A Proposal for the High-Quality Audio Application of High-Density CD Carriers

Technical Subcommittee

Acoustic Renaissance for Audio

Confidentiality. Update Version 1.3, 1 January 1996. This document has been prepared by the Technical Subcommittee of *Acoustic Renaissance for Audio.* An early version (1.1, released 12 April 1995) was circulated to members of *Acoustic Renaissance for Audio* and to members of the syndicates developing high-density CD and related items. With Version 1.2 (23 June 1995), the circulation was extended to interested members of the audio community and the document made available on the World Wide Web. In order that the issue be correctly reported, we request prior consultation if any of this material is to be discussed in the Press.

0. Summary

This document has been prepared as a serious proposal to all concerned in the future of high-density CD formats applied to pure audio.

Its authors present here some consensus views and arguments leading to three important outcomes.

The first is a proposal for a new data type for the compression layer in the *MPEG* system stream encoded on the high-density discs. This data type carries *losslessly compressed* linear PCM (packed) audio[1], and is quite distinct from the currently recognised streams carrying lossy-compressed (data-reduced) channels such as *MUSICAM*, *MPEG Audio*, *PASC* or *AC-3.*

The second is a working proposal for a disc format, which in its essentials is:

- That lossy data-reduction techniques such as *MUSICAM*, *MPEG Audio*, *PASC* or *AC-3* should not be used in any form for the high-quality sound application.
- That higher quality should be based on surround sound and use linear PCM with more channels and a higher data rate than current CD-DA.
- That our proposal does not add significant cost to any currently proposed video players, and is structured in such a way that only the more advanced applications will require extra hardware.

Our proposed format allows from two to eight channels of high-quality linear PCM audio, each of up to 24-bit at 48kHz or 20-bit at 96kHz sampling rates or a combination of both.[2] [3]

---

[1] Lossless compression and decompression is a process that returns the original input data exactly. Its advantage is that in the compressed form, the data occupies less space on the disc, reducing the maximum data rate. The degree of compression is very high at high sampling frequencies. Lossless compression is quite unlike lossy psychoacoustic-based compression, which does **not** return the original data intact. For more details, see the glossary in section 15.

For ease in distinguishing between lossless and lossy compression in this document, we have adopted the term ‘packed’ to describe losslessly compressed linear PCM audio.

[2] We actually propose that the *MPEG*-compatible stream should have complementary fields for carrying both 48kHz and 96kHz components of the signal. The record producer may then set either field to ‘no-content’, or indeed, provide a combination of these components that offers more flexible upwards and downwards compatibility.

[3] At 96kHz the 20-bit limit arises through the peak data rate limits (11.2Mb/s for *MMCD*). Note that 24-bit recordings can be made at 96kHz *providing either* only 6 channels are exercised *or* providing the input channels are filtered to approximately 30kHz (thereby limiting the peak data rate).

The proposed format gives the disc producer the flexibility to make trade-offs between the number of active channels, their bandwidth, their precision, and overall playing time. This flexibility is illustrated in section 11.1 on page 12.

Compatibility for two-speaker stereo is provided by always assigning two of the eight channels to a two-channel mix.

For more advanced applications, the remaining six channels can be used for surround-sound mixes – including those carrying *Ambisonics* information. In this way record producers can simultaneously issue their preferred mixes for two to six channels.

We recognise that there are several benefits to including the lossy-compressed version of the audio on the disc alongside the high-quality audio versions. In particular, this approach allows minimum changes in the player architecture and gives very good compatibility for budget systems.

Some suggestions are made about using the high-quality packed (losslessly compressed) audio with *MPEG2* video for some applications – such as high-quality music-videos – or with other data such as text and graphics. In essence, we see no reason why the disc format should not permit four types of data simultaneously in the compression layer, namely high-quality audio, lossy-compressed audio, *MPEG1/2* video, and associated data.

The final outcome is an expression throughout the document of concerns, items we considered and items we recommend be incorporated into any resulting standard.

A glossary is included in section 15.

Part 1: Background

1. Introduction

In January 1995 two syndicates released information about high-density CD formats under development. Both the *Sony/Philips*[4] and *Toshiba/WEA*[5] syndicates proposed a CD carrier for movie delivery based on *MPEG2* variable-rate picture coding and a multichannel/multilingual compressed audio code, using either*, MUSICAM* (*MPEG) or Dolby AC-3*.

Both syndicates indicated a willingness to discuss other applications of the disc with interested parties; these applications include broadcast, multimedia and pure audio.

This document was written to appeal to both syndicates on the subject of a high-quality pure-audio application of high-density CD, and to raise awareness in the audio community of the issues involved. We refer to these formats as HQAD (High Quality Audio Disc) to distinguish them from the current Red Book CD-DA standard.

In December 1995 these two syndicates agreed on a common set of specifications on which to develop the final DVD standard.

2. Purpose of this document

The purpose of this document is to present the consensus opinion of a sector of the audio industry on suitable format(s) for a high-density CD intended to deliver sound only (HQAD).

The following were important starting points in our thinking:

- There is a real possibility that carriers such as this HQAD will become the de facto standard over a decade, effectively replacing CD-DA.
- The audio community will not be content with a carrier that uses psychoacoustically-based data-reduction schemes such as *AC-3* or *MUSICAM* alone.
- There is a rapidly growing interest in multichannel music systems. A carrier with a data capacity higher than that of CD would make such systems possible.
- Recently audio engineers have begun to explore methods of enhancing CD resolution because it is increasingly understood that the existing CD-DA 2-channel 16-bit/44.1kHz channel is inadequate.

3. Extent of discussion and authorship

This document is a consensus proposal from the Technical Subcommittee of *Acoustic Renaissance for Audio*, a free body dedicated to advancing audio quality.

The members of this committee, their advisors and all affiliations are appended in section 13.

In particular, opinions have been obtained from these industry sectors:

- Recording and record publishing
- Broadcast
- Audio engineering academics
- CD player manufacturing
- Virtual Reality developers

---

[4] *Sony/Philips* propose *MMCD* in the 'Gold Book' standard.

[5] *Toshiba/WEA* propose the *SD* system.

4. Requirements of software producers

Indications from the audio-recording community are:

- Playing time for an audio disc need be no longer than for CD-DA (officially 74 minutes; in later practice, 84min).
- Only one editing session can be afforded for any given session (disc). A two-stage mixdown, first to surround and then to two-channel, may be the preferred method, as shown in Figure 1 on page 19.
- Mastering should be a straightforward process, preferably consisting of only one pass with the minimum opportunity for operator error.
- The disc specification should take realistic account of currently installed hardware. This last point firmly suggests a specification rooted in sampling frequencies based on 48kHz.
- Additionally, broadcasters see a great benefit in a multi-programme ‘archiving’ format particularly for 2-channel material. Several 2-channel programmes in a series could be recorded on the same disc[6] and therefore the format should enable optional and optimal trade-off between number of audio channels and playing time. (See section 11.1).

5. Requirements of HQAD player makers

Indications from the player-manufacturing community are:

- Proposals that allow improved sound quality, whether by increasing resolution, frequency range or the number of channels, should not impose a significant cost penalty on the lowest-common-denominator DVD player, primarily envisaged for movie application.
- Proposals for improved sound-quality delivery should be such that the lowest-common-denominator player can play some part of the audio disc.
- The structure of the disc should be hierarchical, so that only those products that access more advanced or sophisticated features need to use more silicon or signal processing.
- The system should be designed so that high-quality or full implementation can be made using the ‘transport’ plus ‘decoder’ architecture. This means that an interface will be needed to transfer data from the player and should be planned for.
- The system should be designed to give maximum flexibility in developing the signal-processing and coding techniques through the lifetime of the format, giving a very wide set of options and opportunities for manufacturers and recording companies to satisfy niche markets.
- All players should be obliged to access basic two-channel stereo, and all discs should carry this at a minimum.

6. Requirements of the audio community

The audio community is putting considerable effort into pushing the limits of resolution of the current CD-DA channel, using techniques such as noise-shaping, pre-emphasis and buried-data techniques. [10] [2] [15] [21] [1] [24] [25] [22] [27]

An overwhelming requirement of the audio community is to have a carrier that does not use data-reduction methods which essentially throw away those parts of the audio data that are argued to be inaudible, such as *MPEG*, *PASC* or *AC-3*. The HQAD should use linear PCM encoding as a basis.

[6] E.g. ‘The Hitch-hiker’s Guide’ or Beethoven’s Symphonies all on one HQAD

There is a general consensus that 16-bit 44.1kHz linear PCM is inadequate. The question is, how do we balance pushing the envelope outwards?

It would seem to be reasonable to design the channel according to the capabilities of the human receiver, and an obvious set of parameters would accurately encode the entire auditory range of the listener, namely:

1. Full spherical recording
2. Frequency range from near DC to at least 25kHz in air
3. Dynamic range from inaudible to 120dB spl.

6.1 Precision

Here are some relevant factors.

- Stuart has shown in [25] that a linear PCM channel requires 21.5 bits resolution without noise-shaping to encode noiselessly from the average human hearing threshold to 120dB.
- The 1% percentile on threshold for young listeners is up to 15dB lower, although room or recording-venue noise rarely is, and there is little prospect in the lifetime of this carrier of designing associated analogue electronics with an overall dynamic range greater than the equivalent of 22-bit.
- Stuart and Wilson have shown that with psychoacoustically optimised noise-shaping a channel could obtain two bits' extra effective resolution and subjective dynamic range. This subjective advantage can be extended to three bits by combining noise-shaping with pre-emphasis. [24]
- Recordings using noise-shaping are highly vulnerable to corruption by carelessly or inadvertently applied digital filtering, or by other DSP operations in the mastering or replay stages.
- The same problem applies to recordings using buried-data techniques to optimise the channel (*XtraBits*, *HDCD*, *Philips*, *Pioneer*, etc.). [10] [15] [21]
- Fielder has argued that 120dB spl is not an adequately high limit; he proposes 129dB. [4]
- Current-day converter technology has converged on 20-bit, with limited prospects of significant real progress in terms of dynamic range in the next five years.
- Current-day DSP machinery is progressively moving to 24-bit basic precision.

6.2 Frequency range

There is a requirement to sample at a higher rate than that used in the CD-DA, justified as follows:

- The main body of psychoacoustics literature shows that the audible frequency range of steady tones for young adults extends from below 10Hz to 18kHz mean, with 1% able to hear 25kHz tones. On the face of it, therefore, a channel capable of conveying all the sounds people can hear with their ears should be able to convey 25kHz.
- There is a body of opinion, based largely on anecdotal evidence, that spl in the supersonic region has an effect on perception. [19] [20]
- The audio community is in general agreement that a channel limited by 44.1kHz sampling is too tight, leading to 20kHz restriction with a narrow transition band. This tight specification has led to significant implementation difficulties.

In addition, we have to remember that the major investment in recording, production and broadcast is in machinery sampling at 48kHz.

6.3 Full spherical recording

While there may be some disagreement about the desirable frequency or dynamic ranges of an audio carrier, no-one can be in any doubt that there are considerable and immediately obvious benefits to reproducing sound from more than just the frontal horizontal quadrant.

There is an increasing awareness of and demand for the increased realism that results from taking 'stereo' beyond two speakers. There is currently considerable interest in surround-sound techniques, and this is fuelled by the better source in CD-DA and DSP. [26]

- A very satisfactory core technology exists, in the form of *Ambisonics*, that permits an open approach to recording, conveying and reproducing a soundfield. At the heart of the *Ambisonic* approach is the concept that the soundfield can be described by four hierarchical signals, namely W (mono), X (Left – Right velocity), Y (Front – Back velocity) and Z (Up – Down velocity). From W, X, Y, Z any speaker array can be fed optimum signals. [6] [12] [7] [26]
- By far the largest consumer-installed base of surround sound is in home cinema systems. The more demanding DVD customer is expected to have loudspeaker arrays corresponding to the cinema 5.1 arrangement of Left, Centre, Right, Surround Right, Surround Left and Bass Effects (hereafter called the 5.1 layout and abbreviated and paired as L, R; C, E; Ls, Rs). Alternatively in the future, a full bandwidth E channel may be allocated to height information in cinema applications
- The digital soundtracks for movie and TV may well be broadcast or recorded as speaker feeds for these arrays.[7]

6.4 Dynamic Range Control

One of the aims of this proposal is to provide a full-dynamic-range option for those consumers who are in a position to exploit it. However there are circumstances in which, due to high levels of background noise or restricted loudness, the dynamic range may usefully be reduced. Under these circumstances the decoding equipment could control the dynamic range in a manner sympathetic to the programme material.

It has been shown in [13] and [14] that such a control can be provided by analysing the audio during the production process. The full dynamic range is conveyed on the disc, along with control data which can be used to apply dynamic-range reduction when required. This broadcast technique also has advantages in the HQAD application:

- The full dynamics are preserved unchanged on the HQAD.
- The signal analysis, which can be complex, becomes integrated in the mastering process.
- Relatively simple signal processing is needed to implement this feature in the player/decoder.

Such a process has already been developed and included in the specification of the EU147 Digital Audio Broadcasting (DAB) system.

[7] **Note**. There is an important distinction between the signals called 'speaker feeds', i.e. L, R; C, E; Ls, Rs, and the hierarchical signals W, X, Y, Z in *Ambisonics*, which require decoding to a speaker layout. The advantage of the hierarchical signals is that they can be decoded to installations using any number of speakers, starting with one.

Part 2: Suggestions and Possible Strategies

7. Number of channels

The following DVD audio features are assumed:

- The basic movie player will provide outputs for two speaker feeds (*Dolby Surround* encoded as Lt and Rt[8]).
- More advanced player/decoder combinations will use the 5.1 layout fed from a 384kb/s *AC-3, MUSICAM,* or similar lossy-encoded data stream.

7.1 Minimum requirement

The pure audio disc (HQAD) needs to be able to work simply with these speaker layouts. We suggest that five full-bandwidth channels are the *minimum* required.

7.2 Subwoofer feeds

We strongly recommend that music should not be recorded to layouts with a mono subwoofer, since single-subwoofer replay is very inferior in terms of energy response and spaciousness. The subwoofer feed should be generated by the end-user's equipment in the equivalent of the surround decoder function. This function will be present in equipment capable of decoding *Dolby Surround* and should be user-selectable according to customer preference or the capabilities of the loudspeakers.

7.3 Ambisonic process

We have determined that it is possible to take an *Ambisonic* W, X, Y (and Z if necessary) set and to 'decode' these to provide signals for recording onto the HQAD, and therefore, for reproduction via a standardised five-speaker arrangement.

By this method a soundfield recording or mix can be played simply on a standard 5.1 speaker layout.

For more advanced or higher-performance installations, the five feeds can be decoded back to W, X, Y (and Z) for re-operation into other layouts.

7.4 Effects channel

In the current cinema 5.1 systems, five full-bandwidth channels are augmented by a 0.1 (200Hz bandwidth) bass-effects channel. We suggest that this be used as a channel which adds low-frequency power-handling for special contemporary or experimental material. (It is not required for normal acoustic recordings.)

This sixth channel should be defined as a full-bandwidth channel. It could then be flagged in associated data/subcode as an Effects channel, to be used:

- for low-frequency effects,
- as a channel conveying height information, for example the Z component of *Ambisonics,* or
- as an unassigned channel for special applications.

[8] Lt and Rt are the recognised symbols for the Left Total and Right Total two-channel-compatible downmix of surround material. In *Dolby* terminology, Lt and Rt may be encoded as *Dolby Surround* as well as forming a 2-channel compatible feed.

In the event that this channel is unused, or carries only low-frequency information, the data rate on the disc will automatically be reduced by the packing method proposed. Therefore, on material not requiring height information, longer playing times or higher resolution can be chosen at the producer's discretion.

7.5 Two-channel compatibility

In all audio systems there is an issue of down-compatibility.

In considering how two-speaker systems will play surround recordings, we see a number of options and pointers. One less desirable option for two-speaker listening is to use the CD-DA version of the recording. We feel that this option penalises the two-speaker listener, who does not gain access to the highest sound quality. In any event, producers will need to provide a Lt, Rt mixdown for the CD-DA release, and we have examined alternative methods for making this available along with the high-quality surround mix.

We have considered the following options for providing a two-speaker feed:

1. Convey six channels as L, R, C, E, Ls, Rs. Provide DSP downmixing in **all** players to matrix to two channels in a manner analogous to the *AC-3* downmix of 5.1 to 2. This method is not favoured, as it would add signal-processing to all players and disadvantage the two-channel downmix from an artistic point of view.
2. Arrange a matrix such that six channels conveyed on the disc are not speaker feeds as L, R, C, E, Ls, Rs, but Lt, Rt, C, E, Ls, Rs. The Lt and Rt would be downmixed 5 to 2 before mastering in such a way that a sophisticated decoder could extract the original L and R. This method requires either known matrices or a subcode to describe the matrices. If the latter approach is taken, then there is a problem designing and fitting out new downmixing equipment that is able to write the associated mix data. A further problem with this scheme arises in processing and re-issuing original two-channel mixes for surround. [11] [8]
3. Convey eight channels as L, R, C, E, Ls, Rs, with Lt and Rt in addition. This method has the considerable advantage of simplicity for recording companies, mastering houses and hardware makers. There is no real requirement for the Lt and Rt channels to be time-aligned with the surround mix, although that may prove to be beneficial to compression ratios.

We are firmly in favour of option 3, *so long as* packing (lossless compression) is used *and* the options for full surround are not ignored. Although many sophisticated schemes can be considered for downmixing matrices operating in the players, this strategy will always lead to more working difficulties in production and in replay-hardware design.

However, option 2 remains useful when minimising audio data rate is paramount, as in the case of sending three or five high-quality audio channels with *MPEG-2* video.

7.6 Red Book compatibility

It would be very advantageous to evolve to a single inventory disc, by releasing HQAD discs with high-density information on one layer and a conventional CD-DA on the other.

The current DVD proposals permit such mixed-mode discs. It is relatively simple to imagine this working if both layers are viewed from the same side – i.e. placing the CD-DA on the back layer (at 1.2mm) and the HQAD on the nearer layer of a two-layer disc (at 0.6mm), seen from the reading side.

It is not entirely clear to us how Red-Book player compatibility can be obtained if the disc is double sided and the CD-DA layer is only 0.6mm from the reading side.

8. Sampling frequency

Although on the face of it 55kHz is the minimum sampling rate necessary to encode all audible sounds, we recommend a specification that is based on multiples of the 48kHz rate found to be standard in professional audio and in *AC-3*. We see an advantage in permitting the development of quality improvements that higher sampling rates would bring, and suggest that 48kHz and 96kHz are the only options required.

Although at first sight 96kHz may appear to be grossly wasteful of data rate, this is not the case if packing (lossless compression) is used. In section 10.2 we point out that the packed data rate for 96kHz sampling may typically be only 30% greater than that for the same material sampled at 48kHz, and can be less if the full 40kHz bandwidth offered by 96kHz sampling is reduced.

Having carefully considered all the factors, and assuming lossless packing to be used, we conclude that it is not necessary to cater for compromise sample rates such as 60kHz, 66.15kHz, 72kHz or 88.2kHz.

We further propose that *MPEG* packed-audio streams be defined for 48kHz and 96kHz and that these are both always present on the disc. In the majority of cases, one or the other will be sent null data, but there are circumstances in which both may be required and this structure allows for more flexible evolution to 96kHz operation on the part of hardware and software providers.

9. Precision

In view of the known dynamic range of human hearing, recording spaces and analogue electronics, we feel comfortable in recommending channels that can obtain the audible equivalent of 21.5-bit precision when noise-shaping or a combination of noise-shaping with pre- and de-emphasis is used.

As a guideline, a *minimum* requirement using noise-shaping and emphasis would be:

- 20-bit precision for channels sampled at 48kHz
- 14-bit precision for channels sampled at 96kHz

Because pre-emphasis is not helpful to packed audio channels, and because a 14-bit specification is unlikely to find favour, 16 bits at 96kHz is the minimum practical alternative – one which fits in well with existing machinery and interfaces.

10. Channel coding

Decisions about numbers of channels, precision and sampling rate converge on a ‘bit budget’.

10.1 Signal processing

Linear and psychoacoustically correct coding methods are known which can improve the performance of linear-PCM channels. The two principal methods are noise-shaping and pre-/de-emphasis. [10] [2] [1] [24]

Noise-shaping can be made open-choice at the discretion of the recording producer.

Pre- /de-emphasis requires standardisation. Although there are better choices, the standard CD-DA 50/15µs will have to be provided by any player capable of playing CD-DA as well as HQAD, and so should not be ruled out as an option.[9]

[9] It should be noted that the optimum pre- /de-emphasis for 96kHz sampling is different from those optimised at lower sampling rates.

10.2 Lossless coding

We strongly recommend that the high-quality audio channels be losslessly coded (packed). Signal processing has advanced to the state where the data-reduction benefits of such coding are too good to pass by. Unlike perceptual or lossy data reduction, lossless coding does not alter the final decoded transmitted signal in any way, but merely ‘packs’ the audio data more efficiently into a smaller data rate.

Existing lossless audio data compression systems are optimised for reducing average data rate, but not for reduction of peak data rate or for optimum results at high sampling rates such as 96kHz. We have determined simple-to-decode methods optimised for these latter requirements.

The process of packing PCM becomes more efficient as sampling rate is increased. For example, packed 96kHz audio does not double the data rate of packed 48kHz as you would expect; the increase is more like 30%.

Packing offers the opportunity to make a much better product. It allows us to convey more precision on more channels, but also gives a lot of open-ended flexibility to the user – as can be seen in some of the examples quoted in Table 3 on page 13.

10.3 Lossless coding guidelines

We are aware of relatively simple-to-decode packing and unpacking techniques that should allow the lossless data compression shown in the table below for five or more associated channels. We anticipate that higher compression rates can be obtained with development over a relatively short period.(Compression is shown as the saving in bits per sample per channel).

Table 1	**Data-rate reduction: bits/sample/channel**	
Sampling kHz	**Peak**	**Average**
48	0	6
96	5	8

Different musical material compresses by different amounts with lossless packing, with material having narrow dynamic range and high treble energies compressing less well. Lossless coding algorithms can be chosen with less compressible material in mind, giving an overall improvement in degree of data-rate reduction. The degree of data-rate reduction will be greater for highly compressible audio material such as most classical music – for which absolute disc duration is most critical.

10.4 Pre-emphasis and lossless coding

We have determined that, for packed channels, the use of pre- and de-emphasis gives no advantage in coded-data rate for a given noise performance. Therefore, pre- and de-emphasis are only of benefit when used with linear PCM channels that are not losslessly coded, and should not be used with losslessly coded channels.

10.5 Noise-shaping and lossless coding

Psychoacoustic noise-shaping of the PCM audio channels may be used, along with lossless coding, to create a packed channel with perceptual improvements of about 3 bits at 48kHz and 5.5 bits at 96kHz.

For 96kHz recordings some noise-shaping is encouraged, in order to optimise the subjective dynamic range and overall data rate.

10.6 Lossless coding for flexible wordlength

It is possible to design the lossless-coding specification in such a way that at the mastering stage the record producer can make a personal trade-off between playing time, frequency range, number of active channels and precision. The packed channel can convey this choice implicitly in its control data, and the system operation will be transparent to the user.

This scenario has the following benefits:

- A producer mastering at 48kHz can control the incoming precision and trade playing time or channels (e.g. with or without data in E, effects) for noise-floor.
- A producer mastering at 96kHz can also trade bandwidth for playing time, active channels and precision.

By way of examples:

a) playing time or precision may be extended by pre-filtering information above 30kHz

b) playing time or precision may be extended by only supplying a 2, 3 or 4-channel mix.

The technical standard for lossless coding can specify the maximum input and output wordlength, possibly as either 20 or 24 bit. In addition, the standard can be arranged so that choices regarding input wordlength, number of active channels and bandwidth are automatically handled by the coding, without manual intervention by producer or end-user.

10.7 High-quality audio with MPEG-2 video

Consideration should also be given to using the proposed high-quality packed audio alongside *MPEG2* video data and the compressed audio *AC-3*/*MUSICAM*.

It seems likely that there are a number of applications that will benefit from different compromises between audio quality, video quality and playing time from those made in the movie versions of the discs.

For example, slowly-moving or graphical video data could accompany high-quality surround recordings. Alternatively, for some types of music, such as opera, *MPEG-2* pictures could accompany a high-quality sound track using two channels of packed audio (e.g. Lt and Rt at 48kHz and 20-bit nominal).[10]

Within a common standard, producers could choose between a number of viable high-quality options.

10.8 HQAD player concept

Figures 2 and 3 on page 20 show outline player architectures.

[10] In these cases, where minimising audio data rate is paramount, the Lt and Rt mix could be matrix-encoded with a method such as *Dolby Surround*, or alternatively the standard could be extended to allow a smaller number of high-quality sound channels in addition to Lt and Rt, along with more advanced matrix-decoding information, to enable a decoder to play to 5 speakers.

Part 3: Proposals

11. The HQAD bit budget

The table below estimates the bit budget for HQAD and compares it to CD-DA. Within the constraints of close to 74 minutes' playing time and 10Mb/s peak data rate, several options exist, including those shown.

We have allocated eight channels to audio, 448kb/s to a parallel lossy-compressed *AC-3* or *MPEG* channel, and a 176kb/s channel to a parallel data or subcode channel.

Table 2. Outline calculations of HQAD

Lossless compression on all 8 channels

	fs kHz	Audio Channels	Precision Bits	AC-3 or MPEG	Subcode kb/s	Data Rate Mb/s Input	Ch Peak	Ch Ave	Play mins	Capacity GByte
CD-DA	44.1	2	16			1.41	1.41	1.41	74	0.78
HQAD	48	8	20	448	176	8.30	8.30	6.00	104	4.70
	96	8	16	448	176	12.91	9.07	6.77	93	4.70

In the table above there are three columns describing data rate in Mb/s. The first, labelled 'Input', is the worst-case rate of data in the uncompressed recording being fed to the mastering process. The second column, labelled 'Ch Peak', gives the expected maximum data rate in the packed channel – i.e. on the disc. The last column shows the average disc data rate and is used to compute playing time.

These figures assume a lossless packing scheme optimised for peak rate reduction. They also assume 48kHz 20-bit or 96 kHz 16-bit signals having relatively moderate compressibility and the compression ratios given in Table 1 on page 10.

If one or more of the channels is unmodulated, e.g. if the C channel or the E channel or the Ls and Rs channels are not used, or are modulated with a highly compressible signal such as a bass-effects channel, then the data rates will be smaller than those shown, and playing times will be longer – or larger words can be used (up to 24 bit).

11.1 Examples illustrating the flexible disc capacity

Table 2 shows nominal capacity for the proposed HQAD. However, the use of lossless 'packing' gives a very flexible structure to the disc. By specifying a mastering system which can accept:

- between 2 and 8 channels,
- 48kHz or 96kHz sampling rate,
- between 16 and 24 bits,

we effectively construct a carrier in which the producer can make the trade-off between numbers of channels, frequency range, precision and playing time.

The mastering process can embed precision information in the data stream, which has the added benefit that the standardisation process does not need to anticipate all the options – neither is subcode required to control the replay process.

The packing process can effectively provide a continuum of sampling rates between 48kHz and 96kHz, providing the input to the compression process is effectively low-pass filtered – the less information there is at high frequencies, the higher the compression ratio becomes.

Table 3. Examples of HQAD options: new capacity

Lossless compression on all 8 audio input channels

Example	fs kHz	Audio Channels Lossless	Precision Bits	AC-3 or Musicam	Subcode kb/s	Data Rate Mb/s Input	Ch Peak	Ch Ave	Play mins	Capacity GByte
1	48	8	16	448	176	6.77	6.77	4.46	140	4.7
2	48	8	20	448	176	8.30	8.30	6.00	104	4.7
3	48	8	24	448	176	9.84	9.84	7.54	83	4.7
4	48	7.1	20	448	176	7.44	7.44	5.40	116	4.7
5	48	7.1	24	448	176	8.80	8.80	6.76	93	4.7
6	48	2	16	192	176	1.90	1.90	1.33	472	4.7
7	48	2	24	192	176	2.67	2.67	2.10	299	4.7
8	96	8	16	448	176	12.91	9.07	6.77	93	4.7
9	96	8	18	448	176	14.45	10.61	8.30	75	4.7
10	96	7.1	16	448	176	11.53	8.12	6.08	103	4.7
11	96	7.1	18	448	176	12.89	9.48	7.44	84	4.7
12	96	7.1	20	448	176	14.26	10.85	8.80	71	4.7
13	96	2	16	192	176	3.44	2.48	1.90	329	4.7
14	96	2	18	192	176	3.82	2.86	2.29	274	4.7
15	96	2	24	192	176	4.98	4.02	3.44	182	4.7
16	96	6	22	448	176	13.30	10.42	8.69	72	4.7
17	48	9.1	24	448	176	11.11	11.11	8.49	74	4.7

In the table above, we illustrate some extremes of this flexible use. The duration is calculated on the basis of *MMCD* capacity.

- Example 3 shows that the disc can give the full 83 minutes seen in some current CD-DA (beyond official Red Book spec.) with 8-channel 24-bit 48kHz material.
- Example 5 shows that using the E channel for bass effects allows this 24-bit 48kHz material to exceed current CD-DA duration.
- Examples 6 and 7 show that for 2-channel archive applications 16-bit 48kHz material can play for almost 8 hours, reducing to 5 hours with 24-bit input!
- Example 9 shows that the peak data rate allows 1 hour with all 8 channels recorded with 18-bit precision at 96kHz.
- Example 12 shows that 20-bit precision is available at 96kHz providing the eighth channel is used for bass effects. Duration may be just 1 minute less than CD-DA.
- Examples 15 and 16 illustrate reducing the number of channels even allows up to 24-bit recordings to be conveyed at 96kHz.
- Example 17 hypothesises 9 effective channels. This in fact illustrates a mix using 2-channel at 96kHz along with 5.1 channels at 48kHz; both with 24-bit precision.

12. Specification keypoints for High-Quality Audio Disc

12.1 Mandatory

- All players for video must play Lt and Rt channels from the audio disc (HQAD), using either the lossy or the lossless compressed versions.

12.2 Channels

1 Up to eight full-bandwidth (i.e. DC to half-Nyquist) channels of high-quality sound.

The basic data shall be, at the disc producer's discretion:

2 Two speaker feeds Lt, Rt, of any origin but including a mixdown from the surround, ***and either:***

- Six speaker feeds of any origin and paired as L, R: C, E: Ls, Rs, ***or***
- Six channels, including five speaker feeds L, R; C; Ls, Rs suited for five-speaker reproduction coded in a standardised way from *Ambisonic* sources so as to be decodable back to *Ambisonic* format, with Z carried in the E channel, ***or***
- Six speaker feeds of any origin and re-directable by subcode.[11]

These options shall be recognisable in co-temporal subcode and/or in a header at the start of the disc. (There are significant advantages to having both.)

12.3 Channel coding

1 Linear PCM

2 Lossless compression (packing) applied to all eight channels

12.4 Precision and sampling frequency

Two disc-maker options permitted:

1 48kHz, 24-bit maximum, ***or***

2 96kHz, 24-bit maximum

Normally 20 bits would be used at 48kHz and – depending on playing time – maybe only 16 or 18 bits at 96kHz. The producer has the option to use more data with material that compresses well, or when some channels are not used. (Obviously, limited by the maximum allowed data rate in the packed channel.) See the examples in Table 3, page 13.

12.5 MPEG type audio data stream

It is recommended that the *MPEG*-compatible packed audio stream should have, in all eight channels, separate fields for 48kHz-sampled and 96kHz-sampled signal components. These fields may be set to zero or to null status if not used, and will then occupy virtually no data rate in packed form. In the basic use, as described in section 0, either one or other of the fields would be null.

12.6 Pre- and de-emphasis

Not permitted for packed channels.

[11] This method might be advantageous if the standard finally adopted included the non-preferred option of having Lt and Rt be 48kHz non-compressed audio.

12.7 Two-channel compatibility

By provision of Lt and Rt at high precision.

12.8 Budget player compatibility

By provision of a lossy-compressed *AC-3*, *MUSICAM* or similar mix.

12.9 Digital outputs

Several requirements are highlighted:

- A high-speed interface to pass what is effectively the 'video' data stream from the CD decoder IC to an outboard decoder.[12]
- For players using on-board decoding, a multichannel serial interface capable of running at the higher speeds, say up to 96kHz. This serial interface should also carry channel status information; for example, on pre-emphasis and use of channels (Lt, Rt, L, R, C, E, Ls, Rs, Z, bass effects, *Ambisonics* etc.). We further recommend that the serial interface be coded in such a way that correlated jitter distortions are not generated by using the interface.
- For players using on-board decoding, there may be a need to down-sample 96kHz material to 48kHz to provide a standard SPDIF interface[13].

12.10 Additional data channels

- Some kind of header to determine use of subcode and identify use of channels, including E as bass effects, height or other, in addition to pre-emphasis flags.
- Additional data packets conveying copyright details, additional track information such as title, lyrics or musical score information, and multimedia or other applications may be inserted between audio data packets.
- Additional data packets conveying data for dynamic range control.

12.11 3-speaker support

To allow a simple standard method of downmixing the surrounds (Ls and Rs) into three front speakers, additional support is suggested in the data channels or header of 12.10 as follows:

- Coefficients or tables for the preferred mixing level and stereo width of Ls, Rs channels into L and R. These mixing coefficients may be standardised values or chosen by the producer for optimum artistic effect.

12.12 Absolute sound level datum

The header information of section 12.10 should include an indicator of the reproduced sound-pressure level by defining the acoustic gain required in the playback system. A code should be present that indicates 'not known'.[14]

It is possible that players could use this information to 'level' loudness on successive recordings.

---

[12] Any standards for a multichannel audio data stream should be compatible with equipment handling *MPEG-2* data streams, since *MPEG*-stream digital recorders and interfaces will be widely available and will be a convenient format for transferring recordings for mastering.

[13] The SPDIF interface does not consider sampling rates > 48kHz.

[14] In reproducing sound, particularly recordings of original acoustic origin, accuracy is increased if the loudness is the same as that of the original event.

Part 4: Supporting information

13. Who is taking part in this discussion?

Discussion-group members, with their relevant affiliations, are as follows.

13.1 Technical Subcommittee, Acoustic Renaissance for Audio

Tony Griffiths. Consultant Decca Recording Company. Fellow *Audio Engineering Society*, Member *Acoustic Renaissance for Audio*, Chairman *Technical Subcommittee National Sound Archive*, Member *IEE*, Member *Royal Television Society*.

Professor Malcolm Hawksford. University of Essex. Fellow *Audio Engineering Society*, Fellow *Institute of Acoustics*, Fellow *IEE*., Member *Acoustic Renaissance for Audio*.

David Meares. R&D Manager (Audio & Acoustics), BBC Research & Development Department. Fellow *Institute of Acoustics*, Member *Acoustic Renaissance for Audio*, Member *IEE*.

Bob Stuart. Chairman and Technical Director, Meridian Audio Ltd. Visiting Fellow Essex University, Fellow *Audio Engineering Society*, Member *Acoustical Society of America*, Chairman *Acoustic Renaissance for Audio*, Member *XtraBits*, Member *Technical Subcommittee National Sound Archive*, Member *IEE* and *IEEE*. Member *ADA* committee of *Japan Audio Society*.

13.2 Advisors

Peter Craven. Consultant. Member *Audio Engineering Society*, Member *XtraBits*.

Michael Gerzon. Consultant. Gold Medallist and Fellow *Audio Engineering Society*, Member *XtraBits*, Member *Acoustic Renaissance for Audio*.

Hiro Negishi. Director D&D Centre, Canon Inc. Member *Audio Engineering Society*, Member *Institute of Acoustics*, Founder *Acoustic Renaissance for Audio*.

Francis Rumsey. University of Guildford. Member *Audio Engineering Society*, Member *Acoustic Renaissance for Audio*.

Chris Travis. Division Ltd. Member *Audio Engineering Society*.

14. Bibliography and references

1 Akune, M., Heddle, R.M. and Akagiri, K. ‘Super Bit Mapping: Psychoacoustically Optimized Digital Recording’, *AES 93rd Convention* San Francisco, preprint 3371 (1992)

2 Craven, P.G. and Gerzon, M.A. ‘Compatible Improvement of 16-Bit Systems Using Subtractive Dither’ *AES 93rd Convention* San Francisco, preprint 3356 (1992)

3 Dunn, J. ‘High Dynamic Range Audio Applications for Digital Signal Processing’, *93rd AES Convention*, San Francisco, preprint 3434 (Oct. 1992)

4 Fielder, L. ‘Dynamic Range Issues in the Modern Digital Audio Environment’ *Proceedings AES UK Conference ‘Managing the Bit Budget’*, 3–19 (May 1994)

5 Hawksford, M.O. ‘Digital Frontiers’, *HiFi News*, **40**, no.2, 58–59 and 106 (Feb. 1995)

6 Gerzon, M.A. ‘Periphony: With-Height Sound Reproduction’ *J. Audio Eng. Soc*., **21**, 2–10 (Jan/Feb 1973)

7 Gerzon, M.A. ‘Ambisonics in Multichannel Broadcasting and Video’ *J. Audio Eng. Soc*., **33**, 859–871 (Nov 1985)

8 Gerzon, M.A. ‘Optimal Reproduction Matrices for Multispeaker Stereo’ *J. Audio Eng. Soc*., **40**, 571–589 (July/Aug. 1992)

9 Gerzon, M.A. ‘Problems of error-masking in audio data compression systems’ *AES 90th Convention*, preprint 3013 (Feb. 1991)

10 Gerzon, M.A., Craven, P.G., Stuart, J.R. and Wilson, R.J. ‘Psychoacoustic Noise Shaped Improvements in CD and Other Linear Digital Media’ *94th AES Convention*, Berlin preprint 3501 (March 1993)

11 Gerzon, M.A. ‘Hierarchical System of Surround Sound Transmission for HDTV’, *92nd AES Convention*, Vienna, preprint 3339 (March 1992)

12 Gerzon, M.A. and Barton, G.J. ‘Ambisonic Decoders for HDTV’, *92nd AES Convention*, Vienna, preprint 3345 (March 1992)

13 Gilchrist, N.H.C. ‘DRACULA: dynamic range control for broadcasting and other applications’ **18**, 36–47, Tonmeistertagung, Karlsruhe (article in English) (1994).

14 Hoeg, W., Gilchrist, N., Twietmeyer, H. and Juenger, H. ‘Dynamic Range Control (DRC) and music/speech control (MSC)’ *EBU Technical Review*, 56– 70, (Autumn 1994)

15 Komamura, M. ‘Wideband and wide dynamic-range recording and reproduction of digital audio’ *AES 96th Convention,* Amsterdam, preprint 3844 (1994)

16 Meares, D.J. ‘Perceptual Attributes of Multichannel Sound’ *Proceedings of AES 12th International Conference ‘The Perception of Reproduced Sound’*, 171–179 (June 1993)

17 Meares, D.J. ‘High definition sound for high definition television’, *Proceedings of the AES 9th International Conference ‘Television Sound, Today and Tomorrow’*, Detroit, 187– 215 (Feb. 1991)

18 Meares, D.J. and Stoll, G. ‘Sound systems in Digital Television’, Technical Module of DVB, Document 1172, *EBU* (1993)

19 Ohashi, T., Nishina, E., Kawai, N., Fuwamoto, Y. and Imai, H. ‘High Frequency Sound Above the Audible Range Affects Brain Electrical Activity and Sound Perception’, *91st AES Convention*, New York, preprint 3207 (Oct. 1991)

20 Ohashi, T., Nishina, E., Fuwamoto, Y. and Kawai, N. ‘On the Mechanism of 'Hypersonic Effect’, *Proceedings Int'l Computer Music Conference*, Tokyo, 432–434 (1993)

21 Oomen, A.W.J., Groenwegen, R.G., van der Waal, R.G. and Veldhuis, R.N.J. ‘A Variable-Bit-Rate Buried-data Channel for Compact Disc’ *J. Audio Eng. Soc*., **43**, 23–28 (Jan./Feb. 1995)

22 Stuart, J.R. and Wilson, R.J. ‘A search for efficient dither for DSP applications’ *AES 92nd Convention*, Vienna, preprint 3334 (1992)

23 Stuart, J.R. ‘Noise: Methods for Estimating Detectability and Threshold’ *J. Audio Eng. Soc*., **42**, 124–140 (March 1994)

24 Stuart, J.R. and Wilson, R.J. ‘Dynamic Range Enhancement Using Noise-shaped Dither Applied to Signals with and without Pre-emphasis’ *AES 96th Convention*, Amsterdam, preprint 3871 (1994)

25 Stuart, J.R. ‘Auditory modelling related to the bit budget’ *Proceedings of AES UK Conference ‘Managing the Bit Budget’*, 167–178 (1994)

26 Stuart, J.R. ‘Perceptual issues in Multichannel environments’ *97th AES Convention*, San Francisco (1994)

27 Vanderkooy, J. and Lipshitz, S.P. ‘Digital Dither: Signal Processing with Resolution Far Below the Least Significant Bit’ *AES 7th International Conference – Audio in Digital Times*, Toronto, 87–96 (1989)

15. Glossary

AC-3 A system for perceptually encoding at a reduced data rate both two-channel stereo and 5.1-channel surround sound. *AC-3* has been developed by *Dolby Laboratories*; it is used in many motion-picture films and *LaserDisc* releases, and has been selected for television broadcast in the USA.

Ambisonics A method of recording and playing back directional sound over all horizontal directions, or the full sphere of directions including height, based on transmitting directional components of the sound field rather than loudspeaker feeds, and of reproducing the sound field by deriving signals psychoacoustically optimised for the user's specific loudspeaker layout.

CD-DA Red Book CD for Digital Audio.

DVD Originally Digital Video Disc – a high-density disc carrying *MPEG-2* encoded variable-rate video with lossy-compressed audio. More recently redefined as Digital Versatile Disc. A collective term for the new generation high-density CDs.

DSP Digital signal processing.

HQAD High-Quality Audio Disc. New format high-density CD applied to audio, as proposed in this document.

Lossless compression A process by which the data of a PCM audio signal can be more efficiently packed into a channel. Although lossless compression of audio does not work in the same way, users of computers will be aware of algorithms such as ZIP and LZW that allow more efficient use of disc storage. Lossless compression of audio has the same effect: less space is used on the disc, which has the important effect of reducing the data rate. Unlike lossy compression, lossless compression systems return the *input data exactly* from a decoder. For clarity, in this document losslessly compressed PCM is referred to as 'packed audio'.

Lossy compression A process by which an audio signal is examined from a human-psychoacoustic viewpoint. An algorithm attempts to estimate and remove the inaudible components of the signal. The remaining 'audible' component is efficiently coded in the output channel. Lossy compression schemes include *MPEG* audio, *PASC*, *AC-3* and *MUSICAM*. The data recovered from a matched decoder is *not* identical to the original input, although it may sound very similar.

MMCD The trademark for the original *Philips/Sony* proposed high-density disc described in the 'Gold Book'. Now incorporated in DVD.

MUSICAM A system for perceptually encoding at a reduced data rate both two-channel stereo and multichannel surround sound using the *MUSICAM Surround* version. The two-channel version of *MUSICAM* forms layer 2 of *MPEG-1*.

MPEG 'Motion Picture Experts Group' refers to standards for perceptual coding at a reduced data rate of video and sound signals. *MPEG-1* and *MPEG-2* are respectively video-coding standards for medium and high-quality use, and *MPEG-1* layers 1, 2 and 3 are systems for perceptually encoding two-channel stereo sound.

Packed audio The data resulting when a linear PCM audio stream is losslessly compressed.

Packing The process of losslessly compressing linear PCM audio.

PASC A system for perceptually coding two-channel stereo sound at a reduced data rate, developed by *Philips* and used in the Digital Compact Cassette. *PASC* is related closely to *MPEG* layer 1.

PCM Pulse code modulation. A method of coding whereby a signal is represented by a discrete-sampled series.

SD The original code abbreviation for the *Toshiba/WEA* 'Super-Density' disc, now incorporated in DVD.

Unpacking The process of decoding losslessly compressed (packed) audio back into the original linear PCM full-rate data.

16. Acknowledgements

Dolby, *Dolby Surround*, *Pro Logic* and *AC-3* are trademarks of Dolby Licensing Inc.

MUSICAM, *PASC*, *MPEG* and *DVD* are registered trademarks.

HDCD is a trademark of Pacific Microsonics.

Ambisonics is a registered trademark of Nimbus Records Ltd.

MMCD is a registered trademark of *Philips/Sony* licensing.

SD is a registered trademark.

17. Diagrams

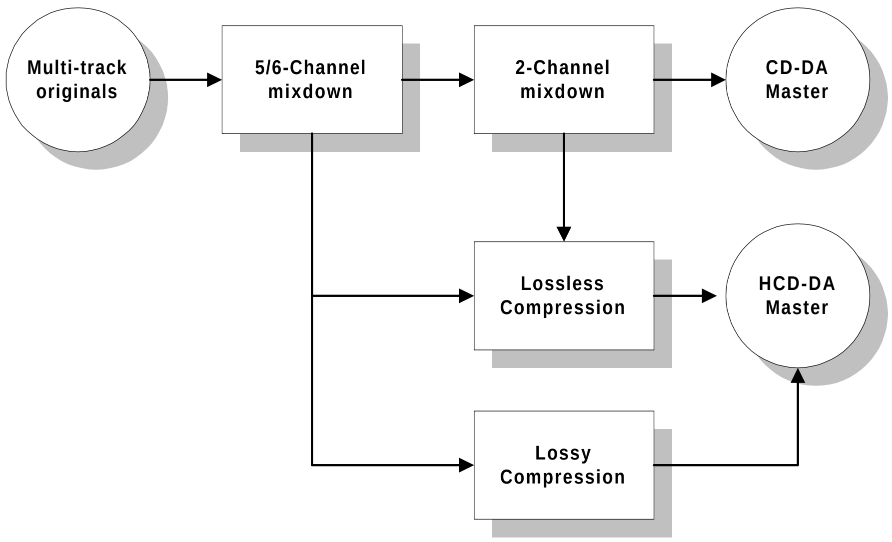

Figure 1. General production scheme of HQAD Master

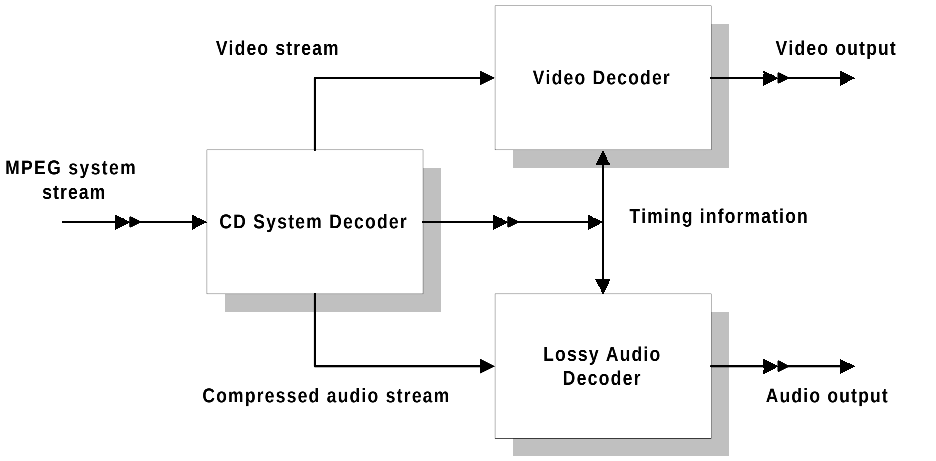

Figure 2. General decoding scheme of DVD video player

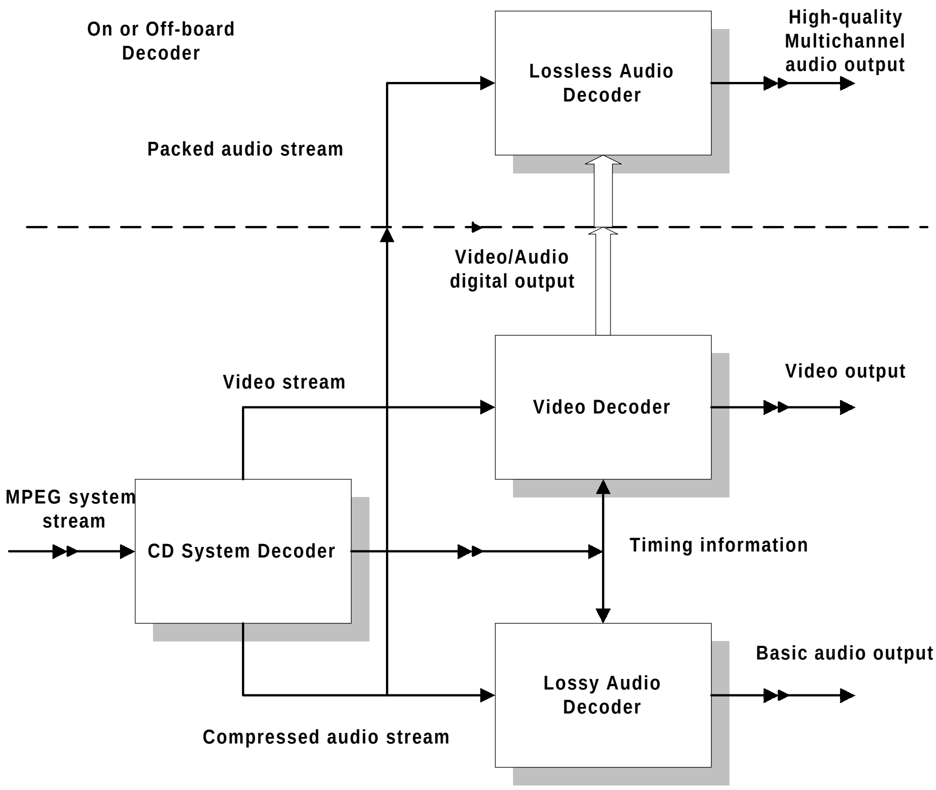

Figure 3. General decoding scheme of HQAD player. Showing possibility of the high-quality audio decoder being outboard.

18. Appendix: High-density formats, facts and assumptions

The following data are taken from public announcements made by the DVD group in December 1995.

18.1 DVD basic parameters

Codename	*DVD*
Disc diameter	120mm
Disc thickness	1.2mm (two 0.6mm bonded)
Memory capacity	4.7GB per layer
Track pitch	0.74 micrometer
Laser	640nm
N.A	Aperture 0.6
Error correction	RS-PC ‘CIRC Plus’
Modulation	Not 8-16
Play time:	
1. Movies	133 min/layer @ 4.69Mb/s average
2. Broadcast	74min/layer @ 9Mb/s average
Picture code	*MPEG*-2 variable rate 1-9Mb/s
Audio code	*AC-3* (5.1 channels) for NTSC issues
	MPEG audio for PAL/SECAM issues
Other features	Multiple aspect ratio, parental lock-out, 2-layer(sided) format

18.2 Conjecture on player architecture

- The sector format will resemble Video-CD to give compatibility with existing CD-DA and CDROM playback for the decoder ICs. These formats all provide 2048 bytes per 2072-byte block. The remaining data used for time-stamping and error-checking/correcting, depending on mode.
- A player has to co-ordinate the data outflow using time-stamps in the *MPEG* transport layer to achieve variable-speed reading from the disc. We anticipate that the variable read rate will be achieved by a constant spinning speed and adaptive skipping. This means that the CD decoder IC will have to perform the function of separating audio, video, data and other streams.
- The CD decoder IC must also be able to read CD-DA, which uses different modulation, error-correction and interleave.
- Following the CD decoder we expect in early machines to see an *MPEG1/2* decoder, allowing the player to process Video CD (*MPEG1*) and Karaoke CD (*MPEG1*) as well as *DVD*. This decoder is well known in the art and will output digital composite video.
- The player may output the *MPEG*, digital video or composite analogue video according to market level.
- Following the CD decoder we expect to see a stock audio *AC-3* or or *MPEG audio* decoders which output two channels (when necessary *Dolby Surround* encoded) to on-board Lt and Rt DACs running at 48kHz.

19. Document revision history

19.1 Version 1.1

Original version released 12 April 1995.

19.2 Version 1.2

The embargo is released and text updated.

- Reflects trademarks/descriptions *SD* (*Toshiba/WEA*) and *MMCD* (*Sony/Philips*).
- Minor editorial changes.
- Table 2 and section 18 reflect recent specification changes/clarifications including higher peak data rate of 11.2Mb/s for *MMCD*.
- Section 7.6 added on Red Book compatibility.
- Section 11.1 added giving examples of disc use.
- Table 3 added illustrating more examples of the flexibility of packing.
- Summary for commentators

19.3 Version 1.3

- Updated to account for *SD* + *MMCD* collaboration in new DVD task force.
- Minor editorial changes.
- Section 7.6 updated to cover double-sided issues with old CD players.
- Table 2 updated to allow for single capacity of 4.7Gb/layer and raising AC-3 to 448kb/s
- Table 3 modified for single capacity of 4.7Gb/layer and raising AC-3 to 448kb/s. More examples added.
- Outline DVD specification brought up to date.
- Summary for commentators updated.

A Proposal for the High-Quality Audio Application of High-Density CD Carriers

Technical Subcommittee
Acoustic Renaissance for Audio

Confidentiality. Update Version 1.3, 1 January 1996. This document has been prepared by the Technical Subcommittee of *Acoustic Renaissance for Audio.* An early version (1.1, released 12 April 1995) was circulated to members of *Acoustic Renaissance for Audio* and to members of the syndicates developing high-density CD and related items. With Version 1.2 (23 June 1995), the circulation was extended to interested members of the audio community and the document made available on the World Wide Web. In order that the issue be correctly reported, we request prior consultation if any of this material is to be discussed in the Press.

Contents

A Proposal for the High-Quality Audio Application of High-Density CD Carriers

Technical Subcommittee

Acoustic Renaissance for Audio

Outline notes for commentators

Background

The attached document has been prepared as a proposal for high-quality audio use of high-density CD carriers that we are making to those involved in standardising the DVD format for audio – notably the 'DVD Taskforce' of *Toshiba, Matsushita, Sony, Philips, Time Warner, Pioneer, JVC, Hitachi* and *Mitsubishi* along with the *Advanced Digital Audio Committee* (ADA) of the *Japan Audio Society*.

An earlier embargoed version was circulated among the original *Toshiba/WEA* and *Sony/Philips* proponents for *SD/MMCD* and it was well received. We were asked by both parties to widen the circulation of this proposal to include some members of the *Audio Engineering Society* (AES), of the *Japan Audio Society* (JAS) and of the *Academy for the Advancement of High-End Audio* (AHEA) and of the audio community in general. The committee of the *Acoustic Renaissance for Audio* (ARA) decided that this amounts to public-domain circulation, and that the process of calling for such a carrier would now benefit from wider publicity. Consequently Version 1.2 was widely circulated and has received considerable press comment, including being reprinted in full in the August 1995 issue of *Stereophile* and the October issue of the *Journal of the Japan Audio Society* (in Japanese translation). This proposal has also been available on the World Wide Web[15].

Although originally made as a submission, the *ARA* have been encouraged by many industry parties to take an active approach in promoting the HQAD proposal.

The urgent purpose of this proposal is to ensure that any special requirements of the HQAD (High Quality Audio Disc) are not excluded in the standardising process for the DVD. It is very important that first-generation video disc players be able to accommodate HQAD should it come to fruition.

It should be clearly understood that the standardisation process for HQAD cannot move on the same time scale as the DVD or CDROM versions of the high-density carriers. Apart from anything else, the audio community does not currently have significant material, mastering equipment or standardised interfaces to achieve the proposals set out. We would estimate that with strong support the HQAD could be available by 1998, and firmly established by the turn of the century. Obviously some companies may well demonstrate or supply custom audio carriers sooner, but we prefer to encourage a wide and universal standard that will meet rich and varied top-quality audio requirements for more than a decade.

Features and Benefits

The essential elements of the HQAD proposal are:

- Audio is coded as linear PCM sampled at either 48kHz or 96kHz. Lossy compression (data-reduction) methods like *PASC, ATRAC, MPEG* or *AC-3* are ***not*** used except for backup.

[15] ARA Home Page at http://www.meridian.co.uk/ara/

- Eight channels are provided. This allows full 3-D surround sound with up to 6 channels *as well as* a separate (conventional) 2-channel feed. The 6 channel surround recording can incorporate bass effects as 5.1 compatible material or effectively encode soundfield methods like *Ambisonics.*
- Each channel can have up to 24-bit precision. Normally 20 bits would be used at 48kHz or 18-bit at 96kHz.
- The proposal centres on encoding the disc using *lossless* compression – a technique we call 'packing'. Lossless compression efficiently packs the data into the available space – in an analogous way to the lossless compression schemes seen on computer discs[16]. The proposal asks for this to be standardised as an option in the *MPEG* compression layer.
- By providing lossless compression (packing), flexible use of the carrier can automatically be provided. The producer can then establish a trade-off between precision, frequency bandwidth, number of channels and playing time.
- 2-channel compatibility through providing a specific mix in addition to the surround version.
- Playing times are dependent on the sample rate, precision and number of channels. The combinations are infinite, but for example, the disc could provide:
  - ⇒ 104 minutes of 8 channel 20-bit material at 48kHz
  - ⇒ 93 minutes of 7.1 channel 24-bit material at 48kHz
  - ⇒ 472 minutes of 2 channel 16-bit 48kHz archive material
  - ⇒ 93 minutes of 8 channel 16-bit material at 96kHz
  - ⇒ 182 minutes of 2 channel 24-bit material at 96kHz.
- The proposal is drawn up in an hierarchical way so that the lowest-common-denominator DVD player carries no real additional cost to accommodate the disc and so that hardware is progressively added to replay machinery to access higher quality.

Recent Activities

Bob Stuart, Chairman of *ARA*, was invited to serve on the *ADA* committee of the *Japan Audio Society*. This focus group aims to comment on the desirable forms for audio for the next century.

ARA formed a Japan wing *'Japan ARA'*, with an inaugural meeting in late October 1995. This group now has ten members – some closely associated with the DVD standardisation process. *ARA* is particularly proud to have Takeo Yamomoto (*Pioneer*) as a founder of the Japan group.

In October, the *ARA* called a special meeting of the *ADA* in Tokyo to present the HQAD proposal along with a series of demonstrations illustrating the benefits of multichannel audio. This meeting, which was graciously hosted by *Pioneer*, was presented by Bob Stuart and Hiro Negishi. It was felt that this presentation helped substantially to ensure that the multichannel options were not ignored in the *ADA* deliberations. The *ARA* 'mission' to Tokyo established important links in the development of DVD audio.

The *ARA* proposal was presented in full by the *Technical Committee* to the *Audio Engineering Society* in London during November.

---

[16] Computer-based lossless packing schemes include 'DoubleSpace' (Microsoft Trademark) and 'ZIP'.

Feedback to ARA

The *ARA* has received a large number of supporting messages/letters from audio manufacturers, journalists and commentators all over the world. There have been no dissenting opinions expressed on the objectives and approach laid out in this document.

It should be understood however that the *ARA* proposal goes beyond the bare minimum we might have expected from standardising committees for DVD. The notable contentious areas have been: multichannel, lossless compression, extension to 96kHz sampling and no provision for 44.1kHz based sampling.

Progress to HQAD

The *ARA* proposal remains the only independent offering to the DVD group. Serious proposals on High-Quality Audio have also been made by *Pioneer* and *Sony*.

Philips – after review of the *ARA* proposal – in effect decided to support it as it stands.

Pioneer'*s* proposal was consistent with that of the ARA in many respects, but was less radical proposing fewer channels without lossless compression. Recent months have seen *Pioneer* move to support many of the aims of the *ARA*.

Competition to HQAD

There are two fundamental competitors to the *ARA* HQAD proposal.

Lossy compression *DTS.* The first proposes to use lossy compression and has been suggested by *DTS.* Whilst the *ARA* fully accept that at some stage psychoacoustically optimised coding may offer more efficient transmission, the following points seem paramount:

1. There is more than adequate data capacity on the HQAD to avoid using lossy compression for high-quality audio
2. None of the currently-available lossy algorithms have withstood sufficient testing in the arena of high-quality high-bit multichannel audio to be suitably chosen for this generation of carrier.

Bitstream coding. A sincere proposal, based on minimum-cost replay hardware has been put forward by *Sony et al.* The proposal, as it is understood, suggests that the DVD Audio disc carries 2 channels of a bitstream (1 bit, 64x oversampled, 7th order Delta-Sigma code). The *ARA* feels quite strongly that whilst bitstream coding clearly works, it is not data-efficient and the provision of 2-channels and denying sensible signal processing at editing and replay are powerful counter-arguments. The *ARA* has published a paper by Prof. Malcolm Hawksford arguing for PCM and against Bitstream at this time[17].

[17] Available from ARA secretariat or on WWW at http://www.meridian.co.uk/ara/