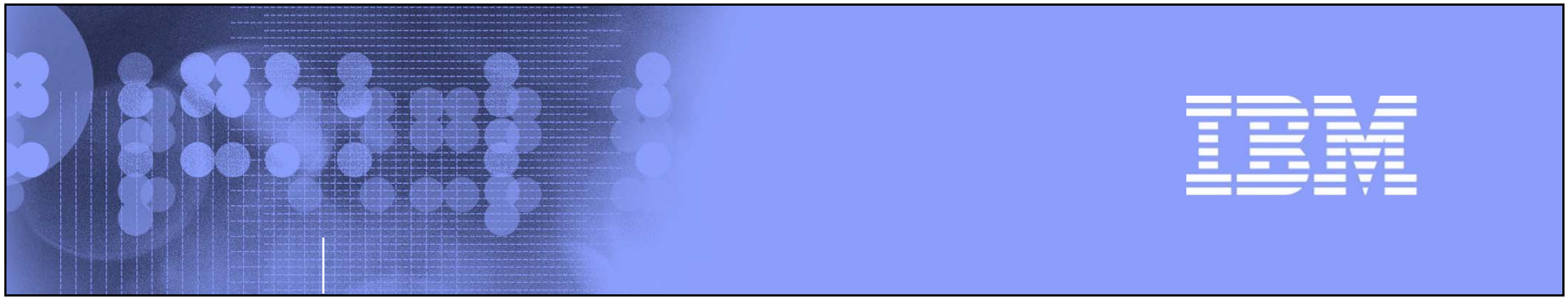
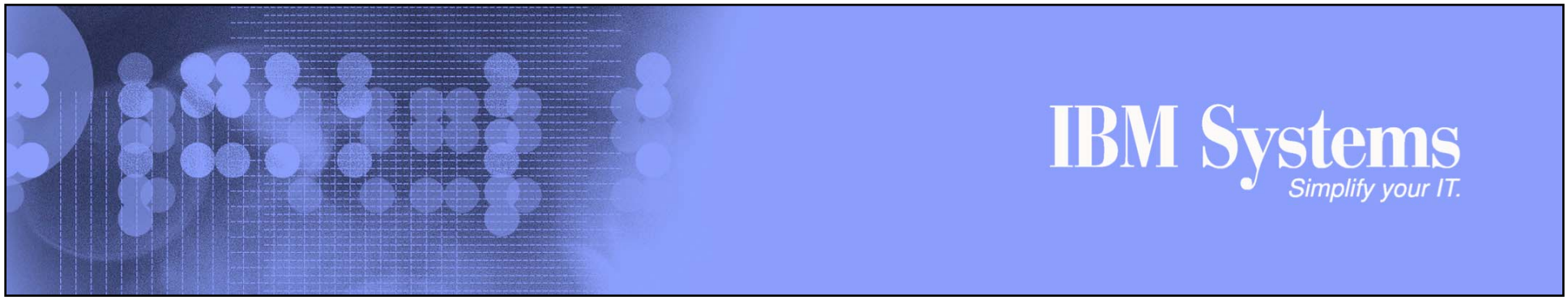


Bringing Heterogeneous Multiprocessors Into the Mainstream

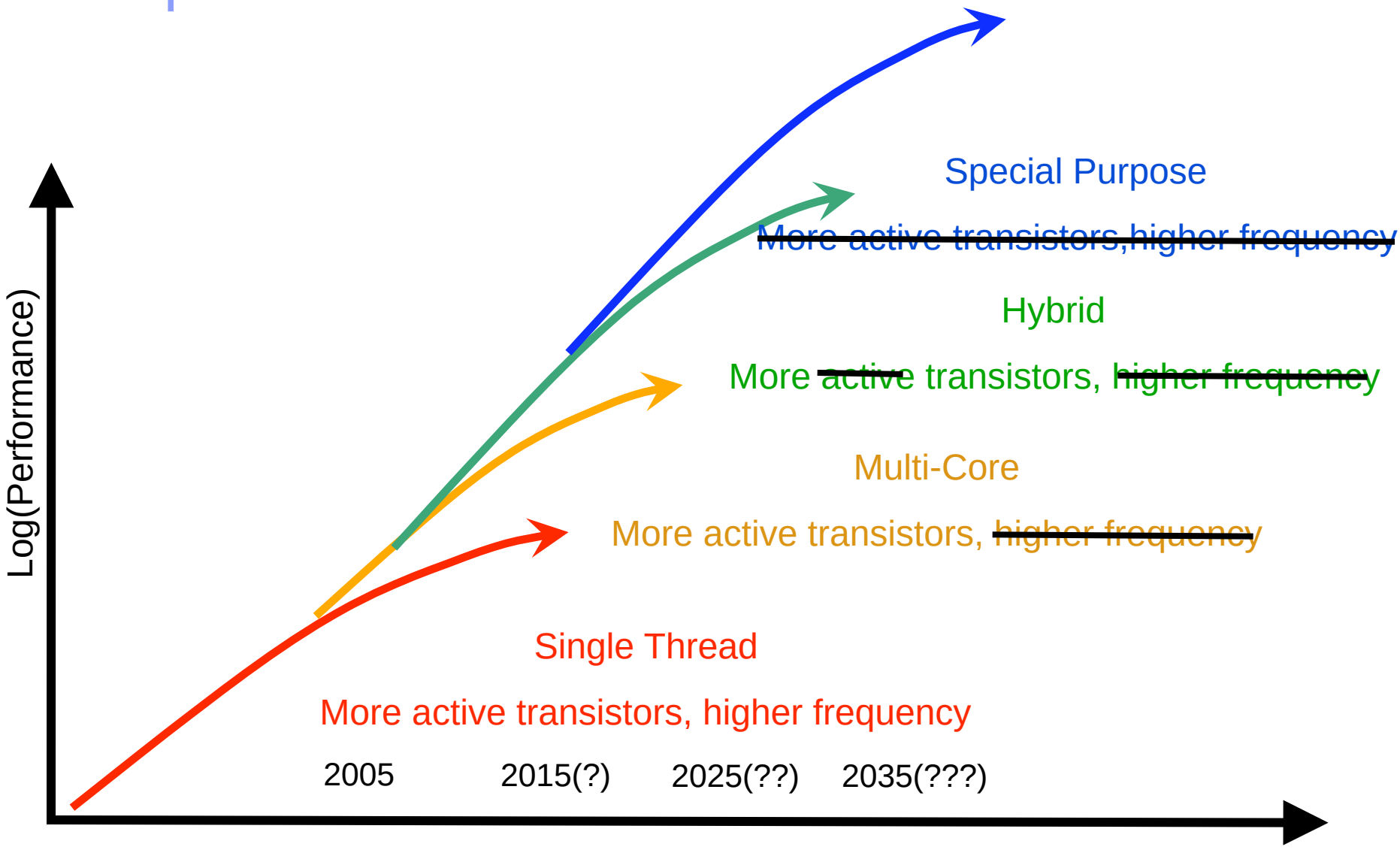
August 9 2009

Brian Flachs
Cell/B.E. Architect

IBM Systems and Technology Group



Microprocessor Trends



Heterogeneity

- Spreading tasks within a multiprocessor
- Hope to divide tasks according to characteristics
  - Dynamic branch prediction valuable vs. static + conditional move
  - Caches effective vs. memory latency tolerance
  - Tight data dependancies vs. control flow/address stream is data independent
- Design processors differently for ***Efficiency***
  - Accelerators
  - Control plane:
    - High voltage, small register file, short pipeline, low frequency with dynamic branch prediction and large caches
  - Data plane:
    - Low voltage, large register file, long pipeline, high frequency with memory latency tolerance

Potential Benefits of Heterogeneity

- More efficient use of memory bandwidth
- More performance per area
- More performance per watt
- Fewer modules, boards, racks, etc.

- Larger portion of interconnect on chip
- More computation per soft-fail / check point
- Probably a big part of how to achieve exa-scale

- Lots of accelerators
  - Graphics, Floating Point, Integer, O/S, Cryptography, Compression, Network, XML
  - Level of programmability varies as does data access model

- But … idle accelerators are not efficient
  - Programmer effort can limit utilization of accelerators
  - Limits specialization of accelerators
  - Limits investment in accelerators

Requirements for Heterogeneous Compilation

(What I have learned from the OpenMP compiler)

- Want to support standard languages: C, C++, Fortran, OpenMP, OpenCL
  - Don’t want separate sources for the accelerators
  - Really don’t want separate #includes
- Assume different Architecture / Micro-architecture benefits from or requires different binaries
  - Optimize for different instruction scheduling, execution latencies, memory system parameters
  - Special instructions & operations
- Want a tool chain to be mostly transparent
  - Need to honor programmer directives
    - Parallel Tasks/Data fetch
  - Runtime & task-queue may become part of ABI
  - Ok to sacrifice portability for better results in high value/benefit codes

Fat Binaries

- Architecture Requirements:
  - Code needs to migrate from any of the processor types to any other processor type
- Compile everything for all cores
  - Analyze compiler output for core type assignment metrics
    - Accelerator compiler could fail
    - have source level pragmas
  - Might have different sources for different processor types
- Fat binaries can run anything anywhere
  - At least via RPC/migration
  - Cost metrics for execution migration
  - Load balancing & task queue architecture

Heterogeneous Code-Gen Problems

- Function pointers
  - Translation table
    - From function pointer address space to actual entry point
    - All constructor code runs the same on all cores
    - Adds extra path to C++ virtual method calls
  - Self-modifying code?
    - architecture independent: llvm
- Data pointers
  - Shared address space
  - Coherency management
- Data formats
  - Little endian/Big endian, etc.
  - Conversions probably impact source

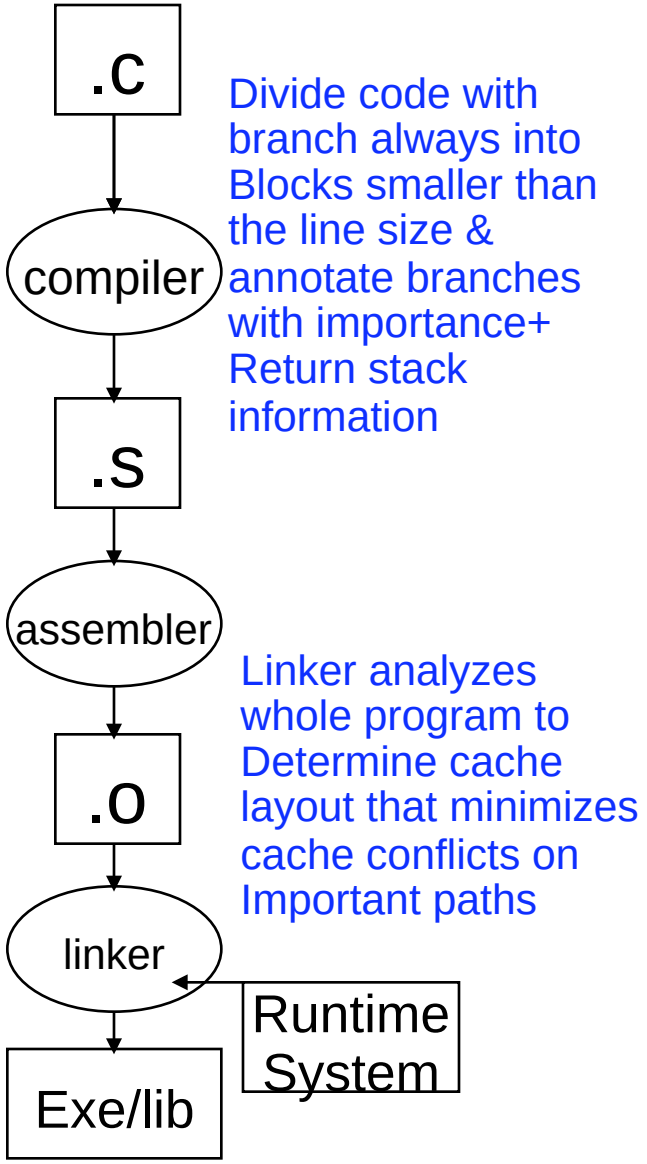
Just In Time Compilation

- Binary distribution does not need to anticipate accelerators that might be present
  - Old programs might be able to take advantage of new accelerators
  - Accelerator architecture might not need to be held constant for long periods of time
- Intermediate needs enough information to do a good job
- How to replace hand-optimized assembly?
  - Some optimizations are ‘hard’ for compilers.

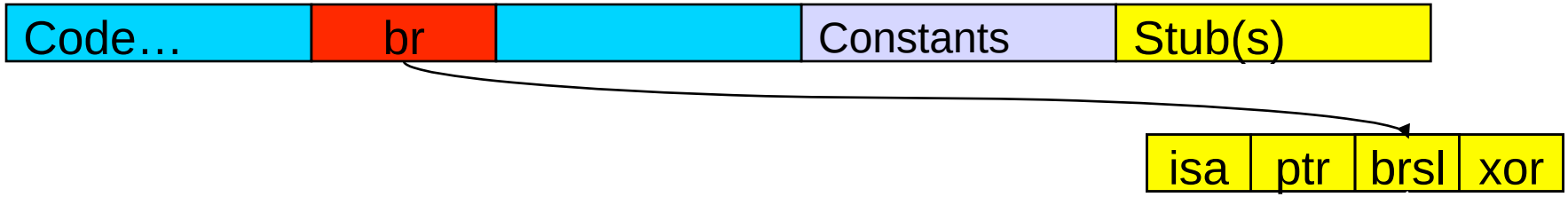
Soft I-Cache for SPE

- Want to run a larger code base
- Limited local store forces small programs
  - Limited libraries
- Users don't like overlay system

- Upto ½ GB of code
- Normal tool-chain flow
  - No detailed knowledge required on the part of the developer.
- 'Small' changes to ABI – good operability with old source.
  - 32 bit virtual address space for code
  - Support code out-side of cache structure



A Cache Block

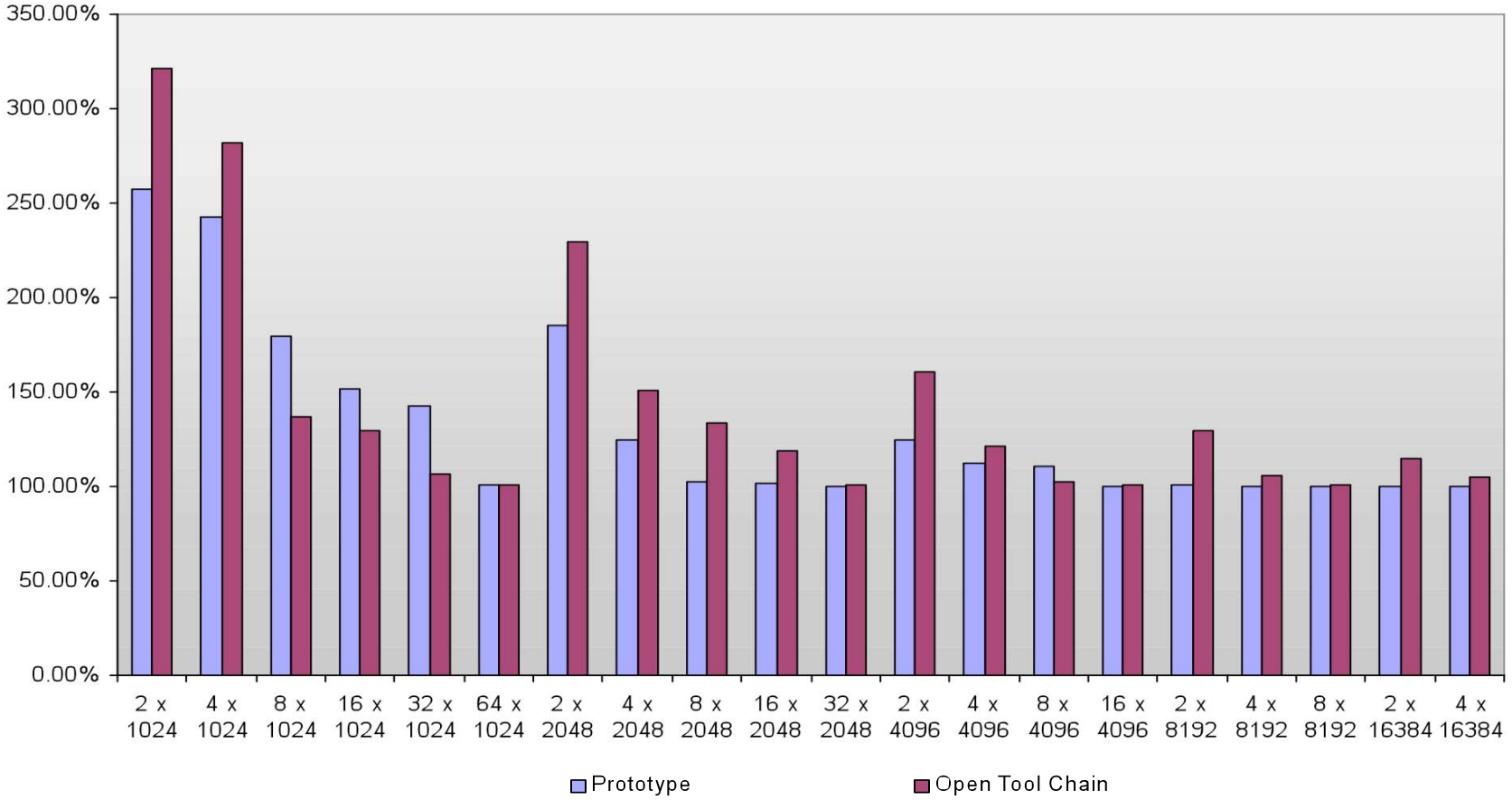


- Linker creates a “stub” at the end of the cache line for each branch immediate
- Stub has 4 word sized fields:
  - Address of target
  - Pointer to the branch
  - Trampoline
  - XOR pattern
- When line is brought from memory into cache branches immediate branches to a brsl (the trampoline (T)) in the stub to the runtime entry point.
  - Brsl records a pointer to the stub

Rewriting

- Runtime patches branch to point to target.
  - Load & rototate xor pattern
    - Converts branch to stub to branch to target
  - Load quadword pointed to by stub (contains the branch)
  - Xor
  - Store back the quadword with the branch
- Next time branch goes directly to intended target
  - no extra delay
- Easy because cache is direct mapped.
- Branch hints work!!!!
- Must un-rewrite if target is evicted
- Indirects require tag check

Performance Relative to LS Static



- Miss penalty = 400 cycle access to main memory in parallel with cache management software
- Less than 10% total runtime penalty for running in small caches.

Next Gen Hardware

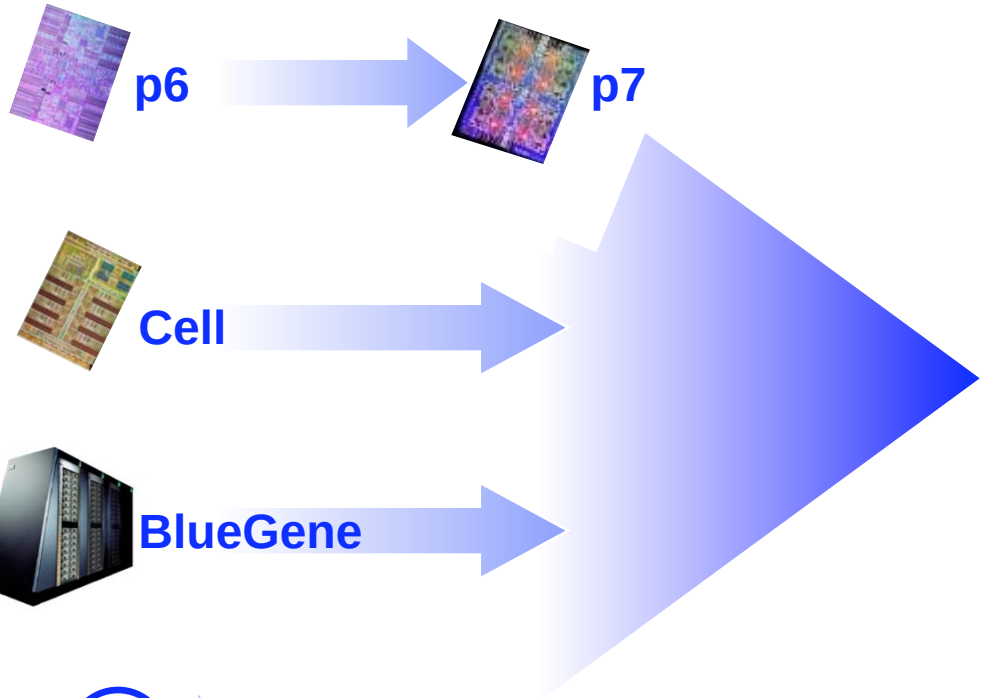
- More chips in systems
- More cores on chips
  - Bandwidth might halt this -> more cores per module
- Slightly better single thread performance or
  - At exa-scale, power limits make reduce single thread performance
- SMP Attached Accelerators
  - Either consolidation driven by standard language OpenCL or
  - Diversification driven by workable heterogeneous tool chain

- Where is the parallelism coming from?
- Where is the latency tolerance coming from?

Next Era of Innovation – Hybrid Computing
The Next Bold Step in Innovation & Integration

Symmetric Multiprocessing Era

Today *pNext 1.0*

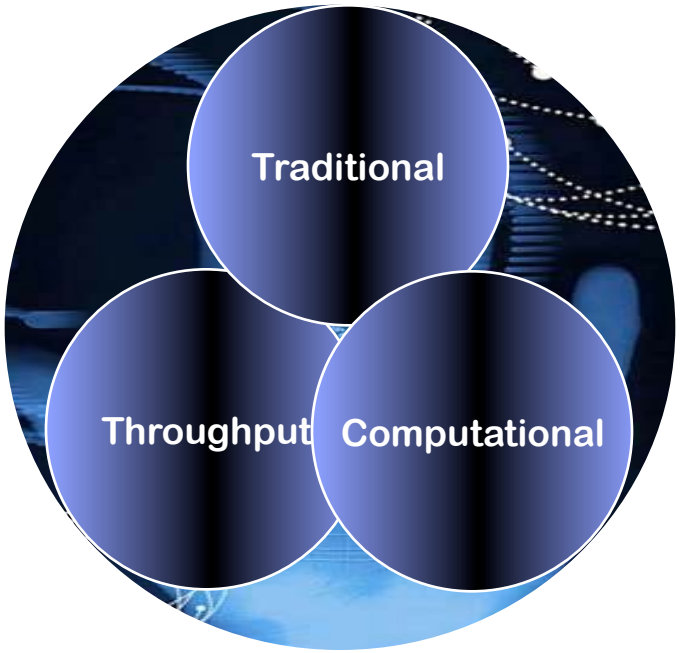


Technology Out

Driven by cores/threads

Hybrid Computing Era

pNext 2.0



Market In

Driven by workload consolidation

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.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

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.)

The following terms are registered trademarks of International Business Machines Corporation in the United States and/or other countries: AIX, AIX/L, AIX/L(logo), alphaWorks, AS/400, BladeCenter, Blue Gene, Blue Lightning, C Set++, CICS, CICS/6000, ClusterProven, CT/2, DataHub, DataJoiner, DB2, DEEP BLUE, developerWorks, DirectTalk, Domino, DYNIX, DYNIX/ptx, e business(logo), e(logo)business, e(logo)server, Enterprise Storage Server, ESCON, FlashCopy, GDDM, i5/OS, IBM, IBM(logo), ibm.com, IBM Business Partner (logo), Informix, IntelliStation, IQ-Link, LANStreamer, LoadLeveler, Lotus, Lotus Notes, Lotusphere, Magstar, MediaStreamer, Micro Channel, MQSeries, Net.Data, Netfinity, NetView, Network Station, Notes, NUMA-Q, OpenPower, Operating System/2, Operating System/400, OS/2, OS/390, OS/400, Parallel Sysplex, PartnerLink, PartnerWorld, Passport Advantage, POWERparallel, Power PC 603, Power PC 604, PowerPC, PowerPC(logo), Predictive Failure Analysis, pSeries, PTX, ptx/ADMIN, RETAIN, RISC System/6000, RS/6000, RT Personal Computer, S/390, Scalable POWERparallel Systems, SecureWay, Sequent, ServerProven, SpaceBall, System/390, The Engines of e-business, THINK, Tivoli, Tivoli(logo), Tivoli Management Environment, Tivoli Ready(logo), TME, TotalStorage, TURBOWAYS, VisualAge, WebSphere, xSeries, z/OS, zSeries.

The following terms are trademarks of International Business Machines Corporation in the United States and/or other countries: Advanced Micro-Partitioning, AIX 5L, AIX PVMe, AS/400e, Chiphopper, Chipkill, Cloudscape, DB2 OLAP Server, DB2 Universal Database, DFDSM, DFSORT, DS4000, DS6000, DS8000, e-business(logo), e-business on demand, eServer, Express Middleware, Express Portfolio, Express Servers, Express Servers and Storage, General Purpose File System, GigaProcessor, GPFS, HACMP, HACMP/6000, IBM TotalStorage Proven, IBMLink, IMS, Intelligent Miner, iSeries, Micro-Partitioning, NUMACenter, On Demand Business logo, POWER, PowerExecutive, Power Architecture, Power Everywhere, Power Family, Power PC, PowerPC Architecture, PowerPC 603, PowerPC 603e, PowerPC 604, PowerPC 750, POWER2, POWER2 Architecture, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, pure XML, Redbooks, Sequent (logo), SequentLINK, Server Advantage, ServeRAID, Service Director, SmoothStart, SP, System i, System i5, System p, System p5, System Storage, System z, System z9, S/390 Parallel Enterprise Server, Tivoli Enterprise, TME 10, TotalStorage Proven, Ultramedia, VideoCharger, Virtualization Engine, Visualization Data Explorer, X-Architecture, z/Architecture, z/9.

A full list of U.S. trademarks owned by IBM may be found at: http://www.ibm.com/legal/copytrade.shtml.
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.
UNIX is a registered trademark of The Open Group in the United States, other countries or both.
Linux is a trademark of Linus Torvalds in the United States, other countries or both.
Microsoft, Windows, Windows NT and the Windows logo are registered trademarks of Microsoft Corporation in the United States, other countries or both.
Intel, Itanium, Pentium are registered tradas and Xeon is a trademark of Intel Corporation or its subsidiaries in the United States, other countries or both.
AMD Opteron is a trademark of Advanced Micro Devices, Inc.
Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.
TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC).
SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECapc, SPEChpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC).
NetBench is a registered trademark of Ziff Davis Media in the United States, other countries or both.
AltiVec is a trademark of Freescale Semiconductor, Inc.
Cell Broadband Engine is a trademark of Sony Computer Entertainment Inc.
Other company, product and service names may be trademarks or service marks of others.