3D Graphics Hardware: Where We Have Been, Where We Are Now, and Where We Are Going

Dan Baum
Silicon Graphics Computer Systems

What can we expect to see in 3D graphics hardware in the coming years? What markets will drive next generation designs? Where will PC graphics be in the year 2000? These questions are frequently debated by users, application vendors and manufacturers of 3D graphics hardware. In this article, I will offer some views and opinions on these and related questions. First, I will briefly review a bit of the history of hardware accelerated graphics. Next, I will give an overview of the current state of graphics hardware. Finally, I will discuss what directions graphics hardware will be going in the future.

History

Most of the advances in 3D graphics hardware have come from industry. Although there are some notable exceptions, universities do not have the capital resources to fund research in hardware development. Fifteen years ago, true interactive 3D graphics systems were primarily calligraphic and only ran using display lists (e.g. the Evans & Sutherland [E&S] Picture System). Although frame buffers for raster graphics had support for 2D graphics operations commonly referred to as "bit blits," they were devoid of acceleration for 3D graphics operations. In the early 1980s, the first accelerated framebuffers had some hardware support for primitive 3D graphics operations and implemented a z-buffer for hidden surface calculations. In the early 1980s, Jim Clark, then a Stanford professor, had a vision of building a 3D graphics engine on a chip. Clark, along with several of his students, took this idea and formed Silicon Graphics, where the first 3D graphics workstation developed. In 1984, the Silicon Graphics IRIS 1400 integrated a host computer with custom VLSI graphics accelerators (GeometryEngines) and a frame buffer.

The IRIS 1400 was an example of what is called a first generation graphics system. The generation refers to the targeted feature set for which the system runs at full performance. In the case of the IRIS 1400, this was flat-shaded polygons. This system supported immediate mode graphics, and was the first interactive raster system used for applications such as CAD and scientific visualization.

Second generation systems raised the realism level by rendering polygons that were Gouraud-shaded, Phong-lit and z-buffered. Additionally, raw transformation and scan-conversion performance jumped considerably. The first example of a second generation system was the Hewlett-Packard SRX, followed by the Silicon Graphics GT, which was the first system to break the 100,000 polygon/second mark. The improved shading capabilities allowed these workstations to become tools for applications such as modeling, animation and molecular modeling.

Third generation graphics systems appeared in late 1992, the first example being the Silicon Graphics RealityEngine, and added texture mapping and full-scene antialiasing. At Silicon Graphics, early engineering initiatives to support hardware accelerated texture mapping were countered with claims that "it is a cool feature, but there's no market!" With the universal acceptance of texture mapping all the way down to the level of commodity graphics, history has certainly proven otherwise. Third generation systems opened the door for general purpose graphics workstations to be used for out-the-window visual simulation applications instead of application-specific flight simulators.

Where Are We Now?

So where are we now? Currently, we are still in the third generation, albeit at higher performance levels. The Silicon Graphics InfiniteReality is the fastest graphics system available. Although it adds features and significant performance over its predecessor, the RealityEngine, it is still fundamentally designed to render textured, antialiased, polygons.

What about PCs? In the last couple of years, PC graphics accelerator chips and boards have made substantial jumps in performance and capabilities. However, PC graphics aren't defining the new features and capabilities. (Given market-imposed cost constraints and competitive pressure to quickly turn higher performance products, they fundamentally do not have the luxury to speculate with new functionality or to experiment with emerging markets.) However, PC graphics are successfully bringing existing 3D technology down to the mass markets. Currently, PC graphics are delivering second generation functionality at low prices. Many PC accelerator cards and even game boxes such as Nintendo 64 are offering some third generation features such as texture mapping support.

Perhaps what is most interesting is to speculate on where 3D hardware graphics is going in the future. Will everything drop down to low price PCs? I don't think so. By virtue of their cost limitations, they will continue to inherit functionality that is originally introduced by higher-end workstations. They will be the vehicle, however, by which these features are moved down-market by redesigning them at lower cost points. The capabilities of PC graphics will be closely tied to trends in semiconductor technology as most PC graphics solutions revolve around a single Application Specific Integrated Circuit (ASIC).

The Future

What trends will we see in higher-end graphics workstations over the next few years? The next machines will introduce fourth generation graphics that will be characterized by significantly enhanced image quality and realism. Greater image realism does not necessarily mean global illumination, however. The cost and memory access requirements associated with full global illumination will not be feasible for a production machine in the next several years. Rather, expect per pixel shading features such as phong shading, bump mapping, improved environment mapping and more complex reflection models to become standard at full performance levels. Some of these features will also show up in PC-based solutions in the not too distant future.

Texture mapping support will continue to evolve. Some specifics include support for very large texture volumes, improved texture filtering techniques and support for applying multiple textures simultaneously. Multiple textures allow applications to efficiently combine effects such as standard texture mapping and bump mapping. Additionally, the realism of atmospheric effects that are primarily used by the visual simulation community will also improve. Semi-global illumination effects such as shadows should also improve both in quality and performance. Possibilities for limited applications of image-based rendering in conjunction with traditional rendering methods also exist. Hopefully, we will also start to see enhancements to how colors are handled such that they more closely match the dynamic range and gamut of the real world.

It is important to remember that application performance does not solely rely on the raw performance of the graphics acceleration hardware. The performance of the integrated computer system is what really counts. When an application needs to move around gigabytes of textures or multiple streams of high definition video in real time, I/O performance both in and out of the graphics subsystem, as well as to disk and main memory, becomes critical. Higher-end graphics workstations will focus more on total integrated system performance and highly scalable and configurable system solutions, whereas PCs will not.

We will see important advances in human-computer interface. Improvements in display devices, especially those of the portable or head-mounted variety, coupled with enhanced input technologies such as haptics, will start to move virtual reality from a research topic to the driving technology for many applications.

For a number of years, people have been predicting that graphics hardware will start to transition from being polygon-based to rendering higher order surfaces as their basic primitives. I do not expect to see this transition in the next four to five years. In the near term, CAD and related markets seem satisfied with surface solutions that rely on more cost effective polygonal tessellation. However, acceleration for additional surface representations, such as subdivision surfaces, could appear.

Finally, given the increasing competition in the graphics hardware market in both the workstation and PC worlds, new graphics features are becoming very customer-driven and market-driven rather than speculative. Sets of features as well as higher level software toolkits will be introduced to address industry-specific needs for markets such as manufacturing, visual simulation, entertainment and data visualization.

The upcoming years will be exciting as we look forward to fourth generation graphics technology. However, not even the current capabilities of third generation machines are being fully exploited by application writers. By making use of standards such as OpenGL and forthcoming market-specific toolkits, application vendors will have the opportunity to take advantage of all that graphics hardware has to offer across multiple platforms. I challenge the graphics community to take this opportunity to bring the features and performance of current and future graphics systems to their customers.

Dan Baum is currently Director of Graphics in the Visual Systems Group at Silicon Graphics. During his 11 years at Silicon Graphics, Baum has been involved in either the design, implementation or management of all of the Silicon Graphics high-end graphics workstations from the IRIS 4D GT to the Onyx2 InfiniteReality. He received his A.B. in Enginering Science from Dartmouth College, and an M.S. in Computer Graphics from Cornell University.

Dan Baum
Silicon Graphics Computer Systems
2011 N. Shoreline Blvd.
Mountain View, CA 94043-1389
Fax: +1-650-932-3671
Email: *drb@sgi.com*