SDRAMs Ready to Enter PC Mainstream

EDO, Today's Best Choice for PC Main Memory, Is Only a Stopgap Solution

by Steven Przybylski

The past few years have been tumultuous ones for the DRAM industry. In the course of five years, the number of DRAM architectures has mushroomed from essentially 2 to more than 15. Several new DRAM architectures that were unknown or even undefined in 1990 are already or soon to become mainstream devices. This proliferation of new devices has resulted in a lot of confusion—and plenty of hard decisions on the part of both the DRAM industry and systems companies.

A metric called fill frequency can be used to compare these new architectures. Any DRAM design that does not meet the minimum fill-frequency requirement for the PC market cannot be effective in that market. This tool allows us to weed out the DRAM architectures that will not be competitive in the future: in particular, the various forms of asynchronous DRAM, including EDO. Fill frequency helps demonstrate that synchronous solutions will be required, and of these, SDRAM has the best shot for near-term success in PC main memory.

DRAMs Must Meet Bandwidth, Granularity Goals

The traditional metrics used to characterize the performance of memories and memory systems are latency—the time from the start of an access to the beginning of a burst transfer—and bandwidth—the peak or average data rate. Bandwidth is increasingly important in both memory system and DRAM design, because the bandwidth needed to sustain processor execution has been growing dramatically with increasing processor performance. In contrast, the bandwidth capabilities of DRAMs grew slowly through the 1970s and 1980s, creating the perception of a bandwidth performance gap, as Figure 1 shows.

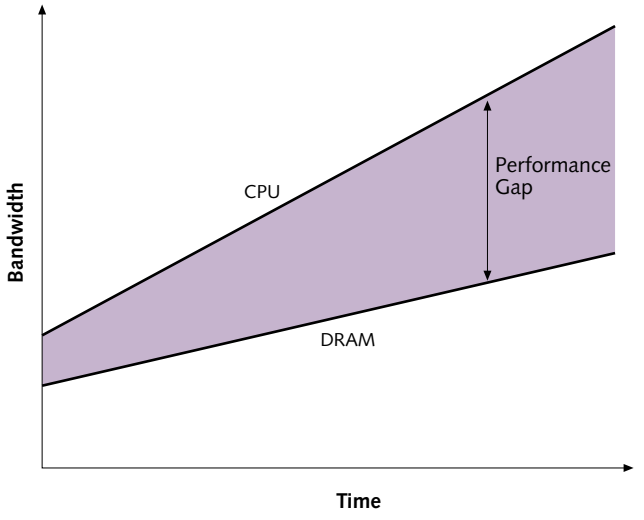

Figure 1. The traditional view of performance gap between the bandwidth needs of processors and the bandwidth capabilities of DRAMs fails to articulate why this gap has become critical now.

But looking at per-DRAM bandwidth alone does not give a complete picture, because as much bandwidth as needed can be generated from a memory system by grouping many DRAMs into wide (×64, ×128, ×256, etc.) parallel structures. The problem with this technique is that the minimum size of the memory system grows with each increase in the width of the memory system. Increasing the minimum memory size raises the cost and price of the entry-level configuration, potentially beyond what the market will bear. So the system markets impose two sets of constraints on the design of a memory system:[1]

1. The memory system must provide adequate bandwidth for the level of CPU performance expected by the end user.
2. The memory system must support an entry-level configuration that is cost effective for the target market.

Bandwidth alone addresses the first market requirement but not the second. As DRAMs continue to get larger and the number of DRAMs per system continues to decline, it is becoming increasingly hard to satisfy both sets of requirements simultaneously. This is the fundamental problem spurring the proliferation of new DRAM architectures.

Fill Frequency Combines Two Metrics

Table 1 lists the approximate peak bandwidth and granularity (i.e., minimum memory size) requirements for several types of memory systems. The granularity requirements are quite concrete, because the minimum memory size directly affects the system price. The bandwidth requirements are less precise, because end users are generally not explicitly aware of the bandwidth of their memory systems. Since system-level performance is dependent on a great many factors—of which memory system bandwidth is a relatively minor component—only where the memory bandwidth is dramatically less than in comparable systems is the ultimate market success of the system affected.

The ratio of the bandwidth and granularity requirements indicates how difficult the memory system is to build. A high ratio indicates that a great deal of bandwidth is

[1]There are, of course, more constraints than this: latency, physical size, form factor, power consumption, and so on. In the PC market, the memory system often must assist the sale of the system by contributing to the list of buzzwords. This article focuses on bandwidth and memory size because they provide a useful way to link DRAM capabilities to system requirements.

Memory System Type	Granularity Requirement (Mbytes)	Peak Bandwidth Requirement (Mbytes/s)	Ratio (Hz)
Workstations	64–128	500	<7.8
Large PC	16	266	16
Small PC	8	133	16
GUI Accelerator	1–4	100–400	100
UMA PC	8–16	200–400	25–33

Table 1. Different system markets impose different granularity and bandwidth requirements on memory system designs. The ratio of the two indicates the difficulty of the memory system design.

needed from a small memory system. In contrast, a low ratio indicates that not much bandwidth is required relative to the memory system's size. It is somewhat counterintuitive that the ratio of market requirements for workstations and other larger computer systems is smaller than the ratio for PCs. Although the bandwidth requirements are greatest for workstations, the granularity requirements are even greater. The ratio is thus reduced in the larger system classes. For unified memory architecture (UMA) PCs, the ratio of bandwidth to granularity lies between those for traditional PCs and GUI accelerators: UMA PCs inherit the bandwidth requirements of GUI accelerators and the granularity requirements of PCs.

A memory system satisfies a system market's bandwidth requirement when the memory system's bandwidth is greater than or equal to the requirement. Similarly, a memory system satisfies a market's granularity requirement when the memory system's minimum size is less than or equal to the granularity requirement.

To satisfy both requirements, the ratio of memory system's bandwidth to granularity must be greater than the ratio of the market requirements. Conversely, if a memory system's bandwidth-to-granularity ratio is below the ratio of marketplace requirements, that memory system will be unacceptable because one or both of the market requirements is not being met. If the memory system's ratio is greater than or equal to the ratio of market requirements, the memory system is not necessarily unacceptable—in other words, it may or may not be acceptable.

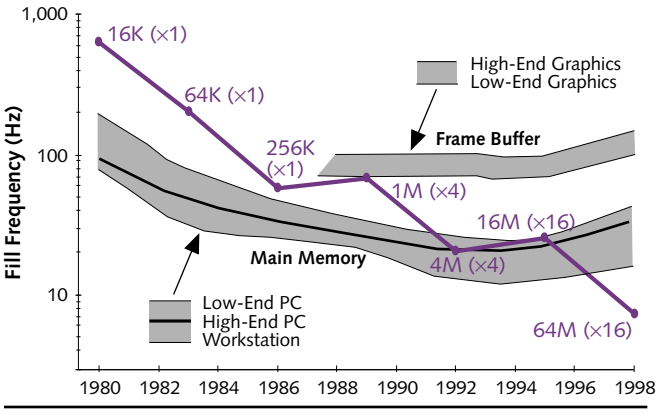

Figure 2. The semiconductor and systems industries have migrated from ×1 to ×4 to ×16 DRAMs to keep the peak fill frequencies of DRAMs (purple) at or above the market requirements (gray).

The ratio of a memory system's granularity to its bandwidth is analogous to the ratio of the volume of a container and the maximum flow rate of water into the container: it is equal to the time required to fill the container or memory system. Consequently, the ratio of the bandwidth to the granularity of a memory system is the frequency with which the memory system can be filled with information. It is therefore called the fill frequency of the memory system, and the ratio of the market requirements is called the market-requirement fill frequency.

DRAM Fill Frequency Proves Unsuitability

DRAM chips also have fill frequencies. They have a capacity and a peak bandwidth, as determined by their width and maximum operating frequency in a system. The ratio of a DRAM's bandwidth to its size is the maximum rate at which the DRAM can be filled.[2] It is necessarily true that the fill frequency of a memory system is less than or equal to the fill frequency of the DRAMs used to construct it. If this were not true, it would be possible to fill the memory system faster than it is possible to fill one of its constituent DRAMs.

A crucial consequence of this observation is this: if a DRAM's fill frequency is less than a market requirement, all memory systems constructed of that DRAM will have unacceptable fill frequencies; in all cases, either the bandwidth or granularity market requirements will not be met. As a result, the DRAM itself is unacceptable for that system market. The ability to compare DRAM capabilities with system market requirements provides us with a powerful tool to directly compare the capabilities of present and future DRAMs with system requirements.

Figure 2 plots market-requirement peak fill frequencies and fast-page-mode (FPM) DRAM peak fill frequencies back to 1980. It shows the DRAM and system industries shifted from ×1 to ×4 to ×16 devices only when it was necessary to keep the fill frequencies from becoming unacceptable. At the 64-Mbit level, however, even ×16 fast-page-mode devices will be unacceptable. With fill frequencies intrinsically dropping by almost a factor of four every three years, the challenge facing the DRAM community is to develop new device architectures that can keep fill frequencies high enough while minimizing system cost.

The market-requirement fill frequencies are fairly constant over time. In general, both bandwidth and granularity requirements increase at roughly the same rate. Since 1980, CPU cache sizes have grown dramatically. This growth has dampened the rate of increase in main-memory bandwidth requirements, allowing many of the market-requirement fill frequencies to decline over that period. In the future, however, market requirements are likely to remain constant or even increase again, because of two factors:

[2]This fill frequency analysis began with the idea of plotting the ratio of DRAM bandwidth versus size over several generations of DRAMs. This idea originated with Mike Farmwald, then of Rambus.

1. Further increases in cache size will not have the same impact as previous increases: much of the intrinsic benefit of caches has already been reaped.
2. New multimedia data types (audio, MPEG video, texture maps, etc.) are difficult to cache effectively because they are inherently transitory. Increases in user expectations in multimedia performance will directly translate into higher main-memory bandwidth requirements.

Only if memory prices continue to fall at greater than historical rates for several years will the growth rate in granularity match the growth rate in required bandwidth.

The peak fill frequencies described here are appropriate for comparing the peak characteristics of DRAMs with the corresponding market requirements. Sustained bandwidth and fill frequencies, however, are more appropriate for comparing the performance of different DRAM architectures in specific system configurations. The sustained metrics are dependent on burst length, the mix of row and column accesses, the frequency of bank conflicts, and a host of system-specific considerations. Typical and best-case sustained metrics are generally 50–100% of the peak values.

The inevitable unacceptability of the cheap and simple alternatives—i.e., narrow FPM and EDO DRAMs—has caused the DRAM industry to offer a host of new architectures that offer better fill frequencies at current and future densities. This is accomplished through a variety of techniques: wider, higher-frequency synchronous interfaces, more efficient protocols, more internal banks, and so forth.

EDO Offers Incremental Improvement

Extended data-out (EDO) and burst EDO DRAMs offer incremental enhancements to the traditional fast-page-mode interface that has been most common until very recently. Both new interfaces are entirely pin-compatible with FPM and derive their improved performance from changes to the semantics of the CAS pin. While FPM devices can typically cycle at 15–25 MHz in real systems, EDO DRAMs can reach 33–50 MHz, and burst EDO can reach 66-MHz burst-transfer rates in moderate-size memory systems.

Figure 3 shows the differences among these three interfaces. FPM devices withdraw data from the output pins as soon as CAS is raised. In EDO devices, data remains on the output pins until the next falling edge of CAS. Consequently, the column precharge time while CAS is high is also useful for transferring data from the DRAM to the memory controller or processor.

Burst EDO (BEDO) DRAMs allow even higher operating frequencies by pipelining the column accesses and providing an on-chip burst counter. With a pipeline depth of one, five CAS edges are needed for a burst of length four. In fact, since BEDO is targeted specifically at PCs, four-word bursts are the only type of column accesses that are possible. Though multiple bursts from within the same DRAM row flow seamlessly together, a substantial penalty is paid for single-word accesses.

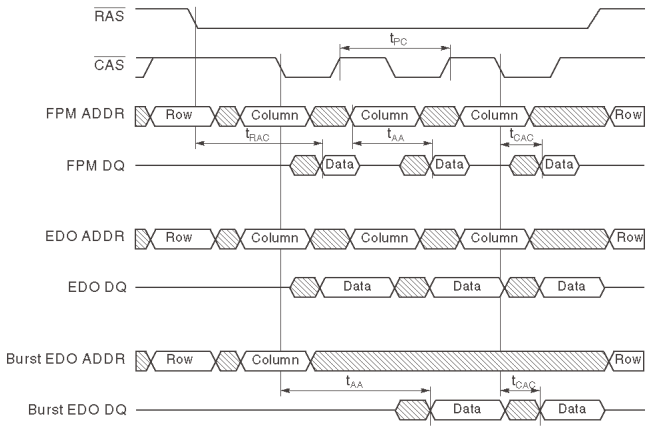

Figure 3. EDO and burst EDO differ from fast-page-mode DRAM only in the semantics of the CAS signal.

Because the operating frequency of EDO is about 50% higher than that of fast-page mode, the fill frequency of an EDO device is correspondingly better. However impressive that sounds, in an environment in which fill frequencies decline by a factor of four with each new generation of devices, a 50% improvement buys the equivalent of approximately 10 months; thus, EDO by itself is not a long-term solution. Only with significant increases in width does EDO remain a viable memory technology. At the 64-Mbit level, the fill frequency of a ×16 EDO DRAM operated at 33 MHz is an unacceptable 8 Hz.

BEDO fill frequencies are naturally better. At the 16-Mbit level, the ×16 device offers a fill frequency of 66 Hz—adequate for conventional and even UMA PCs. At the 64-Mbit level, the ×16 device is marginal at 17 Hz, largely because of the 32-Mbyte minimum memory size for ×64 memory systems. This problem would be mitigated in a ×32 BEDO DRAM, which would have a 33-Hz fill frequency.

EDO Accepted, Burst EDO Is Not

In the past several months, EDO has become widely accepted. Many PC system vendors now routinely offer EDO memory systems for all but their low-end offerings. The abruptness of the shift from FPM to EDO was in part responsible for the dramatic decline in the price of memory that began in 4Q95 and continued through 1Q96. To promote EDO, DRAM vendors emotionally and contractually committed to near equality between EDO and FPM prices. The rapid shift in demand from FPM to EDO, however, caused a shortage of EDO DRAMs and an oversupply of FPM DRAMs—largely because several vendors had not yet converted their production capacity to joint FPM/EDO mask sets. Consequently, the price of FPM chips plummeted and, to a large extent, dragged the price of EDO memory with it.

Burst EDO, promoted by DRAM vendor Micron, attracted a lot of attention during 1995, as it was seriously considered as the next major DRAM technology for the PC market. Much of the DRAM industry, however, was actively

opposed to BEDO because of the relatively fast (52-ns) row-access times needed to reach 66-MHz system operation. Although many chip-set vendors support the new design, the most significant, Intel, recommends against BEDO.

With strong support from only one DRAM vendor and weak support from a few others, plus the growing credibility of synchronous DRAMs, BEDO failed to generate the level of supply necessary for a mainstream PC memory solution. BEDO is unlikely to play a major role in the future.

Synchronous DRAM Backers Resolve Concerns

The past few years have been hard for the proponents of SDRAMs *(see* **070205.PDF** *)*. When they were first introduced in 1993, many people expected these chips to ship in significant volumes in 1995. The list of deterrents to their adoption, however, was initially formidable. The most significant were:

- Price premiums were unacceptably high.
- Significant incompatibilities among vendors at all levels of the specification made it difficult to find multiple interchangeable suppliers.
- There was no SIMM or DIMM standard.
- There was massive confusion with respect to an electrical interface for high-speed operation.
- The 64-Mbit generation was not defined.
- There were no chip sets that supported SDRAMs.

Since then, several of these issues have been resolved. For example, several companies are projecting 0–5% price premiums for SDRAMs over EDO DRAMs by 1997. The number and severity of the logical and temporal incompatibilities across vendors have also been significantly reduced. Those that remain, however, still have serious consequences for the development of a convenient and safe third-party aftermarket for memory expansion. Fortunately, formal and informal efforts are under way to eventually bring the SDRAM world to the same level of cross-vendor compatibility found in the FPM and EDO arenas.

There are now two eight-byte SDRAM DIMM standards—a 200-pin version for very high speed operation and a 168-pin design that is pin-compatible with the eight-byte EDO DIMM. This latter option will allow systems to be populated with either EDO DRAMs or SDRAMs, possibly even on a bank-by-bank basis.

After several attempts, JEDEC has settled on SSTL (Series Stub Transceiver Logic) as a high-speed electrical interface for SDRAM memory systems. Its most significant component is not actually part of the electrical interface: small resistors between the motherboard traces and the DIMM stubs significantly improve signal integrity by raising the impedance of the stubs and thereby decreasing the impedance discontinuity in the motherboard traces caused by the stubs.

The four-bank 64-Mbit SDRAMs have been defined. The allocation of bank, row, and column address bits is such that 64-Mbit and 16-Mbit SDRAMs cannot be ganged together. This makes it impossible to efficiently construct using 64-Mbit SDRAMs an ECC or parity memory system with a minimum memory size less than 64 Mbytes.

Several PC chip sets and embedded systems processors now support SDRAMs, and more will become available soon. In particular, Intel has announced SDRAM support in its 430VX Pentium chip set *(see* **1002MSB.PDF** *)*.

Some Problems Remain with SDRAM

Perhaps the most awkward aspect of the current SDRAMs is the gap that exists between the rated maximum operating frequency of the DRAMs and the effective maximum operating frequency in a system. Given the physical size and capacitive loading of a typical memory system combined with the very long clock-to-data-out times of today's SDRAMs, it is not uncommon to need a "100-MHz" SDRAM for a 66-MHz system. Perhaps worse yet, since the detailed timing parameters still vary among vendors, one vendor's 83-MHz part might be adequate but another's 83-MHz part might not. These problems are of most concern to memory designers building large systems and those using custom or nonstandard memory controllers.

As with all devices, the fill frequencies of SDRAMs depend on the density, width, and speed of the devices. At the 16-Mbit level, the 100-MHz ×16 devices offer fill frequencies of 100 Hz. When operated at 66 MHz, however, the fill frequency drops to 66 Hz, adequate for main-memory or UMA applications but not for graphics.

At the 64-Mbit level, fill frequencies are naturally reduced. If operated at 66 MHz, the ×16 and ×32 64-Mbit SDRAMs offer 17- and 33-Hz fill frequencies, respectively. This means that, unless the price of memory drops sufficiently to support a 32-Mbyte minimum PC configuration, the ×32 part will likely be used in entry-level PCs while the ×16 will be used for larger configurations. If operated at higher frequencies, the 64-Mbit SDRAM becomes a very capable UMA memory. The ×32 device operated at 100 MHz offers a fill frequency of 50 Hz.

SDRAMs are starting to ship in volume. To date, the largest volume application has been Sega's Saturn video game, which uses 4-Mbit and 2-Mbit SDRAMs. A few workstation servers have been shipping with SDRAM main memories since mid-1995. In late 1995, Toshiba began shipping a new high-end notebook computer with an SDRAM main memory and no L2 cache. Most significantly, SDRAM-based desktop PCs from Dell are beginning to ship now, and by the end of 1996, such systems should be quite common.

SDRAM-Lite Not Catching On

The introduction of burst EDO as a possible next-generation memory for PCs jarred the SDRAM community. To address the difference between the then-projected prices of BEDO and SDRAM, a group of DRAM vendors defined a subset of the full SDRAM specification that contains only those features needed by the PC market. This subset came to be

known as SDRAM-lite. It includes bursts of length one, two, and four only and a CAS latency of either two or three.

The net result is a device with the same circuitry as a full SDRAM—all potential cost savings are in reduced test time. Though significant for some companies, many of the main SDRAM vendors have decided to forgo the SDRAM-lite parts and offer only fully tested SDRAM.

Rambus Gains Volume, Targets PC Main Memory

Rambus now has seven DRAM licensees, with four expected to be in production shortly. There are also seven ASIC licensees that can produce Rambus memory controllers for system designers not able or willing to create their own.

In recent months, Rambus has improved its peak bandwidth from 500 Mbytes/s to a baseline of 533 Mbyte/s—to match the peak bandwidth of processors, such as Pentium, with a 64-bit 66-MHz system bus—with 600-Mbytes/s RDRAMs already announced.

Rambus has also defined 64-Mbit devices *(see* **091302.PDF** *)*. For these devices and future 8-, 16-, and 18-Mbit RDRAMs, the protocol by which the memory controller and the RDRAMs communicate has been improved to sustain 425 Mbytes/s for bursts of 32 bytes. This is an improvement of about 30% over the original protocol.

The peak fill frequencies of 8- and 16-Mbit RDRAMs are 533 and 267 Hz, respectively. This capability puts these devices squarely in competition against other high-performance graphics and media-acceleration RAMs. The fill frequency of a 600-Mbytes/s 64-Mbit RDRAM will be 75 Hz. From this perspective, these devices are comparable to the highest-performance ×32 64-Mbit SDRAMs operating at 150 MHz.

In 1995, the first system to use Rambus RDRAMs—a Silicon Graphics workstation—began shipping *(see* **0910MSB.PDF** *)*. Other RDRAM-based products, including the Nintendo 64 (previously Ultra-64) video game, Chromatic's Mpact multimedia processor *(see* **091404.PDF** *)*, and Cirrus Logic's GD5462 graphics accelerator should all be shipping in substantial volumes during the second half of this year. Rambus expects more than 10 million RDRAMs will ship this year, establishing RDRAM as a major player in the graphics market.

The key to Rambus's success in the PC main-memory market is its ability to sign system-logic chip-set vendors—Intel in particular. The 64-Mbit devices promise to be cost competitive for midrange and even high-end UMA systems. But Rambus's strategy of requiring royalties from chip-set and other memory-controller vendors has created at least the perception for many system architects that Rambus's advantages come significantly burdened.

MoSys Aims for Graphics Market

MoSys MDRAMs *(see* **091703.PDF** *)* are marketed primarily as graphics RAMs. There is little about them, however, that is particularly specific for graphics applications. Indeed, several of the early design wins are in embedded systems with no graphics component whatsoever.

Unlike other new DRAM architectures, which wrap a high-speed interface around a generic DRAM core, an MDRAM starts with an unconventional DRAM core consisting of lots of small (32-Kbyte) high-speed banks. The fastest MDRAM has row and column access times of 30 ns and 12 ns respectively. These low latencies are especially effective when combined with a low-overhead 16-bit interface that generates a peak bandwidth of 667 Mbytes/s at 167 MHz.

MDRAM fill frequencies range from 320 Hz for the 10-Mbit 100-MHz device to an outstanding 1,333 Hz for the 4-Mbit 167-MHz part. Even when maximally populated, the peak fill frequency of an MDRAM memory system is a respectable 67 Hz. Though these figures are generally consistent with high-performance graphics systems, MDRAMs are a credible high-performance main-memory solution for applications that require only 8 Mbytes or less per channel.

The primary business limitation: MoSys is a small company with foundry relationships with four second- and third-tier DRAM vendors. The absence of any suppliers among the top ten DRAM vendors is potentially worrisome for anyone considering a very high volume application. Shipments of the Tseng Labs ET6000 *(see* **091703.PDF** *)* should generate a significant volume of MDRAM shipments in 2H96. Although poorly suited for PC main memory, the MoSys chip may find a niche for graphics.

EDRAM and CDRAM Seek Niches

Enhanced DRAMs are a high-speed memory chip sold by Enhanced Memory Systems, a subsidiary of Ramtron *(see* **070205.PDF** *)*. EDRAMs are marketed as a replacement for both an SRAM-based L2 cache and a FPM or EDO main memory. They are able to match or exceed the performance of an SRAM-based cache because of their low latency—the

For More Information

Contact Enhanced Memory Systems (Colorado Springs, Col.) at 719.481.7000; fax 719.488.9095.

Contact Micron (Boise, Idaho) at 203.368.3900; fax 208.368.4431 or on the Web at *www.micron.com*

Contact MicroUnity Systems Engineering (Sunnyvale, Calif.) at 408.734.8100; fax 408.734.8141.

Contact MoSys (San Jose, Calif.) at 408.456.2370; fax 408.321.0780 or send e-mail to *cognac@mosys.com.*

Contact Rambus (Mt. View, Calif.) at 415.903.3800; fax 415.965.1528 or on the Web at *www.rambus.com.*

Contact the SyncLink Consortium at 415.961.0305; fax 415.961.3530.

Most major DRAM vendors can provide additional information about EDO or synchronous DRAM; contact your local sales office.

best row and column access times of 30 and 12 ns, respectively, are roughly half those of mainstream EDO devices—and limited but significant changes to the logical interface. In particular, these changes allow precharges, refreshes, and writes of the DRAM array to be overlapped with column read accesses.

Unfortunately, EDRAMs are available only at the 4-Mbit level and in ×1 and ×4 organizations, though ×8 versions are sampling. At 83 MHz, the ×4 and the ×8 EDRAMs have fill frequencies of 83 Hz and 166 Hz, respectively. Until recently, the supply of EDRAMs has been very limited. However, in mid-1995, IBM signed on as a second source for EDRAMs, and parts from IBM are now entering production. Hopefully, this new source of supply will allow Enhanced Memory Systems to better service the embedded arena that has been its focus and to branch out into larger market segments.

Another way to replace an L2 cache with DRAM is to use Mitsubishi's cached DRAM *(see* **070205.PDF** *)*. There are currently a pair of ×16 CDRAMs that combine 16 Kbits of SRAM on the same die as 4 or 16 Mbits of DRAM. The organizations include separate address and control signals for the two memories, a collection of 128-bit data paths and buffers connecting the memories and the I/O pins, and a synchronous interface capable of speeds up to 100 MHz.

The combination of independent control of the two arrays and wide internal data paths effectively provide opportunities for parallelism and high peak and sustained bandwidths. As a result of its high-bandwidth, low-latency path to the DRAM and in spite of its small size, the SRAM matches the effectiveness of a large SRAM-based L2 cache.

Despite its technical strengths, a modest 15% price premium over EDO DRAMs, and a second source (Samsung), CDRAM has only a few design wins in the arcade videogame market. This device is unlikely to gain significant PC design wins.

SyncLink and MediaRAMs in Development

Two other new DRAM architectures have been proposed. SyncLink begin its life as an appendix to the IEEE RamLink standard (1596.4). Though RamLink defines a ring topology as its primary physical layer, this appendix defines an alternate physical interconnect on which the RamLink protocol could be sent. It consists of two unidirectional buses connecting a memory controller with a collection of SyncLink DRAMs. A byte-wide high-speed bus carries control, address, and write data from the memory controller to the slaves while a separate 16-bit half-speed bus carries read data from the DRAMs to the memory controller.

In late 1994, just as RamLink was being finalized, elements of the DRAM community noticed SyncLink held promise as an alternative to Rambus, but it needed work. Rather than reopen the RamLink definition, a new IEEE working group was formed to refine and essentially commercialize SyncLink. This work proceeded over the course of 1995, partially within the open forum of the IEEE working group and partially within the confines of the SyncLink Consortium, an association of eight DRAM and system companies interested in ensuring the ultimate success of SyncLink.

In recent months, however, the SyncLink Consortium has abandoned the original dual-bus structure of the RamLink SyncLink. The new SyncLink will be released to the IEEE Working Group for review in the near future. Actual products are unlikely to appear in the next two years.

The latest entrant into the DRAM architecture war is MicroUnity. In a 1996 CompCon paper, the company described a MediaRAM DRAM that resides on its MediaChannel, a high-speed byte-wide ring *(see* **091402.PDF** *)*. With point-to-point signals, low voltage swings, and no more than four devices per ring, MediaChannel offers the potential for subnanosecond cycle times and sustained bandwidths of 1 Gbyte/s or greater.

MicroUnity announced it has licensed MediaChannel to one major DRAM vendor. The two companies are co-developing a general-purpose MediaRAM, targeting availability in 1997. Until MicroUnity discloses more details about its licensing plans for DRAMs and memory controllers, however, the applicability of these memories outside MicroUnity's limited sphere of influence is unclear. Many of the important business and technical issues that determine the success of a DRAM architecture are still unknown.

Fill Frequencies Span a Wide Range

With so many architectures in the market, it is increasingly difficult to keep their key characteristics, strengths, and weaknesses straight. Table 2 illustrates the broad range of just a few of those characteristics. The architectures are grouped according to the amount of change they embody over the traditional DRAM architecture: fast-page and EDO are conventional; BEDO, EDRAM, and SDRAM are evolutionary in that they preserve the separate data and multiplexed address buses but otherwise significantly change the logical or temporal interface; RDRAMs and MDRAMs are revolutionary in that address, data, and/or control are carried on a single shared bus. One of the primary tradeoffs open to the system architect is backward compatibility at either the chip- or SIMM/DIMM-level versus the overall bandwidth and granularity of the memory system.

The peak fill frequencies of the devices represented in Table 2 span a factor of more than 50. As expected, fill frequencies decline as the density increases from 4 to 64 Mbits. The sustained fill frequencies, which are more appropriate for comparisons between devices, span a comparable range. The gap between the peak fill frequencies and the typical-case sustained fill frequencies varies by architecture, transfer size, and application and controller characteristics.

At the 16-Mbit level, the revolutionary devices are essentially graphics devices: the level of performance and granularity they provide suit that market well. The 64-Mbit RDRAMs are technically well placed to meet the needs of UMA PC systems, with their high peak and sustained band-

	Conventional		Evolutionary			Revolutionary	
	Fast-page mode	EDO	Burst EDO	EDRAM	SDRAM	RDRAM	MDRAM
Proponent	—	JEDEC	Micron	Enhanced Memory Systems	JEDEC	Rambus	MoSys
Vendor/manufacturer	All	All soon	Micron only for now	NPNX, IBM	Many	Toshiba, NEC, Oki, LG, Samsung, Hitachi	IDT, Oki, TSMC, Siemens
Densities	All	All	16 Mbits	4 Mbits	2, 4, 16 Mbit	8, 16, 18 Mbits	4–10 Mbits
Widths	1–16 bits	1–32 bits	4–16 bits	1, 4, 8 bits	4, 8, 16 bits	8, 9 bits	16 bits
Interface type	Asynchronous	Asynchronous	Pseudo-synchronous	Asynchronous	Synchronous	Synchronous	Synchronous
Best common row access time	60 ns	60 ns	52 ns	30 ns	60 ns	60 ns	30 ns
Data sheet max operating freq	25 MHz	40 MHz	66 MHz	83 MHz	100 MHz	300 MHz	167 MHz
In-system max operating freq	15–25 MHz	33 MHz	66 MHz	83 MHz	66–83 MHz	300 MHz	167 MHz
Peak fill frequency (Hz)	1M×4: 20	1M×4: 33	—	512K×8: 166	256K×16: 264	8 Mbit: 533	4 Mbit: 1,333
	1M×16: 20	1M×16: 33	1M×16: 66	—	1M×16: 66	16 Mbit: 266	10 Mbit: 533
	—	2M×32: 17	4M×16: 17	—	2Mx32: 50	64 Mbit: 75	—
Memory system width	64 bits	64 bits	64 bits	64 bits	64 bits	9 bits	16 bits
Memory system granularity (Mbytes)	1M×4: 8	1M×4: 8	—	512K×8: 4	256K×16: 2	8 Mbit: 1	4 Mbit: 0.5
	1M×16: 8	1M×16: 8	1M×16: 8	—	1M×16: 8	16 Mbit: 2	10 Mbit: 1.25
	—	2M×32: 16	4M×16: 32	—	2M×32: 16	64 Mbit: 8	—
Controller memory interface total pin count	~110	~110	~120	~120	~120	31	~50

Table 2. The many DRAM architectures available for main-memory designs have very different characteristics, strengths, and weaknesses.

width and very good 8-Mbyte granularity. Fundamentally, the revolutionary alternatives offer more bandwidth per pin on both the DRAM and the memory controller. This advantage can be leveraged as either a pin-count advantage at a comparable bandwidth level, or as a bandwidth advantage at a comparable pin count. The option of reducing the memory controller pin count is especially important in tightly coupled UMA systems, because in such systems the chip set is the focal point of several wide, high-bandwidth buses.

Today, at 66-MHz the 16-Mbit burst EDO and SDRAM provide adequate bandwidth to support UMA and good granularity with the ×16, ×8, and ×4 devices for small, medium, and large systems, respectively. At the 64-Mbit level, SDRAMs in particular will be able to maintain a competitive position by increasing the width of the devices—to ×32 for small systems and ×16 for medium and larger systems—and by increasing the frequency of operation. The change from LVTTL to SSTL will play an important role in facilitating frequencies up to or in excess of 100 MHz, even in DIMM-based memory systems.

EDO to Give Way to SDRAM in 1997

Last year marked the beginning of the major switch from FPM to EDO for PC and other system main memories. By the end of this year, that transition should be nearly complete. This transition was particularly easy because of the very limited nature of the change in the EDO DRAM interface. EDO also marked the first time that a main memory technology was chosen, at least in part, for its marketing appeal: "Performance EDO DRAM." As the DRAM market continues to fragment and as more options become technically practical, the tradeoffs among cost, performance, and marketability will become more complicated.

The next major transition—to SDRAMs—is beginning now and should be largely complete by the end of 1997. Most of the major barriers to the adoption of SDRAMs have already been eliminated. Hopefully, those that remain will create only limited confusion and headaches but not adversely affect system reliability.

As the 64-Mbit DRAM becomes cost competitive in late 1997 or 1998, the DRAM producers and consumers face a significant question: can fewer controller and DRAM pin counts yield lower overall system cost for Rambus or one of the other emerging revolutionary alternatives? If so, another transition will begin; if not, the cost burden these alternatives carry must be weighed against the exposure, marketability, and performance they offer. If the cost premium is at all significant, most likely the next major transition in memory technology will be delayed, either until the system requirement fill frequencies rise beyond what high-frequency SDRAMs can deliver—for example, through the need for higher-performance multimedia and 3D graphics in a UMA PC—or until 256-Mbit DRAMs become cost competitive. M

Dr. Steven Przybylski is an independent consultant based in San Jose (Calif.) and author of the second edition of New DRAM Technologies: A Comprehensive Analysis of the New Architectures. *For more information, see* **1006RES.PDF**. *He will also present a full-day seminar at this month's PC Tech Forum.*