MICROPROCESSOR REPORT
THE INSIDERS' GUIDE TO MICROPROCESSOR HARDWARE
VOLUME 12, NUMBER 2
FEBRUARY 16, 1998

Intel Displays 740 Graphics Chip

Auburn Sets New Standard for Quality—But Not Speed

by Peter N. Glaskowsky

After more than a year of waiting (and a great deal of anxiety on the part of current 3D vendors), the Intel740 graphics controller has finally arrived. Unlike the '93-vintage i750, the 740 is sure to be a strong contender in the desktop graphics market. The new chip is no faster than today's best 3D accelerator, but the 740's superior visual quality earns it higher overall test scores on 3D WinBench.

The 740—code-named Auburn—was codeveloped with Chips & Technologies and Real 3D. It turns out to be a fairly conservative design with a few interesting new features, including a tile-based rendering engine. Figure 1 shows the internals of the 740, which looks like most other mainstream 3D chips. The 740's price and performance also show it is aimed squarely at the mainstream market.

As expected, the 740 relies heavily on the Intel-developed Accelerated Graphics Port (AGP) to reduce the chip's local memory requirements and the overall system cost. This reliance goes well beyond that of competing chips, however: the 740 is not available in a PCI-compatible version (though Real 3D created an AGP–PCI bridge chip to enable 740-based PCI expansion cards).

The 740 will face stiff competition from more advanced 3D engines this spring. We expect to see parts with higher performance and better visual quality, but the 740 will have at least one advantage over these chips: the Intel brand.

Tile-Based Rendering Without Software Penalties

Rumors about the 740's architecture and performance have been flying around the industry ever since Intel announced plans for the product. The most intriguing claim was that the 740 used a tile-based rendering architecture, more like VideoLogic/NEC's PowerVR or Microsoft's Talisman than conventional 3D chips.

This rumor was only partially correct. By adding a relatively simple enhancement to the conventional 3D pipeline, Real 3D and Intel achieved many of the performance and memory-bandwidth advantages of tile-based rendering while preserving full compatibility with the conventional Direct3D and OpenGL software interfaces.

Like a conventional 3D chip, the 740 can receive polygons from the host processor in any order. Instead of breaking polygons into horizontal strips of pixels, one for each scan line, the 740's setup engine produces a list of the 4×4-pixel tiles that overlap the polygon.

All of these tiles are rendered before the 740 proceeds to the next polygon. This approach, which can be described as polygon-order dispatch with tile-order rendering, significantly increases the locality of memory references for frame-buffer and texture fetches, allowing Intel to get good results from a relatively small 256-byte texture cache. (Other 3D chips today have much larger texture caches; the 8K cache on Nvidia's RIVA 128, for example, occupies about 15% of its die area.)

AGP Boosts Available Bandwidth

The 740 follows Intel's published recommendations for achieving good performance on AGP graphics cards. Polygon

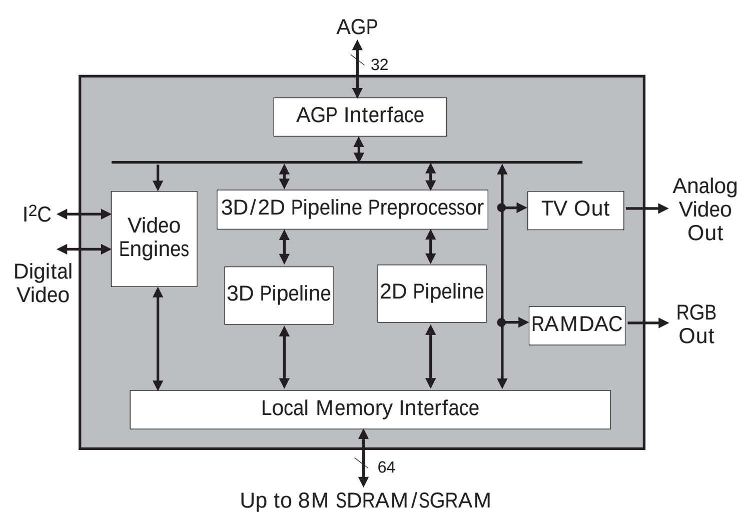

Figure 1. The Intel740 includes separate function units for 2D and 3D processing, video, and RGB output.

Inside: ColdFire 5307 ◇ Lexra 4080 ◇ M•Core ISA ◇ Compaq Buys Digital information from the application is buffered by the driver and processed in small batches, typically 128 to 512 triangles at a time. The host CPU performs geometry and lighting calculations, as required by the application and the 3D API, and writes the resulting vertex data directly to the region of main memory used for AGP accesses. The Windows 95 memory manager configures AGP memory to be noncachable to eliminate cache snoops on AGP transfers, so CPU writes to AGP memory use the write-combining mode to generate burst transfers on the CPU bus.

(The 740 can also use cachable system memory for texture storage, but software developers should avoid using this mode. An AGP fetch from system memory will generate a snoop request to the CPU's L2 cache, during which no CPU or main-memory activity can take place.)

Once the vertex data is in AGP memory, it can be fetched by the 740 in even smaller batches (typically two vertices) and loaded into the 16-vertex on-chip rendering queue. During rendering operations, the 740 can maintain two open pages in local memory (for the color and Z buffers) plus two pages in AGP memory (for vertex and texture data), as Figure 2 shows. With two pending transactions to local memory and up to 164 transactions in the AGP pipeline, the 740 can achieve close to the peak bandwidth of both memory resources under ideal circumstances.

This ideal may be unachievable on current systems, however. While the peak bandwidth of AGP memory is 533 Mbytes/s, this memory is shared with the host processor, and the 740 will rarely see more than 200–300 Mbytes/s on real applications. The 440BX chip set (see MPR 12/29/97, p. 4) will boost main-memory peak bandwidth by 50%, allowing AGP to more closely approach its design target.

Intel may be paying too much attention to its own advice on AGP. Unlike all competing AGP 3D chips, the 740 cannot use local memory for texture storage, and it must use AGP memory for texture maps. Fortunately, the chip demands relatively little AGP bandwidth, no more than about 170 Mbytes/s for command and texture transfers combined.

Like a CPU, the 740 has an on-chip cache to reduce bandwidth demands on main memory. Intel says the 740's tile-based rendering architecture permits high hit rates despite the cache's small size, just 256 bytes. While the 740's cache may satisfy requests for the same texture map on successive polygons, it is not large enough to retain texture data from one frame to the next. Chips that can store or cache texture data in local memory can avoid fetching the same texture data repeatedly for each new frame and leave more main-memory bandwidth for the host processor.

The 740's reliance on AGP texturing also explains Intel's decision to support only 8M of local memory. With no need (or ability) to use local memory for texture storage, 8M will support 3D rendering at 1280 × 1024 resolution with 16 bits per pixel for the front and back render buffers plus 16 bits for the Z buffer. The 740 provides no support for true-color (24 bits per pixel) 3D rendering. This mode is rarely used today aside from CAD software, but true-color graphics are noticeably superior to 16-bit images. True-color rendering is likely to become more popular in the future as 3D accelerators offer it as an option for quality-conscious users.

Design Stresses Quality, But Only to a Point

Intel has chosen to emphasize the visual quality of the 740, and the part certainly does produce good-looking displays. The 740 was designed to achieve peak performance with the most commonly used quality features enabled, including perspective correction, fog, bilinear texture filtering, specular lighting, and Z (depth) buffering. The chip omits features found on some 1997 3D chips from other vendors, however, and we believe the 740 will be be surpassed by many 1998 products.

The 740 supports the commonly used bilinear texture-filtering algorithm, computing a weighted average of four points sampled from a single texture map for each pixel to be displayed on the screen. The chip can choose which texture map to use for each pixel on the screen, producing better results than competing chips like the RIVA 128 or ATI's Rage Pro that make this selection only once for each triangle. Per-triangle MIP-mapping causes inconsistencies in the appearance of an object depending on its distance from the viewpoint; per-pixel MIP-mapping, as found on the 740 and 3Dlabs' Permedia 2, produces a smoother result.

An even better option is trilinear filtering, which combines values from two texture maps for each pixel. Trilinear filtering is not always desirable, because it requires the 3D chip to average eight texture samples instead of four and can reduce pixel-drawing throughput. The extra effort produces better visual quality, eliminating seams between texture boundaries.

The 740 does not support trilinear filtering. Instead, it uses a new technique to approximate the results of trilinear filtering with the lower bandwidth and computational demands of bilinear filtering. Known as level-of-detail (LOD)

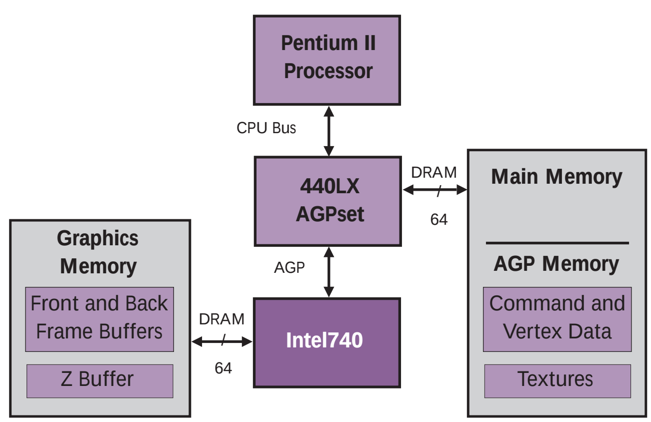

Figure 2. The Intel740 can issue simultaneous requests to local memory as well as to the AGP region of the system's main memory.

dithering, the new approach adds a pseudo-random offset to the texture-map selection criteria for each displayed pixel to dither (blur) the edges between texture maps. The 740 automatically substitutes LOD dithering whenever an application requests trilinear filtering; it isn't really the same thing, but since today's APIs don't distinguish between the two techniques, we can't fault Intel for this choice.

The 740 also supports LOD biasing, which allows the user to bias LOD selections toward coarser texture maps (for improved performance) or finer maps (for improved visual quality).

We expect to see trilinear support on almost all of the other 3D chips introduced this year, and some are likely to provide a higher-quality option known as anisotropic filtering. This alternative, first brought to the attention of the PC graphics market by Microsoft in its Talisman announcement (see MPR 8/26/96 p. 5), uses even more texture samples—typically 16 in these designs—to produce a smoother, more realistic display. While the 740's dithered bilinear option will produce reasonable results for all 3D applications, other chips with true trilinear and anisotropic filtering will produce displays that are markedly superior, giving the 740 a slight competitive disadvantage.

Intel also equipped the 740 with a better approach to Z buffering than most of today's chips use. Instead of common linear 16-bit Z buffering, the 740 uses 16-bit logarithmic values. This provides extra depth resolution for objects that are close to the view point, where depth errors are most conspicuous. Other vendors are moving toward 24-bit or 32-bit Z buffering, but the 740's solution is adequate for now.

Some Advanced Quality Features Are Absent

While the 740's design should yield good performance on today's mainstream applications, it lacks some features that are needed by today's CAD tools and others that are becoming widely used in 3D games.

With its four memory-access pipelines, the 740 can fetch vertex and texture data while rendering into color and Z buffers. This is basically the minimum requirement for 3D rendering, but more sophisticated rendering schemes are used in some OpenGL-based CAD applications. These programs make use of alpha buffers (which manage object translucency) and stencil buffers (which support graphics overlays). The 740 can use an alpha buffer only when Z-buffering is disabled, and it has no support for stencil buffers. The chip handles a special case of alpha blending in conjunction with Z buffering—some 3D games render all the scene's opaque objects first, then render translucent objects in a second pass. The 740 supports this technique, known as sort-dependent or source alpha blending.

Alpha and stencil support and similar CAD-oriented features are likely to migrate into mainstream applications. Stencil buffers, for example, provide an efficient way to implement control panels and status displays in 3D games. True alpha buffering is also a desirable feature in 3D games, where it simplifies the display of natural objects like trees, clouds, and water.

While the 740 lacks some advanced quality features, it meets the industry's de facto minimum quality requirements and should be compatible with most mainstream 3D applications released in 1998. Given the company's emphasis on image quality, we expect Intel to include the full range of quality features in the 740's successor in 1999.

Performance Just Matches Last Year's Best

The 740's designers say the chip's setup engine can process triangles at a sustained rate of about 600,000 triangles/s, roughly 40% faster than necessary to keep up with a 300-MHz Pentium II. This gives the 740 the ability to accommodate more complex 3D scenes when used with the faster Pentium II processors expected later this year. Other 3D chips have even more performance headroom, a feature that may prove valuable with non-Intel processors expected later this year such as AMD's K6 3D (see MPR 10/27/97 p. 19) that will offer much higher 3D geometry performance.

Faster CPUs can't improve a 3D chip's pixel-fill rate, however, and this is where the 740 fails to meet expectations. The 740 runs internally at 66 MHz and can render one pixel per clock, yielding a peak rate of 66 Mpixels/s. This is somewhat lower than the peak rates of some 3D chips, but Intel says the 740's tile-based architecture allows it to render at a sustained rate of 60 Mpixels/s, faster than the current speed king, the RIVA 128.

Tests using the Ziff-Davis 3D WinBench program cast doubt on this claim. Using a 333-MHz Pentium II system, our tests show the 740's rendering throughput is slightly lower than

Price & Availability

The Intel740 graphics accelerator is packaged in a 468-contact BGA package and consumes 5.8 W (maximum), requiring a heat sink for proper cooling. The chip is priced at $37.50 in 10,000-unit quantities. Graphics cards and motherboards using the 740 will be available by April from a number of vendors, including Diamond, Real 3D, and STB. Real 3D will offer 4M and 8M AGP cards with retail prices of $189 and $249. Real 3D will also offer PCI configurations, starting at $229 for a card with 4M of frame buffer on the 740 plus 8M of texture memory on Real 3D's AGP–PCI bridge ASIC. The other board vendors have not yet announced pricing or configurations.

More information on the 740 is available from Intel at *www.intel.com*. Real 3D's Web site may be found at *www.real3d.com*. Information on Diamond's 740-based graphics cards is available at *www.diamondmm.com*. 3D WinBench is available online from the Ziff-Davis Benchmarking Operation, *www.zdbop.com*.

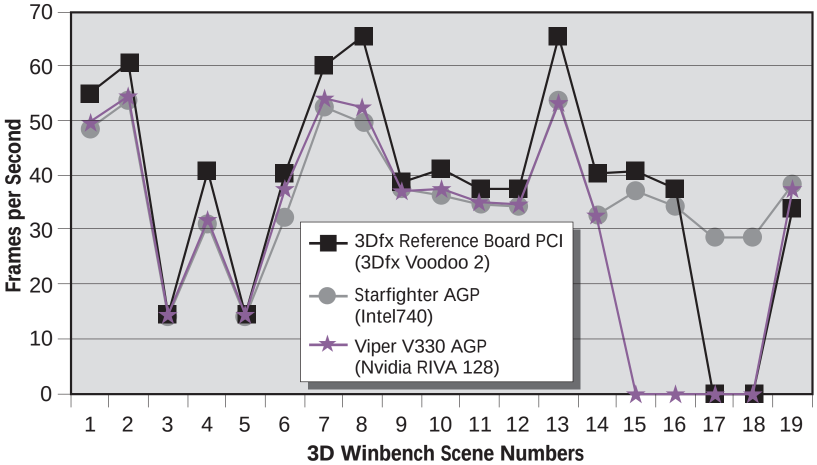

Figure 3. The Intel740 falls behind 3Dfx's Voodoo 2 and Nvidia's RIVA 128 on 3D WinBench scenes that all three can complete. (Source: MDR)

RIVA's, which is rated at 100 Mpixels/s peak but only about 50 Mpixels/s sustained. Figure 3 breaks down the 3D WinBench scores of the 740 and the RIVA 128. The 740's overall score of 691 is higher than RIVA's 576, but only because RIVA gets no score on the four scenes in 3D WinBench that require trilinear filtering or MIP-map blending capabilities. On 14 of the 15 scenes they can both perform, the RIVA 128 is faster than the 740, by about 2% on average.

Table 1 compares the 740 to other leading 3D chips using overall 3D WinBench scores. The 740 completes all 19 of the test scenes and achieves the highest overall score. The second-fastest chip capable of handling all 19 scenes is Chromatic's Mpact 2, which is about 37% slower overall.

While driver refinements are sure to improve the 740's benchmark scores over time, it seems unlikely that the 740 will retain its 3D WinBench performance lead after the arrival later this year of next-generation 3D chips from 3Dlabs, ATI, Nvidia, Rendition, and others, given that these chips are likely to be two to four times faster than current 3D accelerators. We expect to see chips that render multiple pixels per clock at clock rates of 100 MHz and more. Many of these competing chips will use faster, wider local memory than the 740, providing the necessary bandwidth to support these faster rendering engines.

Board (Chip, RAM)	Scenes Passed	Resolution 640 × 480	800 × 600
3Dfx Reference Board PCI (3Dfx Voodoo 2, 4M+8M)	17	724	668
Real 3D Starfighter AGP (Intel740, 8M)	19	691	583
Diamond Viper V330 AGP (Nvidia RIVA 128, 4M)	15	576	501
ATI Xpert @ Play AGP (ATI Rage Pro, 8M)	17	529	446
Chromatic 3DVD PCI (Chromatic Mpact 2, 8M)	19	436	329

Table 1. On 3D WinBench, the Intel740 scores higher overall than all other single-chip 2D/3D accelerators. 3Dfx's forthcoming Voodoo 2 has a better score (despite completing just 17 of the 19 possible scenes), but it is a 3D-only three-chip set. (Source: MDR)

Intel Is API-Neutral

Direct3D isn't the only API of importance to the PC market. Intel has been careful to provide an equivalent level of driver support for OpenGL, which is currently used mostly for professional tools such as CAD programs.

Intel and Real 3D worked together to develop reference implementations of both Direct3D and OpenGL drivers for Windows 95, Windows 98, and Windows NT 4.0/5.0. Intel is providing these reference drivers to all of its customers for the 740. These are actually release-quality drivers, since some 740 customers do not have driver-development resources of their own. Other customers, including the division of Real 3D that will be selling the 740 in board-level products, have done significant work to the reference drivers, boosting performance and improving compatibility.

The feature set of the 740 is closely aligned with that of Direct3D 5.0. Features specific to OpenGL, 3Dfx's Glide API, or version 6.0 of Direct3D (due later this year) are generally not present on the 740. This will also place the part at a disadvantage compared with products like 3Dfx's Voodoo 2 chip set, which was designed for Glide- and OpenGL-based games and thus includes the alpha, stenciling, trilinear texture filtering, and multitexturing capabilities lacking in the 740.

Preliminary tests suggest that Voodoo 2 will be substantially faster than the 740 on Glide and OpenGL games, and it might also have an edge on DirectX titles. Voodoo 2 boards will be significantly more expensive than 740 boards, however, and must be used in conjunction with a separate 2D graphics card, preventing Voodoo 2 from having much of an impact in the volume desktop PC market.

Digital Video Support Also Provided

The 740 provides a fairly typical set of features to support digital-video applications, including DVD decoding. The part can allocate a separate hardware video-overlay window in local memory, up to 720 pixels wide by 480 high. The video window supports RGB and YUV pixel formats and includes rescaling logic. Video can be displayed at full-screen resolution or in a window. The scaling logic includes a three-tap FIR filter for horizontal resizing, while vertical resizing can be performed using line replication, smoothing, or a running average.

The chip also includes a video-capture port compatible with the industry-standard Video Module Interface (VMI) specification. This port is primarily meant to be used with an off-chip DVD decoder. This two-chip solution is necessary for full DVD compatibility on current CPUs, since the 740 does not include motion-compensation logic. Pure software

DVD solutions for the 740 will have to wait for the 350-MHz Pentium II with 100-MHz SDRAM, due in April, which should be able to perform software motion compensation at full speed.

The 740 also contains the necessary logic to support Intel's Intercast technology, used by some television broadcasters to send digital data such as Web pages over the TV signal during the vertical blanking interval (VBI). The 740 does not perform VBI decoding, but it passes VBI data to the host CPU for further processing.

Vendors to Offer Variety of 740-Based Products

Diamond, STB, and Real3D will make AGP cards using the 740, and these are likely to be popular in new systems using Pentium II processors. AGP texturing will allow these vendors to offer a variety of configurations with as little as 2M of local memory for low-end products, though 4M cards are likely to achieve the highest unit volumes.

To serve the large installed base of PCs without AGP, Real 3D designed an ASIC that bridges from a system's PCI bus to a local AGP interface and provides a separate memory interface for local texture storage, eliminating the need to use PCI protocols to access system memory—a slow process due to PCI latency and the need to snoop every access for cache coherency. The AGP driver for Direct3D, developed by Microsoft and Intel, allows the Real3D solution to maintain full software compatibility with AGP cards. Real 3D's PCI–AGP bridge, the R3D-040, supports its own 8M 16-bit 66-MHz SDRAM channel, providing 133 Mbytes/s of peak bandwidth, more than enough to satisfy the 740's demand for AGP textures.

The other key market for the 740 is integrated graphics on PC motherboards. Intel will use the 740 on some of its own motherboards, as will Asus and other vendors. To make these designs more attractive, Intel has solved one of the critical technical hurdles for motherboard AGP implementations.

The original AGP specification effectively eliminated the potential for end-user upgrades by prohibiting motherboard designs with an AGP graphics chip plus an AGP expansion slot. Intel is now offering guidelines to enable what it calls "three-way" AGP, where a motherboard is equipped with an AGP graphics chip that can be disabled by installing an AGP expansion card, allowing the end user to upgrade the system's graphics performance later.

This isn't an ideal solution. Windows NT already supports multiple-monitor configurations, and Windows 98 will follow suit, so it would have been better to allow both the chip and the card to drive separate displays. The three-way configuration, however, meets the basic OEM need to advertise an upgrade option.

Marketing the 740 Should Be Easy

Intel should have no trouble persuading OEMs and end users to buy the 740, even at the chip's higher-than-average volume price of $37.50. Intel already has the most powerful brand name in the PC hardware business, and we expect it to advertise the 740 heavily to help build brand awareness for Intel in the PC graphics business.

Intel's marketing efforts will also make end users more sensitive to 3D performance and quality, boosting demand not just for the 740 but for chips from Intel's major competition. The best 3D-chip vendors will actually benefit from heavy advertising for the 740 if it expands the market instead of just stealing their customers.

Intel will also stress its ability to deliver high-quality device drivers for the 740, provide top-notch technical support to its customers, and meet any reasonable demand for the part. Even though Intel is effectively new to the graphics business, it won't take the semiconductor giant long to establish credibility in these areas.

Barring any unexpected problems with the 740, Intel should be able to sell as many 740s as it is willing to make. This is not a minor point. According to the MDR Cost Model, the 131-mm^2 0.35-micron 740 costs $43 to manufacture, counting fab amortization. If we assume that Intel has already amortized its 0.35-micron fabs using its Pentium line, the estimated cost of the 740 drops to $30. A Pentium II, by comparison, costs almost three times as much to manufacture but sells for at least six times the price. Demand for the 740 is likely to outstrip supply at first as Intel reserves most of its 0.35-micron fab capacity for CPU production. Even so, Intel should be able to sell 5–10 million 740s during 1998, gaining it 10%–20% of the market for 3D chips.

If Intel makes a strategic decision to shift 0.35-micron wafers away from the more profitable Pentium II, the company could possibly make 25 million graphics chips this year, or roughly half the total 3D-chip market. Such a decision might promote wider use of Pentium II, since most 740's will be used in AGP configurations that are not compatible with Intel's Socket 7 designs. The deciding factor may be the speed with which Intel can ramp its 0.25-micron fab lines. If this transition happens quickly, as we expect, Intel may use the 740 to keep its older 0.35-micron factories busy.

The next version of the 740 should arrive in 1H99 with roughly four times the pixel and triangle performance. Intel will presumably make the newer chip on 0.25-micron fab lines, where it will compete even more directly with CPU production. The 740's successor will likely be teamed with Intel's Katmai processor, which will include instruction-set extensions to enhance 3D processing throughput. This will be a potent and popular combination; marketed along with Intel's core logic, it will make life much more difficult for competing CPU and graphics-chip vendors.

The 740 has the feel of a "Version 1.0" product: competently designed but without the refinement that comes from market feedback over successive generations. The 740's most important role may be to establish Intel as a serious vendor of graphics chips, but the company will have to do better if it hopes to dominate this highly competitive market. M