Theme: Towards Reconfigurable HPC
Lecture **5**

Multicores at Work: The CELL Processor

Iris Christadler

Leibniz Supercomputing Centre

Inverted CERN School of Computing, 3-5 March 2008

Introduction

- **Objectives:**
  - Explain the peculiarities of the Cell architecture
  - Show examples of Cell and Playstation clusters
  - Introduce the Roadrunner project
  - Give an overview of software development for Cell

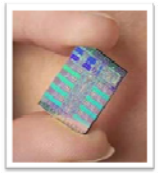


Image: IBM

Cell Broadband Engine Architecture

ARCHITECTURAL DETAILS

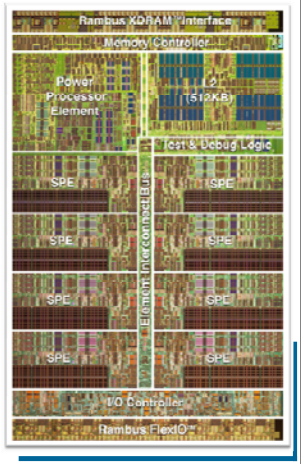
Image: IBM

STI CBEA (= Cell)

- **developed under Sony/Toshiba/IBM (STI) efforts**
- **current Cell chip is used in Sony's PlayStation3 (PS3)**
- **radical departure from previous mainstream processors**
- **8+1 way heterogeneous parallel processor**
- **high performance @ 3.2 GHz**
  - 204.8 GF/s single precision
  - 14.63 GF/s double precision
- **good ratio between performance and floor space as well as power consumption**
- **"nearly" IEEE-754 conform**

A Closer Look at the Hardware

- **1 Power Processing Element (PPE)**
  - general-purpose 64-bit RISC processor (PowerPC)
  - two way hardware multithreading
  - on-chip L1/L2 cache
    - 512 KB L2 cache
    - 32 KB L1 instruction cache
    - 32 KB L1 data cache
- **8 Synergistic Processor Elements (SPEs)**
  - special-purpose RISC processor
  - 128-bit SIMD
  - 256 KB local Memory
- **connected with the Element Interconnect Bus (EIB)**



A Closer Look at the Hardware (2)

SPE: Synergistic Processor Element
SPU: Synergistic Processor Unit
PPE: Power Processing Element

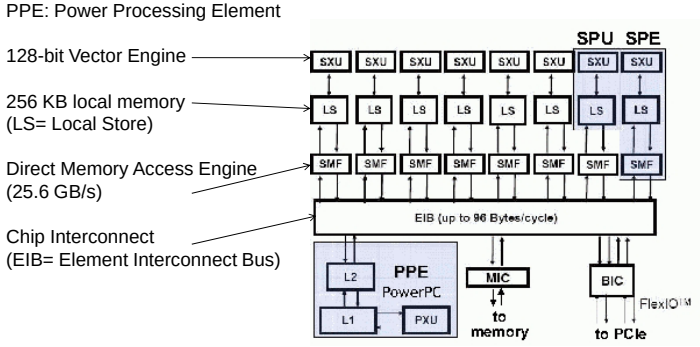


Image: IBM

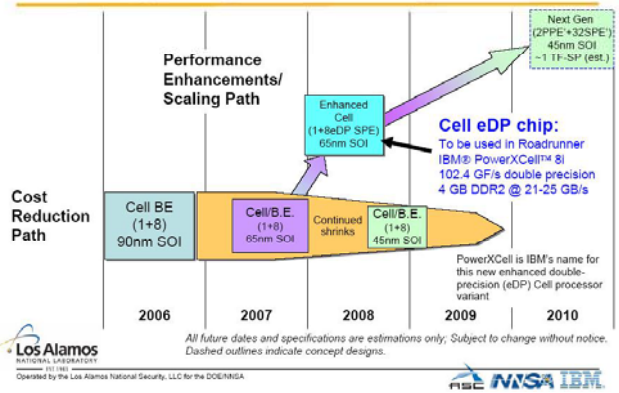
IBM QS20 Cell Blade

- **2 3.2 GHz Cell BE processors**
- **1 GB Memory (512 per processor)**
- **40 GB hard disk**
- **Dual Gigabit Ethernet**
- **InfiniBand 4x adapters**
- **Fedora Linux available**

Image: IBM

"…is especially suitable for computationally intense, high performance workloads across a number of industries including *digital media, medical imaging, aerospace, defense and communications*." (Advertisement claim)

Cell Broadband Engine™ Architecture (CBEA) Technology Competitive Roadmap



Can your Playstation save the world?

(AB)USING YOUR PLAYSTATION 3

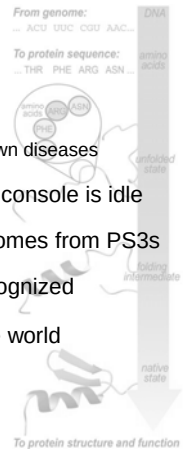
Installing Linux on PS3

- **Different Linux flavours available**
  (Fedora, Yellow Dog, …)
- **Possibility to use a**
  - bootable Linux from CD
  - permanent dual-boot installation setup
- **It's really simple**
  - download the necessary stuff
    (Linux images, PS3 add-on, kernel sources)
  - PS3: Settings -> System Settings -> Format Utility
  - PS3: Settings -> System Settings -> Install Other OS
  - PS3: Settings -> System Settings -> Default System
  - Install Linux

Detailed Installation Instructions e.g. at www.cellperformance.com

Folding@Home

- **Distributed computing project**
  [folding.stanford.edu]
- **Simulates protein folding**
  in order to understand and find cures for many well known diseases
- PS3 is able to perform computations while the console is idle
- ~75% of the combined computational power comes from PS3s
- Thanks to PS3s Folding@Home has been recognized
  by the Guinness World Records
  as the *most powerful distributed network* in the world
  - September 16, 2007, project surpassed 1 PF
  - September 23, 2007, PS3s alone reached 1PF



PLAYSTATION/
CELL CLUSTERS

Overview

- **Two ways of using Cell in HPC**
  - combined in clusters of Cells
  - attached as accelerators to commodity CPUs
- **Two different kind of Cell clusters**
  - "cheap" PS3 clusters with some drawbacks
  - more expensive IBM Blade Server Solution
- **Drawbacks of PlayStation Clusters**
  - 256 MB main memory is not enough
  - Slow network becomes the bottleneck
  - Linux is running under a hypervisor
  - PS3 comes with a special edition NVIDIA graphics card but unfortunately does not allow access to this resource

Univ. of Tennessee: PS3 Cluster

- **PlayStation3 Cluster**
  - 4 Playstations
  - GigaBit Ethernet switch
- **Power Consumption:**
  - ~ 800 Watt
  - ~ 0.5 -1 GF/Watt (sp)
- **Performance:**
  - 800 GF (sp)
- **Price:**
  - ~ 2400 $
  - ~ 330 MF/$

Image: Univ. of Tennessee

JUICE: Juelich Initiative Cell Cluster

- **Configuration**
  - 2 Chassis
  - each equipped with 6 Q20 cell blades
  - each of them include 2 Cell BEs
  - Frontend: Xeon
- **Networks**
  - Gigabit Ethernet
  - InfiniBand
- **Performance**
  (single prec., non IEEE)
  - 204.8 GF per Cell
  - 4.9 TF total

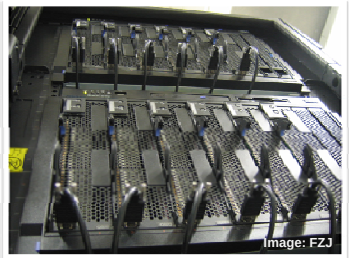


Image: FZJ

Running fast in New Mexico

THE ROADRUNNER PROJECT

http://www.lanl.gov/roadrunner/

The Roadrunner Project

- **Los Alamos National Laboratory (LANL)**
  Department of Energy (DoE)
- **First petaflop system worldwide**
  Peak Performance will be 1.33 PF
- **A hybrid Opteron-Cell system**
  Cells used as accelerators

„Will HPC turn away from homogeneous architectures and go hetero?"
[http://www.hpcwire.com/hpc/897414.html]

„IBM Unlocks the Cell"
[http://www.hpcwire.com/hpc/893353.html]

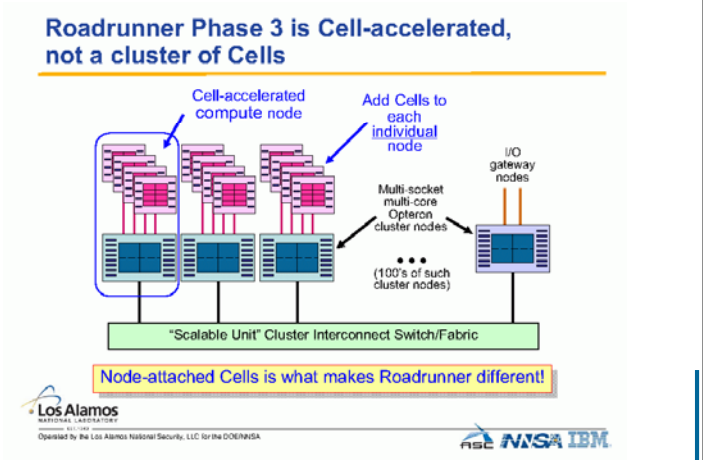
- **due in Q3/2008**
- **Early installation of a migration system to resolve programming issues**

Facts and Figures

- **Cluster of:**
  - 18 Connected Units (CU)
  - 6,912 AMD dual-core Opterons (1.8 GHz)
  - 12,960 IBM PowerXCells (~100 GF dp peak each)
- **Performance:**
  - 9.8 teraflops peak (Opteron)
  - 1.33 petaflops peak (Cell eDP)
  - designed for a sustained 1.0 petaflop LINPACK performance
  - side note: This is an efficiency of 75%

Connected Unit Specification

- **360 1.8 GHz dual-core Opterons**
  - 2.8 TF DP peak Opteron
  - 1.5 TB Opteron memory
- **720 3.2 GHz Cell eDP chips**
  - 73.7 TF peak Cell
  - 2.9 TB Cell memory
  - 15.4 TB/s Cell memory BW
- **192 InfiniBand (IB) 4x DDR cluster links**
  - 768 GB/s aggregate BW (bi-dir)
  - 384 GB/s bi-section BW (bi-dir)
- **24 10 GigE I/O links on 12 I/O nodes**
  - 24 GB/s aggregate I/O BW (uni-dir) (IB limited)



Slide: LANL

Power Consumption, Footprint,…

- **Power Consumption**
  - 3.9 MW Power
  - 0.35 GF/Watt (peak)
  - 0.25 GF/Watt (linpack)
- **Footprint**
  - 296 racks
  - 5500 ft^2
- **OS & SW**
  - OS: RHEL and Fedora Linux
  - IBM SDK for Multicore Acceleration
  - Cell compilers, libraries, SDK tools

IBMs Blue Gene

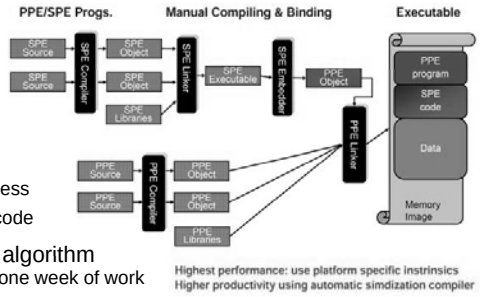
- **IBMs Blue Gene series is currently the second guess if you are talking about petaflop systems**
- **Rank 1 and 2 in the top500 list from Nov 2007 are BlueGenes**
  - Installation at LLNL with nearly 600 TF peak
  - Installation in Juelich with 220 TF peak
- **David Turek (vice president Deep Computing at IBM) in an hpcwire interview mid 2006:**
  "[…] if you look at Blue Gene today, the only thing that separates you from the deployment of a petaflop system is money. The future designs factor in a whole lot of other things - not only how you make a petaflop affordable, but also how do you open the aperture to an enhanced set of applications. […]
  And it's not at all in conflict with what we're doing here with Roadrunner because they're different programming models. For us, that's a key point of differentiation. Right now it looks like they may serve different application sets differently. "
  [http://www.hpcwire.com/hpc/893353.html]

Unfortunately not that easy ☺

The faster your code, the longer the development phase…

SOFTWARE DEVELOPMENT FOR CELL

Software Development for Cell



- **Examples**
  - SpMV
    - one month learning process
    - 600 lines of code
  - Stencil-based algorithm
    - still required one week of work
    - 250 lines of code (scalar version is 15 lines)

[Examples taken from "The Potential of the Cell Processor for Scientific Computing"]

Cell Software Development

- **Cell full system simulator for free download**
  [http://www.alphaworks.ibm.com/tech/cellsystemsim]
- **Important feature I: SIMD**
  To enable vectorization use for example loop-unrolling
- **Important feature II: Software-controlled memories**
  Data movement between registers, local memory and external DRAM is explicitly controlled by the application
- **Use intrinsics**
  Start with “Cell Broadband Engine Tutorial”, end up with the ultimate reference “C/C++ Language Extensions for Cell Broadband Engine Architecture”

Parallelization Models

- **Function offload model**
  main application executes on PPE,
  performance critical functions are offloaded to SPE
- **Parallelization models for the SPEs**
  application is mainly executed on the 8 SPEs
  - *Task Parallelism*
    independent tasks scheduled on each SPE
  - *Pipelined Parallelism*
    large blocks are passed from one SPE to the next
  - *Data Parallelism*
    processors perform identical computation on distinct data (most common, used in many scientific applications)

Software Development Kits

- **IBM SDK for Multicore Acceleration Version 3.0**
- **CorePy**
- **Octopiler**
- **Rapid-Mind**
- **(PeakStream)**
- **Cell Superscalar (CellSs)**
- **The Sequoia Language**
- **Mercury Multi-Core Framework**
- **…**

MIXED PRECISION PROGRAMMING FOR CELL

Motivation

- **Performance**
  - ~ 200 GF/s single precision
  - ~ 15 GF/s double precision
- **Idea**
  - use iterative refinement for some algorithms
  - calculate in double-precision only when necessary
- **Example: Jack Dongarra demonstrated**
  - a Linpack run
  - on a conventional Cell BE
  - with a performance equal to 100 GF

Why mixed precision is not necessary

- **Other people believe that the double-precision power of the accelerators will increase in the future and that there is no need for those algorithms**
- **But: It makes very much sense to think about the necessary precision for your algorithm**

Conclusions & Outlook

- **Conclusions:**
  - The Cell BE is an interesting alternative to commodity homogeneous multicore microprocessors
  - The new Cell eDP, due in 2008, will deliver high double-precision performance
  - Cell programming is still very tedious, alternatives are needed
- **Outlook:**
  - Will roadrunners become an endangered species?
  - Will data-stream programming become widely accepted?

Further Reading

- **A Rough Guide to Scientific Computing On the PlayStation3**
  [http://www.netlib.org/utk/people/JackDongarra/PAPERS/scop3.pdf]
- **The Potential of the Cell Processor for Scientific Computing**
  [http://www.lbl.gov/Science-Articles/Archive/sabl/2006/Jul/CellProcessorPotential.pdf]
- **Technical Report about FZJs Cell Cluster**
  [http://www.fz-juelich.de/jsc/docs/printable/ib/ib-07/ib-2007-13.pdf]
- **IBM Unlocks the Cell**
  [http://www.hpcwire.com/hpc/893353.html]
- **Roadrunner Homepage**
  [http://www.lanl.gov/roadrunner/]
- **IBM Developersite**
  [http://www.ibm.com/developerworks/power/cell/]