DVD-VIDEO PLAYBACK PRIMER

Revision 1.0
December 5, 1997

TOM MARCHOK (TOM_MARCHOK@CCM.SC.INTEL.COM)
CAREY KLOSS

INTEL CORPORATION

Overview: This document provides an overview of topics that are relevant to DVD technology, particularly as it pertains to DVD-Video playback on the personal computer (PC). This document is intended for those readers who have a minimal understanding of DVD.

1. Introduction

DVD is an optical storage medium that is identical in size and similar in appearance to a CD-ROM, but has far greater storage capacity and versatility. The capacity of DVD disks is currently approximately seven times that of a CD-ROM, and technology to quadruple the capacity of current disks is soon expected to become cost effective. DVD technology was originally created by the entertainment industry, for the purpose of distributing full-length motion pictures. A single DVD disk can store more than two hours of audio and video content at a quality level that surpasses all consumer media to date.

Increases in microprocessor performance, in particular the Pentium® II Processor, permit DVD content to be played back on a PC without requiring expensive add-in hardware. The interactive nature of the PC, coupled with the versatility and extensive storage capacity of DVD technology, make possible a new class of interactive applications - which could change the way that consumers are educated and entertained.

This document provides an overview of those topics that are relevant to DVD technology, particularly as it pertains to playback on the PC.

2. Background - DVD-Video

The term *DVD* represents a family of related disks, all of which use a similar storage technology. Available DVD formats include:

- DVD-ROM – This read only format is similar to the CD-ROM format for compact disks. DVD-ROM disks are currently capable of containing 4.7 gigabytes (GB) of information. The capacity of future generation DVD-ROM disks is expected to reach 17 GB. This storage increase will be attained by storing information on both sides of the DVD disk, and storing two layers of information per side instead of one.

- DVD-R disks allow information to be written once, and subsequently read back an unlimited number of times. DVD-R format is analogous to CD-R format. First generation DVD-R disks have a 3.9 GB capacity. Because DVD-R format recorders currently cost more than US$10,000, use of DVD-R format disks is typically restricted to data archiving and distribution, and DVD content development.

- DVD-RAM – This format permits an unlimited number of writes to and reads from the disk (*Read Write* format). DVD-RAM format is analogous to CD-RW format. The capacity of DVD-RAM format is 2.6 GB. This format is used for archiving data, software development, video and audio editing and recording. DVD-RAM format requires a special cartridge to hold disk, since handling (i.e., fingerprints) can mar the sensitive face of the disk.

- DVD+RW - Also a Read Write format. DVD+RW differs from DVD-RAM in that DVD+RW format is 3.0 GB, and also in that it does not require a special cartridge.

- DVD-Audio – Format which is intended to replace audio CDs. Definition of DVD-Audio format has not yet been completed.

- DVD-Video - Specialized DVD-ROM format that has been designed to contain full-length motion pictures.

This document focuses on the playback of DVD-Video format content.

DVD-Video format contains an *MPEG-2 program stream*. The MPEG-2 program stream contains audio, video, and sub-picture streams, all of which are synchronized in time. DVD-Video format can contain as many as 8 audio streams, 8 video streams, and 32 sub-picture streams. These multiple data streams make DVD-Video a very versatile format. For example, multiple video streams can be used to show different camera angles, or each video stream can be mixed with different audio and sub-picture streams (which are commonly used for subtitles).

DVD-Video format is not only more versatile than other video formats to date, it also offers much higher quality. Video streams are compressed and stored in MPEG-2 format, with a pixel resolution of 720 by 480. Audio streams are stored using the Dolby Digital™ (also called *AC-3*) format, which contains 5.1[1] discrete channels of digital audio recorded at 48 kilo samples per second (Ks/s). Table 1 further details the features and capabilities of DVD-Video format.

Table 1: Features of the DVD-Video format

Feature	Explanation
8 Video Streams	Permits one of multiple video streams to be selected and display Used to show different camera angles of the same scene (e.g., in a sports broadcast), or different versions of the same scene.
32 sub-picture streams	Used for multiple language subtitles or interactive screens (such as menus).
8 Audio Tracks	Used for different languages or music to accompany the same video.
Viewer Control	Allows viewer to control which portions of the content may be accessed. One application is parental control, which would permit the user to omit portions of a movie.
Menus, Chapters Indexing	Easy to navigate menus and chapters make it easy to find any portion of the content on the disk.
Aspect Ratio Selection	Aspect ratio is defined as the ratio of the width of the picture to the height. Video streams can be formatted in either a 16:9 or a 4:3 aspect ratio. Televisions and PC monitors have aspect ratios of 4:3, whereas movie theaters and wide-screen televisions typically have an aspect ratio of 16:9. DVD-Video can be displayed in either aspect ratio, depending upon the display

[1] 5.1 stands for five full range channels and one Low Frequency Effects (LFE) channel.

	device.

3. DVD-Video Playback

DVD-Video playback involves converting the compressed binary information that is stored on a DVD-Video disk into video and sound. Figure 1 illustrates those tasks that are necessary for playing back DVD-Video content. DVD-Video playback begins with reading the appropriate amount and sequence of binary information from the DVD disk - a task which is referred to as *navigation*. The binary information is then divided (or split) into the contained audio, video, and subpicture streams. Each stream is then decoded, and the resulting video and audio information is then rendered (displayed).

Various DVD-Video playback solutions are available. These range from dedicated consumer electronics devices, to mechanisms that may be added to a PC - including hardware boards and software solutions. This section overviews those tasks which are involved in DVD-Video playback, independent of the playback method used. Subsequent sections will detail a software implementation for DVD-Video playback on the PC.

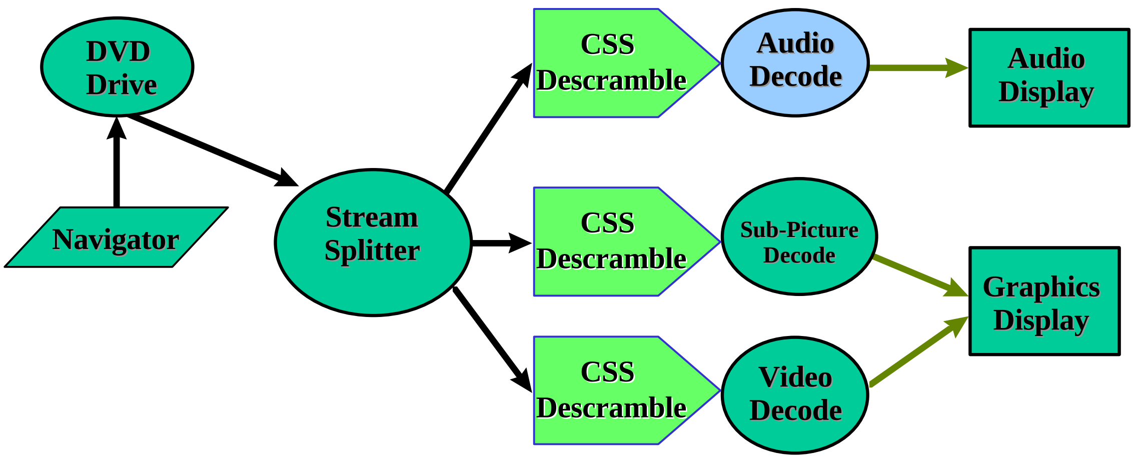

Figure 1. Tasks involved in DVD Playback.

3.1 Navigation

Navigation involves communicating with the DVD drive to scan the DVD disk to locate the appropriate audio, video, and sub-picture data. For playing back movie content, navigation typically looks for the data that corresponds to the next time segment. Depending upon the options presented to the viewer, there may be instances where the navigator will seek data other than the data which corresponds to the next time segment. The navigator is responsible for responding to user-input commands, such as fast forward, rewind, as well as subtitle display.

3.2 Stream Splitting

After the navigator has communicated the location of the desired content to the DVD drive, a process known as *stream splitting* reads and separates the information into the separate audio, video, and sub-picture streams. The user specifies which of the audio, video, and sub-picture streams are selected from the MPEG-2 program stream. Each of these streams is both scrambled (per the Content Scrambling System discussed in Section 5) and *encoded*. The data is encoded, or compressed, to allow a longer duration of content to be contained on a DVD disk. The encoding formats used are MPEG-2 for video, and run length encoding for subpicture. The AC-3 audio format is used in countries which use the NTSC television format, while the MPEG-2 audio format is used in countries which use PAL television format.

3.3 Descrambling

The audio, video, and subpicture streams on DVD-Video disks are typically scrambled according to a copy protection scheme. These streams must therefore be descrambled before they can be decoded. This copy protection scheme was prompted by the motion picture industry, which was initially reluctant to release content on DVD-Video format. Hollywood studios were fearful that the quality and reproducability of DVD technology would make illegal reproduction (*pirating*) of perfect digital copies possible. The motion picture industry worked with the consumer electronics industry, and later the computer industry, to design a system to prevent the illegal copying of content stored on DVD discs. This is known as the *Content Scrambling System* (CSS), and relies on key-based encryption and conditional access to stored content.

The scrambling technique associated with CSS rearranges the content according to a proprietary method. If scrambled information were to be decoded and displayed, severely corrupted audio and video streams would be presented. Figure 2 illustrates a simulated rendered screen shot of scrambled and unscrambled video portions of a movie. To access the encrypted content, a DVD playback device (*DVD player*) must know a set of secret keys and algorithms. A manufacturer must commit that their product will adhere to a set of design guidelines in order to receive this needed information. These guidelines generally require that the DVD playback devices not be designed (or be easily modifiable) to allow illegal duplication. Section 5 provides further details of CSS.

Figure 2: CSS Scrambled vs. Unscrambled Screen

The scrambled audio, video, and sub-picture streams are presented to independent descrambling modules. Each module then applies the CSS keys and algorithms to descramble the incoming streams. The resultant descrambled data streams are subsequently passed to the appropriate decoders.

3.4 MPEG-2 Video Stream Decoding

MPEG-2 Video decoding is the most computationally intensive task involved in DVD-Video playback. It is therefore worthwhile to overview those tasks which are involved in decoding the video stream.

The MPEG-2 video encoding format relies on a number of encoding techniques. A DVD player reverses these encoding actions to decode the encoded video stream. The following sections briefly describes these encoding techniques.

3.4.1 Motion Estimation (Temporal Compression)

The first compression technique which MPEG-2 video uses is *motion estimation*. The associated decompression technique is termed *motion compensation.* Motion estimation is a temporal compression technique. *Temporal compression* techniques rely on similarities between different image frames to reduce image data storage requirements. Motion estimation relies on the property that much of the information in a video segment remains unchanged between successive frames. This is understandable, given that (depending upon the source of the video material and the video display device) either 24 or 30 image frames are displayed per second, and because the video image typically does not change much within any given second. Motion estimation uses this property to significantly reduce the number of bytes needed to represent video information.

Motion estimation divides an image frame into blocks, and compares the content of blocks between frames. Motion estimation searches for similar blocks which exist in multiple frames. The complete information needed to represent such a block need only be stored a single time. Other image frames which contain that same block can then make reference to the previously stored information. This is done by replacing duplicate blocks with *motion vectors*, which indicate the adjusted position of the block. Compression is achieved because the storage space required for the motion vectors is less than that required to store the complete block information.

To further understand motion estimation, consider a short video segment which shows from a distance the side view of a car which is moving across a still backdrop (e.g., a desert landscape). While the position of the car changes, the image of the car itself changes little from one frame to the next. Using motion estimation, the pixel block(s) which define the car need only be stored a single time. Multiple frames can then reference the stored image of the car, using motion vectors to represent the updated position of the car. If the backdrop changes very little, then the same technique can be used to reduce the storage requirements for the still backdrop.

To implement motion estimation, the MPEG2 video encoding process transforms each video frame into one of three possible types of *image frames*:

- *I-frames,* or intra-frames, contain the information required to recreate a complete video frame.
- *P-frames,* or predicted frames, contain references to pixel blocks which are stored in I-frames or other P-frames which occurred earlier in time.
- *B-frames,* which can contain references to pixel blocks from I-frames or P-frames which have occurred either earlier in time or will occur in future frames.

Motion estimation is a computationally intensive task, in part because of the flexibility which it has in searching past and future frames for similar regions. The amount of computation required to encode an MPEG-2 video stream is typically significantly greater than that which is required to play back (decode) the encoded segment. Motion estimation is therefore referred to as an *asymmetric* compression algorithm.

In situations where encoding must be done in real time, such as live television broadcast, *symmetric* encoding techniques (i.e., encoding techniques which have equivalent computational processing requirements for encode and decode) must be used. Since the encode time is constrained, it is typically not possible to minimize the bit-rate of the encoded stream. This helps to explain why the bitrates of other digital video streams, such as digital television broadcast, are significantly higher than those for DVD.

Since the amount of storage space on a DVD disk is constant, minimizing the amount of storage space that an encoded segment requires is often a key objective. The fewer number of bytes in an encoded video segment, the longer duration of content that can be stored on a DVD disk. Achieving minimal MPEG-2 encoding size is an iterative process. During encoding, the video engineer has to trade-off encoding time, the amount of compression attained, and the resultant video quality. Typically, the longer the encode time, the fewer the number of bytes needed to store the resultant encoded segment. Other factors also impact the amount of compression that may be attained, such as the relative amount of motion in the video segment. Less compression is possible for video segments with relatively larger amounts of motion, since there are fewer similarities between successive frames.

3.4.2 Other Encoding methods (Spatial Compression)

After the video content is analyzed and compressed using motion estimation, a number of spatial compression techniques are employed on the remaining blocks of pixels. *Spatial compression* techniques rely on similarities within an image (e.g., that some of the regions of an image frame are the same color) to reduce the number of bytes needed to represent that frame.

The first spatial compression technique is the *Discrete Cosine Transform (DCT)*. The DCT is an arithmetic operation which transforms the color and light intensity information of a block of pixels to an alternate representation. Subsequent transformations are then able to manipulate the transformed information to achieve a more compact representation than otherwise possible.

The DCT computes the spatial frequencies of 8 by 8 blocks of pixels. Here the term *spatial frequency* relates to the rate of change of the color and light intensity of those pixels. The resultant information is typically concentrated among a few spatial frequencies – a condition which facilitates compression. Additional encoding techniques, including quantization, are then used to manipulate the output of the DCT transform, thus allowing the information to be efficiently represented and stored. Q*uantization* groups spatial frequencies which are close to one another. The resultant quantized data is more compact, albeit less precise than the original spatial frequency information. The quantized data is then *run length encoded*, which serves to compress repeated values. Finally, the run-length encoded data is encoded using a *variable length encoding* scheme. Variable length encoding uses a lookup table to map frequently-occuring longer binary sequences into shorter binary sequences. Since the replacement codes vary in length, codes with shorter lengths are assigned to the most commonly repeating binary sequences, which serves to further reduce the size of the compressed video stream.

3.4.3 Decoding

During video stream decoding, each of the corresponding encoding tasks are performed in reverse order to the encoding steps. These steps include variable length decode, run length decode, inverse quantization, inverse DCT, and motion compensation. *Video rendering* is the final task involved in processing the video stream. This involves transforming the video information from its native MPEG-2 video format into a renderable format. The MPEG2 native video format is the YUV4:2:0 digital format. YUV 4:2:0 contains both luminance (brightness, signified *Y*) and chrominance (color, signified *U* and *V*) components. If the rendered video image is to be displayed on a PC monitor, the video information must be transformed from the YUV digital format to *RGB* ("red-green-blue") format which is understood by PC monitors. This process is known as *color-space conversion*. If the resultant video stream is to be displayed on a television, the video stream must be converted into the appropriate analog television format (NTSC or PAL).

3.5 Subpicture Decode

The sub-picture stream is commonly used to implement text subtitles or user menus. The subpicture stream contains two bits for each screen pixel, and is run length encoded. The sub-picture stream is first run length decoded, and then merged (overlayed) on top of the decoded video stream on a per-pixel basis. Each pixel in the subpicture stream assumes one of the following characteristics:

- Transparent
- Any of three non-transparent colors

There are two methods that can be used to merge the bitmaps from the subpicture stream with the frames from the video stream. These include *alpha blending* and *chroma keying*. Both methods determine whether to display the pixel from the video or subpicture stream. In cases where the pixel in the subpicture stream is transparent, the merged pixel data will assume the color of the corresponding pixel from the decoded video stream. Alternatively, when the pixel in the subpicture stream is non-transparent, the displayed pixel assumes the color of that pixel. This is the case whenever that pixel is used as part of a text subtitle of user menu.

3.6 Audio Decode

Audio decode involves transforming the compressed audio data into a format which can be understood by the speakers. Compressed audio can be stored in one of several formats on a DVD-Video disk, including 5.1 channel Dolby Digital™ (AC-3), LPCM, or multichannel MPEG-2 Audio format. This paper focuses on the AC-3 audio format.

The data rate for compressed AC-3 audio on DVD-Video movie titles is typically 384 kbps. This bitstream contain 5.1 channels, which represents five full frequency range channels and one low frequency effects channel of audio information. The five full channels are front right, front left, rear right, rear left, and front center. The *.1* channel is a limited bandwidth channel containing low frequency audio information. Each channel contains 48,000 audio samples per second, where each audio sample is represented using sixteen bits. Compression is achieved using sub-band coding and frequency domain transforms.

Once the audio stream has been decompressed into its native 5.1 channel format, the information can be sent directly to the speakers only if the playback system contains the appropriate number of speakers. If the playback system contains fewer than six speakers, as is typically the case for televisions and PCs, the audio decoder must *downmix* (combine) the six channels into fewer channels of audio data.

Downmixing can be accomplished via a number of techniques. The simplest of these combines all 5.1 audio channels into two. Downmixing may also be performed according to the Dolby Pro Logic™ format, which allows four audio channels to be reconstructed from the two downmixed channels whenever the audio is played on a Dolby Pro Logic™ sound system. Mixing into Dolby Pro Logic™ is therefore primarily useful for feeding into an external speaker system with four speakers. Perhaps more interesting for the PC is downmixing into a virtual surround algorithm such as Intel's 3D Realistic Sound Experience (RSX), which allows the perception of multiple speakers using only two speakers.

4. DVD Playback Solutions

DVD playback devices ("DVD players") first emerged in the form of dedicated consumer electronics devices. These consumer electronics DVD players contain similar playback (but not record) capabilities as video cassette recorders (VCRs), and currently cost several hundreds of dollars ($US). While consumer electronics DVD players have a number of advantages (such as no boot up time, compact size, and simple installation into existing television and consumer electronics environments), they permit a limited amount of interactivity for the user.

The interactive nature of the PC permits it to take advantage of DVD technology in different ways. Thus the synergy between PC and DVD technology has quickly emerged. The interactive nature of PCs make possible a new class of interactive applications which are not possible with a dedicated consumer electronics device. These interactive applications have already begun to emerge in the form of training and entertainment titles. Furthermore, DVD playback is a cost effective use of PC technology. The computational power of the Pentium® II Processor permits DVD playback to be performed in software, essentially at no added cost to the system.

DVD playback solutions for the PC began to appear in the spring of 1997. These solutions initially retailed for approximately US$500, and consisted of a hardware (PCI bus) add-in board and a DVD drive. The cost of such solutions is quickly dropping, and is being driven by economies of scale for both the playback hardware and the DVD drives, as well as higher scales of integration for the DVD playback hardware. By the autumn of 1997, similar solutions were available for less than $400, and the price of such solutions is expected to fall below $200 by the autumn of 1998.

4.1 Host-Based DVD Playback

While hardware add-in solutions are available at added cost, the computational power of the Pentium® II Processor permits full quality DVD playback to be performed as a software application. This form of playback is referred to as *host-based DVD playback*, and is available at no added system cost in 266MHz and higher Pentium® II processor based PCs. This is because host-based DVD playback does not require any additional hardware which would not otherwise already be present in the system.

In host-based DVD playback, the majority of the necessary DVD playback tasks are implemented in software - including the navigator and splitter, as well as a majority of the tasks which are associated with audio, subpicture, and video decode. Though the exact percentage varies between different software implementations, the video decode task typically comprises more than 75% of the computational load for host-based DVD playback. Because many of the tasks necessary for video decode coincide with the built-in capabilities of the graphics controller, host-based DVD playback implementations commonly offload portions of the video decode and rendering tasks to the graphics controller. For similar reasons, some tasks that are involved in the rendering of subpictures are commonly off-loaded to the graphics controller.

4.2 Full- quality of host-based DVD playback

The manner in which the video and subpicture decode tasks are partitioned between the Pentium® II Processor and the graphics subsystem is a key differentiator among various host-based DVD playback implementations. The effectiveness of a given partitioning scheme relies upon the features of the graphics controller, as well as characteristics of the DVD playback software.

While there are multiple partitioning schemes which permit host-based DVD playback, Intel does not advocate any single scheme. Intel instead advocates performance and quality targets for host-based DVD playback. These targets call for implementations which achieve full quality, and which do not require 100% utilization of the Pentium® II Processor. These requirements are in fact related.

The quality of DVD-Video playback is more commonly evaluated in qualitative rather than quantitative terms. Full quality DVD playback encompasses the following aspects:

- full frame rate
- synchronization between display of audio and video streams
- no visual artifacts
- no audio artifacts

- smooth frame delivery

Full frame playback requires that all encoded video frames be decoded and displayed. (Depending upon the content and video display device, full frame rate requires that either 24 or 30 frames per second be displayed.) Displayed frames should be presented in synchronization with the audio stream, which requires smooth frame delivery. Displayed frames should not contain any visual imperfections (*artifacts*), and should be displayed evenly through time. Finally, there should not be any audible artifacts - such as the clicks that can occur when the audio data is not rendered evenly across time.

The requirement that DVD-Video playback not consume 100% of the available Pentium® II Processor allows for spare CPU cycles, often referred to as *CPU headroom*. CPU overhead provides a quality buffer which can ensure that frames will not be dropped during spurious events (e.g., the PC receives a FAX while playing a movie).

The remainder of Section 4.2 describes the platform features which Intel has found to be useful in meeting these performance and quality objectives.

4.2.1 Processor Subsystem

Figure 3 illustrates the four principal components of the PC platform, including the processor, I/O (i.e., chipset, expansion busses), audio, and graphics subsystems. The processor subsystem is considered to encompass the processor, level 2 (L2) cache, and system memory.

The baseline system requirements for full quality host-based DVD playback include:

- Pentium® II Processor, with minimum of 266 MHz operating frequency
- 512K L2 cache
- 32 megabytes (MB) of SDRAM system memory

The Pentium® II Processor has several features that enable it meet the demanding computational requirements for host-based DVD playback. Many of these involve providing the necessary processor-to-memory bandwidth for the data-intensive video and audio decode tasks. The first such feature is the Dual Independent Bus (DIB) architecture. The DIB permits the L2 cache to operate at a frequency that is relative to the processor frequency, rather than at the (lower) frequency of the front side bus. The L2 cache is used to hold working set data for video and audio decode. The size of the L2 cache impacts the portion of the working set that can be contained, and thus the performance of host-based DVD playback.

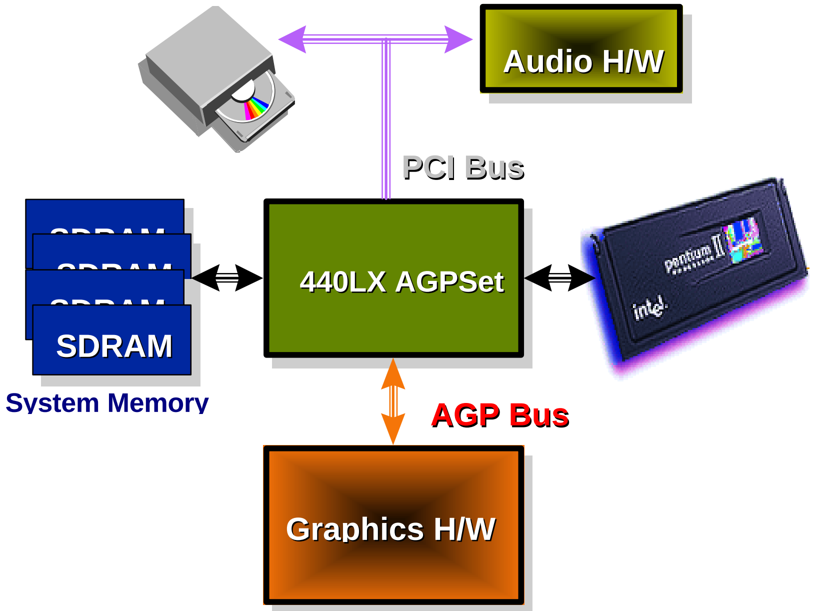

Figure 3. Subsystems involved in DVD-Video Playback

Host-based DVD playback also requires high bandwidth system memory, namely due to the bandwidth-intensive nature of the video stream. The main memory subsystem is commonly used for moving large chunks of data between the processor and I/O subsystem. SDRAM is an ideal memory type because of its low-latency reads for large blocks of data.

4.2.2 Graphics (Video) Subsystem

One of the key differences between various host based DVD implementations is the manner in which video decode is partitioned between the Pentium® II Processor and the graphics subsystem. There is typically a tight coupling between the DVD playback software and the features of the graphics controller. The performance and quality of a given partitioning scheme depends upon the features of the graphics controller, as well as the architecture of the DVD playback software. Similarly, the performance benefit of a particular graphics controller feature varies with the architecture of the DVD playback software. This section describes some possible features of the graphics subsystem that have been found useful in attaining the desired performance and quality objectives.

The following tasks are performed during MPEG-2 video decode and rendering, and can be performed by the graphics subsystem.

- Color-space conversion between YUV 4:2:0 and RGB formats
- Up/down scaling with bilinear interpolation
- Support for display of interlaced video content
- MPEG-2 motion compensation
- AGP bus mastering
- Sub-Picture rendering

It is important to note that the tasks listed here are based on a partitioning which has been found to be effective for a 266MHz Pentium® II processor. As processor speeds increase it will be possible to perform a greater number of the video decode tasks on the host CPU. Hence the above list is subject to change. Each task is explained below.

Other graphics controllers features that are known to benefit performance and quality include:

- Support for two or more MPEG-2 off-screen surfaces
- 4MB of graphics memory (SDRAM preferred)
- Auto-flipping between video frames

In addition, graphics systems which can render directly to televisions (i.e., have a *TV-out* capability) must also support the Macrovision™ analog copy protection system in order to comply with the CSS guidelines. (Chapter 5 provides further details concerning this requirement.)

The remainder of section 4.2.2 further describes these features.

4.2.2.1 YUV 4:2:0 Color Space Conversion

The pixel information in decoded MPEG-2 video frames is represented in YUV 4:2:0 format. Before these frames can be displayed on a standard PC monitor, they must be converted from YUV 4:2:0 to RGB format, which is the format native to PC monitors. Graphics controllers that offer hardware support for YUV 4:2:0 color space conversion are able to perform this conversion most efficiently.

If the graphics subsystem does not support YUV 4:2:0 color space conversion, the Pentium® II Processor is required to perform the conversion via software. This is accomplished by converting the YUV 4:2:0 format data either directly to RGB format, else to some other (e.g., YUV 4:2:2) format that the graphics controller is able to convert to RGB format.

4.2.2.2 MPEG-2 Motion Compensation

Motion compensation can be performed either in software on the Pentium® II Processor, or via dedicated hardware in the graphics controller. Performing motion compensation on the graphics hardware can off-load the CPU, thereby increasing CPU headroom. Because MPEG-2 motion compensation operates on video frames in their native YUV 4:2:0 format, graphics controllers that support hardware motion compensation typically also support YUV 4:2:0 color space conversion in order to realize the full benefit of hardware support for MPEG-2 motion compensation.

4.2.2.3 AGP Bus Mastering

The graphics controller should perform AGP bus mastering in order to manage the transfer of all video data from system memory to graphics frame buffer memory. Doing so off-loads the Pentium® II Processor of this task, thereby increasing overall performance and quality.

4.2.2.4 Up/down Scaling with Bilinear Interpolation

DVD-Video content often exists in an aspect ratio (ratio of screen width to screen height) that is different from that of a standard PC monitor. (A standard PC monitor has an aspect ratio of 4:3, whereas movie content often has an aspect ratio of 16:9.) High quality MPEG-2 video decode requires that the displayed video image be scaled in both the horizontal and vertical directions. Some graphics controllers support this capability, which is known as *up-scaling* and *down-scaling*.

Several different scaling algorithms exist. The most basic up-scaling technique determines the colors of new pixels by simply duplicating the color of existing pixels. Higher quality images result from *bilinear interpolation*, an alternate method that determines the colors of the pixels in the scaled screen by interpolating between the colors of the closest neighboring pixels in the unscaled screen. There exist additional techniques that result in even higher visual quality. Such techniques perform a weighted average of a greater number of neighboring pixels.

Up-scaling and down-scaling are generally regarded as too computationally intensive to perform in software on a 266 MHz Pentium® II Processor. In systems which use a graphics controller that does not offer hardware support for up and down-scaling, a slightly distorted image can result when displaying video content whose aspect ratio differs from that of the display device.

4.2.2.5 Support for Display of Interlaced Video Content

DVD-Video content can exist in either of two forms: interlaced or non-interlaced (also known as *progressive scan*). Interlaced video consists of two temporally distinct fields, where a field contains either all the even or all odd rows of pixels. Non-interlaced frames each contain all rows of pixels.

Displaying interlaced video content in a manner that is free of visual artifacts requires that the graphics controller contain specific functionality. The basic method for displaying interlaced content involves merging the pixel rows which are absent in the previous field using adjusted (interpolated) image data. In this manner, all rows of pixels are updated with each frame refresh.

Displaying non-interlaced content is typically considered too expensive to perform in software. Thus, video quality will significantly degrade when playing back interlaced content on graphics subsystems that do not offer hardware support for display of interlaced content.

4.2.2.6 Sufficient Graphics Memory

The graphics controller must have sufficient memory to hold at least two (off-screen) frames, in addition to the frame that is currently being displayed. This requirement provides a frame buffer that is intended to ensure the smooth delivery of decoded video frames. This buffer compensates for variations in video decode requirements - including differences in computational complexity of various frames, as well as the time availability of needed information.

In systems where the off screen surfaces are stored in YUV4:2:0 format, this requirement typically translates into 4 MB of graphics memory. This amount of memory is sufficient to contain

- three frames in YUV4:2:0 format at a resolution of 720x480 pixels (or 720x576 pixels)
- one frame in RGB format

In addition, off screen buffers may be required to store decoded subtitle data. Using this method, this information can then be merged with the primary surface or one of the off screen video buffers.

4.2.2.7 Sub-Picture Rendering

The subtitle and video streams are typically decoded into separate buffers, and then combined to create the current frame. The Pentium® II processor is able to combine the corresponding sub-picture and video information via software, at some associated overhead to the processor. The sub-picture data is more efficiently merged directly by hardware on the graphics controller. This can be accomplished using alpha-blending, or chroma-keying. These two methods both work for DVD playback, and are described below.

Alpha-blending associates an extra multi-bit field with every pixel in the video image. This bit field tells the graphics hardware how to blend the video image with the sub-picture information to create the final image. As an example, a value of 0.25 in the alpha field of a video image would tell the hardware to mix (0.25 * video pixel color value) with ((1-0.25) * subtitle pixel color value to result in the final pixel color value.

Chroma-keying uses a similar technique, except there is no notion of *blending*. A pixel color value is set up to be the *key* value to determine whether to display the current video image pixel or the sub-picture pixel. Typically, one of the 4 possible bit sequences in the 2bpp sub-picture frame is predetermined to be *transparent*. When the sub-picture is *transparent* the video image pixel is displayed. Otherwise the pixel from the sub-picture is displayed.

4.2.2.8 Auto-flipping between video frames

Graphics controllers that offer an auto-flipping feature permit video frames to be sequenced independent of the Pentium® II Processor. Systems which lack this feature require the Pentium® II Processor to poll the graphics subsystem, thereby degrading system performance and quality.

4.2.3 Audio Subsystem

The audio data on DVD Movie discs typically contains multiple channels of audio information, such as 5.1 channel Dolby AC-3, where each channel contains 48,000 audio samples per second (Ks/s). Since most current PCs typically support two channels of audio output, the audio data must be downmixed to two channels prior to output. The Pentium® II Processor performs this task during audio decode. The main requirement for the audio subsystem is that the audio device drivers support playback of 48 Ks/s content.

Section 4.2.4 describes that Type F DMA must be enabled when the system uses an ISA based audio card for DVD-Video playback.

In the future, PC platforms are expected to support multi-channel digital output using either the Universal Serial Bus (USB) or 1394 formats. In the 1998/1999 timeframe, many platforms are expected to migrate from ISA-based audio devices to PCI-based audio devices. Intel's AC'97 specification describes the architectural requirements for achieving high-quality stereo audio output as well as migrating from ISA-based audio devices to PCI audio.

Table 2. I/O transfers performed during DVD-Video playback.

I/O Task	Method	Bus Master
Transfer DVD-ROM data into system memory	IDE bus master	IDE device
Transfer decompressed MPEG-2 video data into graphics memory for motion compensation	AGP Master (YUV4:2:0 planar)	AGPgraphics device
Transfer decompressed MPEG-2 video data into graphics memory (no motion compensation)	AGP Master	Processor
Transfer decompressed audio data to PCI audio device	PCI bus master	PCI audio device (PCI audio)
Transfer audio data over ISA bus	Type F DMA or other ISA buffering scheme	PCI ISA Controller (legacy audio only)

4.2.4 I/O Subsystem

Performing I/O tasks efficiently is critical to achieving optimal performance for any application. The exact method in which I/O transfers are performed is a function of both the platform chipset and the particular I/O device, so care must be taken when choosing drives and peripherals.

Table 2 lists the I/O transfers that are performed during playback of DVD-Video content, along with the most efficient method of performing each transfer. The remainder of Section 4.2.4 describes these in further detail.

4.2.4.1 IDE Bus Mastering

The DVD-ROM drive should act as bus master when transferring data to system memory. A significant performance degradation will result if the Pentium® II Processor is forced to manage this transfer.

Windows95 OSR 2.x versions and beyond automatically install the required drivers. In the near future, UltraDMA/33 will be the preferred transfer mechanism for IDE DVD-ROM drives.

4.2.4.2 Audio Transfers

In order to off-load the Pentium® II Processor, the PCI audio device should act as the PCI bus master when transferring audio data from system memory to the audio device.

Legacy ISA audio transfers should be performed using a suitable ISA buffer scheme such as Type F DMA. Type F DMA buffers ISA transfers in the chipset, minimizing the PCI utilization for this task. The audio device drivers must proactively enable Type F DMA transfers, since the operating system does not automatically enable them.

4.2.5 DVD ROM DRIVE

The DVD-ROM drive should support the following:

- IDE Interface - Provides compatibility with the Intel 440LX AGPset.
- DMA Mode 0, 1, and 2 (Bus Master DMA, eventually moving to UltraDMA/33) - bus mastering interface requirements which are defined within the IDE interface specification.
- Must be able to read ISO9660 file format, which is currently the standard format for DVD disks. The UDF Physical file format will eventually be required.
- Must support reading at a 1.3 MB/s sustained transfer rate, and a 12 MB/s burst transfer rate. These rates are considered necessary to playback current DVD-Video content.

5. Legal Requirements

DVD-Video content is typically protected according to the Content Scrambling System (CSS), in order to prevent home piracy of movies on DVD. CSS relies on key-based encryption and conditional access to stored content. This section overviews the CSS, and the requirements to which host-based DVD playback implementation must adhere in order to comply with the CSS license. Readers may obtain a more complete overview from another Intel document, titled "*Copy Protection Licensing Requirements for the CSS DVD Method*" (available from http://developer.intel.com/solutions/tech/dvd.htm).

5.1 CSS Overview

CSS includes two principal measures to ensure protection against illegal copying. First, CSS guards access to the content which is stored on DVD-Video disks. CSS establishes guidelines which must be met in order for a DVD-Video playback device to obtain content information from a DVD disc drive. PC-based DVD playback mechanisms require an extra authentication step to create a secure, encrypted pathway for the DVD content. Before the DVD drive will release content to the rest of the system, the DVD playback mechanism must initiate and successfully complete a challenge authorization sequence with the DVD drive. Successfully completing this sequence requires that the DVD player know the keys and algorithms which are associated with the CSS. The result of a successful Authentication is a time-varying key that is shared between the DVD drive and the DVD software player. This key is later required to gain access to the content that is stored on the DVD disk.

Secondly, CSS scrambles the content on the disk, as described in Section 3.4. To unscramble CSS content, a DVD player must know the CSS secret keys and algorithms.

This requires that the DVD player have a CSS License. To receive a CSS License, a manufacturer must commit that their product will adhere to a set of design guidelines. These guidelines generally require that the DVD playback devices not be designed (or be easily modifiable) to allow illegal duplication.

5.2 CSS License Requirements

As a condition for receiving access to the CSS secret keys and algorithms, a manufacturer must agree to adhere to certain design rules. These design rules may be classified into three categories: output requirements, playback control, and architectural requirements. Each requirement is designed with the intent that CSS-protected content not be unreasonably exposed to illegal copying. These requirements are discussed in the following sections.

5.2.1 Output Requirements

CSS content may only be present on the output of devices where the output is authorized, and any required copy-protection technology is applied to the output signal. The CSS guidelines currently require that an authorized analog copy protection system be applied to most analog outputs (NTSC, YUV, PAL, SECAM). The only such authorized system currently available is from Macrovision Corporation.

At the time of writing the CSS guidelines do not permit descrambled CSS video on digital outputs, such as IEEE 1394, Universal Serial Bus, and RS232. The organization that controls the CSS License guidelines is currently working towards agreement on a copy protection method for such digital outputs. The CSS guidelines currently permit digital video to be output to a computer monitor (e.g., SVGA) without being copy-protected; however, it is expected that there will eventually be a copy protection requirement for digital video outputs.

The CSS guidelines currently do not require digital audio outputs to be copy protected; however, there are restrictions on digital audio transmission formats. These require that transmitted digital audio be descrambled, in a compressed format - or using Linear PCM format at no more thatn 48 Ks/sec and no more than 16 bits per sample, and that the SCMS flags not actively be removed.

5.2.2 Playback Control

The CSS guidelines specify three basic requirements for playback control:

- A DVD drive may not release CSS data to the playback system unless authentication has been performed.
- A DVD drive may not play back CSS content from a user recordable disk.
- A DVD player may not play a movie from a region other than the region for which that player has been set.

The first requirement is enforced by the drive itself, and is intended to guard against a software hack. The second requirement aims to reduce the threat of pirated copies of CSS content being distributed on readily available DVD media. The final requirement is referred to as regionalization.

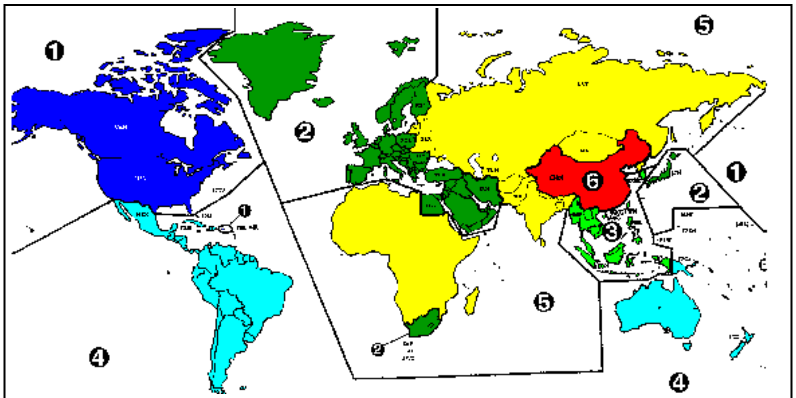

Figure 4: CSS Regionalization Map

Regionalization governs in which geographic regions content can be played, and thereby permits content providers to target products at certain geographics regions. Regionalization is partly intended to address the content provider's fears that DVD would cut into revenues for non-U.S. theaters, as movies are often released on video (and ultimately on DVD) in the U.S. before being distributed to theaters outside of the U.S.. Regionalization divides the globe into six regions, according to Figure 4. Content creators are able to specify in which region(s) their content can be played. While content can be encoded for playback in one or more regions, a DVD drive must be identified as belonging to a single region. For example, regionalization prevents a region 1 (U.S. and Canada) player from playing a disk which has only been released to region 2 (Japan and Europe). Regionalization also serves as a firebreak for pirated content, as content which is pirated from one geographic region cannot necessarily be played back on players outside of that region.

5.2.3 Architectural Requirements

The architectural requirements are designed to limit the misuse of CSS licensed products. These rules are designed to make CSS licensed products resistant to alteration or tampering, or circumvention by devices that may be easily designed. These rules include:

- *No support for "drag and drop" file copying.* This requirement mandates that CSS data not be made available for recording. This limitation prohibits activities such as copying a movie from a DVD to a hard drive.
- *Analog copy protection on internal video signals.* This requirement is intended to prevent manufacturers from circumventing the copy protection plan by making a dual playback/duplication integrated device.
- *Software tamper resistance.* When CSS descrambling is implemented in software, the software must be *tamper resistant*. In general terms, *tamper resistance* means that only a professional using advanced lab equipment and specialized training would be able to defeat the software by exposing the secret CSS keys or algorithms, or redirecting descrambled data to something other than an approved decoder. A good measure of sufficient protection is when the CSS descrambling software is harder to break than the CSS algorithms and keys themselves. Software developers interested in developing software implementations of CSS components should base their designs on a strong cryptographic background and a close reading of the language of the CSS License. Readers may refer to *Tamper Resistant Software: An Implementation* to learn more about one implementation of tamper resistance.
- *Restrictions on data which may be present on various buses.* The CSS guidelines mandate what data formats may be present on what buses, for the purpose of

protecting data on user-accessible buses. The CSS license defines user accessible buses to include the PCI bus, and any card bus that is designed for end-user upgrades of components. The CPU bus and memory buses are not considered to be user accessible buses. In general, the easier it is to use a copied data format, the less accessible data of that format must be. Table 3 provides specific examples. Of note from Table 3 is the fact that the only legal way to transmit data which is both descrambled and compressed over a user-accessible bus is to first re-encrypt the data before transmitting it. The rule that permits transmitting descrambled and decoded CSS data over a user-accessible bus permits fully decompressed video data or for similarly high-bandwidth motion compensation data, to be sent to a graphics card for rendering.

Table 3. Data allowed on user accessible busses

Data Format	*Allowed On User-Accessible Bus?*	*Explanation*
CSS scrambled content	Yes	Scrambling prevents unauthorized playback.
MPEG-2 encoded descrambled video data	No	Low bandwidth: possible to copy and store elsewhere.
MPEG-2 decoded descrambled CSS data	Yes	High bandwidth: data is too voluminous to store, and almost impossible to re-encode in real-time.

5.3 License Mechanics

At the time of writing, there is a fee of 1,000,000 Japanese Yen (approximately US$10,000) required to obtain a CSS License. It is expected that a multi-industry entity will eventually be formed to coordinate and administer CSS licensing issues. In the meantime, the Matsushita Electric Industrial Company has agreed to act as the “interim licensing entity”. For further details of the procedure involved in obtaining a CSS License, readers are encouraged to refer to the Intel document titled “*Copy Protection Licensing Requirements for the CSS DVD Method*” (which can be located on http://developer.intel.com/solutions/tech/dvd.htm).

6. Summary

This document overviews topics which are relevant to DVD-Video playback on the PC platform. Additional information on the status of DVD and related products can be found on the world wide web.