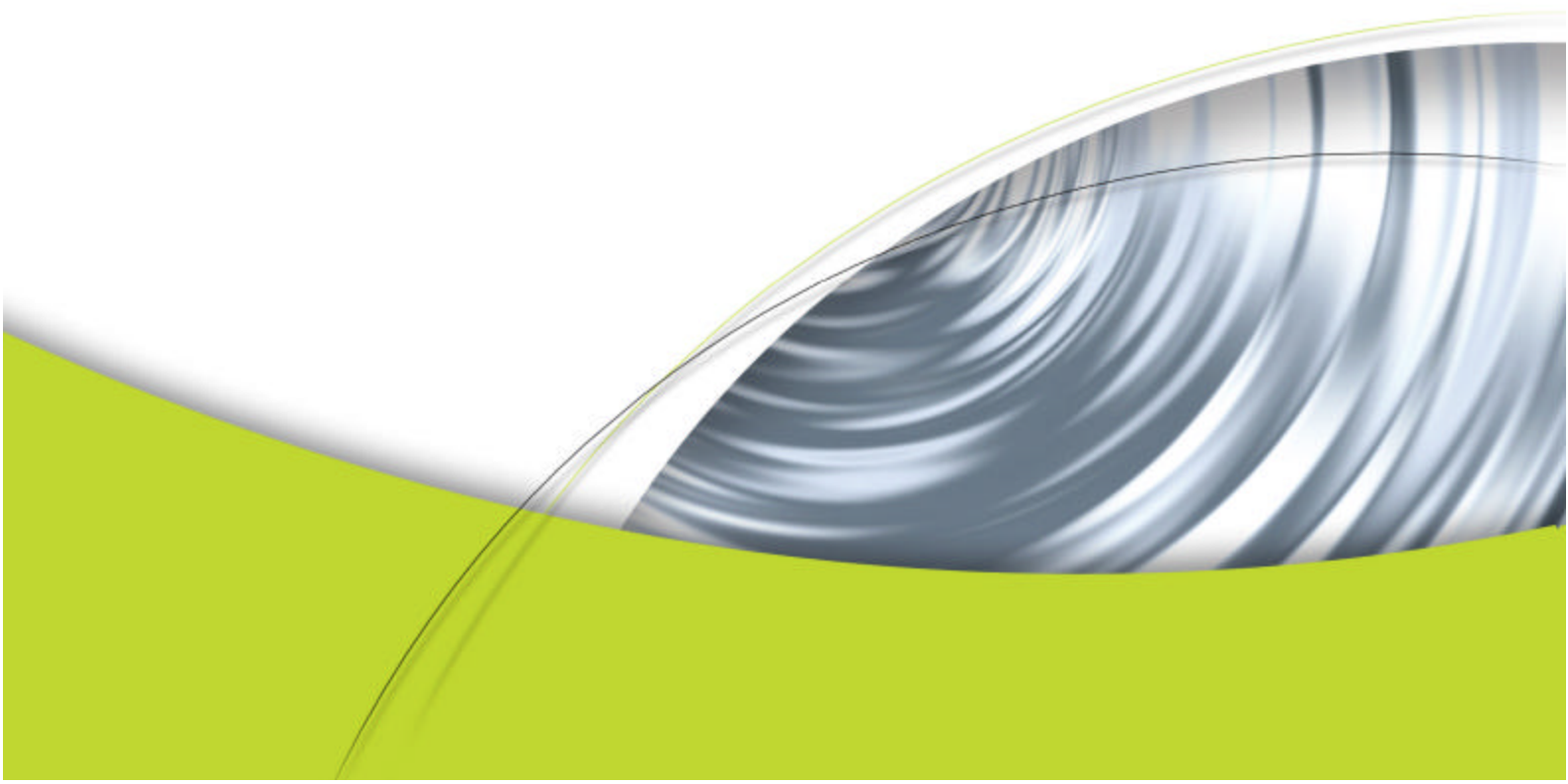

Technical Brief

TurboCache Technology

Redefining Power/Performance and Price/Performance for GPU Solutions

Introduction

The NVIDIA® TurboCache™ technology is a patented hardware and software solution that enables direct rendering to system memory. Through this innovative architecture, NVIDIA can deliver its award-winning NVIDIA GeForce™ 6 Series graphics processing unit (GPU) architecture—including Microsoft® DirectX® 9.0 Shader Model 3.0 and the NVIDIA PureVideo™ technology—to mainstream desktop PCs and notebooks.

This technical brief explains how the TurboCache architecture works and how it compares to non-TurboCache GPU architectures. In addition, it compares the graphics performance and system-level performance of TurboCache to existing graphics solutions to illustrate how TurboCache fundamentally redefines the price/performance and power/performance requirements of mainstream PCs and power-efficient notebooks.

TurboCache Architecture

The key architectural features of TurboCache are:

- A patented hardware and software technology that renders directly to system memory
- A TurboCache Manager (TCM), which dynamically allocates memory for maximum system performance
- Intelligent software algorithms that maximize application performance
- Bidirectional PCI Express® bandwidth, in conjunction with TurboCache architecture, which improves graphics price/performance and power/performance

The 3D graphics pipeline consists of four major stages:

1. Geometry processing (transform and lighting)
2. Setup (vertex processing, the process of converting vertices to pixels)
3. Texture application (the application of textures to pixels)
4. Rasterization (the application of lighting and other environmental effects to produce the final pixel value)

Figure 1 is a block diagram of a typical 3D pipeline.

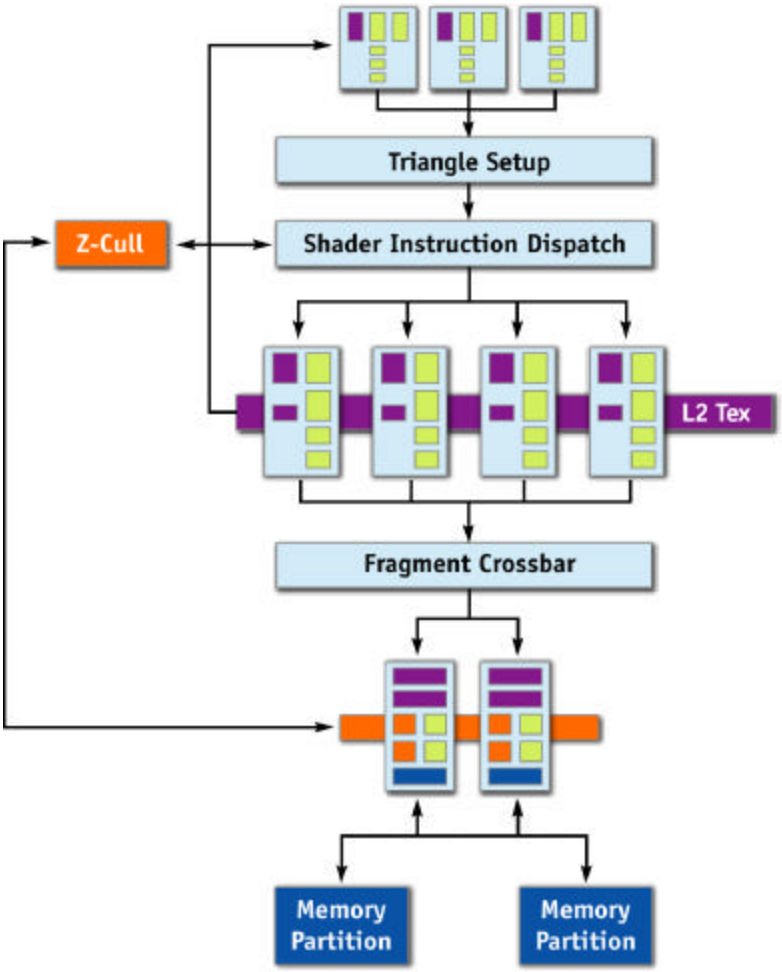

Figure 1. Typical 3D Pipeline

Figure 2 shows a 3D pipeline that is re-architected for TurboCache technology in order for the GPU to:

- Render to system memory with 100 percent efficiency
- Texture from system memory with 100 percent efficiency
- Dynamically allocate surfaces anywhere (locally or in system memory)

A new Memory Management Unit (MMU) allows the GPU to seamlessly allocate and de-allocate surfaces in system memory, as well as read and write to that memory efficiently. Plus, architectural modifications to a variety of pipeline elements handle the increased latency that occurs from accessing memory across the PCI Express interface.

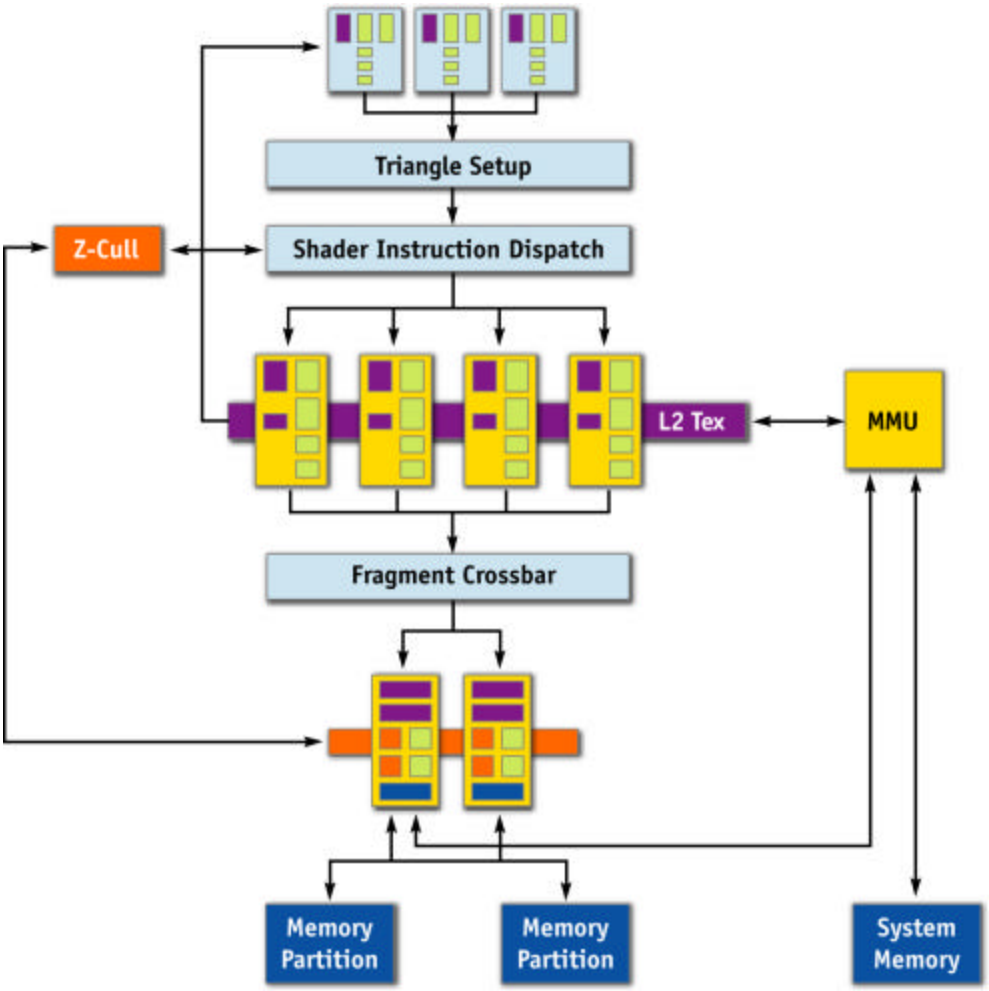

Figure 2. 3D Pipeline Re-Architected for TurboCache

Note: The yellow portions in Figure 2 indicate sections that have been modified.

In addition to the unique hardware built into the GPU to enable TurboCache, modifications were made to the NVIDIA ForceWare™ graphics drivers to intelligently determine where color, texture, and z-buffer data is located. This capability maximizes performance for each application.

The TurboCache Manager (TCM) lets additional memory be allocated for graphics usage based on application need. Once the application is closed, the memory that was allocated for graphics usage is then released and made available for system use. This is done behind the scenes and varies from application to application. In this way, TurboCache technology balances overall system bandwidth utilization during rendering by intelligently distributing the load between local graphics memory and system memory. The scanout buffer for display refreshes is always stored in the local graphics memory.

TurboCache Architecture vs. Non-TurboCache Architecture

Traditionally, render surfaces such as back buffers, depth or stencil buffers, environment mappings textures, or other data have been located in local graphics memory. This required large local frame buffers in order to accommodate the potential render surfaces.

With the unique capability of rendering directly to system memory, a GPU with TurboCache technology doesn't require the large amount of local graphics memory needed by non-TurboCache GPUs. Figure 3 compares the TurboCache architecture to a non-TurboCache GPU architecture in a mainstream PC configuration.

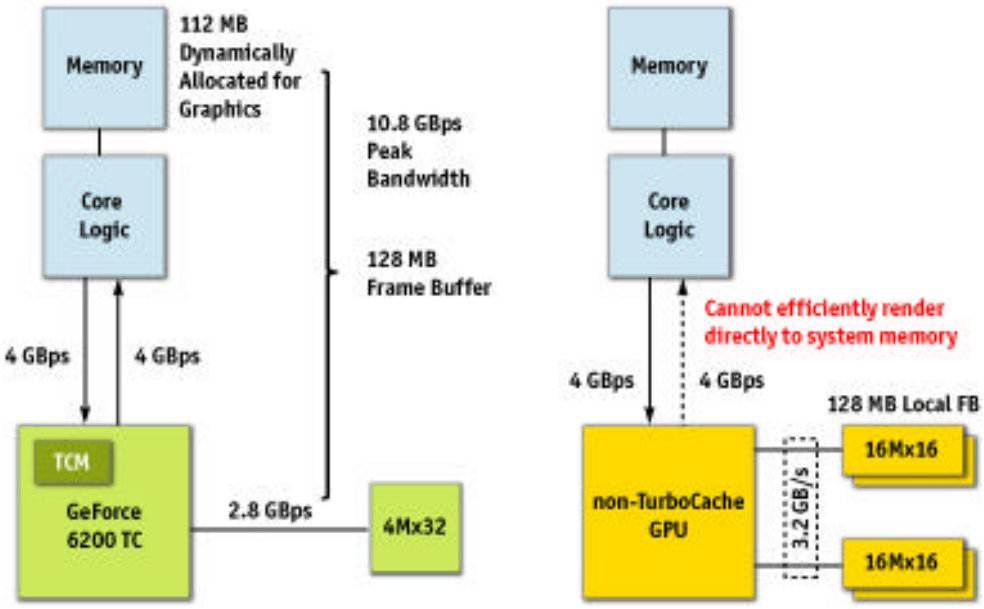

Figure 3. TurboCache Architecture vs. Non-TurboCache Architecture (for a mainstream PC)

Note that the minimum local memory requirement for TurboCache is a single piece of 4M × 32 memory (or 16 MB). However, TurboCache can dynamically allocate system memory for graphics use so the GPU has an entire 128 MB of graphics memory, just like a traditional GeForce GPU.

When memory is shared by both graphics and other system applications, memory bandwidth is important in delivering a quality user experience. By leveraging the bidirectional bandwidth of PCI Express of up to 8 GBps, TurboCache has an effective bandwidth of over 10 GBps (Table 1). This is a 50 percent increase in bandwidth over a non-TurboCache GPU, which cannot render directly to system memory.

Table 1. Memory Bandwidth Calculation

	PCI Express Bandwidth	Local Memory Type and Speed	Local Memory Bandwidth	Effective Bandwidth
GeForce GPU with TurboCache	8 GBps	DDR, 350 MHz	2.8 GBps	10.8 GBps
Non-TurboCache GPU	4 GBps	DDR, 200 MHz	3.2 GBps	7.2 GBps

Performance of TurboCache

NVIDIA TurboCache technology brings a new level of performance and features to mainstream PCs.

Figure 4 and Figure 5 show the performance benefits of Turbo Cache, based on tests performed using this system configuration:

- Platform:
  Intel Pentium4, 3.4 GHz EE CPU
  512 MB DDR system memory
  Windows XP SP1
  Drivers: ATI Catalyst 4.10, NVIDIA ForceWare 71.20
- Graphics:
  Intel Integrated Graphics media Accelerator 900 (i915G)
  ATI Radeon X300 SE (128 MB)
  NVIDIA GeForce 6200 with TurboCache supporting 128 MB, including 16 MB of local TurboCache *or*
  NVIDIA GeForce 6200 with TurboCache supporting 128 MB, including 32 MB of local TurboCache

Figure 4 compares the performance of a GeForce 6200 GPU with TurboCache supporting 128 MB (including 32 MB of local TurboCache) to Intel i915G integrated graphics and to a 128 MB ATI X300 SE.

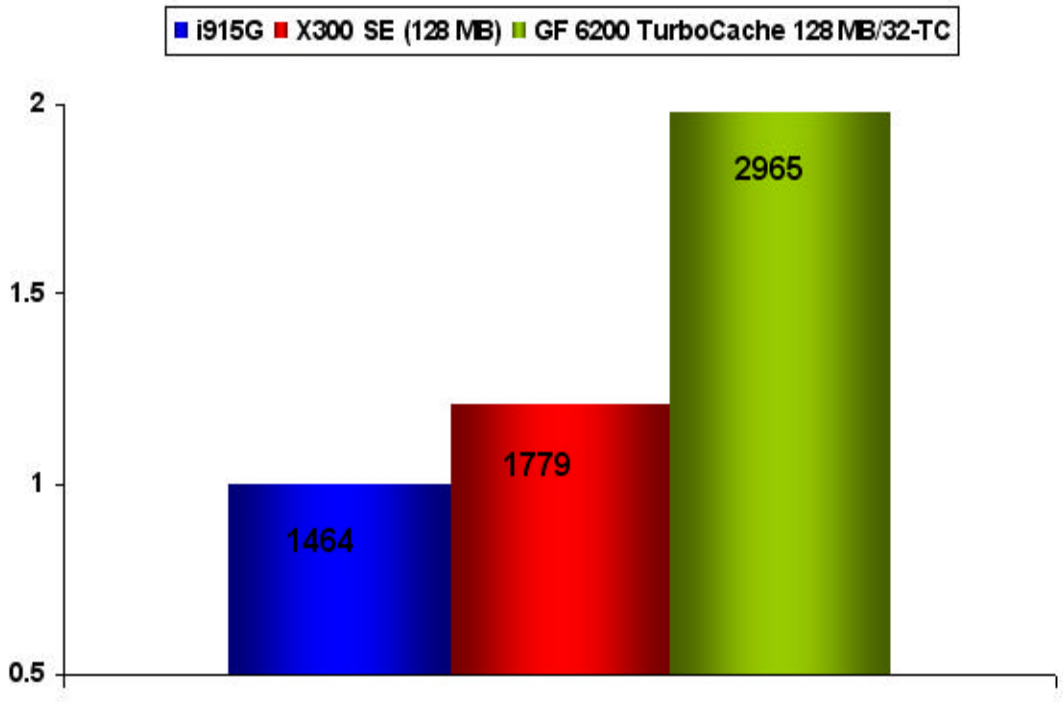

Figure 4. Graphics Performance: 3DMark03, 10 × 7, No AA or AF

The NVIDIA GeForce 6200 with TurboCache has twice the performance of an Intel i915G and outperforms an ATI Radeon X300 SE by over 60 percent. In addition, the GeForce 6200 with TurboCache delivers Microsoft DirectX 9.0 Shader Model 3.0 and NVIDIA PureVideo™ features to the mainstream desktop and notebook PC.

At the system level, the GeForce 6200 with TurboCache also outperforms an Intel i915G and an ATI Radeon X300SE with a variety of system benchmarks (Figure 5).

Note that the integrated solution is 100 percent dependent on system memory, so it will always require more system memory for graphics use than the TurboCache solution. It is the use of system memory and arbitrating for it that causes system performance degradation.

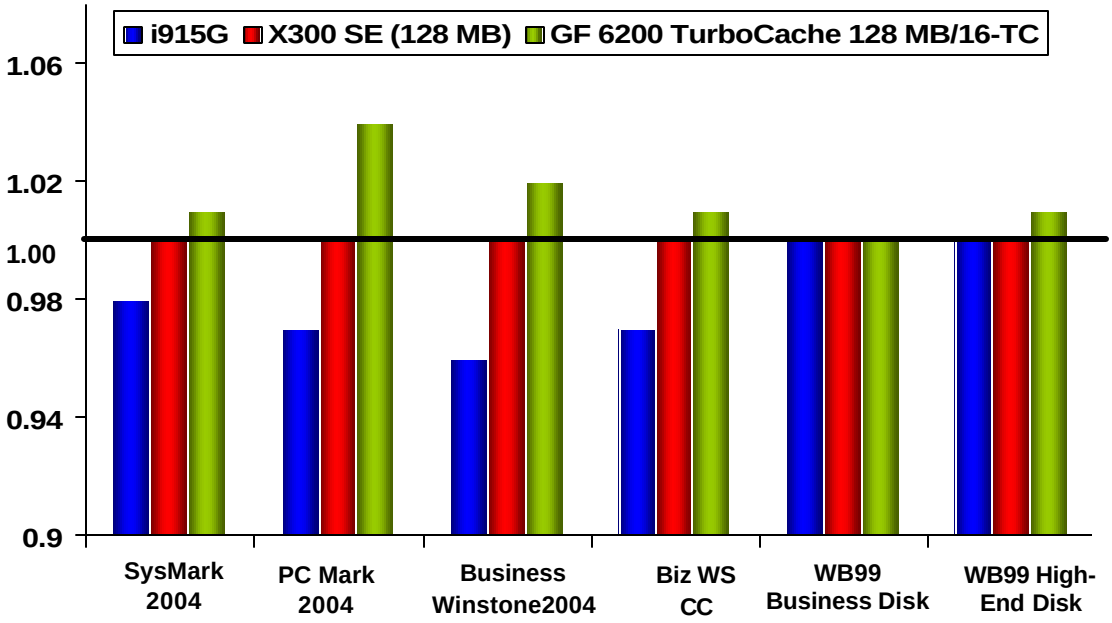

Figure 5. System Performance: 1280 × 1024 × 32 vs. Intel i915G and ATI Radeon X300SE

Conclusion

Through its unique ability to render directly to system memory, the GeForce 6200 GPU with TurboCache fundamentally redefines the price/performance and power/performance of mainstream discrete GPU solutions. With minimal local graphics memory requirements, the GeForce 6200 with TurboCache delivers graphics and system performance exceeding that of non-TurboCache 128 MB graphics solutions.

For notebook PCs, GeForce GPUs with TurboCache technology (GeForce Go 6200/6400) have the added benefit of consuming less power because they require less discrete memory. Using less discrete memory has the advantage of reducing the space required for the graphics subsystem. This allows notebooks with TurboCache GPUs to be smaller and consume less power.

With the innovative TurboCache technology from NVIDIA, users can now experience the latest graphics and video features of the GeForce 6 Series, including Microsoft DirectX 9.0 Shader Model 3.0 support and NVIDIA PureVideo, in mainstream desktop PCs and small, power-efficient notebooks

Notice

Trademarks

Copyright

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