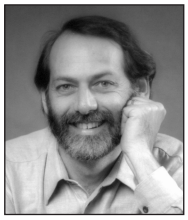
Direct RDRAM (Almost) Arrives in PCs

Memory Bandwidth Boosted—But Benefits Lie in Future



After years of anticipation, the first personal computers with Rambus main memory are on the verge of shipping. The fight to establish Rambus as a mainstream solution has been far longer than its backers anticipated, and it has produced an extraordinary amount of contention and controversy. The Rambus approach offers a lot of long-term benefit, but in the near term it is less compelling. Much of the contention derives from Intel's staunch backing of Rambus with an eye toward the long-term benefit, while critics focus on the short-term costs.

The debut of Direct RDRAM PCs has been seriously marred by Intel's last-minute pull-back on its 820 chip set, the first to support RDRAMs (see MPR 10/6/99, p. 30). Sporadic memory errors in some systems with three RIMM sockets caused Intel to delay the announcement and shipment of the 820, which was already months late. The problems that led to this delay give fuel to those who have criticized the Direct RDRAM electrical interface as pushing the limits of what the PC industry can reliably produce.

Intel's last-minute postponement of the 820 launch and its inability to give a time frame for the 820's debut have caused much unhappiness at some of its partner companies. Dell had already mailed catalogs featuring 820-based systems, and now it is left without a new high-end line going into the Christmas season. Samsung, which invested heavily to be the leading supplier of RDRAMs, now finds its first-to-market advantage squandered. The company has halted all RDRAM wafer starts (but will complete work in process).

The delay means that 133-MHz SDRAM (PC133) systems will ship before RDRAM systems. In fact, the only way to build a system today with a 133-MHz-bus Pentium III processor and AGP graphics is to use Via's PC133 chip set—surely an embarrassment for Intel. The 820 delay is a big win for Via; Intel's 810E supports the 133-MHz bus but not AGP or 133-MHz DRAM.

Benefit Hard to Demonstrate

Initial RDRAM PCs will be limited to high-end systems because of the price premium charged today. The volume of RDRAMs is limited in any case, so they couldn't serve the mass market, even if the price and other attributes were aligned. As more vendors begin shipping and volumes increase, the premium will shrink. But not until late 2000, by the most optimistic estimates, or 2002, by conservative ones, will RDRAM be popular in low-cost PCs.

Ironically, even though RDRAMs have twice the peak bandwidth of PC100 memory, they won't offer any performance boost for most of today's PC applications or for popular benchmarks. RDRAM makes more bandwidth available, but most of today's software, with today's processors, just doesn't need it. Intel developed its own synthetic benchmark to show off the bandwidth offered by the new systems.

Some independent benchmarks show significant performance improvements on data-intensive tasks, such as loading and saving large PhotoShop files. Future PC applications should show more benefit from the higher memory bandwidth. Graphics applications that use very large, complex texture maps, for example, combined with next-generation graphics adapters that use AGP 4× to read these texture maps from main memory, will place a much greater load on the memory system. Some memory bandwidth may be consumed by background tasks, including network activity. And applications that are optimized with SSE's streaming instructions can stress the memory more.

Just how long it takes before popular applications and system-usage patterns deliver benefit to typical PC users from the bandwidth capability of RDRAM systems remains to be seen. It may not occur until after the early RDRAM systems have been retired by their original purchasers.

Heat-seeking gamers are natural candidates for high-end systems, as are certain corporate PC buyers who don't want to take the risk that the life of new systems will be shortened by the emergence of more demanding software. Some such users are likely to buy 820-based systems when they become available. But for many users, the value proposition is unconvincing. Even performance-driven consumers may not be willing to pay more for promised future performance gains, especially if PC133 systems offer equal or better performance for their applications.

RDRAM-based systems will probably have reasonable success—but it will be largely independent of their use of RDRAM, not because of it. Intel's 820 chip set, with AGP 4× and other improvements, primarily supports RDRAM (although SDRAM can be used with an external translator chip). The only other Intel chip set that will support 133-MHz-bus processors this year is the 810E, which has integrated graphics that won't satisfy a 3D-performance-oriented user.

Memory Contention

What many buyers no doubt would like to have is a 133-MHz-bus, PC133 version of the 440BX—but Intel does not offer such a chip set. Intel originally put all its eggs in the

RDRAM basket, but it has belatedly decided to offer a PC133 chip set. The features of this chip set, which is not due until sometime in 1H00, have not been disclosed.

Despite Intel's plans to offer a chip set to support PC133 SDRAM, the company has largely dismissed its benefit. Indeed, PC133 SDRAM has its limitations. Initial PC133 memory will have a three-clock CAS latency, compared with two clocks for the fastest PC100 memory, minimizing its performance gain. But at least some DRAM makers expect to offer faster PC133 memory in 2000, boosting its performance.

Even so, PC133 at best offers a modest incremental improvement over the ubiquitous PC100 SDRAM. The speed probably can't be increased much further without changing the interface, and it requires a 64-bit-wide bank to deliver acceptable bandwidth.

RDRAM's biggest advantage is its high per-pin bandwidth: it delivers more bandwidth on fewer pins. By adding a second Rambus channel in the chip set—which takes far fewer pins than doubling the width of a SDRAM bank—bandwidth can be doubled. This approach is far more expensive with SDRAM. Intel's new 840 chip set (see MPR 10/25/99, p. 28) provides two RDRAM channels for dual-processor workstations. Today's PCs don't need the bandwidth two channels deliver, but with a 1-GHz processor and another generation of software, that bandwidth could become an important advantage, even in the PC market.

The small number of pins required by an RDRAM interface will also make integration of the interface onto the processor chip compelling, eliminating the latency of the chip set. And the fact that a single RDRAM can deliver great bandwidth is the driving force behind the memory's use in small systems, such as Sony's PlayStation2.

Some critics assert that RDRAM performance suffers because, even though it offers higher bandwidth, it has greater latency. Perhaps this concern is left over from first-generation RDRAMs, which did have long latencies. Direct RDRAM, however, roughly matches SDRAM latency and has unquestionably higher bandwidth. It is true that RDRAM doesn't improve latency, which is unfortunate—but there is really no doubt that it is a higher-performance memory than PC133. The question is whether the additional performance buys much initially and is worth the price premium.

Beyond PC133

DDR SDRAM is the next step along the PC133 path, with several vendors planning to offer PC266 chips in 2000. It isn't yet clear how systems with DDR SDRAM will perform (one weakness is that the address and command bandwidth isn't increased), how manufacturable the chips will be, or whether they will be cheaper than RDRAMs. Nevertheless, there is backing for the technology from memory, chip-set, and PC makers, and prototype systems have been shown. DDR SDRAM could become the memory of choice in Athlon-based systems, potentially giving them an additional advantage; Intel has no plans to support DDR SDRAM for PCs.

Much of the anti-Rambus sentiment comes from DRAM makers that resent having to adopt a standard not of their making, and for which they must pay royalties to Rambus. Chip-set makers aren't happy about having to license the Rambus interface, either. These factors alone probably will lead to strong support for PC133 and, in time, DDR SDRAM. Intel and Rambus recently took a step toward appeasing the memory makers by establishing the Rambus Implementers Forum, giving the chip companies more control over the future evolution of the specification.

RDRAM's long-term challenger is DDR-2, which doesn't yet have a public definition. This next-generation design will diverge from SDRAM in a more radical way than DDR. The intellectual property that was developed for the now-abandoned SLDRAM effort has been turned over to the DDR-2 group, which is called Advanced Memory International and has backing from several DRAM suppliers.

Not enough is known to evaluate the technological merit of DDR-2, but given that it will debut a couple years after RDRAM and will have the benefit of hindsight, it has a good chance to offer some advantages. Furthermore, it will be controlled by the DRAM makers, without royalties to a third party, which gives it an emotional and economic advantage. Of course, RDRAM won't stand still either; Intel and Rambus are no doubt already thinking about their next-generation design.

Long Term Still Looks Good

During the coming year or two, there is little doubt that RDRAM will face stiff competition from PC133 and DDR SDRAM. In the long run, however, it still seems likely to dominate in mainstream PCs. Servers will be harder to penetrate, since achieving high bandwidth is easier in a large memory system, and the price premiums and chips-per-channel limits of RDRAM are more troublesome. DDR may have a long-term role in servers; even Intel doesn't have immediate plans for a server chip set with RDRAM support, and the company acknowledges that its future server chip sets might support DDR.

RDRAM should really begin to shine in 2001, when increased supply and optimized designs make price premiums minimal, applications will have evolved, and processors and graphics controllers will have increased in speed. Intel's Willamette will have a much faster bus than Pentium III and will run at speeds above 1 GHz, creating a far greater demand on memory. Also in this time frame, highly integrated processors for mainstream PCs may appear that integrate the RDRAM interface on the processor chip.

The Rambus train certainly isn't moving as fast as its advocates would like, and it has been temporarily derailed by the delays with the 820 chip set. It seems likely, however, that the train will get back on the track, and after a couple years of gathering speed, it may be unstoppable. M

See www.MDRonline.com/slater/rdram *for more on this subject. I welcome feedback at* mslater@mdr.cahners.com.