

Cyrix GX Slashes System Cost

Highly Integrated 5x86 Processor Also Includes Multimedia Features



by Linley Gwennap

Not content with offering chips that are simply software- and pin-compatible with Intel's, Cyrix unveiled at the recent Microprocessor Forum a bold effort to change the way PCs are designed. Cyrix architect Forrest Norrod explained that the forthcoming 5gx86 processor will eliminate the need for an external cache, DRAM controller, PCI interface, nearly the entire graphics subsystem, and various add-in cards—all without reducing performance.

This GX version of today's 5x86 is a highly integrated x86 processor that connects to DRAM and a PCI bus directly. The direct DRAM interface, particularly with EDO memory, increases performance enough to eliminate the need for an external level-two cache. Using a unified memory architecture (UMA), the GX keeps the graphics frame buffer in main memory and performs graphics acceleration on chip; only an external RAMDAC chip is needed to drive the monitor. An innovative compression scheme reduces the bandwidth pressure found in more traditional UMA designs.

The GX will also be the first product to include Cyrix's new virtual system architecture (VSA). These instruction-set extensions and design changes allow the CPU to emulate functions, such as Sound Blaster audio, normally performed by expensive add-in cards. This emulation is fully compatible with existing software. These extensions also improve performance on real-time tasks such as native signal processing (NSP). Cyrix plans to eventually incorporate VSA throughout its entire product line.

Norrod said that his team is currently testing the initial silicon of the 5gx86. It is based on the current 5x86 (see [090901.PDF](#)) and should ship at 120 MHz. The company expects to introduce the new product in 1H96.

Making DRAM Faster Than L2 Cache

It's easy to cut cost by eliminating the external cache; plenty of PC makers do it today. These cacheless systems, however, are 10–20% slower than systems with an L2 cache (see [091305.PDF](#)). As CPU clock speeds continue to accelerate, this performance gap will widen. Moving to EDO main memory reduces but does not elim-

inate this gap.

Norrod noted that today's DRAMs are actually quite speedy: on a page hit, the first data is returned in about 40 ns. In comparison, an external cache typically takes three 66-MHz bus cycles, or 45 ns, to return the first data, since it has to do a tag lookup. Traditional processors take much longer than 40 ns to access DRAM because the address and data must flow through several buffers and a separate DRAM controller, all synchronized to a slow, 66-MHz clock. The same synchronization issues also delay the cache access.

Instead of this complex design, the GX processor puts the DRAM interface right on the processor chip. On a miss from the CPU's on-chip L1 cache, the processor immediately drives the address to the DRAM; about 40 ns later, the data appears on the pins of the processor. With 60-ns DRAM, a 120-MHz GX stalls for a total of six CPU cycles on an L1 cache miss. Note that a 120-MHz 5x86, using an external L2 cache, would stall for 10 CPU cycles, including three 40-MHz bus cycles, in the same situation.

Although faster than SRAM in accessing the critical word, the DRAM is slightly slower in completing the line fill. EDO DRAM will return subsequent data words every 20–25 ns, whereas a synchronous cache will burst at 15–17 ns. The cache also has an advantage on mul-

multiple accesses, as the tag lookup can be pipelined with the previous data transfer, saving two bus cycles. On the other hand, eliminating the L2 caches avoids the overhead of moving data between the L2 and main memory.

Norrod claims the integrated DRAM controller provides memory performance comparable to that of a traditional design with an external L2 cache and DRAM interface. The performance improvement of integrated DRAM controllers has already been validated by RISC CPUs such as MicroSparc, Digital's 21066, and HP's PA-7100LC, supporting this claim. (Although the initial GX design includes an L2 cache controller, this feature probably won't be supported in the final product.)

UMA Cuts Graphics Subsystem Cost

The next step in reducing system cost is in the graphics subsystem. Most PCs today include a stand-alone graphics accelerator that controls its own frame-



Cyrix architect Forrest Norrod explains how the 5gx86 will reduce system cost.

MICHAEL MUSTACCHI

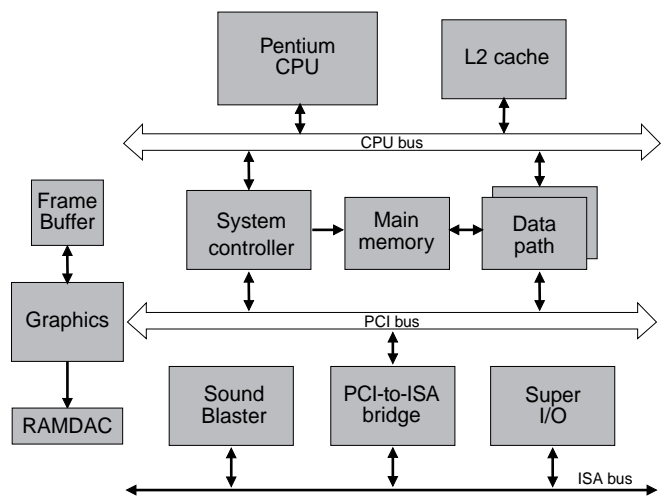


Figure 1. Typical Pentium PC with Triton chip set, PCI graphics accelerator, and Sound Blaster audio.

buffer memory, as Figure 1 shows. In contrast, the 5gx86 system in Figure 2 eliminates both of these components by putting the frame buffer into main memory and the graphics accelerator onto the processor chip.

Several system-logic vendors are planning UMA chip sets that place the frame buffer in main memory (see *090801.PDF*). The problem with standard UMA designs is a decrease in system performance caused by contention between the CPU and the video refresh as both try to access the main DRAM. For example, a $1024 \times 768 \times 8$ display refreshing at 72 Hz requires 57 Mbytes/s, about 30% of the usable bandwidth of an EDO memory system. More color depth, higher resolutions, and faster refresh rates consume even more bandwidth.

Compression Reduces Bandwidth

Cyrix attacked this problem in two ways. First, the GX chip includes a 512-byte refresh buffer that supplies data to the display. As long as this buffer isn't empty, CPU accesses are given higher priority, and the buffer is filled when the DRAM is otherwise idle.

Buffering helps arrange DRAM accesses to avoid contention, but it doesn't solve the problem if the CPU and display together require more bandwidth than the DRAM can supply. This issue is eased by compressing the frame buffer. This technique reduces the amount of data in the frame buffer, and thus the bandwidth needed to access it, by approximately 12:1, according to Norrod. Weitek's UMA chip set uses a more limited form of data compression with a worse compression ratio.

The GX actually keeps two copies of the frame buffer in main memory: one compressed and one uncompressed, as Figure 3 shows. Programs read and write the uncompressed version, making the compression transparent to software. Any time a scan line on the screen is changed, it is marked as dirty. The compressed buffer

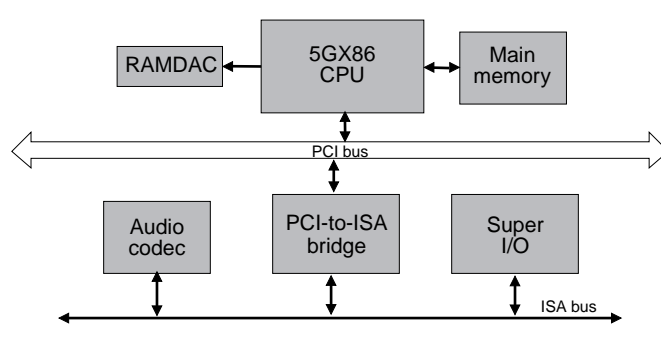


Figure 2. A PC using Cyrix's GX design eliminates the L2 cache as well as most of the graphics subsystem and system-logic chip set.

supplies data to the on-chip refresh buffer, greatly reducing refresh bandwidth. From this buffer, data is decompressed and then sent to the display.

Whenever the chip needs to fetch a scan line that is marked as dirty, it instead reads from the uncompressed buffer to get the most current data. This data is then compressed using four different lossless compression algorithms; the smallest result is written back to the compressed frame buffer. The compressed data is then used on future refresh passes. The scan line tags shown in Figure 3 indicate whether each line is clean or dirty and what compression algorithm is used on that line.

The size of the compressed frame buffer is programmable, since some situations allow greater compression ratios than others. If the software makes the compressed buffer too small, some lines will not fit; these lines must be fetched from the uncompressed buffer on each refresh pass, consuming extra bandwidth.

High-Performance Graphics

The 5gx86 performs common graphics-acceleration functions, such as bitBLTs and raster operations, in the on-chip graphics pipeline shown in Figure 4. The graphics unit also contains some video-acceleration functions, but Norrod would not give any further details. He would not reveal any graphics performance data but claimed that the GX will be comparable to today's high-end DRAM-based graphics accelerators in performance.

Although the graphics pipeline itself is relatively simple, graphics performance is increased by tightly coupling the CPU, graphics unit, and memory controller. As Figure 4 shows, the graphics unit has direct read and write access to the level-one cache, allowing it to quickly and efficiently share data with the CPU. Data is shared at the full CPU clock rate, and no time is lost synchronizing with slow external buses.

Virtual System Architecture

Cyrix's virtual system architecture (VSA) is a set of concepts that will appear first in the 5gx86 and later spread to the company's other processors. Instead of

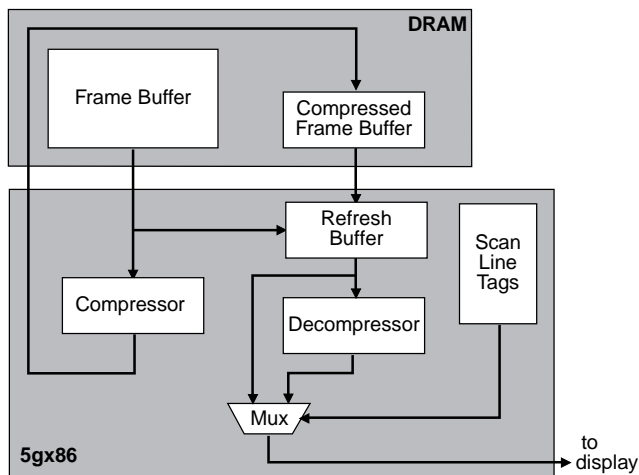


Figure 3. The GX design keeps two copies of the frame buffer in main memory, one compressed and one uncompressed.

adding peripheral hardware, VSA involves changes to the CPU core, making it suitable for nonintegrated processors as well. In fact, at the Microprocessor Forum, Cyrix demonstrated a VSA-enhanced 6x86 processor performing Sound Blaster emulation.

Emulating a legacy peripheral like Sound Blaster isn't easy. Software written for new Windows APIs executes audio functions through drivers that can be configured for any type of hardware, even NSP. But DOS programs access the audio hardware directly, using I/O register reads and writes. Sound Blaster has become the de facto standard for such direct-access audio.

Like Chips & Technologies' SuperState (see [0602VP.PDF](#)), VSA allows the processor to trap accesses to Sound Blaster (and other) registers. Cyrix enhanced its system management mode (SMM) with faster entry and exit procedures and nested interrupts, allowing SMM to be used for peripheral emulation. When software writes to a Sound Blaster register, the CPU enters SMM, where a trap handler analyzes the request and executes an emulation routine to generate the desired audio.

This sequence is completely invisible to the application software and can be extended to other types of legacy I/O devices. In fact, the demonstration system included an emulated, or virtual, DMA controller, because the GX can perform internal transfers faster than the external DMA controller. Cyrix points out that trapping I/O accesses before they are sent to the glacial ISA bus saves an enormous amount of time; in many cases, the emulation routine is finished before the original I/O access would have completed.

The VSA extensions include modifications to the on-chip cache that allow portions to be locked, so the data stored there is never replaced. This feature, common in embedded processors, improves real-time capability by ensuring that important code or data is always in the cache. Because the 5x86 and 6x86 have no load-

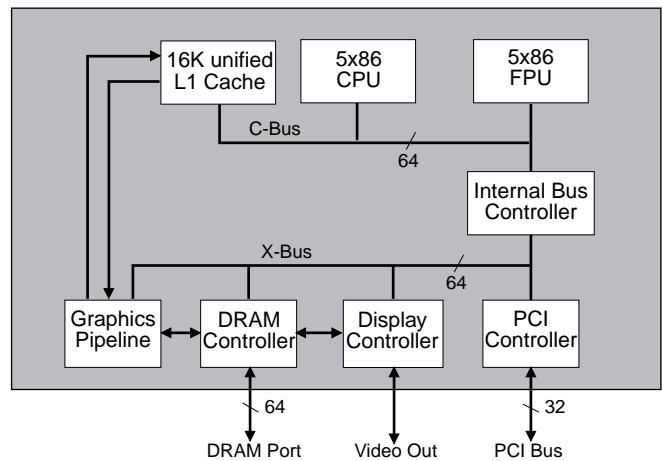


Figure 4. The 5gx86 combines the 5x86 CPU with a variety of system interfaces on a single chip.

use penalty, this feature can also be used to create an extended "register set" in the cache that has the same response time as the physical x86 register set.

New instructions allow block data transfers, which are useful for graphics. These instructions can move transient data through a locked portion of the cache, preventing it from overwriting useful data in other parts of the cache. VSA includes a fast integer multiply-and-accumulate operation, good for multimedia applications. In fact, Norrod hinted that VSA may include a number of new multimedia instructions to compete with Intel's P55C extensions, but he would not elaborate on or even confirm the existence of such extensions.

Avoiding the 486SL Syndrome

Highly integrated x86 processors have not fared well in the past. Texas Instruments' Rio Grande was the most recent such failure (see [0814MSB.PDF](#)), and the path goes back to Intel's 486SL (see [061501.PDF](#)). A big problem with the 486SL was economic: the added DRAM and ISA interfaces doubled the size of the die but replaced only \$10–\$20 of external logic. The cost savings were minimal because the 486SL offered no system-design innovation, just integration.

Cyrix is working to change these economics. Adding system logic to a Pentium-class core makes more sense: since the core is larger than a 486, the relative increment is smaller. Norrod would not reveal the die size of the 5gx86 but said that the new logic adds less than 20% to the size of the standard 5x86.

With the innovative GX design, the system cost savings are far greater than a few dollars. Table 1 summarizes the major cost differences between the Pentium and 5gx86 PCs shown in Figures 1 and 2. The Cyrix design requires only the processor, DRAM, a RAMDAC for the display, and a PCI-to-ISA bridge to provide the main functions. Audio can be delivered through a sim-

	Pentium PC		Cyrix GX PC		Diff
Processor	Pentium-90	\$261	5gx86-120	n/a	n/a
L2 cache	256K asynch	\$42	none	\$0	-\$42
System logic	Intel Triton	\$25	PCI-to-ISA	\$7	-\$18
16M DRAM	fast page	\$436	EDO	\$458	+\$22
Graphics chip	S3 Trio	\$25	none	\$0	-\$25
Frame buffer	1M DRAM	\$27	none	\$0	-\$27
RAMDAC	in S3 Trio	\$0	BtV2487	\$5	+\$5
BIOS ROM	1M EPROM	\$5	1M EPROM	\$5	\$0
Super I/O	National	\$7	National	\$7	\$0
Audio	ESS 788	\$15	codec	\$4	-\$11
Total cost difference					-\$96

Table 1. A comparison of key cost differences between a typical Pentium PC and a system based on Cyrix's GX architecture shows a significant cost savings for the Cyrix design. The table uses 4Q95 pricing for high volumes. (Source: MDR Technology Roadmap)

ple codec rather than a complete Sound Blaster chip. This design gets rid of the L2 cache, graphics accelerator, frame buffer, and most of the system-logic chip set as well as the sound chip, for a net savings of about \$100. These savings would be reduced by any price premium for the 5gx86 itself compared with a standard Pentium.

Although the cost savings are compelling, performance is a concern. Norrod expects that the GX-based system will have application performance comparable to a 5x86 with a standard (non-UMA) chip set. The direct DRAM access appears to satisfy the CPU's demands as effectively as an L2 cache, but doing so requires more DRAM accesses than in a system with an L2 cache. If the DRAM becomes overworked, CPU performance will suffer due to longer apparent memory latency.

The UMA design puts an additional burden on the busy DRAM, particularly in systems using 16- or 24-bit color displays. Cyrix expects its compression algorithm will reduce this burden to just a few percent for most programs. Interference between CPU and video accesses will reduce the page-hit rate of the DRAMs, increasing effective memory latency. Emulation of audio and other system functions may also degrade CPU performance and put an additional burden on the memory system. The company plans to conduct extensive system-level testing to verify the performance of its design.

Based on Cyrix's claims, a 5gx86 system should be comparable in performance to a midrange PC configuration, such as the one in Figure 1. Even if performance doesn't quite meet the company's claims, the 5gx86 would still be compelling for low-end systems, which already cut corners by deleting the L2 cache and will be moving to UMA chip sets with reduced performance. Cyrix did not reveal a price for its device, but the company tends to offer low prices, and the relatively small increase in die size should enable Cyrix to keep the premium for the GX version fairly small.

Price & Availability

The 5gx86 is not an announced product. Cyrix expects to announce pricing and availability for this chip in 1H96. For more information, contact Cyrix (Richardson, Texas) at 800.462.9749 or 214.968.8388; fax 214.699.9857 or access the Web at www.cyrix.com.

Setting a New Standard

The unique features that make the GX attractive will also make it more difficult to establish in the market. Cyrix has made its living with pin-compatible devices that can be dropped into existing PC designs; system makers must instead design a new motherboard specifically for the GX or obtain a design from Cyrix. In doing so, they are committed to the new part, as no other vendor provides a compatible device. Many PC vendors, particularly large ones, may be concerned about relying on Cyrix for such a critical sole-sourced component. The cost advantage of the GX should compel some of them to take the gamble.

One problem is performance: Cyrix expects a 120-MHz 5x86 to match the performance of a 100-MHz Pentium. If the limited DRAM bandwidth saps performance, the GX version might not meet this standard. By mid-1996, this level of performance will be suitable only for the low end of the desktop market. The company will probably address this issue by offering higher clock speeds of the 5gx86 over time; eventually, it could offer the 6x86 core in the GX architecture.

The new chip must also deliver competitive performance on MPEG video and other multimedia applications. We suspect that the company is not so foolish as to ignore this area, but Cyrix would offer no assurances in this regard. The GX probably can't compete with PCs equipped with the Chromatic chip (*see 091404.PDF*), but these systems will cost significantly more to build.

Another issue for Cyrix is building the software needed to support the GX, including the drivers for the graphics system as well as any "virtual I/O devices" that take advantage of the VSA extensions. The company has not tackled such a problem before but is devoting significant effort to software development. For the GX architecture to be accepted, these drivers must be compatible with existing standards, perform well, and cover a range of APIs and functions.

If Cyrix can deliver the required software, we expect the GX to be accepted in low-end desktop systems. The company must keep the price premium for the integrated system logic small, and it must demonstrate adequate performance from its nontraditional design. If so, this innovative new design will help the company maintain its strong growth in 1996. ♦