procedure is now simple to describe. First the data is pseudo randomized, and then used to form a data noise signal as described above. This data noise signal has (discrete M-level) RPDF statistics, and may be used as the dither noise source in figures 6 or 7, as shown in figs 8 and 9. where the quantizer is simply the process of rounding the signal word to the nearest integer multiple of M LSB's (or the nearest level if the levels are placed uniformly at other than the integer multiple of M LSB's). The process shown in figures 8 or 9 subtracts the data noise signal from the audio at the input of the uniform quantizer (which has step size STEP = M LSBs), and adds it back again at the output of the quantizer so as to make the least significant digits of the output audio word equal to the data noise signal. Noise shaping is performed around this whole process.

For best results using the algorithms of figs. 8 or 9 (or equivalent algorithms such as that in figure 10 below), it is best if the input audio word signal is available at a higher resolution or wordlength than that used in the output, since this will avoid cascading the rounding process used in figs. 8 or 9 with another earlier rounding process. By making the input signal available at the highest possible resolution, any overall degradation of signal-to-noise ratio is minimized.

Since the output equals the output of the quantizer plus the data noise signal, the noise shaping has no effect on the information representing the data in the output audio word, but merely modifies the process by which the quantization of the audio is performed so as to minimize the perceptual effect of the added data noise on the audio. It is remarkable that this output signal, being the output

of a noise-shaped subtractively dithered quantizer, automatically incorporates all the benefits of noise shaped subtractive dither without the audio-only listener needing any special subtractive decoding apparatus.

Moreover, because the information received by the data-channel user is not dependent on the noise shaping process, the noise shaping can be varied in any way desired without affecting reception of the data (provided only that no overflow occurs in the noise shaping loop near peak audio levels - fitting a clipper in the signal path before the quantizer to prevent this may be desirable). Thus the noise-shaping process does not affect the way the signal is used by either audio or data end-users of the signal, and so does not need any standardization, but may be used in any way desired by the encoding operative to achieve any desired kind of static or dynamic noise shaping characteristic.

Other equivalent noise-shaped dithering architectures may be used in place of those shown in figs. 8 and 9 for encoding the data signals into the output audio word, using the kind of equivalent architectures discussed in ref. [4]. Purely by way of example, fig. 10 shows yet another implementation having identical performance to that shown in figs. 8 or 9. It is also evident that in a similar way, the data noise signal can be added and subtracted outside the "outer" noise shaper of fig. 9 rather than inside the noise shaper as shown.

Although noise shaping is preferably used for systems of adding buried data according to the methods of subtractive dither by the buried data as described above with reference to figs. 8 to 10, it may also be applied to those systems in which the buried data is not subtracted before the quantiser but only added after the quantizer, for example as in figs. 8b and 9b, where the subtraction node of

Figures 8a and 9a immediately before the quantizer is omitted, or where the signal fed to this node is conventional additive pseudo-random dither noise rather than the psuedo-randomised data signal. Omission of the subtracted data noise signal or its substitution by conventional dither noise at the node before the quantizer typically loses some of the quality advantages of the preferred process for burying data, typically increasing reproduced noise levels in the modified digital word by 6dB. Nevertheless, if such a procedure is adopted, its subjective performance will be maximised if a noise shaping such as illustrated in Figures 8 or 9 is used around the process in which pseudo-random dither is added before the quantizer and the data noise signal is added only after the quantizer.

Explicit coefficients for the noiseshaper filter $H(z^{-1})$ that may be used to reduce the audibility of buried data on compact disc and other audio media at sampling rates of 44.1 or 48 kHz, with or without audio pre-emphasis are described in reference 12.

It will of course be realised that subtracting a data noise signal before a quantizer and adding it after the quantizer is equivalent to adding a polarity-reversed data noise signal before the quantizer and subtracting it after the quantizer. Since a polarity reversed data noise signal conveys exactly the same data information as the original data noise signal, it will be seen that all descriptions of the invention are equivalent to the case where a data noise signal is added before the quantizer and subtracted after it. Which realisation is adopted in practice is purely a question of convenience.

4.2 Buried Channel Decoding

Optimum recovery of the audio channels involves no need for any kind of decoder in this proposal. Playback is

conventional, with the effect of subtractive dither by the data noise signal being automatic as described above.

Recovery of the buried data is also straightforward, simply being recovery of the data noise signal by rejecting highest bits of the received audio word, or in the case of M-level data, the inverse process to the encoding, of reading the remainder of the audio word after division by M, i.e. resolving the least significant digits of the audio word via modulo M arithmetic. This is followed by the inverse pseudo-random decoding process to recover the data before pseudo randomization, and then the data is handled as data in the usual way. This decoding process is shown schematically in figure 11.

In the case that the data is encoded as integer coefficients $w_{(k)}$ with more than one base M_j as in Eq. (2-7) above, the data is recovered by K successive divisions by M_1 to M_K, at each stage discarding the fractional part, the K coefficients $w_{(k)}$ being the integer remainders of the division by M_{k+1}. This is the same process shown in figure 11, but with K stages of the modulo division.

5. Vector quantization and dither

5.1 Reasons for digression

The above descriptions of the use of noise shaped subtractive dithering also apply to the stereo parity coding case as well. To see this, we need first to look at vector quantization and vector dithering, and show that exactly the same ideas for subtractive dithering, noise shaping and data encoding can be applied to the vector quantizer case as the scalar case described above.

The description here is given in greater generality than needed just for the stereo parity coding case, since it has applications to coding information in the parity of the

corresponding bits in 3 or more channels in transmission media carrying more than two audio or image channels, for example in the 3 channels containing the 3 components of a color image.

5.2 Uniform vector quantizers

As briefly indicated in earlier papers [1], [3], [9], the concepts of additive and subtractive dither can be applied to vector as well as scalar quantizers. Vector quantizers quantize a vector signal $\boldsymbol{y}$ comprising n scalar signals $(y_1,\ldots,y_n)$ in geometrical regions covering the n-dimensional space of n real variables. As in the scalar case, we shall say that a vector quantizer Q is a uniform quantizer if the signal y is quantized to a point of a discrete grid G of quantization vectors $\{\boldsymbol{y}_g : g \in G\}$, where there exists a region $\boldsymbol{C}$ around $(0,\ldots,0)$ of n-dimensional space such that the regions $\boldsymbol{y}_g + \boldsymbol{C} = \{\boldsymbol{y}_g + \boldsymbol{c} : \boldsymbol{c} \in \boldsymbol{C}\}$ cover without overlap (except at their boundary surfaces) the range of signal variables y being quantized. Thus a uniform vector quantizer divides the n-variable space into a grid of identical vector quantization cells that are translates of the cell $\boldsymbol{C}$ to the points of the grid G, and quantizes or rounds any point in the cell $\boldsymbol{y}_g + \boldsymbol{C}$ to the point $\boldsymbol{y}_g$.

There are many examples of uniform vector quantizers, the simplest of which has a hypercubic cell $\boldsymbol{C}$ = the region $\{(c_1,\ldots, c_n) : |c_i| \leq \frac{1}{2}\ \text{STEP}\ \forall\ i = 1,\ldots,n\}$, i.e. separate scalar quantization of the n variables. The grid G in this case is simply points of the form $(m_1\text{STEP}, m_2\text{STEP},\ldots,m_n\text{STEP})$ for integer m_j's, and the associated vector quantizer is simply that that takes $(y_1,\ldots,y_n)$ to $m_j = \text{integer}(y_j/\text{STEP})$ for $j = 1,\ldots,n$. This case is trivial in the sense that it is equivalent to using separate uniform scalar quantizers on each of the n channels.

A more complicated but easily visualised example is the 2-channel case where C is a regular hexagon in the plane, for example the region consisting of the points (c_1, c_2) in the plane such that

$$|c_1| \le \tfrac{1}{2}\,\mathrm{STEP},\ |-\tfrac{1}{2}c_1 + (\sqrt{3}/2)c_2| \le \tfrac{1}{2}\,\mathrm{STEP},\ |-\tfrac{1}{2}c_1 - (\sqrt{3}/2)c_2| \le \tfrac{1}{2}\,\mathrm{STEP}, \quad (5\text{-}1)$$

and the grid G is the centers of the hexagons in the honeycomb grid covering the plane, i.e. G is the points

$$((m_1+\tfrac{1}{2}m_2)\mathrm{STEP},\ (\sqrt{3}/2)m_2\mathrm{STEP}) \quad (5\text{-}2)$$

for integer m_1 and m_2.

A uniform vector quantizer of particular interest and practical use in n dimensions is what we shall term the rhombic quantizer. This starts off with a conventional hypercubic grid G_C of points at positions $(m_1\mathrm{STEP}, m_2\mathrm{STEP}, \ldots, m_n\mathrm{STEP})$, where STEP is a step size, and m_1 to m_n are integers, which of course has the hypercube quantizer cell described just above and corresponds to the use of n separate scalar uniform quantizers. However, we the produce a new grid $G \subset G_C$ which consists of just those grid points in G_C with $m_1+\ldots+m_n$ having even integer values. This new grid only has half as many points as the original, and can be equipped with a new vector quantization cell C as follows, which we shall term the n-dimensional rhombic quantizer cell.

The rhombic quantizer cell can be described geometrically by thinking of the original hypercubic cells as being colored white if $m_1+\ldots+m_n$ is even and black if $m_1+\ldots+m_n$ is odd, forming a kind of n-dimensional checkerboard pattern of alternately black and white hypercubes. Then attach to each white hypercube that "pyramid" portion of each adjacent black hypercube lying between the center of the black hypercube and the common "face" with the white hypercube. The resulting solid is the rhombic cell C.

It is evident, since the pyramid portions taken from adjacent black hypercubes are in total enough to form one black hypercube if pieced together, that the volume occupied by the rhombic quantizer cell is twice that occupied by the original hypercube quantizer cell, and that the versions of the rhombic quantizer cell translated by the grid G indeed cover the n-dimensional n-parameter vector signal space.

For n = 2, the rhombic quantizer cell C is a diamond-shape, being a square whose sides are rotated 45° relative to the channel axes, as shown in fig.12. For n = 3, the rhombic quantizer cell C is a rhombidodecahedron, a 12-faced solid whose faces are rhombuses. For n = 4, the rhombic quantizer cell C is a regular polytope unique to 4 dimensions termed the regular 24-hedroid .

Calculations involving quite complicated multidimensional integrals, which we shall not detail here, show, for a given large number of quantizer cells covering a large region of n-dimensional space, that for n = 2, rhombic quantization has the same signal-to-noise ratio (S/N) as conventional independent quantization of the channels, but that for $n \geq 3$, rhombic quantizers give a better S/N than conventional independent quantization of the channels. The improvement reaches a maximum of about 0.43 dB when n = 6. This improvement in the S/N is maintained when subtractive dither is used as described below. The hexagonal 2-channel quantization described above gives a 0.16 dB better S/N than independent quantization of 2 channels.

Mathematically, the rhombic quantizer has grid G consisting of the points

$$(m_1 \mathrm{STEP}, m_2 \mathrm{STEP}, \ldots, m_n \mathrm{STEP}), \qquad (5\text{-}3a)$$

where the m_i have integer values with

$m_1+\ldots+m_n$ having even integer values. (5-3b)

The rhombic cell C is that region of points $(c_1,\ldots,c_n)$ satisfying the n(n-1) inequalities

$$|c_i+c_j| \le \text{STEP},\ |c_i-c_j| \le \text{STEP},$$

(5-4)

for $i \neq j$ selected from $1,\ldots,n$. The associated uniform vector quantizer rounds a vector signal $(y_1,\ldots,y_n)$ by an algorithm whose outline form might be

```
m'_i := integer(y_i/STEP),
If m'_1+...+m'_n is even
     then m_i := m'_i for all i = 1,...,n,
     else      c_i := y_i/STEP - m'_i,
(*)       d_j : = sgn(c_j) if |c_j| > |c_i| for all i < j and
|c_j| ≥ |c_i| for all i > j
          d_i := 0 for all other i,
          m_i := m'_i + d_i for all i = 1,...,n.
     End If
```

(5-5)

The function x → integer (x), where, e.g., x=y/STEP in the above formula, here is that "rounding" function that takes a number x to the nearest integer value, i.e. which takes x to the integer part of x + 1/2 by discarding the fractional part of x + 1/2.

There are, of course, various equivalent forms for this kind of rhombic quantizer algorithm, a computationally demanding aspect on typical signal processors being the determination in line (*) of that j for which $|c_j|$ is biggest.

In the n = 2 case, there is a simpler rhombic quantization algorithm as follows

```
x_1 := y_1+y_2,   x_2 := y_1-y_2 ,
m'_1 := integer(x_1/((⌊2)STEP))
m'_2 := integer(x_2/((⌊2)STEP))
m_1 := m'_1 + m'_2,   m_2 := m'_1 - m'_2,
```

(5-6)

which is based on the observation that the rhombic quantizer cell for n = 2 is the same shape as the square

cell used for ordinary independent quantization of the two channels, but rotated by 45° and with an increase of the step size by a factor $\sqrt{2}$. (See fig. 12).

5.3 Subtractive vector dither

The concepts of dithering for uniform quantizers developed in refs. [1-4] for scalar uniform quantizers may be applied also to the vector case by using appropriate vector dithers. An n-signal dither noise vector $(n_1,..,n_n)$ is said to have a uniform probability distribution function in a region C of n-dimensional space if its joint probability distribution function is constant within the region C and zero outside it. This is the n-dimensional generalization of rectangular PDF dither for vector signals, and we denote the associated n-vector dither signal by $\mathbf{r}_C$.

It can be shown (we omit any proofs here) that if the subtractive dither arrangement of figure 4 is used for modifying an input vector signal, where the "uniform quantizer" becomes a vector uniform quantizer with quantization cell C, and the dither noise becomes a uniform PDF vector dither $\mathbf{r}_C$ on the region C, then the output vector signal of the system is free of all nonlinear distortion and modulation noise effects i.e. the first moment of the output signal error is zero, and the second moment independent of the input signal [4]. Moreover, this is still the case if any statistically independent additional noise is added to the uniform PDF dither noise $\mathbf{r}_C$ on the region C.

Moreover, noise shaping can be applied around such subtractive dither in exactly the same way as before, as shown in figs. 6 and 7, or in equivalent noise shaping architectures, the only difference being that any filtering is now applied to n parallel signal channels. It is also possible, if desired, to use an n x n matrix error feedback filter $H(z^{-1})$ or $H'(z^{-1})$ to make the noise shaping dependent

on the vector direction, for example to optimize directional masking of noise by signals [9], [10].

It is possible to generate uniform PDF vector dither $\mathbf{r}_C$ over the rhombic cell C described above by an algorithm such as the following: First generate, for example by the well-known congruence method, n statistically independent rectangular PDF dither signals r_i (i = 1,...,n) with peak values ±½ STEP, and also generate an additional two-valued random or pseudorandom signal u with a value of either 0 or 1. Then the values of the noise signal $\mathbf{r}_C = (v_1,..,v_n)$ are given by:

If u = 0

then $v_i := r_i$ for all i = 1,...,n,

else $d_j := \mathrm{sgn}(r_j)$ if $|r_j| > |r_i|$ for all i < j and $|r_j| \geq |r_i|$ for all i > j

$d_i := 0$ for all other i,

$n_i := r_i - d_i\mathrm{STEP}$ for all i = 1,...,n.

End If. (5-7)

However, in applications of subtractive dither, this algorithm may involve unnecessary complication, since it can be shown that with the subtractive dither arrangement of fig. 4 with a uniform vector quantizer with quantization cell C, that a uniform PDF vector dither signal $\mathbf{r}_D$ may be used for *any other* uniform quantization cell D sharing the same grid G, and will still eliminate nonlinear distortion and modulation noise in the output. Whatever the shape of the other quantization cell D used for the dither signal, the resulting error signal from the subtractive dither arrangement of fig. 4 is a noise signal with uniform PDF statistics on the quantizer cell C of the uniform vector quantizer used.

This can allow a much simpler algorithm to be used for generating the vector dither in which uSTEP is added to (or subtracted from) just one of the n rectangular PDF noise

components. For example, a uniform PDF vector dither noise signal $\mathbf{r_D} = (v_1,..,v_n)$ given by

$v_1 := r_1 - uSTEP$

$v_i := r_i$ for $i = 2,...,n$.

(5-8)

may be used to subtractively dither the above rhombic quantizer.

6. Refinements of the basic proposal.

6.1 Further developments

The encoding process described above will work well as it stands, but does not incorporate various desirable refinements which we shall now describe. These include methods to take account of the fact that the data noise signal has a discrete and not a continuous PDF dither, and applications involving stereo parity coding.

6.2 Non-discrete dither

The fact that the dither given by the data noise signal has an M-level discrete probability distribution function rather than a continuous RPDF means that there is still unwanted quantization distortion at the level of the LSB of the audio word which is not properly dithered. Preferred methods of adding "non-discrete" dither or, strictly speaking, dither at a significantly high arithmetic accuracy such as implemented using 24 or 32 bit arithmetic are now described. The method of adding such dither shown in fig. 5 is not preferred for three reasons:

(i) Optimum playback requires subtractive decoding of the ±½ LSB RPDF dither signal, with all the usual problems of implementing subtractive dither [1], since unlike the discrete data noise signal, this is not explicitly transmitted in the audio word.

(ii) the ±½ LSB RPDF dither signal added before the quantizer does not eliminate modulation noise in non-

subtractive playback, having the wrong statistics for this purpose [2] , and

(iii) if the whole system is noise shaped as in figs. 6 or 7, the nonsubtractive listener will hear the ±½ LSB RPDF dither signal as having a white spectrum not affected by the noise shaping, so will perceive an increase in noise level.

A correct way of adding extra dither to avoid nonlinear quantization distortion and modulation noise at the ±½ LSB level is shown in figure 13. The dither used has a triangular PDF with peak levels ±1 LSB (so-called TPDF dither) with independent statistics at each discrete time instant, so as to eliminate modulation noise in nonsubtractive playback [2], and is added before the quantizer in the noise shaping loop, but not subtracted in the noise shaping loop. This ensures that the added noise in nonsubtractive playback is noise shaped.

Subtractive playback of the extra dither is done, also as shown in fig. 13 by reconstituting the triangular ±1 LSB PDF dither at the playback stage, passing it through a noise shaping filter $1 - H(z^{-1})$, and subtracting the filtered noise from the output audio word. Subtractive playback of course reduces the extra noise energy caused by the non-discrete dither by a factor 3, although this will only be highly advantageous when the data noise signal has fairly low energy, e.g. at a data rate of 1 BPSS.

The triangular dither signal may be generated, in encoding, as proposed in the "autodither" proposal of ref. [3] by means of a pseudo-random logic look-up table or a logic network having the effect of a pseudo-random look-up table, from the less or least significant parts of the *output* audio word in the last K previous samples, where typically K may be 24, and can be reconstructed from the same audio word at the input of the system by the same look-up table

or logic in the decoding stage. This is shown in the system of fig. 13 in fig. 14.

Although figures 13 and 14 are shown for the particular noise shaping architecture of fig. 6, similar ways of adding the extra triangular dither can be used with any other equivalent noise shaping architecture such as the outer form of figure 7 and fig. 10, again by adding the triangular dither just before the quantizer and subtracting it, via a noise shaping filter $1 - H(z^{-1})$, only at the output of the decoder. It is clear that the points at which dither signals are added can be shifted around in various ways without affecting the functionality of the system.

A disadvantage of the methods for adding ±1 LSB triangular PDF dither shown in figs. 13 and 14 is that in these schematics, the noise shaping filter $1-H(z^{-1})$ used for the triangular PDF dither and for the quantizer is identical. Especially in systems of subtractive dither where the noise shaping of the subtracted dither in the decoder is desirably standardised (see ref. [3]), this would not allow use of noise shaping around the quantizer with a non-standardised characteristic, such as for example noise shaping adaptive to the signal waveform level and spectrum to take advantage of auditory masking by the signal.

An alternative shown in fig. 15 avoiding this disadvantage uses a first possibly fixed or standardised filter $1-H_1(z^{-1})$ for the ±1 LSB triangular PDF dither noise subtractive decoding, but now uses the same filter in the encoding, and instead adds this filtered ±1 LSB triangular PDF dither noise to a point before the noise shaping loop. The noise shaping loop around the quantizer may then use a second possibly different error feedback filter $H_2(z^{-1})$ in place of $H(z^{-1})$ to achieve any desired predetermined quantizer

noise shaping characteristic $1-H_2(z^{-1})$, including ones adaptive to the signal waveform.

6.3 The stereo parity case

Suppose we have 2-channel stereo signals in which data is encoded pseudorandomly in bit N for all N = 15 to say 15-h+1 (where the integer h may typically be any integer from say 0 to perhaps 6 or 8, the case 0 being the case of no bits being encoded) of the left and right audio words, and data also being encoded in the stereo parity (Boolean sum) of bit 15-h of the left and right audio words, as described in subsection 2.2 above.

Based on the results on uniform vector quantization and subtractive vector dither of section 5 above, the noise-shaped subtractive encoding of the data described above in the scalar case for individual audio channels may be applied to this case too with just two reinterpretations of the above:

(i) The uniform quantizer used in figs.6-10 now becomes a uniform 2-dimensional rhombic quantizer (such as described in Algorithm (5-6) and illustrated in fig. 12) with STEP = 2^h LSB.

(ii) the "data noise signal" used for dithering is given, for example, by Eqs. (5-8) where r_i is the data noise signal of the last h bits of the i'th-channel audio word (with the first channel being say left and the 2nd channel being say right), and u being the parity of bits 15-h of the left and right audio words. In units of LSB, the data noise signal for the left channel is then $L_0 - 2^h u$ and for the right channel is R_0, where L_0 and R_0 are the respective integer words represented by the last h bits of the audio word formed by the data in the two channels.

Any alternative data noise signal may be used that represents an appropriate uniform PDF vector dither as

described in section 5.3, such as for example that given by Algorithm (5-5).

The residual nonlinear distortion and modulation noise effects at the LSB level caused by the fact that the vector data noise is discrete rather than continuous can be removed by using exactly the same technique described in subsection 6.2 and figs. 13 and 14 above by adding and, where appropriate, subtracting ±1LSB triangular PDF dither in each channel separately, the only difference being that the uniform quantizer has become a rhombic vector quantizer and the data noise signal has a modified vector form as just described.

The particular case h = 0, where data is transmitted only in the parity of the LSB of the audio word in 2 channels, simply uses the parity signal itself at the LSB level as a "data noise signal" in one of the two channels in the encoding process - it does not matter which of the two channels is chosen. With subtractively dithered playback, it turns out that the use of properly designed stereo parity coding of data, using a rhombic vector quantizer in the encoding process, gives a total noise level 1 dB lower than would the process of coding the data into the LSBs of the words of just one of the two audio channels. Thus stereo parity coding at low bit rates not only ensures audio left/right symmetry for added noise, but gives a significant noise level advantage.

6.4 Generalized stereo parity coding

There are various generalizations of the particular stereo parity coding case just described. We outline these briefly to show the applicability of these ideas to other cases.

A first generalization is that the same process may be applied to other audio wordlengths besides the 16 bit

wordlength of CD, for example to the 10 bit wordlengths of NICAM encoded digital signals or to the 20 bit or 24bit wordlengths used in some professional audio applications when it is desired to hide data in the audio words. For example, in ref. [3], the inventors described a proposal to add data at the 24th bit in studio operations on signals to detect whether or not they had been modified, and the data encoding techniques of this invention can be used in that application to minimize the audibility of the modification of the signal proposed there.

The second generalization is that one can also apply stereo parity coding to the case where one replaces the 2^h-level data in the last h bits by an M-level case for any integer $M > 1$. In this case, data is coded into the residue of the audio words of the two channels after division by M, and the "stereo parity" data channel is coded into the Boolean sum of the binary LSB in the two channels of the integer parts of the audio words divided by M. This case is handled identically to that in the previous sub-section 6.3 except that 2^h is replaced throughout by M, and the phrase "last h bits" is replaced by "residue modulo M").

A third generalization instead considers n channels rather than two. As before, this uses a rhombic quantizer in the encoding process for STEP = M LSBs, but now the n-dimensional rhombic quantizer described in (5-3) to (5-5) above, and a vector data noise signal comprising the n M-level data noise signals generated for the residue modulo M data conveyed in each of the n audio channels, to just one of which at each instant is added or subtracted uSTEP, where u is the parity (i.e. Boolean sum) of the binary LSB in the n channels of the integer parts of the audio words divided by M. Other than replacing the ordinary uniform quantizers with STEP step size by a rhombic quantizer and using the modified data noise signal, the descriptions given earlier for coding data still apply to this case.

Note that the choice of which channel of the vector data noise signal to add or subtract uSTEP, and the choice of whether to add or subtract, can be made freely, and that this choice can be made adaptively instant by instant to minimize data noise energy if desired, e.g. by making that choice which minimizes the maximum of the data noise signals in the n channels at each instant. This choice is (a discrete approximation to) that described in (5-7) for uniform PDF vector dither over a rhombic quantizer cell.

6.5 Low bit-rate case

If one has n transmitted channels of audio, then the parity of their LSBs can be used to transmit a 1 bit per n-channel-sample data channel, with remarkable little loss of S/N, especially in the case that full subtractive dithering is used at the LSB level. One might expect a loss of S/N of 6.02/n dB because the loss is shared among n channels, but for n > 2, one gets a smaller loss, typically between 0.3 and 0.4 dB better, because of the fact noted in section 5 that rhombic vector quantization has a better S/N than independent channel quantization for a given density of quantization points in the quantization grid. For n = 6, a 1 bit per n-channel-sample subtractively dithered buried data channel causes a S/N degradation of under 0.6 dB compared with a properly dithered case with no buried data channel.

Exactly the same techniques can be used to convey data via q successive samples of a monophonic signal, for example by coding into the parity of the LSBs of each successive block of q samples, as described in section 2.3. What has been shown is that by using the parity signal as a subtractive dither for any one sample with a q-dimensional rhombic quantizer, plus normal triangular additive or subtractive dither, that this fractional rate channel can be coded with a very small loss of S/N (e.g. 0.6dB for a block length q = 6), and yet with no nonlinear distortion or modulation

noise in either nonsubtractive or subtractive reproduction.

This kind of efficient low bit-rate culling of data capacity could be used, for example, with successive samples within individual sub-band channels of a sub-band data compression system. Its application is not confined to audio; culling say 1 bit per 6 10-bit video samples in a digital video recorder with a video data rate of 200 Megabits per second would give a data rate typically enough for 4 16-bit audio channels or a consumer-grade additional data-reduced video signal while losing only 0.6 dB in video S/N in the original video channel.

References

[1] M.A. Gerzon & P.G. Craven, "Optimal Noise Shaping and Dither of Digital Signals", Preprint 2822 of the 87th Audio Engineering Society Convention, New York (1989 Oct. 18-21)

[2] S.P. Lipshitz, R.A. Wannamaker & J. Vanderkooy, "Quantization and Dither: A Theoretical Survey", J. Audio Eng. Soc., vol. 40 no. 5, pp. 355-375 (1992 May)

[3] P.G. Craven & M.A. Gerzon, "Compatible Improvement of 16-Bit Systems Using Subtractive Dither", Preprint 3356 of the 93rd Audio Engineering Society Convention, San Francisco (1992 Oct. 1-4)

[4] M.A. Gerzon, P.G. Craven, R.J. Wilson & J.R. Stuart, "Psychoacoustic Noise Shaped Improvements in CD and Other Linear Digital Media", Preprint presented at the 94 Audio Engineering Society Convention, Berlin (1993 March.)

[5] W.R. Th. Ten Kate, L.M. Van De Kerkhof & F.F.M. Zijderveld, "A New Surround-Stereo-Surround Coding Technique", J. Audio Eng. Soc., vol. 40 no. 5, pp. 376-383 (1992 May)

[6] M.A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo", J. Audio Eng. Soc., vol. 40 no. 9, pp. 692-705 (1992 Sept.)

[7] M.A. Gerzon, "Hierarchical System of Surround Sound Transmission for HDTV", Preprint 3339 of the 92nd Audio Engineering Society Convention, Vienna (1992 Mar.)

[8] M.A. Gerzon, "Compatibility of and Conversion Between Multispeaker Systems", Preprint 3405 of the 93rd Audio Engineering Society Convention, San Francisco (1992 Oct. 1-4)

[9] M.A. Gerzon, "Problems of Error-Masking in Audio Data Compression Systems", Preprint 3013 of the 90th Audio Engineering Society Convention, Paris (1991 Feb.)

[10] M.A. Gerzon, "Directional Masking Coders for Multichannel Subband Audio Data Compression", Preprint 3261 of the 92nd Audio Engineering Society Convention, Vienna (1992 Mar.)

[11] M.A. Gerzon, "Problems of Upward and Downward Compatibility in Multichannel Stereo Systems", Preprint 3404 of the 93rd Audio Engineering Society Convention, San Francisco (1992 Oct. 1-4)

[12] J.R. Stuart & R.J. Wilson, "Dynamic Range Enhancement Using Noise-Shaped Dither Applied to signals with and without Pre-emphasis", Preprint presented at the 96th Audio Engineering Society Convention, Amsterdam, 26 Feb to 1 March 1994

[13] M.A. Gerzon, "Periphony: With-Height Sound Reproduction", J. Audio Eng. Soc., vol. 21, pp. 2-10 (1973 Jan./Feb.)

[15] M.A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., vol. 33 no. 11, pp. 859-871 (1985 Nov.)

[16] R.A. Wannamaker, S.P. Lipshitz, J. Vanderkooy & J.N. Wright, "A Theory of Non-Subtractive Dither", submitted to IEEE Trans. Sig. Proc. (1991)

[17] S.P. Lipshitz, J. Vanderkooy & R.A. Wannamaker, "Minimally Audible Noise Shaping", J. Audio Eng. Soc., vol. 39 no. 11, pp. 836-852 (1991 Nov.) Corrections to be published in JAES

[18] S.P. Lipshitz R.A. Wannamaker & J. Vanderkooy, "Dithered Noise Shapers and Recursive Digital Filters", to be presented at the 94th Audio Engineering Society Convention, Berlin (1993 Mar.)

[19] J.R. Stuart & R.J. Wilson, "A Search for Efficient Dither for DSP Applications", Preprint 3334 of the 92nd Audio Engineering Society Convention, Vienna (1992 March)

[20] P.A. Regalia, S.K. Mitra, P.P. Vaidyanathan, M.K. Renfors & Y. Nuevo, "Tree-Structured Complementary Filter Banks Using All-Pass Sections", IEEE Trans Circuits & Systems, vol. CAS-34 no. 12, pp. 1470-1484 (1987 Dec.)

[21] R.E. Crochiere & L.R. Rabiner, "Multirate Digital Signal Processing", Prentice-Hall Inc., Englewood Cliffs, New Jersey (1983), especially chapter 7.

[22] J.R. Emmett, "Buried Data in NICAM transmissions", Preprint 3260 of the 92nd Audio Engineering Society Convention, Vienna (1992 March)

CLAIMS

1. A method of encoding digital data within digital words representing signal waveforms, including the step of modifying least significant digits of said digital words representing signal waveforms in dependence upon said digital data,

characterised by pseudo-randomising said digital data thereby forming data noise words having levels small relative to those of said waveform words,

subtracting the pseudo-randomised data words from said waveform words thereby producing a dithered waveform word, and

quantizing said dithered waveform word and adding said data noise word to said quantized word thereby forming an output of reduced noise carrying information representing digital data in the least significant digits thereof.

2. A method according to Claim 1 further comprising applying noise shaping around said step of quantizing, thereby modifying the spectrum of the difference between output and input waveform signals.

3. A method according to Claim 1 or 2, in which said digital data is pseudo-randomised with a probability distribution function such that the difference between the output and input waveform signals has the form of a noise signal substantially free of non-linear distortion products related to the input waveform signal.

4. A method according to Claim 3, in which the probability distribution function is such that the difference between the output and the input waveform is also substantially free of variations in statistics dependent on the input waveform.

5. A method according to Claim 3 or 4, in which the probability distribution function is such that the difference between the output and input waveform signals is also substantially free of variations in statistics dependent on the encoded data.

6. A method according to any one of claims 1 to 5, in which said digital words represent audio signal waveforms, and said digital data comprises compressed data representing additional audio channels.

7. A method according to any one of claims 1 to 5, in which said digital words represent audio signal waveforms, and said digital data comprises signals for use in controlling the gain of said audio signals when reproduced.

8. A method according to any of claims 1 to 5, in which said digital words represent audio signal waveforms, and said digital data comprises one or more continuous control signals continuous in time.

9. A method according to claim 8, wherein said continuous control signals are used to modify the reproduced parameters of the output audio waveform signal as a function of time.

10. A method according to claim 9, wherein said continuous control signal is used to modify the gain of the reproduced audio waveform signal as a function of time.

11. A method according to any of claims 8 to 10, wherein said continuous control signals are MIDI control signals.

12. A method according to claim 8 or 11, wherein said continuous control signals are used to control sound production and control devices.

13. A method according to claim 12, wherein said continuous control signals are used to control MIDI-controlled musical instruments or MIDI-controlled effects devices.

14. An encoder for encoding data within digital words representing signal waveforms, comprising

means for receiving input digital waveform words representing input waveform signals,

means for receiving input digital data,

means for outputting output digital waveform words representing said waveform signal and incorporating said digital data,

means for pseudo-randomising said data and for forming a data noise signal having a level or range of levels small relative to that of the waveform words,

means for subtracting said data noise signal from said digital waveform words thereby producing dithered waveform words,

means for uniformly quantizing said dithered waveform words, and

means for adding said data noise signal to the output of said means for uniformly quantizing thereby producing output digital words representing said signal waveforms and carrying said digital data in the least significant digits thereof.

15. An encoder according to Claim 14, further comprising noise shaping means connected around said means for quantizing effective to modify the spectrum of the difference between output and input waveform signals in a desired predetermined manner.

16. An encoder according to Claim 14 or 15, in which the means for uniformly quantizing said dithered waveform words comprise a uniform vector quantizer, as herein defined.

17. A method according to any one of claims 1 to 5, in which said digital words represent audio signal waveforms over a predetermined bandwidth and said digital data represents said audio signal in the frequency range extending at least partly beyond said predetermined bandwidth represented by the digital waveform word.

18. A method of encoding data within digital words representing signal waveforms, including the steps of quantizing said signal waveform word and modifying least significant digits of said digital words representing signal waveforms in dependence upon said digital data,

characterised by applying a reversible pseudo-random function to said data prior to modifying said signal waveform word and applying noise-shaping to said quantized and modified signal waveform word thereby modifying the spectrum of the difference between the input signal

waveform word and the output modified and quantized signal waveform word.

19. A method of encoding data within digital words representing signal waveforms, including the steps of quantizing said signal waveform word and modifying least significant digits of said digital words representing signal waveforms in dependence upon said digital data,

characterised by applying noise-shaping to said quantized and modified signal waveform word thereby modifying the spectrum of the difference between the input signal waveform word and the output modified and quantized signal waveform word.

20. A method according to claims 18 or 19, in which in the step of modifying the signal waveform word data words small in level relative to the waveform words are added to the signal waveform words subsequent to the quantization of said signal waveform words.

21. A method of decoding digital data encoded within digital words representing signal waveforms by a method according to any one of claims 1 to 13, 18 to 20, comprising

receiving said digital waveform words, separating least significant digits from said digital waveform words, applying a function effective as the inverse of the pseudo-randomising function used in encoding the digital data to the least significant digits thereby recovering the data, and outputting the data.

22. An encoder for encoding data within digital words representing signal waveforms, including means for quantizing said signal waveform words and means for modifying the least significant digits thereof in dependence upon said digital data,

characterised by means for applying a reversible pseudo-random function to said data prior to modifying said signal waveform word and means for applying noise-shaping to said quantized and modified signal waveform word thereby modifying the spectrum of the difference between the input

signal waveform word and the output modified and quantized signal waveform word.

23. An encoder for encoding data within digital words representing signal waveforms, including means for quantizing said signal waveform words and means for modifying the least significant digits thereof in dependence upon said digital data,

characterised by means for applying noise-shaping to said quantized and modified signal waveform word thereby modifying the spectrum of the difference between the input signal waveform word and the output modified and quantized signal waveform word.

24. An encoder according to the claim 22 or 23, in which the means for modifying are arranged to add data words small in level relative to the waveform words to the signal waveform words subsequent to the quantization of said signal waveform words.

25. A method or apparatus according to any one of claims 18 to 20, or 22 to 24, in which said digital words represent audio signal waveforms, and said digital data comprises compressed data representing additional audio channels.

26. A method or apparatus according to any one of claims 18 to 20, or 22 to 25, in which said digital words represent audio signal waveforms, and said digital data comprises signals for use in controlling the gain of said audio signals when reproduced.

27. A method or apparatus according to any of claims 18 to 20, or 22 to 26, in which said digital words represent audio signal waveforms, and said digital data comprises one or more continuous control signals continuous in time.

28. A method or apparatus according to claim 27 wherein said continuous control signals are used to modify the reproduced parameters of the output audio waveform signal as a function of time.

29. A method or apparatus according to claim 28, wherein said continuous control signal is used to modify the gain

of the reproduced audio waveform signal as a function of time.

30. A method or apparatus according to any of claims 27 to 29, wherein said continuous control signals are MIDI control signals.

31. A method or apparatus according to claim 27 to 30, wherein said continuous control signals are used to control sound production and control devices.

32. A method or apparatus according to claim 31, wherein said continuous control signals are used to control MIDI-controlled musical instruments or MIDI-controlled effects devices.

33. A method of encoding and decoding data information within digital words representing signal waveforms comprising

encoding said digital data by a method according to any one of claims 1 to 13, 18 to 20, or 25 to 32,

outputting the resulting signal via a recording or transmission channel,

receiving the signals from the recording or transmission channel, and

decoding the received signals by a method according to Claim 21.

34. A method according to Claim 33, in which the signals output from the encoder are recorded on an audio CD.

35. An audio decoder for decoding a digital word representing an audio signal waveform encoded by a method according to claim 6, or claim 25, said digital data encoded within said digital word containing additional audio channels, the decoder comprising:

means for receiving digital waveform words,

means for separating least significant digits from said digital waveform words,

means for applying a function effective as the inverse of the pseudo-randomising function used in encoding the data thereby recovering compressed data representing additional audio channels,

means for decompressing said recovered data, and

means for outputting said audio signal waveform and said recovered additional audio channels for reproduction via a multi-channel audio reproduction system.

36. An audio decoder for decoding a digital word representing an audio signal waveform encoded according to claim 6 or 25, said digital encoded within said digital word containing additional audio channels, the decoder comprising:

means for receiving digital waveform words,

means for separating least significant digits from said digital waveform words,

means for applying a function effective as the inverse of the pseudo-randomising function used in encoding the data thereby recovering compressed data representing additional audio channels,

means for decompressing said recovered data, and

means for outputting said recovered additional audio channels for audio reproduction

37. An audio decoder for decoding a digital word representing an audio signal waveform encoded by a method according to claim 26 or 29, said digital data encoded within said digital word containing level control signals for controlling the dynamic range of said audio signal waveform when reproduced, the decoder comprising:

means for receiving digital waveform words,

means for separating least significant digits from said digital waveform words,

means for applying a function effective as the inverse of the pseudo-randomising function used in encoding data thereby recovering level control signals, and means for outputting said audio signal waveform and said level control signals for reproduction via an audio system responsive to said level control signals to modify the dynamics of the reproduced audio signal.

38. An audio decoder for decoding a digital word representing an audio signal waveform encoded by a method

according to any one of claims 1 to 5 or 17 to 20, 22, 25 to 32, said digital data encoded with said signal representing said audio signal waveform at least partially outside the bandwidth represented by said digital word, and means responsive to said decoded digital word and digital data for synthesising an audio signal of extended bandwidth for reproduction.

39. An audio decoder for decoding a digital word representing an audio signal waveform encoded according to any of claims 1 to 6 or 25 to 32, said digital data encoded within said digital word containing additional audio channels, the decoder comprising:

means for receiving digital waveform words,

means for separating least significant digits from said digital waveform words,

means for applying a function effective as the inverse of the pseudo-randomising function used in encoding the data thereby recovering compressed data representing additional audio channels,

means for decompressing said recovered data, and

means for outputting said recovered additional audio channels for audio reproduction.

40. An audio decoder according to claim 39, incorporating

means for outputting said audio signal waveforms, and

means for combining said audio signal waveforms and said additional audio channels for audio reproduction.

41. An audio decoder according to claim 40, wherein said combining means feeds an adjustable mixing means for altering the mix or sound balance of the reproduced audio signals.

42. An audio decoder according to claim 40, wherein said combined audio signals are recovered for directional sound reproduction via a multi-channel audio reproduction system.

43. A method according to any one of claims 18 to 20, in which said digital words represent audio signal waveforms

over a predetermined bandwidth and said digital data represents said audio signal in the frequency range extending at least partly beyond said predetermined bandwidth represented by the digital waveform word.

5

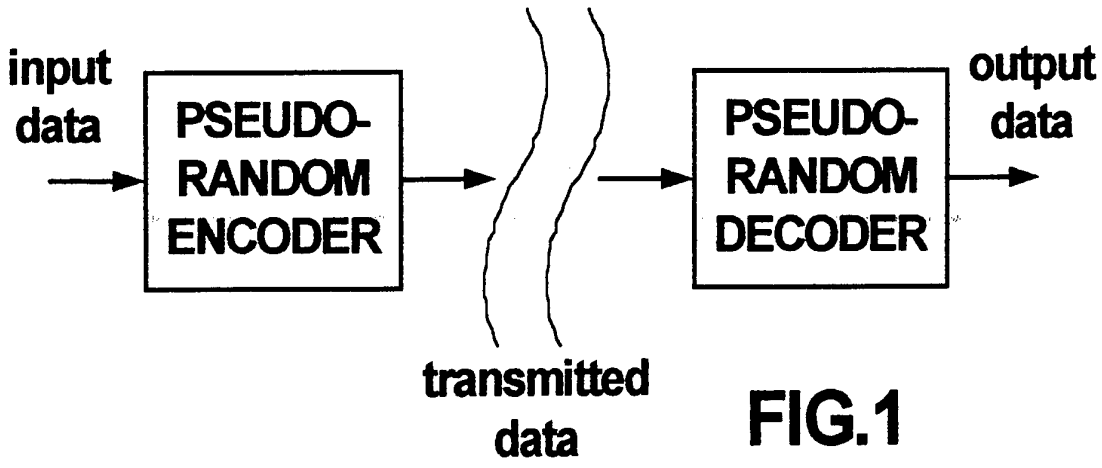


FIG.1

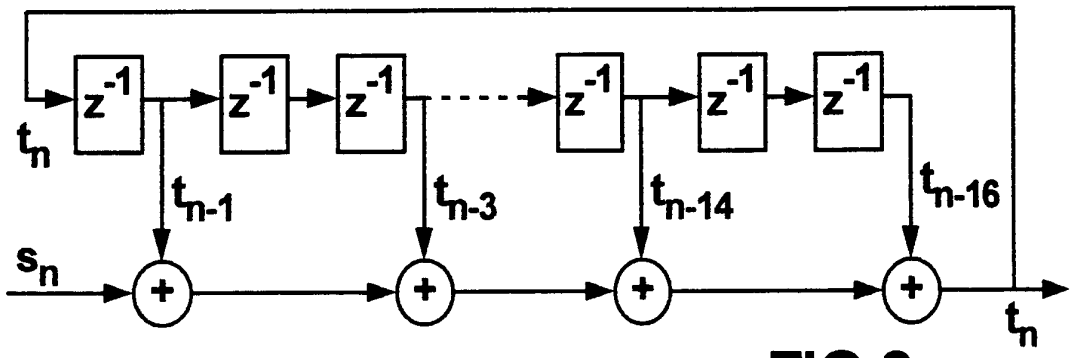


FIG.2a

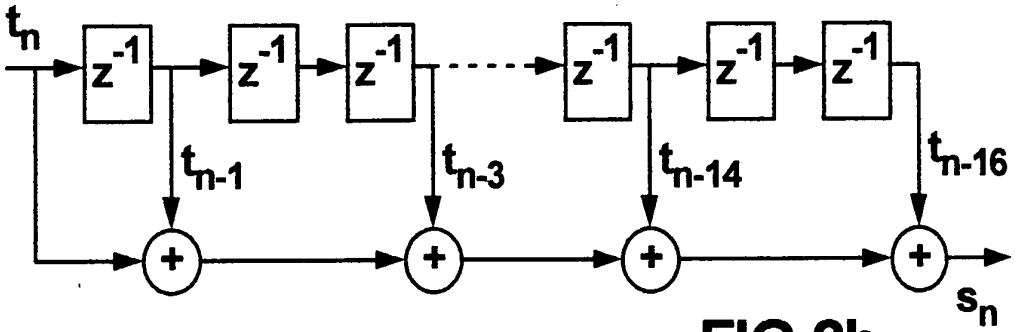


FIG.2b

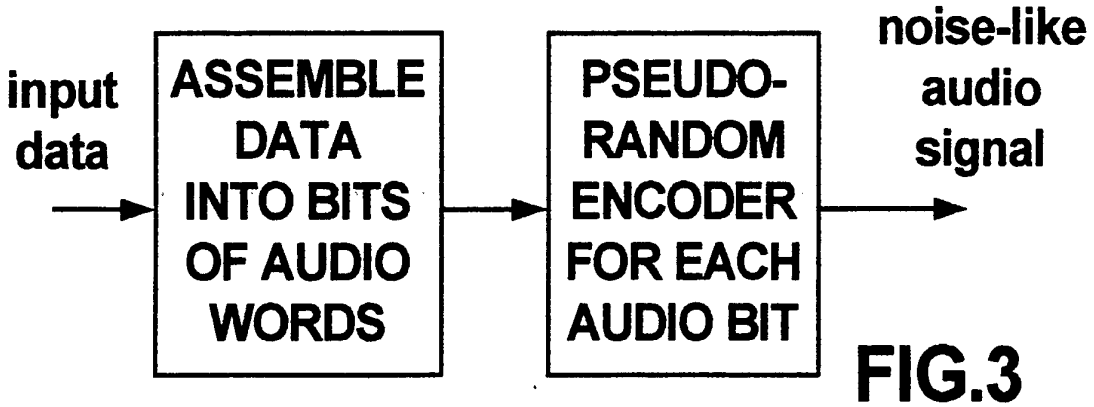

input data
ASSEMBLE DATA INTO BITS OF AUDIO WORDS
PSEUDO-RANDOM ENCODER FOR EACH AUDIO BIT
noise-like audio signal

FIG.3

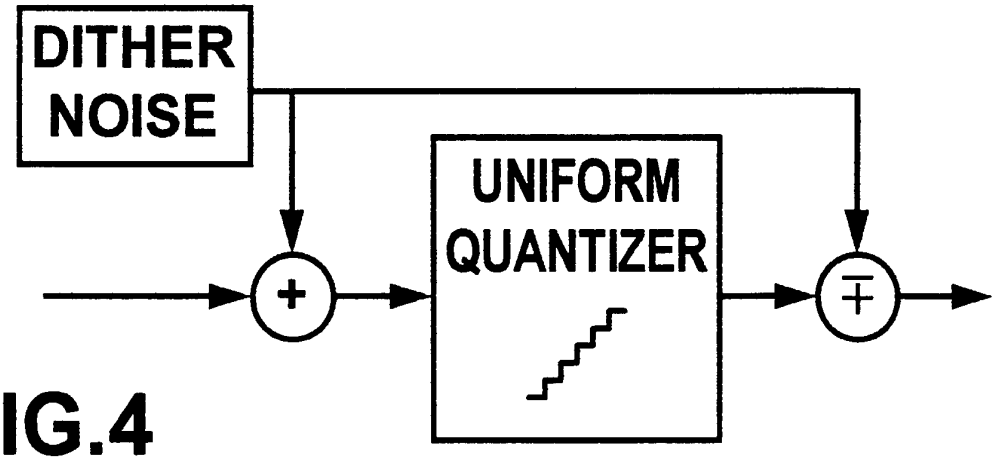

DITHER NOISE
UNIFORM QUANTIZER
+
∓

FIG.4

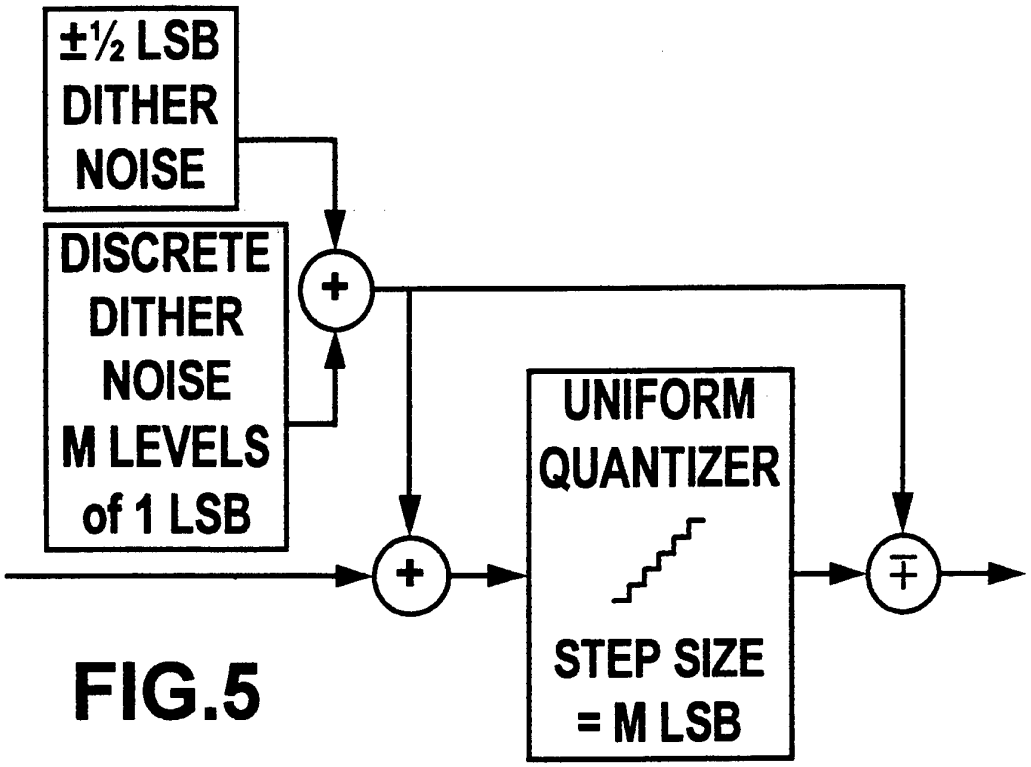

±½ LSB
DITHER
NOISE
DISCRETE
DITHER
NOISE
M LEVELS
of 1 LSB
+
+
UNIFORM
QUANTIZER
STEP SIZE
= M LSB
∓

FIG.5

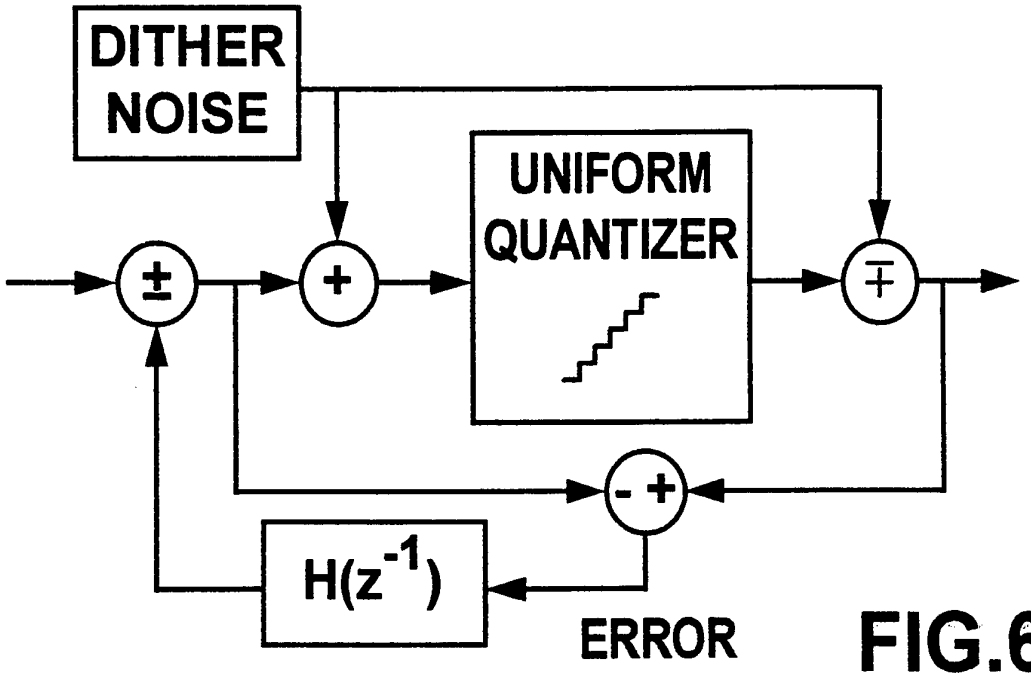

DITHER
NOISE
±
+
UNIFORM
QUANTIZER
∓
- +
H(z-1)
ERROR

FIG.6

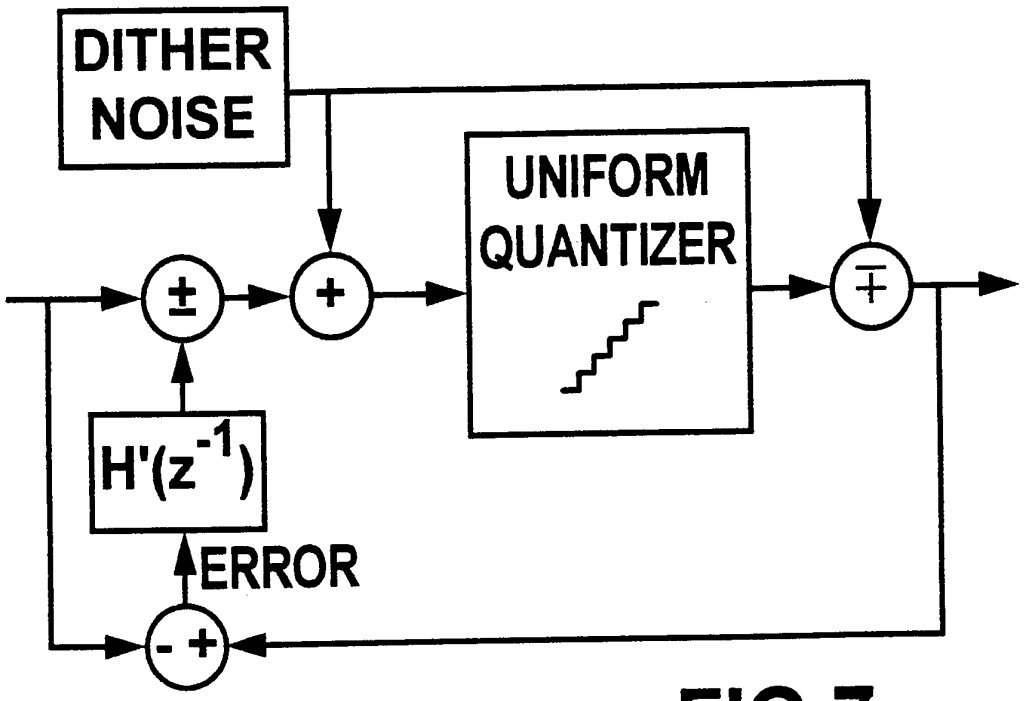

DITHER NOISE
UNIFORM QUANTIZER
±
+
∓
H'(z-1)
ERROR
- +

FIG.7

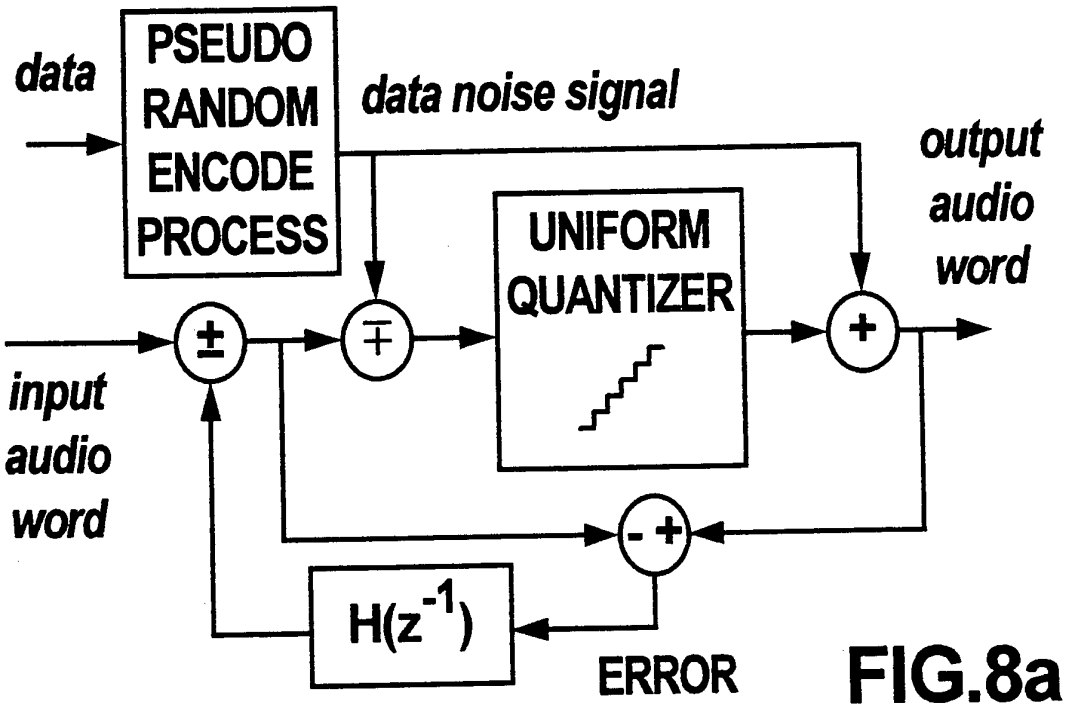

data
PSEUDO RANDOM ENCODE PROCESS
data noise signal
output audio word
UNIFORM QUANTIZER
±
∓
+
input audio word
- +
H(z-1)
ERROR

FIG.8a

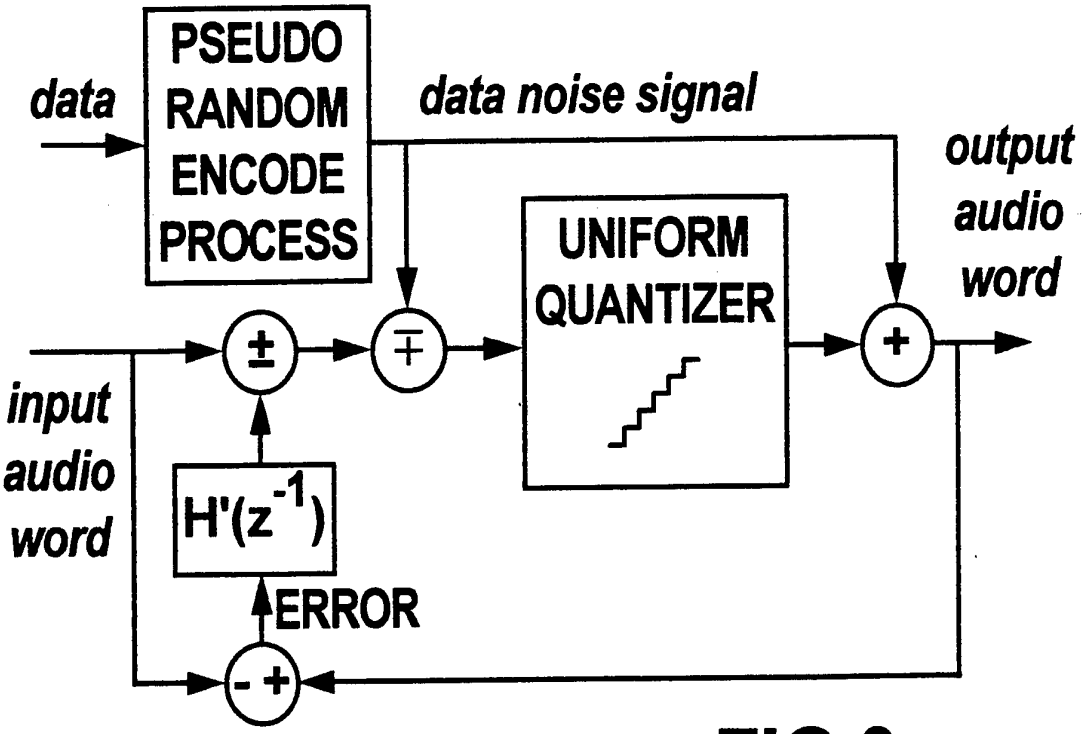

data
PSEUDO RANDOM ENCODE PROCESS
data noise signal
output audio word
UNIFORM QUANTIZER
input audio word
H'(z-1)
ERROR

FIG.9a

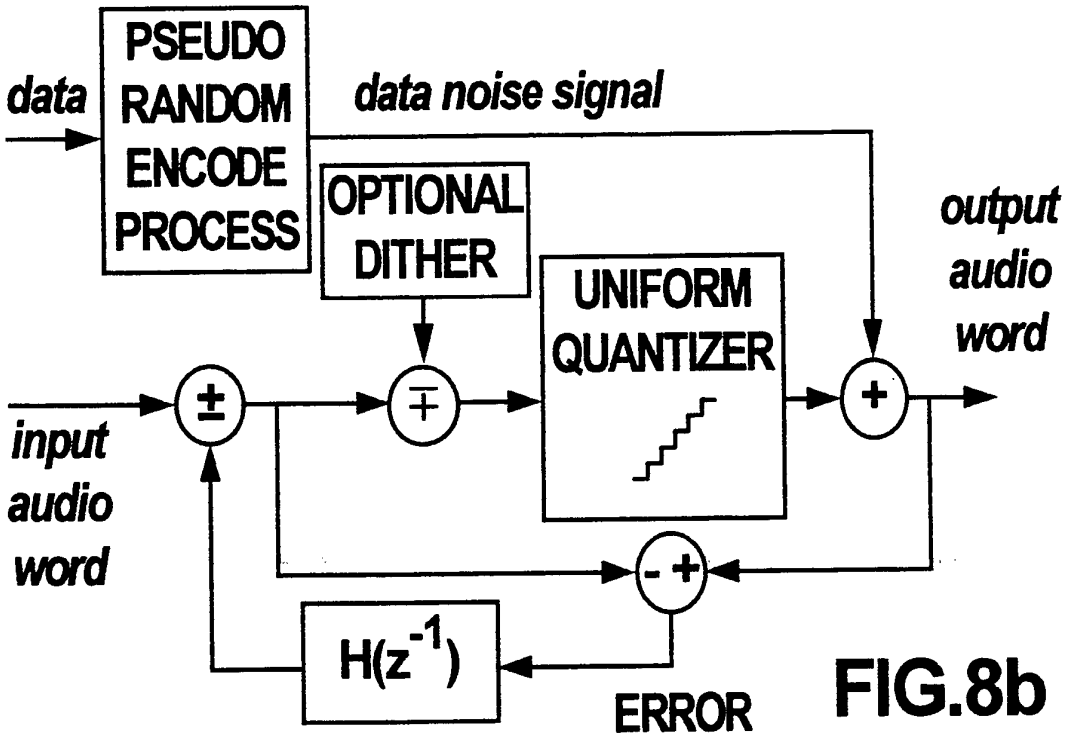

data
PSEUDO RANDOM ENCODE PROCESS
data noise signal
OPTIONAL DITHER
UNIFORM QUANTIZER
output audio word
input audio word
H(z-1)
ERROR

FIG.8b

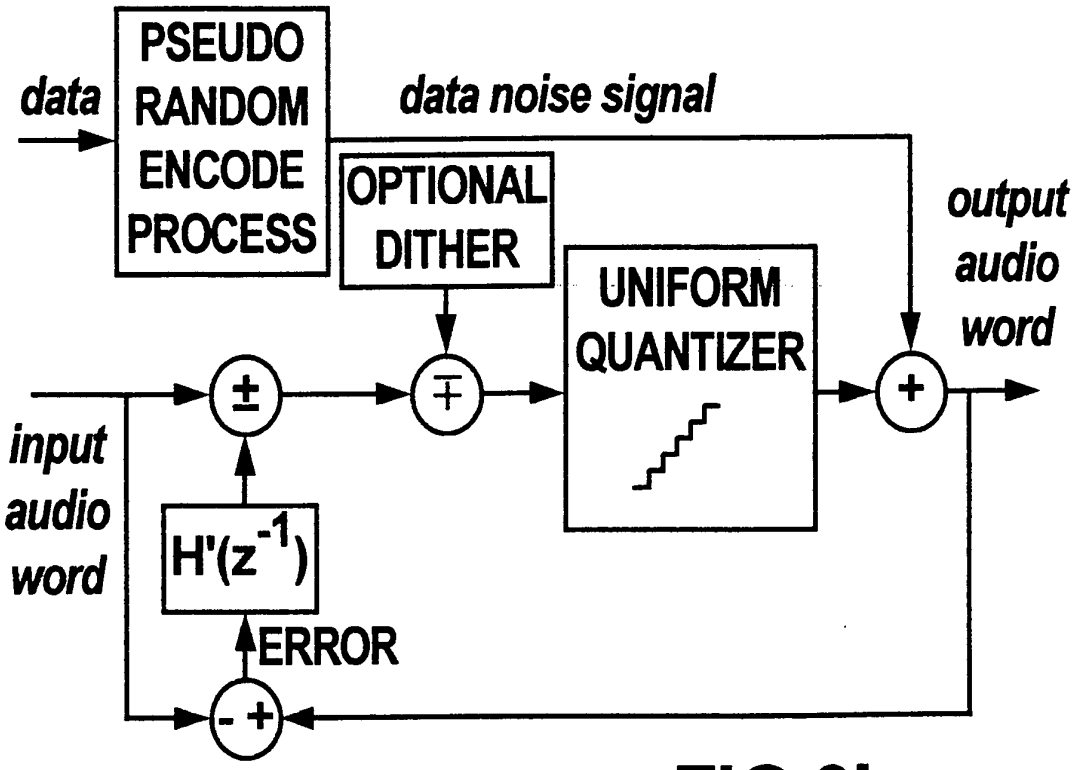

data
PSEUDO RANDOM ENCODE PROCESS
data noise signal
OPTIONAL DITHER
UNIFORM QUANTIZER
output audio word
input audio word
H'(z-1)
ERROR

FIG.9b

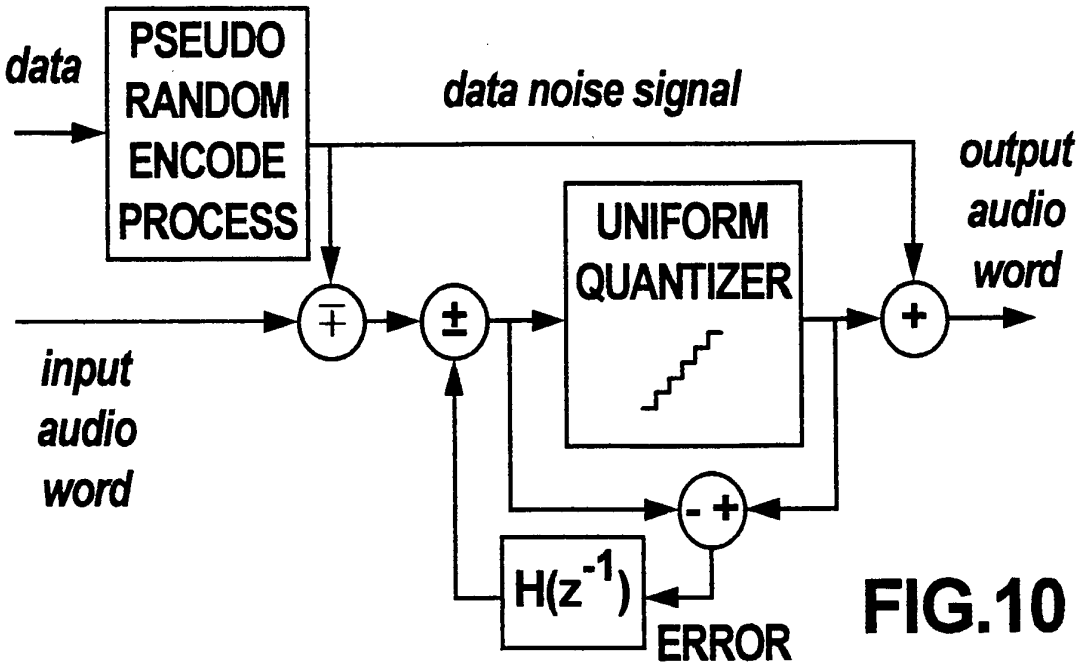

data
PSEUDO RANDOM ENCODE PROCESS
data noise signal
UNIFORM QUANTIZER
output audio word
input audio word
H(z-1)
ERROR

FIG.10

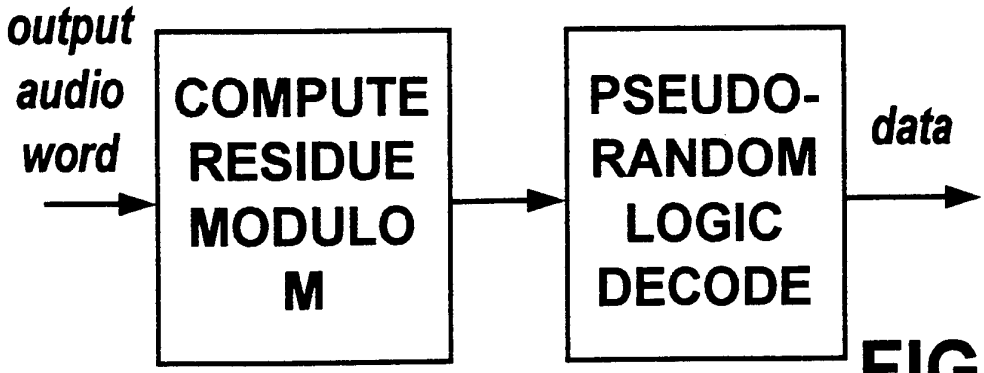

output audio word
COMPUTE RESIDUE MODULO M
PSEUDO-RANDOM LOGIC DECODE
data

FIG.11

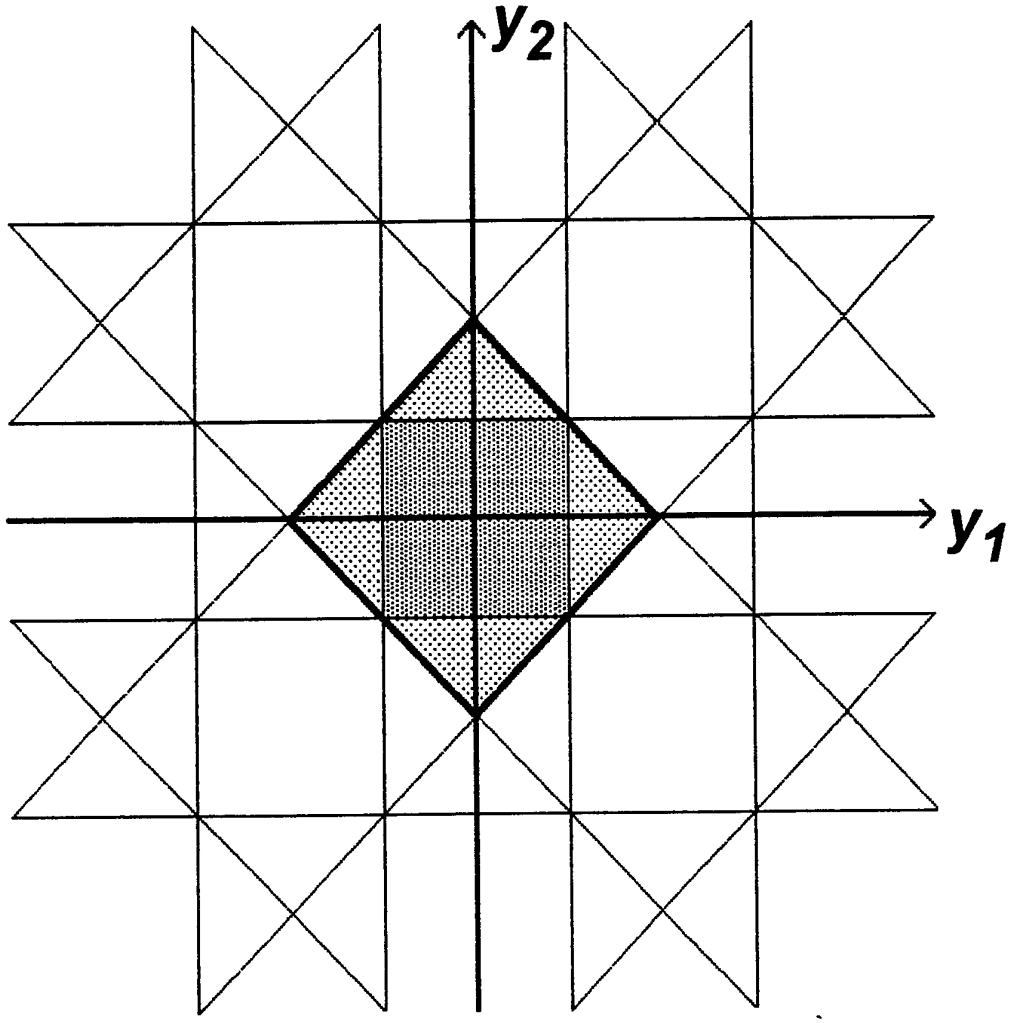

y_2
y_1

FIG.12

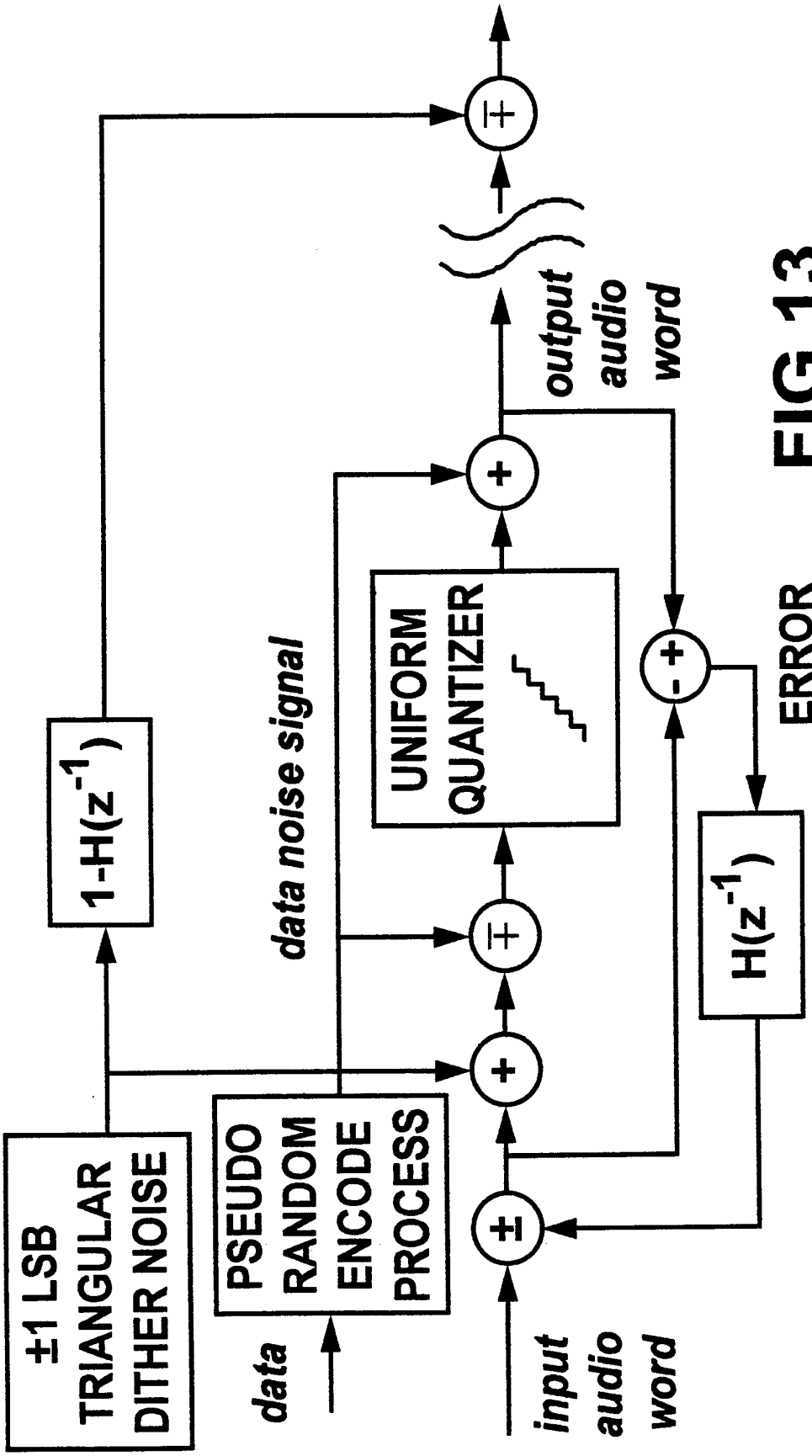

±1 LSB TRIANGULAR DITHER NOISE
PSEUDO RANDOM ENCODE PROCESS
data
input audio word
1-H(z-1)
data noise signal
UNIFORM QUANTIZER
output audio word
ERROR
H(z-1)

FIG.13

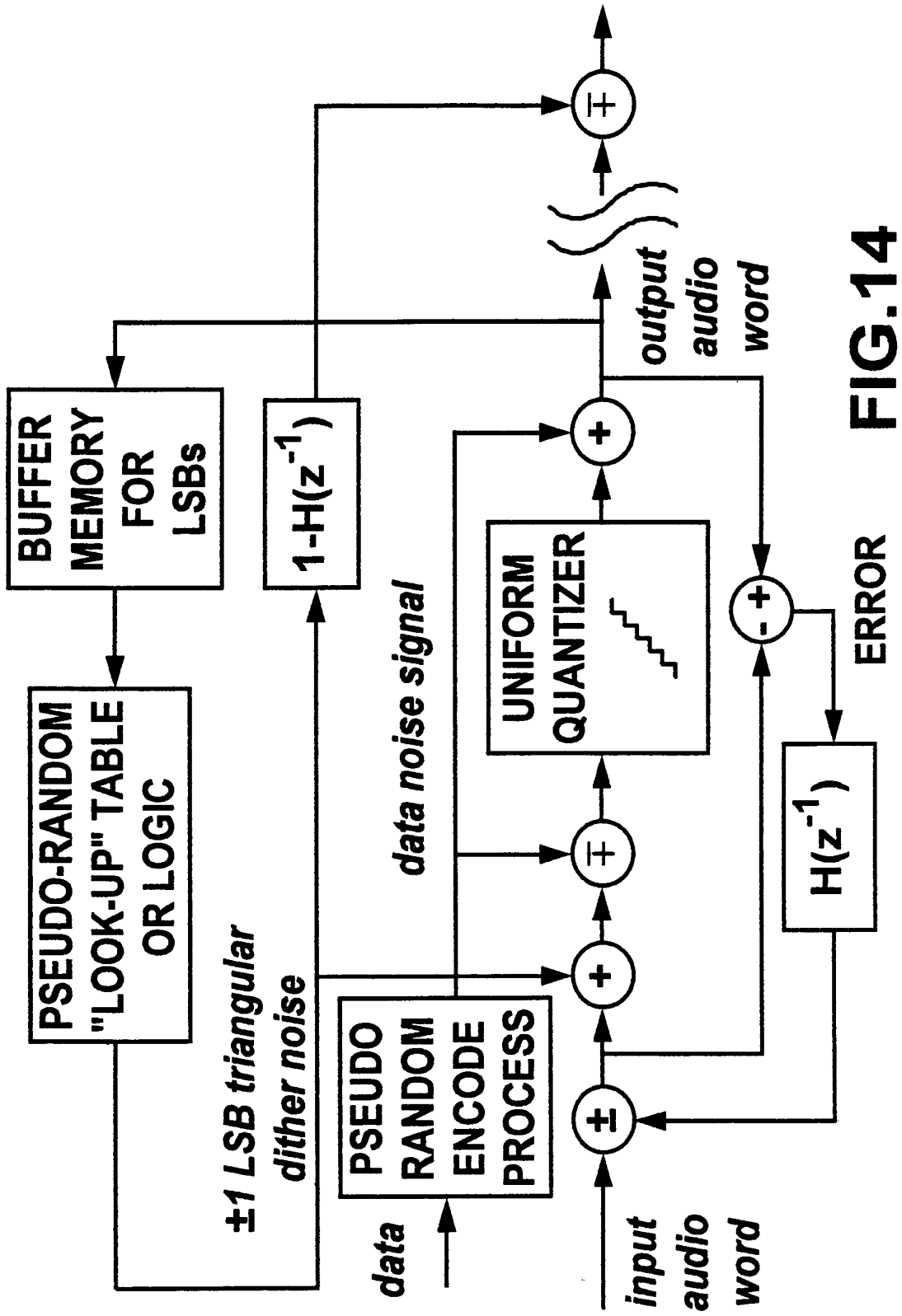

BUFFER MEMORY FOR LSBs
PSEUDO-RANDOM "LOOK-UP" TABLE OR LOGIC
1-H(z-1)
±1 LSB triangular dither noise
data noise signal
PSEUDO RANDOM ENCODE PROCESS
data
UNIFORM QUANTIZER
output audio word
ERROR
H(z-1)
input audio word

FIG.14

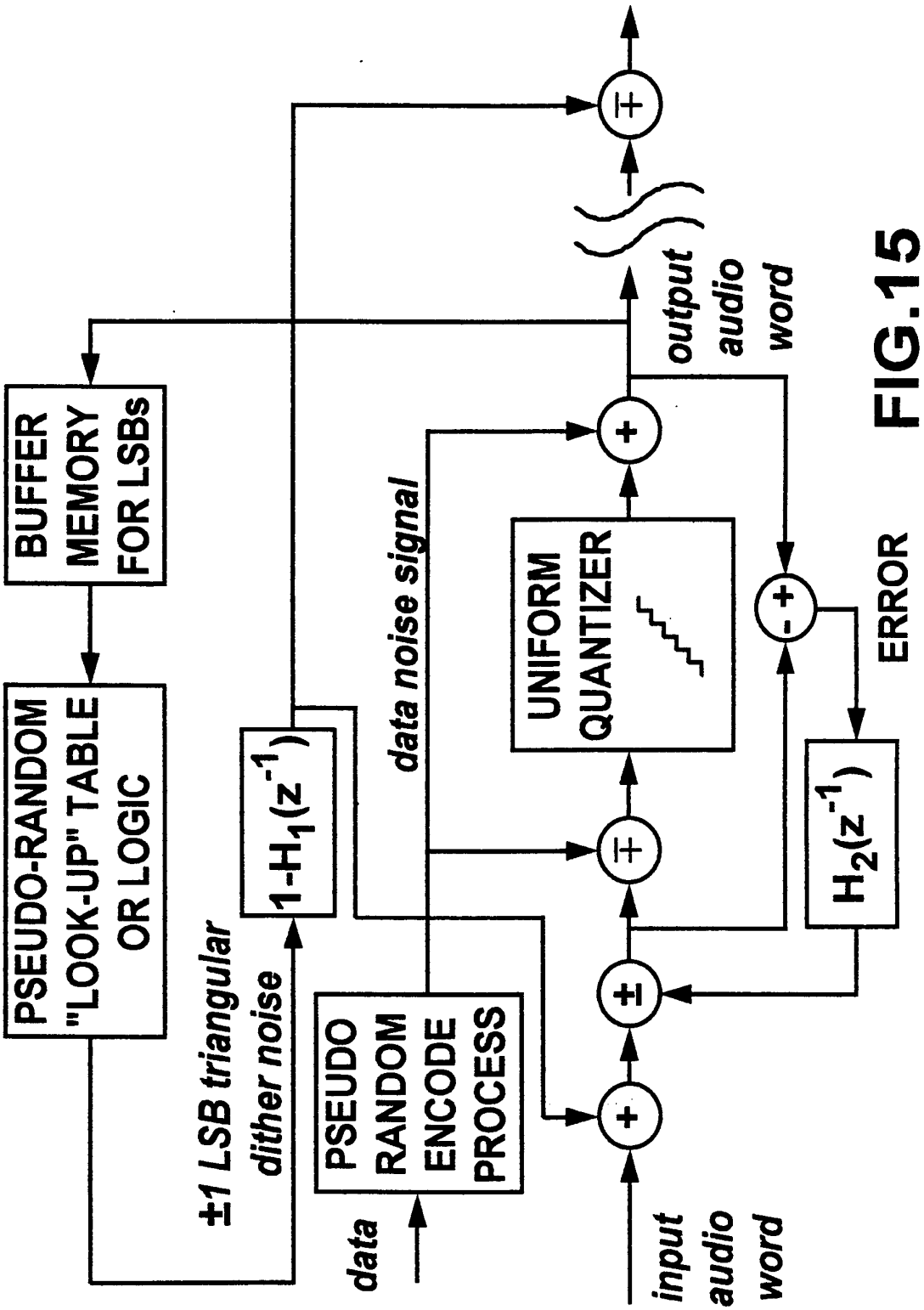

BUFFER MEMORY FOR LSBs
PSEUDO-RANDOM "LOOK-UP" TABLE OR LOGIC
±1 LSB triangular dither noise
1-H1(z-1)
data noise signal
UNIFORM QUANTIZER
output audio word
ERROR
H2(z-1)
PSEUDO RANDOM ENCODE PROCESS
data
input audio word

FIG.15

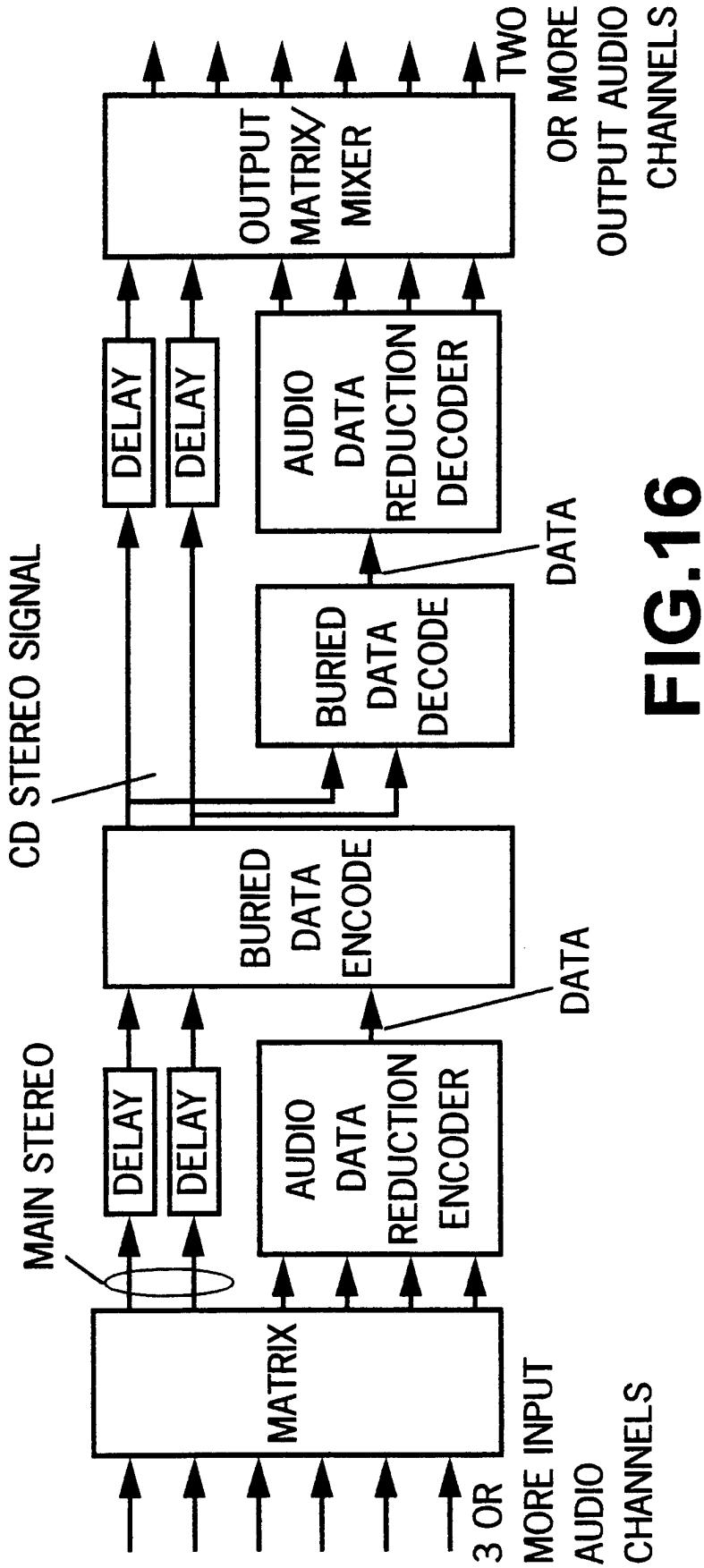

3 OR MORE INPUT AUDIO CHANNELS
MATRIX
MAIN STEREO
DELAY
DELAY
AUDIO DATA REDUCTION ENCODER
DATA
BURIED DATA ENCODE
CD STEREO SIGNAL
BURIED DATA DECODE
DATA
AUDIO DATA REDUCTION DECODER
DELAY
DELAY
OUTPUT MATRIX/ MIXER
TWO OR MORE OUTPUT AUDIO CHANNELS

FIG.16

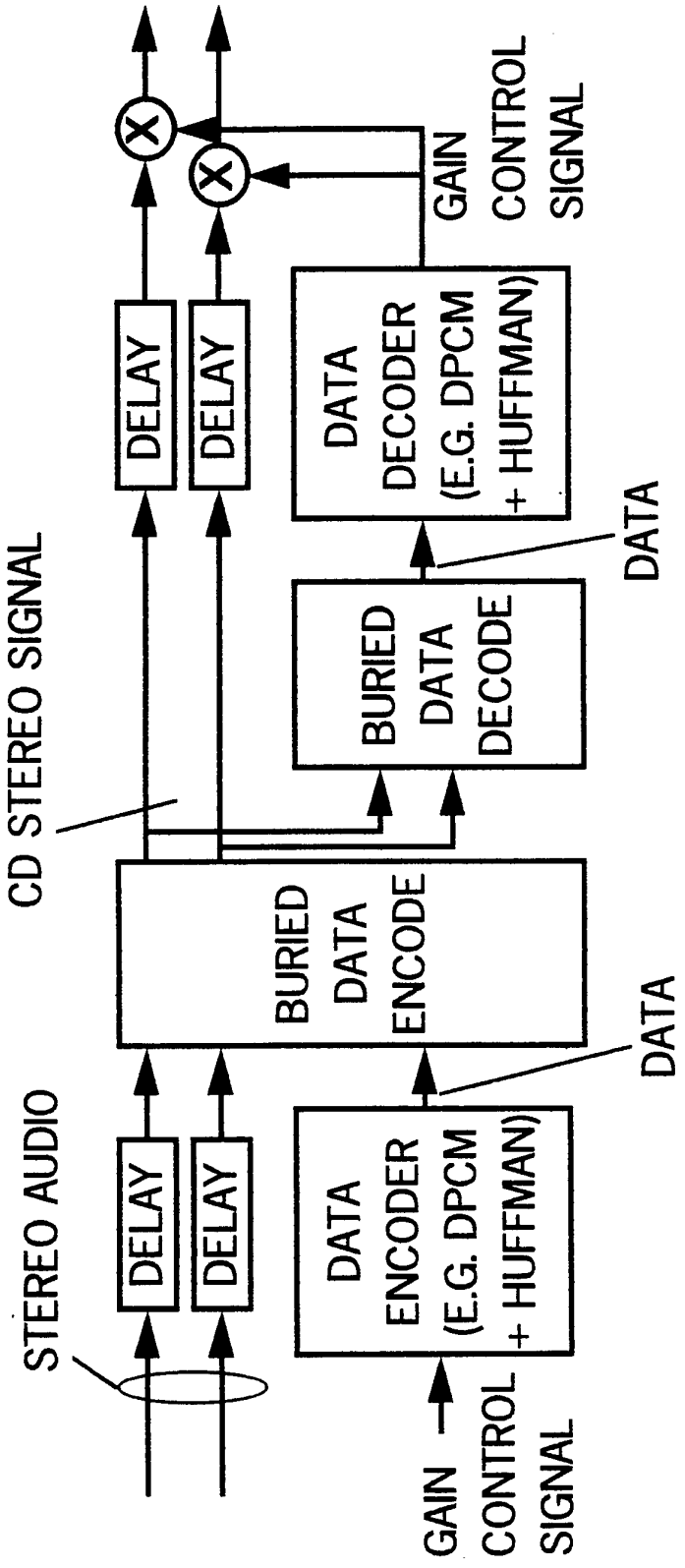

STEREO AUDIO
DELAY
DELAY
DATA ENCODER (E.G. DPCM + HUFFMAN)
GAIN CONTROL SIGNAL
CD STEREO SIGNAL
BURIED DATA ENCODE
DATA
BURIED DATA DECODE
DATA
DATA DECODER (E.G. DPCM + HUFFMAN)
DELAY
DELAY
GAIN CONTROL SIGNAL

FIG.17

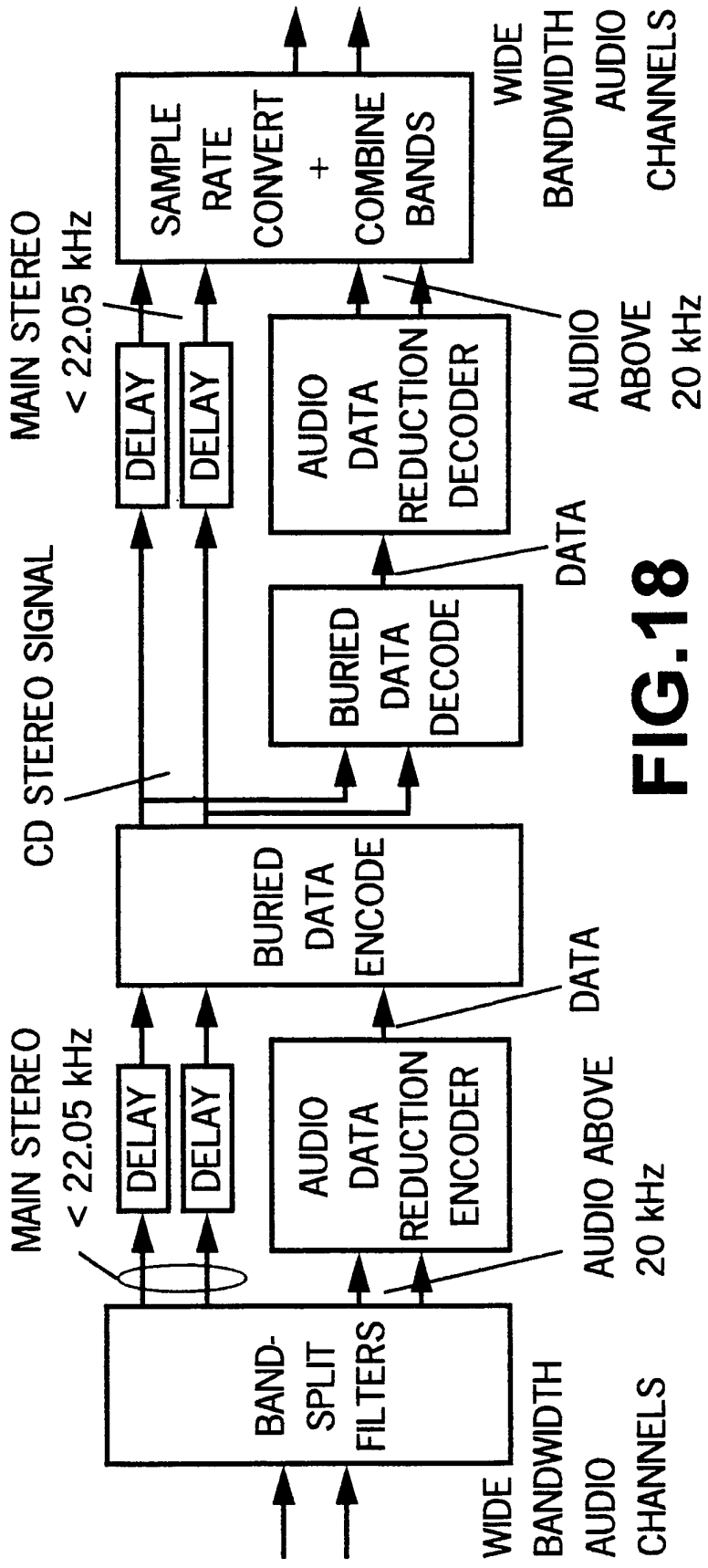

WIDE BANDWIDTH AUDIO CHANNELS
BAND-SPLIT FILTERS
MAIN STEREO < 22.05 kHz
DELAY
DELAY
AUDIO ABOVE 20 kHz
AUDIO DATA REDUCTION ENCODER
DATA
CD STEREO SIGNAL
BURIED DATA ENCODE
DATA
BURIED DATA DECODE
AUDIO DATA REDUCTION DECODER
AUDIO ABOVE 20 kHz
DELAY
DELAY
MAIN STEREO < 22.05 kHz
SAMPLE RATE CONVERT + COMBINE BANDS
WIDE BANDWIDTH AUDIO CHANNELS

FIG.18

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB 94/00297

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5 H04B14/04 H04B1/66

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 H04B H03M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DE,A,38 06 411 (DEUTSCHE THOMSON-BRANDT GMBH) 7 September 1989 see column 2, line 1 - column 3, line 4	1
A	US,A,4 857 927 (TAKABAYASHI) 15 August 1989 see column 3, line 13 - line 21; figure 1	1
X	WO,A,92 22060 (PACIFIC MICROSONICS, INC.) 10 December 1992	18-24, 26-29
A	see page 3, line 19 - line 27 see page 4, line 10 - line 13 see page 13, line 33 - page 15, line 37; figure 2	1

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 May 1994	06. 06. 94

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+ 31-70) 340-3016	Bossen, M

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

information on patent family members

International application No.
PCT/GB 94/00297

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
DE-A-3806411	07-09-89	NONE		
US-A-4857927	15-08-89	NONE		
WO-A-9222060	10-12-92	AU-A-	2161892	08-01-93
		CA-A-	2110182	10-12-92
		EP-A-	0586565	16-03-94