Graphics Introductions Accelerate

Memory Architecture Differentiates Six New Controllers

By Dean McCarron

At the recent MicroSystems Forum, Cirrus Logic and Weitek unveiled new graphics accelerators that push the boundaries of graphics performance. In addition, Chips and Technologies, Oak, and Tseng Labs also have new accelerators with unique, powerful memory-interface designs.

While past accelerators from these companies have competed using impressive feature sets and powerful graphics engines, a new competitive metric is emerging as each attempts to address the heart of the graphics performance problem: display-memory bandwidth.

Although having a fast graphics engine and bit-block-transfers (BitBLTs) certainly boosts performance, the ultimate determination of graphics performance is the available bandwidth between the graphics engine and the display RAM. The fastest graphics engine in the world is limited by the bandwidth that remains in the graphics memory system after display refresh has taken its share.

Recognizing the bandwidth limitations of older parts, each of these companies has made a break from past designs to pursue the Holy Grail of CPU-to-display bandwidth. Remarkably, virtually every manufacturer chose a different solution to the bandwidth problem. This is most likely the result of differences in both design experience and design goals.

Even as the manufacturers are innovating in memory architectures, their graphics engines are converging upon a common feature set with very little differentiation. All of the new controllers support BitBLTs, color expansion, all 256 Windows raster operations, hardware cursors, line drawing, and a host of other graphics features that have become standard in all but the lowest-cost graphics accelerators. While a few of these features improve performance in certain situations, their performance impact is minor compared to that of raw memory bandwidth.

Alpine Offers Integrated Performance

Cirrus' newest GUI accelerator, the CL-GD5434, is the first member of a new, high-performance family Cirrus has named "Alpine." The new chip is faster than any other Cirrus part to date, and is the company's first entry into the high-end controller market. All devices in the family, including the GD5434, will share the same pin-out in a 208-pin PQFP package, and Cirrus expects the same drivers and BIOS code to operate the entire Alpine product range. These devices will differ in the widths of frame buffer and CPU interface buses.

The GD5434 supports both PCI and VL-Bus interfaces. Like Cirrus' other products, it also has an integrated clock synthesizer and a 24-bit color RAMDAC (triple 8-bit DAC). Unlike these products, the GD5434 also has an additional 8-bit "alpha" channel that can be used for color-keyed overlays and has wide application—such as video special effects and titling—in multimedia systems.

Taking the Wide Road

The width of the frame-buffer memory interface is especially critical in the Cirrus part. The company opted to remain with traditional DRAM instead of pursuing a VRAM-based solution. While consistent with the overall product focus—affordable mid-range graphics applications—staying with the 32-bit DRAM interface used by past parts would have limited performance.

Faced with the decision of sticking with DRAM but needing to significantly increase display bandwidth, Cirrus had but two options: a 32-bit interleaved memory system or a 64-bit memory system. Cirrus felt that a 32-bit interleaved design would require three clocks per access to meet standard DRAM timing requirements, while a 64-bit non-interleaved design could return data every two cycles. It thus chose the 64-bit design for its 50% greater bandwidth.

Implementing a 64-bit memory with standard DRAMs results in a granularity problem: the minimum 64-bit design using four $256K \times 16$ DRAMs results in a display buffer of 2M. This is a significant cost disadvan-

tage when compared to today's omnipresent 1M graphics controllers, but all of the new controllers suffer from the same 2M granularity problem except the C&T DGX. The GD5434 does support a 32-bit-wide, 1M memory option, but at the cost of losing virtually all of the performance enhancements of the 64-bit architecture.

Based on a 2M configuration, the GD5434 effectively offers twice as much memory with the same cost as a VRAM design, because VRAM costs about twice as much as DRAM. The 64-bit DRAM implementation offers some bandwidth advantages over a 32-bit VRAM configuration as well. The 64-bit DRAM interface offers approximately 180 Mbytes/s of bandwidth; after servicing a 256-color display at a 1024 × 768 resolution with 72 Hz refresh, 120 Mbytes/s of bandwidth remains for the CPU and graphics engine, as shown in Table 1.

In contrast, a 32-bit VRAM memory system offers approximately 90 Mbytes/s of bandwidth to the CPU and graphics engine through its parallel port. The serial port handles display refresh and hence places no load on the parallel interface, so the full 90 Mbytes/s is available to the CPU. While a DRAM implementation's CPU bandwidth depends on the display's color depth and resolution, in the modes most common to Windows, a 64-bit DRAM design offers about 33% more bandwidth than a 32-bit-wide VRAM design.

The added memory also enables the GD5434 to support additional true-color modes, and with the concurrent improvement in memory bandwidth, those modes will operate with sufficient performance to be usable. (DRAM-based implementations from most manufacturers offer truly glacial performance in true-color modes because display updates saturate the available memory bandwidth.)

Big Engine

The 64-bit memory interface offers significant

Cirrus Logic's Bill Chu discusses the new Alpine accelerator at the MicroSystems Forum.

MICHAEL MUSTACCHI

bandwidth improvements. Merely increasing memory bandwidth, however, can cause the controller to become bandwidth-limited someplace else. Within graphics controllers, the most bandwidth-intensive operation is BitBLT, which moves blocks of memory at the highest rate possible. A 32-bit BitBLT engine connected to a 180 Mbytes/s memory subsystem would represent, in the words of Weitek's Allen Samuels, an "impedance mismatch." Rather than shifting the bandwidth limitation from the memory interface to a 32-bit BitBLT engine, Cirrus chose to implement a 64-bit BitBLT engine in the GD5434, thus providing high bandwidth throughout the device.

The GD5434 is intended for desktop systems, but it sports a number of power-saving features intended to help manufacturers of so-called "green" or Energy-Star PCs *(see 070905.PDF)*. When the software of an Energy Star PCs enters sleep mode after a period of inactivity, the GD5434's internal DAC may be shut down, and clocks for the graphics subsystem may be reduced in frequency to lower power consumption. The device also supports the VESA display power-management signal (DPMS) interface *(see 070905.PDF)*, which can be used to place DPMS-compliant monitors into a power-saving mode at the same time that the system goes to sleep.

Power 9100 Has Bandwidth to Spare

At the extreme high-end of the new graphics controllers is Weitek's Power 9100, which was announced at the MicroSystems Forum. While the device has an interesting graphics engine that includes some more advanced features such as polygon fill, the Power 9100's true claim to fame is its memory architecture.

Weitek's solution to the bandwidth problem extends the high-end of memory architectures for PC graphics. Weitek implements a 32-bit interleaved design but uses VRAM instead of DRAM. An on-chip delay line for the
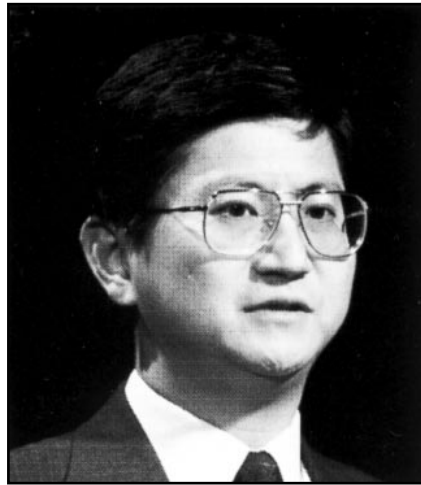
Controller	Memory Architecture	Graphics Engine Bandwidth	Max PCLK	Max MCLK	RAMDAC	Frequency Synthesizer	Price
Cirrus Logic GD5434	64-bit DRAM	120-160 MB/sec	85 MHz	60 MHz	24-bit	yes	$39
Oak OTI-107	64-bit DRAM	120 MB/sec	110 MHz	60 MHz	none	no	$55
Tseng W32i/p	32-bit interleaved DRAM	115-170 MB/sec	85 MHz	50 MHz	none	no	$25
Weitek Power 9100	32-bit interleaved VRAM	200 MB/sec	170 MHz	50 MHz	none	no	N/A
Chips & Tech Wingine DGX	32-bit DRAM with "Cache"	40 MB/sec	80 MHz	72 MHz	24-bit	yes	$29

Table1. Architectures and bandwidth available at a benchmark resolution of 1024 × 768 × 8-bit color with a 72-Hz refresh. PCLK is the maximum pixel clock rate. MCLK is the maximum memory clock rate. Note that the bandwidth listed is for an average memory subsystem using 70-ns DRAMs (except for Weitek's VRAM design). In some cases the device's memory clock can be run faster to increase bandwidth. Higher-speed memories can also be used, resulting in higher bandwidths for the DRAM-based controllers.

VRAM control signals allows fine timing adjustments and enables the VRAM timing to be pushed right to the specifications, maximizing the bandwidth that can be obtained from the memory.

As a result of these timing adjustments, Weitek achieves a stratospheric bandwidth of 200 Mbytes/s for the memory system. While other controllers mentioned here claim memory bandwidths in the 200 Mbytes/s range, because they are DRAM implementations, about 70 Mbytes/s of that bandwidth is taken for display refresh for a typical $1024 \times 768 \times 8$-bit color mode. In high-color and true-color modes, even more bandwidth is used for display refresh.

Weitek's VRAM memory system doesn't suffer from refresh bandwidth loading, since the VRAM's serial port services the display. The net result is that the full 200 Mbytes/s is available to the CPU and graphics engine. This is significantly more bandwidth than any other accelerator.
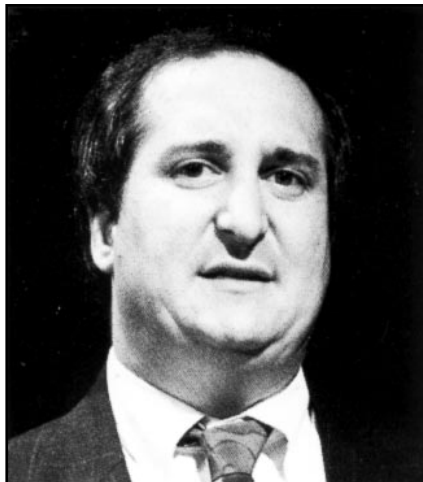
This level of performance exceeds that of PCI—the fastest local-bus standard to date—which has a maximum bandwidth of 132 Mbytes/s. In a dumb frame-buffer graphics system, the remaining 68 Mbytes/s of bandwidth would be wasted. The 9100, like most graphics accelerations, has an internal engine that's designed to take advantage of such a large bandwidth.

Allen Samuels, of Weitek, at the MicroSystems Forum, sharing tips on how to maximize graphics performance.

MICHAEL MUSTACCHI

Dual Graphics Engines

The Power 9100's graphics engine shows its workstation heritage—it is unlike traditional Windows accelerator engines. The 9100 includes two independent units: a parameter engine and a drawing engine. The parameter engine handles command validation and exception processing of commands. Once a command is validated, the parameter engine hands it off to the drawing engine, which then executes the command.

The two engines operate concurrently, that is, the next command's parameters are validated during execution of the current command. The split engines relieve the host CPU of the many graphics parameter checks that must be performed, and the drawing engine is designed to provide maximum performance without checking for error conditions.

The commands to the 9100 are memory-mapped; a write to a given address implies the command, with the data that is written carrying the parameters. This results in a very fast command interface, utilizing the full bandwidth of the Power 9100.

Oak's Spitfire

Oak Technology's GUI accelerator, the OTI-107, has a multimedia focus. It incorporates a video port to allow mixed video and graphics displays. This port reduces chip count for integrated graphics and video systems. The Oak "Spitfire" chip does need an external RAMDAC and clock synthesizer. The '107 has a 110-MHz pixel clock, which enables it to support higher-performance RAMDACs than are typically used in competing parts with integrated DACs. The fast pixel clock can thus support higher resolutions—such as non-interlaced 1280×1024—than most integrated parts. Spitfire also has a glueless interface to the VL-Bus and PCI local buses.

Oak uses the 64-bit memory bus for the '107 for mostly the same reasons Cirrus does: it's an easily implemented, high-bandwidth memory subsystem. Unlike Cirrus, however, Oak needs the extra bandwidth for more than just the CPU and graphics engine—it also supports a bandwidth-hungry full-motion video input.

The '107's memory system is further enhanced with many FIFOs, which enable the chip to handle high-bandwidth bursts of activity without impeding performance. Three FIFOs are on-chip: a four-level, 32-bit-wide CPU-interface FIFO, which improves the system-interface data transfer rate; a 64-bit-wide, eight-deep graphics memory FIFO, which improves the data transfer rate between the graphics memory and engine; and an 8-bit, 32-deep video-port FIFO, which improves the data transfer rate across the video channel.

Full-Motion Video Channel

The '107's 8-bit video port feeds into the video FIFO. The input typically comes from a Brooktree Bt812 or Philips SAA7191B video decoder, as shown in Figure 1. These parts accept a standard analog video input in NTSC format (a US television standard) and generate 8-bit YUV (luminance and chrominance) data and appropriate synchronization signals.

With other chips, the output of the video decoder is fed into a video controller that has its own video memory, as shown in Figure 2. A digital video mixer then combines the graphics controller and video controller data streams and feeds the result to a RAMDAC. In such an arrangement there is a duplication of expensive video memory and two controller chips.

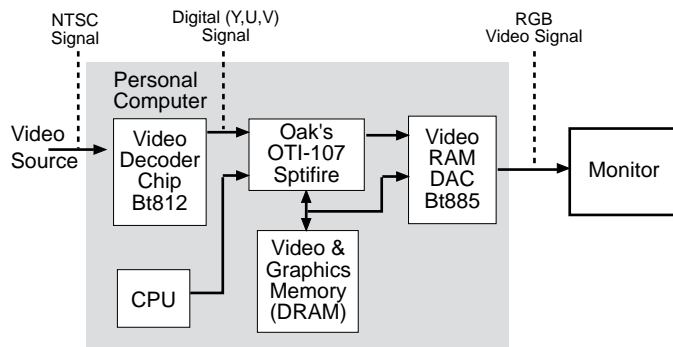The '107 eliminates most of the components associ-

Figure1. Oak's Spitfire chip can be used with a video decoder and two-channel DAC to provide a complete video subsystem.

Figure2. Typical video implementations require a video controller and video memory on an separate add-in board, increasing cost.

ated with past PC video implementations. It accepts signals from the video decoder and then translates the 8-bit YUV data into 24-bit RGB (red-green-blue) and feeds it into the same display memory that the CPU and graphics engine use. The video image is stored in a separate memory area from the active graphics display, to allow for correct windowing—if they were stored in the same area, video would overwrite the graphics image underneath. This method obviates the need for both a video controller and video memory, substantially reducing the cost of a video implementation.

In a typical video design, the '107 will be attached to a Brooktree Bt885 dual-ported video RAMDAC, which accepts both graphics data from the controller and video data directly from the display buffer. Registers within the '107 control the size, placement, and memory locations of the video image. Based on the register settings, the '107 provides appropriate control signals to the Bt885.

The size of a full-motion video image depends upon the display RAM size. A 1M memory system—which will undoubtedly be a rare beast, since it requires the oddball $128K \times 16$ (2-Mbit) DRAM—enables true-color video images up to 320×240 to be displayed on top of a $1024 \times 768 \times 8$ graphics image. In this example, the graphics display uses 768K of the frame buffer while the video image uses 225K. A larger, 2M memory system allows a standard VGA resolution (640×480) video image to be displayed in a larger, $1280 \times 1024 \times 8$ window.

The OTI-107 graphics engine includes some fairly advanced features. The engine is a full 64-bit design and, in addition to the standard graphics features, there are four independent hardware bitmaps. The first three correspond to the standard source and destination bitmaps and pattern register of Windows. The presence of a bitmap instead of a pattern register increases bitmap processing performance. The fourth bitmap handles the Windows NT "mask" operand, allowing the '107 to support both Windows 3.x and Windows NT in hardware. The '107 also supports a 64×64 hardware cursor. The
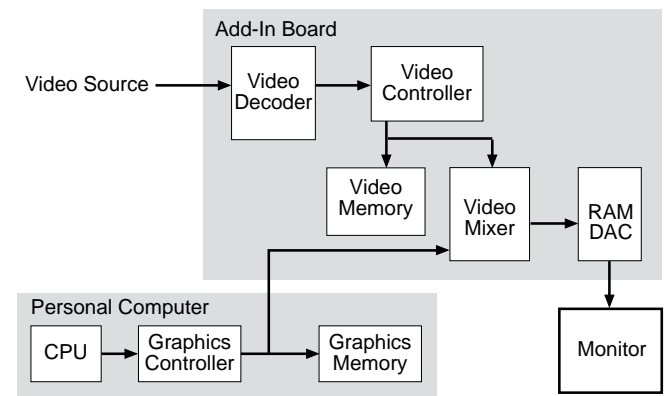
hardware cursor and a video window cannot be present at the same time, however, as these functions share circuitry.

Tseng Labs W32i and W32p

Tseng Labs has introduced two new, closely related components, the W32i and the W32p. The primary differences between the two are that the W32i provides direct interfaces for the ISA/EISA/MCA expansion buses and the VL-Bus, while the W32p supports only the VL and PCI local buses. The W32p also supports 16-bit external RAMDACs and has a hardware line-draw engine, while the W32i has neither feature. The W32p is in a 208-pin package, instead of the W32i's 160-pin package, and the W32p also supports higher resolution and color-mode combinations.

The W32i continues Tseng's pin-compatible history; boards based upon the older W32 controller can be quickly upgraded to the W32i simply by changing the controller. Neither of the Tseng components integrates the RAMDAC or clock synthesizer. Like Oak, Tseng supports high-speed external RAMDACs with a 114-MHz pixel clock, allowing it to be used with non-interlaced 1280×1024 monitors.

The Interleaved Approach

Tseng cites its long experience with DRAM-based designs—the company has yet to introduce a VRAM-only controller—as the reason for its selection of a DRAM-based memory subsystem for the W32i/p controllers. Interestingly, rather than using a 64-bit wide DRAM interface to boost bandwidth, Tseng chose an interleaved 32-bit design.

The Tseng interleaved memory design is a two-way, one-clock-per-transfer design, which offers more bandwidth than the three-clock design considered by other vendors. Tseng indicates that while the one-clock interleaved design is more difficult to implement when designing a chip than a 64-bit solution—the DRAM timing is very tight—the controller uses the available DRAM

Sierra Swift

Most of the recently introduced components are from established controller manufacturers and are high-end, high-cost devices. Sierra Semiconductor, which has long had a presence in the graphics market as a RAMDAC and clock-synthesizer supplier, makes its first foray into the graphics-accelerator market with Swift, which is an inexpensive, low-end integrated SVGA controller with accelerator features and an on-chip 24-bit color RAM-DAC and clock synthesizer. The component is a direct competitor to Cirrus Logic's low-end GD5422 accelerator.

Swift offers the standard BitBLT, color expansion, and pattern fill features common to virtually all accelerators today. The Swift memory system is a typical 32-bit DRAM implementation, though it is limited to 1M in size. Swift supports a VRAM-based memory subsystem as well. Due to the VRAM option, Swift's memory bandwidth (and hence performance) can be increased substantially by a change in the memory implementation.

The Swift is sampling now with volume production in August. It is priced at $15 in large quantities. Contact Sierra Semiconductor, 2075 North Capitol Avenue, San Jose, CA 95132; 408/263-9300, fax 408/263-3337.

bandwidth to maximum advantage without violating any timing specifications.

While the 32-bit interleaved approach may have been more difficult during chip design, it does result in simpler graphics-subsystem designs. Specifically, the memory system remains 32-bits wide—albeit in two banks. In contrast to 64-bit designs, this reduces the need for the very widest DRAMs and eases board layout and skew considerations. Systems using the W32i also have the advantage of a 160-pin package, which is easier to handle than the 208-pin package used by W32p and many other controllers.

Tseng's W32i achieves a bandwidth of 170 Mbytes/s with a 2M interleaved memory system implemented as two symmetric 1M banks, using standard 70-ns DRAMs. As with any DRAM-based design, a portion of the bandwidth is used for display refresh. In the case of the W32i/p, 115 Mbytes/s of bandwidth remains after servicing a 1024 × 768 × 8 display. The maximum memory clock of the controller is only 50 MHz, so there is still room to increase bandwidth with faster devices.

Viper Video

Tseng was the first to support an 8-bit video data port on the graphics controller with the original W32, and has continued this feature in the W32i and W32p. Tseng's approach to video is somewhat different than Oak's, however. While Oak has a strategy allowing high-integration designs, Tseng places much of the video pro-

cessing into a second chip, called Viper.

The Viper chip gluelessly connects to any member of the W32 family. It provides the video scaling and YUV-to-RGB conversion necessary for video on a PC. The Viper also accepts 15/16- and 24-bit RGB inputs. Similar to the Oak solution, the Viper/W32 combination allows a single frame buffer to be shared by the graphics and video functions.

While utilizing a second chip results in a less integrated video solution than Oak's, it does offer a flexible graphics architecture. W32-based designs can include a daughter-card connector, and two levels of video can be offered as add-on options to the traditional accelerator card. One level of video supports Intel Indeo and Microsoft Video for Windows. The Viper chip can accept a data stream from the CPU and perform the video-in-a-window display. This offers performance advantages over a software-only video window. Another level of video support, which adds analog-to-digital conversion, allows NTSC, PAL, and S-VHS video sources to be added. Viper also enables still-frame capture from external sources.

Chips and Technologies Wingine DGX

Chips and Technologies' new accelerated graphics controllers offer some new twists. This controller is similar to the highly integrated Cirrus GD542x series. Like the GD542x, the 64300 has an on-chip 24-bit color RAM-DAC, integrated clock synthesizer, and the standard graphics features. A deeper look reveals that the C&T 64300 family (there are two parts) is significantly different from other accelerators. The company has taken an approach to increasing memory bandwidth that has never before appeared in a desktop graphics controller.

C&T is using a cache-like technique to increase the available bandwidth. The new controllers have an internal "cache," and, optionally, an external DRAM that the company calls an "XRAM." The cache concept is a misnomer—the devices operate by buffering displayed scan lines, and not by caching data accesses. While the company is secretive about the specifics of how the additional memory increases performance—a patent on the proprietary technique is pending—enough of it has been made public to understand the general operation.

The internal cache is 1.2K. The optional XRAM, a 256K × 4 DRAM, can be added to the 64300; the 64301 supports only the internal RAM. These cache RAMs are used to store portions of the current display memory on a scan-line basis. As the display is refreshed, when it reaches a scan line that is stored in either the internal memory or the XRAM, the controller reads the contents out of that RAM instead of reading the data from the actual display DRAM. As a result, the graphics DRAM bandwidth that would have been used for refresh is now available to the graphics engine. The added bandwidth translates into higher performance. The cache can be up-
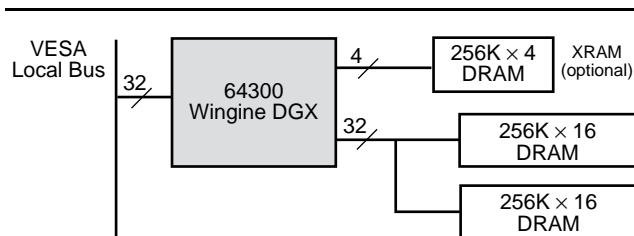
Figure 3. Chips and Technologies' 64300 uses an extra 256K × 4 DRAM (called an "XRAM") in a unique way to boost the performance of a 32-bit DRAM memory system.

dated with new scan lines simultaneously with actual display refresh from the display DRAM.

How much the cached-display approach helps performance will vary. On a purely mathematical basis—adding back to the display memory the bandwidth the cache memory supplies—the 256K×4 XRAM should free up approximately 11–12 Mbytes/s of bandwidth from the primary display memory. The basic memory of the 64300 is configured as a standard 32-bit DRAM design. Therefore, the bandwidth of the system is limited to about 90 Mbytes/s. In a typical bandwidth-intensive mode, only about 30 Mbytes/s of bandwidth will remain once the display is serviced. Adding a 256K×4 DRAM results in an impressive 35% increase in display-memory bandwidth, to nearly 42 Mbytes/s. More importantly, with correct timing, memory contention may be avoided, resulting in an even greater performance improvement.

One potential pitfall of a display cache is with multimedia applications; caching a scan line depends on the data not changing. In high frame-rate scenarios typical of video-in-a-window displays, the caching technique could either fail or result in poor display quality, where a few scan lines would lag behind on an image. C&T has managed to avoid this problem, and in fact demonstrates the 64300 in a system displaying three video windows without losing frames. The company indicates that caching with a finer granularity than scan lines enables their controller to enhance video performance.

Summary

The advent of local-bus graphics clearly has shifted the performance bottleneck in graphics controllers from the CPU-bus interface to the display-memory interface.

The new controller architectures demonstrate that many controller manufacturers are addressing the issue with legitimate technical innovations rather than simply tweaking the graphics engine.

System and board manufacturers have a wide variety of options. Highly integrated solutions, such as those from Cirrus, C&T, and Sierra (see sidebar) are no longer low-performance choices, but now offer a full range of options from low to high end. Manufacturers that wish to differentiate on the basis of performance and video capability will be best served with the Tseng and Oak controllers, which are video-ready. For those interested in performance at any cost, the Weitek 9100 offers great possibilities.

No matter what their ultimate design goal is, system manufacturers should pay close attention to the memory bandwidth of their controllers, because in many instances—such as the memory-to-display and display-to-display BitBLT operations that occur frequently in Windows—memory bandwidth more closely reflects real-life performance than any of the often-exaggerated artificial benchmarks currently in vogue.♦