

Anatomy of a Cell

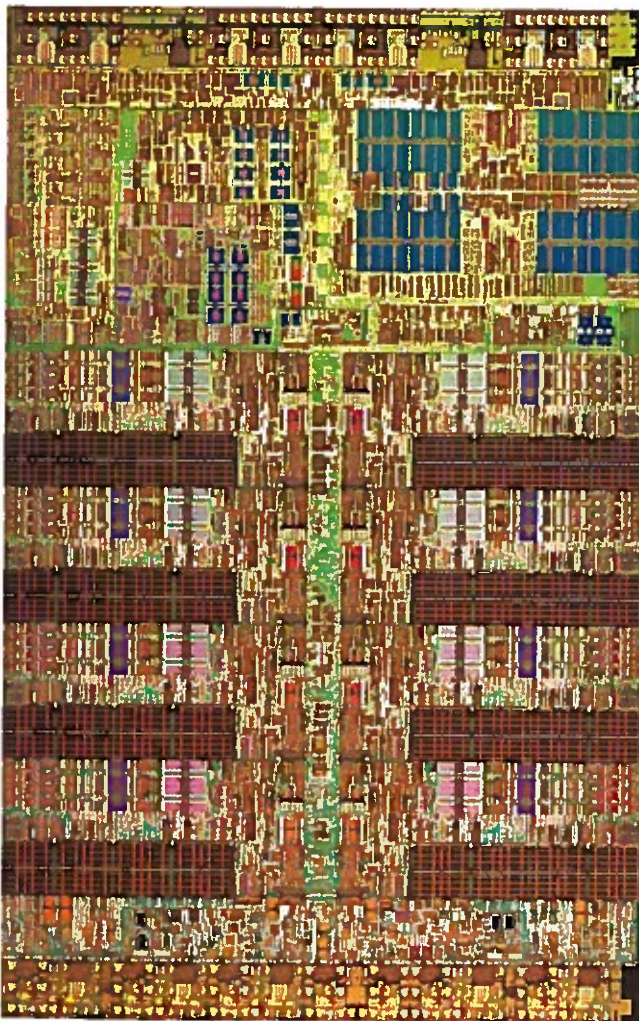
The multi-core revolution has arrived. Will the Cell processor, one of the first multi-core-like chips to hit the market, make waves in bioinformatics?

By Jennifer Crebs

The gaming market continues to drive hardware innovations that are snapped up and remodeled for bioinformatics applications. Joining the fate of graphics cards that have been re-implemented to speed up search heuristics, the Cell microprocessor might just be the next piece of technology to change the way bioinformatic programming is done.

Known in full as the Cell Broadband Engine, the microprocessor powers Sony's PlayStation 3 gaming console, which is scheduled to debut in the US later this month. But Cell was not designed to address gaming tasks alone. IBM says that its architecture is general purpose enough to find a home in a range of data-processing-intensive applications, from cryptography to genome alignment. This is due to its components — one main general purpose core augmented by eight GPU-like coprocessors.

The Cell processor consists of a chip containing a 64-bit IBM Power Architecture core along with eight coprocessors based on a 128-bit single-instruction multiple data architecture. The main core, also known as a power processor element or PPE, coordinates the computational work of the eight coprocessors. Each of these coprocessors, termed syner-



STI's Cell features nine on-chip processors. In this image, the PPE sits at the upper left and the eight SPEs appear below it.

gistic processing elements in Cell-speak, features 256KB of local memory; together, they do the computational heavy lifting on instructions orchestrated

by and received from the PPE. These components share memory and are linked together by a high-speed bus.

Cell's four-year gestation took place at the STI Design Center, a Texas-based outfit formed by Sony, Toshiba, and IBM. According to Cell's chief architect, Peter Hofstee of IBM, the microprocessor was designed from the outset to work as a component in interconnected environments. Hofstee says that the name came from Ken Kutaragi, chief of Sony Entertainment, who wanted it to evoke "a human-oriented kind of processor."

In explaining the major innovation of Cell's architecture, Hofstee relies on another human-oriented metaphor he picked up from Stanford's Bill Dally. Essentially, he says, today's microprocessors work a bit like a daft worker doing a plumbing project. For every part that's needed for the project, this plumber scuttles across town on separate trips to the hardware store. That is, typical multi-gigahertz microprocessors may have a memory latency of several hundred cycles. "It's all very, very inefficient," Hofstee says, "but this is the way in which programs are written."

Why are they written this way in the first place? Hofstee notes that back when microprocessors were first devel-

oped, memory was just a few cycles away. Sequential programming made sense, as the metaphorical hardware store was next door. The fundamental decision in the development of Cell was, in a way, to address the urban sprawl of hardware configurations. To do this, developers optimized the way instructions are executed on the SPEs. Each SPE contains three major units, Hofstee says: a DMA processor, a local store of memory, and an execution core that integrates floating point and media register files.

Cell's multiple cores, 128-bit registers, and local memory improvements all collude to improve processing speed by an order of magnitude on many applications, Hofstee says. However, optimal speed gains require writing code to explicitly take advantage of the parallel processing afforded by the SPEs. Hofstee says that "if you take a piece of sequential code ... and you just recompile it on this processor, obviously you're not going to get the best possible result." Instead programmers must write code directly to the hardware, or use a middleware program to do so. Rewriting code is "basically the price you have to pay," Hofstee says, "but I think we're going to have to do that anyway, because the entire industry is going multi-core."

One such middleware provider is RapidMind, a Canada-based company founded by Michael McCool, developer of the Sh metaprogramming language. The company has commercialized a software development platform that allows developers to use standard C++ programming to create applications targeted for high-performance processors, including Cell, GPUs, and other multi-core CPUs. In Cell, the RapidMind platform does so by way of a data parallel-type programming environment compiled to efficiently make use of the processor's SPEs, Hofstee says.

IBM also offers a Linux-based development platform to help programmers make the Cell switch. In order to try out a particular program on Cell using

Linux, interested users can download architecture documentation and a full system simulator from IBM's developer-Works website.

Luckily, you won't need to rip open a PlayStation 3 to get at the processor itself. By the end of September, Cell became globally available in a full computing system commercialized as IBM's BladeCenter QS20. The standard configuration of IBM's blade system features two Cell processors running at 3.2 GHz and working in a Linux Fedora operating system. Mercury Computer System also has Cell-based blade systems available, as well as a software development kit aimed at high-performance computing markets.

Despite its youth, the Cell processor is already being test-driven on bioinformatics applications. According to Hofstee, "on computationally and memory intensive operations, we tend to do quite well." For instance, researchers from IBM have evaluated the performance of Clustalw and HMMer on Cell, detailing their results in a poster set for the SC2006 conference this month.

Stanford University's Folding@Home distributed computing project is also looking forward to reaping the benefits of Cell via the PlayStation 3. Software from Sony enables users to donate processing power while watching protein folding simulations happen in real time. According to an estimate on the project's website, it would take 10,000 machines plugged in "to achieve performance on the petaflop scale."

The current incarnation of Cell is optimized for single precision floating point computations, which are not ideal for intensive scientific computing, but Hofstee says that his team is currently working on a version of the processor that "will deliver 100 gigaflops of double precision, as well as 200 gigaflops of single precision." This is the version of the processor that is slated for use in the third generation of blades, Hofstee says, as well as the upcoming Roadrunner 1.6 petaflop supercomputer at Los Alamos National Laboratory. **GT**

Clustal W

EMBOSS

Glimmer

HMMER

Bioinformatics Power Cluster

Moab

mpiBLAST

MrB

NCBI BLAST

56 - 2.5GHz IBM 970 (G5) cores.

ProCoS

7U rackmount chassis.

T-Coffee

32-bit AltiVec / 64-bit native.

Wise

4x performance density over Xserves.



Now available, the IBM BladeCenter JS21 with Y-Blo gene sequence analysis tools pre-installed for a compact, high performance bioinformatics cluster.

Options include gigabit, Myrinet, or Infiniband Interconnect; IBM XL compilers, Moab, and a mobile rack.



U2BIO



www.terrasoftsolutions.com

sales@terrasoftsolutions.com

970.278.9243