Productivity Commission Inquiry into Broadcasting

Submission by:

Dolby Laboratories Inc.

San Francisco, CA

November 29, 1999

1. Preface

1.1 Dolby Laboratories conceives, designs, develops and deploys audio signal processing products and technologies. The company manufactures professional audio equipment for the motion picture, broadcasting, and music recording industries. Dolby also licenses audio signal processing technologies for a wide variety of entertainment applications in the consumer electronics and computer industries. Providing the best possible audio for any entertainment environment, including music, movies, television and multimedia, is Dolby Laboratories' principal mission. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo, and European headquarters in England.

1.2 Dolby has studied the Productivity Commission's draft report on Broadcasting, as well as many of the submissions that provided information to the Commission. We believe that some of the information provided to the Commission is incorrect, and that the draft report contains inaccurate statements about Dolby AC-3 audio. Dolby Laboratories appreciates this opportunity to provide factual information for consideration by the Commission.

1.3 Three mischaracterizations regarding Dolby AC-3 technology are particularly egregious, namely:

- that it is "proprietary"
- that it adds significant cost to decoders
- that inclusion of AC-3 is not "true DVB"

1.4 AC-3 technology is an international standard, and Dolby has made the customary commitment to open and fair licensing. The cost of AC-3 decoder circuit implementations is currently low, and is rapidly becoming negligible as decoder chips become more highly integrated. The royalty cost of AC-3 decoders is modest. While slightly higher than MPEG-1 stereo audio decoders, product manufacturers (Dolby licensees) and consumers using end products with AC-3 technology receive good value for the modest royalty expense. At the request of Australian broadcasters and other countries, the DVB Project recently incorporated AC-3 technology into the DVB Standard. Equipment with AC-3 decoders and DTV transmissions with AC-3 audio (even transmissions sans MPEG audio) are fully DVB system compliant.

1.5 Section 2 of this submission provides background information on AC-3 technology and its rapid worldwide adoption for a myriad of important emerging applications. Section 3 addresses the issue of intellectual property (IP) in standards noting that there is no significant difference between the IP situation with AC-3 and other technologies such as MPEG video or audio, COFDM modulation, GSM phone technology, etc. Section 4 addresses the cost of AC-3 in terms of the physical silicon (chip) cost as well as the patent royalty cost. In Section 5 the issue of spectrum efficiency and the impact of making certain receiver capabilities optional are discussed. In order to keep the body of this submission brief, additional supporting information is included in a series of numbered attachments.

2. AC-3 Multi-Channel Audio Technology

2.1 Dolby Laboratories has led advancements in multi-channel audio technology since the introduction of the Dolby Surround system in cinema applications in the 1970's. In the 1980's Dolby introduced this same matrix surround sound system into the home and launched the home theater revolution. In the 1990's Dolby introduced the AC-3 digital multi-channel sound system—first in the cinema, then in domestic (home) applications.

2.2 The domestic version of AC-3 (called "Dolby Digital" in the consumer marketplace) is not simply a delivery "pipe" for audio signals: it is a comprehensive audio entertainment delivery system wherein one encoded bit stream simultaneously provides for optimized audio reproduction in listening scenarios that span the range from the monophonic television receiver to the multi-channel home theatre. The AC-3 system includes an extensive set of sound control features that constitute solutions to the real problems of optimal sound reproduction in diverse consumer listening environments. System features provide means to normalize variations in dialogue volume between programs, to optimize reproduced dynamic range (e.g., in quiet vs. somewhat noisy listening environments), and to optimize reproduction with different configurations of loudspeakers (e.g., from mono to multi-channel sound).

2.3 Multi-channel audio is not just for the home theater enthusiast! An enjoyable multi-channel effect can be synthesized from an AC-3 encoded signal using just two front loudspeakers and a cost-effective signal processing technique known as "Dolby Virtual" (attachments 1 and 2), or via headphones using another technique known as "Dolby Headphone"[1] (attachments 3 and 4). A detailed technical overview of the AC-3 system (attachment 5) is provided for further information.

2.4 MPEG-1 audio technology was designed to provide a pipe for two audio channels. With MPEG-2 audio, this pipe was extended to be multi-channel, but with the constraint of backwards compatibility with the earlier 2-channel MPEG-1 system. This restriction imposes a devastating compromise in performance that prevents MPEG-2 audio from matching the audio performance of AC-3 audio at comparable, practical bit rates. The performance benefits of AC-3 have been underscored in every critical listening test of the two systems—this includes comparisons performed by the U.S. Grand Alliance, MPEG, ARIB (Japan, formerly BTA), the European Broadcasting Union (EBU), and the Canadian Research Center (CRC).

2.5 The audio quality delivered by AC-3, the compelling feature set, the availability of a multitude of cost-effective consumer decoders, and the high level of applications engineering support provided by Dolby Laboratories to build a robust, supporting infrastructure have led to widespread adoption of AC-3 for applications including:

- U.S. digital terrestrial television
- U.S. digital cable television
- LaserDisc
- DVD (initially NTSC, later PAL)
- U.S. direct broadcast satellite including DirecTV (attachment 6), and Echostar
- DVB (attachment 7)

[1] Dolby Headphone technology was invented by Lake DSP Pty., an Australian company that engaged Dolby Laboratories to license and commercialize the system.

- Digital television in Australia and Singapore (attachment 8)
- Digital television in Europe (attachments 9, 10, 11 and 12)

2.6 The adoption of AC-3 for DVD applications *worldwide* is an interesting case to explore further. Initially it was thought that the NTSC television world would adopt AC-3, while the PAL television world would select MPEG audio. The initial DVD specification specified AC-3 for the NTSC (525-line/60 Hz) versions of DVD players, while MPEG audio was specified for PAL (625-line/50 Hz) players. However, multi-channel audio became a "must-have" feature for DVD, and the overwhelming advantages of the AC-3 system (i.e., the combination of audio performance, availability of professional encoders, consumer integrated circuit decoders, reference designs and technical support) led to a change in the DVD specification to include AC-3 in the PAL version. The current situation is that an NTSC DVD player may have a single audio decoder (AC-3), while the PAL player must have two audio decoders (AC-3 and MPEG). This is a good example of the free (from imposed regulation) market choosing the very slight increase in cost/complexity of dual decoding in order to gain the compelling benefits delivered by AC-3 technology. Excerpts of relevant industry news items (attachment 13) describe then current developments in DVD audio specifications and market reactions.

3. AC-3 Technology – A Recognized International Standard

3.1 AC-3 is an internationally recognized sound coding standard. The technology is fully documented in the customary form for international standardization (e.g., equivalent to MPEG-2 audio), including a complete specification of the audio bitstream syntax and decoder definition in a publicly available specification.[2] AC-3 international standard technology is readily accessible to product manufacturers worldwide under license, and is widely acknowledged as the pragmatic solution for a host of consumer applications including sound services for digital television.

3.2 Examples of regulatory and standards setting bodies that have embraced or adopted AC-3 technology for professional and consumer applications in digital television include:

- The U.S. Advanced Television Systems Committee (ATSC)
- The FCC Advisory Committee on Advanced Television Services (ACATS)
- The U.S. Federal Communications Commission (FCC)
- The International Telecommunications Union (ITU-R)
- The Digital Video Broadcast Consortium (DVB)
- The Digital Versatile Disc (DVD) Forum
- The Digital Audio-Visual Council (DAVIC)

One very relevant example of international standardization is ITU-R Recommendation BS.1196, which recommends AC-3 technology[3] for terrestrial digital television services.

3.3 Effective standards efforts often benefit from creative technical contributions from commercial enterprises. MPEG Layer II audio was largely developed through the efforts of three research organizations including Philips (Netherlands), CCETT (France) and the

[2] The specification is available at http://www.atsc.org/Standards/stan_rps.html.
[3] MPEG-2 audio is also a subject of this ITU-R Recommendation.

IRT (Germany). In the case of AC-3, Dolby Laboratories was responsible for the research and development efforts. In both instances, however, the technologies were discussed and debated in open standards processes before being finalized, documented and ratified.

3.4 MPEG-2 and AC-3 audio technologies are the subject of international patents held by their respective developers. AC-3 is no more “proprietary” than MPEG-2 audio in this regard, as royalty-bearing patent licenses are required to practice either of the respective technologies in the marketplace. There is a great benefit with AC-3 licensing, however, as Dolby Laboratories provides worldwide technical support to ensure that consumers benefit from the interoperability of products using AC-3 technology.

3.5 Since the practice of either of the MPEG-2 or AC-3 audio standards necessarily involves practice of intellectual property rights protected by patents, the standardization processes for both technologies necessitated statements from rights holders indicating their willingness to license their relevant patents on a non-discriminatory basis under reasonable licensing terms. A list of the rights holders making such claims for MPEG-2 audio is provided (attachment 14). A letter from Dolby that underscores Dolby’s willingness to license AC-3 patent rights to enable widespread use of the technology in reference to the ITU-R BS.1196 Recommendation is also provided (attachment 15).

4. The Cost of AC-3 Technology

4.1 As is the case with any valuable technology, the substantial benefits of AC-3 are accompanied by some costs. These costs include the circuitry required in the receiver to decode the AC-3 encoded signal, and patent royalties paid to the owners of the intellectual property (IP) that is employed to implement the decoder in the receiver.

4.2 Like most modern technologies, the cost of AC-3 circuitry started out high, has fallen to moderate levels, and is rapidly trending towards insignificance due to advances in semiconductor technology, increases in production volumes, and the high level of system integration in consumer products. The initial implementation in the cinema industry (circa 1991) employed a set of five expensive digital signal processing (DSP) chips in a professional product that initially sold for nearly $20,000 U.S. Within a few years the AC-3 decoder was implemented on a single DSP chip, and began to penetrate the high-end consumer marketplace in products selling for a couple of thousand dollars. Today there are numerous AC-3 decoder chips that sell for less than $10 in volume. Many of these chips also include other complementary audio functions. Home theatre decoding products—even some that include six channels of power amplification and a modest complement of loudspeakers—sell for less than $350 U.S.

4.3 The cost of embedded AC-3 decoders in audio-visual (A/V) and television receiver products, DVD players, and set-top boxes is much lower. Continued systems integration has led to inclusion of the AC-3 decoder function on the same chip as the MPEG-2 video decoder. In the supporting example, a portion of the data sheet for a complete HDTV decoder chip is provided (attachment 16). This moderate cost combination video/audio decoder chip accepts the entire HDTV bit stream from a data demodulator (VSB modulation in the case of the ATSC DTV system), and produces both video and audio outputs. For cost sensitive product applications the HDTV picture is down-converted for display at SDTV resolution, while the AC-3 multi-channel audio is down-mixed for presentation via stereo loudspeakers in a way that preserves an enjoyable multi-channel effect (Dolby Virtual, as noted above). Consumer receiver/decoder products using this chip have appeared in the U.S. marketplace at an initial selling price of less than $300 U.S.

4.4 The results of a study of the basic silicon cost of the AC-3 decoding function are captured in a document originally submitted to the ITU-R (attachment 17). This study—based on the results of actual chip designs—illustrates that the true silicon cost of a combined MPEG and AC-3 audio decoder is on the order of $0.30 U.S. today. More importantly, the cost is already trending downward toward a cost of $0.10 in the not too distant future. The argument that dual audio decoders (MPEG plus AC-3) are prohibitively expensive for consumer products is fallacious and misleading as such dual decoders are used today in cost effective consumer products.

4.5 The other cost issue is that of intellectual property. There are modest IP costs associated with both MPEG audio and AC-3 audio. The licensors of each technology (Philips in the case of MPEG audio, Dolby in the case of AC-3) have established a sliding scale where royalty payments are based on volume: as production volume increases, the per unit royalty cost decreases. Both audio technologies have reached high volumes of production, and the royalty costs for manufacturers of mass-market receiving equipment that will be used in Australia will be paying relatively low incremental royalty costs.

4.6 Dolby Laboratories understands that the high volume incremental royalty for an MPEG stereo decoder is $0.40 U.S. Dolby anticipates that the comparable incremental royalty for an AC-3 decoder with a 2-channel output would be in the range of $0.54 to $0.62 U.S. While the AC-3 royalty is somewhat higher than the MPEG audio royalty, Dolby respectfully submits that the greater value and longevity of AC-3 technology, and the substantial support provided by Dolby to its licensees and the industry in general defines a very workable value proposition for all constituents. If the burden of two audio royalty payments is deemed too high for Australian consumers to bear, Dolby respectfully suggests that the MPEG audio function be disabled in decoder equipment and that the MPEG audio royalty be eliminated.

5. Spectrum Efficiency

5.1 Dolby reviewed arguments submitted in this inquiry suggesting that Australian broadcasters be subject to mandatory transmission of SDTV video with MPEG audio (mono or stereo) so as to enable reception by lower cost receivers. These arguments suggest that multi-channel audio (and HDTV video) could be transmitted as additional, perhaps optional, services in a “simulcast” mode. Dolby also reviewed arguments suggesting that multi-channel audio (and HDTV video) would use too much spectrum, thus potentially inhibiting growth of datacasting services.

5.2 AC-3 operates with superior spectrum efficiency compared to MPEG audio. AC-3 delivers near-transparent stereo audio at a bit rate of 192 kb/s. The MPEG audio coder requires a bit rate of 256 kb/s—33 percent higher bit rate (i.e., poorer spectrum efficiency)—to deliver audio of similar quality. The spectrum efficiency advantage of AC-3 technology can be leveraged to economic advantage (e.g., data broadcasting) if broadcasters are permitted to transmit only AC-3 audio (i.e., AC-3 mono-cast) without a requirement to simulcast the less spectrum efficient MPEG audio.

5.3 While AC-3 delivers a useful gain in spectrum efficiency for stereo audio transmission, the spectrum efficiency of an AC-3 mono-cast (no MPEG audio) becomes overwhelming when multi-channel audio is broadcast. Multi-channel AC-3 bit streams serve those listening in mono and stereo via down-mixing;[4] i.e., there is no need to transmit a separate bitstream for those listeners. In the suggested AC-3 mono-cast scenario, a mandate to simulcast an MPEG mono or stereo audio bitstream wastes the bit rate dedicated to transmission of the MPEG audio.

5.4 Direct broadcast satellite (DBS) services in the U.S. and many of the digital video broadcast (DVB) services in Europe began operation prior to the time that AC-3 was ready for market application. These broadcast services, of necessity, began with MPEG stereo audio. Now that multi-channel audio is becoming an important feature,[5] these services are adding AC-3 audio in a simulcast mode. Existing services are, perhaps forever, saddled with the efficiency robbing requirement to simulcast stereo with multi-channel sound, because they began operation with MPEG stereo and must keep that original stereo service active so as not to silence early receivers.

5.5 In terrestrial broadcasting in Australia, where spectrum efficiency is of great importance, the situation of a “legacy” mono or stereo simulcast should be avoided. If every terrestrial DTV receiver can decode AC-3 audio, the terrestrial service can benefit from the improved spectrum efficiency of AC-3 technology for mono and stereo audio, as well as avoid a need to simulcast stereo audio whenever AC-3 multi-channel sound is delivered.

6. Conclusion

6.1 Dolby Laboratories respectfully suggests that Australian authorities look to the future by selecting and leveraging the substantial and compelling benefits of AC-3 technology to the greatest economic and technological advantage of the Australian population. AC-3 technology is the ideal choice for Australian terrestrial digital television where a new transmission standard must withstand the test of time by simultaneously serving a broad cross section of evolving consumer and commercial market interests.

6.2 The overall implementation cost of AC-3 technology is modest today, and will have a nearly negligible impact on prices of consumer DTV receivers in the long run. AC-3 multi-channel audio can benefit those listening with stereo speakers or headphones by the employment of “virtual” surround techniques. AC-3 technology is fully supported by the DVB standard. AC-3 technology delivers improved spectrum efficiency for mono or stereo audio, and significantly improved spectrum efficiency when multi-channel sound is broadcast. AC-3 technology offers a valuable, future-proof feature set to consumers and broadcasters alike. And finally, AC-3 technology is no more “proprietary” than many other technologies that are the subject of international standards. AC-3 is the best audio technology choice for Australia.

[4] Down-mixing involves reproduction of *all* of the artistic content of the “N” channels in the transmitted multi-channel program, where N for a modern motion picture film might include five full-bandwidth channels plus an optional low frequency effects channel (i.e., 5.1 channels), using just two (stereo) or one (mono) loudspeaker.
[5] The importance of multi-channel sound is increasing in the minds of consumers driven by the growth in surround sound systems, home theater and new multi-channel delivery formats such as DVD.

7. List of Attachments

No.	Attachment Title/Description	Pages
1	Dolby Laboratories Information: “Virtual Surround Reproduction”	4
2	Fall 1999 Information Package: “Virtual Technologies”	2
3	Dolby Press Release: “Personal surround sound arrives via *Dolby Headphone*,” December 1998.	2
4	Dolby Press Release: “Dolby Headphone brings cinema-quality surround sound to the aircraft cabin,” April 1999.	2
5	Technical Paper: Steve Vernon, “Dolby Digital: Audio Coding for Digital Television and Storage Applications,” AES 17th Conference on High Quality Audio Coding.	18
6	Dolby Press Release: “Starz! & DirecTV launch 5.1 Dolby Digital Theatre-quality audio on a premium channel,” June 1999.	2
7	Dolby Press Release: “DVB Project Recognizes Dolby Digital as a Digital Audio Standard,” July 1999.	2
8	Dolby Press Release: “Dolby Digital selected as a key factor in Singapore’s DVB future,” June 1999.	2
9	Dolby Press Release: “Pro Seiben is first broadcaster in Europe to transmit Dolby Digital 5.1-channel sound,” July 1999.	2
10	Dolby Press Release: “Set-top box manufacturers prepare for European broadcasts with Dolby Digital 5.1 audio,” August 1999.	2
11	Dolby Press Release: “Canal + and Sony introduce a new generation of digital TV set-top boxes with Dolby Digital 5.1-channel sound at IFA’99,” August 1999.	2
12	Dolby Press Release: “First DVB demonstration in France with Dolby Digital 5.1-channel audio,” October 1999	1
13	News letters: DVD Intelligence, Feb. 1998, July 1998	1
14	ISO/IEC Document 13818-3:1994(E): “Annex E,” List of patent holders for MPEG-2 technologies, 1994.	2
15	Dolby Letter to ITU-R, patent statement, June, 1995.	1
16	LG Electronics’ GDC21S802A: “HDTV All-format Single Chip Decoder,” April 1999.	12
17	ITU-R Document 10C/25E: “Complexity of Decoders for the Systems Specified In Recommendation ITU-R BS.1196,” May 1999.	4

Virtual Surround Reproduction

Vir-tu-al

1. *Being such practically or in effect, although not in actual fact or name. –Webster's*

Dolby has seen a wide acceptance of multichannel reproduction based on its multichannel audio technologies using a basic 5-speaker audio system. It is clear, however, that no matter how well these systems work, there remain both practical and physical barriers to adopting full home theater systems universally. It has been shown that, for example, televisions merging Dolby Pro Logic with the JVC 3D Phonic virtual surround process can offer consumers a useful degree of surround sound involvement from a cost-effective, low-impact product design. The concept has also found its way to the multimedia PC both in the form of sound cards using the same Pro Logic 3D Phonic process, and by way of special loudspeakers pioneered by Altec Lansing that use a multiple drivers in a controlled directivity pattern fed from a Pro Logic decoder.

Dolby is introducing Virtual product category based on combining Pro Logic or Dolby Digital with a virtual speaker processor.

With the introduction of Dolby Digital 5.1-channel sound in DVD and other digital delivery mediums, consumers are becoming even more aware of the multichannel audio those systems can deliver. DVD appears to be focusing on the same market occupied by VHS (90% of U.S homes) as opposed to Laser Disc (2% of U.S. homes), while Dolby Surround decoders are estimated to exist in 10% of U.S. homes. It may be fair to conclude that many DVD owners will not necessarily be equipped with a home theater system. A virtual surround processor may therefore be the key to opening their ears to new possibilities in sound reproduction, and inviting consumers into taking those first steps to becoming more involved with true multichannel systems in the future.

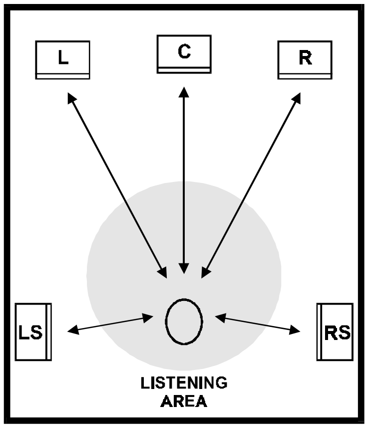
How it Works

Dolby Surround systems in home theater settings use five speakers to reproduce sound: three across the front (left, center and right) and two more around the sides (left-surround and right-surround), as shown in Fig. 1. In a virtual surround system, all the sounds are recreated by only the main left and right speakers as shown in Fig. 2.

Dolby
Signal Processing and Noise Reduction Systems

100 Potrero Avenue
San Francisco, California 94103-4813
www.dolby.com
Telephone 415-558-0200
Facsimile 415-863-1373

Wootton Bassett
Wiltshire SN4 8QJ
England
Telephone 01793-842100
Facsimile 01793-842101



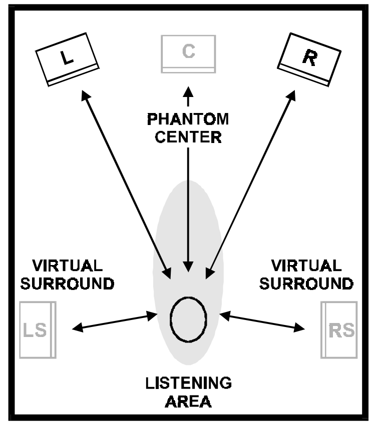


Fig. 1. Multispeaker home theater. Fig. 2. Virtual surround system.

The virtual surround effect is achieved through the use of special processing of the LS and RS signals in a "virtualizer" before reproduction by the main speakers. Several such psychoacoustic processing techniques have been developed over the years, and have only recently become more prevalent because of the wide availability of powerful digital audio processing circuits. Fig. 3 shows the typical signal flow used to create virtual surround from soundtracks decoded with either Pro Logic or Dolby Digital.

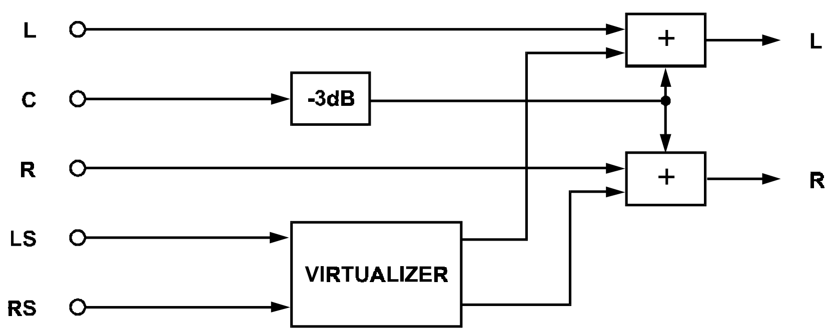


Fig. 3. Virtual surround downmixing process.

Virtual Products

A class of products has been defined that requires the use of virtual surround processing to enable users to experience the effect of Dolby Surround Pro Logic or Dolby Digital sound. In order to be approved for use with Dolby multichannel audio technologies and to display specific Dolby logos, the virtualizer must meet the necessary performance criteria when evaluated by Dolby Laboratories:

1. The virtual surround circuit must be able to work with mono surround signals from Dolby Pro Logic decoded soundtracks. If the product also includes multichannel Dolby Digital, it must work effectively with stereo surround signals.
2. The effect must not produce overt side effects such as "clicking", disturbing "null zones", severe timbral colorations or comb filter effects, or dynamic effects or noise modulation.
3. The result must be consistent with the original content as rendered by a properly calibrated surround system, both in terms of the spatialization effect and localization characteristics. A surround level control is recommended.
4. A stereo bypass mode must be offered that provides Dolby Surround compatible outputs.
5. The process shall not reduce the headroom of the audio system preceding the volume control. Other general audio quality criteria shall apply.

Virtual surround processing is intended for use in multimedia PCs, televisions, DVD players, VCRs, mini and portable stereos, and set-top cable and satellite receivers. It may also be included as an option in any Pro Logic or Dolby Digital multichannel decoder product.

The Players

Several companies offer virtual surround technology for this application. Some of the technologies that have been found to meet the requirements are listed in the following table. Dolby Laboratories also offers a virtualizer process under license.

Company	Technology
JVC	3D Phonic
Matsushita	Virtual Sonic
Harman	VMAx
Aureal	A3D
QSound	QSurround
Spatializer	N-2-2 DVS
SRS	TruSurround 6-2
EMI/CRL	Sensaura

The above technologies are not identical in performance. Just as the sound quality and imaging varies greatly between loudspeakers made by different manufacturers, so does the sound quality of different virtual surround speakers. Some of the key areas where performance will differ are:

1. Timbre. The system should give the impression of flat frequency response in the various channels, especially from the surround "speakers".
2. "Sweet spot" size. While all virtual surround processors tend to exhibit their best effects within a narrow width of seating area, usually suitable for no more than one person, the behavior of the system toward the boundary edges of the effect can differ significantly. Where some systems gently diminish in effectiveness, others cause abrupt changes that are

easily heard, and which may constrict the listening area. In some cases the surround channel image will collapse toward the front unless the listener is seated in the exact center position.

3. Processor requirements. As the algorithms improve in subjective quality, they often increase in their processing requirements. This may translate into increased costs for more powerful processors or for a separate processor circuit.

Logo issues

There are two product types that apply: products based on Pro Logic, and products based on Dolby Digital. The logo design uses "virtual text" to proclaim the feature.

For the Pro Logic products, we wish to prevent the implication that the product is equivalent to real Pro Logic, so the term Pro Logic shall not be used. The proposed logo is:

VIRTUAL DOLBY SURROUND

For Dolby Digital products, it is important to let the consumer know that Dolby Digital technology is included, as it will be a key distinguishing feature. The proposed logo is:

VIRTUAL DOLBY DIGITAL

Virtual Technologies

The following is an updated chart of currently approved Virtual Technologies. Please contact the individual companies for information on licensing their respective technologies.

Company	Technology	Virtual Dolby Surround (VDS)	Virtual Dolby Digital (VDD)
Aureal	A3D	✓	✓
Dolby	Dolby Laboratories' Virtualizer	✓	✓
EMI/CRL	Sensaura	✓	✓
Harman	VMAx	✓	✓
Intel	RSX	✓	✓
JVC	3D Phonic	✓	
Matsushita	Virtual Sonic	✓	✓
Micronas Intermettal	3D-Panorama	✓	✓
Onkyo	Theater-Dimensional	✓	✓
Philips	Incredible 3D Surround	✓	
QSound	QSurround	✓	✓
Sanyo	VASIL	✓	✓
Sony	Sony Virtual	✓	✓
Spatializer	N-2-2 DVS	✓	✓
SRS	TruSurround	✓	✓
Yamaha	Ymersion-VS	✓	✓

The following table is a list of implementations approved at the time of publication. For questions or updated information, please contact Dolby Laboratories.

Company	Technology	IC Name	Virtual Dolby Surround (VDS)	Virtual Dolby Digital (VDD)	Pro Logic Decoder	Dolby Digital Decoder
Matsushita	Virtual Sonic	MN19425B	✓	✓	✓	
Micronas	3D-Panorama	DPL3519A	✓		✓	
Micronas	3D-Panorama	MSP3451G	✓	✓		
NJRC	SRS TruSurround	NJM2180	✓	✓		
NJRC	Dolby's Virtualizer	NJU25006	✓		✓	
NJRC	Harman VMAx	NJU25007	✓		✓	
NJRC	SRS TruSurround	NJU25008	✓		✓	
NJRC	Spatializer N-2-2	NJU25009	✓		✓	
Philips	Incredible 3D Surround	SAA7712H	✓		✓	
QSound	QSurround	QS7777	✓	✓		
Sanyo	VASIL	LV1018	✓		✓	
Sony	Sony Virtual	CXD-2724Q	✓		✓	
Toshiba	SRS TruSurround	TC-9332	✓		✓	
Toshiba	SRS TruSurround	TC-9337	✓		✓	
Toshiba	SRS TruSurround	TC-9447	✓		✓	
Yamaha	Ymersion-VS	YSS902	✓	✓	✓	✓
Zoran	Aureal A3D	ZR36710/ ZR36215	✓		✓	
Zoran	Aureal A3D	ZR38600	✓	✓	✓	✓
Zoran	Dolby's Virtualizer	ZR38600	✓	✓	✓	✓
Zoran	Harman VMAx	ZR38600	✓	✓	✓	✓
Zoran	QSound QSurround	ZR38600	✓	✓	✓	✓
Zoran	SRS TruSurround	ZR38600	✓	✓	✓	✓
Zoran	Spatializer N-2-2	ZR38600	✓	✓	✓	✓

Contact:

Jim Arnold
Director of Public Relations
Dolby Laboratories Inc
415-558-0116
jja@dolby.com

Lynne Daily
Miller Shandwick Technologies
310-203-0550
ldaily@miller.shandwick.com

Personal surround sound arrives via *Dolby Headphone*

Dolby Laboratories and Lake DSP join forces to introduce multichannel audio system for headphones

San Francisco, December 7, 1998 - Dolby Laboratories Inc. and Lake DSP of Sydney, Australia announced today that the two companies have agreed to develop and market a new technology that allows consumers to enjoy multichannel surround sound on conventional stereo headphones.

Dubbed *"Dolby Headphone,"* the system will be featured primarily in multichannel audio/video products incorporating Dolby Digital or Dolby Pro Logic Surround decoding, but will also work with two-channel stereo programs.

With *Dolby Headphone* technology, consumers can use their favorite headphones to experience programs as if they were listening in a cinema or a fine home theater. Roger Dressler, Director of Technology Strategy for Dolby Laboratories, explains, "Traditional stereo playback on headphones, while pleasant, creates a shallow but extremely wide stereo effect that appears 'in your head.' *Dolby Headphone* allows the listener to hear the sound much more naturally, out in front, and out of head – and due to the full 5.1-channel surround capability of the system, more than one person can watch a movie from 'the best seat in the house.'"

The sophisticated *Dolby Headphone* technique is performed in a digital signal processor chip (DSP), which makes it easy to include in DVD Video players, upcoming DVD Audio players, set-top cable boxes, satellite receivers, digital televisions, VCRs, personal computers, video game consoles, A/V surround decoders, auto sound systems, multimedia speaker systems, and even headphone portables.

Brian Conolly, Managing Director and CEO of Lake DSP, stated, “We are delighted to be working with Dolby Laboratories on this project. We have long held the belief that the ability to listen to surround sound in headphones would have a wide appeal.”

“The technology Lake created leaps over the final barriers to 5.1-channel audio,” said Dolby’s John Kellogg, General Manager of Multichannel Audio and Music Production. “It brings the multichannel audio experience to an entirely new audience who, until now, were either unable or unwilling to install the necessary equipment to enjoy 5.1 audio in their homes. Now everyone can enjoy multichannel audio.”

Leonard Layton, Marketing Director at Lake DSP explains, “Our experience in headphone technology development coupled with Dolby’s dedication to multichannel audio delivery creates a winning combination. We are excited about the tremendous impact this will have on many consumers.“

Lake DSP, based in Sydney, Australia, has for eight years been developing hardware and software products and licensable technologies for the automotive, telecommunications, professional audio and acoustic research fields. More information on Lake DSP can be obtained at the company’s website : www.lakedsp.com.

Dolby Laboratories is the developer of audio signal processing systems used worldwide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting, and motion-picture sound. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo, and European headquarters in England.

S98/12347

Media Contacts:
Jim Arnold, *Dolby Laboratories*, 415-645-5116
jja@dolby.com

Rick Clements/Ameer Toor, *SIA Public Affairs*,
Tel.: (65) 541-5880, Fax: (65) 545-6083

Lynne Daily, *Miller/Shandwick Technologies*, 310-203-0550
ldaily@miller.shandwick.com

FOR IMMEDIATE RELEASE

Dolby Headphone brings cinema-quality surround sound to the aircraft cabin

Singapore, April 5, 1999—From 1 May 1999, the revolutionary *Dolby Headphone* technology will be available to Singapore Airlines (SIA) passengers on all classes of travel. This will be the first time the new technology will be available anywhere in the world.

Dolby Headphone technology provides cinema-quality surround sound to airline passengers using standard stereo headphones. Originally developed by Australian company Lake DSP, *Dolby Headphone* will soon be used in consumer electronics and personal computers. It will be SIA customers, however, who will experience this new technology for the first time, while enjoying SIA's in-flight entertainment system, KrisWorld.

Said SIA's Executive Vice President (Commercial), Mr. Michael Tan: "Singapore Airlines has been in close liaison with the reputable Dolby Laboratories and the original creator, Lake DSP for some time. We see this exciting innovation as breaking the sound barrier in in-flight entertainment. Our customers will have the privilege of being the first in the world to experience this amazing high quality surround sound. We are confident that this innovation will clearly make SIA's in-flight entertainment more enjoyable."

John Kellogg, General Manager of Multichannel Audio Production at Dolby Laboratories in Hollywood and Director of the *Dolby Headphone* In-Flight program said: "Dolby Laboratories is very excited about the application of *Dolby Headphone* to in-flight entertainment and is very happy to be working with a very progressive airline in Singapore Airlines."

Dolby Headphone was born out of pioneering work by Australian company Lake DSP over a period of eight years. Lake DSP made a series of breakthroughs in the simulation of acoustics using computers, and licensed the technology to Dolby Laboratories in October 1998.

"Lake DSP is very pleased to be working with Singapore Airlines to introduce the technology, and I am sure SIA passengers will appreciate this new advancement in in-flight sound," said Leonard Layton, Marketing Director at Lake DSP in Sydney, Australia. "It has been gratifying to work with an airline with such a strong commitment to in-flight entertainment and services."

SIA's KrisWorld system offers personal video screens, a selection of movies and in-seat telephones in all three cabins of the aircraft. KrisWorld is available on all MEGATOP 747s, CELESTAR A340s and JUBILEE B777s.

In September 1998, SIA set new standards in air travel by unveiling a new product worth about US$300 million, encompassing its First, Raffles and Economy Class. From revolutionary SkySuites in First Class, which transform into beds, to fine gourmet meals by some of the world's top chefs, and exquisite cabin interiors designed by Givenchy and James Park Associates, SIA has created an unprecedented level of comfort, service and luxury in commercial air travel.

Lake DSP, based in Sydney, Australia, has for eight years been developing hardware and software products and licensable technologies for the automotive, telecommunications, professional audio and acoustic research markets. More information on Lake DSP can be obtained at the company's website: www.lakedsp.com.

Dolby Laboratories is the developer of audio signal processing systems used worldwide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting, and motion-picture sound. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo, and European headquarters in England.

S99/12497

DOLBY DIGITAL: AUDIO CODING FOR DIGITAL TELEVISION AND STORAGE APPLICATIONS

STEVE VERNON

Dolby Laboratories Inc., San Francisco, California, USA
sv@dolby.com

Dolby Digital is a worldwide audio coding standard, with applications that include digital television and home theater. Its underlying perceptual coding engine provides high quality multichannel audio at low bitrates, without requiring excessive computational complexity. It also supports a number of system features that improve the overall listening experience, such as volume normalization and dynamics processing. This paper presents an overview of Dolby Digital, focusing on the design elements that differentiate it from competing systems and that have contributed to its success.

INTRODUCTION

What is Dolby Digital?

Dolby Digital is a high quality multichannel audio coding system, capable of delivering up to 5.1 channels at data rates roughly equivalent to half of one PCM audio channel [1], [2], [3]. The system was first introduced by Dolby Laboratories in 1992 with the Dolby SR*D digital cinema sound format. It has since been adopted by several consumer applications, including ATSC and DVB digital terrestrial television broadcasting, several cable and satellite television standards, worldwide DVD audio, compatible laserdisc audio, and emerging PC applications.

The coded audio format offers a wide degree of flexibility. It is designed to accommodate sample rates of 32, 44.1, or 48 kHz, and eight different channel configurations ranging from mono or stereo up to 5.1 channels (shown in Figure 1). The ".1" refers to an optional low frequency enhancement (LFE) channel, capable of carrying low-frequency material up to a maximum bandwidth of 120 Hz. Allowable data rates range from as high as 640 kbps for multichannel audio down to 64 kbps or lower for mono material.

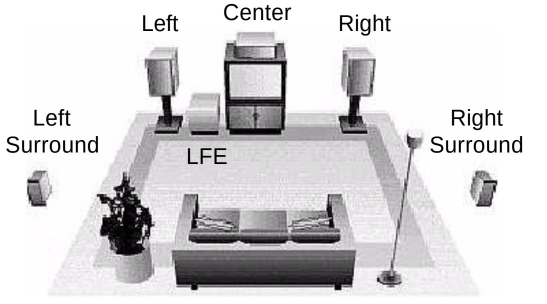


Figure 1: The standard 5.1 channel arrangement.

High quality at low data rates is accomplished by way of a sophisticated perceptual coding engine, which uses psychoacoustic principles and source coding techniques to reduce the necessary bitrate without impairing the perceived sound quality. Several unique design approaches have been taken to provide efficient audio coding, while keeping the implementation costs low enough for a wide range of applications.

In addition to being a powerful perceptual coder, Dolby Digital is also a fully-integrated audio delivery system, with built-in support for several important consumer features that will improve the listening experience in next-generation consumer equipment. These include volume normalization, dynamics processing, and compatibility with existing loudspeaker playback configurations. With these features, the consumer is given the opportunity to optimize the audio playback environment based on their preferences.

Keys to Success

Dolby Digital is not the only audio coding system; several competing systems are commercially available from a variety of developers [4], [5]. Neither is it the most efficient coder available; other coders have been developed that perform equally well at lower data rates [6]. However, it has proven to be a remarkably successful technology, selected for use in a number of different applications.

The keys to this system's success lie in the fact that it addresses all of the different attributes previously listed, including high quality coding at low data rates, a flexible format, integrated support for important consumer features, and moderate implementation complexity.

It is the position of this paper that in order to be successful in today's marketplace, an audio coding system must address all of these criteria evenly. As audio coding has evolved from a young to a mature field, standards organizations, application developers, and consumers have all come to expect a fully-integrated package from the audio subsystem. Dolby Digital was designed from the ground up to be a complete system, striking a careful balance between all of these different attributes.

In the remainder of this paper, I will present a more detailed introduction to the various elements of Dolby Digital. Particular attention will be paid to those elements that are unique to this system, or that substantially contribute to its overall success.

1. PERCEPTUAL CODING

At the heart of Dolby Digital is its perceptual coding engine - the component responsible for reducing the audio data rate without introducing audible impairments. This is accomplished via a combined approach of applying psychoacoustic principles (removing signal components that are inaudible to the human ear) and source coding techniques (using as few bits as possible to convey the information that remains).

There are many well-known approaches to perceptual coding, and dozens of techniques available for audio data rate reduction. Each of these techniques has a benefit (reduced data rate, improved sound quality) and a cost (implementation complexity, memory storage requirements). As this system was developed to meet the needs of cost-sensitive consumer applications, special care was taken in evaluating these cost/benefit tradeoffs. The result is a very powerful low bitrate coder with only moderate implementation cost.

A simplified overview of the Dolby Digital encoder is shown in Figure 2. Each of the processing stages is discussed in more detail in the following sections.

1.1 Time-to-Frequency Transform

Many of the most powerful psychoacoustic phenomenon that are exploited by perceptual coders occur in the frequency domain (e.g., frequency masking, hearing threshold limitations, etc.). As a result, most perceptual coders start out with a time-to-frequency transform to convert input PCM samples into blocks of frequency samples. The key design

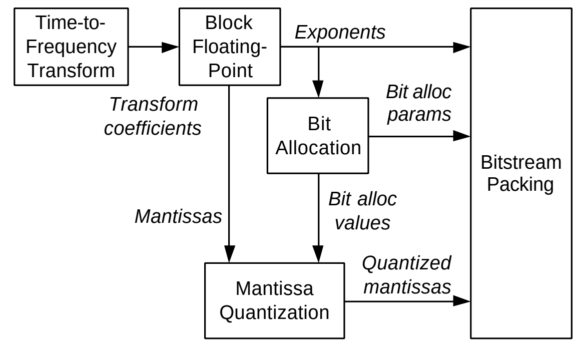


Figure 2: Encoder overview

parameters are the transform type, the transform window type, and the transform length.

Dolby Digital uses a time-domain alias cancellation (TDAC) transform approach built around a modified discrete cosine transform (MDCT), where adjacent transform blocks are overlapped by 50% [7]. This method has the useful property that it is critically sampled (i.e., the number of new frequency samples generated is the same as the number of new time samples required per block). In addition, the 50% overlap at the reconstruction stage in the decoder provides a form of automatic crossfade, ensuring smooth transitions from one block to the next.

The TDAC overlap process is illustrated in Figure 3. In this figure, the time samples corresponding to three consecutive transform blocks are shown. The samples for a given block are made up of a combination of windowed and time-aliased PCM samples. The aliased samples at the end of one block are constructed to be a negated version of the aliased samples at the start of the next block. As a result, when the blocks are overlapped and summed in the decoder, the aliased

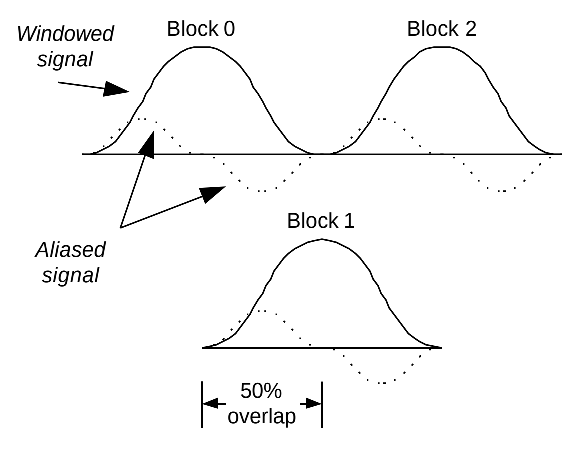


Figure 3: The TDAC transform overlap

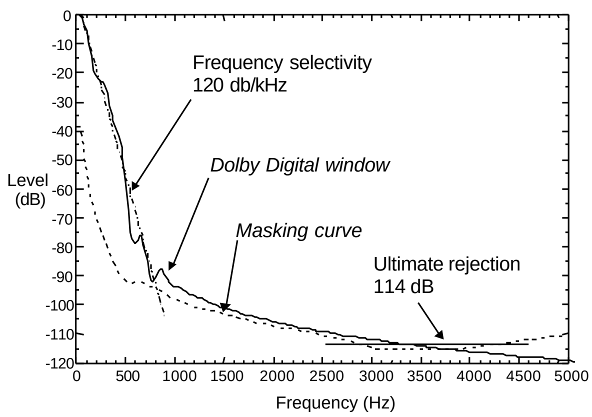


Figure 4: Transform window characteristics

signal components cancel and disappear, while the original PCM signal is properly reconstructed.

The transform window type used is a variant of the Kaiser-Bessel family of windows. This window type allows the designer to trade off frequency selectivity against ultimate rejection in order to tailor the window performance. For Dolby Digital, the specific window chosen has frequency selectivity such that the contribution from a tone into neighboring bins drops off at a rate of approximately 120 dB per kHz, while the ultimate rejection measures about 114 dB. These numbers compare favorably to those required by psychoacoustics, as the slope and ultimate rejection of typical frequency domain masking curves produce nearly the same results, as described in [3] and shown in Figure 4.

The final parameter, transform length, has a direct impact on the performance of the coding system. Shorter transforms offer better time resolution, improving the ability of the system to distinguish between distinct events in time. This is important for percussive signals, where the coder must faithfully reconstruct transients without time smearing.

On the other hand, longer transforms offer better frequency resolution, improving the ability of the system to distinguish between distinct signal frequency components. This is crucial to efficient perceptual coding, as processes such as frequency masking require the coder to properly distinguish between different frequency components of the signal.

Perceptual coders generally address this dilemma by offering two or more transform lengths and switching between longer and shorter transforms dynamically based on signal conditions. Dolby Digital supports this approach and offers two transform lengths: 512-point and 256-point.

Under most signal conditions, the 512-point long transform is selected. It produces 256 frequency sample bins (also known as transform coefficients) per block, with a linear frequency resolution of 93.75 Hz per bin. The time resolution for this transform is 10.7 msec.

When sudden transients occur, the 512-point transform is replaced by two 256-point transforms, as shown in Figure 5. Each of the two short transforms produces 128 transform coefficients, with a frequency resolution of 187.5 Hz per bin. Importantly, the time resolution of each short block is reduced to 5.3 msec, which ensures that transient signals fall within the range covered by time-domain masking.

In this system, transform block switching is performed in a rather unique way. While the long transforms have traditional tapered windows, the short transforms are computed using a window that is simply the long transform window split down the middle. The sharp edge at the middle of this window pair leads to several interesting circumstances, including poor frequency selectivity for the short transforms, and an absence of the smooth crossfade at the point of transition between short blocks.

These features would result in significant audible impairment if the short transforms were to be used for non-percussive audio signals; however, the use of the short transforms is reserved exclusively for transient signal conditions. In these cases, since the signal spectrum itself will be relatively flat across a wide bandwidth, high frequency selectivity is not essential. Furthermore, the only time the short transforms will be selected is when a loud transient signal is located in the time interval corresponding to the second short block (Block 1b in Figure 5). The guaranteed presence of a loud transient attack in this block serves as a time domain masker that will minimize the audibility of any distortion that may arise due to the absence of crossfade smoothing.

Finally, this approach to block switching enables a significant simplification to the overall design of the

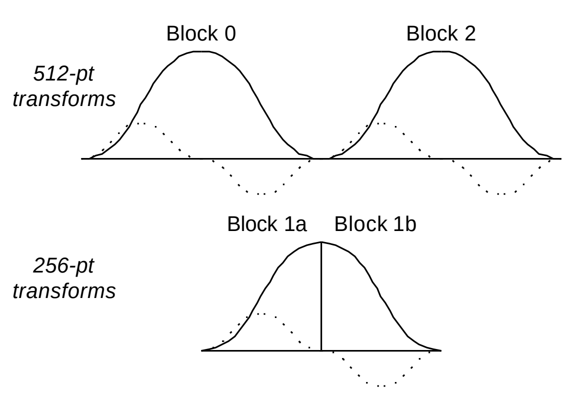


Figure 5: Long and short transforms

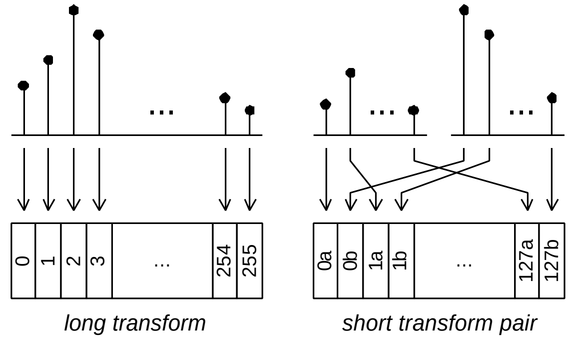


Figure 6: Transform coefficient ordering

perceptual coding engine. Specifically, the two groups of short block transform coefficients are interleaved into a single array, as shown in Figure 6. From this point on in the coder, the interleaved array is processed as if it were a group of long block transform coefficients. This reduces the complexity of the subsequent processing stages, as there is no need for special case handling of different transform lengths.

It is worth noting that during the design of this system, transforms of length longer than 512 points were considered. Longer transforms would have allowed for even better frequency resolution, and thus more efficient frequency-domain coding; however, they would have also resulted in a significant increase in the overall memory requirements. Furthermore, since the coder would still have required 256-point short transforms for transient signals, the block switching technique would have had to be more complex than the simple window splitting approach used here.

The use of 512-point transforms sets an upper bound on the overall system frequency resolution. In response to this, the next stage - coding the transform coefficients - has been designed to retain as much of this resolution as possible through the rest of the encoding process.

1.2 Block Floating-Point Encoding

Once the time-to-frequency transforms have been performed, the next step is to convert each transform coefficient into a normalized fractional mantissa and an integer-valued exponent. This process is illustrated by the diagram in Figure 7.

First, "ideal" exponents are computed by determining the magnitude of each transform coefficient rounded up to the nearest power-of-two, and thus represent a +6/-0 dB approximation of the spectral energy in each frequency bin. Exponents are constrained to lie in the range from 0 to 24, corresponding to transform coefficient magnitudes from 1.0 to 2^{-24} (transform coefficients with magnitudes less than 2^{-24} are defined to have an exponent of 24).

The mantissas are computed by dividing each transform coefficient by the equivalent magnitude of its exponent (with a 2's complement processor, this is accomplished by shifting the transform coefficient left by the value of the exponent). Given an ideal exponent, each mantissa value will be guaranteed to have a magnitude in the range from 0.5 to 1.0. Mantissas are signed, and may take on positive or negative values.

In Dolby Digital, exponents are extremely important to the rest of the coding algorithm. They are used to normalize the mantissa values in the encoder and denormalize them in the decoder. As discussed in the next section, they are also used in the bit allocation process to determine the amplitude precision required by each mantissa. As such, they must be sent to the decoder in the bitstream. Unfortunately, since ideal exponents are computed individually for each of 256 frequency bins (for each of up to 5.1 channels), and since each exponent requires 5 bits to represent the allowable range of values, the total exponent data rate quickly proves to be impractical.

One approach to reducing the exponent data rate, used by a number of perceptual coding systems, is to group sets of frequency bins together into bands. Within each band, the transform coefficient with the largest magnitude is used to compute a single shared exponent for the entire band. Commonly, 30 to 40 bands are designated per transform block, with boundaries chosen that correspond to logarithmically-spaced psychoacoustic critical bands.

While this technique succeeds at reducing the overall exponent data rate, there is an associated cost. Specifically, since the band exponent is based on the transform coefficient with the largest magnitude, there will generally be a number of other transform coefficients in the band for which the exponent is overly-conservative. If the band exponent is used to normalize these transform coefficients, one or more

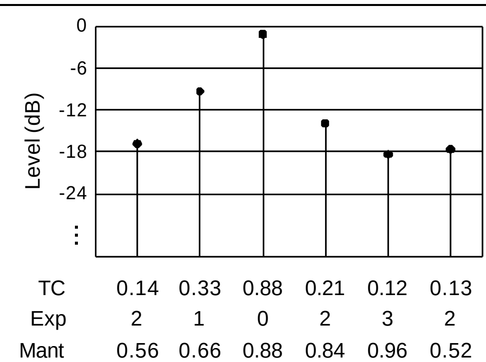


Figure 7: Exponent/mantissa conversion

of the leading bits in the coded mantissas will not contain useful information, but rather will only contain sign-extension bits. Thus, the savings gained by reducing the exponent data rate is offset by the extra data rate required to convey the suboptimally-normalized mantissas.

Rather than take this approach, Dolby Digital recognizes the benefit of fine-grain exponents (i.e. individual exponents for each transform coefficient), and instead looks for ways to reduce the data rate required to represent them. This is accomplished three ways: by coding them differentially, by sharing them across frequency, and by sharing them across time.

Differential exponent coding reduces the exponent data rate by transmitting not the actual exponent values, but rather the difference between adjacent exponents. In this system, these differentials are constrained to take on only a limited set of values, namely ±2, ±1, and 0. This imposes a slew limit constraint on how quickly exponents can change of 12 dB-per-bin. In practice, this limit is already effectively in place due to the frequency selectivity of the transform window (120 dB/kHz ≈ 12 dB/bin, see Figure 4).

In the example shown in Figure 7, the sequence of differential exponents would be as follows: {2, -1, -1, +2, +1, -1, ... }. Note that the first exponent is kept in absolute form, as there is no previous value.

Since these differential exponents only take on one of five distinct values, they could be coded using 3 bits instead of the 5 bits that would be required for absolute exponents. However, this system takes an additional step and uses arithmetic grouping to further reduce the necessary data rate.

Specifically, the differential exponents are coded in groups of three. Given three separate variables, each of which may take on one of five different values, the total number of combinations is $5^3 = 125$. This group may be coded with a 7-bit codeword. The net per-exponent data rate has thus been reduced to 2.33 bits per exponent. This method of exponent coding is referred to as the D15 exponent strategy, as it codes the exponents differentially, allocates each exponent to only one transform coefficient, and allows five separate difference levels.

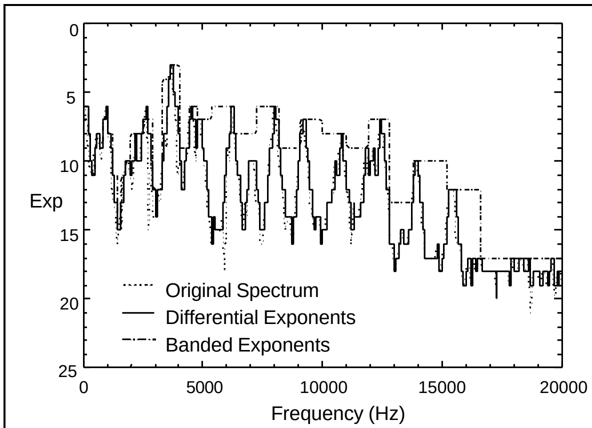


Figure 8: Differential vs. banded exponents

Two other exponent coding options are also supported. One starts by sharing exponents across frequency in pairs (i.e., each pair of transform coefficients share a common exponent); the other starts by sharing exponents in groups of four. In either case, the shared exponents are differentially coded, grouped into triplets, and coded with 7-bit codewords. The method that shares exponents in pairs is referred to as the D25 exponent strategy, while the method that shares in groups of four is called the D45 exponent strategy. With the D25 method, the data rate cost is 1.17 bits per exponent, while D45 requires only 0.58 bits per exponent.

In addition to differential coding and frequency sharing, the exponent data rate may be substantially further reduced by sharing exponents across time. The transform length was chosen to be 512 samples; however, many audio signals are stationary for time periods that exceed this duration. It is common for adjacent transform blocks to contain similar signal spectral content, which results in similar exponent values. This system exploits this fact by allowing the exponents for a given block to be reused by up to five subsequent blocks. In the best case, this reduces the overall exponent data rate by a factor of six, to as low as 0.10 bits per exponent for D45 coding.

In comparison with frequency banding, fine-grain exponent coding proves quite affordable. The equivalent exponent data rate for 30 bands of 5-bit exponents is 0.59 bits per bin, and differential coding is able to meet or exceed this level of efficiency when time sharing across blocks is used. In a coder such as this one, where transform lengths are short enough to support time sharing, differential coding is especially attractive.

Furthermore, differential coding is much better at tracking the peaks and valleys of complex signal spectra than banded exponents, as shown in Figure 8. This results in fewer leading sign bits in the coded mantissas, and a remarkably good spectral estimate from the exponents themselves. Of course, due to time and frequency sharing, it is still possible to have exponents that are not set to their "ideal" values; however, with fine-grain differential coding, the extent to which suboptimally-normalized mantissas adversely affect system performance is greatly reduced.

1.3 Bit Allocation

The bit allocation routine is the part of the coder responsible for determining the amplitude resolution needed by each frequency sample (i.e., how much precision to use to code each mantissa value). Frequency components that are clearly audible will require high precision, while those that are inaudible do not require any precision at all. The bit allocation routine uses a psychoacoustic model of the human hearing system to determine which frequency samples are audible, and how much precision is required to represent them well enough for human ears.

Most perceptual coders use a forward-adaptive bit allocation approach, in which the encoder examines the signal spectrum in fine detail, determines the bit allocation requirements for each mantissa, and sends the result as side-chain information in the coded bitstream. The term "forward-adaptive" refers to the fact that the bit allocation values are computed in the encoder and sent forward to the decoder, and are derived adaptively based on the signal spectrum.

Forward-adaptive systems have the benefit that the psychoacoustic model in the encoder may be revised without sacrificing compatibility with existing decoders. Since the bit allocation values are sent explicitly to the decoder, they may be computed using arbitrarily complex hearing models.

On the other hand, the cost of this approach can be substantial. The bit allocation values must be sent in the bitstream to the decoder, eating into the overall data rate. For this reason, bit allocation values in these systems tend to be shared across critical bands. Unfortunately, this has a further negative effect on the overall data rate, as all mantissas in a band will be given the same degree of precision, regardless of whether or not they are all equally audible.

Dolby Digital takes a different approach to bit allocation. Rather than computing the bit allocation values in the encoder based on the detailed signal spectrum, Dolby Digital computes these values based on the exponents (which, as was shown in the last section, provide a close approximation to the signal spectrum). Because the exponents are already sent to the decoder, there is no need to simultaneously send the bit allocation values, as they may be recomputed in the decoder. This technique is referred to as "backward-adaptive" bit allocation, indicating that the bit allocation values are computed in both encoder and decoder.

Since the exponents are sent in fine-grain form, the bit allocation data is also fine-grain. This retains much better frequency resolution than would be possible using a critical band approach. As a result, the bit allocation routine is able to assign bits only where they are needed, eliminating the mantissa data rate penalty suffered by the banded approach.

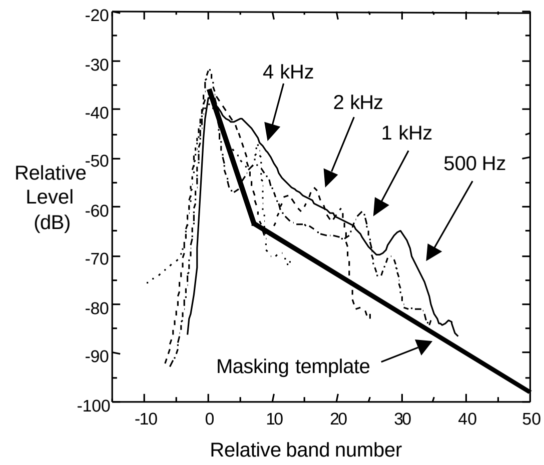


Figure 9: Simplified masking model

The most serious objection to backward-adaptive bit allocation is that it does not allow for improvements in the psychoacoustic model. Since the decoder must compute the bit allocation in the same way as the encoder, it is not generally possible to update the model in the encoder and still retain compatibility with existing decoder implementations. Dolby Digital addresses this concern in two ways: by using a parametric psychoacoustic model, and by allowing for a forward-adaptive correction factor if necessary.

The psychoacoustic model is based on a masking template designed to approximate measured data on frequency-domain masking, as shown in Figure 9. In this figure, the template is shown superimposed upon measured masking data for a variety of single-frequency tones. For simplicity, the template is constructed as a two-segment linear curve, designed to conservatively match the data. Only upward masking is modeled by this template, as downward masking is not a significant enough effect to justify its corresponding increase in complexity.

The characteristics that define this masking template are the slopes of the two linear segments, and the dB offsets of each segment relative to the level of the tonal masker. Rather than set these to fixed values, the encoder selects these parameters and sends them to the decoder in the bitstream. In this way, much of the flexibility of forward-adaptive bit allocation is preserved, as the encoder psychoacoustic model can be improved while retaining compatibility with existing decoders.

It is certainly possible that an improvement could be found to the psychoacoustic model that could not be incorporated by simply altering the masking template parameters. For this case, one further technique has been added to the bit allocation routine - the use of delta bit allocation. This technique allows the encoder to send forward-adaptive correction factors to the decoder that are used to modify the bit allocation results.

Since delta bit allocation does have an impact on the data rate, this technique is only used when the resulting gain in mantissa efficiency is greater than the data rate needed to send the correction factors. Note that it is not necessary to send correction factors for the entire spectrum - they may be sent for just one segment of the spectrum. As a result, the overhead associated with delta bit allocation can be quite small.

Backward-adaptive bit allocation has one major cost, and that is decoder complexity. Rather than simply extracting the bit allocation data from the coded bitstream, the decoder must rederive the bit allocation values. Even with a reasonably simple psychoacoustic model, this computation can be quite lengthy, especially if it must be done for each fine-grain frequency bin. As can be seen in Section 4, bit allocation calculations are one of the main contributors to decoder complexity. However, the overwhelming benefit in data rate savings and the preservation of frequency resolution make this a cost worth paying.

1.4 Mantissa Quantization

Once the bit allocation results have been computed, the next step is to quantize the mantissa values based on the bit allocation data. Each mantissa is given a variable amount of precision ranging from 0 bits to 16 bits, depending on the audibility of the associated frequency component.

At the low-precision end of the scale, odd-symmetric quantization is used to avoid unnecessary biases in mantissa coding. The term "odd-symmetric" quantization refers to the fact that the mantissa range is divided into an odd number of quantizer steps, constructed to be equal in width and symmetric about the value zero. Figure 10 shows an example of odd-symmetric quantization for a 3-level quantizer.

In this figure, the input mantissa range from -1 to +1 is divided into three equal-width regions. All input mantissa values that fall into the first of these regions, from -1 to -1/3, are quantized to the same level. This is reproduced in the decoder as the value -2/3, which is the midpoint of the input range. Odd-symmetric quantization has the useful properties that all quantizer steps have the same width, and the value zero may be coded without introducing any error.

A simpler alternative to odd-symmetric quantization is even-asymmetric quantization. This technique is shown in Figure 11, for the case of a 4-level quantizer. In this technique, all quantizer steps are not equal width - the first is 50% shorter than the rest, and the last is 50% longer. This is done to ensure that an input value of zero may be faithfully coded, a constraint that is especially important given that exponent sharing may result in suboptimally-normalized mantissas. This technique is more natural for 2's complement processing, as it simply requires rounding the input mantissa to the specified number of bits. The downside of this approach is that the asymmetry that exists in the first and last quantizer steps tends to introduce biases in the reconstructed mantissas.

Dolby Digital uses odd-symmetric quantization for all low-precision quantizer levels, including 3-level, 5-level, 7-level, 11-level, and 15-level. Beyond these cases, even-asymmetric quantization is used (from 32-

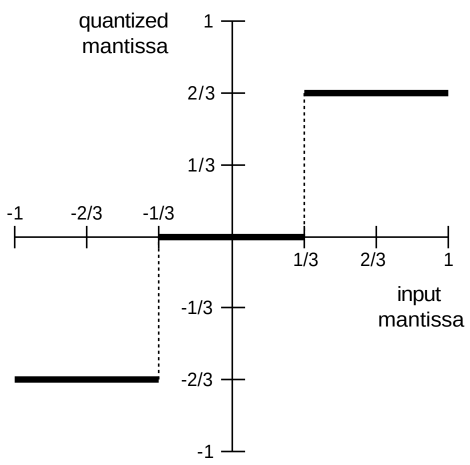


Figure 10: Odd-symmetric quantization

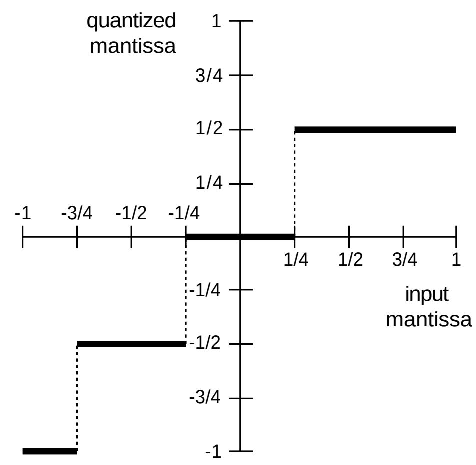


Figure 11: Even-asymmetric quantization

level through 65536-level quantizers). For these higher-precision cases the bias problem becomes much less significant, as the large number of quantizer levels makes the end cases substantially less likely to occur.

Using odd-symmetric quantization for low-precision quantizers has the side benefit that it enables more efficient mantissa coding. In the case of 3-level quantizers, each mantissa may take on one of three different quantized values. These could be coded using 2-bit codewords, but since not all states are used there would be an inherent inefficiency.

Instead, 3-level quantized mantissas are gathered together in groups of three and jointly coded. Three mantissas, each taking one of three values, makes 3^3 = 27 different combinations which can be coded with a 5-bit codeword. Using the same principles, groups of three 5-level quantized mantissas and groups of two 11-level quantized mantissas are coded with 7-bit words. 7-level and 15-level quantized mantissas are coded with 3-bit and 4-bit codewords, respectively.

1.5 Bitstream Packing

The last stage of encoding is bitstream construction, where all elements are packed together into the coded bitstream.

The Dolby Digital bitstream format is organized into frames, which are the smallest independently-decodable access unit. Each frame contains six transform blocks for each coded channel, and represents a time duration of 1536 audio samples (32 msec assuming a 48 kHz sample rate).

Within each frame, the coded data is organized in a carefully-configured structure, as shown in Figure 12. At the start of the frame lies a sync word, which allows decoders to recognize frame boundaries and begin decoding. Next comes the bitstream information (BSI) segment, which contains control data that indicates the number of coded channels, sample rate, data rate, and several other parameters.

Following the BSI segment are six audio blocks, each of which represents a time duration of 256 samples. The audio blocks contain information that describe the coded transform data, including the block switch flags (long transform or short transform pair), dither flags (see Section 1.6), coupling data (see Section 1.7), rematrixing data (see Section 1.8), exponent strategy, differentially-coded exponents, bit allocation parameters, and quantized mantissas. Each block contains values for all of the coded channels, as many as 5.1 channels in all.

For a given data rate, the overall frame is assigned a constant number of bits; however, the audio block boundaries within the frame are not fixed. This allows the encoder to globally allocate bitstream resources from a common bit pool. As a result, resending exponents in a particular block does not disadvantage the mantissa bitrate in that block. Also, the mantissa bitrate itself can be dynamically allocated based on relative signal requirements of each block.

Interestingly, the way in which the different elements are ordered in the bitstream has a significant impact on the implementation complexity of the decoder. Because the audio block data is sequenced according

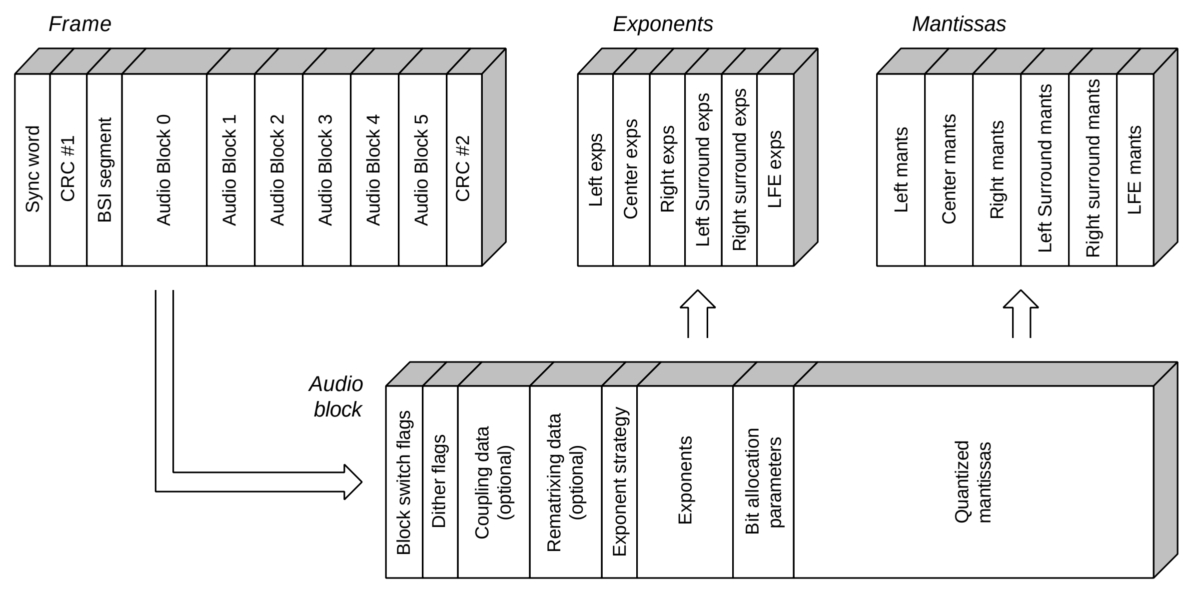


Figure 12: The Dolby Digital bitstream structure

to time, it is possible for the decoder to reconstruct and start playing out the first block's PCM samples before starting to decode the second block.

This results in reduced decoder memory requirements, as the decoder does not need to store exponent, bit allocation, or mantissa data for more than one block at a time. In fact, it is even possible for the decoder to avoid storing more than one channel at a time [8]. If the audio frame had been constructed differently, this complexity reduction may not have been possible.

Finally, each frame contains two separate CRC words, which allow decoders to detect the occurrence of bitstream errors. In the case of errors, frame decoding is replaced by concealment, which is accomplished by repeating previous known good audio for short error bursts, or muting for longer outages.

It is important that the decoder not attempt to decode a frame in the presence of errors. If the unpacked exponents or bit allocation parameters are not the same as were used in the encoder, the bit allocation results in the decoder will be invalid, and thus the mantissas will be unpacked incorrectly. This will have a ripple effect on subsequent audio blocks, as the audio block boundaries will almost certainly not be correctly recovered.

1.6 Decoder Processing

Within the decoder, the signal reconstruction process follows the inverse of the encoding process, as shown in Figure 13. The block switch flags, dither flags, coupling data, rematrixing data, and exponent strategy values are first unpacked from the bitstream. If new exponents are indicated in a given block they are unpacked, otherwise the exponents from the previous block are reused.

The bit allocation routine is then performed to determine the precision for each mantissa. Using the bit allocation data, the quantized mantissas are unpacked from the bitstream and denormalized by their corresponding exponents. The resulting reconstructed transform coefficients are then passed through an inverse TDAC transform, and overlap-added with the samples from the previous block. Finally, the decoded PCM samples are played out of the decoder.

Aside from these core tasks, there are a few interesting functions that the decoder performs that merit special mention. If the desired output format is different than the number of coded channels, the decoder is capable of downmixing the coded audio to an arbitrary number of output channels. The decoder also performs the important system functions of volume normalization and dynamic range control.

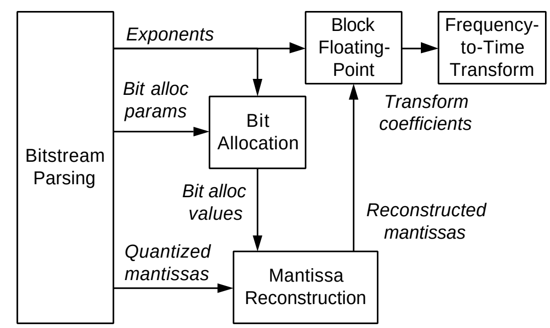


Figure 13: Decoder overview

Each of these topics are discussed at greater length in Section 3.

The decoder also performs one algorithmic function that has a substantial effect on the overall sound quality of the coder. Specifically, in cases where mantissa values were quantized with 0 bits, the decoder may substitute a dither value for the missing mantissa. The dither values are random numbers within the range from -0.707 to +0.707. This process is illustrated in Figure 14.

On first glance decoder dither may seem unnecessary. Mantissas that have been allocated no bits indicate frequency components that were deemed inaudible, and thus it should be possible to reproduce them using any arbitrary value (e.g., 0). However, this logic can fail for two reasons.

First of all, if the coder is operated at uncomfortably low data rates, there may be cases where audible frequency components are quantized with 0 bits in order to meet the data rate constraint. By replacing

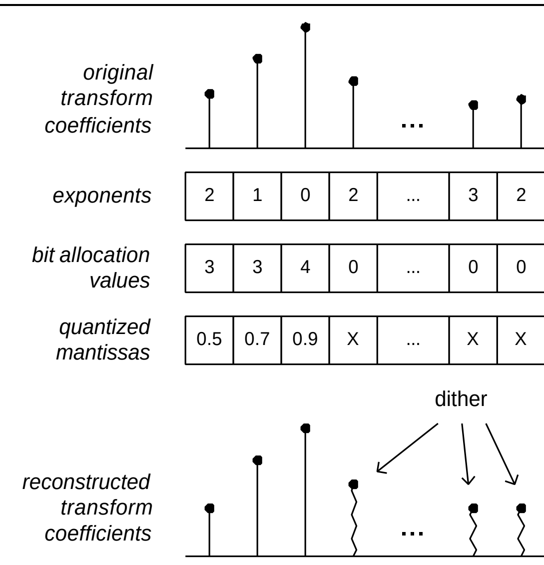


Figure 14: Decoder dither processing

these mantissas with dither values, the decoder is at least able to regenerate these frequency components within a fair range of their actual value. Since the dither-substituted mantissas are denormalized by the fine-grain exponents, which have an amplitude uncertainty of +6/-0 dB, the reconstructed transform coefficients should be in the same range as their original values. While the result may not be true high fidelity, it will be better than removing the frequency component altogether.

Secondly, it is sometimes not the absence of a particular transform coefficient that gets noticed, but rather the modulation effect that occurs when the transform coefficient is present in one block, then absent in the next, and then present again. This kind of modulation can be very audible, even if the psychoacoustic model predicts that the underlying frequency component is not. By regenerating 0-bit quantized mantissas with dither rather than silence, this modulation artifact is much less likely.

There are some cases where the use of dither for 0-bit quantized mantissas is not appropriate. One interesting example is in the case of a short transform pair. In this case, the assumption is that there is a loud transient signal in the second short block, while the signal is comparatively quiet in the first short block.

Under these conditions, the transform coefficient array, which is constructed by interleaving the transform results from the two short blocks, will tend to contain alternating small and large values. The exponents will similarly follow this pattern. Since differential exponent coding only allows for up to 12 dB of change from one exponent to the next, the exponents for the first short block will be substantially modified by the exponents for the second short block. As a result, the mantissas for the first short block are virtually guaranteed to be suboptimally normalized.

The net result of this is that if dither is used to replace a 0-bit quantized mantissa belonging to the first short block, it will not be scaled to the level of the original transform coefficient, but will rather be reproduced at a much higher level than desired. This will introduce a noise burst in the first short block, which is exactly the effect that block switching is trying to avoid!

To address cases where dither is inappropriate, it is left to the encoder to decide whether or not the decoder should apply dither to 0-bit quantized mantissas. This is done using the dither flags, which are bitstream parameters sent for each audio block.

1.7 Channel Coupling

One topic that deserves mention in any discussion of Dolby Digital perceptual coding is channel coupling. This is an optional technique that is available when two or more channels are being coded. Although channel coupling does have the potential to introduce audible artifacts, it also has the power to dramatically reduce the overall data rate, and thus can be a very successful tool if wisely used.

The psychoacoustic principle behind channel coupling is that the human ear is not sensitive to signal phase at frequencies above 3 kHz, but rather is only sensitive to signal amplitude [9]. Given this assumption, it should

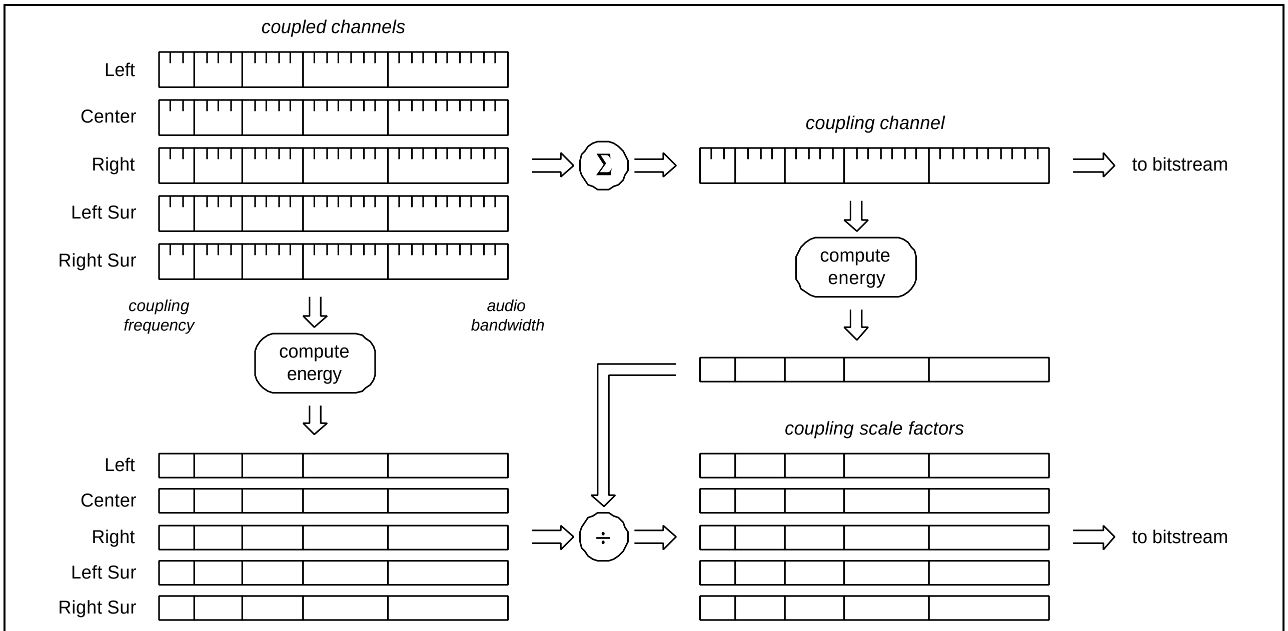


Figure 15: Channel coupling

not be necessary to code high-frequency mantissa values accurately, so long as their magnitude is correct. Furthermore, given the critical band frequency resolution of human hearing, it follows that the high-frequency magnitudes need not be correct for individual frequency bins, so long as the overall signal energy within each critical band is correct.

This is a very powerful argument, as it implies that the coder can introduce large amounts of error in individual high-frequency bins. Given that 85% of the transform coefficients represent frequency bins above 3 kHz, this presents a tremendous data rate reduction opportunity.

Channel coupling may be applied to some or all (but at least two) of the coded channels, following the process shown in Figure 15. Channels that participate in coupling are referred to as "coupled channels". The channel coupling process takes all high-frequency transform coefficients for all coupled channels and adds them together to form a single mono sum, called the "coupling channel".

Next, the individual coupled channels and the coupling channel are measured to determine the signal energy in each critical band. From these measurements, scale factors are computed by dividing the energy measurement for each band in each coupled channel by the energy measurement for corresponding band in the coupling channel. These scale factors represent the amount that the coupling channel band will be scaled by in the decoder in order to recreate a band with the same energy as the original signal.

Finally, the coupling channel and the scale factors are quantized and packed into the bitstream frame. The high-frequency exponents and mantissas for the coupled channels are discarded.

The frequency at which coupling is allowed to begin is a variable parameter controlled by the encoder, and can range from 3 kHz up to almost 20 kHz. Assuming a modest coupling frequency of 10 kHz, this technique single-handedly discards nearly half of all exponent and mantissa data, resulting in a dramatically lower overall data rate. For each coupled channel, half of the exponents and mantissas are eliminated, replaced simply by a handful of coupling scale factors and the exponents and mantissas that correspond to the coupling channel itself.

Channel coupling is a very aggressive coding technique, and not surprisingly it does have the potential to introduce artifacts. On the positive side, the audibility of these artifacts diminishes greatly as the coupling frequency is raised, and thus coupling can be a useful technique when constrained to higher frequencies (e.g., above 10 kHz). In certain circumstances, it can be used successfully at lower frequencies as well, however this requires careful analysis by the encoder based on dynamic signal characteristics.

1.8 Rematrixing

Finally, it is worth mentioning the role of rematrixing within Dolby Digital. Rematrixing is an optional process that is only available if the audio being coded is a two-channel program. It provides a means for the encoder to perform sum/difference coding in place of individual channel coding, as shown in Figure 16. This technique is especially relevant in the event that the two-channel program contains multichannel matrix-encoded material.

When the two channels of a stereo program are coded independently, they will invariably be coded with different exponents, different bit allocation data, and different mantissa quantization between the two channels. As a result, the loss of precision applied to each channel, which can be thought of as an added noise source, will be different. Generally speaking, this added noise will be inaudible, as it is masked by the coded signal content.

With the rise in popularity of matrix-encoding systems (e.g., Dolby Surround), it has become common to encounter two-channel program material that actually contains a matrixed multichannel soundfield, rather than simple stereo. Unfortunately, if independent noise is added to a matrix-encoded signal, it can damage the matrix encoding beyond repair, even if the noise is not audible when the matrixed signal is listened to as stereo. To see how this is possible, a few words need to be said about matrix encoding.

Matrix encoders such as Dolby Surround reduce four channels down to two by routing the left and right input channels to their corresponding output channels, applying the center input channel equally and in-phase in the two output channels, and applying a filtered version of the surround input channel equally and out-of-phase (see Figure 17). This process is reversed by a matrix decoder, which reconstructs the four channels using a complementary matrix, in which the left and right output channels are derived directly from the

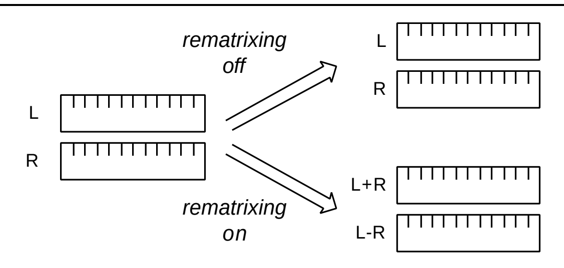


Figure 16: Rematrixing

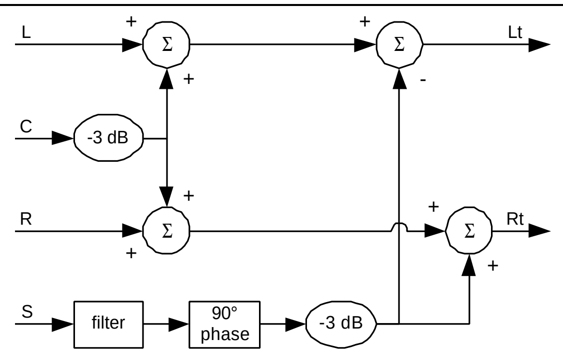


Figure 17: Dolby Surround encoding

stereo input, and the center and surround output channels are computed from the sum and the difference of the two input channels.

Suppose a center-only signal (such as a mono dialogue program) is matrix encoded, resulting in a two-channel mono signal. Suppose this program is then perceptually coded while treating the two channels as independent. Despite the strong correlation between the two channels, there will probably be slight differences between them. These may be due to imbalances in the A/D converters used for the two channels, gain imbalance in the system, or rounding differences within the perceptual encoder itself. Whatever their source, these slight differences will cause the noise added by the perceptual coder to be different for the two channels.

Now suppose this signal is then decoded by a perceptual coder, and the reconstructed two-channel signal is passed to a matrix decoder. The surround output will be derived as the difference between the left and right input channels. This is the difference of the original signals (which is essentially zero) and the difference of the noise sources (which is not zero, and could be quite significant). As a result, the surround channel output will exhibit significant audible noise. Furthermore, the level of this noise will be proportional to the center channel signal, leading to a noise modulation artifact that will be quite objectionable.

In order to address this problem, rematrixing recognizes cases where the two channels being coded are highly correlated (either in-phase or out-of-phase). In these cases, the encoder performs sum/difference coding rather than individual channel coding. In the decoder, the sum/difference operation is reversed.

The net result of this approach is that when the input signal is highly correlated, the noise that is added to the two channels will also be correlated. Thus, the center and surround output channels of the matrix decoder will both sound as expected, without the potential for annoying noise modulation.

Within the rematrixing routine, the sum/difference decision is made independently for each of up to four separate frequency bands. This ensures that correlation in one part of the spectrum can be recognized and rematrixed without requiring the entire wideband signal to be rematrixed. The only overhead required by this process is the bits that indicate to the decoder whether or not the signal has been sum/difference coded in each of the rematrixing bands.

2. ENCODER SYSTEMS FEATURES

Built around Dolby Digital's core perceptual coding engine are a number of important and useful system features. In the encoder, the list of available features includes input signal conditioning, reversible dynamic range processing, and carriage of audio metadata. These features are not required for high-quality audio coding; however, by addressing the system-level needs of audio coding applications, they provide significant additional value.

2.1 Input Signal Conditioning

In order to improve the overall audio sound quality, Dolby Digital encoders offer a selection of time-domain filters that may be used to condition the input PCM samples prior to encoding. These filters are optional, and their application is controlled entirely by the mastering engineer.

The first available filter is a digitally-implemented 50/15 µS deemphasis filter, which may be used to remove preemphasis from the digital audio source material. As the use of digital audio preemphasis is becoming less common, this filter is infrequently used in practice. However, if the input source material happens to be preemphasized, it must be deemphasized prior to encoding; otherwise the perceptual coder's frequency masking estimates will not be correct. On most implementations, this filter is automatically engaged based on the preemphasis flag in the digital audio channel status field, but it may also be manually enabled if desired.

The next set of filters offer signal conditioning for the surround channel(s), including a 90 degree phase shift filter and a 3 dB attenuation function. The 90 degree phase shift filter makes it possible to create a 5.1 channel bitstream that may be downmixed to a form compatible with a Dolby Pro Logic decoder. Since the phase shift operation is computationally intensive, it is offered in the encoder as a preprocessing stage rather than performed in the decoder. The 3 dB attenuation function addresses an alignment issue in which some

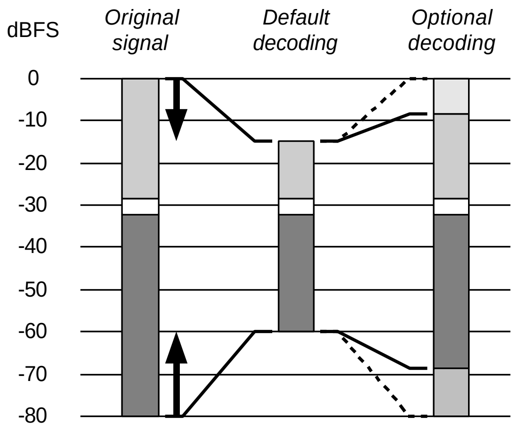


Figure 18: Reversible dynamic range processing

cinema post-production facilities master 5.1 channel material with the surround channels 3 dB higher than the television/home video standard.

Next, a trio of filters are offered to help band-limit the signal to the useful audio range. A DC-blocking highpass filter helps to remove DC offset biases which might otherwise lead to thumps when switching between programs. A bandwidth lowpass filter provides a clean way to roll off the signal content just below the bandwidth limit of the audio coder. Finally, an LFE lowpass filter may be used to lowpass filter the LFE channel to its maximum bandwidth of 120 Hz.

After all of these filters have been applied, the resulting PCM signal is routed to the input of the perceptual coder, as described in the Section 1.

2.2 Dynamic Range Processing

One of the most important system features surrounding audio broadcast and distribution is the use of dynamic range compression. With digital audio formats providing upwards of 100 dB of useable dynamic range, it is now very easy to create audio signal content for which the dynamic range exceeds the capabilities and/or the desires of the consumer.

For example, someone listening to an audio program in a car or via a portable device would have to strain to hear quiet passages that fall 30 dB below the average reference level. Meanwhile, someone watching a movie late at night in a small apartment would probably not like the loud passages to reproduce 20 dB higher than the reference level. Of course, there are situations where full dynamic range reproduction is desirable, such as in a high-end home theater room.

In the past, television broadcasters and home video producers have applied dynamic range compression to the source material to create a one-size-fits-all soundtrack designed for the average consumer. While this may be acceptable for some, it is far from the ideal solution.

Dolby Digital turns this entire problem around by offering a new approach to dynamic range compression. In this new design, the amount of desired dynamic range compression boost or cut is computed in the encoder, but not actually applied to the audio. Instead, these boost and cut values are sent to the decoder in the bitstream. At the decoder, the compression values may be applied fully to the audio, applied partially, or ignored, based on the preference of the consumer and the capabilities of the decoder. This behavior is illustrated in Figure 18.

Dolby Digital encoders offer a variety of pre-configured dynamic range compression profiles tailored for common soundtrack types such as music, speech, or film material. The mastering engineer is able to select a compression profile that meets their artistic needs and matches their content, thus retaining a strong degree of control over how the audio will sound to most consumers. At the same time, at the other end of the chain, consumers who want to hear the full dynamics of the original audio material have the freedom to do so.

2.3 Audio Metadata

The term "metadata" loosely refers to any information about the audio material that is not part of the audio signal itself. As such, the dynamic range boost and cut values described in the previous section are a form of audio metadata. In addition to these values, Dolby Digital encoders are capable of inserting a number of other important metadata parameters into the coded bitstream.

Perhaps the most important metadata parameter aside from the dynamic range compression values is the dialog normalization value. This parameter establishes an absolute reference for the loudness of the dialog component of the audio program. As will be seen in Section 3.2, it is used to enable volume normalization in decoders.

Another important class of metadata parameters are the downmix level scale factors, which allow the mastering engineer to control the relative level of the different coded channels when they are downmixed in a decoder. For example, these values may be used to reduce the level of the surround channels in the downmix, so that the front channels and dialog are more clearly audible.

Several metadata parameters provide a way to describe the audio material, such as the type of program (complete main, hearing impaired, emergency, etc.), whether or not it has been Dolby Surround encoded, and whether or not it contains copyrighted material. Finally, two parameters are available that describe the type of room used for mastering and the reference mixing level, allowing the consumer to reproduce the audio as it was heard by the mastering engineer.

Most of these parameters are located in the BSI portion of the bitstream, which makes them easy to retrieve without actually decoding the entire bitstream. Their impact on the overall coded audio data rate is very small, and well worth the added utility that they provide.

3. DECODER SYSTEMS FEATURES

In the decoder, three important system features are available that help solve a number of potential configuration problems, while remaining flexible enough to meet the needs of different types of listeners. These features include dynamic range control, volume normalization, and downmixing.

3.1 Dynamic Range Control

As described in Section 2.2, Dolby Digital supports reversible, configurable dynamic range control. Rather than directly modifying the audio content, the encoder generates dynamic range compression boost and cut values which are sent to the decoder in the bitstream.

In the decoder, there are two distinct compression modes that may be used to configure the dynamics processing. Only one of these modes may be selected at a time, and which one is used depends on the nature of the decoder application. The first, termed "line mode", is used by the decoder to provide a high-fidelity line-level output, as in the case of a home A/V receiver or high-end DVD player. The second, termed "RF mode", is used when the decoder is intended to drive an RF remodulator that connects to the antenna input of a television set, as in the case of a set-top box.

The dynamics processing requirements of these two modes are quite different and even contradictory. In line mode, the objective is to provide a high quality baseband signal with reasonable headroom and scaleable dynamic range control. In RF mode, the goal is to match the level and dynamics that are typical in television broadcasting practice. In order to address both of these cases, the bitstream provides two different sets of boost and cut values - one for line mode and one for RF mode.

In line mode, the decoder makes use of the first set of boost and cut values. These values may then be scaled to reduce the degree of dynamic range reduction. If desired, these values may be scaled down to zero, which recovers the original dynamics of the source audio material.

In RF mode, since the need to match conventional television broadcast levels is paramount, compression scaling is not allowed. Instead, decoders must use the RF mode boost and cut values without modification. This guarantees that the antenna demodulation circuitry within the television will be driven with signals that are at the appropriate levels.

3.2 Volume Normalization

In addition to dynamics processing, Dolby Digital decoders also implement volume normalization. With this feature, program material mastered in different facilities and with different reference levels can be made to reproduce at a common playback level. This reduces the need for the consumer to reach for the volume control every time they change the channel on their TV, or when they switch between a DVD movie and a broadcast television program.

The most obvious way to implement volume normalization is to introduce a common industry-wide absolute reference level. Unfortunately, there is very little consensus within the audio industry as to what the best mastering reference level should be. While some facilities use -20 dBFS, others use -12 dBFS or even higher. To try to bring uniformity to the industry would require changing the habits and preferences of audio engineers around the world - an endeavor doomed to failure.

Another challenge with volume normalization is defining exactly what part of the signal should be normalized. With music CDs, peak level normalization seems to be the norm, as mastering engineers generally adjust levels such that the signal peaks reach (or exceed!) 0 dBFS many times per disc. This is clearly not appropriate for volume normalization, as equal peak levels do not guarantee that two programs will have equal loudnesses.

Dolby Digital's approach is to use the dialog component of a program as the basis for establishing uniform levels. Two programs are said to be playing at the same level if their dialog components have a common loudness. In this context, dialog loudness is measured using a long-term-averaging frequency-weighted meter.

In order to implement volume normalization, the bitstream carries a metadata parameter called "dialog normalization" which indicates the dialog level for the program. This parameter is specified as a value between -1 and -31 dBFS in 1 dB steps. By indicating the dialog level in the bitstream, there is no need to change mastering preferences or standard reference levels within facilities.

In a decoder operating in line mode, program material with a dialog level higher than -31 is normalized downward, such that all content is reproduced with a uniform output level of -31 dBFS. If the decoder is operating in RF mode, the reference output level is raised to -20 dBFS, which sacrifices headroom dynamics but is necessary to match standard television program material.

By providing a means for volume normalization, Dolby Digital lays a path for a much improved listening experience for consumers. As broadcasters and content providers begin to use this important feature, loudness differences between different programs will become the exception rather than the rule.

3.3 Downmixing

No matter how compelling 5.1 channel discrete audio may be, there will always be consumers that do not have 5.1 channel reproduction equipment, and instead choose to listen in stereo or mono. As a result, in order to succeed, any multichannel audio system must have the ability to reproduce the full program content over fewer than 5.1 channels. Dolby Digital decoders accomplish this using a technique called "downmixing". With downmixing, the number of reproduced output channels is completely unrelated to the number of coded channels in the input bitstream. Instead, the number of output channels is a function of the number of loudspeakers connected to the decoder.

In order to implement downmixing, the decoder must unpack and decode all of the coded channels. As each channel is unpacked, it is routed and mixed into one or more of the output channels, using downmix scale factors that control its relative level in the output channels. These downmix scale factors have default values, but may be modified using the audio metadata parameters described in Section 2.3.

Dolby Digital supports eight different output configurations, including two different two-channel options. The first, termed the "Lo/Ro" downmix, provides a two-channel output that may be listened to directly as stereo or alternately summed together to form a mono output. The other two-channel option, termed the "Lt/Rt" downmix, is used to create a Dolby Surround compatible signal that may be fed to a Dolby Pro Logic decoder. This is a useful option for product categories such as portable DVD players that only offer a two-channel interface, but might be connected to a home theater system for multichannel listening. The equations used for the Lo/Ro and Lt/Rt downmixes are shown in Figure 19. Note that the LFE channel is not included in these equations. This points out the importance of treating the LFE channel as an enhancement channel, and not a carrier for essential program material.

Lo/Ro Downmix

$$Lo = L + (cmix * C) + (smix * Ls)$$
$$Ro = R + (cmix * C) + (smix * Rs)$$

Lt/Rt Downmix

$$Lt = L + (0.707 * C) - (0.707 * (Ls + Rs))$$
$$Rt = R + (0.707 * C) + (0.707 * (Ls+ Rs))$$

Figure 19: Lo/Ro and Lt/Rt downmixing equations

One consideration that must be addressed when downmixing is the possibility for output channel overload. If too many of the input channels are at or near full scale, then the downmixed output will almost certainly run into clipping. One way around this problem is to introduce a fixed attenuation level that guarantees that clipping cannot occur. Unfortunately, this approach sacrifices signal fidelity by lowering the program reference level relative to the noise floor of the D/A converters; furthermore, it undermines the goal of volume normalization. Given that clipping requires several channels to be near full scale simultaneously - a reasonably rare condition - this tradeoff is not particularly attractive.

The approach taken by this system is to downmix channels together without introducing a fixed attenuation offset. Instead, clip protection is provided by the dynamic range compression values. When the signal conditions are such that clipping might occur in a decoder downmix, the encoder is given the responsibility to ensure that the dynamic range compression boost and cut values provide enough attenuation to avoid clipping. Decoders that are downmixing (and thus are potentially capable of clipping) are restricted from scaling the dynamic range boost and cut values, and must apply them without modification. Decoders that are not downmixing may scale the boost and cut values, allowing the dynamic range compression to be disabled.

As a result of this approach, the dynamic range compression boost and cut values have a dual purpose - they provide artistic compression based on the profile selected by the mastering engineer, and they also provide clip-protection limiting based on the potential for multiple high-level channels to clip when downmixing. Using the compression boost and cut values to protect against clipping has the positive effect that peak limiting will only occur when the input channels are near full scale. Most of the time, when this condition is not true, limiting will be avoided and noise floor performance will not suffer.

4. COMPLEXITY SUMMARY

Throughout the design of Dolby Digital, algorithm and system features were balanced carefully against their implementation complexity. This includes the processor computational load measured in millions of instructions per second (MIPS), as well as memory storage requirements and processing latency.

It is interesting to note that the costs associated with these complexity parameters varies depending on the implementation platform. For example, a software implementation running on a personal computer might be less sensitive to memory storage needs, whereas MIPS usage could be a critical parameter to support proper operation in a multitasking environment. On the other hand, a dedicated hardware integrated circuit implementation might be willing to use more MIPS in order to reduce the memory requirements, which have a direct impact on overall IC area.

In the sections that follow, the numbers shown are taken from real-time hardware implementations, and reflect a design approach that favors reducing memory requirements rather than the processor load. In some cases, implementation tradeoffs are available that would bias these results in the other direction.

4.1 Encoder Complexity

The Dolby Digital encoder is by necessity a rather complex system. It is the encoder's responsibility to determine the optimal settings for parameters that are sent to the decoder, such as the bit allocation model. Furthermore, the encoder must decide which type of transform to use in each block, and whether to resend or reuse exponents. The encoder must determine whether or not it is appropriate to use coupling and/or rematrixing. Finally, the encoder must support selection of any or all of the signal conditioning filters, and compute the proper dynamic range compression boost and cut values.

Because the bitstream frame supports arbitrary block boundaries and global bit pool allocation, the encoder must generally consider all six blocks of the frame simultaneously, in order to compute the best distribution of bits. This has a direct impact on memory, as six blocks worth of transform coefficients, exponents, etc. must be held in memory simultaneously. It also has a direct impact on latency, since the encoder cannot begin to transmit a given frame until it has processed all of the information contained within it.

Processor load (MIPS)	5.1-chan	2-chan
Input signal conditioning	33.4	4.6
Dynamic range processing	2.2	1.2
Forward transform	23.4	8.8
Coupling/rematrixing	4.3	4.6
Block floating-point encoding	8.9	3.8
Bit allocation	42.0	18.1
Mantissa quantization	6.4	2.5
Bitstream packing	6.7	3.0
Total	*127.3*	*46.6*

Read/write memory (words)	5.1-chan	2-chan
Input PCM buffer	18432	6144
Output bitstream buffer	1792	768
Internal buffers	66674	25578
Total	*84.9k*	*31.7k*

Read-only memory (words)	5.1-chan	2-chan
Internal tables	4593	3121
Program code	17255	16700
Total	*21.3k*	*19.4k*

Latency (msec)	5.1-chan	2-chan
Encoding latency	178.6	69.3

Figure 20: Encoder complexity summary

Figure 20 shows the processor load, memory, and latency requirements of the encoder. Two different profiles are considered: a 5.1-channel encoder capable of supporting all channel modes, and a two-channel encoder only capable of encoding mono or stereo program material. For the two-channel encoder, a few simplifications can be made to the algorithm, such as removal of any surround channel or LFE processing operations.

In Figure 20, all numbers shown assume an input sample rate of 48 kHz. The 5.1-channel encoder assumes a data rate of 448 kbps, while the two-channel encoder assumes a data rate of 192 kbps. All input conditioning filters are assumed enabled. The input and output buffers are assumed to be double-buffered for simplicity of design.

4.2 Decoder Complexity

In comparison with the encoder, the complexity of the decoder is remarkably low. This reflects a conscious and deliberate process of minimizing the decoder complexity without sacrificing the power of the algorithm. Wherever possible, complexity has been shifted from the decoder to the encoder (such as in determining optimal coding parameters, or performing the 90 degree phase shift needed for Dolby Surround compatibility). As a result, decoding complexity is low enough to make Dolby Digital an attractive

Processor load (MIPS)	5.1-chan	2-chan
Bitstream parsing	5.0	5.0
Bit allocation	8.1	8.1
Mantissa unpacking	5.5	5.5
Block floating-point decoding	1.5	1.5
Inverse transform	6.6	5.8
Downmixing	0.6	0.6
Total	*27.3*	*26.5*

Read/write memory (words)	5.1-chan	2-chan
Input bitstream buffer	1792	1792
Output PCM buffer	3072	1024
Internal buffers	2944	1408
Total	*7.6k*	*4.1k*

Read-only memory (words)	5.1-chan	2-chan
Internal tables	2018	2018
Program code	3500	3500
Total	*5.4k*	*5.4k*

Latency (msec)	5.1-chan	2-chan
Decoding latency	5.3	5.3

Figure 21: Decoder complexity summary

solution for a wide range of consumer electronics applications.

Figure 21 shows the complexity summary for the decoder. Two different decoder profiles are shown: a 5.1-channel decoder capable of full 5.1 channel reproduction, and a two-channel decoder that is limited to only reproducing two output channels.

Note that the two-channel decoder is not constrained from receiving a 5.1-channel bitstream. On the contrary, two-channel Dolby Digital decoders are able to decode 5.1-channel content, and use downmixing to reduce the output channel count. As a result, the two-channel decoder has very close to the same processor load as a 5.1-channel decoder. Of course, since the output buffer requirements are much lower, there is a significant difference in memory usage.

As was the case in the encoder summary, the numbers shown in Figure 21 assume an input sample rate of 48 kHz, and an input 5.1-channel bitstream coded at 448 kbps. Input and output buffers are assumed double-buffered. Also, the latency figures for the decoder do not include transmission latency.

5. SOUND QUALITY

Dolby Digital's sound quality has been evaluated in several listening tests, dating back to the Grand Alliance HDTV audio selection process in 1993. The test results have consistently shown that Dolby Digital is a very high quality coding system, with a demonstrable sound quality advantage over a variety of competing systems.

One of the most comprehensive recent tests was conducted by the Communications Research Centre [10]. This test compared two-channel Dolby Digital against other commercially available coding systems such as AAC, MPEG Layer 2, MPEG Layer 3, and PAC. The results showed that Dolby Digital at 192 kbps performed as well or better than all other coders at their respective tested data rates, with a significant lead over MPEG Layer 2 at the same data rate.

In this test, the performance of Dolby Digital at 192 kbps was found to be comparable to AAC at 128 kbps, a data rate difference of 33%. AAC is a very powerful coding technology, generally recognized as the best system available in terms of coding efficiency. However, this power comes at a substantial cost, as AAC requires significantly more memory resources than Dolby Digital, leading to more expensive decoder implementations.

6. STANDARDS AND APPLICATIONS

Dolby Digital has been successfully designed into a number of applications, including digital television, digital video media products, and emerging application areas such as PC/multimedia and video gaming.

One of the earliest design wins for Dolby Digital was as the sound format for ATSC high-definition television. Since then, it has been adopted by a number of other television applications, including several satellite and cable systems. Most recently, the DVB consortium has specified Dolby Digital as an allowable option, enabling broadcasters to use this technology without simulcasting any other audio format.

Digital video storage media has also turned to Dolby Digital as the preferred audio coding solution. This trend began with the laserdisc format, which added Dolby Digital multichannel audio as a compatible format extension in 1995. More recently, DVD-video has selected Dolby Digital as its audio format for use throughout the world.

Digital television and digital video media share the common characteristic that they both allocate the vast majority of their data rate to digital video, with only a small portion going to digital audio. Within reasonable limits, this fact tends to reduce the importance of coding efficiency relative to other factors, such as consumer features and implementation complexity. Using an audio coding system that has a moderate efficiency advantage (such as AAC) has little impact on the overall data rate for the audio+video program; however using a system with fewer features or a considerably higher implementation cost can have a significant negative impact on the overall application. Finally, Dolby Digital is beginning to find application in PCs as a high-quality sound format for DVD-ROM content and video games. These applications are able to take advantage of Dolby Digital decoding that is already present for DVD-video playback.

7. CONCLUSIONS

This paper has presented the Dolby Digital audio coding system, including its perceptual coding engine, as well as the consumer features that help make it a complete audio solution.

Dolby Digital was developed with a number of unique features not found in competing systems. These include design decisions made in the perceptual coding core, as well as external features not commonly associated with low bitrate audio coding. Some of these design decisions were made to help reduce overall complexity, while others were made in spite of the added complexity, in order to provide greater utility or improved performance.

Considered as a whole, Dolby Digital is a carefully constructed, well-balanced technology, combining high sound quality, flexibility, important consumer features, and moderate implementation complexity. This balance has had a direct impact on its success in the marketplace.

In conclusion, as audio coding becomes a readily-available commodity, and as consumer awareness of desirable system features grows, it will not be possible for any audio coder to be successful without addressing all of these criteria. Sound quality, format compatibility and flexibility, customization to end user requirements, and cost of implementation will all combine to determine the ultimate success of new formats, with no one of these elements able to make up for the lack of support for another.

REFERENCES

[1] U.S. Advanced Television Systems Committee, "Digital Audio Compression (AC-3) Standard," Doc. A/52, Dec. 20, 1995.

[2] C. Todd et. al, "AC-3: Flexible Perceptual Coding for Audio Tranmission and Storage," Presented at the 96th Convention of the Audio Eng. Soc., May 1994, Preprint 3796.

[3] L. D. Fielder et. al, "AC-2 and AC-3: Low-Complexity Transform-Based Audio Coding," Collected Papers on Digital Audio Bit-Rate Reduction, Audio Eng. Soc., Sept. 1996, pp.54-72.

[4] K. Brandenburg and G. Stoll, "ISO-MPEG-1 Audio: A Generic Standard for Coding of High-Quality Digital Audio," Collected Papers on Digital Audio Bit-Rate Reduction, Audio Eng. Soc., Sept. 1996, pp. 31-42.

[5] SMF Smyth et. al., "DTS Coherent Acoustics: Delivering high quality multichannel sound to the consumer," Presented at the 100th Convention of the Audio Eng. Soc., May 1996, Preprint 4293.

[6] K. Brandenburg and M. Bosi, "ISO/IEC MPEG-2 Advanced Audio Coding: Overview and Applications," Presented at the 103rd Convention of the Audio Eng. Soc., Sept. 1997, Preprint 4641.

[7] J. P. Princen and A. B. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time-Domain Aliasing Cancellation," IEEE Trans. Acoust., Speech, Signal Process., Vol. 34, Oct. 1986, pp. 1153-1161.

[8] S. Vernon, "Design and Implementation of AC-3 Coders," IEEE Trans. Consum. Elec., Vol. 41, No. 3, August 1995, pp. 754-759.

[9] W. A. Yost and G. Gourevitch, "Directional Hearing," Springer-Verlag, New York, 1987.

[10] G. A. Soulodre et. al., "Subjective Evaluation of State-of-the-Art Two-Channel Audio Codecs," J. Audio Eng. Soc., Vol. 46, No. 3, March 1998, pp.164-177.

Contacts:

Paul Jacobson
Encore Media Group
Phone (303) 267-5821

Jim Arnold
Dolby Laboratories, Inc.
(415) 645-5116

Robert Mercer
DIRECTV
(310) 726-4683

STARZ! & DIRECTV launch 5.1 Dolby Digital Theatre-quality audio on a premium channel

Premium channel first on DIRECTV with June 29 airing of “Armageddon” at 10:15 PM ET/PT

Englewood, CO—June 29, 1999 Marking the latest digital first for Encore Media Group LLC, Encore’s new hit movie premium channel STARZ! will transmit *Armageddon* starring Bruce Willis June 29 at 10:15 PM ET/PT over DIRECTV with a Dolby Digital 5.1 (also known as Dolby AC-3) signal, providing theatre-quality audio for the first time to standard definition premium channel viewers.

STARZ! and DIRECTV will continue the Dolby Digital (AC-3) signals with upcoming films such as *Six Days Seven Nights*, *Lost in Space*, *ConAir*, *Blade*, *54*, and *The Wedding Singer*.

Dolby Digital 5.1-channel surround, also known simply as Dolby Digital, offers sensational sound by delivering five discrete, full-fidelity (3Hz–20,000 Hz) audio channels (Left, Center, Right, Left Surround, Right Surround) plus a sub-woofer channel. Viewers with Dolby Digital audio receivers can re-create the same spectacular audio environment when viewing STARZ! at home that they enjoy at movie theatres.

Dolby Digital is the latest digital service breakthrough to be offered by EMG, which pioneered the now popular multiplexing of flagship premium channels to better serve the increased channel capacity of direct broadcast satellite and digital cable systems.

“Dolby Digital raises the entertainment value for STARZ! customers and underscores Encore Media Group’s position as the digital programming leader in DBS and cable,” said John J. Sie, chairman, and CEO of Encore. “We now have thirteen premium networks tailored for the digital environment as dependable destinations, and we will continue to push new digital technology into the market so that the capabilities of increasingly sophisticated home entertainment systems are fully exploited.”

“This is exciting news for DIRECTV customers who have come to expect nothing less than the best and the latest in digital picture and sound technology,” said Stephanie Campbell, senior vice president of programming for DIRECTV. “With the STARZ! Dolby Digital launch we are able to deliver more movies with Dolby’s superior sound technology for our customers who own Dolby Digital enabled receivers.”

“We’re excited that STARZ! will now offer Dolby Digital to its standard definition premium channel viewers” said Tom Daily, Dolby’s marketing manager, broadcast products. “With nearly two million Dolby Digital 5.1 channel decoders in use, the popularity of theatre-quality sound in the home is growing fast.”

Dolby Digital is backward compatible with the existing Dolby Surround Pro Logic receivers currently in place. STARZ! will continue to broadcast movies in Dolby Surround while adding the new Dolby Digital.

The Dolby Digital service on the STARZ! East feed will be available to all DIRECTV customers equipped with the latest Dolby Digital DIRECTV System receivers. Currently two models with the Dolby Digital technology are available to DIRECTV customers, the RCA DRD-515-RB and Sony SAT-A4. Other models will follow soon. The STARZ! West feed will carry Dolby Digital later this year.

Encore Media Group will make Dolby Digital available to digital cable operators, whose customers will be able to enjoy theatre-like sound if their cable company chooses to provide digital set top boxes equipped with the Dolby Digital S/PDIF output option.

Consumers will know when a movie is being transmitted on the STARZ! East feed in Dolby Digital by the familiar Dolby “trailers” that are seen in Dolby-equipped movie theatres. Dolby will provide the trailers to EMG for airing before each Dolby Digital transmission on STARZ! The availability of Dolby Digital will also be noted on the rating and content screen that appears before each film.

Dolby Laboratories, the world leader in multichannel audio technology, develops audio signal-processing systems for use in consumer audio and video products, consumer audio and video entertainment media, and professional sound applications that include music recording, broadcasting, and film sound. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo, and European headquarters in England.

DIRECTV is the largest digital television service in the country with more than 7 million customers, including customers subscribing to the recently acquired PRIMESTAR satellite television service. DIRECTV is a registered trademark of DIRECTV, Inc., a unit of Hughes Electronics Corporation. The earnings of Hughes Electronics are used to calculate the earnings per share attributable to GMH (NYSE symbol) common stock. STARZ! is the first-run premium movie service from EMG, offering first-run exclusive movies from Universal, Miramax, Touchstone/Hollywood Pictures, New Line Cinema and Fine Line Features, among others.

EMG is the largest provider of cable and satellite-delivered premium movie networks in the United States, currently counting more than 50 million pay units through its ownership of 13 domestic networks.

S99/12635

Press Release

Media Contacts:
Jim Arnold, Dolby Laboratories
415/645-5116
jja@dolby.com

Catherine Unwin, Dolby Laboratories UK office
44 1793 842 133
cmu@dolby.co.uk

Lynne Daily, Miller/Shandwick Technologies
310/203-0550
ldaily@miller.shandwick.com

For Immediate Release

DVB Project Recognizes Dolby Digital as a Digital Audio Standard

DVB Broadcasters Now Able to Transmit Exclusively in Dolby Digital

San Francisco, CA and Wootton Bassett, UK—July 6, 1999 Dolby Laboratories, the world leader in multichannel sound, today announced its Dolby Digital audio technology has been recognized as an accepted audio transmission format for Digital Video Broadcasting (DVB). The DVB Project, a group of 240 organizations responsible for determining the design standards for DVB, has given broadcasters the option to add Dolby Digital to existing DVB applications, and to transmit exclusively in Dolby Digital in new applications where all receivers are guaranteed to be equipped with Dolby Digital decoding.

“We believe the future of audio is clearly in multichannel surround sound,” said Tony Spath, Marketing Director, Technology, for Dolby Laboratories’ European operation. “We are delighted that broadcasters can now choose Dolby to provide high quality multichannel audio with DVB transmissions.”

The DVB Project states that while all DVB-compliant Integrated Receiver-Decoders (IRDs) support MPEG stereo audio, they may additionally support Dolby Digital audio decoding. Dolby’s acceptance by the DVB Project directly follows the recent announcement that, like Australia, Singapore’s broadcasters will transmit a DVB service that includes only Dolby Digital audio.

S99/12636

"The detailed work on documenting exactly how Dolby Digital could be best carried in the MPEG-2 transport stream so as to be fully DVB compliant was requested by the DVB's Commercial Module and carried out by the DVB Technical Module," commented Peter MacAvock of the DVB Project office. "Broadcasters can now choose the option to transmit Dolby Digital audio secure in the knowledge that receiver manufacturers will be able to make IRDs to a clear and standardized specification."

Dolby Digital–Becoming A Worldwide De Facto Standard
Currently 14.9 million products incorporating Dolby Digital audio technology have been sold worldwide. Dolby Digital is a perceptual audio coding algorithm that takes advantage of auditory masking and both intra- and inter-channel redundancy to enable the efficient storage and transmission of high-quality digital audio. Conceived as a multichannel system, it was first introduced in 1992 for cinema sound. Due to its unmatched combination of audio quality, low data rate and flexibility, Dolby Digital is now available on laser discs, is a mandated audio format for digital versatile disks (DVD) worldwide, and has become the audio standard for ATSC digital broadcast television and SCTE digital cable television.

About The Digital Video Broadcasting Project
The Digital Video Broadcasting Project encompasses over 240 well known organizations in more than 30 countries worldwide. Members include broadcasters, manufacturers, network operators and regulatory bodies committed to designing a global family of standards for the delivery of digital television. Distinguished by the now instantly recognizable DVB logo, DVB-compliant digital broadcasting and reception equipment for professional, commercial and consumer applications is widely available on the market. Numerous broadcast services using DVB standards are now operational in Europe, North and South America, Africa, Asia and Australia.

About Dolby Laboratories
Dolby Laboratories is the developer of audio signal processing systems used worldwide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting and motion picture sound. The privately held company is headquartered in San Francisco, with offices in New York and Los Angeles and European headquarters in England.

Media Contacts:
Jim Arnold, Dolby Laboratories
415/645-5116
jja@dolby.com

Kristin Thomson, Miller/Shandwick Technologies
310/203-0550
kthomson@miller.shandwick.com

For Immediate Release

Dolby Digital selected as a key factor in Singapore's DVB future

SAN FRANCISCO, June 10, 1999—Dolby Laboratories, the world leader in multichannel sound, announced its Dolby Digital audio technology is a key constituent of Singapore's chosen digital television transmission system. In view of Dolby Digital's market penetration in Singapore and across the world, the Singapore Broadcasting Authority demanded the technology be incorporated in the city-state's new digital television system. Singapore's decision last week to adopt the European digital television standard, Digital Video Broadcasting (DVB), is the second consecutive announcement to use DVB with Dolby Digital audio, the first being Australia.

Singapore was the first country to test all of the options available for digital television, before making their final decision in favor of Dolby Digital. Roland K C Tan, the primary audio expert on the Singapore DTV Technical Committee, now on the research staff of Kent Ridge Digital Labs, Singapore explains: "Besides the existing sound format in stereo or Dolby Surround sound, we recommended that a 5.1 multichannel audio format should be considered in our DTV transmission, to add realism in the reproduced sound effect. We also pointed out that with the growing popularity of home theatre systems in Singapore, such a multichannel setup should not be a problem for the new digital TV sound. Moreover, with the exposure to multichannel sound in cinemas and home theatre systems, we believe local home viewers would be ready to accept the higher sound standard at 5.1 channels, if not expecting it as a de-facto standard in the new digital TV technology."

"Dolby Digital (AC-3) has been in active use for more than seven years now," Mr. Tan continued. "We felt that it is the more established technology, especially in the entertainment industries, having been available in the cinema, on laser discs, as well as on DVDs for some time. To include Dolby Digital in our proposed DTV standard has an added advantage in terms of the availability of Dolby Digital processors worldwide as well as in Singapore."

A committee of representatives from government agencies, the broadcast industry, research institutions and commercial suppliers unanimously recommended the DVB standard. Singapore is expecting to launch its DVB system quickly, with digital television services starting as early as 2000. Singapore is reported to be the first country in Southeast Asia to have reviewed and selected a digital television standard.

“Singapore’s demand for the best in digital television technology has laid the foundation for other Southeast Asian countries,” said Tony Spath, Marketing Director, Technology, Dolby Laboratories. “We look forward to establishing Dolby Digital as the digital audio standard for digital television broadcasting worldwide.”

About Singapore Broadcasting Authority

The Singapore Broadcasting Authority (SBA) was formed following the privatization of Singapore’s broadcasting industry. The Singapore Broadcasting Corporation (SBC), the former national broadcaster, was incorporated on October 1, 1994, with the passing of the Singapore Broadcasting Act 1994. This act also provided for the formation of a new statutory board—SBA—under the Ministry of Information and the Arts to regulate and promote the broadcasting industry in Singapore.

The SBA works in association with the following organizations:

- Asia Pacific for Broadcasting Development
- Asia Broadcasting Union
- Commonwealth Broadcasting Association
- Asian Media Information and Communication Centre
- Cable and Satellite Broadcasting Association of Asia
- National Association of Broadcasters
- Commonwealth Broadcasting Association
- World DAB Forum

About Dolby Digital

Dolby Digital is a perceptual audio coding algorithm that takes advantage of auditory masking and both intra- and inter-channel redundancy to enable the efficient storage and transmission of high-quality digital audio. Conceived as a multichannel system, it was first introduced in 1992 for cinema sound. Due to its unmatched combination of audio quality, low data rate, and flexibility, Dolby Digital has since become available on laser disc, a mandated audio format for DVD-Video discs worldwide, and the audio standard for ATSC digital broadcast TV and SCTE digital cable TV.

About Dolby Laboratories

Dolby Laboratories is the developer of audio signal processing systems used worldwide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting, and motion picture sound. The privately held company is headquartered in San Francisco, with offices in New York and Los Angeles and European headquarters in England.

S99/12602

Press Release

Dolby

FOR IMMEDIATE RELEASE

Media Contact: Elmar Stetter
Dolby Laboratories
(44) 1793 842100
es@dolby.co.uk

ProSieben is first broadcaster in Europe to transmit Dolby Digital 5.1-channel sound

Dolby Digital brings digital surround experience to viewers of DVB digital TV format

Wootton Bassett, 19 July 1999—Dolby Laboratories announced today that German Broadcaster ProSieben Media AG will be transmitting a Dolby Digital 5.1 audio stream with prime-time movies in the near future. The broadcaster is planning to transmit its first feature film with Dolby Digital multichannel sound around the time of IFA '99 (28.8 to 5.9.99 in Berlin). Set-top boxes suitable for receiving this transmission are also expected to be on show at IFA, by several manufacturers.

"ProSieben, as an innovative and progressive media company, is proud to be the first in Europe," said Erich Merkle, Technical Director of ProSieben Media AG. "Dolby Digital multichannel surround sound brings genuine extra value to digital television and is an improvement that is clearly recognised by our audiences. We are also planning to set up, in conjunction with Astra, a programme sample loop that will allow the retail trade to demonstrate Dolby Digital 5.1-channel digital surround sound from ProSieben everywhere, and at any time."

"Market figures for home entertainment equipment clearly demonstrate that consumers like multichannel sound, and the demand is growing fast. With the rapid acceptance of DVD with Dolby Digital, the public will start to expect new digital television services to deliver digital multichannel audio," commented Dolby's Tony Spath, Marketing Director, Technology.

Dolby engineers are now working with ProSieben to establish the necessary in-house infrastructure for digital multichannel work, using Dolby E, and liaising with the film production and post-production studios concerned.

W99/124

Dolby Laboratories Inc. 100 Potrero Avenue, San Francisco, CA 94103-4813 Telephone 415-558-0200 Fax 415-863-1373
Wootton Bassett, Wiltshire, SN4 8QJ, England Telephone (44) 1793-842100 Fax (44) 1793-842101 www.dolby.com

Dolby Digital is the audio format for the ATSC digital television standard, and the digital cable standard SCTE. For DVD (Digital Versatile Disc), Dolby Digital is a standard audio format worldwide, meaning every DVD player in the world can decode it. Dolby Digital, with its 5.1 separate channels, is now the backbone of the rapidly growing home theater trend.

Dolby Laboratories is the developer of audio signal processing systems used worldwide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting, and motion picture sound. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo and European headquarters in England.

Press Release

FOR IMMEDIATE RELEASE

Media Contacts: Catherine Unwin
Dolby Laboratories
(44) 1793 842133
cmu@dolby.co.uk

Jim Arnold
Dolby Laboratories
(1) 415-645-5116
jja@dolby.com

Set-top box manufacturers prepare for European broadcasts with Dolby Digital 5.1 audio

Lemon, Panasonic, and Radix step up with products to enable digital surround sound for viewers of DVB digital TV format

Wootton Bassett, 3 August 1999—Following the recent announcement that German free-to-air satellite broadcaster ProSieben Media AG will be transmitting a Dolby Digital 5.1-channel audio stream with prime-time movies in the near future, three set-top box manufacturers will have product on view at the Internationale Funkaustellung (IFA) '99 (28.8.99 – 5.9.99) in Berlin.

Lemon Electronics will present two new DVB receivers using the latest Volksbox technology. The first is called DreamMachine, a free-to-air version with a coaxial S/PDIF output, and the second one is called Volksbox gamm@-CI DAD with a unique combination of Digital, Analogue, ADR and Dolby Digital in one device, including a Common Interface for Pay-TV modules and two S/PDIF outputs (optical and coaxial) for the Dolby Digital 5.1-channel audio.

Panasonic Germany will launch their first DVB-Digital satellite receiver at IFA 99. The universal home entertainment satellite decoder S3 features integrated CI (Common Interface) for both free-to-air and Pay-TV, EPG (Electronic Program Guide) and Bookmarkings. Interactive services can be enjoyed over the digital platform F.U.N. Free Universe Network. The S3 features analogue stereo outputs and optical Dolby Digital 5.1 pass-through via S/PDIF.

W99/132

Radix will be showing the EPSILON 3 CI, which offers integrated Common Interfaces, and can receive both free-to-air and Pay-TV. One of the main features of the EPSILON 3 CI is the S/PDIF output, which can connect directly to a Dolby Digital 5.1-channel audio decoder.

All of these products will allow the Dolby Digital 5.1-channel audio to be extracted from the DVB broadcast stream and sent to a digital output, allowing full 5.1 decoding in any Dolby Digital home cinema system. “The manufacturers of Dolby Digital home cinema systems that we have spoken to are naturally delighted with this development,” said Tony Spath, Marketing Director Technology for Dolby Laboratories. “There is now a second important source of high-quality Dolby Digital 5.1 audio to complement DVD playback, where Dolby Digital is now the accepted standard in Europe.”

ProSieben has access to a large catalogue of major film titles and plans regular DVB transmissions of blockbuster films. For maximum impact, these will be broadcast in widescreen format with Dolby Digital 5.1 audio. The broadcaster is planning to transmit its first feature film with Dolby Digital multichannel sound around the time of IFA ‘99.

Dolby Digital is the audio format for the ATSC digital television standard, and the digital cable standard SCTE. For DVD (Digital Versatile Disc), Dolby Digital is a standard audio format worldwide, meaning every DVD player in the world can decode it. Dolby Digital, with its 5.1 separate channels, is now the backbone of the rapidly growing home cinema trend.

Dolby Laboratories is the developer of audio signal processing systems used worldwide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting, and motion picture sound. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo, and European headquarters in England.

Press Release

FOR IMMEDIATE RELEASE

Media Contacts: Catherine Unwin
Dolby Laboratories
(44) 1793 842133
cmu@dolby.co.uk

Jim Arnold
Dolby Laboratories
(1) 415-645-5116
jja@dolby.com

Jean-Louis Erneux
CANAL+
(33) 1 4425 7581
jerneux@canal-plus.com

CANAL+ and Sony introduce a new generation of digital TV set-top boxes with Dolby Digital 5.1-channel sound at IFA'99

IFA '99, Berlin, Hall 18, Sony Stand

Wootton Bassett, 17 August 1999—European pay-TV leader CANAL+ and the Sony Corporation today announced that they will together introduce set-top boxes with Dolby Digital 5.1-channel sound during the Internationale Funkausstellung (IFA '99) in Berlin.

This new format will be presented to the public by CANAL+ TECHNOLOGIES, CANAL+'s technology division and the world's leading supplier of digital broadcasting software solutions, and the Sony Corporation. The two companies will jointly show a new generation of digital television set-top boxes which will also offer enhanced web-related services and a wealth of advanced interactive and multimedia applications.

CANAL+ and Sony are convinced that digital multichannel sound has a major role to play in the development of digital television, particularly in Europe, where CANAL+ alone already counts over 3.1 million digital subscribers. The new generation of set-top boxes will allow features such as access to the Internet in an environment specifically designed for television viewers, a full e-mail service and enhanced interactive and multimedia applications. The Dolby Digital multichannel surround sound will greatly contribute to this totally new television experience.

Dolby Digital is the audio format for the Advanced Television Systems Committee (ATSC) digital television standard, and the digital cable standard of the Society of Cable Telecommunications Engineers (SCTE). It has also recently been recognised as an

W99/151

accepted audio transmission format for Digital Video Broadcasting (DVB). For DVD (Digital Versatile Disc), Dolby Digital is a standard audio format worldwide, meaning every DVD player in the world can decode it. Dolby Digital, with its 5.1 separate channels is now the backbone of the rapidly growing home cinema trend.

Dolby Laboratories is the developer of audio signal processing systems used world-wide in consumer audio and video products, on consumer audio and video entertainment media, and in professional sound applications that include music recording, broadcasting and motion picture sound. The privately held company is headquartered in San Francisco, with offices in New York, Los Angeles, Shanghai, Tokyo and European headquarters in England.

CANAL+ TECHNOLOGIES, the technology arm of CANAL+, Europe's largest pay-TV operator with over 12 million subscriptions and revenues in excess of FRF 12 billion in 1998, offers a wide range of field-proven and flexible systems to broadcasters and digital operators around the world. The world's leading supplier of digital broadcasting software solutions with over three million set-top boxes currently in use, CANAL+ TECHNOLOGIES has licensed its interactive middleware MEDIAHIGHWAY and conditional access system MEDIAGUARD to 25 international manufacturers and consumer electronics suppliers.

Sony Corporation is a leading manufacturer of audio, video, communications and information technology products for the consumer and professional markets. Its music, pictures and computer entertainment companies make Sony one of the most comprehensive entertainment companies in the world. Sony recorded consolidated annual sales of over $51 billion for the fiscal year ended March 31, 1998, and it employs approximately 170 000 people worldwide.

Press Release

Dolby

Media Contacts: Catherine Unwin, Dolby Laboratories
(44) 1793 842133
cmu@dolby.co.uk

Jim Arnold, Dolby Laboratories
(1) 415 645 5116
jja@dolby.com

First DVB demonstration in France with Dolby Digital 5.1-channel audio

Paris, October 21, 1999—Paris-based digital satellite broadcaster TPS la Télévision Par Satellite has successfully transmitted film trailers with Dolby Digital 5.1 multichannel soundtracks, via Eutelsat Hot Bird 4. French broadcaster TF1 hosted the demonstration in their auditorium, as part of the 2nd International Multichannel Sound Forum held in Paris from 21-22 October 1999.

The demonstration showed, in practice, how Dolby Digital multichannel sound can be transmitted within the DVB standard to DTV audiences in France. The Dolby Digital bitstream was extracted from the DVB transmission, at the receiving end by a commercially available consumer set-top box. The Dolby Digital bitstream was then decoded by an AV amplifier with Dolby Digital decoding, a consumer electronics item that is already widely available from many manufacturers.

Dolby Digital – Becoming A Worldwide De Facto Standard
Currently 19.6 million products incorporating Dolby Digital audio technology have been sold worldwide. Conceived as a multichannel system, it was first introduced in 1992 for cinema sound. Due to its unmatched combination of audio quality, low data rate and flexibility, Dolby Digital is the standard multichannel audio format for digital versatile disks (DVD) worldwide and has become the audio standard for ATSC digital broadcast television and SCTE digital cable television. In July this year Dolby Digital was recognised by the DVB Project as an accepted audio transmission format for new and existing DVB applications; German broadcaster ProSieben subsequently began transmitting feature films in Dolby Digital in September.

About Dolby Laboratories
Dolby Laboratories is the developer of signal processing systems used worldwide in applications that include motion-picture sound, consumer entertainment products and media, broadcasting, and music recording. Based in San Francisco with European headquarters in England, the privately held company also has offices in New York, Los Angeles, Shanghai, and Tokyo.

W99/185

DVD Intelligence

9 February 1998 Vol 1/No 3

First with news and analysis of the global market for DVD products

Dolby AC-3 Wins Mandatory Status On PAL/SECAM Discs

DOLBY LABORATORIES CELEBRATED victory when, last December, in a development under-reported, the DVD Forum upgraded the company's AC-3 audio system from an option to a mandatory inclusion in DVD Video discs produced for the PAL/SECAM territories. Officially, the DVD discs for these markets are no longer required to include MPEG Multichannel audio, which its progenitor, Philips, fought hard to impose as primary system. *Jean-Luc Renaud reports.*

At the Intercontinental Hotel in Tokyo on December 5, the 10 members of the Steering Committee of the DVD Forum voted 8-2 to change the DVD-V specifications (see official tables p. 2).

They agreed to include Dolby Digital (AC-3) in the list of audio types, at least one of which must be on 625/50 PAL/SECAM discs. This means that in effect 625/50 discs can be made with a Dolby Digital track alone and obviating the need for MPEG audio.

The member companies supporting the change comprised Hitachi, Matsushita, Mitsubishi, Pioneer, Thomson Multimedia, Victor Company of Japan (JVC), Time Warner and Toshiba. Philips and Sony were the dissenting voices.

In a press statement, Dolby was quick to characterise this as "a welcome change for those content providers who wish to provide content which contains a 5.1 channel Dolby Digital track and, for reasons of limited disc capacity, do not wish to also have to include a 2-channel MPEG-1 audio track." Making the point clear, Dolby stated: "With the specification change, the MPEG-1 track is no longer required for conformance."

Philips had announced at the IFA consumer electronics fair at Berlin in August that agreement had been reached mandating MPEG Multichannel for PAL countries. As it turned out, Philips' triumph was to be short-lived. The Dutch company was reportedly behind schedule in delivering MPEG-2 Multichannel encoders to DVD developers.

Philips, naturally, has deplored the result of the December vote. "We are sincerely convinced that in the long run the decision of the DVD Forum is not in the interest of the consumer," said Jan Oosterveld, senior director of corporate strategy at Philips.

continued on page 2

DVD Intelligence

Euro $720bn forecast to year 2010, p 4

LABORATORIES INC.

3 - 1998

RECEIVED

July 1998 Vol 1/No 13

First with news and analysis of the global market for DVD products

DVD Publishers Turn Deaf Ear To MPEG Multichannel Sound

SIX MONTHS AFTER THE DVD FORUM UPGRADED Dolby AC-3 audio from an option to a mandatory inclusion in DVD Videos published in PAL/SECAM territories, the rival MPEG-2 Multichannel Audio system appears to be on the way out. And the simpler MPEG-2 stereo option fares no better. *Barry Flynn reports.*

Initially, Philips succeeded in mandating MPEG Multichannel alone as the audio standard (see *DVD Intelligence*, April/May 1997), but the policy was reversed by the DVD Forum in December 1997 (see *DVD Intelligence*, February 1998).

Of the publishers surveyed by *DVD Intelligence*, only Columbia TriStar has had a consistent policy of using both types of audio on its European discs. But now the studio is thinking again. According to Steve Brown, the UK operations director: "Columbia are reviewing their options on what to do with sound."

Brown will not reveal which way the studio is leaning, but the implication is that one of the two systems could be ditched. Given current market sentiment, *DVD Intelligence* guestimates that MPEG Multichannel is the most likely casualty. Indeed, Columbia's forthcoming titles such as ***Sense And Sensibility*** and ***Midnight Express*** are Dolby-only.

Warner Home Video's Neil McEwan says that the studio's current policy for European-aimed DVD Videos is "all Dolby AC3 and only AC-3 – either stereo or 5.1."

Suggesting that this is likely to be Warner's policy for the foreseeable future, he adds: "It's our view that at the moment there is no market demand for (MPEG Audio). The rest of the market is going AC-3. At this point in time there is no reason for anybody to buy kit in MPEG-2."

Although the music products of Warner Vision, a separate division within Time-Warner, might seem to dictate a different policy, it is eschewing MPEG audio also. Frank Brunger, Warner Vision's senior director of international marketing and sales, says: "All our discs are DVD10s: we use (Dolby AC-3) 5.1 on one side and LPCM (Linear Pulse Code Modulation] on the other." His division is not planning to use MPEG Audio, he adds, because "everybody is using AC-3 where they're putting multichannel sound on."

It is so far unclear what Buena Vista's policy is. The

continued on page 2

Annex E

(Informative)

List of patent holders

The user's attention is called to the possibility that - for some of the processes specified in this Recommendation | International Standard - conformance with this International Standard/Recommendation may require use of an invention covered by patent rights.

By publication of this Recommendation | International Standard, no position is taken with respect to the validity of this claim or of any patent rights in connection therewith. However, each company listed in this annex has undertaken to file with the Information Technology Task Force(ITTF) a statement of willingness to grant a license under such rights that they hold on reasonable and non-discriminatory terms and conditions to applicants desiring to obtain such a license.

Information regarding such patents can be obtained from the following organisations.

The table summarises the formal patent statements received and indicates the parts of the standard to which the statement applies. The list includes all organisations that have submitted informal statements. However, if no "X" is present, no formal statement has yet been received from that organisation.

Company	ISO/IEC 13818-2	ISO/IEC 13818-3	ISO/IEC 13818-1
AT&T	X	X	X
BBC Research Department			
Bellcore	X		
Belgian Science Policy Office	X	X	X
BOSCH	X	X	X
CCETT			
CSELT	X		
David Sarnoff Research Center	X	X	X
Deutsche Thomson-Brandt GmbH	X	X	X
France Telecom CNET			
Fraunhofer Gesellschaft		X	X
GC Technology Corporation	X	X	X
General Instruments			
Goldstar			
Hitachi, Ltd.			
International Business Machines Corporation	X	X	X
IRT		X	
KDD	X		
Massachusetts Institute of Technology	X	X	X
Matsushita Electric Industrial Co., Ltd.	X	X	X
Mitsubishi Electric Corporation			
National Transcommunications Limited			
NEC Corporation		X	
Nippon Hoso Kyokai	X		
Nippon Telegraph and Telephone	X		
Nokia Research Center	X		
Norwegian Telecom Research	X		
Philips Consumer Electronics	X	X	X
OKI			
Qualcomm Incorporated	X		
Royal PTT Nederland N.V., PTT Research (NL)	X	X	X
Samsung Electronics			
Scientific Atlanta	X	X	X
Siemens AG	X		
Sharp Corporation			
Sony Corporation			
Texas Instruments			
Thomson Consumer Electronics			
Toshiba Corporation	X		
TV/Com	X	X	X
Victor Company of Japan Limited			

Dolby
Dolby Laboratories Inc
Dolby Laboratories Licensing Corporation
Signal Processing and Noise Reduction Systems
100 Potrero Avenue
San Francisco, California 94103-4813
Telephone 415-558-0200
Facsimile 415-863-1373
Telex 34409

June 2, 1995

International Telecommunications Union
Radiocommunications Sector
c/o Robert W. Jones, Director
Place des Nations
CH-1211 Geneva 20
SWITZERLAND

Fax: 011-41-22-730-5785

AC-3 Intellectual Property Rights

Dear Sir,

Dolby Laboratories Licensing Corporation (Dolby) has been engaged in licensing intellectual property rights (IPR) to manufacturers of consumer and professional audio and audio/video equipment for nearly 30 years. This letter is meant to disclose Dolby's policy with respect to licensing IPR for AC-3 technology which may be included in an ITU-R Recommendation.

Should AC-3 be included in an ITU-R Recommendation, Dolby will license any IPR it owns and which is required to implement AC-3. The license will be offered on a non-exclusive and non-discriminatory basis, at reasonable terms, to all qualified manufacturers. Dolby will not charge broadcasters or network operators when transmitting signals which correspond to the AC-3 specification.

Sincerely,

Ed A. Schummer
Vice President, Licensing

cc: Steve Forshay
Vice President, Engineering

GDC21S802A
(HDTV All-format Single Chip Decoder)

for
HDTV Set-top box &
PC Add-on card for HDTV receiving

Version 1.0

April, 99

LG Electronics

LDS-GDC21S802A-9904 / 10

TABLE OF CONTENTS

1. General Description

This single chip decoder, GDC21S802A, is implementing ATSC standard (A/53) defined for the Digital TV broadcasting in the U.S.A. GDC21S802A receives digital transport packet data from VSB/FEC block or IEEE1394 port, parses the TP packets, decodes video and audio bit streams in a chip. In decoding ATSC system (A/65), GDC21S802A parses 32 TP packets with different identifications, such as PID, table_id, and others, with automatic CRC check. GDC21S802A also supports the Data Broadcasting specification (T3/S13 Doc. 010) under development within ATSC working group. For the video decoding, GDC21S802A receives and decodes the ATSC defined 18 different video formats for the display on the NTSC TV monitor. In decoding the 18 different video formats for display on a 720 x 480i TV monitor, GDC21S802A uses only 4 Mbytes of external memory to make the device feasible for low cost set-top box or PC add-on card. Since the Digital TV allows maximum encoded data rate of 19.4 Mbps and 38.8 Mbps for terrestrial and cable applications, respectively, the system decoder can handle incoming bit stream of up to 38.8 Mbps. For audio decoding, GDC21S802A supports AC-3 full specification (A/52) defined by Dolby Lab as well as Pro-Logic decoding and 3-D virtual surround. The decoder GDC21S802A consists of three major decoding blocks, i.e. system, video, and audio decoding blocks, and other interface blocks as shown in figure 1.

2. Key Features

- Single chip HD to SD all format down converter including ATSC system, video, and audio
- Fully comply on the ATSC specifications (A/53, A/52, A/65, and others)
- Hardware filtering up to 32 TP packets with CRC checking
- Supporting Data Service defined at ATSC T3/S13 (on-going specification)
- IEEE1394 interface to 1394 Link Layer Controller (LLC)
- 16 bit host processor interface
- I2C master controller
- VIP ver. 1.1 host port interface
- Automatic lip synchronization between video and audio
- Connect to an external PCI bus interface IC
- Decoding ATSC 18 video formats within only 4 Mbytes external SDRAM
- Video display format conversion for NTSC TV monitor with LG’s proprietary algorithms
- Support VGA graphics controller video port interface of VIP ver. 1.1
- Dolby AC-3 decoding, Pro-Logic decoding, and 3-D virtual surround
- Dual audio service, main and associated audio service, is handled.
- Support IEC958 (S/PDIF) audio output format

At the current silicon, AC-3 dual audio service is implemented, but it is not fully tested due to lack of support from Dolby Lab. Due to ambiguity of A/65 document, PSIP filtering with EIT source_id and ETT ETM_ID are implemented, but not recommended to use them.

3. Block Diagram

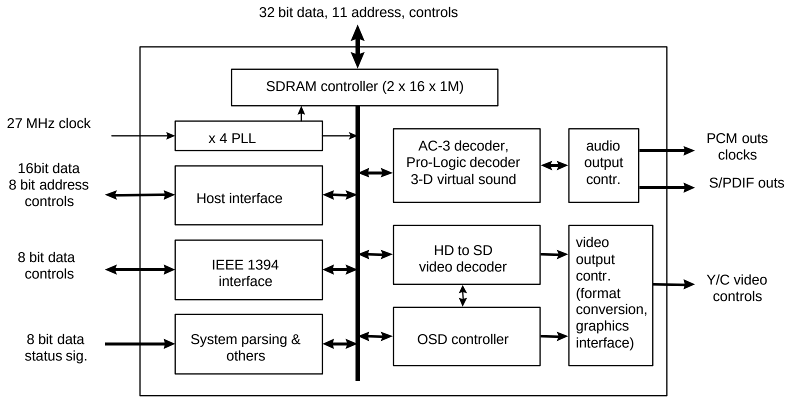


Figure 1. Block Diagram of GDC21S802A

4. Block Descriptions

4.1 System Decoding Block

The system decoding block consists of five sub-blocks, TP packet parsing, system memory interface, PSIP packet parsing with CRC check, IEEE1394 interface, and host interface. For the PC add-on card applications, an external PCI interface IC is going to be used along with the decoder GDC21S802A. At the moment, PCI interface IC from PLX Technology, PCI9080, is under evaluation for its feasibility.

4.1.1 TP Parsing

This block parses the incoming TP packets based on the user control, such as PID, table_id, section number, version number, and others which are set by the control program running on the external processor, using the register sets in the host interface block in the decoder. GDC21S802A allows up to 32 different TP packets being parsed simultaneously by writing the selected IDs at the corresponding registers. It also allows hardware CRC checking for the parsed PSIP packets. The video and audio elementary streams are stored at the encoded bit buffers at the external DRAM for decoding by the video and audio decoding blocks. The time stamps, PCR, PTS and DTS coming with the bit streams are decoded and kept internally in order to keep automatic lip synchronization. The PSIP packets being parsed based on the selected IDs are stored at the corresponding buffers in the external DRAM for later access by the external processor. Upon the parsing of a PSIP packet, the packet is gone through CRC checking operation in the unit of section and its readiness of a PSIP section is informed to the external processor via interrupt if there is no error after CRC checking. During the process, the block checks if there are errors or loss in the TP stream and performs proper error handling operations.

For the Data Broadcasting defined at ATSC specification (T3/S13 Doc. 010), the block parses the selected data service packets and, either stores at the pre-assigned buffer in the external DRAM or feeds them out directly via IEEE1394 port or 16 bit host interface port for the processing of the data service by the external processor.

When data service packets are being parsed, the external device connected to the IEEE1394 or 16 bit host interface should reads out the data service packets at the maximum slew rate in order to not loose any data due to overflow. The figure 2 shows the internal data flow of the system block. To the IEEE1394 port, either incoming TP streams of full bit rate without any parsing, or all TP packets corresponding to the selected channel which is on decoding by video and audio decoding blocks, or selected data service packets are out. For the same port, either TP packets or A/V elementary streams can be fed. The input source is selected using the system control flags. To the 16 bit host interface port, the full bit rate TP streams are not allowed to be out, but other two input sources, selected TPs or data service packets, can be out. However, only one source, either selected TP packets or data service packets, is available at only one output interface at a time, either IEEE1394 or 16 bit interface. Selected TP streams under decoding is not allowed to be stored at the external DRAM for access through 16 bit interface by the external processor. For the same 16 bit host interface, A/V elementary streams can be fed as shown. For the A/V elementary stream feeding, the system parsing block is by-passed by setting a control flag at the system control register (address: 0x00). The system block just receives the elementary bit stream and stores it at the corresponding bit buffer (video or audio). This feature allows GDC21S802A decoder to be used as a decoding engine for DVD bit stream even though the decoder can not parse Program Stream.

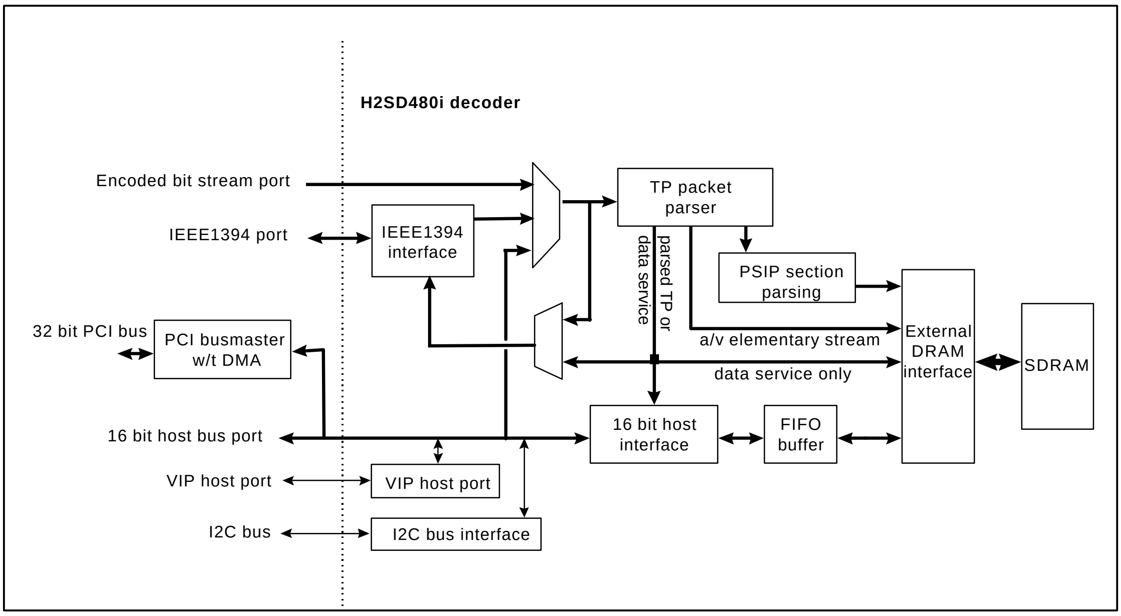


Figure 2. Data Flow Diagram of the System Decoding Block

The block can parse the incoming TP stream at a rate of 80 Mbps from the 8 bit TP data port or IEEE1394 port.

4.1.2 System Memory Interface

The sub-block interfaces with the SDRAM controller in the GDC21S802A in order to write encoded video/audio bit streams, selected PSIP packet data, and data service data at the external DRAM. It also provides a path for the external processor to access the external DRAM for On Screen Display data downloading and PSIP packet data reading.

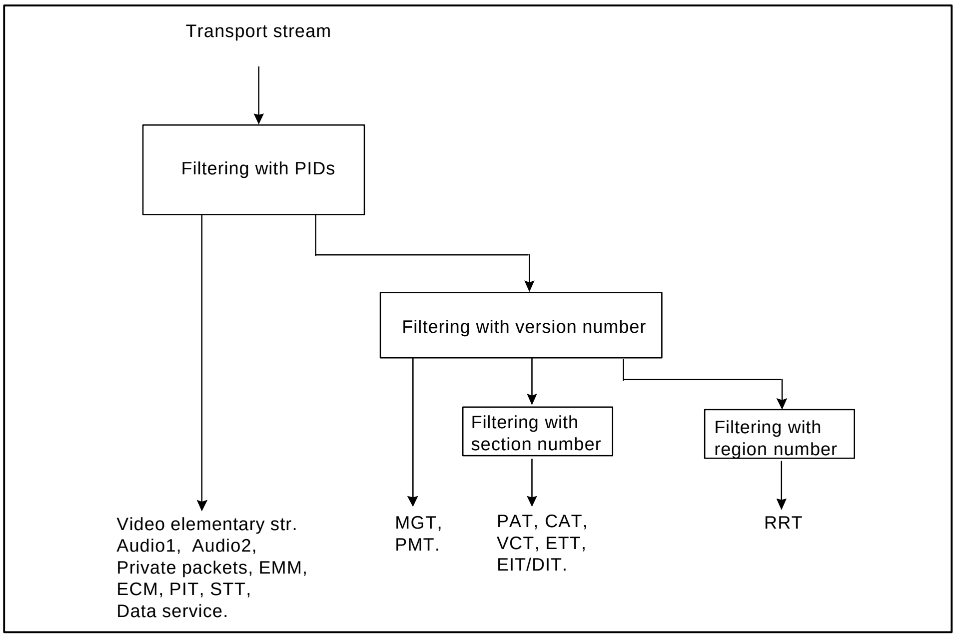


Figure 3. Filtering Flow of Transport Stream

4.1.3 PSIP Packet Parsing with CRC Check

Since PSIP packets defined at A/65 are transmitted repeatedly, they are needed to be parsed based on the external control. For this, each PSIP packet can be parsed based on its PID, version number, section number, and others as shown in Figure 3. There are corresponding registers in the decoder for setting proper values. The Section syntax in the 13818-1 specification has CRC field and the CRC checking is required for most PSIP Tables. The CRC checking can be performed either by the external processor after reading the data from the decoder GDC21S802A, or can be performed within the decoder GDC21S802A upon request from the external processor. There is a control register flag for CRC check enabling. Only the section data passed the CRC checking are stored at the external DRAM and informed to the external processor for its readiness. An interrupt is generated when at least one unread PSIP section is stored at the external buffer space.

4.1.4 IEEE 1394 Interface

This port is supported for interfacing with external devices, such as Digital VCR, network devices, and others. The port supports 8 bit parallel data bus and its related control signals for direct connection with TI IEEE1394 Link layer controller (TSB12LV41). It is bi-directional port, either for outputting the incoming TP packets either selected or not, or selected data service TP packets, or for receiving TP packets for decoding in the GDC21S802A. There are control register flags for the selection of input data and its transfer direction.

4.1.5 Host Interface

It is an interface for an external processor, such as Motorola 68000 series, Hitachi SH series, and others. There are two different interface protocols supported, asynchronous and synchronous protocols. It is selectable by a signal. The interface can be used for accessing the internal registers or data at the external DRAM connected to the decoder. It can be also used for downloading the On Screen Display data. This interface is good for set-top box implementation where the external processor is directly connected with the decoder GDC21S802A. The decoder GDC21S802A also supports I2C master interface in order to control devices, such as Tuner and VSB demodulator, without using a separate I2C controller IC. The decoder also supports VIP host port, version 1.1 for the VGA controller interface in order to have the VGA controller control the decoder through the port without using the PCI interface.

4.1.6 PCI Bus Interface

In order to provide the PCI interface for PC add-on card, an external PCI interface IC is used along with the GDC21S802A decoder. At the moment, PCI interface IC from PLX Technology, PCI 9080 is under evaluation. In order to enable 16 bit word burst transfer of encoded bit streams between PCI9080 interface IC, the decoder GDC21S802A provides separate interface signals, such as *pci_clk, pci_rdrdy, pci_rden/, pci_wren/, and pci_wrwait/*. In this mode of data transfer, the 16 bit word data are connected to the Transport Stream parsing block (refer to Fig. 2). This interface is used only for the encoded A/V elementary bit stream feeding or Transport stream outputting via PCI bus using its DMA master mode. For the other I/O operations through PCI interface IC (slave mode PCI operations), such as accessing the internal registers or the external DRAM spaces, the general 16 bit host interface signals are used with minor glue logic. Since the same 16 bit data bus is shared between general I/O operations with its control signals and burst transfer with its control signals, either general I/O or burst transfer should be enable one at a time, not both at the same time.

4.1.7 Audio and Video Lip Synchronization

In addition to the above operations, the system decoding block does communicate with video and audio decoding blocks in order to keep automatic lip synchronization and proper decoding operations. For the automatic lip synchronization between audio and video, the system decoding block maintains all time stamps, such as PCR, STC, and PTSs for video and audio access units. Decoding of audio and video bit streams and their playback are controlled by comparing the corresponding PTS of a given frame with the STC
value. Depending on the comparison results, each decoding block keeps synchronization by either skipping or repeating a decoded frame with smooth signal transition. To enable the synchronization operation in the decoder, GDC21S802A, there is a control register flag. All the time stamps can be accessed through 16 bit host interface by the external processor.

4.1.8 Error Handling

The decoder GDC21S802A has many different level of error handling operations in order to give robust error immunity. The incoming encoded bit stream goes through error handling processes at system, video, and audio decoding blocks. The system decoding block checks if the incoming bit stream is error-free at packet unit. If it is error corrupted, the packet is discarded. If the packet is video elementary stream, an error code is inserted for later treatment at video decoding block. If it is PSIP packet, the status register of the PSIP is cleared. All these cases, an interrupt is issued. If it is either audio elementary stream, private data, EMM/ECM, or data service packet, there is no action except generating an interrupt. In the video decoding block, there are error handling operations at each layer, sequence, GOP, picture, and slice layer. In the audio decoding block, when there is an error in the decoded fame, the previous frame is repeated or muted down depending on the error conditions.

There can be overflow at the bit stream buffers even though it should not happen during normal operation. When the PSIP data are not retrieved in time and there is no empty slot or empty section buffer, then the incoming PSIP data are discarded without being filtered until there is empty space. But, if user data slots are filled with unread user data when there is new user data, then all the user data slots are cleared and the new user data is stored at the clear slot space. In any cases, an interrupt is issued. When there is buffer overflow at video or audio encoded bit buffers, then the corresponding decoding block is reset and started again. For video, the current picture under display is kept until new decoded picture is ready to be displayed. For audio, the output audio signal is muted down smoothly.

4.1.9 Locking to the Encoder System Clock

The decoder, GDC21S802A, also supports features to lock the decoder system clock (27MHz) to the encoder system clock. To allow the clock to be locked to the encoder clock, the chip generates error term between incoming PCR and STC at GDC21S802A. Depending on the existence of an external VCO logic for the system clock generation, the clock locking scheme is different. If there is no VCO logic outside, the decoder GDC21S802A updates the decoder's system time clock (STC) with the incoming PCR value since there is no way to control the decoder clock frequency on the fly. If there is an external VCO, the error term from the GDC21S802A is fed to the VCO input to control the decoder's system clock frequency on the fly as shown in the Fig. 4. So, the GDC21S802A does not update the STC with the incoming PCR value upon its arrival. However, if there is discontinuity in the PCR value, such as discontinuity_indicator field at TP packet, the STC is updated by PCR upon its arrival and its error term is set to zero even though the external VCO logic is in use. Since the discontinuity in PCR value can be set at the current packet or one packet prior to the real change in the PCR value, the PCR and PTS comparison, lip synchronization, is automatically disabled for the transition period.

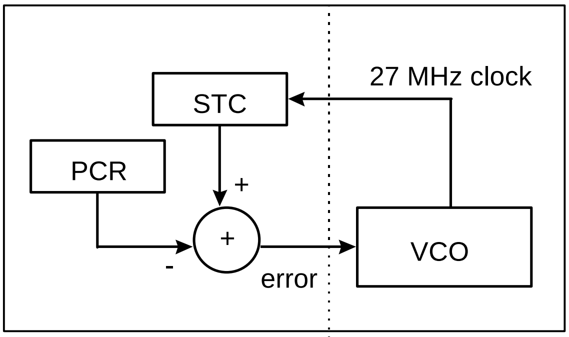


Figure. 4 Timing Recovery Using VCO Logic

The decoder does not support Program Streams nor Still Picture model of the ISO/IEC 13818-1 which are excluded in the ATSC specification (A/53). The system decoding block also supports some error handling operations to be robust on incoming error conditions, some unknown errors and known errors.

4.2 Video Decoding Block

The video decoding block consists of three sub-blocks, HD to SD decoding block, On Screen Display controller, and video display controller.

4.2.1 HD to SD Video Decoding Block

This block is a main engine of the all format decoding for the 18 different video input resolutions defined at ATSC. For any input format, the decoding block decodes the encoded bit stream within 4 Mbytes of the external DRAM. In order to decode the high definition video resolutions, such as 1920 x 1080 or 1280 x 720, with limited resources while maintaining excellent image quality, LG's proprietary down converting algorithm is used, which gives almost the equal visual quality as that of the decimated images from the fully decoded ones. The block also supports VCR-like control commands, such as Slow motion, Single step, and Play, assuming the input bit stream being fed properly. It also has robust error concealment operation for the error corrupted input bit stream.

4.2.2 Video Display Controller

For any input video format, the block converts it into 720 x 480 interlaced format to display on an NTSC TV monitor using LG's proprietary decimation algorithms optimized to generate the best visual quality. Since there are three different aspect ratios on the input image, its aspect-ratio conversion, either letter box or pan scan conversion, is performed for displaying depending on the images aspect ratio and the display monitor type (4x3 or 16x9 TV monitor). Assuming that the 16 x 9 TV monitors have the aspect ratio conversion control for incoming images with 4 x 3 aspect ratio, aspect ratio conversion is performed only for input images with 16 x 9 aspect ratio when a 4 x 3 TV monitor is used. For input image of 640 x 480, it is expanded to fill the 720 x 480 display resolution. 60 Hz progressive images are decimated temporarily to 60 Hz interlaced with minimizing the degradation on image quality. The decoded video data is intermixed with OSD data based on the OSD commands at the final stage of the block. The video data is output at either CCIR656 or CCIR601 format with the corresponding video sync. signals at 27 MHz pixel clock rate. The block also supports the VGA Graphics controller video interface protocol, VIP video port version 1.1.

4.3 Audio Decoding Block

The block consists of two sub-blocks, audio decoding engine and audio output control unit. The audio decoding engine decodes audio elementary stream encoded with Dolby AC-3 specification (A/52) and generates linear PCM data at 48 KHz sampling frequency. Besides decoding the main audio service, the engine also decodes associated service and intermixes it with the main audio if needed to generate a complete audio program for the speakers. For flexible control of the dual audio services, there are control flags for selection of the services, output to the given speaker set, and others. In addition to the AC-3 decoding, the engine also decodes Pro-Logic encoded linear PCM data. The engine also supports the virtual surround algorithm from Dolby Lab. There is a flag for on/off control of the feature.

For outputting the decoded linear PCM data, there are two output ports, linear PCM data with the corresponding control signals and IEC958 (S/PDIF). For the linear PCM output, there are two different output modes, "left justified mode" or "I2S mode" which is selectable with a control flag. Since the output channel is up to 5.1 channels, there are three signal lines of linear PCM data. For an additional headphone channel output, there is one more linear PCM data line carrying Lo/Ro channel data. On the IEC958 port, either two channel linear PCM data or encoded audio bit stream can be carried.

INTERNATIONAL TELECOMMUNICATION UNION

RADIOCOMMUNICATION STUDY GROUPS

Document 10C/25-E
14 May 1999
Original: English only

Received: 14 May 1999

Chairman, Working Party 10C

INFORMATION FOR WP 10C

COMPLEXITY OF DECODERS FOR THE SYSTEMS SPECIFIED IN RECOMMENDATION ITU-R BS.1196

During the discussions leading up the approval of Recommendation ITU-R BS.1196 there were suggestions that a decoder could be implemented that would accept a signal that was encoded per either Annex A or Annex B. The following information regarding the complexity of dual decoders was provided to the DVB Technical Module in January 1999 in response to an agenda item of audio decoder complexity. This information may be of interest to some members of WP 10C.

This contribution presents a cost/complexity analysis of AC-3 and MPEG audio decoding in VLSI implementations optimized for high volume and cost sensitive consumer applications. The results presented here come from six years of experience in designing VLSI MPEG and AC-3 audio decoders for the consumer market. Silicon area data is derived from ICs implementing a portable core cell design from Jacobs Pineda, Inc. (JPI) that is capable of decoding both AC-3 and MPEG using approximately 3.0 sqmm of silicon area in 0.35 micron CMOS technology. For comparison, the silicon area required for a complete MP ML MPEG-2 video decoder IC in 0.35 micron technology is around 100 sqmm, so the incremental cost of audio decoding is only 3% of the total decoder IC cost. When compared to future MP HL (HDTV) video decoders, the relative cost of audio decoding will approach 1%. The table below shows sizes AC-3 and MPEG decoder cells:

Audio Decoding function	**Area (sqmm) in 0.35 u CMOS**
AC-3 and MPEG (dual mode)	3.0
AC-3 only	2.4
MPEG only	1.8

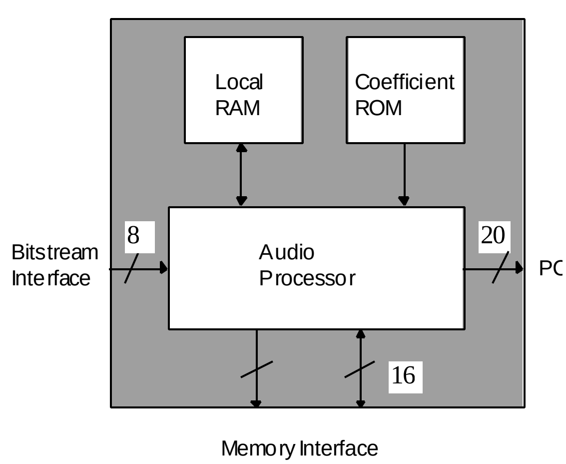
HDL design methodology for audio processing in VLSI circuits

While most implementations for audio processing are based on Digital Signal Processor (DSP) techniques, Jacobs Pineda, Inc. uses a Hardware Description Language[1] (HDL) methodology that results in much smaller designs. The efficiency of the HDL methodology is important for high volume consumer VLSI products where cost is critical. In the HDL methodology, the audio processing algorithms are expressed behaviorally at a high language level, and compiled directly into logic gates using logic synthesis tools[2]. This differs from DSP implementations where the algorithm is expressed in C or assembly language and compiled into a program ROM from which a centralized processing unit executes instructions to perform the computations. The DSP hardware itself is more general-purpose, and as such, not necessarily optimized to any particular algorithm. Like DSP based designs, HDL designs are programmable, but that programmability is at the HDL level, and the program is translated directly into a network of operations implemented in logic gates instead of into instructions in a program ROM.

The efficiency of the HDL approach comes from the fact that an HDL description allows for arbitrary data path configurations. This allows HDL designs to have data paths optimized to match the bandwidth and precision required for specific algorithms. Further, HDL designs naturally implement parallel processing, and allow bandwidth to be balanced efficiently across processes, since the amount of logic dedicated to each process can be tuned to match the bandwidth. Thus HDL designs generally maintain their logic modules at a high level of utilization, and this translates to a reduction in the amount of logic required to execute an algorithm when compared to DSP implementations.

As a data point for the efficiency of the HDL methodology, the JPI dual mode AC-3/MPEG core cell (product designation “J1”) requires only 3.0 sqmm of silicon area in 0.35 micron technology. This includes RAM, ROM, control, and data path, except for an external low bandwidth (4 MHz) 9 KByte of memory needed for buffering purposes. The cell architecture allows that 9 KB memory to be located in memory that is shared with other functions, such as DRAM used in video decoding, as this partitioning works well with VLSI system architectures. Consequently, the cost of the 9 KB memory is typically negligible, either residing off-chip in a DRAM (which is megabytes in size), or in an efficient high-density on-chip RAM. The J1 core cell is typically 3 to 4 times smaller than DSP based designs with the same function. Other examples of the efficiency of HDL methodology applied to designs at JPI are a multi-sample-rate Audio D/A Converter, which requires 1.0 sqmm, and a 3D audio processor requiring 0.5 sqmm in 0.35 micron CMOS technology.

Complexity of arithmetic computation in audio decoders



It had initially been expected that the cost of an AC-3 decoder capable of receiving a 5.1 channel bitstream and producing a down-mixed two channel output would be much greater than an MPEG-1 layer II two channel decoder. This is logical, since AC-3 needs to decode

[1] Verilog HDL and VHDL are two commonly used Hardware Description Languages.

[2] JPI uses Synopsys synthesis tools and Verilog HDL.

five channels before down-mixing to two channels, while MPEG needs to decode only two channels. However, this turns out not to be the case, because the AC-3 algorithm is actually more efficient per channel. AC-3 makes use of FFT techniques to fold the transform and thus reduce computation, while such optimizations are limited for MPEG because of the 16 point FIR used for windowing cannot be folded. Excluding down-mixing, the AC-3 computation requires 15 multiplies/sample, while the MPEG computation requires 36. The following table shows the number of multiplies required for two channel AC-3 and MPEG decoding:

AC-3 5.1 channels down-mixed to 2 channel

Operation	Multiplies/sample	channels	Total
Dequantization	2	5	10
Pre/IFFT/Post	2+7+2	5	55
Down mix	2 * 5 input channels	2	20
Window	2	2	4
Total per stereo pair			89

MPEG 2 channel

Operation	Multiplies/sample	channels	Total
Dequantization	4	2	8
32 point Transform	16	2	32
16 point Window	16	2	32
Total per stereo pair			72

Different algorithm choices may result in other totals. In particular, there are many ways to fold the MPEG 32-point transform that may have different benefits depending on the processing architecture. The number of multiplies can be reduced to 8 or less per sample, but with an increase in data movement, register usage, and addition operations, thus the number of clock cycles may not be reduced proportionally. For this reason, the 16 multiply folding is chosen for this analysis, and in practice has been found to be a good choice in many cases. As shown, the MPEG two-channel algorithm requires approximately 81% of the computation required for the down-mixed AC-3 algorithm.

The total arithmetic bandwidth isn't that large, requiring only 4.3M multiplies per second for AC-3. With clock rates of 50 MHz typical for DSP designs in 0.35 micron technology, only 9% of available bandwidth is utilized, leaving the multiplier logic idle for 91% of the time. In contrast, HDL based designs are more optimal because the amount of logic dedicated to multiplication can be optimally matched to the bandwidth requirement, thereby reducing silicon area and cost.

While the cost of non-down-mixed multi-channel decoders has not been addressed here directly, it should be apparent that because of relative transform efficiencies, the cost of a 7.1 output channel MPEG decoder would be greater than for a 5.1 output channel AC-3 decoder, since the arithmetic throughput required is approximately 2-3 times that for AC-3.

Complexity of parsing and bit allocation

One of the more challenging portions of an AC-3 decoder is the design of the parser and bit allocation logic. The AC-3 algorithm operates in many different modes that are encoded in 105 different kinds of fields in the bitstream, each of which has to be handled differently. Furthermore, AC-3 uses a complex algorithm to compute the bit allocations from parameters passed in the bitstream. By contrast, MPEG has only 20 different field types, only a handful of modes, and the bit allocations that are transmitted directly.

As such, MPEG is an easier algorithm to understand. However, in terms of cost, the AC-3 decoder does not suffer proportionally, and the additional cost for AC-3 is less than might initially be thought. This is because, first, the parser and bit allocation logic are only part of the overall decoder logic. And second, since the complexities of the AC-3 bitstream are more logical than arithmetic, they are efficiently handled by logic functions and state machines that are relatively low cost and well expressed in HDL.

Most of the additional cost of an AC-3 decoder over MPEG is in the parsing and bit allocation areas because of the increased algorithmic complexity, and because AC-3 needs to parse and decode six channels of audio data compared to two channels for MPEG. Thus, most of the 25% silicon area difference (2.4 vs. 1.6 sqmm) between AC-3 and MPEG can be attributed to parsing and bit allocation. One additional note: because parsing and bit allocation can run in parallel with transform and windowing computations, the time to compute the AC-3 bit allocations need not adversely affect the cost or throughput of the main computational logic.

Conclusion

To summarize, experience at Jacobs Pineda, Inc. with several MPEG and AC-3 VLSI implementations using the HDL design methodology demonstrates that a dual mode AC-3/MPEG decoder requires 3.0 sqmm in 0.35 micron CMOS technology. A 2-channel, non-down-mixing, MPEG-only decoder requires approximately 60% of the silicon area of the dual mode decoder, while an AC-3-only decoder requires approximately 80%.

To put the cost of audio decoding into perspective, it should be considered that high volume VLSI circuits in the 100 sqmm range used for consumer products can sell for less than $ 10. Thus the price of the audio decoding function when implemented with HDL methodology is today on the order of $ 0.30. Consideration of cost versus function in a comparison of MPEG and AC-3 might do well to focus more on function and audio quality than cost, especially since the cost will continue to decline with IC technology feature sizes.

Additional reference information and on the function and architecture of the Jacobs Pineda AC-3/MPEG audio decoder, may be found on the internet at: **http://www.jacobspineda.com**.

____