The Digital Compact Cassette*

GERARD C.P. LOKHOFF

Philips Consumer Electronics, 5600 JB Eindhoven, The Netherlands

One of the first implementations of ISO-MPEG audio coding is the Digital Compact Cassette. A system overview of both the audio coding and the magnetic tape recording function is given, and the design considerations and operation are described.

0 INTRODUCTION

The Digital Compact Cassette (DCC) system is a digital extension of the well-known Compact Cassette system. The main differences are:

- Highly efficient digital source coding of the audio signal, called precision adaptive subband coding (PASC), resulting in excellent sound quality
- Digital multitrack recording, including error-correction techniques
- Thin-film techniques for head production
- Autoreverse feature in all players
- New features (such as text display).

Fig. 1 shows the schematic diagram of a DCC recorder, including the items mentioned. Three specific parts may be identified:

1) The audio input–output section. This includes analog-to-digital and digital-to-analog converters as well as a digital audio interface circuit.

2) The PASC coding circuitry, transforming broadband audio data into coded data.

3) The tape drive, a tape recording and playback system that allows the storage of information from three different digital sources—audio, auxiliary code, and system information.

In the recording mode the input signal from the analog or the digital input is processed by the PASC processor. In the tape drive the error-correcting code is added to the PASC encoded signal. The signal is formatted into tape frames and recorded onto eight parallel tape tracks using 8-bit to 10-bit modulation (or channel coding).

In playback the signals from the tape are amplified and equalized before demodulation, error detection, and correction. The PASC processor reconstructs the input signal, which is fed to the digital and analog outputs.

* Manuscript received 1995 March 16.

Auxiliary code and system information can be recorded in addition to the audio. The auxiliary code is recorded on a separate track. It contains such information as track and index number, time codes, and table of contents. The system information is recorded in the same tracks as the audio, interleaved with the PASC signal bytes. It is mainly used for the text display feature.

1 DCC DESIGN PHILOSOPHY

As the compact cassette entered the most recent phase in its life cycle, an update of the system became attractive. Detailed market research and close cooperation with the music industry and another major hardware manufacturer determined the key issues of the DCC concept:

- Backward compatibility, to the extent that conventional compact cassettes can be replayed on DCC equipment
- Digitization to provide improvements in audio quality, with some new features
- A modern high-technology cassette, closely resembling the familiar compact cassette and with the same convenience in use and handling.

DCC is a combination of the proven compact cassette technology with new digital processing and recording techniques. It uses tape produced in high volumes (that is, video tape). This limits the minimum wavelength to about 1 μm. As it needs to be compatible with existing analog mechanisms and recordings, the tape speed remains 47.6 mm/s. Normal (video) chrome or equivalent tape having a width of 3.78 mm is used.

The number of tracks has to be minimal to allow for the same kind of track and azimuth tolerance as with the compact cassette. This also limits the total bit rate of the system. Fortunately new subband coding techniques offer excellent quality sound at bit rates even lower than 4 bit per sample. So the 384 kbit/s available in DCC is sufficient to code two audio channels at the maximum sampling frequency of 48 kHz. The audio bit rate is doubled by adding bits for the error detection and correction, synchronization, and modulation. Using eight tracks the bit rate recorded is 96 kbit/s per track, corresponding to a minimum wavelength of 0.99 μm.

Two-channel audio signals can be recorded with sampling frequencies of 48, 44.1 (the default frequency), and 32 kHz. Optionally, emphasis (50/15 μs) may be used. The dynamic range is equivalent to a system with a signal-to-noise ratio better than 105 dB. The reproduced signal is not affected by speed variations introduced by the tape transport (wow and flutter). Its timing is determined by a crystal-controlled clock within the equipment. The recording time is normally up to two times 45 min, with provision for two times 60 min if a thin tape is used. Additional recorded information may include several time codes (absolute, track, remaining time), track numbering, table of contents, and text information. In the search mode track start locations can be found.

2 CASSETTE

The basic dimensions of the DCC are the same as those of the compact cassette. However, it is flat with a cover that slides away when the cassette is loaded into the deck, protecting the tape against dust, soiling, and damage. When closed, the slider also locks the hubs. All players will be equipped with an autoreverse mode, so it is no longer necessary to have holes for driving the reels on both sides of the cassette. One side is closed, allowing the display of graphic artwork or a list of music titles. On blank cassettes the tape length is indicated by holes in the casing, and a record protection switch is incorporated in the design.

3 HEADS AND RECORDED TRACKS

The DCC system requires a new thin-film head design in which three sets of elements are combined—one set of nine integrated recording heads (IRHs) for digital recording, one set of nine magnetoresistive heads (MRHs) for digital playback, and finally one set of two MRHs for analog playback. The digital heads are located on one-half of the head surface. The analog heads are located in the other half. As a consequence, a DCC tape will have the signals recorded on the opposite side (of the same surface) of the tape from a compact cassette recording in the same direction. The head arrangement is shown in Fig. 2.

The digital signals are recorded on nine parallel tracks, each 185 μm wide, with a track pitch of 195 μm. The vertical alignment of the signals recorded on the tracks should be within 5 μm between any two tracks. The height of each playback head is 70 μm. This means that the sensitivity of the DCC to azimuth errors is less than that of the compact cassette.

4 TAPE FORMAT

Two different kinds of data are recorded on the tape, main data (PASC and system information) in eight tracks and auxiliary data in 1 track. The data formats are similar.

For the main data channel the signals recorded on tape are divided into Tape Frames. Between tape frames an inter-frame gap (IFG) with variable length is provided to accommodate small deviations (such as those caused by clock jitter) from the sampling frequency used during recording. Its nominal length is 64 bit periods, corresponding to about 0.4% of the nominal tape frame length (including IFG) of 16 384 bit periods. The IFG carries a signal that has alternating polarity at every bit position.

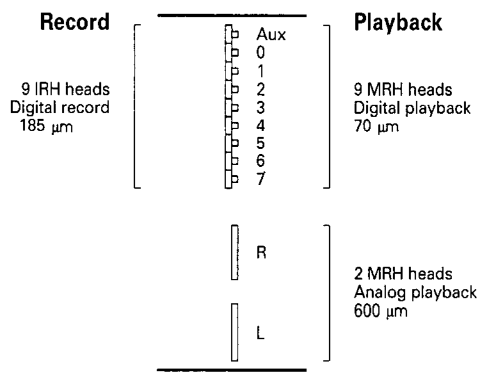

Fig. 2. Schematic representation of DCC head as seen from tape.

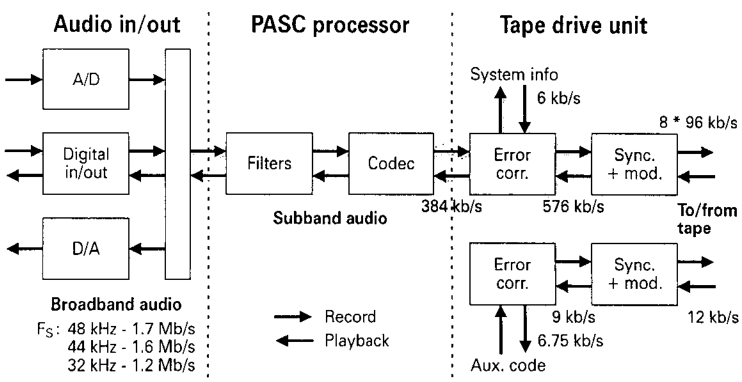

Fig. 1. Block diagram of DCC recorder.

Each of the eight main data tracks carries 32 tape blocks per frame. A tape block contains 51 tape symbols of 10 bit. These are generated according to the rules of an 8- to 10-channel coding table. The channel coding adapts the signals to the characteristics of the recording channel. Fig. 3 shows the relationship between the tape frame and its constituent tape blocks.

The first three symbols in a block constitute a block header, comprising synchronization pattern, frame, and block addresses. The synchronization pattern is used to identify the start of a tape block. The addresses are needed in the playback process to place the recovered data in the correct RAM location before error correction. The remaining 48 symbols (called the body) carry the PASC audio data, system information, and parity symbols for the error detecting and correcting code.

A cross interleaved Reed–Solomon code is employed to protect the data against random and burst errors. During playback a first code (called C1) is evaluated to detect, and if possible correct, errors in a tape block. The interleaving of the symbols comprising the two C1 code words is shown in Fig. 4. The C1 code may correct a maximum of four error symbols per tape block. Any errors that cannot be corrected are passed to the second correction phase (called C2) as erasure symbols, indicating just the location of the erroneous symbol. The C2 code can correct up to six of these erasure symbols per code word. The distribution of the symbols for a C2 code word is such that an almost optimum physical distance between the symbols is achieved, resulting in a "honeycomb" pattern. The interleaving of the C2 symbols is shown in Fig. 5. This enables the correction of dropouts with a diameter of up to 1.45 mm. The system can even operate if a complete track is missing, although in this case a long dropout in the other tracks is more likely to cause audible impairment of the sound. The PASC bytes are distributed among the tracks in such a way that any failure of the error-correction system does not disrupt consecutive PASC bytes. An uncorrectable C1 or C2 code word will not result into a burst error in the PASC signal. This helps the PASC processor to conceal these errors.

In total each main data tape frame contains 12 288 bytes, of which 8192 form the PASC signal, 3968 are parity bytes, and 128 bytes are for the system information. With the additional synchronization and addressing signals the number of bytes increases to a total of 13 056, or 1632 per track per tape frame.

For auxiliary data just one track is available. This limits the possibility for any error detecting and correcting code. To compensate for this, the bit rate is reduced to one-eighth the bit rate of the main data tracks (that is, 12 kbit/s). As a consequence, the nominal length of the IFG becomes eight bit periods and the number of tape blocks is reduced to four per tape frame. Fig. 6 shows the auxiliary data tape frame and tape block structures.

A single Reed–Solomon code protects the auxiliary data, with the capacity for correcting up to six error symbols in a tape block.

To enable easy detection of special locations on the tape during the search mode, the body of tape blocks 1 and 3 is recorded only at these locations (markers). In this way it is possible to detect the beginning of a music track without the need to decode the auxiliary code.

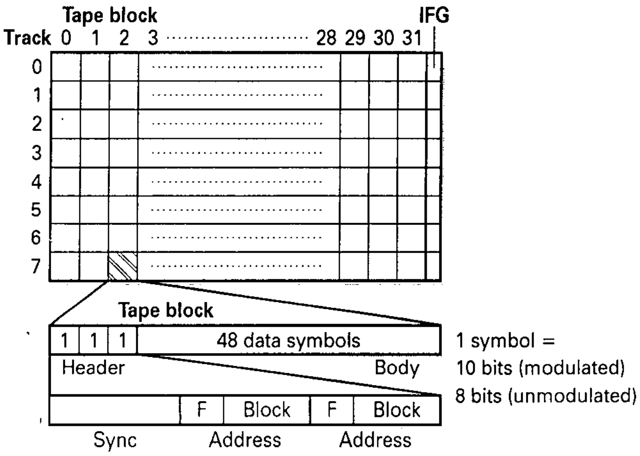

Fig. 3. Main data tape frame and tape block.

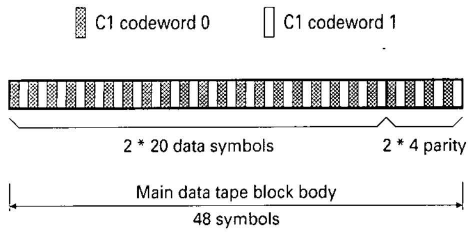

Fig. 4. Interleaving of C1 code words in main data tape block.

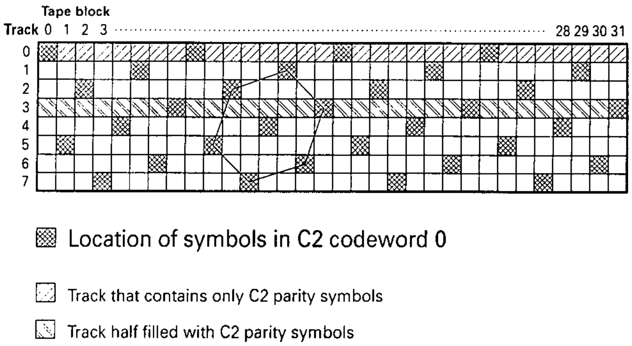

Fig. 5. Location of symbols of a C2 code word in main data tape frame. Note that track 0 is completely filled, and track 3 is half filled, with C2 parity symbols.

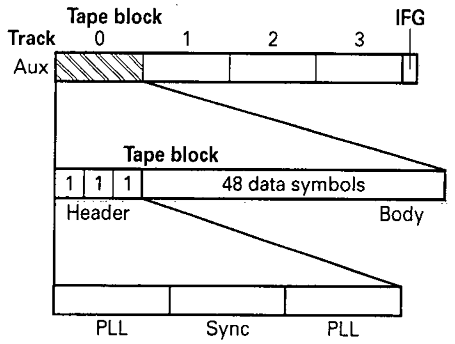

Fig. 6. Auxiliary data tape frame and tape block.

Fig. 7 shows diagrammatically the coding of the search markers in the envelope of the auxiliary data at the transition from unlabeled to labeled frames. The markers have a minimum length of four labeled frames. In addition, the location of the synchronization pattern in the header is different than for the main data, and the other header symbols are used to indicate whether or not the adjacent body is recorded.

5 PRECISION ADAPTIVE SUBBAND CODING

In order to be able to record high-quality stereo digital audio signals using the DCC with the tape format and speed described in the previous sections, it is necessary to use highly efficient source coding. PASC was developed specifically for DCCs and is based closely on ISO-MPEG layer I coding [1].

5.1 Principles

PASC is based on the following phenomena of the human ear:

1) One can hear sound only when it is above a certain level, called the threshold of hearing. This depends on the frequency of the sound and is different for every individual. In the PASC system, the threshold of hearing for a very sensitive person is assumed. Fig. 8 shows how for normal hearing the threshold of hearing changes with frequency.

2) Loud sounds hide, or mask, softer sounds in their vicinity. They dynamically adapt the threshold of hearing. An example is shown in Fig. 9, in which the threshold of hearing of Fig. 8 is modified by a tone at about 500 Hz to the extent that a quieter tone at about 900 Hz is completely masked.

These effects facilitate the efficient coding of audio signals by computing the masking threshold and transferring only those signal components that extend above this threshold. In addition the samples transferred are coded very efficiently using a floating-point representation, which decreases the bit rate dramatically.

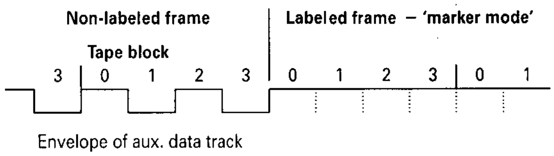

Fig. 7. Search markers coded in envelope of auxiliary data.

During encoding the PASC processor, shown in the block schematic diagram of Fig. 10, analyzes the broadband audio signal (with sampling frequency f_s) by splitting it into 32 subband signals of equal bandwidth (with sampling frequency $f_s/32$). More conventional methods follow the theory of critical bandwidths, which states that the human auditory system analyzes sounds by the use of small frequency bands centered around the frequency of the sound. The bandwidths of the critical bands vary with the frequency (approximately 100 Hz for frequencies below 500 Hz and one-third of an octave for frequencies above 500 Hz). This leads to the use of subband filters that split the frequency spectrum into subbands with unequal bandwidths.

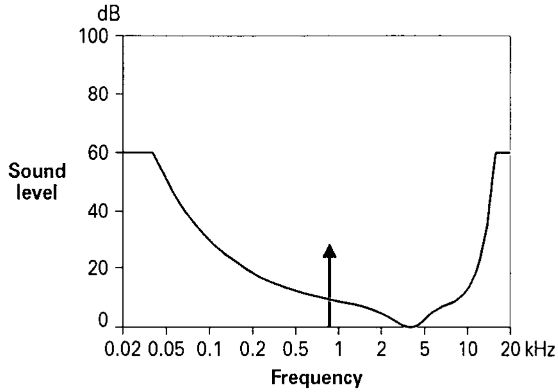

Fig. 8. Only sounds above absolute hearing threshold are audible.

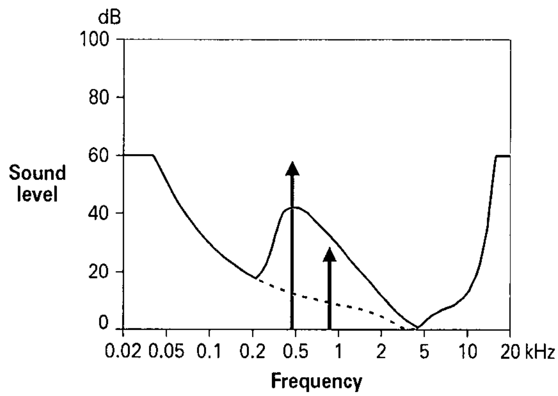

Fig. 9. Loud signals mask soft signals by raising hearing threshold.

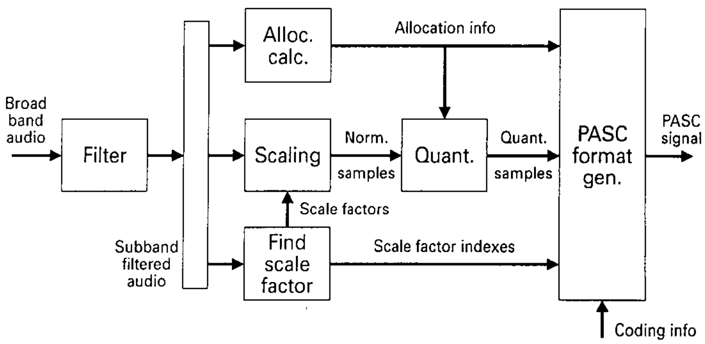

Fig. 10. Block diagram of PASC coding.

In the PASC system filters with subbands of equal bandwidth are used. The advantages of these over critical-bandwidth subbands are:

- Equal time resolution in all subbands
- Equal delay in all subbands
- No spreading of aliasing components over entire frequency spectrum
- Simpler implementation possible.

The equal bandwidths can be compensated for by changing the bit allocation algorithm. With critical bandwidths the number of bits allocated to each subband may, on average, be equal. However, with equal bandwidths this number will typically be higher for the lower frequency subbands and lower for the higher frequency subbands. This is illustrated in Fig. 11.

PASC is able to work with three different sampling frequencies, and the bandwidth of the subbands changes with the sampling frequency.

Sampling Frequency (kHz)	Subband Bandwidths (Hz)
48	750
44.1	689
32	500

The broad-band signal is filtered using a window representing a 512-tap finite impulse response (FIR) filter. The window is shifted 32 samples for every calculation of 32 new subband samples; one for each subband. This filtering is illustrated in Fig. 12. Because of the large overlap, any subband signal is still a PCM signal, but bandlimited to the subband.

The subband signals are represented in fixed-point notation. In the PASC-coded signal a floating-point notation is used, in which the length of the mantissa may vary between 2 and 15 bit, so the resolution of the system is up to 15 bit, equivalent to a 92-dB signal-to-noise ratio when performing an IEC total harmonic distortion plus noise (THD + N) measurement. However, the dynamic range is governed by the exponent in the floating-point notation. This is chosen so that the range of +6 dBFS to −118 dBFS (where FS denotes for full scale, an indication of the maximum permissible signal level) is covered in steps corresponding to 2-dB changes in signal level. The 6-dB headroom is needed because of the nature of the subband system. A complex signal may have a component having an amplitude that extends above the 0-dBFS limit of the broad-band signal in one of the subbands.

The PASC signal is composed of PASC frames. The frame period has to be short for the coding of transient signals, although this increases the number of bits needed for nonsample data transfer, and it is actually made slightly less than the delay through the subband filters (which is 10–15 ms). The frame rate corresponds with the rate of the subband samples, and each PASC frame period corresponds to the time occupied by 384 samples of each of the two channels, or 12 samples of each of the 32 subbands. It is thus 384 divided by the sampling frequency:

Sampling Frequency (kHz)	Frame Period (ms)
48	8
44.1	8.7
32	12

For the 48- and 32-kHz sampling frequencies this presents no particular problems, but for the 44.1-kHz mode the length of the PASC frame sometimes has to be adjusted (by adding additional 0 bits) in order to achieve a suitable average frame rate.

Because the first set of subband samples in a frame is calculated from 512 samples (by the 512-tap FIR filter) and the "filter window" is shifted by 32 samples at a

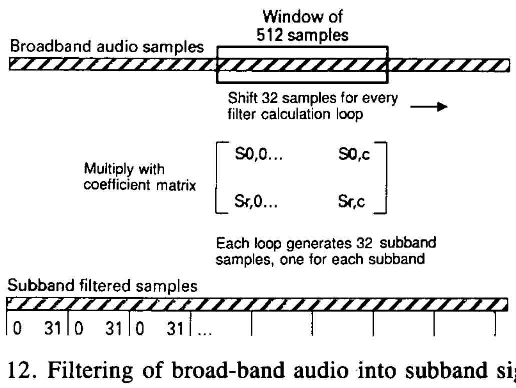

Fig. 12. Filtering of broad-band audio into subband signals.

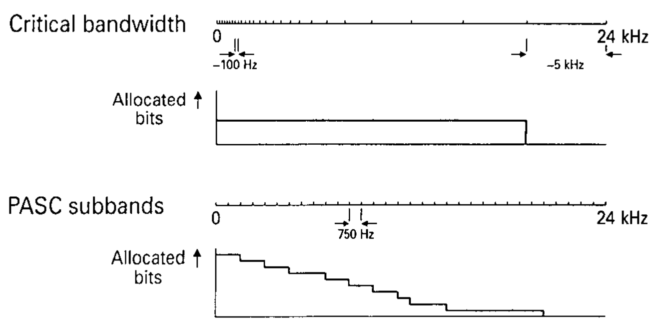

Fig. 11. Difference in bit allocation between systems based on critical bandwidths and PASC system with equal-bandwidth subbands.

time into 11 more positions during the frame period, each PASC frame is based on the information from 864 broad-band audio samples per channel. The frame window shifts 384 sample positions for each new frame.

A synchronization pattern is included in the PASC frame to identify the beginning, and some coding information is provided for correct decoding at the receiving end (original sample frequency, emphasis, and so on). Bit allocation information for each of the 32 subbands comprises a 4-bit value, which tells the decoder how many bits are used for the coding of the mantissa. If the bit allocation value is 0, it means the subband is not included in the PASC signal because its peak power was below the masking threshold. For all the subbands in the PASC signal the scale factor indexes are represented as 6-bit values. The PASC-coded samples complete the frame. If there are any remaining bits, these are set to 0.

The masking characteristics mentioned earlier can be exploited by varying the length of the mantissa. The resolution of the coding need be no higher than is necessary to keep the quantization noise below the masking threshold. In an extreme situation, the whole subband may be masked by others so that there is no need to code the signal at all.

In order to determine the masking threshold, first the power of the subband signals in the time interval corresponding with the PASC frame is calculated. Then these are used to find the contribution in the masking of one subband to another, or the masking within the subband itself. To these the absolute hearing threshold is added. The resultant masking threshold is compared with the peak powers of the subband signals. Peaks below the masking threshold indicate subbands that need not be coded. Peaks above the masking threshold will be coded with a mantissa having a length proportional to the distance between the peak power and the masking threshold. As the process is repeated for every PASC frame transferred, it is called the adaptive allocation of the available capacity.

After the length of the mantissa has been fixed, the actual coding of the PCM subband samples to PASC notation is performed. For every subband the absolute value of the peak value of 12 samples is compared with a table of multiplication coefficients, called scale factors, which represent the 2-dB steps in signal level. The samples of the subband are normalized by division by the next higher scale factor. The index of the scale factor in the table will be included in the PASC bit stream to enable decoding. This index may be considered to represent the exponent of the floating-point representation and is the same for all the coded samples of the subband for the duration of a frame. After normalization the samples, now comprising the mantissae of the subband sample values, are quantized. The quantization depends on the number of bits allocated to the subband. For example, if just 3 bit are allocated, a seven-level quantization is performed, and if 14 bit are available, the samples are in a range of 16 383 values.

A block schematic diagram of a PASC decoder is shown in Fig. 13. In the decoder the PASC-coded samples are reconverted into PCM subband samples by multiplication with the correct scale factor. The subbands are then merged to form a broad-band signal in the synthesis filter.

5.2 Implementation

The minimum processing capacity needed for PASC coding and decoding is quite substantial. Until recently it was necessary to use four large-scale-integrated (LSI) circuits to implement a PASC coder and decoder, as shown in Fig. 14. This comprises two subband filters (SBF), one adaptive allocation and scaling (ADAS) calculating unit, and one subband codec (SBC) which performs all the formatting and deformatting functions. It is now possible to implement the codec in two LSI circuits, a subband filter and codec (SFC) integrated circuit, and the ADAS. A decoder for playback-only appli-

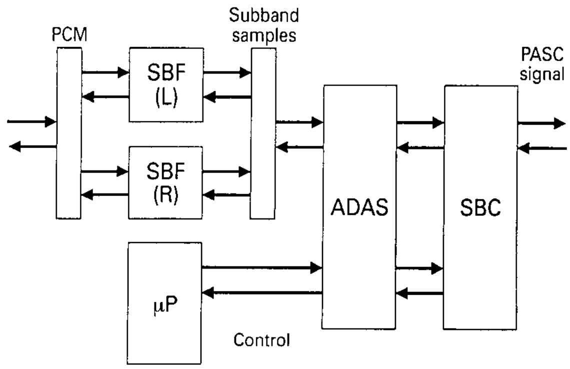

Fig. 14. Block diagram of PASC codec using large-scale-integrated (LSI) circuits.

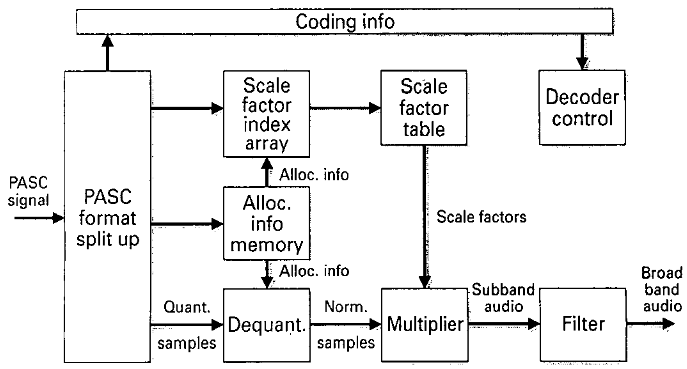

Fig. 13. Block diagram of PASC decoding.

cations requires only the SFC.

The subband filter architecture is an adapted version of the polyphase filter structure. For this a prototype low-pass filter (LPF) is used. If a digital broad-band signal passes through such an LPF, the output may be sampled at a lower frequency than the original, without the loss of information in the output signal. If the LPF characteristics are transferred to higher frequencies, a bandpass filter (BPF) is created. This may be effected by multiplying the impulse response $h(n)$ of the LPF with a cosine factor:

$$h_i(n) = h(n) \cos\{\theta_i + (i + 0.5)\ \pi n/M\}$$

where n is the PCM sample number, i is the subband number, and M is the total number of subbands. This is illustrated in Fig. 15. The phase angle θ_i is the initial phase of the subband filter, and is chosen to ensure that the following requirements are fulfilled:

1) Sufficient attenuation in the stopband

2) A flatness constraint,

$$|H_i(\theta)|^2 + |H_{i+1}(\theta)|^2 = 1$$

where $H_i(\theta)$ is the amplitude/frequency response of the filter for subband i.

3) The attenuation slopes of adjacent pairs of subband filters must be mirror images.

Such filters are known as quadrature mirror filters (QMFs). The reasons for the constraints can be explained as follows:

1) Aliasing components in the stopband of a certain filter must be prevented from appearing in another filter's passband. This is one of the main advantages of this type of filter over critical bandwidth filters.

2) To ensure a 0-dB gain over the whole bandwidth of the combined filters.

3) Because the transition bands of adjacent filters overlap, a frequency component in this region will be present in the output of both filters. The constraint guarantees correct reconstruction, eliminating alias products.

The third constraint is accomplished automatically when using the "modulation" method in which all subband filters originate from the same prototype filter, so the attenuation slopes will be exact mirror images. The result is a filter with almost optimal characteristics: a transition band of 375 Hz with about 100 dB suppression at 750 Hz [2], [3], as shown in Fig. 16.

During encoding the ADAS circuit calculates the power of the subband signals. From this and the absolute hearing threshold an estimate is calculated for the current masking threshold. The difference between the peak signal and the masking threshold is taken as a measure for the resolution requirement of the subbands. The subbands with the highest requirement will be allocated the most bits. The result of this calculation, a table expressing the number of bits allocated to each of the subbands, is passed to the subband codec. During decoding the ADAS does not perform any function relevant to the PASC system, and in a playback-only application it can be omitted.

The subband codec is the formatting device in the system. It performs this operation in encode or decode mode, according to settings it receives from a system microcontroller.

In the encode mode it receives the subband signals and the bit allocation table, from which it calculates scale factor indices that express the maximum amplitude of a set of subband samples that are to be coded in the PASC frame. The subband samples are normalized with respect to the maximum amplitude and requantized with the number of bits allocated by the ADAS. Finally the PASC frame is constructed. It includes synchronization information, coding information (to enable correct decoding), allocation information, scale factor indices, and

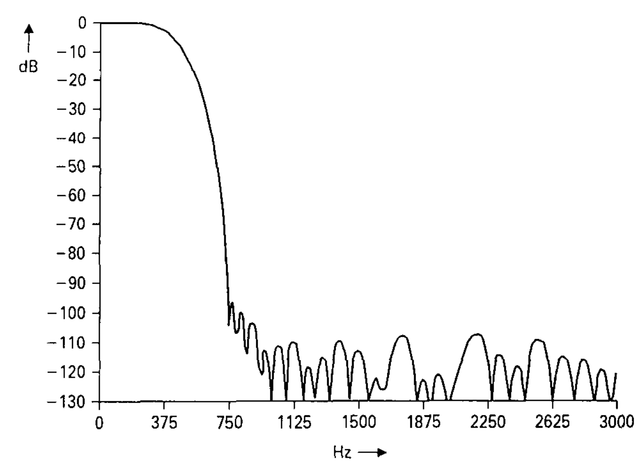

Fig. 16. Frequency response of prototype filter. Horizontal axis shows frequencies for sampling frequency of 48 kHz.

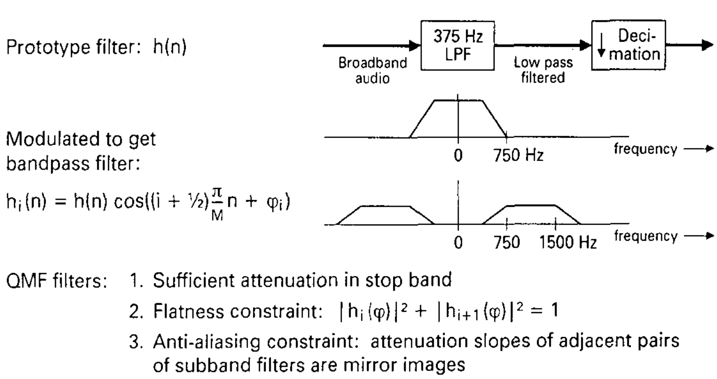

Fig. 15. Use of prototype low-pass filter to generate bandpass filters.

the quantized subband samples.

In the decode mode a PASC frame is split into components again, a process for which the coding and allocation information is essential. The subband samples are reconstructed by multiplying them with the correct amplitude factor (which has been identified using the scale factor index), after which they can be sent to the subband filters.

6 CONCLUSIONS

With the DCC advanced digital audio bit-rate reduction techniques are used to enable high-quality digital audio signals to be recorded on, and reproduced from, tape carried in a cassette with dimensions identical with those of the proven (analog) compact cassette. The tape speed, too, is identical with that for the compact cassette, so that DCC equipment is able to replay conventional analog compact cassettes in addition to recording and replaying DCCs.

7 REFERENCES

[1] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for the Coding of High-Quality Digital Audio," *J. Audio Eng. Soc.*, vol. 42, pp. 780–792 (1994 Oct.), reprinted in this issue, pp. 31–42.

[2] J. H. Rotweiler, "Polyphase Quadrature Filters—A New Subband Coding Technique," presented at the 1983 IEEE International Conference on Acoustics, Speech, and Signal Processing, Boston, MA.

[3] P. L. Chu, "Quadrature Mirror Filter Design for an Arbitrary Number of Equal Bandwidth Channels," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, pp. 203–218 (1985 Feb.).