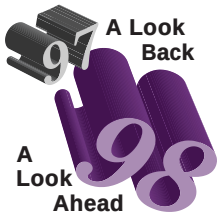
3D Vendors Prepare for Rough Seas in ’98

Intel’s i740 and Microsoft’s Talisman Will Make Waves and Sink Some Chips



by Peter N. Glaskowsky

For most of us, the Christmas season ended a few days ago. For 3D-chip vendors, however, the final Christmas presents were shipped more than a month ago. While Santa’s elves are enjoying some time off, workers in graphics-chip engineering departments are already hard at work on next year’s gifts, which must be wrapped up by spring.

The rapid pace of advancing technology led to some impressive new chips in 1997, such as Nvidia’s benchmark-leading Riva 128 and ATI’s flexible Rage Pro. We’ll see more of the same in 1998 as Intel introduces the eagerly awaited i740 graphics accelerator and new chips arrive from all of today’s major players in the PC 3D market.

S3, Intel on Collision Course

Even without a performance-competitive 3D engine, S3 still has the lion’s share of the graphics-chip market. The company earned and protects its position with top-notch applications engineering and driver development as well as a broad array of devices that allow OEMs to use S3 chips across entire product lines.

Starting early next year, S3 will face a competitor that already provides all of these elements to all of S3’s customers, but for a different socket—the CPU. Intel’s i740, formerly known by its code name Auburn, is expected to offer performance on a par with other mainstream chips, combined with high-quality rendering and high-quality driver software. By hitting S3 where it is weak as well as where it is strong, Intel is likely to stagger the graphics giant.

We don’t expect Intel to knock S3 out, however. S3 has at least one, and possibly two, high-performance 3D chips under development. The company’s recent Trio3D chip announcement also described a new 336-pin BGA package that S3 plans to use for faster, hotter chips. We expect to see a new conventional 3D chip in the spring, and possibly a second chip based on Microsoft’s Talisman architecture later in the year.

S3 will also retain several critical advantages over Intel, at least for the next year. Intel will have only one device, possibly with a few options in the area of memory size and type (SDRAM vs. SGRAM, etc.) to allow limited differentiation by its OEM customers. In contrast, S3’s 1998 lineup will cover a wide range of price points, with at least four desktop chips in separate families for consumer and business graphics, plus two mobile graphics accelerators. S3’s lead in the areas of driver quality and customer support is likely to last until 1999, though Intel clearly understands these areas and is likely to catch up quickly.

There’s also the power of the Intel name. That unique advantage, combined with Intel’s other strengths, should help Intel carve into S3’s market share immediately. Matrox is also likely to be hard hit, due to its shallow product line and rapidly declining reputation in the 3D market, and the company has already lost key people to Intel’s 3D group.

Other leading vendors will also be hurt by Intel’s attack, but to a lesser extent. ATI and Trident, for example, have strengths similar to S3’s but offer stronger 3D engines in all of their products. Nvidia and Rendition have also built good reputations in this market and will mount strong defenses against Intel in the form of new chips later in the year.

Capacity Constraints Likely to Worsen

Graphics-chip companies have enough to worry about without considering the imminent arrival of Intel in the mainstream 3D market. Most 3D chips are designed by fabless chip companies, and many of these have found themselves unable to get enough production capacity to meet demand, resulting in lost sales to the tune of millions of units. The best protection against this problem is establishing a strategic relationship—by means of cash payments or rights to 3D-chip designs—with one or more fab vendors, a plan that has allowed companies like 3Dlabs and Trident to meet demand in 1997.

Most shortfalls are less a technical problem than a failure of forecasting. Today’s 3D chips require the very latest fab technology, and few chip makers really believed they would be called upon to provide millions of 3D chips in 1997. This situation will only get worse in 1998 as mainstream chips move to even more severely constrained 0.25-micron technology. This time, however, fab vendors can’t say they weren’t warned.

Due to the long lead time for new fabs, we expect to see a repeat of the situation this year, where 3D vendors have had to turn away new business in order to protect their existing customers. The usual market response to limited supply in the face of high demand is increased prices, as we predicted earlier this year (see MPR 7/14/97, p. 3). We hope this prediction comes true, giving 3D vendors the R&D funds they need to design better and faster chips.

3Dlabs, ATI, Nvidia Do Well

Despite capacity constraints, 3Dlabs, ATI, and Nvidia had a good year in the mainstream market. All introduced new chips (Permedia 2, Rage Pro, and RIVA 128, respectively) that set speed records and established new standards for

ZDBOp Debuts 3D WinBench 98

The release of 3D WinBench 98 from the Ziff-Davis Benchmark Operation (ZDBOp) has confirmed Nvidia's performance leadership, but new quality tests have only heightened awareness of the shortcuts and approximations used by Nvidia and other leading chip vendors to achieve their impressive results.

The new benchmark package offers 41 quality tests, 25 of which are used to control whether particular operations are performed using the 3D hardware or instead emulated in software with a corresponding decrease in performance.

A total of 19 different scenes are rendered at a fixed 640×480-pixel, 16-bit color resolution to arrive at the overall 3D WinBench result. The frame rates for each scene are added together to produce the final result. If a chip can't complete a test correctly even with software emulation—not an uncommon problem—it receives a zero for that test. Tests may also be run at other resolutions and color depths to stress local-memory bandwidth and pixel-fill rates.

Results reported by *PC Magazine* cover a wide range of performance, from a score of 66 for a Matrox Millennium II PCI card to a high of 493 for a Diamond Viper V330 AGP card powered by the Nvidia RIVA 128. Few other PC subsystems show 7:1 ranges of features and performance; clearly, the choice of a 3D chip is one of the most important an end user can make.

Some disagreements have already surfaced over the interpretation of the new quality tests. Chip vendor 3Dlabs says Nvidia's implementation of several Direct3D features fails to meet the guidelines in 3D WinBench 98, and the company believes its own Permedia 2 passes more quality tests than it is being given credit for. If these decisions were made in 3Dlabs' favor, the company claims, Nvidia's score would drop to 360, and Permedia 2 would take over the top spot with a 479.

There is some merit to 3Dlabs' claims, though not as much as the company would like. The Ziff-Davis policy has been to accept implementations that work well on existing 3D software, and most of the contested features in the RIVA 128 do produce the expected results in the real world. While 3D WinBench is a more demanding test than today's games, new titles are likely to put more stress on the 3D subsystem. Other problems may be fixed by future driver updates.

One of the functions of quality-oriented benchmarks like 3D WinBench 98 is to highlight potential problems before end users find them, and we hope this debate will serve the same purpose.

More information on 3D WinBench is available from ZDBOp's Web site: *www.zdbop.com*.

visual quality. For markets where 3D performance was a serious consideration—mainly home entertainment and gaming PCs—chips from these vendors won most of the major OEM bundling agreements in 1997.

3Dlabs remains the only mainstream 3D vendor to succeed at the high end of the market, where the company's Glint family shared the spotlight with the 3DPro accelerator from Mitsubishi's Vsis subsidiary. 3Dlabs's Glint Gamma, eagerly awaited in the CAD market as the first geometry-processing chip capable of higher effective throughput than Intel's best CPUs, experienced significant delays. These problems forced the company to continue selling the less capable Delta setup accelerator with 3Dlabs' newest Glint MX rendering chip. Many customers have chosen to wait for the more potent Gamma/MX combination, now expected to ship in 1Q98.

Gamers Remain Loyal to 3Dfx

Perhaps the most surprising lesson from the busy 1997 season has been the persistence of 3Dfx's Voodoo products. The original Voodoo chip set was introduced more than two years ago but was so far ahead of its time that it still outperforms many of today's newest chips on 3D games, and it remains the preferred choice of the most avid PC gamers. Indeed, Voodoo has kept 3Dfx alive and kept competitors Quaking in their boots.

Voodoo's continued popularity is more remarkable considering its lack of 2D features. Most Voodoo cards must be used with a separate 2D or 2D/3D graphics card and a pass-through connection for the monitor cable. For game addicts, this is a small price to pay for access to the large base of 3D games written to take advantage of the high-speed Voodoo architecture.

The Voodoo family will receive its first speedup early next year with the release of Voodoo 2, a 192-bit AGP accelerator said to offer about three times the performance of Voodoo. Like Voodoo, Voodoo 2 will handle only 3D graphics, but users will be able to install a second Voodoo 2 card to double pixel-fill rates. Since AGP allows only one expansion card, VGA compatibility must be provided via a PCI card; the performance of this configuration on nongame software is likely to limit its acceptance among users who spend any significant amount of time in nongame applications. Similar limitations will apply to multicard Voodoo 2 configurations; the PCI bottleneck for additional cards will reduce the long-term scalability of this solution.

3Dfx hopes to reach a wider market with Banshee, a combined 2D/3D chip planned for mid-1998. The company has not characterized the expected performance of Banshee; we expect it to fall somewhere below Voodoo 2, dropping 3Dfx into the thick of the commodity 3D-chip fray—a market where the company has not seen much success.

3Dfx's main competition for the hearts and minds of PC gamers has come from Rendition. Rendition's Vérité V1000 (see MPR 5/6/96, p. 1) has enjoyed an unusually long

life for the same reasons that have kept Voodoo alive—good performance and a wide variety of game titles. Although Rendition's V2100 and V2200 experienced delays in engineering and manufacturing, arriving too late for most Christmas '97 OEM design wins, they rank among the top three 3D chip families in benchmark performance, and they have perhaps the second-best set of game titles after 3Dfx.

In spite of the delays and lost revenue, Rendition is in a good position to survive the coming shakeout in the PC 3D market, having received $22 million of private venture capital last August. These funds should allow Rendition to remain solvent and independent while it develops its next generation of 3D products, which we expect will include a substantial upgrade to its current Vérité architecture.

The PowerVR family, a cooperative effort of VideoLogic and NEC, is also aimed at the 3D game market. The PCX2, announced earlier this year, has achieved reasonable success in low-cost add-in cards from Matrox (where it helps to compensate for the poor native 3D performance of Matrox's other products) and other card vendors. Like Voodoo, the PCX2 lacks 2D support, but the PowerVR architecture is more difficult for software developers to support, limiting the number of titles available for PCX2 and reducing the chip's desirability for 3D gamers.

Earlier this year, NEC announced plans to combat both of these deficiencies by spending $16 million on software development at independent software vendors (ISVs) plus another $9 million for advertising and other promotional activities. This tactic will no doubt boost the market for PowerVR products, but probably not by more than a million units. As a result, NEC is spending roughly $25 per chip—roughly the same as PowerVR's selling price.

NEC has been highly successful in at least one market, however; PowerVR is said to have won a socket in the Sega 64 video-game system.

Cirrus's Future Looks Cloudy

While the fastest 3D chips can command relatively high prices, not all OEMs want or need leading-edge 3D chips. For more cost-sensitive systems, OEMs can choose from a number of second-tier products and vendors, but they will insist on second-tier pricing as well. Most 3D-chip vendors find themselves in this position, unfortunately; as dynamic as the 3D market is, it simply isn't large enough to support all the vendors currently competing for a share of it (see sidebar).

Some companies have already been caught in this squeeze. Cirrus Logic, once the clear market leader in the graphics-chip business, has been reduced to selling its top-of-the-line Laguna3D chip for less than $10, close to its manufacturing cost. Laguna3D-based cards are available for less than $60, suggesting that Cirrus's customers aren't making any money either. With little or no desktop revenue coming in, Cirrus has been unable to sustain the development effort for the Laguna 2 family. These chips, originally scheduled for

PC 3D Market Continues to Grow

Just since June, the PC 3D market has grown from 33 to 45 companies:

3Dfx (Voodoo, Banshee); 3Dlabs (Glint, Permedia); Accelerix (PhantASM); Acer Labs (M3151); Advanced Rendering Technology (AR250); Alliance (ProMotion, Cobra); ATI (Rage Pro); Avance Logic (ALG 27000); Chromatic (Mpact); Cirrus (Laguna3D); Cyrix (MXi); Dynamic Pictures (Oxygen); Equator; FourFold (LightWing); Fujitsu (Pinolite, Wizard); GigaPixel; Hitachi (Spherix); IGS (CyberPro); Intel (i740); Intergraph (RealiZm); IXMicro (TwinTurbo 128-3D); Matrox (Millennium); Microsoft (Talisman); Neomagic (MagicGraph); Number Nine (Ticket to Ride); Nvidia/SGS-Thomson (RIVA 128); Oak (Warp 5); Philips (Big Cats); PixelFusion (VPF/PixelFlow); Raycer Graphics; Real3D (R3D/100); RealVision (GA330); Rendition (Vérité); S3 (ViRGE); Silicon Graphics; Silicon Magic (Magic F/X); Silicon Motion (Lynx); Silicon Reality (Taz); SiS (SiS63x6); Stellar (PixelSquirt); Trident (3DImàge); TriTech (Pyramid3D); Tseng Labs (ET6300); VideoLogic/NEC (PowerVR); and Vsis (3DPro)

Gone from this list are several vendors who have faded from the scene. Artist, Sigma Designs, and Yamaha have ceased marketing efforts for their 3D-chip designs and do not appear to be working on new products. Cirrus Logic remains, though some observers believe the company has discontinued all 3D-chip development efforts.

Of the new additions to our list—Acer Labs, ART, Cyrix, Equator, FourFold, GigaPixel, Hitachi, PixelFusion, Raycer, RealVision, Silicon Graphics, and Silicon Motion—none are planning to compete directly with today's mainstream PC 3D makers. These companies all plan to use unique architectures or target specific market niches, staying out of the fight for desktop supremacy.

Clearly, those who expected the 3D market to collapse in 1997 have been disappointed—and we may not yet have reached the high-water mark. While total revenue from PC 3D chips will not exceed $3 billion in 1998, the overall 3D market is much larger. Additional income for 3D-chip vendors will come from 3D-capable video games and other consumer products as well as from an increasing share of the workstation market, which is now dominated by proprietary solutions.

Even these new applications will not be enough to support the dozens of companies listed here. Sooner or later, a shakeout is coming, and the arrival of Intel's i740 will cause some significant tremors.

Key 3D Events of 1997

Silicon Magic (*www.simagic.com*) announces the F/X256, a 2D graphics controller with 1.25M of embedded DRAM built by partner **Oki**. Eschewing the laptop market for such chips, the F/X256 is aimed at desktop systems (2/17/97, p. 5).

Cyrix's MediaGX (*www.cyrix.com*) integrates 2D graphics with a P133-class processor core and other peripherals (3/10/97, p. 1), later followed by the MXi with fast 3D processing and rendering features (12/8/97, p. 16).

ATI unveils the world's first 3D chip with a full-featured 2×-mode AGP interface, floating-point setup engine, and 100-MHz frame buffer (3/31/97, p. 15).

S3 *(www.s3.com)* responds with the Virge/GX2, which lacks a setup engine and provides only rudimentary AGP support (4/21/97, p. 5).

Rumors emerge about **Sega's** *(www.sega.com)* plans for a new video-game platform based on **Hitachi**'s SH-4 processor and **Microsoft**'s Windows CE with the DirectX multimedia extensions (6/2/97, p. 5), but **3Dfx** loses this design win to a chip based on **VideoLogic/NEC**'s PowerVR architecture (7/14/97, p. 5).

3Dlabs *(www.3dlabs.com)* debuts its Permedia 2 and describes Glint MX and Gamma at PC Tech Forum, while **Rendition** (*www.rendition.com*) previews the V2000 family and **Trident** announces new chips for desktop and mobile systems (6/2/97, p. 16) (6/23/97, p. 1).

Number Nine (*www.nine.com*) and **Nvidia** (*www.nvidia.com*) bring their chips to market; Nvidia's Riva 128 wins top honors on the new 3D WinBench (6/2/97, p. 16). **Rendition**'s V2000 family appears, and PowerVR's PCX2 does too (6/23/97, p. 1).

Trident's first and **NeoMagic**'s fourth embedded-DRAM graphics chips arrive, aimed squarely at the profitable laptop-graphics market (6/23/97, p. 5). Trident's chip provides 3D acceleration; NeoMagic's doesn't, but NeoMagic's experience ensures good sales anyway.

AGP graphics chips finally gain a home as **Intel**'s 440LX chip set ships (8/25/97, p. 4).

Siggraph sees new rendering architectures from **ART** and **Hewlett-Packard** (9/15/97, p. 9).

BOPS describes its ManArray DSP at the Microprocessor Forum, while **S3** announces its first embedded-DRAM graphics controller. **AMD, Cyrix** describe plans for host-based 3D-geometry acceleration (10/27/97, p. 4). Also at the Forum, **Centaur** tips its own 3D-geometry plans (11/17/97, p. 17).

S3 introduced 3D to its Trio family of business graphics chips, previously 2D only (11/17/97, p. 4).

ATI's Rage LT Pro claims mobile-3D performance lead (12/8/97, p. 5).

an August '97 announcement, are nowhere to be seen, and we understand that almost all 3D-chip engineering efforts at Cirrus have been discontinued.

Talisman Teeters But Doesn't Topple

The popular pick for an early death, Microsoft's Talisman architecture suffered a serious setback in 1997 with the cancellation of the original Talisman reference design—to which Cirrus was a major contributor. Talisman also, however, saw major progress toward long-term success in the form of public commitments from Trident and Fujitsu. Both of these companies plan to introduce single-chip Talisman renderers in the first half of 1998. S3 has also announced plans to offer a Talisman-derived graphics accelerator, but it is not clear how much of the Talisman architecture will be implemented in the new chip, or when S3 might release it.

Some reports have suggested ATI, Matrox, and Philips are interested in Talisman, but we do not expect to see Talisman-derived chips from any of these vendors in 1998. At most, they may offer support for Talisman-style texture compression or other minor features of the architecture.

In our original coverage of the Talisman announcement (see MPR 8/26/96 p. 5), we predicted that Talisman could not become successful without low-cost implementations, and that such products would not be possible before 1998. From this standpoint, the Talisman hardware effort is still essentially on schedule, though the cancellation of the reference design has delayed software development by about six months.

If Talisman trips up, other revolutionary architectures are waiting in the wings. Like Talisman, NEC's PowerVR and Oak's Warp 5 products include region-based rendering engines, boosting pixel-fill rates. Oak's part also improves visual quality by rendering the scene at a higher effective resolution, a technique known as subpixel antialiasing. Pixel-based rendering is used by Stellar's PixelSquirt and the PixelFlow architecture developed by the University of North Carolina, Chapel Hill. Hewlett-Packard's record-setting PxFl system is based on PixelFlow, and startup PixelFusion hopes to bring the same technology to PCs.

Some Vendors Thinking Outside the (PC) Box

Unconventional thinking marks the efforts of several of the new entrants to the PC 3D business. Advanced Rendering Technology (ART) has focused on ray tracing, a neglected but important style of 3D rendering. While moderate image quality is more easily and predictably achieved by the real-time polygon-oriented rendering algorithms used in mainstream 3D chips, the photorealistic images needed by the motion-picture industry are usually rendered by ray-tracing software. ART is attracting a lot of interest, as would any vendor promising to replace hundreds of expensive computer systems with dozens of inexpensive chips.

Cyrix hopes to counter the one-two punch of Intel CPUs and Intel 3D chips with a single sharp jab from its MXi

processor, a tightly integrated combination of CPU, graphics chip, memory controller, and PCI bridge. When the MXi debuts, it is likely to provide high-end performance at a low-end price—a popular combination in any industry.

Toward the end of 1998, we should see the debut of the first x86- and Windows NT-based "Visual PC" workstations from Silicon Graphics. Whether SGI is just expanding its business, or preparing for an eventual move to IA-64, SGI's new systems will come equipped with SGI's own 3D hardware, the company's most significant contribution to workstation design. These systems, and the 1999 debut of new Fahrenheit 3D APIs developed with Microsoft (see sidebar), will vault SGI into an immediate position of influence in the PC 3D market. SGI may not attempt to compete on price, but it will certainly deliver a system that offers leading-edge performance on real-world applications.

New Vendors, New Chips to Spice Up New Year

The most interesting announcements to come out of the 3D market in 1997 have also been the most tantalizingly vague. Several new vendors have cropped up, including Equator, FourFold, GigaPixel, and Raycer Graphics. We know only that Equator is working on a media processor of some kind (presumably based on the VLIW technology Equator inherited from Multiflow), FourFold plans a rendering engine based on parallel processing for PCs and embedded applications, GigaPixel's founders have a strong background in PC multimedia, and Raycer is *not* working on a ray-tracing chip. We cannot predict when, or if, these new companies will introduce products, but we expect to hear some details about these design efforts in 1998.

The mainstream market is much more predictable. While 3D vendors are notoriously tight-lipped about future products, we expect to see new chips next spring from all the major players. PC OEM schedules are not negotiable, nor are the rapidly increasing performance demands of OEMs as well as end users.

For 1998, new 3D chips must provide roughly four times the peak throughput of 1997 chips in the same market segment. High-end CAD solutions should be able to process about two million triangles per second in OpenGL applications, requiring hardware geometry acceleration. Mainstream chips running Direct3D software need to sustain about one million triangles/s to keep up with the geometry-processing capabilities of Intel CPUs. New processors from AMD and Cyrix may raise this bar, providing some demand for even faster 3D engines. Even low-end chips in 1998 must provide as much performance as 1997's midrange products—and will, since today's midrange chips naturally become tomorrow's low end.

Despite increased competition from the experienced teams at 3Dlabs, ATI, and Nvidia, and new threats from industry giants Intel and SGI, 3D-chip vendors must continue to meet their customers' expectations on their customers' schedules—or accept the consequences. M

SGI, Microsoft Ink 3D API Pact

Resolving many months of controversy over the relative merits of OpenGL and Direct3D, SGI and Microsoft have agreed to work together on Fahrenheit, a pair of 3D application programming interfaces (APIs) to be included in future versions of Microsoft Windows and SGI's Unix-based Irix operating system.

Fahrenheit's new low-level API combines the capabilities of OpenGL and Direct3D as well as 3D-rendering technology being developed by Intel to provide optimal performance on the Pentium II processor.

Another result of the partnership will be a new high-level API based on SGI's current Performer interface and HP's DirectModel. Working directly with 3D scene descriptions (scene graphs), the new API will make it easier for developers to add 3D capabilities to application software. Low-level APIs like OpenGL and Direct3D are little more than thin insulation between applications and 3D-chip registers. Software must still produce long lists of polygon-vertex coordinates, an unnecessarily complicated task for developers who would rather focus on abstract goals like modeling the appearance of a new automobile.

Separate extensions to the scene-graph API will support large-model visualization applications with features such as complex geometric modeling, automatic tessellation, occlusion culling, and other techniques to simplify the design and streamline the rendering of 3D scenes with millions of polygons.

While we are pleased that the wrangling over low-level 3D APIs is now all but over, Fahrenheit's high-level APIs will be of greater long-term significance to the industry. SGI and Microsoft have set their priorities accordingly; the high-level API will be released first, in the first half of 1999. This first Fahrenheit release will run on top of OpenGL on both Windows and Irix. The new low-level API will appear in 1H00 with the second revision of the high-level APIs, but the low-level API will be available only on Windows. While SGI and Microsoft discussed porting the low-level DirectX APIs to Irix, no agreement was reached. Consequently, Irix implementations will continue to use OpenGL, somewhat diminishing Fahrenheit's cross-platform appeal.

SGI declined to comment on the implications of this arrangement, but it may suggest the company is planning to move its entire product line to Windows NT—presumably a 64-bit NT designed around Intel's IA-64 processor architecture—where limited compatibility in Irix is of little importance. Even if DirectX were available for Irix, developers would still be required to accommodate significant differences in other APIs to port applications from NT to Irix or vice versa.