



# **Using Cell Broadband Engine Technology to Improve Molecular Modeling Applications**

---

## Contents

---

- 2 *Executive overview*
- 3 *Introduction*
- 4 *The need for speed (and, by extension, power)*
- 6 *Cell/B.E. architecture: breakthrough performance*
- 8 *Success stories*
- 12 *Coding differences*
- 13 *Cost savings*
- 14 *Green computing*
- 14 *Cell/B.E. technology roadmap*
- 15 *Conclusion*
- 15 *For more information*

### Executive overview

For many years, organizations have relied on performance gains from increasing clock speeds of “traditional” microprocessor architectures. However, in recent years this approach has been challenged by the physical limitations of semiconductors and by traditional processor architecture implementations. Issues with power consumption, heat dissipation and memory latencies have led to diminishing returns on performance. High performance computing (HPC) applications such as those used in computational chemistry may need a fundamentally new technology and approach to the system-level architecture to achieve the desired level of performance.

IBM®, in collaboration with Sony Computer Entertainment and Toshiba Corporation, established a relationship with the primary objective of developing a new processor with dramatically increased performance, responsiveness and security. The group understood that conventional microprocessors have performance limitations and traditional improvements were not going to meet current and future demands for greater processor performance. Another equally important goal was to overcome the increasing power consumption, cooling resources and floor space needed by systems based on conventional microprocessors that were trying to meet the ever-increasing processing demands of organizations.

The results of the relationship is Cell Broadband Engine™ (Cell/B.E.™) technology, an advanced IBM Power Architecture® technology-based microprocessor. It helps companies to overcome the limitations of today’s microprocessors and targets current and future compute-, data- and memory-intensive application workloads. Currently, Cell/B.E. technology is available from IBM on select IBM BladeCenter® products and from IBM Business Partners in industry-standard form factors.

This paper examines the real-world advantages of Cell Broadband Engine technology for improving performance within molecular modeling applications

---

---

Highlights

---

---

and specifically the results of our collaboration with SimBioSys to produce the eHiTS® Lightning application running on the Cell/B.E processor. We discuss the benefits to pharmaceutical companies and biomedical researchers, and demonstrate scalable performance and superior performance-per-watt for high-performance computing (HPC) applications and breakthrough improvements in performance of up to 60X for the eHiTS application relative to a standard Intel® processor. The paper also explores how systems that are based on Cell/B.E. technology can overcome the increasing power consumption and cooling resources required by conventional microprocessors.

**Introduction**

Bigger is better. More hard drive space, more RAM, more GHz. We have become obsessed with the components and functions crammed into our electronic devices whether they are cameras, phones, or music players. And, of course, we demand increasing compute power. But while the general public is not necessarily concerned with petabyte performance from their electronics, in the world of high performance computing algorithms, performance and scalability are crucial.

***IBM delivers innovative, powerful, open HPC solutions to address demands for intense computation, visualization, or manipulation and management of massive amounts of data***

IBM knows this and delivers innovative, powerful, open High Performance Computing (HPC) solutions to address demands for intense computation, visualization, or manipulation and management of massive amounts of data. In particular, computational chemistry requires massive processing power to run algorithms crucial for helping develop new drugs and materials, and assisting the chemist in understanding data or challenging him to find entirely new chemical objects.

Flexible ligand docking helps chemists identify molecules, or ligands, that interact with—or dock—to a receptor associated with a disease-state. SimBioSys' eHiTS® is an industry leading software application for flexible ligand docking and utilizing this tool provides an in silico approach to the discovery and design of new drugs offers to potential cures. The processing power necessary

---

---

Highlights

---

---

for such computations presents challenges. Thousands of ligands at a time are docked and the processing time for each ligand computation can take up to two minutes.<sup>1</sup> One way to speed algorithms is to use large clusters of computers—but this increases costs for hardware, cooling, and system management. This is why computational chemistry may need a fundamentally new technology and approach to the system-level architecture to achieve the desired level of performance.

Enter Cell Broadband Engine (Cell/B.E.) technology. The Cell/B.E. is a microprocessor architecture jointly developed by Sony, Toshiba, and IBM. Originally used in cutting edge gaming applications to speed up physics simulations and catch up with the 3D graphics rendering speeds provided by advanced graphics processing units (GPUs), Cell/B.E. architecture has now been integrated into enterprise-class technology systems.

***The ‘supercomputer-like performance’ of the Cell/B.E. processor is especially suitable for high-performance workloads***

The Cell/B.E. processor’s breakthrough multi-core architecture and ultra high-speed communications capabilities deliver vastly improved, real-time response. Effectively delivering “supercomputer-like performance” by incorporating advanced multiprocessing technologies used in sophisticated IBM servers, Cell/B.E. is especially suitable for high-performance workloads. Matching the superior performance of the eHiTS algorithms with the acceleration derived from using the Cell/B.E processor allows the eHiTS Lightning application to provide unprecedented accuracy and throughput for flexible ligand docking to the life sciences Industry.

**The need for speed (and, by extension, power)**

In 1965, Intel co-founder, Gordon E. Moore observed that the number of transistors that can be inexpensively placed on an integrated circuit increases exponentially, doubling approximately every two years. Dubbed Moore’s Law, this trend has continued for more than half a century and the exponential growth of transistor density has directly translated into exponential performance increase of the CPUs, until a few years ago. As

---

<sup>1</sup> See “Computational chemistry” section for more information.

---

---

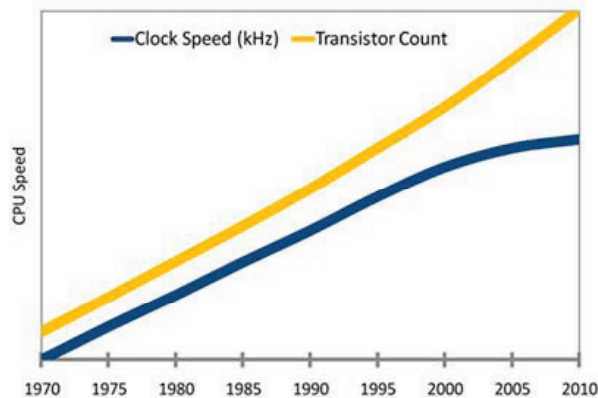
Highlights

---

---

shown in the figure below, since the year 2000, CPU clock speed increases have slowed down. This performance lag is mainly due to three limiting obstacles:

- The frequency wall: the standard speed up technique of processor instruction pipelining is not returning any more improvements in speed because of code branching.
- The memory wall: the processor frequency has now surpassed the speed of the DRAM and the current workaround of using multilevel caching leads to increased memory latency.
- The power wall: both heat generation and power consumption can become unmanageable



***Slow main memory access on traditional x86 architectures results in much lower sustained performance than the theoretical peak of the CPU***

The slow main memory access on traditional x86 architectures creates a data flow bottleneck causing processor idle times. This results in much lower sustained performance than the theoretical peak of the CPU. To combat the bottleneck, state of the art processors have significant cache (L1, L2, L3), typically several megabytes on the processor chip. This uses up space that would otherwise be available to allow more transistors (and more processing power, as well as more heat). This “wasted” cache memory area is one explanation for why Moore’s law no longer translates into equivalent performance increases.

---

---

**Highlights**

---

---

Alternatively, special purpose computational hardware such as graphics processors (GPUs) on 3D video cards and various hardware accelerators continue to increase performance exponentially. These processors use streaming system architecture, many simple coprocessor cores, wide data parallelism and localized dedicated data storage to avoid the previously highlighted performance walls.

Conventional microprocessors also consume more power as more transistors are placed in the same space in an effort to gain more compute power. However, placing more transistors in a given area causes an increase in the power and cooling that is needed to operate the microprocessor. Special purpose computational hardware also consumes more power as they require not only their own power and cooling, but typically also need an x86 life support box to control them. Organizations are becoming more sensitive to the increasing cost of energy used by their computer centers. Using more power and cooling is causing energy bills to take a larger piece of the IT budget.

**Cell/B.E. architecture: breakthrough performance**

The Cell/B.E. processor is an asymmetric multi-core processor that is optimized for parallel processing and streaming applications. Unlike symmetric multi-core, cache-based architectures which may not be able to efficiently handle streaming applications, Cell/B.E. technology is designed to offer very high performance and fast response. The Cell/B.E. processor includes a Power Processor Element (PPE) and eight highly optimized enhanced double precision (eDP) SIMD engines called Synergistic Processor Elements (SPE).

***Cell/B.E. processor performance is about an order of magnitude better than traditional processors for applications that can take advantage of its SIMD capability***

Cell/B.E. processor performance is about an order of magnitude better than traditional processors for applications that can take advantage of its SIMD capability. The PPE is intended to run the operating system and coordinate computation. Each SPE is able to perform mostly the same as, or better than, a General Purpose Processor with SIMD running at the same frequency. A key performance advantage comes from its eight decoupled eDP SPE SIMD engines with dedicated resources including large register files and DMA channels.

---

Highlights

---

***Another important design difference is the memory architecture. Cell/B.E processor memory architecture uses dedicated, on-die local memory storage for each SPE***

Another important design difference is the memory architecture. Cell/B.E processor memory architecture uses dedicated, on-die local memory storage for each SPE with a separate Memory Flow Controller (MFC) unit for managing data transfers, plus the use of XDR RAMBUS memory that can operate at CPU clock speeds (>3GHz compared to <1GHz of DDR RAM used in PCs), eliminating the need for large L2 cache memory and data latency.

Performance specification highlights of the Cell/B.E. processor:

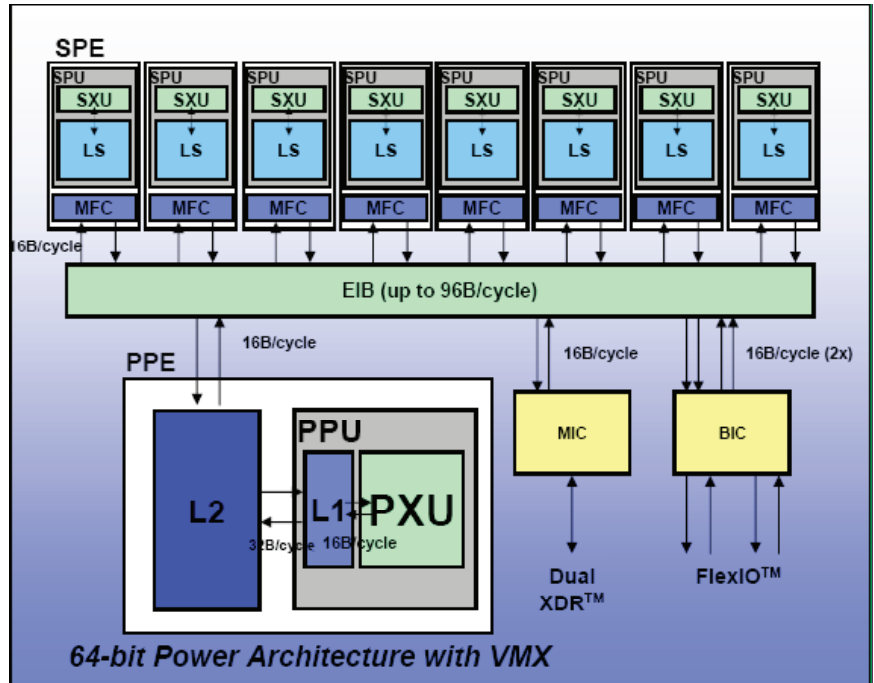
- 241 million transistors
- 9 cores, 10 parallel execution threads
- >230 GFlops single precision
- Up to 25 GB/s memory bandwidth
- Up to 75 GB/s I/O bandwidth
- >230 GB/s bandwidth on the Element Interconnect Bus (EIB)
- Frequency > 4GHz

The following figure shows the architecture diagram of the Cell/B.E. processor. The PPE accesses the main memory via the L1 and L2 cache in the traditional way, thus accessing the whole main memory. There is a latency penalty caused by two layer caching. However, the SPEs can only operate directly on their own 256 KB local memory. Each SPE has a programmable MFC that allows fast data transfer between the main memory and the local store. The SPEs can also communicate with each other and with the PPE via the very high speed EIB. Most of the compute power of the Cell/B.E. processor comes from the SPEs. Each SPE is capable of executing up to 8 floating point or 32-bit integer operations per cycle, a total of 64 operations in a single cycle. Alternatively, they can execute 128 16-bit operations or 256 8-bit operations in parallel. A good application design for Cell/B.E. technology uses the PPE for main control and disk I/O tasks, while all computation intensive tasks are distributed to the SPEs.

---

Highlights

---



**Success stories**

*High performance technical computing*

High performance computing, also referred to as high performance technical computing or high productivity computing, is most commonly associated with computing used for scientific research, or to the engineering applications of computational fluid dynamics and simulation. Recently, the term has also been applied to workloads in Digital Content Creation & Distribution, Seismic Analysis, Financial Analysis, Medical Imaging and Electronic Data Analysis.

***The Cell Broadband Engine offers up to 10X performance improvement with high levels of energy efficiency***

For many of these workloads, the Cell Broadband Engine offers a 2X to 10X performance improvement over traditional architectures while providing high levels of energy efficiency that make the Cell/B.E. a logical choice for any “green” computing initiative.

---

---

Highlights

---

---

*Financial analytics*

Innovative trading concepts, data explosion, and the need to speed up complex financial calculations are driving the requirement for increased computing power as never before. Financial firms are competing at new levels and having to do more with less.

Platform Computing developed Symphony™ for BladeCenter to take full advantage of the Cell/B.E. multi-core architecture and support the deployment of compute-intensive risk and analytical applications. Platform Symphony optimizes application performance by enabling near 100 percent utilization of all 18 cores on a blade containing two Cell/B.E. processors. The joint solution is extremely energy efficient, effectively reducing hardware requirements while increasing resource utilization and throughput.

For two specific financial workloads, European Options using Black-Scholes<sup>2</sup> and American Options using Binomial Trees, the IBM PowerXCell™ 8i processor performed 3.57X and 5.57X faster respectively than a single socket 2.66 GHz quad-core Intel X5355 processor and a single socket Intel E5345 quad-core processor.

*Medical Imaging*

Medical and molecular imaging systems are revolutionizing the way that medicine is practiced today. Traditional imaging modalities (such as magnetic resonance imaging, or MRI; computerized tomography, or CT; positron-emission tomography, or PET; and ultrasound) are being combined with new classes of imagery and information from digital pathology and molecular screening devices, and each can provide different kinds of information for faster and superior diagnosis. These systems generate a huge amount of information that needs to be analyzed, visualized, and interpreted for diagnostic and treatment planning purposes.

The IBM PowerXCell 8i multi-core processor architecture helps the QS22 accelerate key algorithms such as 2D/3D rendering, image analytics,

***The IBM PowerXCell 8i multi-core processor architecture helps accelerate key algorithms and enables highly visual, immersive, real-time applications***

---

<sup>2</sup> Refer to <http://www.risklatte.com/features/quantsKnow050905.php> and [http://www.global-derivatives.com/index.php?option=com\\_content&task=view&id=52&Itemid=31](http://www.global-derivatives.com/index.php?option=com_content&task=view&id=52&Itemid=31)

---

Highlights

---

***MRI image reconstruction algorithms on the IBM Cell/B.E. processors demonstrated up to 16x improvements***

compression and encryption, and enables companies to create and run highly visual, immersive, real-time applications—this performance offers significant potential benefit to companies in health care, life sciences and other industries, and reduces floor space, power consumption and cooling costs.

Recently, a team at GE Global Research implemented three MRI image reconstruction algorithms on the IBM Cell/B.E. processors and demonstrated up to 16x improvements.<sup>3</sup> IBM worked with Mayo Clinic to implement a mutual-information-based three-dimensional (3D) linear registration algorithm on the Cell/B.E. processor. Using 97 sets of image data, the registration code running on Cell/B.E. was about 11X faster than the same code running on an Intel Woodcrest PC.<sup>4</sup>

*Bioinformatics*

Sean Eddy's Profile Hidden Markov Models (profile HMMs) is an important bioinformatics application for protein sequence analysis and is used for sensitive database searching. IBM Research recently implemented a version of Sean's code<sup>5</sup> on an IBM PowerXCell 8i processor. It outperformed a 3 GHz dual-core Intel X5160 system (single socket) by a factor of 12X.

*Molecular modeling*

The eHiTS flexible ligand docking software developed by SimBioSys is an industry leading application for the purpose of flexible ligand screening. It was ported to run natively on the Cell/B.E. architecture and fully utilizes the vector processing power of the SPEs. The port is currently in beta-testing, and commercial release is planned for Q3 2008.

The following table shows execution times and performance improvement reached for one example program, the Lennard-Jones 6-12 Potential calculation, part of the eHiTS program.

---

<sup>3</sup> Source: GE Research. MR IMAGE RECONSTRUCTION ON THE CELL PROCESSOR, SMRM conference in Berlin May 2007

<sup>4</sup> Source: Mayo Clinic. <http://www.mayoclinic.org/news2007-rst/3996.html>

<sup>5</sup> <http://hmmer.janelia.org/> and <http://powerdev.osuosl.org/project/hmmerAltivecGen2mod>

---

**Highlights**

---

| <b>System</b>                      | <b>PPU only<br/>(no SPUs)<br/>~ Power Mac</b> | <b>Dual-core Intel<br/>Xeon® Processor,<br/>2.4 GHz</b> | <b>IBM BladeCenter<br/>QS21</b> |
|------------------------------------|---|---|---------------------------------|
| <b>Time</b>                        | 3m 34s  | 1m 14s  | 1.0s                            |
| <b>Performance<br/>improvement</b> | 1x  | 3x  | 214x                            |

The next table illustrates the performance gains achieved in various calculations on the IBM BladeCenter QS21 with two 3.2GHz Cell/B.E. processors compared to traditional CPUs at equivalent clock speeds.

|   |            |
|---|------------|
| eHiTS Scoring (with rotamer optimization) | 122x       |
| Rigid Fragment Docking                    | 58x        |
| Pose Matching                             | 37x        |
| Flexible Chain Fitting                    | 92x        |
| Conformation Minimization                 | 84x        |
| Final Optimization                        | 55x        |
| <b>Total (Complete Flexible Docking)</b>  | <b>62x</b> |

***Due to the nature of the Cell/B.E. architecture, redesigned molecular modeling algorithms can perform more brute force calculations***

The complete application performance improvement depends on the input structures (such as the size of the cavity compared to the size of the rigid fragments, and the number of rigid fragments) and different cases require a different number of function calls of various types and, therefore, range between 26-fold and 60-fold. It is important to note that due to the nature of the Cell/B.E. architecture, the redesigned algorithms perform more brute force calculations compared to smart cutoffs, and this paradigm shift leads to higher accuracy which is yet another significant benefit to the port to the Cell/B.E processor.

### **Coding differences**

The performance exhibited by the Cell/B.E. is not delivered simply by recompiling existing C/C++ code for the Cell/B.E. processor. There are several key differences in coding for the Cell/B.E. processor compared to traditional CPUs:

- The PPE and the SPE are not binary compatible. The code needs to be compiled with a different compiler to generate code fragments that run on the SPE. Simple POSIX threads cannot be scheduled by the OS to run on the SPE.
- A thread running on an SPE can only directly access the 256 KB local storage of that SPE. This small amount of memory must contain the code, the data and the execution stack at any one time. Of course, data or code can be shuffled back and forth between the main memory and the local store via DMA calls, but those have to be explicitly programmed and managed by the code. Furthermore, the DMA calls in the code have to be designed to use double buffering or similar tricks to streamline data and avoid stalling due to latency.
- The power of the SPEs comes from SIMD vector operations, where the same instruction is executed for multiple data entries. An SPE using dual pipe can execute 8 instructions per cycle. These are two different instructions each executed on 4 parallel data units and those data units must reside in a consecutive block of memory or in a single 128-bit register. High performance can therefore only be reached if the data is organized in a very specific way that is suitable for SIMD calculations.
- SPEs do not have branch prediction hardware. This was one of the simplifications compared to traditional CPUs that allowed the engineers to pack more computation cores into the chip. The downside of this is that branches in the code are more costly and lead to a loss in pipeline efficiency. A single branch miss costs an 18-cycle delay and 144 (from 18x8) potential instruction executions are lost! Therefore, code must be written by minimizing branches. If a choice needs to be made, it is often worth computing both alternatives and selecting the most appropriate one from the results rather than inserting a branch to compute only what is necessary.

---

Highlights

---

***IBM SDK for Multicore Acceleration enables programmers to develop or modify applications to effectively use Cell Broadband Engine technology***

To support the programmer who is modifying code to take full advantage and reap the rewards of the Cell Broadband Engine architecture, IBM has developed a Software Development Kit for Multi-core Acceleration. The SDK, based on Eclipse and developed to run on Linux®, enables programmers to develop or modify applications to effectively use Cell Broadband Engine technology. This SDK is also for a robust software development community that engages in the development of applications that are optimized to Cell/B.E. technology. Tools and solutions that are created by IBM Business Partners ease implementation of applications. Currently, the SDK includes:

- IBM XL C/C++ compiler
- Enhanced application libraries, including *BLAS (Basic Linear Algebra Subroutines)* and *LAPACK (Linear Algebra Package)*
- Example source code with samples, libraries, workloads and prototype code
- Support for Hybrid Model programming wizard

**Cost savings**

IBM BladeCenter QS21 and QS22, based on Cell/B.E. technology, are standard blade form factors. Coupled with the robust BladeCenter H chassis, which offers advanced high-speed communication fabric, the processing power of the Cell/B.E. technology can be fully leveraged by compute-intensive applications. And because the overall BladeCenter infrastructure uses super energy-efficient components and shared infrastructure architecture, you can realize even lower power consumption when compared to many alternative designs.

***BladeCenter QS21 and QS22 boost performance on fixed-point arithmetic through the use of the eight powerful SPEs and their SIMD execution units***

Compared to the rigid pipelines and fixed functions of many graphics processors, the BladeCenter QS21 QS22 are more robust. Unlike specialized add-in floating-point accelerators the QS21 and QS22 are able to also boost performance on fixed-point arithmetic through the use of the eight powerful SPEs and their SIMD execution units.

There are several potential cost savings for users of Cell/B.E. technology compared with traditional HPC solutions such as clusters, CPU farms or supercomputers. These savings include:

- Initial hardware cost (per GFlops)
- Electricity cost: Cell/B.E. technology requires far less power
- Cooling cost: Heat emission is much lower
- Space saving: Smaller CPU and fewer boards mean less equipment—both to house and manage

The same advantages also apply to other hardware accelerator technologies, such as GPUs or FPGAs.

#### **Green computing**

Many corporate and academic computing users seek to make their installations as energy efficient as possible, to lower their carbon footprint and their overall costs. Thus, they evaluate alternative computing architectures based on the amount of computation that can be delivered per watt of power consumed.

The Green500.org publishes a ranking of the most energy-efficient supercomputers in the world and serves as a complementary view to the TOP500 supercomputer list. In the third annual Green 500 list published in 2008, IBM holds all top 10 places with the top three built using IBM PowerX 8i processors.

#### **Cell/B.E. technology roadmap**

IBM is committed to continuing their innovation with Cell/B.E. technology. In 2003, we opened the SDI design center in conjunction with Sony and Toshiba and, in 2005; we released the first Cell/B.E. processor inside the Sony PS3. In August of 2006, IBM released the IBM BladeCenter QS20 built with Cell/B.E. technology—the first enterprise-class Cell/B.E. solution designed to help yield application results faster and with more fidelity.

---

Highlights

---

*The QS22 offers extraordinary floating point computing power to accelerate key algorithms such as 3D rendering, compression and encryption*

In 2007, continuing our momentum, IBM released the BladeCenter QS21, based on second-generation Cell/B.E. technology. With two high-performance Cell/B.E. processors, and a potential 10x or greater performance acceleration of targeted workloads, the QS21 enables organizations to get information faster to facilitate important business decisions.

In 2008, IBM announced the BladeCenter QS22, based on the new IBM PowerXCell 8i multi-core processor. The QS22 offers extraordinary single precision and double precision floating point computing power to accelerate key algorithms such as 3D rendering, compression and encryption. Combined with memory support up to 32 GB, the QS22 can offer significant potential benefit to companies in a wide range of industries.

**Conclusion**

In this white paper we have provided an overview of the Cell/B.E. processor and the obvious potential that this advanced technology offers by improving speed, performance, and energy efficiency of many HPC applications, including docking experiments. SimBioSys is the first company to offer a commercial flexible ligand docking solution compatible with the Cell/B.E. processor and the matching of their best-in-class software with the IBM platform offers exciting opportunities for drug discovery. The need for computational power is only going to increase and researchers seeking to unravel the mystery of disease, and discover new drugs and materials must look to solutions that help them meet their need for increasing computational power. SimBioSys eHiTS running on BladeCenter QS21 can deliver the power organizations need to accelerate key algorithms and speed solutions to tomorrow's problems today.

**For more information**

Cell/B.E. technology: [ibm.com/technology/cell](http://ibm.com/technology/cell)

IBM BladeCenter: [ibm.com/systems/bladecenter](http://ibm.com/systems/bladecenter)



© Copyright IBM Corporation 2008

IBM Systems and Technology Group  
Route 100  
Somers, New York 10589  
U.S.A.

Produced in the United States of America  
10-08  
All Rights Reserved.

IBM, the IBM logo, BladeCenter and IBM PowerXCell are trademarks or registered trademarks of IBM Corporation in the United States, other countries, or both. For a list of additional IBM trademarks visit **ibm.com/legal/copytrade.shtml**

Cell Broadband Engine and Cell/B.E. are trademarks of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.

Intel, Xeon and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

eHiTS, SimBioSys and the SimBioSys logo are registered trademarks of SimBioSys, Inc. in Canada and other countries.

Other company, product and service names may be trademarks or service marks of others.

This publication could include technical inaccuracies or typographical errors. This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. References herein to IBM products and services do not imply that IBM intends to make them available in other countries. Consult your local IBM business contact for information on the product or services available in your area.

All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

The IBM home page on the Internet can be found at **ibm.com**

BLW03019USEN