IEEE SPECTRUM ONLINE
Tomorrow's Technology Today

Winner: Multimedia Monster
By: Samuel K. Moore
http://www.spectrum.ieee.org/jan06/2609

We're flying at about Mach 1.5 around Mount Saint Helens, in Washington state. IBM Corp. senior programmer Barry L. Minor is at the controls, rocketing us over the crater and then down to the lake at its base to skim over the tree trunks that have been floating there since the volcano exploded over 25 years ago. The flight is exhilarating, even though it's just a simulation projected on a widescreen monitor in a cluttered testing lab.

Then, at the flick of a switch, Minor turns the simulation over from his new Cell processor to a dual-processor Apple Power Mac G5, and the scenery freezes. The G5 almost audibly groans under the burden, though it's no slouch. In fact, it's currently the top of the line for PCs. But Cell is something different entirely. It's a bet on what consumers will do with data and how best to suit microprocessors to the task—and it's really, really fast. Cell, which is shorthand for Cell Broadband Engine Architecture, is a US $400 million joint effort of IBM, Sony, and Toshiba. It was originally conceived as the microprocessor to power Sony's third-generation game console, PlayStation 3, to be released this spring, but it is expected to find a home in lots of other broadband-connected consumer items and in servers too.

Executives at Sony Corp., in Tokyo, wanted more than just an incremental improvement over PlayStation 2's processor, the Emotion Engine. What they got was a 36-fold acceleration, to a whopping 192 billion floating-point operations per second (192 gigaflops). Because Cell is a combination of general-purpose and multimedia processors, it defies an exact comparison with other upcoming chips, but it's thought to be more powerful than the chips driving competing game systems.

Cell can calculate at such blazing speed, in part, because it's made up of nine processors on a single chip of silicon, optimized for the kind of real-time calculations needed in today's broadband, media-rich environment. A specially designed 300-gigabit-per-second bus knits the processors into a single machine, and interface technology from Rambus Inc., Los Altos, Calif., gives it fast access to memory and other off-chip systems.

So far, microprocessor watchers have been impressed with what they've seen of Cell. "To bring huge parallel processing onto a single chip in a clean and efficient way is a real accomplishment," says Ruby B. Lee, a professor of electrical engineering at Princeton University and an IEEE Fellow.

A graphics-heavy item such as PlayStation 3 isn't just a showcase for an unusual chip. For IBM it's a philosophical statement. "Gaming is the next interface driving computing," says James A. Kahle, Cell's chief architect with the IBM Technology Group, in Austin, Texas [see photo, "Multicellular"]. Just as moving from punch cards to electronic displays changed what people expected of computers, the highly collaborative, real-time realism of today's games will set the standard for what people want from computers in the future.

But even now, the sheer desire for power in the gaming market guarantees that Cell will be made in volumes that more than make up for the loss last year of IBM's highest profile customer, Apple Computer Inc. Market research firm iSuppli Corp., in El Segundo, Calif., predicts that 37 million game consoles will be sold this year alone worldwide. By 2007, when all three game console makers will have released their next-generation products, the market will have grown to 44 million. And though Cell is exclusive to the PlayStation 3, IBM has a lock on the rest of the console market. Its microprocessors will power both of Sony's competitors, Microsoft's Xbox and Nintendo's GameCube.

The Cell-powered PlayStation 3 can expect to pick up a little less than half of what could become a market worth up to $9.5 billion in 2007, according to iSuppli senior analyst Chris Crotty. And, of course, there are other high-volume plans for Cell.

Toshiba Corp., in Tokyo, for one, plans to build television sets around it. The company has already shown that a single Cell processor can decode and display 48 compressed video streams at once, potentially allowing a television viewer to choose a channel based on dozens of thumbnail videos displayed simultaneously on the screen. And in a smaller market, Cell has already found its first outside customer in medical- and military-systems maker Mercury Computer Systems Inc., in Chelmsford, Mass., which is developing a two-Cell blade server due out by April.

With two such massive consumer electronics makers as Toshiba and Sony behind it, Cell is an obvious attempt to control the "digital living room," as technology executives have dubbed their dream of a home where all the media players are intelligent and networked together. "[Sony's] goal is to make a computer fun...to make it an entertainment platform," says Sony's Cell director Masakazu Suzuoki. "But even if we make the Cell system an entertainment platform, there's nothing if there's no content."

Indeed, experts say Cell's success hinges on whether programmers outside IBM, Sony, and Toshiba will be able to exploit the gigaflops that Cell has to offer. Tony Massimini, chief of technology at the consulting firm Semico Research Corp., in Phoenix, puts it bluntly: "Cell has strong potential, assuming that the game developers satisfy their customers' needs. But if the games suck, who wants to buy it?"

That Cell has more than one processor core on a single chip is more a sign of the times than a revolution. All the microprocessor stalwarts are moving to multicore design. The principal reason is that the old way of doing things—increasing the number of calculations per second by shrinking the processors into a tighter knot of tinier transistors and then dialing up the clock speed-has essentially crashed headlong into the brick wall of heat generation.

Because transistors using today's technology are so small, even when they are supposed to be in the "off" state, infinitesimal currents still leak through them. That leakage warms them constantly, and with the extra heat generated when transistors switch "on" or "off," it produces a microfurnace on a chip. If chip makers had continued on their old path, by the year 2015, microprocessors would be throwing off more watts per square millimeter than the surface of the sun.

As a result, the industry has shifted from maximizing performance to maximizing performance per watt, mainly by putting more than one microprocessor on a single chip and running them all well below their top speed. Because the transistors are switching less frequently, the processors generate less heat. And because there are at least two hot spots on each chip, the heat is spread more evenly over it, so it's less damaging to the circuitry and easier to get rid of with fans and heat sinks.

IMAGE: IBM CORP.
CELL CITY MAP: The Cell microprocessor that will power Sony's PlayStation 3 game console has nine processor cores. The core making up the left quarter of the chip is similar to the processors in Apple computers. The other eight cores, notable by their columns of memory [brown], are designed to do multimedia tasks.

Multicore processors on the market today are generally symmetrical—that is, they have two copies of essentially the same core on one chip. Cell, on the other hand, has an asymmetric architecture that contains two different kinds of cores [see photo, "Cell City Map"]. One, the Power processing element, is similar to the CPU in a Mac; it runs the Linux operating system and divides up work for the other eight processors to do. Those eight—called Synergistic processing elements—are designed specifically to juggle multimedia applications: video compression and decompression, encryption and decryption of copyrighted content, and, especially, rendering and modifying graphics.

The Synergistic elements were built from the ground up to do what are called single-precision floating-point calculations—the kind of operations needed for dazzling three-dimensional graphics and a host of other multimedia tasks. The design traded flexibility—a Synergistic element is not versatile enough to run the Linux operating system on its own—for eye-popping speed. When pushed to its 5.6-gigahertz limits, a single unit can do 44.8 billion single-precision floating-point calculations per second. Not wanting to cut Cell off from a role in scientific computing, its designers included circuitry in each Synergistic element that can do the more exacting calculations, called double-precision, that scientists demand, but its performance is only about one-tenth that of the single-precision unit.

In fact, the Synergistic elements are so fast that a single one could easily consume the entire bandwidth on the interconnects to the off-chip memory, leaving its siblings starved for data and stalled out. IBM and its partners had to design a special chunk of circuitry into Cell just to prevent that problem.

Apart from its raw power, Cell has content-protection tricks that should make it attractive to multimedia applications makers. For instance, the Synergistic element's architecture prevents any application or external device from accessing the element's local memory, so that, for instance, a program cannot steal a music file that is being decrypted by the processor. "Once you bring your code in and decrypt it, it can execute in a virtually trusted environment," says IBM's Cell architect Charles R. Johns. "All the data it calculates on, sends out, and brings in is fully protected."

The isolation function can be used in several ways, says Kahle. "We knew we couldn't anticipate all the different security needs in the future, but we wanted to know we had the right hardware to support a very robust security system."

Barry Minor's Mount Saint Helens simulator is a good example of how Cell's different processors work together. His program takes a satellite photo of the volcano, lines it up with an elevation map, and then turns it into a detailed 3-D terrain on the fly. The Mount Saint Helen's data has a resolution of 2.4 meters. The city of Austin, where the Cell design center is, once gave Minor access to its 15.4-centimeter-resolution satellite map. "You could land in Michael Dell's backyard and check out his view," Minor says with a grin.

What's happening inside the processor is a finely choreographed dance. The Power processing element starts by figuring out where the joystick is pointing the simulator in the stored 2-D maps. Then it divides that scene into 32 portions, four for each Synergistic element. Though perfectly capable of it, the Power processing element does no calculations on the actual data. Instead, it plays to its strength as a controller, figuring out which chunk of work should go to each of the other cores according to how complex the scene is and which cores have more or less time on their hands.

The Synergistic elements then go to work. They pull their portion of the data into their local memories, which they can access at great speed. Then each runs a rendering algorithm on the data and stores it off the chip in the system memory. When the processors are done, they signal the Power element, which instructs one of the synergistic units to run a video compression algorithm. That processor compresses its sister units' finished products and then pushes them out to be displayed on the screen or streamed to a PDA or some other device.

Because the compression takes less time than rendering the graphics, the compressing processor automatically switches gears when it's finished and runs the rendering algorithm on a portion of data until it's needed for compression again. With each frame, the process starts over.

This dance works so well for two reasons. The first has to do with the way Cell handles memory. Rather than waste several clock cycles waiting for the right data to arrive from memory, a

Synergistic element works only on data stored in its own 256 kilobytes of memory, to which it has a high-bandwidth connection. More important, Cell's memory-handling engines can be programmed to keep data streaming through the processor. "We can get over 128 memory transactions going in flight at once," boasts Michael N. Day, a distinguished engineer at IBM.

The memory-access engine takes in new data and sends out the old just in time for the synergistic unit to perform the necessary calculations. When Cell runs Minor's volcano simulator, it waits for data to arrive from memory for only 1 percent of the time; the G5, in contrast, stands idle for about 40 percent of the time.

Cell's other key to speed has to do with breaking problems into parts that can be done in parallel. In Minor's simulation, it probably seems obvious that an image can be divided up into eight strips and these worked on independently. What wasn't so obvious was that the 3-D rendering could be done four pieces of data at a time within each synergistic processor. Such four-way parallel computing is called single instruction multiple data, or SIMD, and it is particularly well suited to the manipulation of graphics and other multimedia.

In these problems, you typically want to perform the same operation on each of the elements in a large chunk of data. For example, to increase the brightness of an image, you'd want to add the same number to every pixel in it. Since around the mid-1990s, general-purpose processors such as the Intel x86 architectures have been doing SIMD computing using a set of multimedia-specific instructions, explains Princeton's Lee, a multimedia instructions pioneer.

But SIMD instructions run far faster on Cell's Synergistic processors, because the Cell processors were designed from the start to handle them. And don't forget: there are eight such processors on each chip. Cell programmers spend most of their time turning complex algorithms into efficient SIMD algorithms, says Minor. "Once you've done that, you're 80 percent done."
The chip's commercial success will depend on whether programmers can learn to exploit its full potential. To that end, the developers have from the beginning put a high priority on crafting the appropriate software tools.

One of the key deadlines the Cell development team had to meet was having its software ready and tested in time for the arrival of the first chips, in spring 2004. The software team was running programs on a Cell simulator two full years before it got the first chip—and when the chip finally arrived, both the operating system and the applications worked on the first try. "Had we waited to do software development until the chip came back, it would have been a disaster," says Theodore R. Maeurer, software manager at IBM.

With such a head start on the software, the group could focus on how to familiarize new programmers with Cell. "A programmer has to do a really nice job of laying out the data transfers and so forth," says Day. But soon that job will be turned over to the compiler and the programming tools. IBM software engineers are also developing tools that will make it easier for programmers to divide tasks between the Power element and the Synergistic cores, and they're making others to automatically find solutions to problems that fit well with the Synergistic units' SIMD strengths. The company has already released more than 700 pages of documents to applications developers and will begin releasing tools and compilers, as well.

Cell's asymmetric architecture signals the beginning of a big shift in how computers are programmed, says Craig Steffen, a senior research scientist at the National Center for Supercomputing Applications, Urbana-Champaign, Ill., who gained some fame lashing together 70 PlayStation 2 consoles to form a $50 000 supercomputer.

"How do you program with eight engines running full speed without them constantly stopping and waiting for data?" Steffen asks. Cell will force mainstream programmers to wrestle with that question. But ultimately, parallel programming will become fairly routine, he predicts. "Over the next several years, we won't think of an asymmetric processor as anything different."

Indeed, some think Cell is an indication of what's to come in other microprocessors. "In the future, we'll see convergence of general-purpose multiprocessors and game- and media-oriented processors," says Princeton's Lee. "Media processors will become more general purpose, and general purpose, more multimedia." And with any luck, that will make your living room a more entertaining place.

Cell Microprocessor

Goal: Make a new microprocessor architecture that beats all others at handling graphics and broadband multimedia.

Why it's a winner: It met that goal and is being designed into high-volume mass-market items like game consoles and televisions.

Organizations: IBM, Sony, and Toshiba.

Center of activity: Austin, Texas.

Number of people on the project: 400 at its peak.

Budget: US $400 million.