Profusion Lowers Cost of Eight-Way Servers

Intel's High-End Chip Set From Corollary Finally Ships

by Peter N. Glaskowsky

When we first wrote about Corollary's Profusion chip set (see MPR 9/16/96, p. 9), it was expected to ship in 1997 as an eight-way Pentium Pro chip set. Shortly thereafter, Intel purchased the company, and Corollary has spent the last two years adapting its original design to the faster Pentium III bus interface. Earlier this year, Intel began shipping Profusion board sets (the chips are not available separately) to a number of major server OEMs. Profusion-based systems are already setting records for price/performance on key server metrics such as the Transaction Processing Council's TPC-C benchmark.

These results confirm the robustness of the original Profusion design. In the last three years, very little has changed in the design apart from the shift from 66-MHz buses to 100-MHz buses.

Profusion Breaks Four-CPU Barrier

Profusion still combines three processor buses, two SDRAM buses, and a programmable cache-coherency mechanism in just two ASICs, as Figure 1 shows. Two of the three Pentium III buses each accommodate up to four processors, while the third CPU bus connects the system's four PCI bus bridges.

Multiple buses are necessary for several reasons. The Pentium III system interface can scale to only four processors because of electrical-loading, signal-length, and bus-protocol limits (see MPR 5/30/95, p. 1). The available bandwidth on a single segment is sufficient for only four processors, so extending the bus beyond this point would not be useful.

Corollary could have distributed the PCI bridges between the two processor buses, but in a high-performance server, a single PCI bus can require as much bandwidth as a processor. Corollary's approach separates PCI transactions from the processor buses, giving PCI devices independent access to main memory and more sustained throughput. While the original Profusion design used Intel PCI bridges from a Pentium Pro chip set, Intel never updated these devices to support the 100-MHz Pentium III bus. Instead, Corollary worked with Compaq to develop new PCI bridge chips. Each of the new chips provides a 64-bit, 66-MHz PCI interface with four times the throughput of the original 32-bit, 33-MHz bridges. In 66-MHz mode, only two slots are supported; the chips can also be used in a 33-MHz mode that allows six slots.

Two banks of SDRAM, operating independently and concurrently, satisfy the combined main-memory bandwidth demands of eight CPUs and four PCI buses. Up to 32 DIMMs are supported by Corollary's reference design. With 1G DIMMs made from 256-Mbit DRAMs—currently available, but at very high prices—up to 32G of main memory can be configured. This is the same maximum array size as the original design, which supported up to 64 DIMMs using a proprietary stub series-terminated logic (SSTL) interface. Today's Profusion systems use commodity PC100 DIMMs with error-correcting code (ECC), allowing much lower system costs.

The two banks of SDRAM are interleaved on cache-line boundaries, with even cache lines in one bank and odd cache lines in the other. Since memory accesses under Windows NT and Unix show no significant preference for even or odd cache lines, this interleaving allows efficient sharing of memory accesses between the two banks.

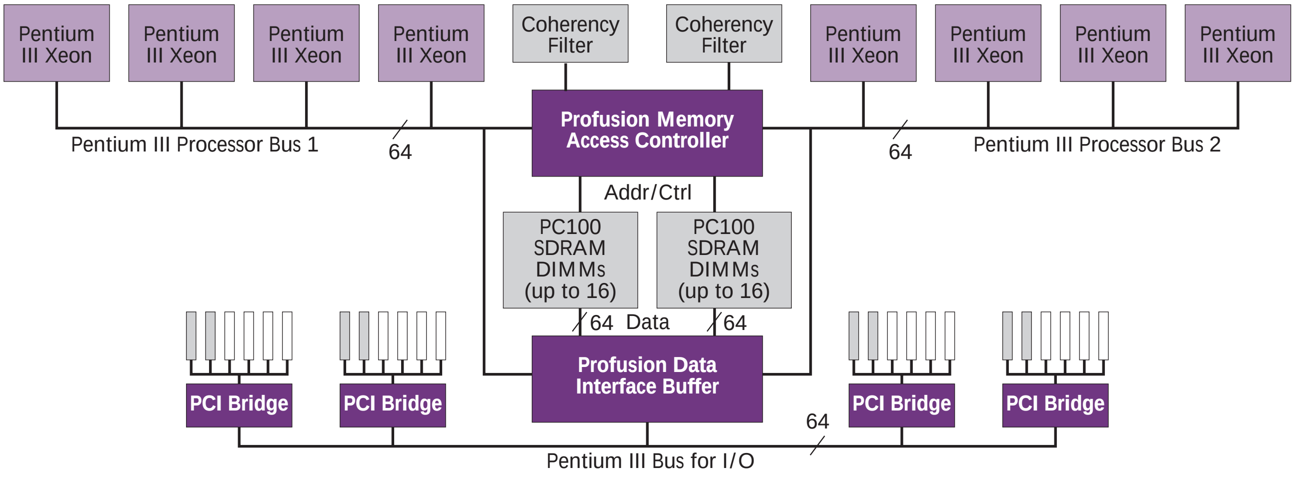

Figure 1. The Profusion reference design consists of two custom ASICs and four PCI bridge chips, eight Pentium III Xeon processors, and commodity SRAM and DRAM memory devices. The PCI bridges may be configured for two 66-MHz slots or six 33-MHz slots.

The dual-bank arrangement also helps improve system availability. The two banks are electrically and logically isolated. Even if one bank fails, the system can be rebooted to run on the other bank of memory and one CPU bus. OEMs can take advantage of this mode to ship entry-level systems with four processors and just one bank of DRAM. Such systems can later be upgraded to 8-way configurations.

Coherency Filters Boost Bus Efficiency

Except for their higher operating frequency, the cache-coherency filters shown in Figure 1 are virtually identical to those found in the original Profusion design. These filters consist of SRAMs that work like the tags for a direct-mapped L3 cache—but without the data. The filters are used by Profusion to check transactions on one processor bus against the contents of L2 caches on the other bus.

Coherent reads and writes that miss in a processor's L2 cache and appear on its bus are tested against the coherency filter for the opposite bank of processors. A main-memory transaction is initiated in parallel.

If the transaction hits in the coherency filter, Profusion starts an intervention request on the opposite bus to read or invalidate the L2 cache line(s) that may be held by one or more of those four processors. It is possible the corresponding L2 cache line(s) may already have been invalidated, since the Pentium III does not always generate a bus transaction when it alters the status of L2 cache lines. If the processor L2 caches do not have the requested data, or if the original transaction missed in the coherency filter, the transaction is completed to main memory.

Each filter is implemented in one, two, or four late-write 100-MHz synchronous SRAMs. Each 256K × 18 SRAM represents 8M of this virtual cache, for a total of up to 32M for each bank of four processors.The coherency algorithm is stored in SRAM inside the Profusion chip set and controls a simple protocol engine, allowing the algorithm to be tuned or changed later. This internal SRAM is initialized by the system BIOS before the caches are turned on.

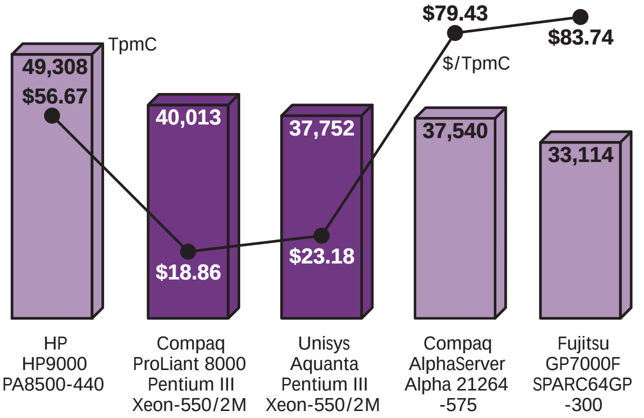

Figure 2. Profusion-based systems from Compaq and Unisys compare well with other 8-way systems on the TPC-C benchmark, and they have a sizable price/performance advantage. (Source: TPC)

New Packaging Techniques Reduce Chip Count

Corollary considered and rejected other multiprocessor architectures, such as the classic hierarchy of buses that would have been simpler to implement. The company's simulations showed these alternatives would have lost too much performance to bus arbitration and cache snooping operations. Corollary also rejected complex point-to-point switched architectures such as that used by HotRail (see MPR 7/12/99, p. 12) and some high-end RISC server vendors.

Instead, Profusion uses modern high-pin-count packages to support eight processors with just two devices, one for control and one for data. Rather than two major buses, the Profusion ASICs connect to five. Instead of layers of dual-ported cache and memory controllers, Profusion has just one five-port memory controller.

This results in highly complex devices: the Profusion Memory Access Controller (MAC) has 800,000 gates. The Data Interface Buffer (DIB) has 80,000 gates plus a 64-cache-line (2K) 10-port SRAM array. The MAC is fabricated in a 0.25-micron process and packaged in a 624-contact ceramic BGA. The less-complex DIB is built in a 0.35-micron process and fits in a 655-contact plastic BGA. The current chips have more than twice as many logic gates as the original Profusion designs and are said to be much more efficient.

Integration Improves Performance

The proof of the pudding, as usual, is performance. Figure 2 shows the TPC-C benchmark results for the Profusion system against the scores of other 8-way systems. Only one other 8-way system exceeded the TpmC score of Compaq's Profusion-based ProLiant 8000, and that HP system costs 3.7 times as much. The fastest system to undercut Compaq's price, a 4-way Xeon machine from Fujitsu, offers 36% less performance for 96% of the price of the 8-way system.

Multiple Operating Systems Supported

Profusion is compatible with Intel's MPS 1.4 symmetric multiprocessing specification (see MPR 5/9/94, p. 12). This specification allows Profusion to support essentially all available x86 multiprocessor operating systems. Implementations for Windows NT 4.0, Windows 2000, SCO UnixWare 7.1, and Novell NetWare 5.0 are fully supported by Intel. Profusion is also compatible with Linux, Solaris, and various older versions of the supported OSs.

Profusion will compete against hierarchical 8-way SMP, switched architectures, CC-NUMA, clustering, and other techniques. Because of its low chip count, Profusion should offer better price/performance than most other MP architectures with up to 16 processors. Intel plans to explore clustering technology to achieve even higher performance on distributed applications.

Given Corollary's excellent track record with PC-based SMP and Intel's ability to create effective products, we expect Profusion to deliver on its promises and become a popular choice for high-end servers. M