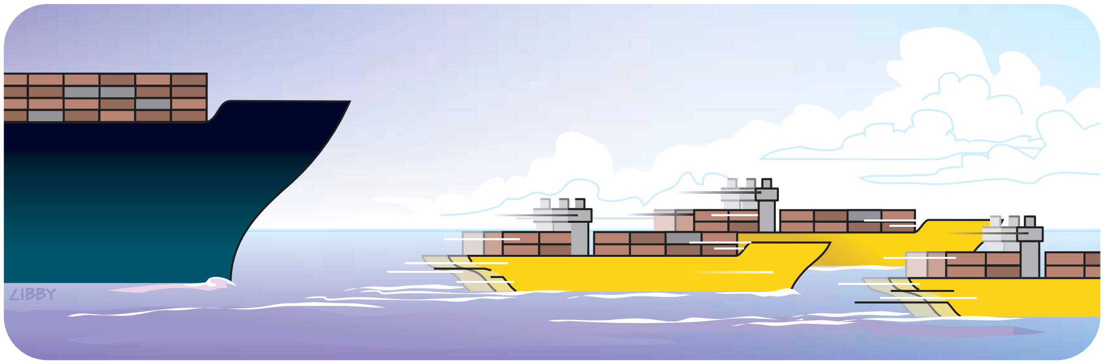
Sun: Better Computing Through Threads

Quick Note

Gordon Haff
9 July 2003

A fundamental shift is underway among designers who create microprocessors for servers. For years, their primary goal was to extract as much performance as possible from a single stream of instructions using a grab-bag of techniques collectively known as "instruction level parallelism" (ILP). But that hole is getting mined out; refinements of old techniques are having a tougher and tougher time delivering system-level performance gains.[1] As a result, designers are starting to move toward a complementary set of techniques called "thread level parallelism" (TLP). Unlike ILP, the goal of TLP isn't necessarily to run a single task as fast as possible, but rather to make the large and varied collection of tasks on a typical server run as fast as possible in the aggregate.

Sun Microsystems is the company that has most clearly made TLP a focus for its future products. Not only has it incorporated thread-centric approaches into its future designs, it has also taken the idea of increased thread-level parallelism, wrapped it in a mantle it calls "Throughput Computing," and begun to evangelize it not only as a computing technique, but as a movement of which Sun will be the most visible champion.



But there's more to Throughput Computing than just marketing. Even if it's not alone in pursuing TLP, Sun is bringing a particular focus and outsized development energy to what it calls CMT (Chip MultiThreading)[2]—including a future processor codenamed "Niagara" that will be able to handle a full 32 threads on a single processor.

1. An upcoming Illuminata note will detail the technical issues and performance considerations underlying the shift from instruction level parallelism to thread level parallelism, the most problematic of which is the increasing gap between CPU performance and memory latency.

Taking the Wraps Off

Even talking about the future of its processors is a departure for Sun—which is famously close-mouthed about its plans, as much to avoid criticism as to protect the advantage of surprise.[3] Today, however, it's almost impossible to get Sun to shut up about all it's doing to advance the future of microprocessor design. Of course, eventually it will have to live up to all the talk; how well it does so will start to become clearer in system-level performance benchmarks when Sun starts delivering the results of this quest in about a year.

However, at least in the near-term, the action really isn't on the flagship UltraSPARC processors that run in Sun's mid-range and high-end servers. The next-generation "s-Series"[4] CPU, UltraSPARC IV, will take advantage of a process shrink to put dual UltraSPARC III cores on a single chip. This CMP approach is what IBM does with POWER4 today, and what HP will do with its PA-8800 around the end of this year. The UltraSPARC V, due in 2006,[5] gets more technologically interesting. It will have a sort of dual personality; software will be able to make it act like a single large core for ILP-friendly workloads, or like two smaller cores for those that benefit from a more TLP-amenable environment. Still, there are no way-out changes to be found in that design approach. For that we have look to Sun's plans that start on the scale-out end of the computing spectrum.

2. CMT combines CMP (Chip Multi Processing, or implementing multiple processor cores on a single chip) and a variant of SMT (Simultaneous Multi Threading, or running multiple threads on a single processor; also called hyperthreading by Intel) that focuses on rapid thread switching within relatively simple processor cores. Collectively, these techniques tilt designs toward maximizing the total number of threads that they can concurrently handle, rather than how speedily they can process a single thread.
3. Criticism of UltraSPARC III shipment delays only exacerbated those secretive tendencies—after all, if no one knows when a chip is scheduled for delivery, it's harder for anyone to accuse you of being late.
4. S as in "scalability." Sun has adopted this nomenclature for its flagship processor line that goes into its largest servers. I-Series processors will be optimized for 1-4 way servers and h-Series for blades.
5. This date, as with other SPARC processor availability dates in this note, is Sun's estimate of when systems using the processor will ship in volume.

Although Sun plans to come out with a dual-core chip ("Gemini") in 2004 for one- to two-way servers, the real focus of its thread-level parallelism plans is a blade-oriented processor codenamed "Niagara" that is slated for late 2005 or early 2006. It will have eight cores—each of which can simultaneously execute four threads—on a single die. That's a whopping 32 logical processors on a single die. While each one might have just 50 percent of the performance of today's UltraSPARC III, Sun estimates that the large number of cores and the optimization for TLP could yield a 16x faster processor in the aggregate.

Niagara will also incorporate a memory controller and some network interfaces on-chip. Limiting the system to just one of these multi-core processors helps makes this "system on a chip" level of integration possible. All the communications needed to ensure coherency among the different executing threads can therefore take place within the confines of a single chip.[6] This eliminates the electronics and pins that would be needed to conduct this coordination across multiple chips and also speeds it up. Physical proximity, if nothing else, ensures that multiple cores on a single chip can communicate more quickly than multiple chips can communicate with each other. Because threads often need to synchronize with each other—perhaps one thread processes the result produced by another—system level performance depends on quick coordination.[7]

Niagara will be particularly suitable for workloads that are inherently highly multithreaded. First and foremost among these are network servers and networked applications with their many manners of network services, from simple Web (HTTP) and IP (FTP, NFS, NTP, iSCSI, etc.) protocols on through Java-based application servers and Web Services engines. Some classically-programmed business applications could also easily take advantage; SAP's R/3 is a prime example. Applications such as these

6. A single coherent view of all the cache contents throughout the system is necessary to ensure that a thread is always working on the most up-to-date copy of data.
7. Limiting the degree to which threads interact with each other is an omnipresent software development concern. But some dependencies and coordination are always present, so it's important to handle them efficiently.

are currently bound by memory latencies. As an aggressively TLP-optimized design, Niagara tackles memory latency head-on; should any instruction stream need to wait for data—the most common activity in today's microprocessors—Niagara will swap out the current thread for another one that's already ready-to-run. Because it keeps multiple threads' state loaded into the processor core, switching between threads carries nearly zero cost. The next thread starts running almost immediately.

Niagara's multiple-threads-per-core approach is similar in concept to other SMT implementations such as the hyperthreading found on Intel's 32-bit Netburst architecture processors (i.e. the current Pentium 4 and Xeon generation chips). However, SMT on a "super-scalar" processor like the Pentium 4 is more about increasing the utilization of its multiple execution units by using instructions from more than one thread in one cycle. By contrast, Niagara will have relatively simple cores and will use SMT more as a fast "thread switcher" to shuffle threads in and out. If a thread is waiting for memory to deliver some data, it gets swapped out. If it's ready to run, it gets swapped in.

Not Just (or Even Mostly) Big Iron

Sun's current throughput efforts have both external and internal origins. To jumpstart the effort, in mid-2002, Sun purchased Afara Websystems, which had a 32-thread, SPARC-architecture processor similar in concept to Niagara already in the works. The companies were a good fit; in addition to being SPARC-based, Afara founder Les Cohen was the architect for the first generation of UltraSPARC. Sun is augmenting this acquisition with in-house technology. For example, Sun already has a graphics chip—MAJC—with multiple cores.

In recent years, one of Sun's distinctive competencies has certainly been in the Big Iron scale-up system designs. Therefore, the considerable attention Sun is paying to TLP for horizontal scale-style computing is particularly interesting in the light of other recent Sun goings-on. Most notably, on May 19, CEO Scott McNealy took the stage with Oracle's Larry Ellison to trumpet the wonders of low-cost/high-performance compute elements combined with a complete middleware stack from Sun—and capped with Oracle's distributed database clustering technology, Oracle 9i RAC. Ellison made clear his position that the time is coming when scale-out architectures will be able to effectively supplant all of the roles that Big Iron fills today. McNealy certainly holds a less extreme view—one that sees a continuing role for a wide range of system shapes and sizes for some time to come. Nevertheless, Sun's thinking has clearly moved much closer to the camp that believes distributed computing is good for more than just the network edge.[8] For McNealy to share the stage (*sans* grimace) while Ellison trumpeted the coming demise of scale-up systems is a very real shift indeed.

At the heart of this shift is Sun's belief that it can leverage computing trends like TLP to eventually beat out Intel in price/performance—and to do so within a SPARC architecture that's already 64-bits with the large memory support that implies. Sun's never been a timid company and, like IBM in its own way,[9] Sun sees the inevitable transition to broader-based 64-bit computing as one that will create opportunities as some flavor or flavors of 64-bit inevitably displaces today's ubiquitous 32-bit x86 computing.[10] Never mind that Intel rules high volume computing today.

And while it only speaks of them in vague ways for now, Sun has other processor instances in the pipeline that extend the TLP approach. For example, it has spoken tantalizingly about a Niagara follow-on

8. Although network edge applications will continue to be among the best matches for TLP. It's worth noting that Intel, whose Itanium 2 is *all* about ILP, also has a very different network processor design that takes an aggressively multi-core approach.
9. See Illuminata note "IBM Marches to its own 64-Bit Linux Drummer" (February 2003).
10. AMD's alternative is to just extend the x86 instruction set to 64-bit, maintaining backward compatibility at full performance for 32-bit binaries. See Illuminata notes "AMD Extends its Engine to 64-bits" and "x86: Not Dead Yet" (January, 2003). Sun plans to produce a version of Solaris for the AMD64 architecture (including Opteron and Athlon 64 processor families) but says that it has no current plans to sell hardware running 64-bit AMD chips.

able to process each thread at perhaps 90 percent, rather than 50 percent, of the single-thread (ILP) speed of today's UltraSPARC III, as well as other features that suit it for "data facing" rather than "network facing" applications. Though such a processor is unlikely to see the light of day before 2006 or 2007, it's very clear how that might make a formidable database engine—and one very much in-tune with both Sun's "large SMP" heritage and the "increasingly scale-out" designs of heavyweight middleware such as Oracle and DB2. This post-Niagara processor could also be the point at which Throughput Computing and traditionally ILP-friendly High Performance Technical Computing (HPC or HPTC) have their most natural intersection.

While the other features of this as-yet-unnamed follow-on are still sketchy, clearly it will need to support vast memory complements—on the order of 1 TB of RAM or more—and will probably be complemented with much more on-chip partitioning and fault-tolerance than is currently common. Given its design point, it is likely to have a lower level of integration than Niagara, therefore requiring more external support chips and subsystems. However, in the long run, even high-end data-facing systems may eventually drift toward uniprocessor designs. As TLP vastly cranks up the throughput of a single microprocessor, there will be less and less need for hugely expensive, hugely multi-chip systems in a world dominated by network-oriented computing and increasingly distributed databases.

All of these plans, from Niagara forward, depend on the next generation or two of advanced semiconductor fabrication. Even though a single TLP-optimized core needs far fewer—perhaps 10 to 12 percent by Sun's reckoning—transistors than one of today's ILP-focused CPUs, other factors weigh in. The multiplicity of cores, the logic to maintain multiple sets of processor status information, and the very high level of system-on-a-chip integration all are transistor-hungry assumptions. You can start with the low hundreds of millions of transistors per die of today's 0.13 micron (130 nanometer) fabrication, but you really want the 500 million, 1 billion, or the more than a billion transistors enabled by the 90nm fabrication now coming on-line, or the 65nm and 45nm processes just over the horizon. TLP, then, is as much about optimizing for Moore's Law and the inexorable advance of semiconductor fabrication as it is about optimizing for latency and multithreaded workloads.

Running With Threads

Of all the major vendors, Sun is the one most aggressively touting TLP as a cornerstone of its microprocessor roadmap. In many respects, this strategy is a logical next step for Sun. Even though UltraSPARC hasn't led the benchmarks for fastest processor for a number of years, Sun has generally maintained performance parity with competitors at the system and applications level. It accomplishes this with various techniques, such as by putting relatively more CPUs into its systems, and by focusing more intensively on software optimization—both approaches well-suited to its SPARC/Solaris architecture. Thread-level parallelism takes this reliance on high-scale SMP to an even more granular level, while simultaneously eliminating some of the infrastructure and coordination overhead associated with a very large number of discrete microprocessors.

Certainly Sun will need to pair its architectural vision with strong execution if it is to make this strategy fly. One need only look at Intel's 32-bit x86 line to see how relentless advances in processor frequency and cache size—enabled in large part by manufacturing prowess bought and paid for with volume economics—can wring impressive performance from the creakiest of architectures.[11] Or take Itanium 2, which Sun holds up to ridicule as a nefarious example of ILP *über alles*—but which nonetheless significantly outperforms UltraSPARC III on a chip-by-chip basis, and which faces real performance competition only from IBM's POWER4.

11. Of course, Intel and AMD have also updated the architecture in subtle and not-so-subtle ways. See Illuminata note "x86: Not Dead Yet" (January, 2003).

However, there can be no doubt that Sun has successfully identified and aligned itself with a significant multiprocessor design shift in the early days of what will undoubtedly be a long, gradual transition. While the trend toward increasingly mixing instruction parallelism with a judicious dose of thread parallelism may be widely recognized among microprocessor designers, no other vendor is embracing it with the enthusiasm and single-mindedness of Sun. With a dual-core POWER4 today and more multi-core and multi-thread designs on the way, IBM comes closest. But its POWER strategy is anchored by high-end capabilities that trickle down to lower-end products. By contrast, Intel—although it has aggressively promoted on-chip multithreading under the name "hyper-threading"—has been relatively reticent about how it plans to take its designs toward a more fundamentally thread-oriented approach.[12] Intel's 64-bit Itanium Product Family, especially, seems as ill-matched with TLP as it is well-matched with ILP.

For Sun, TLP is about the low end at least as much as the high end. And it's not just an interesting trend, it's a *fundamental* part of its strategy.

12. The only other piece of TLP on Intel's public roadmap is dual-cores for Montecito, the next generation of Itanium processor, due in 2005.

Through subscription research, advisory services, speaking engagements, strategic planning, product selection assistance, and custom research, Illuminata helps enterprises and service providers establish successful infrastructure in five key areas: Server Technologies, Information Logistics, Application Strategies, Enterprise Management, and Pervasive Automation.