3Dlabs Flies With Jetstream

New Geometry, Rendering Chips Show Future Direction

by Peter N. Glaskowsky

Though last month's Microprocessor Forum featured several new CPUs optimized for floating-point 3D geometry processing, one presentation showed that programmable solutions lag well behind hardwired logic for peak 3D performance. The Gamma G3 geometry processor, part of 3Dlabs' new Jetstream architecture, is expected to handle as many as 44 million vertices per second by the end of 1999—about four times the throughput of any workstation CPU available during the product's lifetime.

The G3 is also the first announced chip to include three 4× AGP interfaces: one upstream port for the host computer's core logic plus two downstream connections to 3D rendering engines. This allows better parallelism for 3D rendering than 3Dlabs' current solution, which uses multiple Glint MX chips on a single shared PCI bus. It also gets around the limitation created by the point-to-point nature of AGP, which doesn't allow bused configurations. In his presentation, Neil Trevett, 3Dlabs' vice president of workstation graphics, also described the new Glint R3 and R4 rendering chips designed for use with the G3.

MICHAEL MUSTACCHI

Neil Trevett, VP of workstation products for 3Dlabs, describes the new Jetstream architecture.

Gamma G3 Is Six Chips in One

Apart from its unique AGP-interface architecture, the G3 contains six identical fixed-function geometry processors that are derived from the core in the company's pioneering Gamma G1 (see MPR 6/23/97, p. 1). Processing is distributed among the six cores according to the tasks to be performed.

The G3's cores include minor improvements over that of the Gamma G1, some of which are also included in the Gamma G2 chip that 3Dlabs will soon begin shipping as a pin-compatible upgrade to the Gamma G1. The Gamma G2 is almost twice as fast as the G1, a speedup achieved in part by increasing the chip's clock speed from 85 to 125 MHz. The remainder come from enhancements to the G2's setup processing. The G3 will run on an internal 133-MHz clock and offer significantly faster lighting calculations.

Each core in the G3 has seven basic stages: a command unit to process DMA operations; a vertex machine to assemble graphics primitives; a coordinate-transform unit; a geometry unit that performs clipping, culling, and fog calculations; a normalization unit; a lighting unit; and a final material unit to combine lighting effects with material properties such as specular highlights. Vertices pass through a total of 150 pipeline stages on their way through the G3. With its six parallel pipelines, there are 900 pipeline stages operating simultaneously on vertex data.

So many pipeline stages would be impractical in a general-purpose programmable CPU, because pipeline hazards would cause frequent stalls or pipeline flushing. The G3 has no such hazards. This illustrates one of the reasons that dedicated geometry processors can outperform host-based solutions, even on high-end CPUs. The G3 is effectively 12 times faster than the original Gamma G1, with a peak rating of 26 GFLOPS. With no lighting calculations to perform, a single G3 can process as many vertices as can pass over a 4× AGP interface—about 44 million per second.

When Six Cores Aren't Enough …

Professional 3D applications often apply multiple lighting effects to individual polygons. A single G3 would slow down linearly in such a situation. However, 3Dlabs realized that its two-port AGP bridge, designed to support parallel rendering chips, could also help solve this problem. Creating a two-level or three-level hierarchy of G3 chips allows the transform and lighting tasks to be spread among two or four devices. Figure 1 shows the high-end configuration, with seven G3 geometry engines driving eight Glint R4 rendering chips.

In a cascaded configuration, the first G3 chip performs trivial vertex culling and other simple tasks, then broadcasts all incoming vertices to both of its downstream ports. The second-level G3 chips (if present) perform no processing at all, acting as simple AGP bridges to fan out the vertex data. All transform and lighting operations are performed in the final layer of G3 chips, allowing up to four devices to share the processing required for these tasks.

Working together, a three-level G3 pyramid has up to 115 GFLOPS of 3D processing power, about 50× faster than the original Gamma G1. The relative ease of achieving such parallelism is another advantage of dedicated geometry processors compared with general-purpose CPUs. The scalability of workstation processors depends heavily on caching schemes and bandwidth to shared memory. The G3 does not use an external main memory, and even the G3s at the same level in a single subsystem share no data and need no form of cache coherency.

Rendering Engines Keep Pace

The impressive geometry-processing performance of the Gamma G3 will allow it to handle 3D objects or scenes of unprecedented complexity. The company's current Oxygen GMX add-in board can display a scene with two million polygons about once per second with one light source. A high-end Jetstream subsystem should be able to redraw such a scene 20 times per second with four light sources.

Scene complexity and frame rate are just two of the four essential performance metrics for 3D rendering. The other critical measurements are resolution and display quality. Geometry acceleration can't help with these aspects of the task; it takes a different kind of chip. The Glint R3 and R4 are 3Dlabs' solution to this part of the problem.

The Glint R3 renderer is essentially equivalent to the mainstream Permedia 3 rendering engine announced earlier this year (see MPR 8/3/98, p. 1). These chips, and the R4, provide a peak fill rate of 125 Mpixels/s with trilinear filtering, as fast as any rendering chip announced to date.

The Glint R4 incorporates small but significant improvements to the P3/R3 design, such as the ability to manage up to 64M of local memory. The R4's greatest improvement is its support for parallel rendering. Up to eight chips can work together to achieve fill rates as high as one gigapixel per second.

This parallelism is achieved by dividing the screen into horizontal stripes of pixels, each 4, 8, or 16 pixels high. R4s are assigned to each stripe in rotation, with one G3 controlling each pair of R4s. With two R4s, some polygons will fit entirely within a single stripe, and some must be processed by both R4s, as Figure 2 shows. The G3 driving the pair determines the coverage of each polygon by inspecting the transformed vertex coordinates, then it sends each polygon to one or both R4s as needed.

If a polygon will not be visible on either R4, it is discarded, eliminating lighting calculations in the G3 and any further processing by the R4. This stripe-based solution, 3Dlabs believes, boosts setup-processing performance by about 50% with two R4s and up to 600% with eight R4s.

Pricing and Availability

3Dlabs did not announce pricing for Jetstream chips. The company will offer its own Jetstream-based expansion cards in 1999. Midrange boards using the Glint R3 should appear by midyear, with prices starting at $300. High-end cards are planned for the second half of 1999. We expect these cards to range in price from $2,000 to $8,000, typical of products in this market. For more information, visit the 3Dlabs Web site, *www.3dlabs.com*.

This stripe-based parallelism works best on small polygons, which are less likely to overlap multiple stripes. The high geometry-processing throughput of the Gamma G3 supports real-time rendering of 3D models with several million polygons. In a rendered frame of such a model, each polygon will necessarily be very small, typically less than 10 pixels in size and therefore likely to fit into a single stripe. When processing connected strips of such polygons (such as the strip shown in Figure 2), the G3 is likely to break the strip as it passes across a stripe boundary. In this case, the G3 will automatically begin a new strip for the next R4 rasterizer.

The Jetstream solution represents a significant advance over the original Glint Gamma/MX architecture, which allowed only one Gamma to drive just two or four MX rendering chips over a shared 66-MHz PCI bus. The older approach offered no scalability for geometry processing and limited overall vertex traffic to the capacity of the PCI bus.

The new 3Dlabs scheme also invites comparison with that of 3Dfx's Voodoo2, which allows a pair of Voodoo2 cards to handle alternate scan lines—a method similar to Jetstream's, except that 3Dfx's stripes are just one pixel high. The Voodoo2 approach shares the pixel-drawing workload, but 3Dfx does not have a geometry processor. All polygons, furthermore, are sent to both Voodoo2 cards, and both cards must perform setup calculations for all

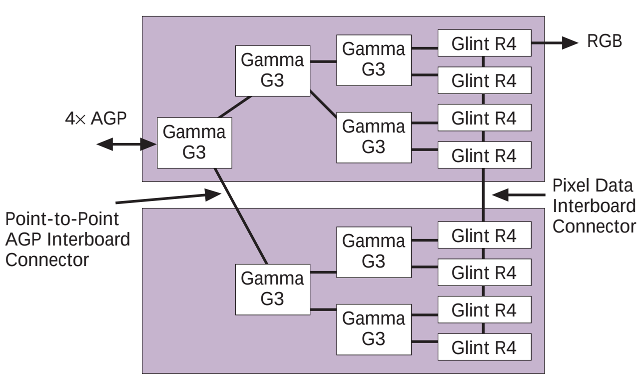

Figure 1. A high-end Jetstream configuration uses seven Gamma 3 geometry processors and eight Glint R4 rendering engines to draw 44 million polygons and 1 billion pixels per second.

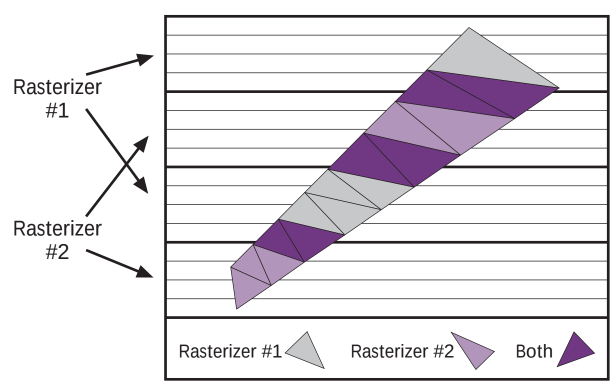

Figure 2. The Glint R4 rendering chip can be used in parallel configurations where each chip rasterizes only a portion of the polygons in the 3D model.

polygons. Few polygons will fit into a single scan line, so there was little to be gained from testing vertex coordinates against scan lines.

Jetstream Won't Blow Away Competition

The scalability of the Jetstream architecture should allow 3Dlabs to remain at or near the top of the performance ladder in the highly competitive professional-3D market. Though its products outsell those of other independent chip vendors, 3Dlabs has lost some market share in this segment to Evans & Sutherland, and it will face even more competition in the coming year from new entrants such as Raycer Graphics (*www.raycer.com*), PixelFusion (*www.pixelfusion.com*), and SP3D (*www.sp3dtech.com*). System vendors with their own graphics products, including HP, Intergraph, and soon SGI, control a majority of the workstation market, and 3Dlabs cannot sell its chips or boards to those companies.

Intergraph, which recently began shipping the first card based on its new Wildcat architecture, says Wildcat will scale to more than 1,000 on the industry-standard ViewPerf CDRS-03 benchmark. According to 3Dlabs, Jetstream will scale to more than 1,300 on the same test. These comparisons must be taken with a grain of salt, since the CDRS-03 dataset has significant flaws and will soon be replaced (see MPR 8/3/98, p. 4). Even so, CDRS-03 remains a popular basis for comparing new professional 3D products.

As long as 3Dlabs' performance is close to that of other vendors—or superior, as the company hopes—it will remain a major supplier of professional 3D solutions to independent workstation vendors. This market is not particularly large (fewer than 100,000 systems per year) or fast-growing (about 10% per year), but it still provides good profit margins. These margins may erode due to increased competition in the next year, but they're still enough to justify 3Dlabs' investment. By designing just two chips with the ability to cover an 8:1 performance range, 3Dlabs has done a good job of leveraging its development costs. We expect Jetstream to be impressive, pricey, and popular when it arrives next year. M