Technical Brief

NVIDIA Video Processing Engine

The Ultimate Digital Entertainment Experience

Video Processing Engine

With DVD technology already reaching over 32 million[1] U.S. households, consumers continue to expect more from the "ideal" digital video experience. The NVIDIA® video processing engine (VPE), an integral part of NVIDIA's next-generation desktop GeForce4™ family of graphics processing units (GPUs), delivers the ultimate "Digital Entertainment" experience on today's desktop PCs. From DVD playback on a high-definition TV (HDTV) to personal video recorders (PVRs) that allow you to time-shift live television broadcasts, NVIDIA's VPE integrates the necessary hardware to deliver compelling, feature-rich video experiences on PCs that don't require a top-of-the-line CPU.

Offloading the CPU as much as possible is key to providing end users with the best video experience. By designing the VPE to handle the majority of the video decoding and playback tasks, NVIDIA satisfies user's wishes to simultaneously run multiple applications—3D gaming, Web browsing, or traditional office applications—without overloading the CPU and encountering impaired performance.

However, driving the highest-quality video output to any device—be it a PC monitor, TV, or HDTV set—is only part of the solution. In addition to offloading the CPU as much as possible, the VPE integrates more MPEG2 decode algorithm than any previous desktop GPU. The VPE also takes DVD playback to the next level by offering an advanced adaptive de-interlacing engine, 5 horizontal x 3 vertical taps filtering, digital vibrance control, 1024x768 TV output, as well as HD and progressive-scan DVD output.

[1] Source: Gartner Dataquest

Hardware MPEG2 Decode

The VPE has a full-hardware MPEG-2 decoding engine that puts into silicon the technology to process and decode MPEG-2 video streams. However, playing DVD content also involves navigation of the DVD, parsing of the data streams, and decode and playback of the video and audio streams.

Navigation refers to the graphical interface that allows a user to play, pause, fast-forward, reverse, select chapters, view angles, select language, and use parental controls. Several DVD navigation applications are available and may become part of the standard desktop user interface in the near future.

The **parsing** function includes decrypting data from the DVD disc and separating video data from audio data. The audio data is typically decoded by the CPU and represents about a 5 percent overhead for most current CPUs. The Content Scrambling System (CSS) decryption also represents overhead of around 5 percent on current CPUs.

NVIDIA's VPE handles video decoding using hardware that implements the industry's standard **MPEG2 decode algorithm** (Figure 1), which includes inverse quantization (IQ), inverse discrete cosine transform (IDCT), motion compensation, color space conversion (CSC) functions, as well as hardware subpicture alpha blending.

The entire video decode process can consume up to 45 percent of today's CPU overhead if handled entirely in software. NVIDIA's VPE however, performs these functions in hardware, using only one-tenth of the same CPU processing power.

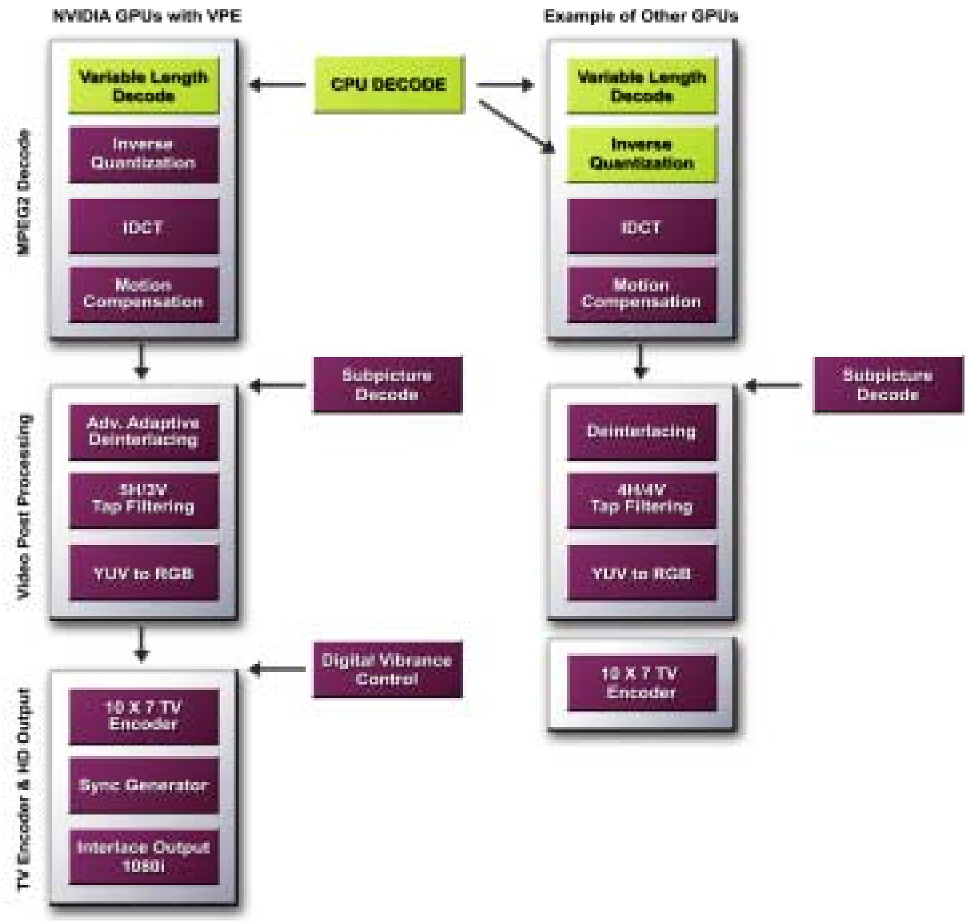

Figure 1. **NVIDIA GPUs with VPE (vs. other GPUs)**

Advanced Adaptive De-interlacing

The second significant piece of the VPE is the advanced, adaptive, de-interlacing engine, which is ideal for processing and displaying interlaced content on the typically progressive PC monitor.

Analog TV, 1080i HD, and some DVD content originating from rendered computer graphics or music videos, are all interlaced. Interlaced displays, or content, render one field using the odd scan lines, and render the next field using the even scan lines. This “every other line” mode of processing is a form of compression, allowing more data and higher resolutions to be displayed in the same bandwidth. De-interlacing is required to display interlaced content on a progressive display, such as a computer monitor where interlaced fields (odd and even scan lines) come in at a rate of 60 fields per second and have to be converted to progressive frames at 30 frames per second.

The two most simple and common methods of de-interlacing are *weave* and *bob*.

- **Weave** combines the two adjacent fields to generate a frame. While this works with still images, it creates annoying feathering (combing) artifacts with motion video.
- **Bob** is the process of zooming in on each field (enlarged by a factor of two) in the vertical direction, displaying them successively, and shifting the bottom field by one scan line. The bob method works well—even with motion video—but suffers from a shimmering artifact due to a loss in picture resolution. This is most obvious with text and stationary objects/logos overlaid on top of a live video.

Adaptive de-interlace looks at the fields on a pixel-by-pixel basis and decides whether to bob or weave, based on motion for that pixel. The problem is the determination of motion. Typically, only luma (image brightness) is used for the determination. Most implementations today look at the surrounding pixels to determine if they are in rank order. If so, it considers that as no motion. Others look at motion vectors and attempt to make decisions in that fashion. Both methods are complex and can make mistakes unless a great deal of hardware is applied.

NVIDIA’s implementation of per-pixel adaptive de-interlacing leverages the power of its powerful 3D GPU engine, resulting in extremely accurate and superior video quality over any competitive solution.

Advanced 5 Horizontal x 3 Vertical Taps Scaling & Filtering

Scaling and filtering are necessary to convert one resolution of content into a different resolution of display. HDTV content, for example, can be output in a 1920 x 1080 resolution, but many computer desktops are only capable of showing resolutions up to 1280 x 1024 or perhaps 1600 x 1200. It is critical to provide extremely smooth scaling of the video and to convert the millions of possible colors from a single frame into a high-fidelity image. To achieve this, the VPE incorporates a 5-tap horizontal x 3-tap vertical scaling engine that takes input from any of the DVD or HD formats, and accurately scales them to the current display resolution with smoothing in both the vertical and horizontal domains. Just as antialiasing is important to 3D visual quality, VPE's smooth scaling can reduce the jaggies when users are watching video content.

Independent Hardware Color Enhancements and Digital Vibrance Control

Content created for TV often looks dark on computer displays because of the different gamma characteristics between computer displays and TVs. The VPE has a separate hardware control for video gamma, which allows a user to lighten video content without washing out the colors for their 2D and 3D computer applications.

In addition to video overlay color controls, NVIDIA's exclusive Digital Vibrance Control (DVC) allows users to modify the display characteristics of any display and enhance the visual appearance to match any lighting environment. DVC can be set independently for each attached display to make images look crisp, clean, and bright, regardless of where the images are output.

HD Component Output

To output to HDTV sets and TVs with 480i component or 480p progressive scan inputs, video must be converted from the traditional RGB format to YPrPb (a broadcast-specific display format). NVIDIA is the first graphics company to integrate a 1024x768 TV encoder, along with hardware, to support HD component outputs.

Note: Standard composite video combines luminance and chrominance information into one signal, causing visual artifacts such as color smearing and fuzziness. S-video produces sharper images by using two signals for luminance and chrominance. However, S-video only has the bandwidth to handle low-resolution analog signals. Component video uses one signal for luminance (brightness) and two signals for chrominance (color) to retain maximum bandwidth and superior visual quality.

To support YPrPb output, three main hardware pieces are required:

- Master sync generator to control the sync levels.
- Interlacer to output 480i and 1080i interlaced modes. Most graphics processors today support only progressive output.
- TV encoder, which operates in digital-to-analog converter (DAC) mode with tri-level sync.

NVIDIA's VPE includes the first two key enablers to support HD component or YPrPb output. To ship a graphics board with an integrated HD component output, the board manufacturer can simply populate one of the approved Conexant or Philips' TV encoders with tri-level sync support and replace the S-video 4-pin connector with a 9-pin connector (the same one that is used with NVIDIA Personal Cinema). This 9-pin-to-component output cable provides the components a consumer needs to connect their NVIDIA-based graphics card to an HDTV or TV set with component input.

With VPE capability, consumers with TV sets that support either 480i component or 480p progressive-scan component input can enjoy watching superior quality DVD playback without buying a more expensive HDTV set.

Note: Competitive solutions require end users to purchase a $20-$40 DVI-component adapter, which may be difficult to find, and may not be fully tested to support all component output formats.

Accelerating PC PVR Applications

PVRs, such as Tivo or ReplayTV, allow end users to record live TV shows, while offering advanced features such as time-shifting, instant replays, commercial bypass, and interactive programming guides. PVRs come with a significant price tag, however, with prices ranging from $200 to $2000 for the hardware, plus additional monthly fees for programming information.

Time-shifting playback heavily taxes a PC. It requires the CPU to encode (compress) the incoming video into the MPEG2 format, store it to the hard drive, and simultaneously decode and play back the same video. With NVIDIA's VPE, however, half of that process is handled by the VPE's dedicated hardware, which offloads those functions from the CPU. The VPE handles all the decode and video post-processing in hardware, allowing the CPU to fully concentrate on encoding high-quality video at the full resolution (see Figure 2).

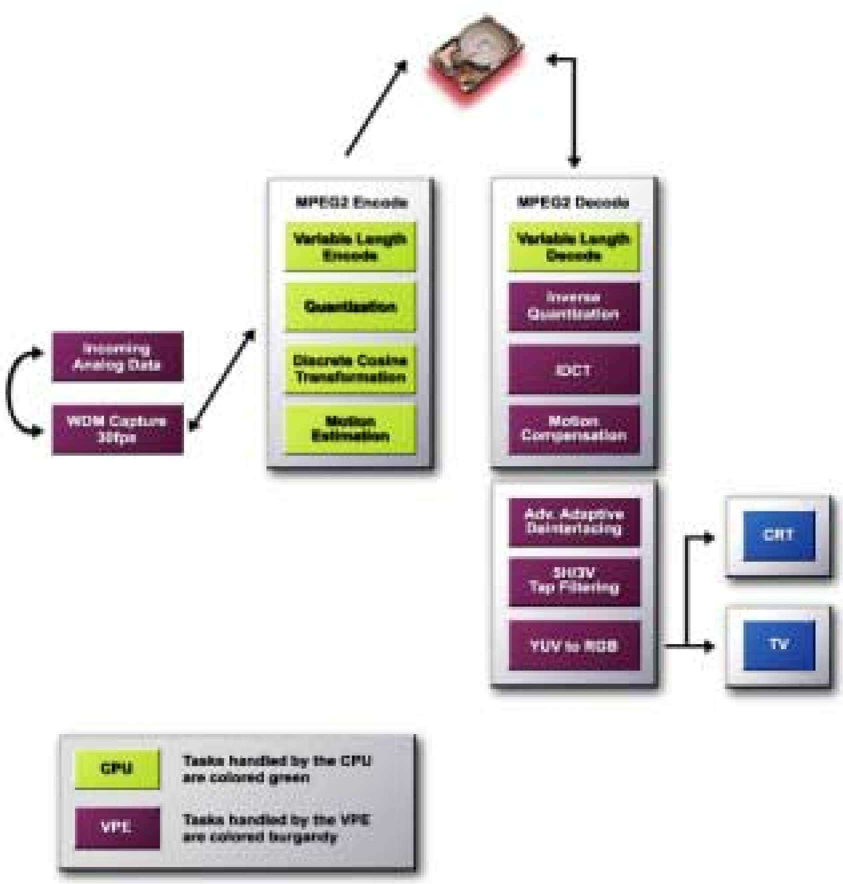

Figure 2. **Time-Shifting: Simultaneous MPEG2 Encode and Decode**

To replace a consumer PVR, the PC-based PVR solution must integrate full remote-control support for all the PVR's advanced features, including pausing, instantly replaying a scene, commercial bypassing, and automatic programming of future recordings.

When NVIDIA shipped the NVIDIA® Personal Cinema™ solution in September 2001, it delivered the first PC-based PVR and digital entertainment solution with a wireless multifunction remote. In addition to controlling PVR's advanced features, the Personal Cinema remote also replaces the consumer DVD remote by integrating the most commonly used DVD functions. From its simple hardware setup to its intuitive software, Personal Cinema was the first mainstream product to combine exceptional video functionality with NVIDIA's award-winning 3D performance.

Future Personal Cinema products with the NVIDIA VPE will make the PVR experience even more compelling by delivering the highest-visual quality, ease of use, and lowest cost.

Accelerating the PC HDTV Transition

With 221 Digital Television (DTV) stations in 78 markets reaching over 70 percent of all U.S. households, and entry-level HDTV sets dropping to the $2000-$4000 range, many broadcasters are motivated to simulcast in high definition during primetime and throughout the day. By May 2003, broadcasters are required to broadcast at least 50 percent of the content in digital format[2].

NVIDIA's VPE further accelerates the HDTV transition by integrating hardware, MPEG2 decoding, advanced scaling, de-interlacing, and HD component output. When coupled with a low-cost DTV receiver, VPE turns any desktop PC into a fully featured HDTV machine that receives, decodes displays, and outputs at the native HD resolution.

The VPE's advanced scaler can downscale HD resolution (720p or 1080i) to standard definition resolution (720 x 480) while retaining the highest-quality image. This capability allows the current 250-plus million analog TVs to display HD content without upgrading to a more expensive HDTV set.

Figure 3 on the next page shows a possible HDTV implementation with the NVIDIA VPE. A low- cost PCI DTV receiver receives the HD broadcast and transfers the transport stream to a software HDTV decoder. The software splits the stream into audio and video streams. The VPE handles most of the HD MPEG2 decoding, processes the decoded content and outputs it to the monitor or TV set via its integrated TV encoder, or HDTV set through YPrPb (HD component) using a supported tri-level sync encoder.

[2] Source: FCC

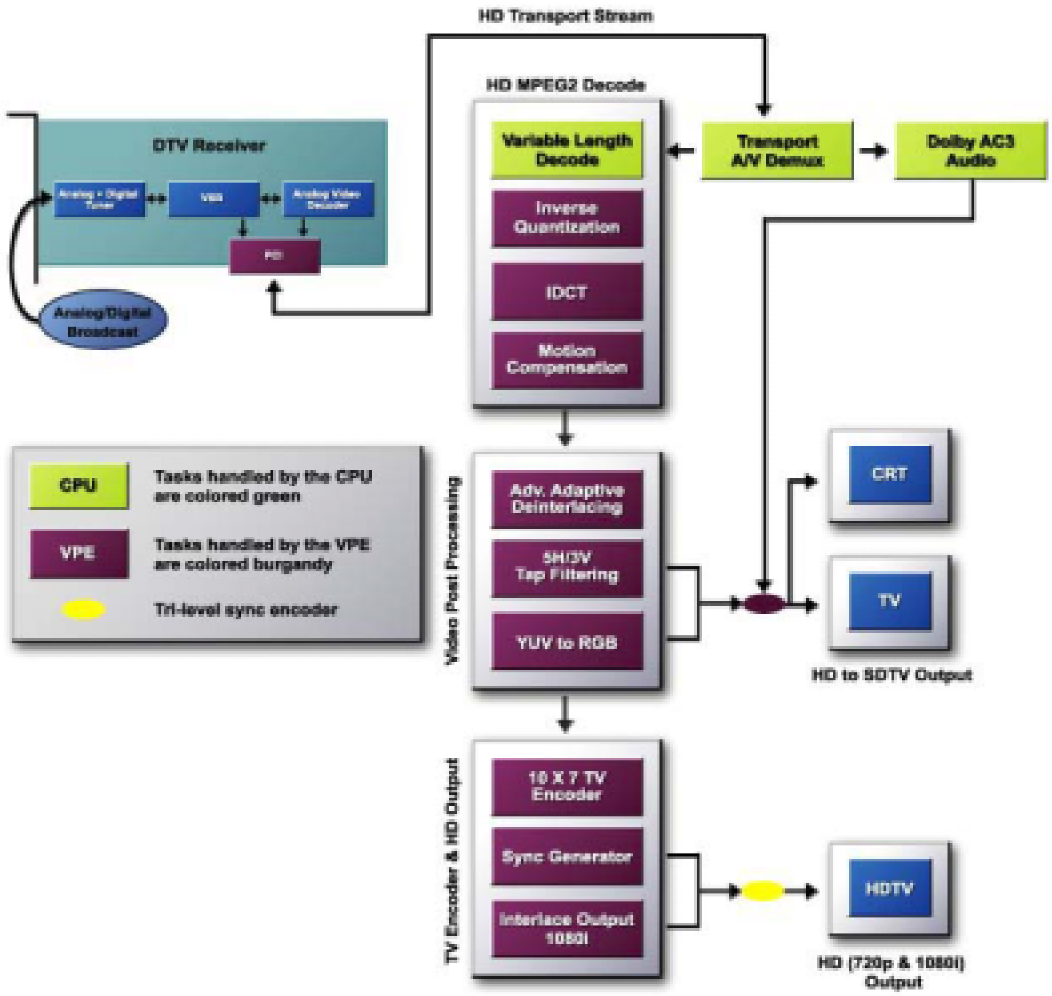

Figure 3. **HDTV Application Flow using NVIDIA VPE**

Note: Low-cost DTV receiver and software HDTV decoders are expected in the market in Q2 2002 as stand-alone products. Initially, a complete solution may ship in a fully configured system with a GeForce4 GPU-based board, a DTV receiver card, and the optimized software HDTV decoder. NVIDIA is working with these HDTV solution providers. More information will be provided when the product is readily available.

HDTV Benefits and Details

HDTV leverages a high-resolution digital television (DTV) format capable of reproducing a 16:9 aspect ratio and Dolby Digital sound. HDTV products can reproduce 720 and 1080 resolution (progressive and interlace) and receive all 18 possible digital TV (ATSC) broadcast formats.

These capabilities result in the clearest picture possible today, with minimal scan lines, less flicker, and greater depth of field (see Figure 4).

Analog TV – 210,000 pixels **HDTV – 2.1 million pixels**

Figure 4. **Pixels in Analog TV vs. HDTV**

HDTV will also introduce different viewing dimensions (see Figure 5). Today's television programs are designed to fit TV screens with a width-to-height ratio of about 4:3, which can also be expressed as 16:12. By comparison, HDTV programming will be broadcast in a ratio similar to movie theater screens: 16:9. As a result, broadcast HDTV movies will not have to be reformatted or cropped when transmitted to our homes. Sports viewers will also enjoy the 16:9 format because it will allow an expanded view of the field and action.

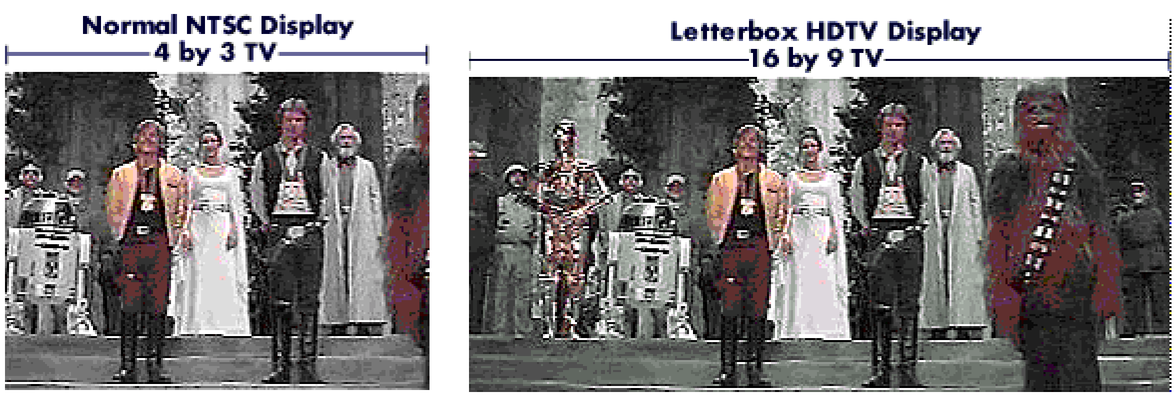

Figure 5. **Comparison of Viewing Dimensions**

Conclusion

The combination of NVIDIA's award-winning GPUs and the technical innovation and integration of the Video Processing Engine pave the way for dynamic, visually superior digital entertainment experiences. From DVD playback to the time-shifting and recording of live TV, HDTV, and advanced 3D gaming, NVIDIA is the first company to make the concept of the "Entertainment PC" a reality for millions of PC users worldwide.

Appendix A: ATSC Formats

This appendix contains the formats specified by the Advanced Television Systems Committee (ATSC) standards. All formats listed in the tables are supported by the NVIDIA VPE. Included in this appendix is the standard-definition TV (SDTV) modes and the high-definition modes for HDTV.

Table 1. ATSC Formats

Size		Aspect Ratio		60 pfs	30 pfs	30 p[fs	24 pfs	NVIDIA
Horizontal	Vertical	16:9	4:3	Progressive	Interlace	Progressive	Progressive	VPE
1920	1080	Yes	No	No	Yes	Yes	Yes	Yees
1280	720	Yes	No	Yes	No	Yes	Yes	Yes
720	480	Yes	Yes	Yes	Yes	Yes	Yes	Yes
640	480	No	Yes	Yes	Yes	Yes	Yes	Yes

Table 2. Common Abbreviations

Acronym	Meaning
i, p	Interlace Progression
480i	640 x 480 Interlace
480p	640 x 480 Progressive
720p	1280 x 720 Progressive
1080i	1920 x 1080 Interlace
1080p	1920 x 1080 Progressive

Table 3. Typical Usage of the Different Formats

Formats	Typical Usage Model
1080i30	High definition live action/sports event
1080p24	High definition film-oriented content
720p60	High definition live action/sports event
720p24	High definition film-oriented content
480p60	Standard definition live action/sports event
480i30	Standard definition content

Glossary

Bit Depth

The bit depth refers to the number of bits of precision for the color and z-values associated with each pixel on the screen. More bits of precision improve the visual realism and accuracy of the rendered frame. The two most common bit depths in modern graphics hardware are 16-bit and 32-bit. Each of these values can be associated with color or Z-values. Color that is 32 bit (for example) typically is used to represent red, green, blue and alpha (or transparency) values with up to 8 bits per component, or 256 "values" for each of those components. A 32-bit z-value is typically allocated as 24 bits of Z precision (or depth precision) and 8 bits of stencil or "mask" precision.

Depth Complexity

Depth complexity is a measure of the complexity of a scene. It refers to the number of times any given pixel must be rendered before the frame is done. For example, a rendered image of a wall has a depth complexity of one. An image of a person standing in front of a wall has a depth complexity of two. An image of a dog behind the person but in front of the wall has a depth complexity of three, and so on. As depth complexity increases, more rendering horsepower and bandwidth is needed to render each pixel or scene. The average depth complexity of today's graphics applications is two to three, meaning that for every pixel you end up seeing, it gets rendered two or three times by the graphics processor.

Fill Rate

Fill rate is the rate at which pixels are drawn into the screen memory. Fill rate is a common measure used to illustrate the pixel processing capabilities of today's 3D graphics processors. Fill rate is usually measured in millions of pixels/sec. (Mpixels/sec.) In 1997, 50-70 Mpixels/sec. was considered state of the art. In 2002, the leading 3D graphics processors will be capable of more than 1200 Mpixels/sec. While this improvement is an incredible achievement, it is still barely enough to create a compelling 3D environment. Rendering pixels at such a high rate consumes enormous amounts of memory bandwidth.

Frames per Second

Frames per second (fps), or frame rate, refers to how many times per second the scene is updated by the graphics processor. Higher frame rates yield smoother, more realistic animation. It is generally accepted that 30fps provides an acceptable level of animation, but increasing the performance to 60fps results in significantly improved interaction and realism. Beyond 75fps it is difficult to detect any performance improvement. Displaying images faster than the refresh rate of the monitor results in wasted graphics computing power, because the monitor is unable to update its phosphors (or display) that fast, wasting frame rate beyond its refresh rate.

Memory Bandwidth

Memory bandwidth refers to the rate at which data is transferred between the graphics processor and graphics memory. Memory bandwidth limitations are one of the key bottlenecks that must be overcome to deliver truly realistic 3D environments. To deliver truly stunning 3D requires high-resolution, 32-bit color depth at high frame rates, with rich geometry, sophisticated texture mapping, and complex vertex and pixel shading.

Resolution

Resolution is the number of pixels on a screen. Higher resolutions can create a more realistic 3D environment because more scene detail can be displayed. Most modern displays are capable of at least 1280 horizontal pixels x 1024 vertical pixels, while many larger or more expensive displays are capable of 2048x1536 pixels. Most graphics applications support a variety of resolutions, allowing the end user to run at higher resolutions (and hence higher level of detail) with the trade-off being increased load on the graphics processing system.

Texture Mapping

Texture mapping is the technique of projecting a 2D image (typically a bitmap) onto a 3D object. Texture mapping allows substantial increases in visual detail without significant increases in polygon count. Because of the improved realism that can be obtained with a very small increase in computational cost, texture mapping is one of the most common techniques for displaying realistic 3D objects. In order to render a texture-mapped pixel, the texture data for that pixel needs to be read into the graphics processor, consuming memory bandwidth.

NVIDIA Corporation
2701 San Tomas Expressway
Santa Clara, CA 95050
www.nvidia.com