Profusion Adds Processors and Performance

Corollary Creates Credible Chip Set for Eight-CPU Pentium Pro Servers

by Peter N. Glaskowsky

Profusion, a new multiprocessing architecture from Corollary, is the first to allow OEMs to scale Pentium Pro servers beyond four processors on a single motherboard. Corollary, an old hand in x86-based symmetric multiprocessing, developed C-bus I for 486-based multiprocessor systems and the widely used C-bus II for Pentium SMP *(see* **0817MSB.PDF***)*. The company's new architecture combines three Pentium Pro buses, two SDRAM buses, and a programmable cache-coherency mechanism. By taking advantage of the latest packaging technology, all of the proprietary Profusion logic fits into just two 599-pin ASICs.

Profusion Breaks Four-CPU Barrier

Figure 1 shows the Profusion architecture as implemented in Corollary's reference design. Two of the three Pentium Pro buses can each accommodate up to four CPUs each, while the third bus is used for peripherals and PCI expansion.

Multiple buses are necessary for several reasons. The Pentium Pro system interface can scale to only four CPUs because of electrical-loading, signal-length, and bus-protocol limits *(see* **090701.PDF***)*. Four Pentium Pro CPUs alone demand about 400 Mbytes/s of bus bandwidth, close to the effective performance of a single bus segment.

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, yielding a more symmetrical design and allowing better scalability to future 64-bit and/or 66-MHz PCI implementations.

Two banks of SDRAM, operating independently and concurrently, satisfy the combined main-memory bandwidth demands of eight CPUs and four PCI buses. Up to 64 DIMMs are supported by Corollary's reference design. With DIMMs made from 256-Mbit DRAMs, this will allow 32G of main memory in future systems.

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. This scheme requires both banks to be configured identically, however, so the granularity of upgrades is two DIMMs at a time.

Although the initial ASICs have low-voltage TTL memory interfaces and the memory buses in the reference design operate at only 66 MHz, Corollary has chosen to specify Stub Series-Terminated Logic (SSTL) DIMMs that can operate at up to 100 MHz because of SSTL's reduced voltage swing. If Intel releases P6 processors with faster system buses, as seems inevitable, customers will be able to upgrade to faster motherboards and reuse these faster DIMMs. This capability is considered important by server users, who may invest far more in DRAM than in the rest of the system.

Coherency Filters Boost Bus Efficiency

Two cache-coherency filters, shown in Figure 1, are used by Profusion to check transactions on one processor bus against the contents of L2 caches on the other bus. Each filter is implemented in one to four late-write 66-MHz synchronous SRAMs and works like the tags for a direct-mapped L3

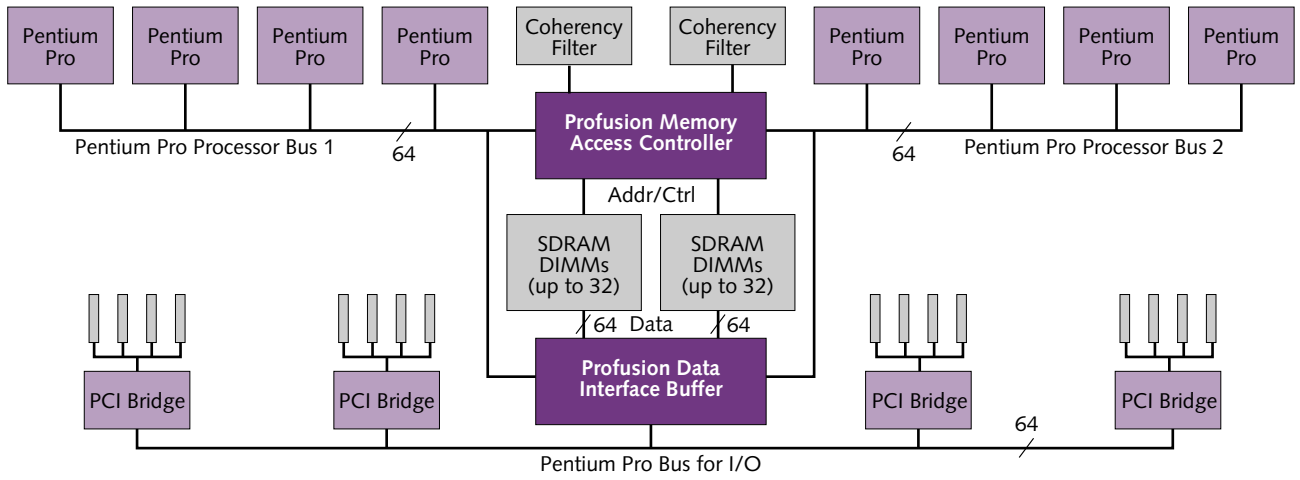

Figure 1. The Profusion reference design consists of two custom ASICs (shown in dark purple), eight Pentium Pro CPUs and four PCI bridges from Intel (shown in light purple), commodity SRAMs for the coherency filters, and up to 32G on a total of 64 SSTL DRAM DIMMs.

For More Information

Pricing and availability for the Profusion ASICs and reference design have not been announced. Contact Corollary (Irvine, Calif.) at 714.250.4040 or on the Web at *www.corollary.com.*

cache—but without the data. Each 256K × 18 SRAM represents 8M of this virtual cache, for a total of up to 32M of L3 for each bank of four processors.

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 Pentium Pro does not always generate a bus transaction when it alters the status of L2 cache lines. In this case, or if the original transaction missed in the coherency filter, the transaction is completed to main memory.

The coherency algorithm itself is stored in SRAM inside the Profusion chip set, allowing the algorithm to be tuned or changed later. This internal SRAM is initialized by the system BIOS before the caches are turned on.

New Packaging Techniques Reduce Chip Count

Corollary investigated several multiprocessor architectures that would have supported eight Pentium Pro CPUs. A classic bus hierarchy, shown in Figure 2, would have been simpler to implement. Intel's existing PCI bridge and DRAM controllers could have been used. Also, the L3 cache controllers would have been connected to only two buses, permitting the use of inexpensive packages.

Instead, Profusion uses modern high-pin-count packages to collapse the entire hierarchy into 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.

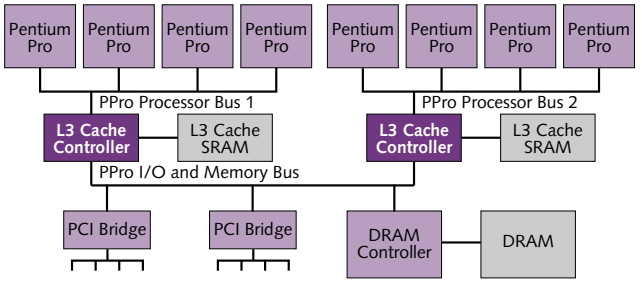

Figure 2. Corollary considered and rejected this conventional hierarchy-of-buses architecture, which is simpler than Profusion but slower because of extra bus delays for read requests.

This results in highly complex devices: the Profusion Memory Access Controller (MAC) has 350,000 gates. The Data Interface Buffer (DIB) has 20,000 gates plus a 64-cache-line (2K) 10-port SRAM array. The MAC and DIB are fabricated in a 0.5-micron process and packaged in 599-pin tape ball grid array (TBGA) packages.

Profusion also uses four of Intel's 82454GX PCI bridge chips from the commodity 450GX chip set, further leveraging Corollary's use of the Pentium Pro bus standard.

Integration Improves Performance

This partitioning also produces a substantial performance advantage over conventional approaches. For example, in the hierarchy-of-buses architecture, L3 cache misses on one processor bus could result in snoop requests on the other. These would have to transit three buses—each separately arbitrated—and then return. Corollary's simulations show that this case would represent about 20% of all external requests and would result in a typical delay of 40 bus-clock periods. Thirty percent of requests would be satisfied from main memory in 24 cycles, and 50% of the requests would hit in the L3 cache and take 7 clocks to satisfy. The average latency, therefore, would be 18.7 cycles.

In the Profusion architecture, all read requests are satisfied from main memory or by snoops to the opposite processor bus. The same 20% of all read requests will result in interventions to the other bus, but because they cross only one device and two buses, they complete in only 24 clocks. The other 80% will result in main-memory transactions that are completed in 7 cycles, as fast as allowed by the Pentium Pro's system bus. These improvements reduce the average latency for Profusion to just 10.4 cycles, 56% of the time required by the hierarchical approach.

MPS 1.4 Compliance Guarantees Compatibility

Profusion is compatible with Intel's MPS 1.4 symmetric multiprocessing specification *(see* **080603.PDF***)*. This will allow Profusion to eventually support all available x86 multiprocessor operating systems, but Windows NT 4.0 and SCO UnixWare 2.1.0 will be ported first.

Corollary has developed a complete reference system for Profusion. OEMs can use the design as is or use it as an example of how to solve the difficult mechanical and thermal design issues in high-end SMP servers.

Profusion will compete against hierarchical eight-way SMP, CC-NUMA, clustering, and other techniques. It should offer better price/performance than other MP architectures with up to 12 to 16 processors. Beyond that, Profusion can be combined with clustering to achieve higher performance on distributed applications. Given Corollary's excellent track record with PC-based SMP, we expect Profusion to deliver on its promises and become a popular choice for high-end PC servers. M