An introduction to POWER8 processor

Joel M. Tendler
Executive IT Architect
jtendler@us.ibm.com

Georgia IBM POWER User Group

January 16, 2014
The experts look ahead

Cramming more components onto integrated circuits

With unit cost falling as the number of components per circuit rises, by 1975 economics may dictate squeezing as many as 65,000 components on a single silicon chip

By Gordon E. Moore

Director, Research and Development Laboratories, Fairchild Semiconductor division of Fairchild Camera and Instrument Corp.

Electronics, Volume 38, Number 8, April 19, 1965
A “Classical” Silicon Technology Roadmap (Moore’s Law)

Year

1.8V

1.5V

1.2V

1.0V

< 1.0V

Vdd

Increasing Performance

0.22µm

0.18µm

0.13µm

0.7

0.7

0.7

300mm

90nm

65nm

45nm

200mm

1.0V

7S

7S - SOI

8S 7SF

8S2 8SE

8S3

9S 9SF

9S2

10S 9SF

10S2

11S 10SF

11S2

12S 11SF

© 2014 IBM Corporation

16 January 2014
Classical CMOS scaling

**Scaling**

Voltage: \( V/\alpha \)

Oxide: \( t_{\text{ox}}/\alpha \)

Wire width: \( W/\alpha \)

Gate width: \( L/\alpha \)

Diffusion: \( x_d/\alpha \)

Substrate: \( \alpha \cdot N_A \)

**Results**

Higher Density: \( \sim \alpha^2 \)

Higher Speed: \( \sim \alpha \)

Power/ckt: \( \sim 1/\alpha^2 \)

Power Density: \( \sim \text{Constant} \)
What is Classical Scaling?

- The **coordinated** reduction, year on year, of a fixed set of device dimensions governing the performance of silicon technology
Classical CMOS scaling

Scaling

- Voltage: $V/\alpha$
- Oxide: $t_{ox}/\alpha$
- Wire width: $W/\alpha$
- Gate width: $L/\alpha$
- Diffusion: $x_d/\alpha$
- Substrate: $\alpha * N_A$

Results

- Higher Density: $\sim \alpha^2$
- Higher Speed: $\times$
- Power/ckt: $\sim 1/\alpha^2$
- Power Density: $\sim$Constant
An introduction to POWER8 processor

Innovation Drives Performance

Physics is not permitting performance gains by technology scaling; however it is still enabling more transistors on a node to node basis.
CEOs consider technology the single most important external force shaping their organization’s future

Source: Question E8–What are the most important external forces that will impact the enterprise over the next 3 to 5 years?; n=884 [CEO only]
Client Business needs have evolved

- Data-driven insights
- Flexible, responsive IT environment
- Secure from external and internal threats
- Simplified end user experience
- Compelling ROI

8 zettabytes
digital content by 2015
(90% unstructured)

1 Trillion
Connected devices by 2015

25+
average # of mobility
applications to be deployed
by CIOs in next 2 years

69%
of IT cost is server
management &
administration
(est $247B)

80%
of new applications will
be developed with cloud
characteristics

8 zettabytes
digital content by 2015
(90% unstructured)

1 Trillion
Connected devices by 2015

25+
average # of mobility
applications to be deployed
by CIOs in next 2 years

69%
of IT cost is server
management &
administration
(est $247B)

80%
of new applications will
be developed with cloud
characteristics

1 ZB =
1000000000000000000000bytes
= 10^21 bytes
= 1000 exabytes
= 1 billion
terabytes

Client Business needs have evolved

- Data-driven insights
- Flexible, responsive IT environment
- Secure from external and internal threats
- Simplified end user experience
- Compelling ROI

Business needs are Transforming → Driving infrastructure Transformation

69%
of IT cost is server
management &
administration
(est $247B)

80%
of new applications will
be developed with cloud
characteristics

1 Trillion
Connected devices by 2015

25+
average # of mobility
applications to be deployed
by CIOs in next 2 years

8 zettabytes
digital content by 2015
(90% unstructured)

69%
of IT cost is server
management &
administration
(est $247B)

80%
of new applications will
be developed with cloud
characteristics

1 Trillion
Connected devices by 2015

25+
average # of mobility
applications to be deployed
by CIOs in next 2 years

8 zettabytes
digital content by 2015
(90% unstructured)
Expanding Power into Big Data, Analytics, Cloud, Linux

Power recognized leader for
- Business Applications – ERP, OLTP, Enterprise Web Applications
- Data Warehousing/Mining/Analytics
- Data Store

Expansion Play
- Big Data & Analytics
- Industry Solutions
- Linux & Open Source
- Cloud computing
- Cognitive Computing / IBM Watson

High Performance  Scalability  Data Capabilities  Security
Efficiency  Reliability and Availability  Advanced Virtualization
Industry Affiliation  Mature and Growth Markets  Large Scale Adoption
Data and Analytics are transforming business… does your analytics infrastructure enable you to keep pace?

Power and IBM Software are partnering to provide a comprehensive suite of capabilities

Cognitive
Learn Dynamically?

Prescriptive
Best Outcomes?

Predictive
What Could Happen?

Descriptive
What Has Happened?

Exploration & Discovery
What Do You Have?

power systems: the information engine

Data in Motion

Data at Rest

Data in Many Forms

Information Ingestion and Operational Information

Real-time Analytics

Exploration Enterprise Warehouse and Mart Analytics Appliances

Landing Area, Analytics Zone Archive

Analytic Platform and Applications

16 January 2014 © 2014 IBM Corporation
POWER8 Vision

**Leadership Performance**
- Increase core throughput at single thread, SMT2, SMT4, and SMT8 level
- Large step in per socket performance
- Enable more robust multi-socket scaling

**System Innovation**
- Higher capacity cache hierarchy and highly threaded processor
- Enhanced memory bandwidth, capacity, and expansion
- Dynamic code optimization
- Hardware-accelerated virtual memory management

**Open System Innovation**
- Coherent Accelerator Processor Interface (CAPI)
- Agnostic Memory interface
- Open system software

Optimize Analytics & Big Data  
Enhance Cloud Efficiency  
Enable Open Innovation on POWER
POWER8 Processor

Technology
• 22nm SOI, eDRAM, 15 ML 650mm²

Cores
• 12 cores (SMT8)
• 8 dispatch, 10 issue, 16 exec pipe
• 2X internal data flows/queues
• Enhanced prefetching
• 64K data cache, 32K instruction cache

Accelerators
• Crypto & memory expansion
• Transactional Memory
• VMM assist
• Data Move / VM Mobility

Caches
• 512 KB SRAM L2 / core
• 96 MB eDRAM shared L3
• Up to 128 MB eDRAM L4 (off-chip)

Memory
• Up to 230 GB/s sustained bandwidth

Bus Interfaces
• Durable open memory attach interface
• Integrated PCIe Gen3
• SMP Interconnect
• CAPI (Coherent Accelerator Processor Interface)

Energy Management
• On-chip Power Management Micro-controller
• Integrated Per-core VRM
• Critical Path Monitors
An introduction to POWER8 processor

POWER8 Processor

Technology
• 22nm SOI, eDRAM, 15 ML 650mm2

Cores
• 12 cores (SMT8)
• 8 dispatch, 10 issue, 16 exec pipe
• 2X internal data flows/queues
• Enhanced prefetching
• 64K data cache, 32K instruction cache

Accelerators
• Crypto & memory expansion
• Transactional Memory
• VMM assist
• Data Move / VM Mobility

Caches
• 512 KB SRAM L2 / core
• 96 MB eDRAM shared L3
• Up to 128 MB eDRAM L4 (off-chip)

Memory
• Up to 230 GB/s sustained bandwidth

Bus Interfaces
• Durable open memory attach interface
• Integrated PCIe Gen3
• SMP Interconnect
• CAPI (Coherent Accelerator Processor Interface)

Energy Management
• On-chip Power Management Micro-controller
• Integrated Per-core VRM
• Critical Path Monitors
POWER8 Core
POWER8 Core

Execution Improvement vs. POWER7
• SMT4 → SMT8
• 8 dispatch
• 10 issue
• 16 execution pipes:
  • 2 FXU, 2 LSU, 2 LU, 4 FPU,
  2 VMX, 1 Crypto, 1 DFU,
  1 CR, 1 BR
• Larger Issue queues (4 x 16-entry)
• Larger global completion,
  Load/Store reorder
• Improved branch prediction
• Improved unaligned storage
  access

Larger Caching Structures vs. POWER7
• 2x L1 data cache (64 KB)
• 2x outstanding data cache misses
• 4x translation Cache

Wider Load/Store
• 32B → 64B L2 to L1 data bus
• 2x data cache to execution dataflow

Enhanced Prefetch
• Instruction speculation awareness
• Data prefetch depth awareness
• Adaptive bandwidth awareness
• Topology awareness

Core Performance vs. POWER7
~1.6x Thread
~2x Max SMT
**POWER8 Core**

**Execution Improvement vs. POWER7**
- SMT4 → SMT8
- 8 dispatch
- 10 issue
- 16 execution pipes:
  - 2 FXU, 2 LSU, 2 LU, 4 FPU, 2 VMX, 1 Crypto, 1 DFU, 1 CR, 1 BR
- Larger Issue queues (4 x 16-entry)
- Larger global completion, Load/Store reorder
- Improved branch prediction
- Improved unaligned storage access

**Larger Caching Structures vs. POWER7**
- 2x L1 data cache (64 KB)
- 2x outstanding data cache misses
- 4x translation Cache

**Wider Load/Store**
- 32B → 64B L2 to L1 data bus
- 2x data cache to execution dataflow

**Enhanced Prefetch**
- Instruction speculation awareness
- Data prefetch depth awareness
- Adaptive bandwidth awareness
- Topology awareness

**Core Performance vs. POWER7**
- ~1.6x Thread
- ~2x Max SMT
POWER8 On-chip Caches

- L2: 512 KB 8 way per core
- L3: 96 MB (12 x 8 MB 8 way Bank)
- **“NUCA” Cache policy** (Non-Uniform Cache Architecture)
  Scalable bandwidth and latency
  Migrate “hot” lines to local L2, then local L3 (replicate L2 contained footprint)
- Chip Interconnect: 150 GB/sec x 12 segments per direction = 3.6 TB/sec
Cache Bandwidths

- GB/sec shown assuming 4 GHz
  Product frequency will vary based on model type

- Across 12 core chip
  4 TB/sec L2 BW
  3 TB/sec L3 BW
Memory Organization

- Up to 8 high speed channels, each running up to 9.6 Gb/s for up to 230 GB/s sustained
- Up to 32 total DDR ports yielding 410 GB/s peak at the DRAM
- Up to 1 TB memory capacity per fully configured processor socket
Memory Buffer Chip … with 16MB of Cache

Intelligence Moved into Memory
• Scheduling logic, caching structures
• Energy Mgmt, RAS decision point
  – Formerly on Processor
  – Moved to Memory Buffer

Processor Interface
• 9.6 GB/s high speed interface
• More robust RAS
• “On-the-fly” lane isolation/repair
• Extensible for innovation build-out

Performance Value
• End-to-end fastpath and data retry (latency)
• Cache → latency/bandwidth, partial updates
• Cache → write scheduling, prefetch, energy
• 22nm SOI for optimal performance / energy
• 15 metal levels (latency, bandwidth)
Centaur Memory DIMM

Memory DIMM
Form factors
2-hop Interconnect

48-way Drawer

38.4 GB/s  12.8 GB/s
P7 Routing

Source

Target

P7 node to node peak is 20 GB/sec
An introduction to POWER8 processor

Direct Route

Source → Target Bandwidth = 12.8 GB/sec

P7 node to node peak is 20 GB/sec
Source → Target Bandwidth = 25.6 GB/sec  
P7 node to node peak is 20 GB/sec

+Secondary Route
Source → Target Bandwidth = 51.2 GB/sec  P7 node to node peak is 20 GB/sec
Source → Target Bandwidth = 153.6 GB/sec  P7 node to node peak is 20 GB/sec
Socket Performance

- POWER7+ baseline
- Memory Bandwidth
- Commercial
- Java
- Integer
- Floating Point

Bar chart showing performance comparison.
Native PCIe Gen 3 Support
- Direct processor integration
- Replaces proprietary GX/Bridge
- Low latency
- Gen3 x16 bandwidth (16 Gb/s)

Transport Layer for CAPI Protocol
- Coherently Attach Devices connect to processor via PCIe
- Protocol encapsulated in PCIe
CAPI Coherent Accelerator Processor Interface

Virtual Addressing
- Accelerator can work with same memory addresses that the processors use
- Pointers de-referenced same as the host application
- Removes OS & device driver overhead

Hardware Managed Cache Coherence
- Enables the accelerator to participate in “Locks” as a normal thread
- Lowers Latency over IO communication model

Customizable Hardware Application Accelerator
- Specific system SW, middleware, or user application
- Written to durable interface provided by PSL

Processor Service Layer (PSL)
- Present robust, durable interfaces to applications
- Offload complexity / content from CAPP
## An introduction to POWER8 processor

<table>
<thead>
<tr>
<th></th>
<th>POWER5 2004</th>
<th>POWER6 2007</th>
<th>POWER7 2010</th>
<th>POWER7+ 2012</th>
<th>POWER8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>130nm SOI</td>
<td>65nm SOI</td>
<td>45nm SOI eDRAM</td>
<td>32nm SOI eDRAM</td>
<td>22nm SOI eDRAM</td>
</tr>
<tr>
<td>Compute</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cores</td>
<td>2 SMT2</td>
<td>2 SMT2</td>
<td>8 SMT4</td>
<td>8 SMT4</td>
<td>12 SMT8</td>
</tr>
<tr>
<td>Threads</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caching</td>
<td>1.9MB 36MB</td>
<td>8MB 32MB</td>
<td>2 + 32MB None</td>
<td>2 + 80MB None</td>
<td>6 + 96MB 128MB</td>
</tr>
<tr>
<td>On-chip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Off-chip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bandwidth</td>
<td>15GB/s 6GB/s</td>
<td>30GB/s 20GB/s</td>
<td>100GB/s 40GB/s</td>
<td>100GB/s 40GB/s</td>
<td>230GB/s 96GB/s</td>
</tr>
<tr>
<td>Sust. Mem.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak I/O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
An introduction to POWER8 processor

POWER8 Enabling: …Big Data, Analytics, Cognitive Computing…

POWER8 Differentiation for Analytics

- Massive capacity and bandwidth to memory and IO
- Large caches with massive bandwidth
- Strong Single thread
- SMT8, Many threads to hide memory latency
  - Graph traversals
  - Transactional memory enables efficient thread scaling

CAPI Accelerators

- Enables heterogeneous compute (GPU, FPGA, etc.)

Synergy with IBM Software, Driving Optimization Across the Stack
**OpenPOWER** will enable hyper-scale cloud data centers to rethink their approach to technology.

Member companies will use **POWER** for custom open servers and components for Linux based cloud data centers.

For the first time, **OpenPOWER** ecosystem partners can optimize the interactions of server building blocks – microprocessors, networking, I/O & other components – to tune performance.
An introduction to POWER8 processor

POWER8

- Significant Performance at Thread, Core, and System
- Optimization for VM Density & Efficiency
- Strong Enablement of Autonomic System Optimization
- Excellent Big Data Analytics Capability
Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.
Special notices (cont.)

IBM, the IBM logo, ibm.com AIX, AIX (logo), AIX 5L, AIX 6 (logo), AS/400, BladeCenter, Blue Gene, ClusterProven, DB2, ESCON, i5/OS, i5/OS (logo), IBM Business Partner (logo), IntelliStation, LoadLeveler, Lotus, Lotus Notes, Notes, Operating System/400, OS/400, PartnerLink, PartnerWorld, PowerPC, pSeries, Rational, RISC System/6000, RS/6000, THINK, Tivoli, Tivoli (logo), Tivoli Management Environment, WebSphere, xSeries, z/OS, zSeries, Active Memory, Balanced Warehouse, CacheFlow, Cool Blue, IBM Watson, IBM Systems Director VMControl, pureScale, TurboCore, Chipoppper, Cloudscale, DB2 Universal Database, DS4000, DS6000, DS8000, EnergyScale, Enterprise Workload Manager, General Parallel File System, GPFS, HACMP, HACMP/6000, HASM, IBM Systems Director Active Energy Manager, iSeries, Micro-Partitioning, POWER, PowerLinux, PowerExecutive, PowerVM, PowerVM (logo), PowerHA, Power Architecture, Power Everywhere, Power Family, POWER Hypervisor, Power Systems, Power Systems (logo), Power Systems Software, Power Systems Software (logo), POWER2, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, POWER7, POWER7+, Systems, System i, System p, System p5, System Storage, System z, TME 10, Workload Partitions Manager and X-Architecture are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries.

A full list of U.S. trademarks owned by IBM may be found at: http://www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

AltiVec is a trademark of Freescale Semiconductor, Inc.

AMD Opteron is a trademark of Advanced Micro Devices, Inc.

InfiniBand, InfiniBand Trade Association and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.

PowerLinux™ uses the registered trademark Linux® pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the Linux® mark on a worldwide basis.

Microsoft, Windows and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries or both.

NetBench is a registered trademark of Ziff Davis Media in the United States, other countries or both.

SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECapc, SPECchpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC).

The Power Architecture and Power.org wordmarks and the Power and Power.org logos and related marks are trademarks and service marks licensed by Power.org.

TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC).

UNIX is a registered trademark of The Open Group in the United States, other countries or both.

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

16 January 2014 © 2014 IBM Corporation