VOLUME 10, NUMBER 6
MAY 6, 1996

MICROPROCESSOR REPORT

THE INSIDERS' GUIDE TO MICROPROCESSOR HARDWARE

# Verite: A Programmable 3D Chip

## *Rendition Uses Internal RISC CPU for Versatility, Performance*

*by Linley Gwennap*

Rendition, a fabless graphics-chip vendor, has taken a unique approach to accelerating 3D graphics. Rather than the fixed-function designs favored by most vendors or the fully programmable processor developed by Chromatic, Rendition's Verite (pronounced "verity") chip combines a programmable RISC processor core with a fast pixel engine. This hybrid may ultimately deliver better performance on real applications than either of the other approaches.

Verite's internal CPU is similar in many ways to the R3000 and its contemporaries. Rendition has added several enhancements to improve performance on graphics code, such as an expanded register file, SIMD operations, and specialized drawing and arithmetic. At 50 MHz, it can sustain 100 million operations per second (MOPS), similar to a low-end Pentium processor.

The Verite CPU executes the 3D setup function, converting polygon vertices to horizontal spans so the host CPU need not perform these calculations; Verite's pixel engine handles 3D rendering. Most other 3D accelerators require the host CPU to perform setup, creating performance bottlenecks in the host CPU and the PCI bus that prevent these 3D chips from achieving their rated performance. By avoiding these bottlenecks, Verite should outperform most other 3D chips on advanced 3D games and similar applications.

Verite also handles 2D graphics acceleration and has some video functions, but it does not handle audio, fax/modem, or MPEG decompression, so it is not in the same class as Chromatic's media processor, for example. Verite is aimed at the add-in card market; Rendition assumes audio and fax/modem chips are on the motherboard already, and MPEG-1 decoding can be handled in software on most Pentium processors.

Rendition—which is led by Michael Boich, former president and founder of Radius—is currently sampling Verite and intends to have production silicon available at the end of this month. At a list price of $65, the chip can be used in add-in cards that sell for about $250.

### Full-Featured Graphics Accelerator

Figure 1 shows a block diagram of the Verite chip. The two main processing blocks are the RISC CPU core and the pixel engine. The CPU has its own small instruction and data caches, while the pixel engine has a small texture cache. These two processors perform 2D and 3D graphics acceleration; a VGA block handles legacy DOS programs. Both the pixel engine and the VGA block are hardwired; the RISC CPU is the only programmable portion of the chip.

Verite connects directly to the PCI bus. To increase performance, the bus-mastering PCI interface includes a DMA controller that supports chained DMA. This feature allows a large block of data to be transferred from main memory to Verite using a stream of consecutive DMA transactions with no dead time. Rendition has measured PCI bus utilization as high as 96% using this technique; thus, the chip can accept data at nearly the peak speed of the PCI bus. While several recent chips support PCI bus mastering, only a few perform chained DMA, although we expect this feature to become more common in the future.

The memory controller supports fast-page-mode or

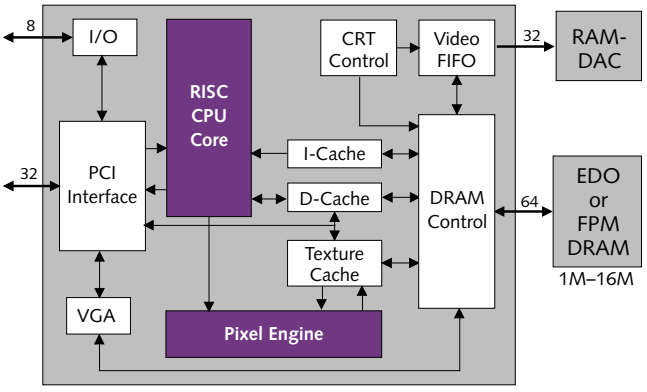


**Figure 1.** The Verite processor combines a programmable RISC CPU with a pixel engine for efficient 3D performance. It connects directly to PCI, EDO DRAM for the frame buffer, and a RAMDAC for video output.

EDO DRAM, although the latter is recommended for best performance. This memory holds the frame buffer, Z-buffer (if needed), and texture storage; it is also used by the RISC CPU to store code and data. A typical implementation will have 2M of off-chip memory, supporting double-buffered 640 × 480 × 16-bit mode (typical for 3D games) and all necessary RISC CPU code while leaving about 700K available for texture caching. For GUI applications, the 2M memory can handle a single-buffered display at up to 1024 × 780 × 16. Higher resolution or more colors can be used by increasing the size of the frame buffer, which can be as large as 16M.

Rendition eschewed more advanced memory types such as SDRAM or RDRAM *(see* **100605.PDF** *)*, which provide more bandwidth than EDO DRAM but also, at least today, carry a significant price premium. EDO prices are approaching parity with fast-page-mode prices, making this type of memory very cost-effective. Rendition believes that the 400-Mbyte/s peak bandwidth from its 64-bit-wide EDO memory system is adequate for the 3D performance the market requires today.

The chip includes a simple 8-bit I/O bus. Add-in card vendors can attach an ASIC through this port to provide value-added features. The protocol is flexible enough to allow a standard sound chip to be added.

## RISC Engine Handles 2D Acceleration

In addition to 3D setup, the RISC engine handles 2D graphics acceleration. According to Rendition, 2D performance is around 30 Winmarks, fairly speedy but a bit behind the leaders. The RISC engine can process 64 bits of data at once, comparable to most leading 2D chips. The RISC CPU also handles SVGA emulation; in this mode, it can execute graphics commands as fast as the PCI bus can deliver them.

A 2K instruction cache and a minimal data cache support the RISC CPU. Rendition would not specify the size of the data cache but noted graphics data generally does not cache well because it is used once and discarded, so there was no reason to add a large data cache. A second small cache stores texels (texture pixels) for the pixel engine.

The pixel engine handles 3D rendering. It performs the gamut of per-pixel drawing functions, including texture filtering, Z-buffering, fog, and perspective-correct bilinear and trilinear texture mapping *(see* **100103.PDF** *)*. These effects can be combined to produce high-quality 3D images. The pixel engine is optimized for executing these functions; performance increases only slightly when they are disabled. Other 3D rendering engines may deliver high performance using simple drawing functions but bog down when called upon to produce high-quality polygons that are textured, shaded, and accurately rendered.

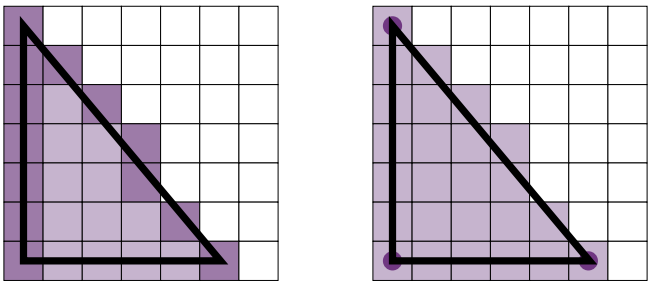

**Figure 2.** A 25-pixel right triangle typically requires 14 pixels to specify seven horizontal spans, but in the setup process, only three vertices define the entire triangle.

Rendition rates its pixel engine at 500,000 flat-shaded polygons per second, or 200,000 textured polygons per second. Both specifications assume 25-pixel triangles that are perspective correct, subpixel accurate, Gouraud shaded, and bilinear filtered in a 640 × 480 × 16-bit frame buffer. Although other vendors have announced significantly faster rendering engines, those chips will be unable to achieve their quoted peak rates without a hardware setup engine.

The pixel engine also provides some video acceleration. In particular, it handles image scaling and YUV-to-RGB conversion. These features allow the chip to display full-screen digital video from MPEG, Indeo, or other video sources. Verite can also apply video texture maps to 3D objects.

The video subsystem also includes a CRT controller and a simple FIFO, allowing the memory controller to handle accesses from the internal CPU and pixel engine without delaying the video stream. This FIFO sends pixel data to an external RAMDAC.

## Programmability Provides Flexibility

Including a programmable processor gives Verite two key advantages. First, the device is more flexible than a hardwired design. The current 3D environment is very immature, with varying hardware and software standards proposed but none firm. Even Microsoft's Direct3D, expected to become the dominant API, has just achieved beta release and is likely to continue evolving over time. Programmability allows the hardware to quickly adapt to changing standards.

This flexibility also allows Rendition to test certain new algorithms and techniques quickly and easily, letting the small company iterate its design without spinning chips or buying expensive emulators. Even after it begins shipping its chips, Rendition could improve performance or add entirely new features to devices in the field simply by supplying new software.

The second major advantage is the ability to repartition tasks between the host CPU and the graphics chip as needed for optimum performance. Verite's RISC processor is a general-purpose device with roughly the performance of a low-end Pentium, so it can be programmed to do almost anything the host CPU can. In particular, Verite is designed to offload the triangle setup stage of the 3D pipeline, freeing the host CPU to focus on geometry calculations. Only a few of the current fixed-function 3D chips handle setup *(see* **100304.PDF** *)*.

Partitioning could be applied to other tasks as well. The

Verite RISC engine is capable of handling video processing, for example, although Rendition has not yet developed the necessary software for this area. The RISC engine also simplifies the drivers that run on the host CPU; Verite can accept parameters in any order, for example, whereas a hardwired graphics chip depends on the host to arrange parameters and generate the correct control words. In short, the programmable Verite engine can work hand in hand with the host CPU to most effectively handle graphics tasks.

The disadvantage of programmability is that, in general, hardwired logic for a specific function is more compact than a general-purpose processor. Realizing this, Rendition hardwired several portions of the chip, such as the VGA logic and the pixel engine. These sections execute well-defined functions, so there is little risk in giving up flexibility in these areas. Performance is high, and the amount of code that Rendition must generate is reduced.

## Setup Required for High 3D Performance

The issue of 3D setup is key in today's PC market. In the rush to get first-generation devices to market, most 3D vendors have chosen to leave setup to the host CPU. These chips are simple rendering engines, in the tradition of workstation graphics accelerators. A rendering engine, also known as a span engine, requires the CPU to define the horizontal spans of a triangle or polygon on a scan-line basis, as Figure 2 shows; the engine then performs anti-aliasing and applies textures, shading, and other effects. The figure shows that a simple 25-pixel polygon usually covers seven scan lines and thus requires 14 pixels to define its spans.

A setup engine, in contrast, requires only the absolute subpixel addresses of the three triangle vertices. It maps these subpixel addresses onto the pixel grid and calculates the position of each edge pixel, thus defining the spans. At that point, the spans can be processed by a rendering engine as in the previous example.

Traditionally, 3D graphics have been performed on workstations using rendering engines. But applying the workstation model of a rendering engine onto a Pentium PC creates two problems: CPU overhead and bus bandwidth.

In a typical workstation application like CAD, the host CPU is doing little other than creating a highly realistic 3D image; thus, it is reasonable to devote 100% of the host CPU to the task of setup. In a PC environment, however, the typical 3D application is a game that is updating a complex environment in real time. Games often include intelligent algorithms that simulate realistic opponents. Game designers want as much of the CPU as possible to be available for these tasks rather than 3D imaging.

The speed of the game is determined by the number of CPU cycles available. By moving setup to the Verite chip, Rendition maximizes the response time of the game. Alternatively, Verite allows game designers to execute more complex algorithms without moving to a faster CPU.

The Rendition design offers high performance even on slower processors. Because a rendering engine relies on the host CPU for setup, the frame rate will be lower on, say, a Pentium-75 than a Pentium-133. Verite provides a constant level of setup and rendering performance, keeping the frame rate more consistent across a range of processors. With Verite, gamers will see a better response time from the faster CPU, but the visual output will be similar.

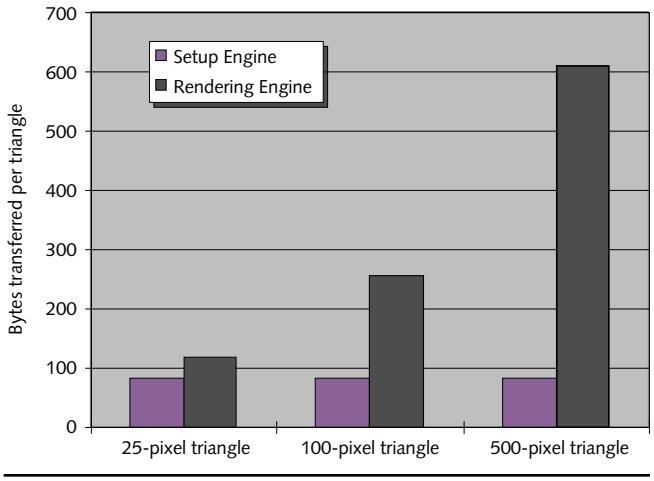


**Figure 3.** Moving setup from the host CPU to the graphics chip reduces bus bandwidth by 30% for small, randomly placed triangles, with even greater gains for larger polygons. These figures are for Microsoft's Direct3D API. (Source: Rendition)

Setup occupies a significant portion of the host processor even with a Pentium-166, although the fraction varies depending on the complexity of the 3D image and the frame rate. Intel agrees that even its fastest Pentiums will not be well suited to performing setup. Even the forthcoming MMX-enhanced P55C gains little in this area, as setup typically involves 32-bit or greater precision, whereas MMX is optimized for 8- and 16-bit calculations. With its fast floating-point unit, the P6 may have the CPU power to handle setup for at least some applications.

## Verite Avoids "Graphics Deceleration"

Even if the host CPU has enough cycles available for setup, the amount of data that must be transferred in this arrangement is much greater than when the graphics chip performs setup. In the 25-pixel triangle example, a rendering engine needs an average of seven spans, or 14 pixels, per triangle, whereas a setup engine needs only the three vertices of the triangle. A typical workstation contains a high-bandwidth connection dedicated to the rendering engine, enabling it to transfer the large amount of data required. A PC rendering engine connects via the PCI bus, however, which is quickly swamped by the rendering data created by a typical 3D game and also must be shared with other system resources.

Figure 3 graphs the bus-bandwidth advantage achieved by a setup engine. These figures take into account all data that must be transferred for each triangle, including the vertices (or spans, for a rendering engine), depth, and color

information. For the common case of a small 25-pixel triangle, the setup engine reduces bandwidth by 30%. This benefit becomes greater as the triangle size increases, since the setup engine needs the same amount of information but the span engine requires more spans to define the larger triangles. These larger polygons are often used to create the background (sky, grass, etc.) in a 3D game.

Because of this bandwidth problem, initial tests performed by Microsoft and others using PC rendering engines show these chips are, in some cases, "graphics decelerators"; that is, 3D performance is better using full software rendering than when using these rendering chips. Note that the 25-pixel triangle requires more than half of its pixels to be communicated to the rendering engine; assuming a two-to-one overlap among various triangles, a reasonable value for a 3D game, software rendering actually sends fewer pixels to the graphics chip than hardware rendering does. When PCI bandwidth becomes the bottleneck, the number of pixels sent across the PCI bus is the critical performance issue.

Of course, software rendering is rather slow; the point is that, for many 3D games, a rendering engine cannot deliver adequate 3D performance in today's PCs due to the PCI bottleneck. The polygon rates quoted for these rendering engines represent peak rates that are not achievable or sustainable with typical 3D games in a PCI-based PC. These rendering chips are suitable for 3D programs that have less complicated images: fewer polygons, less texturing, lower frame rates. Most applications that are adopting 3D graphics, however, are games pushing the limits of frame rate and image realism; these applications are better suited for chips that handle setup.

As a result, both 3Dfx and 3Dlabs have setup chips that complement their rendering chips, and others are likely to follow. Rendition is the only vendor to date with a single chip that performs both setup and rendering. We believe designs that provide hardware setup will be needed for adequate 3D performance for the next 12–18 months. After that, it is possible that a PC with, for example, a 300-MHz P6 processor and a rendering engine based on the high-bandwidth AGP *(see* **1005MSB.PDF** *)* could deliver strong 3D performance with setup done by the host CPU.

## RISC Core Features Large Register File

The Verite CPU core is similar in many ways to a classic RISC processor such as the R3000. The instructions and registers are 32 bits wide. Instructions use a load/store model with three-operand arithmetic and a straightforward encoding. Loads and branches each have a single delay slot, although the hardware provides load interlocks, unlike the R3000. Data is stored in big-endian format, but the DMA engine supports endian swapping for compatibility with little-endian x86 processors. As in the MIPS architecture, only a single addressing mode, register+offset, is provided. These features give the Verite CPU a general-purpose flavor and make it simple to program.

One might ask why Rendition did not simply use an existing MIPS (or other RISC) core, but the company found it necessary to add a few optimizations to enhance graphics performance. A key feature is the large register file: each instruction can access any of 256 registers for its source or destination fields. Of these, 128 are general-purpose registers, giving Verite more than four times the register capacity of a typical RISC processor. Another 32 hold fixed values (zero, one, minus one, bit masks, and other useful constants); the remainder are allocated to special functions such as flags, control bits, and input/output or are reserved for future use.

The 128 general-purpose registers come into play when rendering fully featured polygons. Even a simple 3D polygon requires several registers per vertex: X, Y, and Z (location); RGBα (color); U, V, and Q (texture); plus slope and interpolation values. With three vertices, a simple triangle consumes most of the registers of a typical RISC processor, leaving only a few for working values. More complex triangles require additional parameters for shading, texturing, light sources, fog/transparency, and other features, quickly overflowing 32 registers. In fact, Rendition claims a fully featured polygon cannot be stored in even 64 registers, causing it to allocate 128 for Verite.

With this design, Verite can set up and render these complex polygons without a single load once the parameters are placed in the register file. In contrast, a processor with fewer registers must frequently reload parameters during setup, significantly reducing throughput.

## Special Instructions Optimized for Graphics

To further improve performance, the Verite CPU can operate on two 32-bit registers at once using parallel 32-bit ALUs. This feature allows limited SIMD (single-instruction multiple-data) operation similar to that of Intel's MMX instruction set *(see* **100301.PDF** *)*.

Rendition did not, however, include parallel operations on values smaller than 32 bits, a key feature of MMX. The Verite CPU is optimized for 3D setup, which requires 32-bit precision. Rendering, which operates on smaller values, is performed by the pixel engine. In contrast, MMX is designed for rendering, image processing, and audio processing, all of which use smaller values. Verite does not handle audio data at all, leaving this function to a support chip.

The lack of 8- and 16-bit parallel operations will not hamper Verite on its intended task of 3D setup. On other tasks, however, this lack would restrict the chip's performance relative to an MMX processor. If Rendition wishes to add, for example, audio capabilities to future products, its current RISC CPU design would be inefficient. The Verite core could be fairly easily modified, however, to operate on these smaller data types.

The chip includes two other parallel instructions for drawing pixels. These instructions can generate up to four pixels per cycle, a total of 64 bits, with various options. They are used for 2D graphics to perform bit block transfers and

are also helpful in certain 3D operations.

Another optimization for graphics is the inclusion of some unusual arithmetic operations. In addition to the standard integer and logical functions, Verite performs minimum, maximum, ceiling, and floor operations. These operations are useful for clipping algorithms and color functions. Unlike MMX, the Verite CPU does not include saturating arithmetic; the maximum and minimum functions can simulate saturating arithmetic, albeit with extra operations. The pixel engine, which handles most color functions, does perform saturating arithmetic, reducing the need for this feature in the RISC engine.

Verite also includes a $32 \times 32 \rightarrow 64$-bit multiply instruction with a two-cycle latency. These multiplies can be issued every cycle. This function speeds many image-processing algorithms. The multiplier includes an unusual post-shift normalizer for variable-radix multiplication. Although Verite does not support full floating-point operations, the variable-radix option allows programs to use block floating-point, in which software keeps track of the location of the binary point within the register.

Rendition does not plan to publish the full instruction set, as this would require supporting many vendors that might wish to develop code. Instead, the startup plans to develop its own software, providing drivers for Direct3D, Quickdraw 3D, and other emerging 3D APIs as well as 2D drivers for Windows 3.1, Windows 95, and Windows NT. The company will allow a few third parties to develop custom drivers as necessary.

## Simple Pipeline Achieves 100 MOPS

The initial implementation of the Verite processor is a simple scalar processor that can issue one instruction per cycle at 50 MHz. Using the parallel arithmetic instructions, Verite achieves a peak execution rate of 100 MOPS. Since Verite is usually executing code that takes advantage of parallel arithmetic, for these algorithms it should deliver performance similar to that of a low-end Pentium.

Verite implements a simple five-stage pipeline, again like the R3000: fetch, decode, execute, load, writeback. The only instructions that cause pipeline delays are loads and multiplies, as branches have a software delay slot. Load and multiply interlocks can be avoided by inserting a single instruction between the load/multiply and the instruction that requires its result.

Figure 4 shows the Verite RISC core. Not including the large register file, this simple design consumes less than 10 mm$^2$ in a 0.6-micron two-layer-metal process, yet it provides adequate performance for 2D graphics and 3D graphics setup. While the Verite CPU can match the performance of a low-end PC processor, its 100 MOPS pales in comparison to the 2,000 MOPS reported for Chromatic's Mpact multimedia processor *(see* **091404.PDF** *)*, but these figures represent two very different design styles. For one thing, the Chromatic MOPS are for 8-bit data; when operating on 32-bit data, as in a 3D setup operation, Mpact peaks at 500 MOPS.

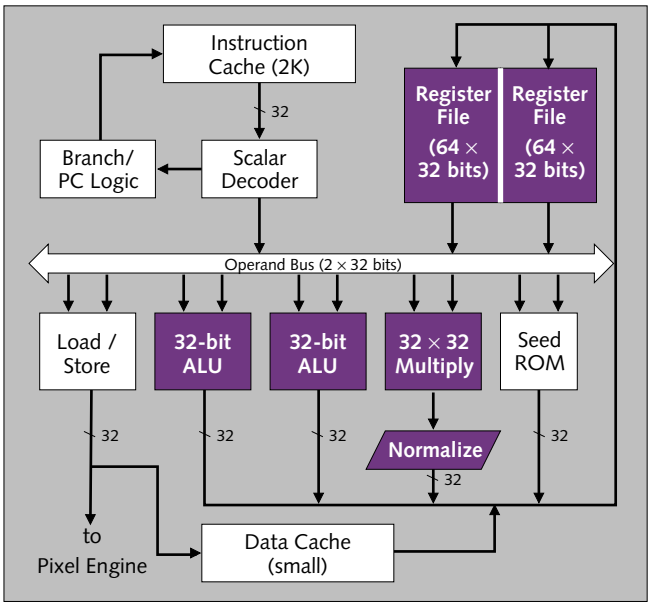


**Figure 4.** The Verite RISC processor executes one instruction per cycle but can perform that instruction on two sets of operands drawn from the parallel register files using the parallel ALUs. The multiplier includes a post-shifter to support floating-point emulation. The seed ROM supports random-number generation.

More important, the Mpact engine must also perform 3D rendering, GUI acceleration (2D graphics), and audio in a typical system, requiring immense processing power. Verite includes hardwired function blocks for rendering and 2D graphics, and it relies on an external audio chip, so its limited general-purpose processing power can be focused entirely on the critical 3D setup task. When heavily loaded with other tasks, Mpact can run out of cycles for 3D setup, forcing it to shift that task to the host CPU.

## Designed for Add-In Cards

The differences between the Verite and Chromatic designs highlight different target markets for these devices. Chromatic is intended to be placed on the motherboard in low-end and midrange systems, providing an all-in-one multimedia solution with a base level of performance. Verite is designed for add-in cards that deliver a superior level of 2D and 3D graphics performance while relying on motherboard components for other multimedia functions.

In the add-in market, Verite will compete with 3D accelerators from vendors such as 3Dlabs, 3Dfx, and S3. Of these, only the 3Dfx Voodoo chip set performs triangle setup, like Verite. But Voodoo does not handle 2D graphics or video, so a complete add-in card would require additional chips, increasing the minimum price for a Voodoo card to about $350, compared with around $250 for Verite.

The 3Dlabs and S3 chip sets provide more complete solutions than Voodoo but, without handling setup, offer

### Price & Availability

The Verite V1000 is currently sampling; Rendition expects volume production by the end of May. The company now has alpha versions of 2D drivers for Windows 3.1 and Windows 95 (DirectX) as well as 3D drivers for Direct3D, Speedy3D, CGL, and Heidi; it expects beta code by the end of May and final code by the end of June.

In a 240-pin PQFP, the V1000 carries a 1,000-piece list price of $65. This price includes driver software. For more information, contact Rendition (Mountain View, Calif.) at 415.335.5900; fax 415.335.5999 or check the World Wide Web at *www.rendition.com*.

poor 3D performance on many applications. Recognizing this problem, 3Dlabs has recently developed a setup engine to couple with its Permedia rendering chip. This new chip allows 3Dlabs to offer performance comparable to Rendition's, but its two-chip solution costs more than Verite.

Other vendors are working on powerful rendering engines due to ship late this year; these companies include Oak, Trident, and NEC/Videologic. Nvidia is also upgrading its NV1 graphics/audio accelerator *(see* **090904.PDF** *)*. Without on-chip setup engines, however, these devices are likely to run into the same bottlenecks as current rendering chips. Cirrus is the only other vendor to announce plans for a 3D accelerator that includes a setup engine; this device, based on technology licensed from 3DO *(see* **1005MSB.PDF** *)*, should ship by the end of this year.

## Room for Growth

Rendition, of course, is not standing still. The initial Verite is built in 0.6-micron CMOS by Chartered Semiconductor (Singapore) and measures 100 $mm^2$. The company is already working with a new fab to produce a 0.5-micron version that should reduce the die size and increase the clock speed from the current 50-MHz rating; this part could be available late this year. Rendition is negotiating with other fabs to secure 0.35-micron capacity, allowing further improvements.

Verite is designed using gate-array technology from Silicon Architects, making the design easily portable to a number of fabs. Gate arrays are typically larger than standard-cell designs, but Rendition claims its design would have been no smaller using the commercially available standard cell libraries that it evaluated. The company plans to redesign some of the data path by hand to further improve performance and reduce die size.

Shrinking the current design would also allow integration of additional functions. For example, the RAMDAC could be added, reducing the external chip count. Audio or video functions could also be added with some new software and possibly some minor changes to the RISC engine. Integrating new functions would make the chip more attractive for PC motherboards. Rendition has announced no plans for any of these specific improvements, but we expect at least some will be present in the next-generation device.

Finally, there are several longer-term options to improve performance. The throughput of the RISC CPU could be enhanced through proven techniques such as deeper pipelining, superscalar dispatch, and out-of-order execution. The pixel engine, which lags the performance of leading-edge PC rendering engines, could be enhanced to deliver 2–3× faster rendering. To support this rendering speed, the memory controller could be adapted for synchronous memory, and the PCI interface could be upgraded to the higher-bandwidth AGP.

## Rendition Stands Alone

The partitioning of Verite between a programmable CPU and hardwired logic appears to be a wise tradeoff, enabling Rendition to take advantage of both design styles. Among the early 3D entrants, Rendition stands alone in delivering the feature set needed for high-performance PC graphics: a combination of 2D, 3D, and video acceleration with an on-chip triangle setup engine. This combination has excited makers of 3D games and add-in cards, building initial market acceptance for Verite.

Several key game makers are working with Rendition. Initial titles will include Zone Raiders (Virgin Interactive), Quake (Id Software), Fury 3 (Terminal Reality), IndyCar II (Papyrus), Flight Unlimited (Looking Glass), Actua Soccer (Gremlin), and Krazy Ivan (Psygnosis). Microsoft is using Verite as a development platform for Direct3D and is also developing games for it. Creative Labs has selected Verite for the next version of its 3D Blaster card, and Number Nine will also supply add-in cards using Rendition's chip.

The fledgling chip vendor is developing a wide range of drivers for Verite, but not all will be available at first release. Rendition is first focusing on 2D drivers for Windows 3.1 and Windows 95 plus Direct3D. Because Direct3D was only recently released in beta form, most of the initial games are coded to Speedy3D, a DOS-based interface developed by Rendition. The company is also seeing interest from CAD software makers and is currently supporting them using the CGL and Heidi interfaces but not OpenGL. Rendition expects to deliver Windows NT drivers at a later date.

The company will have to work hard to maintain its edge. The 3D market is just in its infancy, and a number of big companies are jumping into the fray. Although Verite's feature set is unique today, other vendors are developing their own setup engines and integrated features. Verite's programmable architecture should give Rendition an edge in a rapidly changing market, but the small company will be challenged to keep ahead of the big-name players. M