

Chip Makers Turn to Multicore Processors

David Geer

Computer performance has been driven largely by decreasing the size of chips while increasing the number of transistors they contain. In accordance with Moore's law, this has caused chip speeds to rise and prices to drop. This ongoing trend has driven much of the computing industry for years.

However, transistors can't shrink forever. Even now, as transistor components grow thinner, chip manufacturers have struggled to cap power usage and heat generation, two critical problems. Even performance-enhancing approaches like running multiple instructions per thread have bottomed out.

For these reasons, processor performance increases have begun slowing. Chip performance increased 60 percent per year in the 1990s but slowed to 40 percent per year from 2000 to 2004, when performance increased by only 20 percent, according to Linley Group president Linley Gwennap.

"We could build a slightly faster chip, but it would cost twice the die area while gaining only a 20 percent speed increase," noted Marc Tremblay, chief architect for Sun Microsystems' Scalable Systems Group.

In response, manufacturers are building chips with multiple cooler-running, more energy-efficient processing cores instead of one in-



creasingly powerful core. The multicore chips don't necessarily run as fast as the highest performing single-core models, but they improve overall performance by handling more work in parallel, as Figure 1 shows.

"Multicore chips are the biggest change in the PC programming model since Intel introduced the 32-bit 386 architecture," stated Gwennap.

"Multicores are a way to extend Moore's law so that the user gets more performance out of a piece of silicon," said John Williams, Advanced Micro Devices' technical director for server microprocessor planning.

Chip makers AMD, IBM, Intel, and Sun are now introducing multicore chips for servers, desktops, and laptops.

DRIVING MULTIPLE CORES

Current transistor technology limits the ability to continue making single processor cores more powerful.

For example, as a transistor gets smaller, the gate, which switches the electricity on and off, gets thinner and

less able to block the flow of electrons. Thus, small transistors tend to use electricity all the time, even when they aren't switching. This wastes power.

Also, increasing clock speeds causes transistors to switch faster and thus generate more heat and consume more power. Gwennap said thermal-design advances have mitigated some problems. However, he noted, this approach can't keep pace with processors' increasing power and heat buildup.

These and other challenges have hurt manufacturers' plans for new, faster single-core processors. For example, Intel canceled two next-generation Pentium 4 processors last year, noted Jeff Austin, the company's desktop product manager. Intel also postponed and then cancelled a 4-GHz, current-generation Pentium. And IBM could build so few of its G5 chips that Apple Computer had to delay last year's introduction of its new iMac G5 desktop, which uses the processor.

INSIDE MULTICORE CHIPS

A dual-core chip running multiple applications is about 1.5 times faster than a chip with just one comparable core, according to University of Texas assistant professor Steven Keckler. He said each core in a typical multicore chip includes everything a microprocessor has except level-2 cache memory and the memory hierarchy, which is located elsewhere on the silicon for all the cores to use.

Divvying up the work

"The compiler handles the scheduling of instructions for a program," said Bill Roth, vice-president of product marketing for software vendor BEA Systems.

The operating system controls the overall assignment of tasks in a multicore processor. Based on this, either the OS or a multithreaded application parcels out work to the multiple cores.

Generally, when a multicore processor has completed a task, one core takes the completed data from the other cores and assembles the final result.

Industry Trends

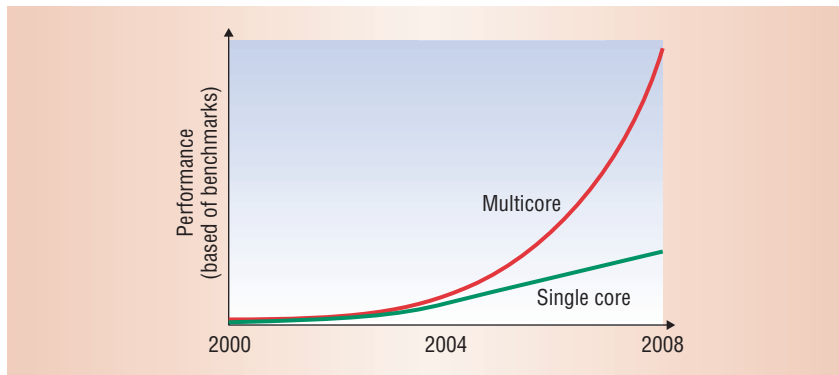


Figure 1. Multicore chips perform better—based on Intel tests using the SPECint2000 and SPECfp2000 benchmarks—than single-core processors. And, Intel says, multicore chips’ relative advantage will increase during the next few years.

Working with applications

To take advantage of multicore chips, vendors must redesign applications so that the processor can run them as multiple threads. “It is more challenging to create software that is multithreaded,” noted AMD’s Williams.

Programmers must find good places to break up the applications, divide the work into roughly equal pieces that can run at the same time, and determine the best times for the threads to communicate with one another.

Vendors also must redesign applications so that they can recognize each core’s speed and memory-access capabilities as well as how fast cores can communicate with one another.

Intel provides a Threading Toolkit to help game and other software developers design multithreaded applications to be used on its new chips.

Memory cache approaches

Each of the two cores in AMD’s Opteron and Intel’s Itanium chips for servers and workstations will have its own cache. IBM, on the other hand, doesn’t use separate caches in its multicore server chips.

Separate caches eliminate the extra work needed to design chips so that multiple cores can work with a single, centralized cache. In some chip designs, though, single caches can function more rapidly than multiple caches.

Advantages

When a single-core chip runs multiple programs, it assigns a time slice to work on one program and then assigns different time slices for others, noted assistant professor Keckler. This can cause conflicts, errors, or slowdowns when the processor must perform multiple tasks simultaneously.

“If you have multiple tasks that all have to run at the same time, you will see a boost with multicore processors,” said Keckler. For example, the chips could use a separate core for each task.

Because the chips’ cores are on the same die, they can share architectural components, such as memory elements and memory management. They thus have fewer components and lower costs than systems running multiple chips. Also, the signaling between cores can be faster and use less electricity than on multichip systems.

MULTIPLE MULTICORE EFFORTS

Several companies are making or planning to make multicore chips.

AMD

AMD will release Opteron enterprise-server multicore first, then the Athlon 64 and Sempron desktop chips, and finally Turion mobile chips. “We will ship them all this year,” said Williams.

AMD says it didn’t have to change its chip architecture to accommodate mul-

ticore capabilities because it anticipated that the technology would become viable and developed its architecture several years ago with that in mind.

Intel

Intel is working on 16 multicore-chip projects.

Two Pentium chips will use the dual-core technology code-named Smithfield. Intel’s high-end PC chip, the Pentium Processor Extreme Edition 840, has begun shipping and runs at 3.2 GHz but outperforms today’s high-end, single-core, 3.8-GHz Pentium 4. The Pentium D, slated for release this year, will be a mainstream desktop chip.

Intel plans to release two dual-core Xeon server and workstation processors early next year: Xeon MP chips for servers running at least four processors and Xeon DP chips for servers and workstations.

A dual-core Itanium server chip scheduled for release later this year will contain 1.7 billion transistors. It will be Intel’s first processor with more than 1 billion transistors. Intel has not released information about the chip’s clock speed.

Intel has developed energy-saving dynamic power-coordination technology that, when workloads permit, lets the OS tell one processing core to sleep or slow down while the other works. Intel plans to integrate DPC, which would extend battery life, into Yonah, the company’s first dual-core laptop chip. Yonah is slated for release later this year.

In 2002, Intel introduced its hyperthreading technology, now supported by Windows XP and most Linux releases. Hyperthreading lets multithreaded software’s threads execute in parallel on a single core, thereby improving performance.

Hyperthreading accomplishes this by enabling more efficient use of all execution units—including arithmetic logic units and floating-point units—in a core. The technology also informs the OS that it supports multiple threads and coordinates their execution.

IBM

IBM released the industry's first dual-core server chip, the Power 4, in 2001. Last year, it introduced the dual-core Power 5, which runs four times faster than its predecessor.

IBM, Sony, and Toshiba have completed design of the Cell processor optimized for compute-intensive workloads, broadband data transmission, and multimedia processing. The companies plan to begin production during the second half of this year, said Ted Maurer, IBM's lead Cell software engineer.

They designed Cell for use in consumer-entertainment devices such as the Sony PlayStation III game console. The companies plan to implement the chip this year in an IBM-Sony workstation primarily for handling computer animation and other demanding graphics tasks, and next year in a Sony-Toshiba high-definition TV and a Sony server.

Cell will use one 64-bit Power core to run the operating system, divide up tasks, and assign them to eight 128-bit processing cores optimized for the floating-point matrix algebra associated with computer entertainment and rich media. The processor will have considerable bus bandwidth between cores and to memory.

The first version will run at about 4.6 GHz and perform 256 Gflops. It will use IBM's silicon-on-insulator technology, in which pure crystal silicon sits on pure silicon-oxide insulation. The purity lets the chips operate faster, more efficiently, and cooler.

Sun

In 1999, Sun released the Micro-processor Architecture for Java Computing dual-core, multimedia, desktop chip. MAJC was never widely adopted as a desktop processor but has been used as an embedded-systems chip. Sun has since built UltraSparc IV dual-core server chips.

Now, Sun's Tremblay said, the company is working on the Niagara multi-core chip for high-end servers. Planned for release early next year, Niagara will

Are Multicore Processors One or Many Chips?

Software vendors charge customers in various ways for using their products. One prominent practice is to charge customers for each processor that will run the software. A customer running an application on 20 machines with single-core processors would thus pay a set amount per chip.

A key issue for multicore-chip makers such as Intel is whether software vendors will consider a processor to be a single core or an entire chip.

Intel defines a processor as a unit that plugs into a single socket on the motherboard, regardless of whether it has one or more cores, and advocates that software vendors charge accordingly, explained Jeff Austin, the company's desktop product manager.

Microsoft and Novell agree and don't charge extra for using their software on multicore processors.

BEA Systems and Oracle, on the other hand, charge more to use their software on multicore chips for per-processor licensing. "Customers get added performance benefit by running our software on a chip with two cores, so we charge a fraction of the single CPU price for additional cores," said Bill Roth, the company's vice president of product marketing.

Multicore-chip makers are concerned that this type of policy will hurt their products' sales.

have eight cores, each handling four threads. It will also feature compilers that will generate parallel threads automatically from an application. The OS could then map these threads to the hardware automatically, Tremblay explained.

In the future, manufacturers will make their multicore chips faster by increasing the speed of each core, as Sun is already doing. During the next few years, said AMD's Williams, the decrease in chips' feature sizes from today's 90 nanometers to 65 nanometers will leave room for more cores.

Multicore processors will find a natural home in servers, said Keckler, but won't be very useful on the desktop until vendors develop considerably more multithreaded software.

Until this occurs, Williams said, single-core chips will continue to compete. Also, he added, single-core chips are inexpensive to manufacture, so they will continue to be popular for lower-priced PCs for a while.

According to the Linley Group's Gwennap, the widespread transition

to multicore chips will occur during the next two years. However, this could give the semiconductor industry time to find new ways to improve single-core chip technology via, for example, exotic materials and advanced manufacturing techniques.

The per-processor fees that enterprise-software vendors charge their customers could be a challenge to multicore chips' success, as the "Are Multicore Processors One or Many Chips?" sidebar explains.

Nonetheless, Williams expressed optimism and said, "Multiple cores are the new megahertz. Multicore will be the transition from brute-force performance to architectural elegance." ■

David Geer is a freelance technology writer based in Ashtabula, Ohio. Contact him at david@geercom.com.

Editor: Lee Garber, *Computer*,
l.garber@computer.org