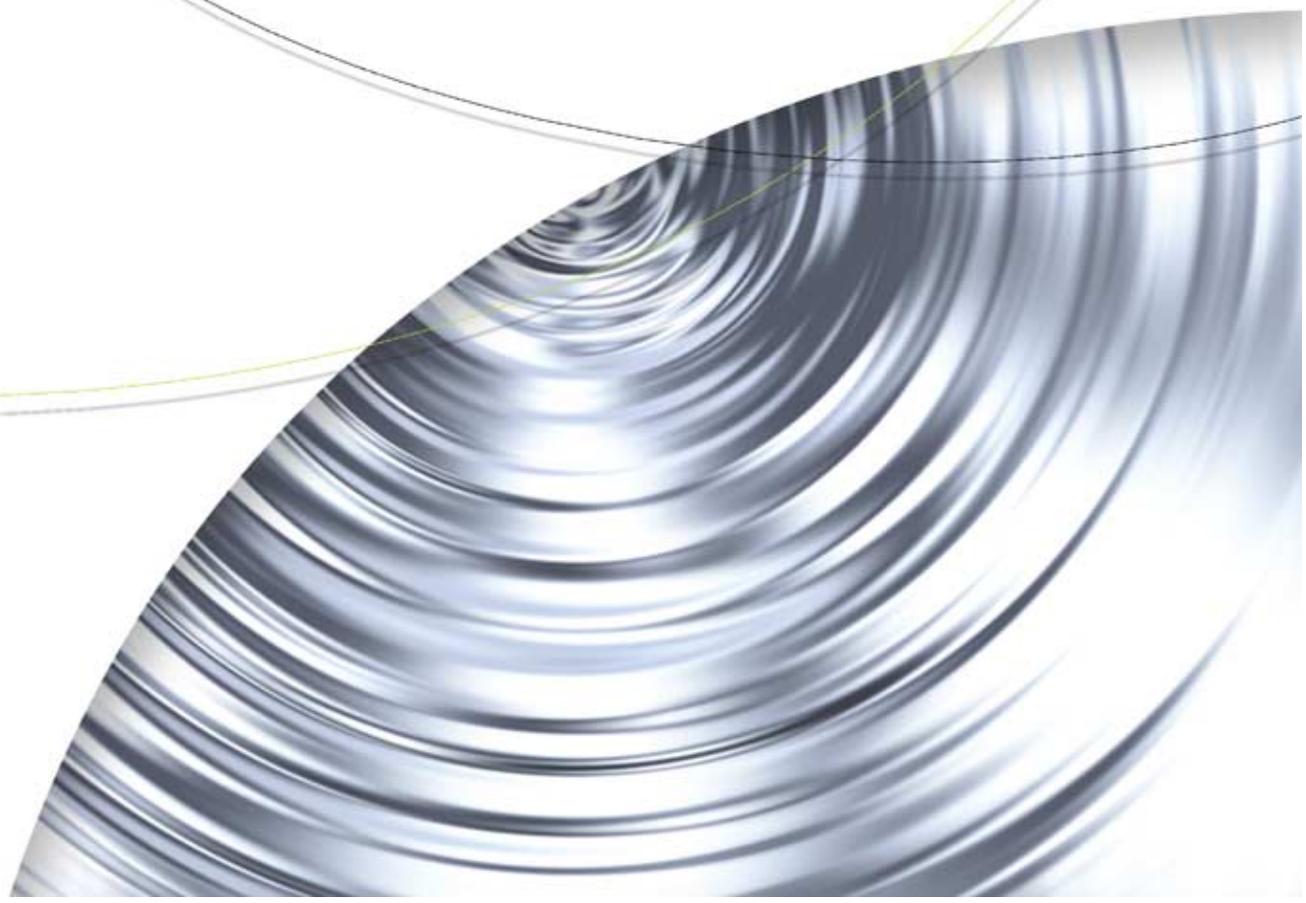




**NVIDIA.**

## Technical Brief

**NVIDIA nfiniteFX II Engine**  
From Research to Real Time



The top of the page features a background image of water ripples on the left side, with the word "NVIDIA" in a large, light, serif font centered across the top. A thin green vertical line runs down the right edge of the page.

# NVIDIA nfiniteFX II Engine

---

## Research and the Need for Processing Power

Computer graphics research has always pushed the boundaries of software and hardware, forcing us to ask the question: “If the technology were fast enough, what else could be done?”

Thankfully, as hardware and software technologies advance, some of these questions can now be answered. The GeForce4 family of Graphics Processing Units (GPUs) with dedicated high-speed memory pipelines are capable of calculating billions of operations per second, and enable the development of unbelievably complex effects through their powerful shading instruction sets. Through the use of their integrated dual vertex and advanced pixel shaders, the GeForce4 GPUs have unleashed ferocious graphics power on many of the computer graphics research topics currently being studied.

A great example of the power behind the GeForce4 GPUs are their inherent ability to solve one of the classic computer graphics problems of all time: how to render realistic hair and fur. Animals that have skin textures have traditionally been easier to render than fur. For example, the dinosaurs in *Jurassic Park* were easier to make lifelike and believable than the animals in the safari movie *Jumanji*. Trying to render hair and fur in real time has historically resulted in animals that looked plastic, stiff, and constrained. Computer-generated hair and fur should give the appearance of texture and movement, and it should show off the subtle reflections and light absorption that real hair or fur would in a given situation.

For the first time ever, and only through the power of NVIDIA’s nfiniteFX II engine, which includes support for dual vertex shaders, advanced pixel shader pipelines, 3D textures, shadow buffers and z-correct bump mapping, is it now possible to render, among other things, incredibly complex, high quality hair and fur while delivering very fast frame rates.

---

## A Little Bit of History

### California Institute of Technology's Fuzzy Bear

In the early days of research, scientists at the California Institute of Technology created a new approach for rendering hair and fur by using a 3D hair texture volume. This 3D texture would have a surface frame, and normal tangent and binormal 3D vectors, used to calculate lighting and reflections. The parameters of the lighting model would be freely distributed throughout the volume. This volume was then rendered in software by walking through the volume and accumulating the densities along the viewing vector (a ray-casting approach).



Figure 1. **Fuzzy Bear**

This approach produced the “Fuzzy Bear” picture, which became a reference of sorts in the computer graphics industry (see Figure 1). One of the interesting facts to realize is that when this research occurred in 1989, a single frame took over two hours to render, running on a network of large IBM mainframes, containing 12 3090 processors and four 3081 processors. And, while the bear was a graphics breakthrough for its time, the rendering time of two hours per frame was not acceptable for any real-time applications.

### Princeton and Microsoft's Furry Bunny

In early 2001, researchers at Princeton University and Microsoft Corporation published a paper on the subject of real time fur. The principal contribution of this paper was the concept of rendering the 3D fur volume by generating a series of concentric “shells”. These shells were created by scaling the base “skin” layer along the vertex normal. The textures for these shells would be created from samples of the 3D hair texture at different depths in the volume. These textures would then be applied to their corresponding concentric shell layers, and the concentric shells would then be blended together to produce the final appearance of hair or fur.



Figure 2: **Furry Bunny**

Lastly, in addition to the shell geometry, the researchers added the use of specially created fin geometry and textures to help enhance the object if viewed in silhouette. These fin textures are used to produce the wispy bits of hair and fur that add realism to the animal, especially when viewed in silhouette.

The result of their research became known as “Furry Bunny” (see Figure 2). This furry bunny, which pushed the technological envelope about a year ago, was rendered on a NVIDIA GeForce 256 Graphics Processing Unit (GPU).

The furry bunny contained 5,000 faces (or Polygons), 7,547 edges, 306 patches and ran between 12 and 23 frames per second (depending on the number of concentric shells that were turned on).

One of the most interesting parts of this research paper was the rendering discussion and future work sections. In these sections, the researchers discussed the current state of hair and fur rendering as well as directions of where the technology was going. The researchers specifically mentioned two key technologies:

**Programmable vertex shaders.** Within the coming year, commodity graphics hardware will have programmable vertex shaders. Such shaders will be ideal for accelerating shell rendering, since they can directly evaluate the geometric offsets between shells, in fact using only two additional shader instructions.

**Programmable pixel shaders.** Future programmable pixel shaders may be able to perform per-pixel lighting, which would be useful for wavy and curly hair patterns.

## The Future Is Here Now

### NVIDIA's Wolfman

The NVIDIA GeForce4 GPU family and nfiniteFX II engine represents the first time that realistic fur with per-pixel lighting can be applied to a highly complex animated character and run at high frame rates.

The nfiniteFX II engine's dual vertex shaders are able drive more than 100 million vertices per second. This power is needed as the Wolfman contains over 100,000 polygons, or more than 20 times the complexity of the "Furry Bunny" research project of a short year ago. Figure 3 below contains a wire mesh of the Wolfman before the concentric shells have been applied; the inset version is fully rendered.

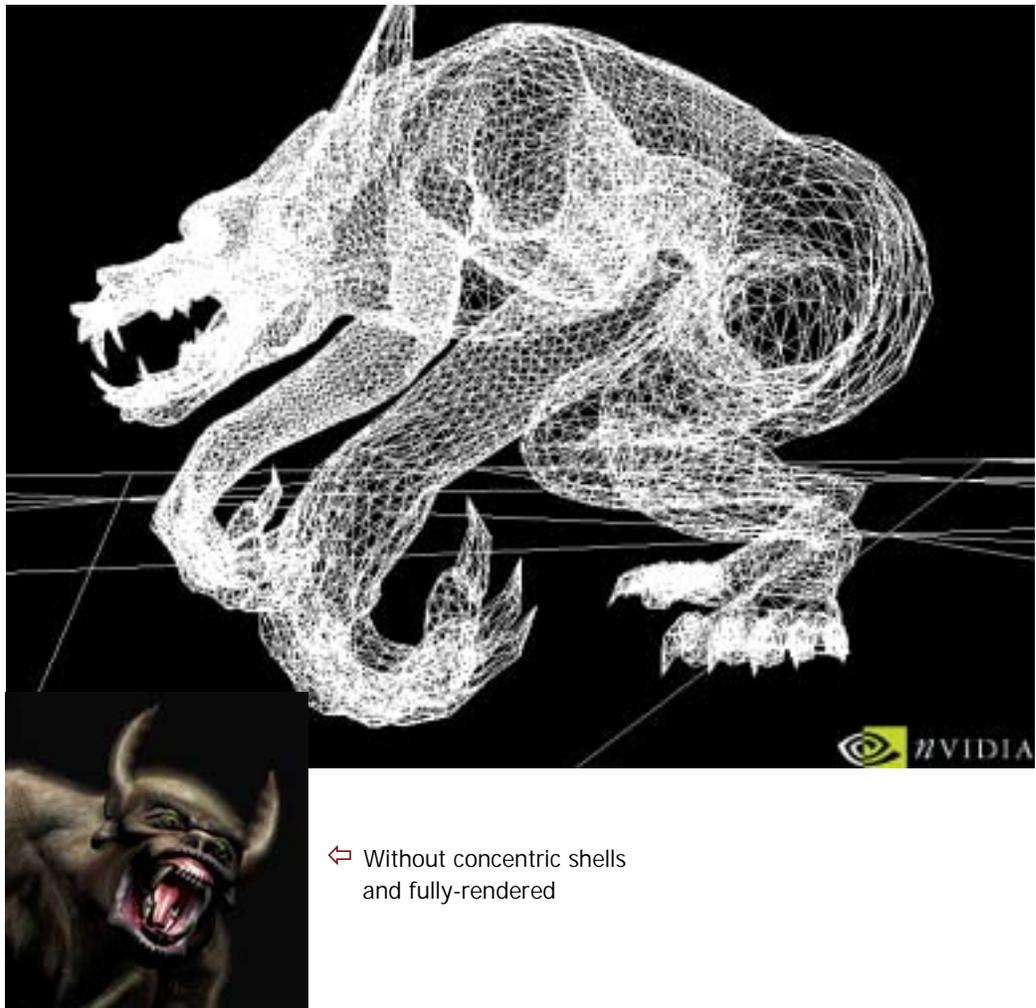
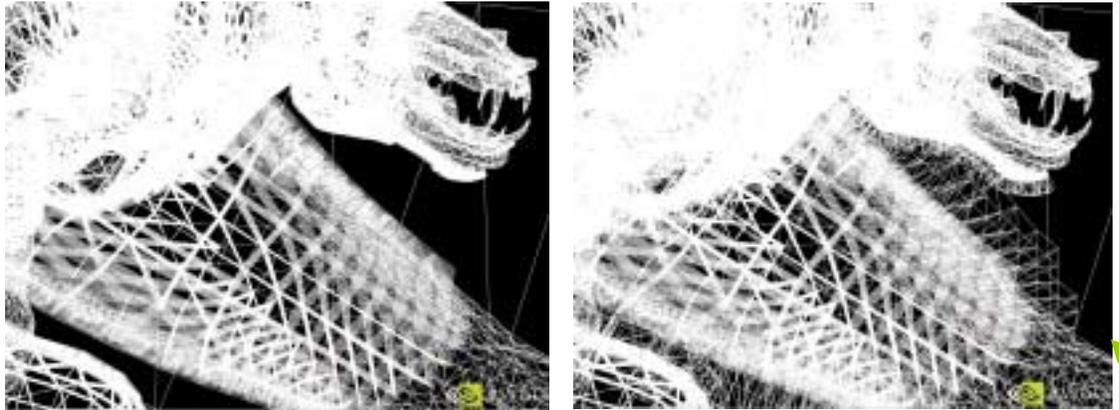


Figure 3. **Wireframe Model of NVIDIA's Wolfman**

The Wolfman uses eight concentric fur shells that can be seen in Figure 4. The color and density of the fur is controlled using a separate texture map that covers the entire body, which gives the fur its distinct look, rather than a uniform pattern (see Figure 7). The nfiniteFX II engine's advanced pixel shader support for 3 and 4 textures accelerates this type of rendering.



Notice the extra layers of geometry around the arm. This helps with the wispy hairs and improves the image when viewed in silhouette.

Figure 4. **With Concentric Shells and Fins**

The Wolfman is not a mere static model. Rather, it is a completely skinned animation. This Wolfman contains a 61-bone skeleton (see Figure 6). The complexity of this model is on par with that used in television and film special effects production. Each and every vertex of the skin, fur layers, and fin geometry are deformed in real-time to match the movement of the underlying skeleton. The complexity of this task is amazing, as the nfiniteFX II engine needs to handle these vertex deformations for each of the eight layers.

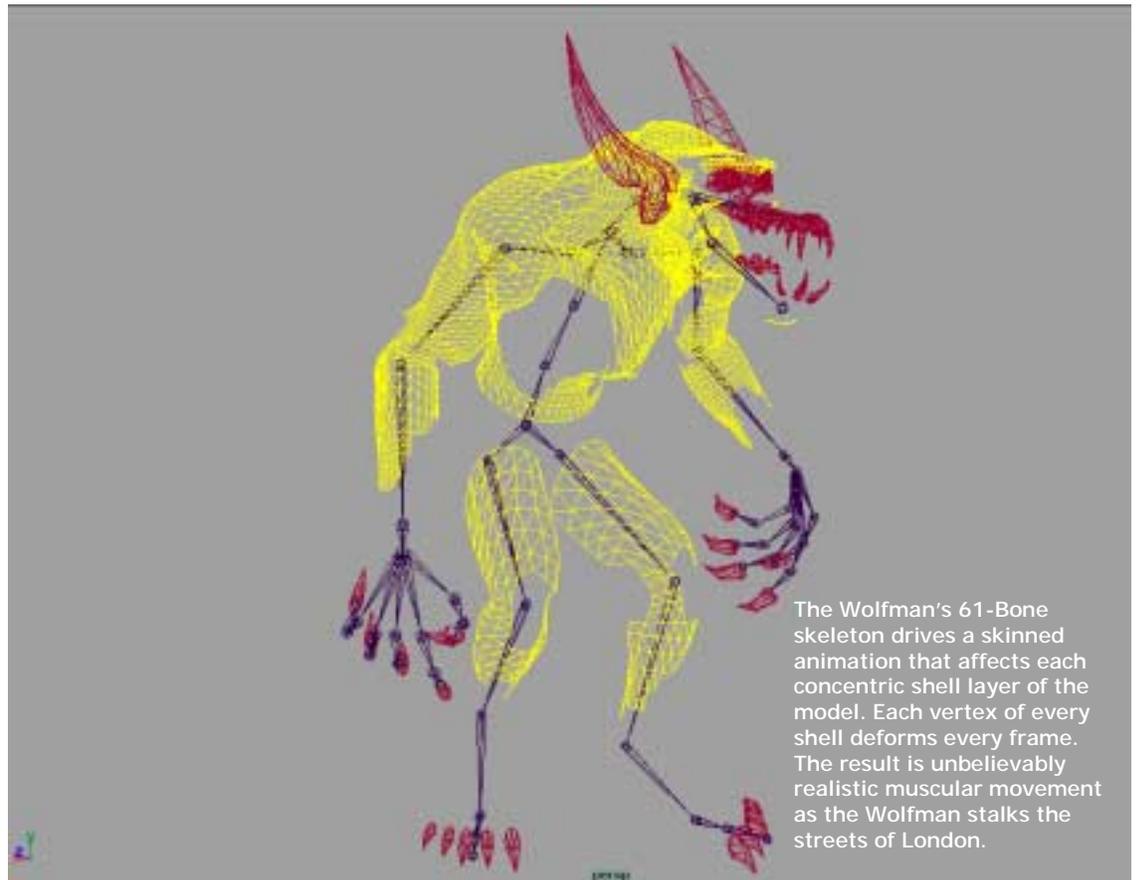


Figure 6: **The Wolfman's Bone Skeleton**

One of the unique properties of stranded material such as hair and fur is that it reflects light more in some directions than others. This is known as “anisotropic” lighting, and is computationally expensive to reproduce.

The nfiniteFX II engine has advanced pixel shaders that help the GeForce4 GPUs deliver 50% more performance than the GeForce3. These improvements allow the GPUs to deliver anisotropic lighting to the Wolfman while maintaining fast frame rates. Individual strands of hair and patches of fur react based on the position and the intensity of the light and the angle that the light strikes the fur. Figure 7 shows the fully rendered Wolfman.



Individual hairs react to the light based on orientation of the light and structure of hair. This occurs on a per-pixel basis.

Figure 7. **Rendered with Concentric Shells and Anisotropic Lighting**

Table 1 shows the content and performance of various fur projects:

Table 1. **Furry Creatures Comparison Chart**

	Processor	Polygons	Frame Rate
Fuzzy Bear	IBM 3090/3080	NA	2 Hrs/Frame
Fuzzy Bunny	GeForce 32M	5,000	12-23 fps
Furry Wolfman	GeForce4 Ti 4600	100,000 +	60+ fps

---

## Conclusion

As hardware and software technologies evolve, so do the research and remedies. In the case of rendering hair and fur, the nfiniteFX II engine provides a vehicle to approach the latest research and display the results in real time. In addition, these renderings occur at very high frame rates with stunning antialiased visuals.

With the nfiniteFX II engine, NVIDIA has once again raised the bar for high quality visuals, while simultaneously delivering ferocious graphics power. The nfiniteFX II engine's dual vertex shaders are capable of driving the complex type of geometries required for a skinned character like the Wolfman (100k+ polygons); and its advanced pixel shaders deliver blistering performance, up to 50% faster than the GeForce3. In addition, its increased performance in 3 and 4 texture situations is especially vital when dealing with multiple texture layers of fur, hair and light.

The nfiniteFX II engine has truly delivered on its promise of bringing movie quality graphics—once only topics of research papers—running at high speed, real-time frame rates to the PC.



To view a streaming movie of the Wolfman demo, visit:

[http://www.nvidia.com/view.asp?PAGE=power\\_demos](http://www.nvidia.com/view.asp?PAGE=power_demos)

---

## 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.



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, GeForce4, and the NVIDIA logo are trademarks of NVIDIA Corporation.

Other company and product names may be trademarks of the respective companies with which they are associated.

**Copyright**

Copyright NVIDIA Corporation 2002.



**NVIDIA.**

NVIDIA Corporation  
2701 San Tomas Expressway  
Santa Clara, CA 95050  
[www.nvidia.com](http://www.nvidia.com)