Technical Brief

The GeForce 6 Series of GPUs:
High Performance and Quality
for Complex Image Effects

Richer, More Vibrant Images in the GeForce 6 Series

The NVIDIA® GeForce™ 6 Series of graphics processing units (GPUs) pushes high-quality imaging to higher levels of performance and precision, enabling developers to create more stunning real-time effects. These next-generation GPUs introduce an innovative superscalar architecture that supports more operations per cycle, eliminating trade-offs between quality and speed as it raises the standard for achievable image effects. With full 32-bit floating point support through the entire pipeline, the GeForce 6 Series GPUs power cinematic-quality images with full 128-bit color. Programmers can work in the more memory-efficient 16-bit format, or easily switch to full precision when the action or scene calls for the cleanest, highest-impact effects.

The new NVIDIA High-Precision Dynamic-Range (HPDR) technology, part of the GeForce 6 Series architecture and based on the OpenEXR standard by Industrial Light & Magic (http://www.openexr.com/), further improves static and moving image quality. With NVIDIA HPDR, motion is smooth and texture detail increases. The GeForce 6 Series products also include a new rotated-grid antialiasing system that helps polygon edges by supporting more effective subpixel coverage values. The result is a more accurate pixel color representation, giving polygon edges crisp, clear, definition.

This paper provides an overview of the NVIDIA GeForce 6 Series architecture, its advanced image quality, and examples of the effects and techniques it enables.

Superscalar Design

The NVIDIA GeForce 6 Series introduces an innovative shader architecture that can double the number of operations executed per cycle (Figures 1 and 2). Two shading units per pixel deliver a twofold increase in pixel operations in any given cycle. This increased performance enables a host of complex computations and pixel operations. The result is stunning visual effects and a new level of image sophistication within fast-moving bleeding-edge games and other real-time interactive applications.

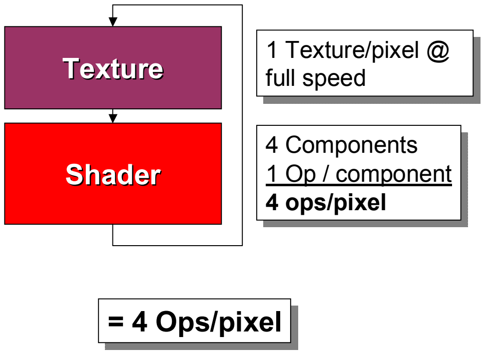

Figure 1. Traditional nonscalar shader architectures provide one shader unit and only process up to four operations per cycle.

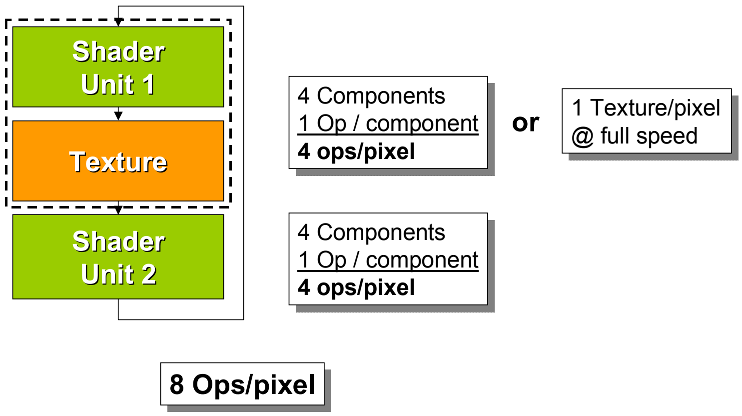

Figure 2. Each GeForce 6 Series GPU features a superscalar architecture, with a second shader unit, to double pixel operations per cycle.

With two shader units, the GeForce 6 Series architecture supports true dual processing—two instructions executing in the same cycle on different shader units. Some architects try to elevate a single-shader nonscalar design by claiming support for two instructions in the same cycle. The difference between these approaches is significant. On single-shader architectures, only two instructions execute on the same shader unit (Figure 3), and the instructions operate on components of the same word or pixel. The architecture for the GeForce 6 Series, however, provides more total throughput for mathematical computations carried out on pixel components. During each cycle, the dual shader units can execute up to four instructions per cycle and up to eight operations per pixel.

Note: "Instructions" are the commands, delivered to hardware, which can operate on multiple components of a pixel and require multiple operations. "Operations" are the mathematical functions performed to carry out an instruction.

In addition to improving throughput, the GeForce 6 Series architecture increases programming flexibility. Pixel components can be operated on individually, or in groups of two, three, or four components per operation. This ability to define groupings introduces many new programming techniques and speeds up the implementation of complex mathematical operations that create next-generation effects.

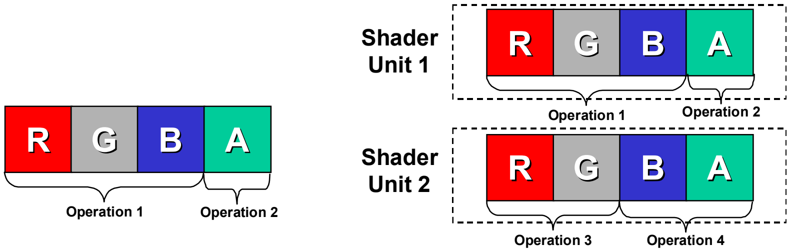

Figure 3. Traditional architectures (left) vs. NVIDIA superscalar architecture (right).

To summarize, the superscalar architecture is capable of up to four instructions and eight operations per pixel, compared to just two instructions and four operations in traditional architectures. In Figure 3, notice how traditional architectures (left side of figure) can only support two instructions per pixel per cycle, with limited grouping of the pixel components for each operation. On the other hand, NVIDIA's superscalar architecture (right side of figure) can execute four instructions per pixel per cycle, with complete flexibility for grouping the pixel components.

32-Bit Native

The GeForce architecture has always enabled developers to choose the necessary level of precision for each image or scene. Now the choice is simpler, because the performance degradations associated with full 32-bit floating-point precision have been eliminated.

Developers can still use 16-bit mode when memory space efficiency is a priority, but now they can achieve higher-quality images in a broader range of real-time situations. Users will appreciate and enjoy a richer, more vibrant graphics experience, and game developers will be able to set themselves apart from their competitors.

NVIDIA HPDR Technology

With NVIDIA HPDR technology, graphics will continue to make strides towards more photorealistic rendering. This type of high dynamic-range rendering lets users experience realistic effects and environments, such as the intense brightness of the sunlight, or the rich color of the dark images illustrated in Figure 4.

Image courtesy of Paul Debevec

Figure 4. Example of high dynamic-range lighting. The range of white light is very bright, and the detail in the dark marble slabs at the bottom of the image is preserved.

In the past, many barriers limited this type of rendering. Computer frame buffers had linear display scales for various color intensities. In addition, standard 32-bit per-pixel frame buffers only gave 255 values to represent the entire range of color—not enough to represent dynamic imagery. So, developers had to develop creative solutions to deliver these types of graphics—including using expensive conversions (such as RGBE) in the pixel shader; avoiding incompatible techniques that would have been desirable (such as dynamic lighting); and emulating functionality that was missing from graphics hardware (like high-precision texture filtering).

The GeForce 6 Series architecture features the NVIDIA HPDR technology, which meets the needs of the incredibly powerful, high dynamic-range rendering technique. NVIDIA HPDR delivers greater data precision during shading, blending, and filtering operations, improving both static and dynamic images.

With NVIDIA HPDR, high dynamic-range rendering is implemented with efficiency and beauty. The rendering task is broken into three phases—light transport, tone mapping, and color and gamma correction.

Light Transport

Light transport takes the geometry, texture maps, light positions, and light radiances and outputs a high-dynamic per-pixel radiance value for the reflected light, based on the surfaces that are visible. This value can be anything. This information must be stored in a frame buffer with enough precision and range to properly represent the vast range of intensity values of the light. The larger the range that can be maintained, and the higher precision that data can be stored in, the more accurate the visual result.

NVIDIA HPDR follows the same method of encoding as the OpenEXR standard, SM10e5. This means 1 sign bit, 10 bits of mantissa, and 5 bits of exponent are used to represent light transport information.

Depending on the values of *e* and *m*, one can see the range of this powerful format.

$x=(-1)^s 2^{e-15}\left[1+\frac{m}{1024}\right]$	$x=(-1)^s 2^{-14}\left[\frac{m}{1024}\right]$	$x=(-1)^s \infty$	$x = NaN$
e > 0	e= 0	e=31, m=0	e=31,M>0

> In this example the dynamic range = 12 dB, with the largest value 65504 and the smallest value 2^{-24}.

This amount of range and precision is needed to accurately capture the vast range and precision of light transport data.

Tone Mapping

The output data of the light transport phase is then mapped to color values and operates on the 2D image. This process is known as tone mapping.

Color and Gamma Correction

The final stage, known as color and gamma correction, maps these color values from a standard "color space"—where red, green, and blue are defined in a given way—to the monitor's red, green, and blue "color space." In addition, a gamma correction is applied so that the logarithmic differences in color intensity calculated in the beginning of this rendering process make it to the final display device.

The human eye responds to light logarithmically. In fact, the human eye is more sensitive to lower intensities of light, seeing darker shades with more detail than higher intensity or brighter light. This final stage maps the data to the monitor while maintaining the proper visual effect.

GPU Requirements for NVIDIA HPDR Rendering

To execute this type of rendering approach, the GPU needs to be capable of floating point shading, blending, filtering, and texturing. Lastly, it must be able to store colors so that the logarithmic nature of the data can be preserved.

Floating Point Shading

As mentioned earlier, 32-bit shading is the native mode of operation for the GeForce 6 Series GPUs. Shading operations can be performed at maximum speed while still maintaining maximum precision. Effects like physically correct lighting, iridescence, and subsurface scattering are all rendered magnificently at uncompromised speed.

Floating Point Blending

The blending operation combines previously rendered pixels with the newly calculated fragment value that exists at a given location. Depending on the effect being rendered, the values will be mixed to get a final color value. The higher the precision, the more accurate and higher the quality of the blended pixel.

Effects that take advantage of floating point blending include motion blur and soft shadows, as well as accumulation that results from multiple dynamic light calculations in a scene.

Floating Point Filtering

The filtering operation filters pixels to sharpen an object or smooth edges in a scene. Filtering improves visuals during motion, as with bilinear and trilinear filtering. It can also sharpen image quality when pixels depict an object from an extreme viewing angle (this is known as anisotropic filtering). Other floating point filtering effects include tone mapping and glow, which are required for high dynamic-range rendering.

In addition, NVIDIA GeForce 6 Series GPUs support a higher level of precision and 16× anisotropic filtering.

Floating Point Texturing

The texturing operation applies a texture to a given polygon. The ability to use floating point textures allows for unique effects such as omnidirectional shadow maps, depth of field, and ray tracing.

The image on the left in Figure 5 was taken without high dynamic range. There is only a 100:1 difference in the light source intensities. The result is a blown-out look on the window and the floor lighting. The image on the right was taken with high dynamic range. There is more than a 9000:1 difference in light source intensities. Note the subtle lighting variations on the floor and on the nature scene on the right.

Image courtesy of Microsoft

Figure 5. Image taken without high dynamic range (left) vs. with high dynamic range (right).

Storing Colors

The ability to map color values during the gamma correction phase is a crucial part of the high dynamic-range rendering technique. In order to preserve the wide range of values, a logarithmic format of some type must be adopted.

sRGB, which is an 8-bit gamma color space, is the standard for the Microsoft® Windows® operating system. sRGB is a low-storage-cost solution that matches CRTs and is implemented in GeForce hardware. However, sRGB is not enough by itself. While sRGB does give a logarithmic representation of data, it does not have enough range and precision to accurately represent data calculated during the light transport phase of rendering.

In Table 1, note the difference in range between sRGB and OpenEXR. OpenEXR provides a larger range for calculations such as light transport. However, for any type of storage and mapping (like those used during tone mapping or color correction phases of high dynamic-range rendering), sRGB is an intelligent choice.

Table 1. Color Ranges

	Range	**Precision**	**Storage***	**Notes**
RGBE	76.8 dB	9-bit log	189.8 Mb	Radiance- compressed 32-bit float
32-Bit TIFF	76.8 dB	24-bit log	759.4 Mb	IEEE-754 32-bit floating point
OpenEXR	12.0 dB	11-bit log	379.7 Mb	ILM-developed 16-bit floating point
e-sRGB 12	4.6 dB	12-bit poly	213.6 Mb	Clamped at [–0.53..1.68]
16-Bit int	4.8 dB	16-bit linear	379.7 Mb	Clamped at [0..1]
sRGB	3.5 dB	8-bit poly	189.8 Mb	Clamped at [0..1]
RGBA	2.4 dB	8-bit linear	189.8 Mb	Clamped at [0..1]

***Note**: The storage information is based on a single 1600 × 1200 resolution frame of 1080p ATSC video.

In addition, state-of-the-art games use a technique called "dynamic lighting," where the dynamic range and reflection data for each light source is calculated separately and then added together in a buffer. Unfortunately, sRGB values cannot be added together. To do this, the values would need to be converted, added, and then converted back to the sRGB format. The result is a compromise in performance. Refusing to convert to another format would result in unsightly artifacts.

The NVIDIA HPDR technology solves the problem of high dynamic-range rendering. It provides studio-quality 16-bit floating point formats for storage, blending, shading, texturing, and filtering during the light transport phase. Plus, it allows the use of the efficient sRGB format in the tone mapping and color and gamma correction phases.

> *"Accurately representing the huge range of colors and lighting seen in the real world has always been a challenge for computer graphics. Now that NVIDIA has complete support for floating point textures, floating point blending, and sRGB gamma correction, accurate color and lighting reproduction in high dynamic range rendering is easily implemented."*
>
> **Herb Marselas, CEO/Director of Technology**
> **Emogence, LLC**

And best of all, the NVIDIA HPDR technology is implemented in hardware. There is no pixel shader encode or decode to deal with. Furthermore, it is already exposed in Microsoft DirectX® 9.0 and OpenGL® APIs.

Rotated-Grid Antialiasing

The newest generation of NVIDIA GeForce GPUs introduces a rotated-grid antialiasing sampling algorithm. Based on four samples per pixel, the new scheme maintains industry-leading performance while significantly increasing color accuracy. Previously, the four subpixels were sampled in a two-by-two grid pattern for each pixel. By slightly rotating the pattern of the four subpixels, the new antialiasing scheme provides sampling from a four-by-four diamond-shaped grid. In Figure 6, notice how the GeForce 6 Series subpixel pattern (right) has been rotated to a diamond shape.

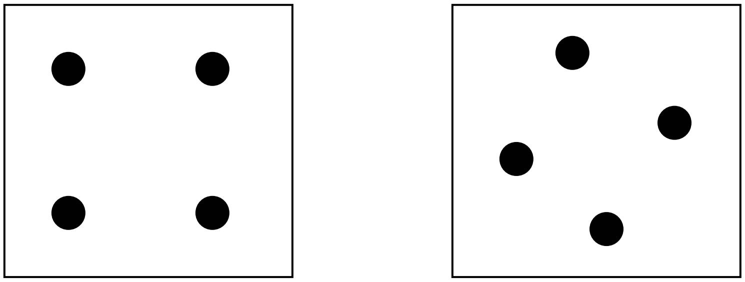

Figure 6. A GeForce FX pixel (left) and a GeForce 6 Series subpixel pattern (right).

The rotated-grid configuration allows superior subpixel coverage in horizontal and vertical dimensions. In Figure 7, notice that the GeForce FX architecture provides coverage for two vertical values and two horizontal values, but the GeForce 6 Series coverage spans four values for the horizontal and vertical subpixel positions. The increased coverage produces higher color accuracy at the edges of polygons.

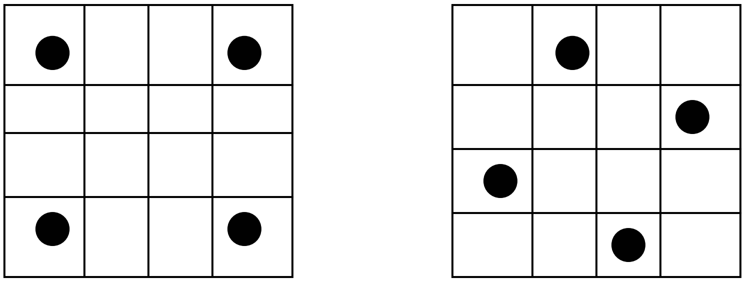

Figure 7. Pixel patterns for GeForce FX (left) and GeForce 6 Series (right) architectures showing horizontal and vertical values.

For a more detailed explanation of antialiasing and sampling techniques, please refer to the NVIDIA technical brief, "NVIDIA Accuview Technology: High-Resolution Antialiasing Subsystem (TB-00311-001)" at www.nvidia.com.

The New Era of Effects

The superscalar architecture for the new GeForce 6 Series GPUs, with its native 32-bit pipeline and imaging advancements, improves speed and precision across a broad range of imaging operations and effects. Many operations become practical for the first time in real-time applications and games—texture filtering, high dynamic-range effects, depth of field, blurs, and 16× anisotropic filtering—bringing life and cinematic realism to the PC (Figure 8).

© 2004 NVIDIA Corporation

Figure 8. The GeForce 6 Series delivers unmatched realism to leading-edge graphics applications.

Conclusion

The NVIDIA GeForce 6 Series brings unprecedented realism to the next generation of leading-edge graphics applications. Without compromising speed for quality, developers can implement stunning visual effects throughout complex scenes and digital worlds.

Revolutionary innovations like NVIDIA HPDR rendering allow higher visual quality and more unique effects by maintaining floating point precision in all aspects of rendering, including shading, texturing, filtering, and blending. Rotated-grid antialiasing adds to the overall image quality by providing more levels of coverage on polygon edges.

To summarize, the latest architecture provides an enhanced pixel pipeline (Table 2) and enables real-time floating point operations in the following areas:

- 2D graphics
- 2D textures with mipmaps
- Cube maps
- Volume maps
- Shading
- Texture filtering
- Blending
- Filtering

Table 2. GeForce 6 Series Architecture Characteristics

Architecture Characteristics of the GeForce 6 Series	
Pixel pipelines	16
Superscalar shader	Yes
Pixel shader operations/pixel	8
Pixel shader operations/clock	128
Pixel shader precision	32 bits
Single texture pixels/clock	16
Dual texture pixels/clock	8
Adaptive anisotropic filtering	Yes
Z-stencil pixels/clock	32

Soon, even experts will be doing double takes trying to discern computer-generated scenes from filmed sequences. With a superscalar architecture, a native 32-bit pipeline, and state-of-the-art imaging capabilities, the NVIDIA GeForce 6 Series is breaking through the final obstacles for achieving cinematic realism in virtual worlds.

Notice

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE.

Information furnished is believed to be accurate and reliable. However, NVIDIA Corporation assumes no responsibility for the consequences of use of such information or for any infringement of patents or other rights of third parties that may result from its use. No license is granted by implication or otherwise under any patent or patent rights of NVIDIA Corporation. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all information previously supplied. NVIDIA Corporation products are not authorized for use as critical components in life support devices or systems without express written approval of NVIDIA Corporation.

Trademarks

NVIDIA, the NVIDIA logo, and GeForce are trademarks or registered trademarks of NVIDIA Corporation. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2004 by NVIDIA Corporation. All rights reserved.

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