



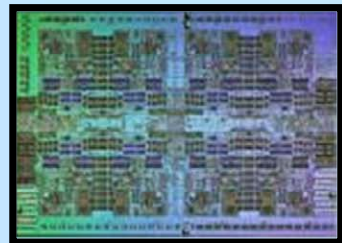
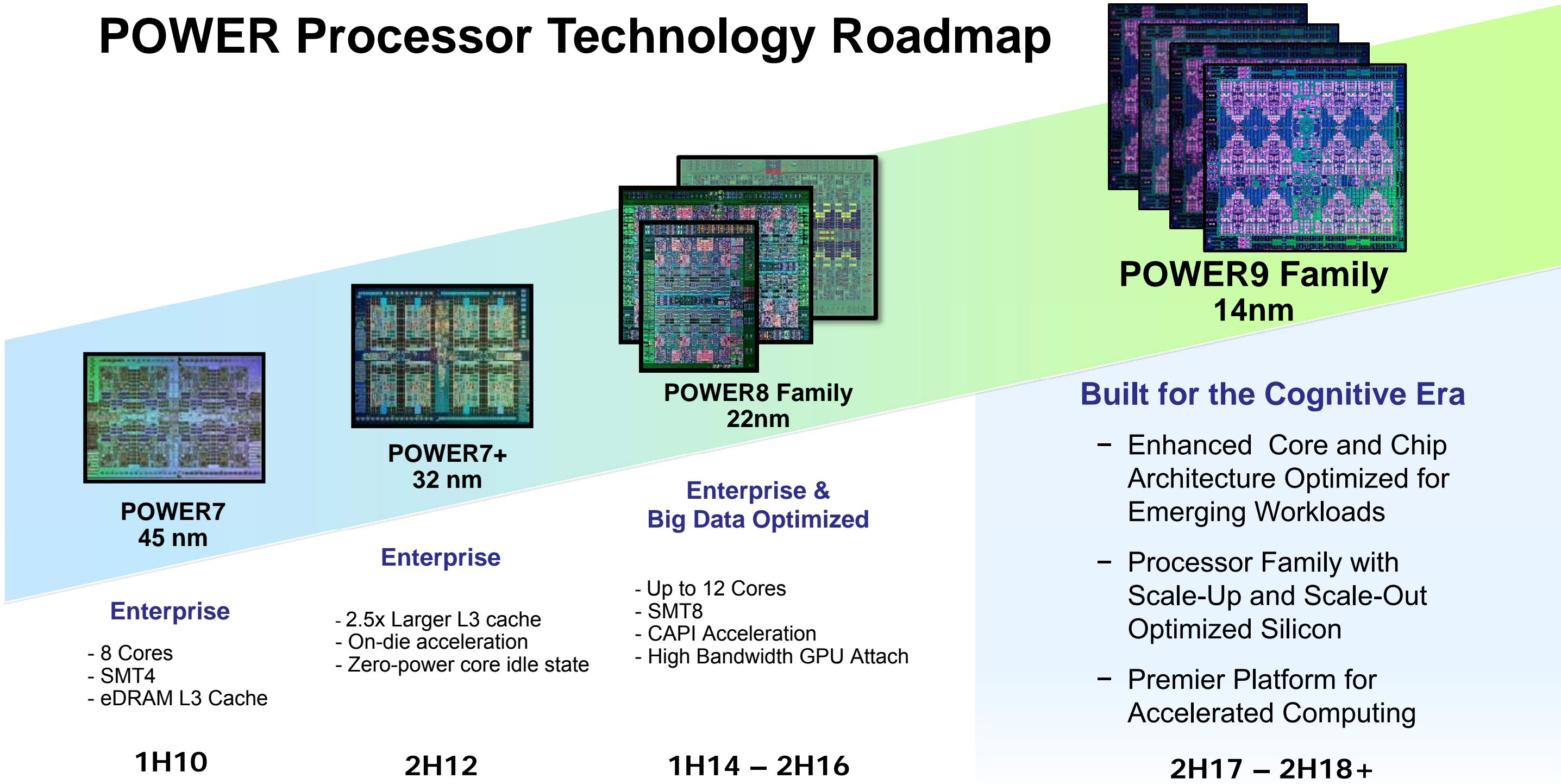
# POWER9 Processor for the Cognitive Era

**Brian Thompto**

POWER Systems, IBM Systems



# POWER Processor Technology Roadmap

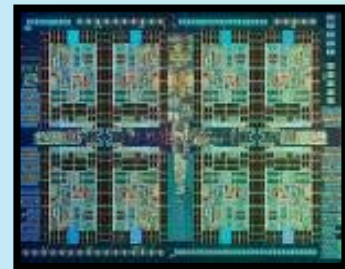


**POWER7**  
45 nm

**Enterprise**

- 8 Cores
- SMT4
- eDRAM L3 Cache

**1H10**

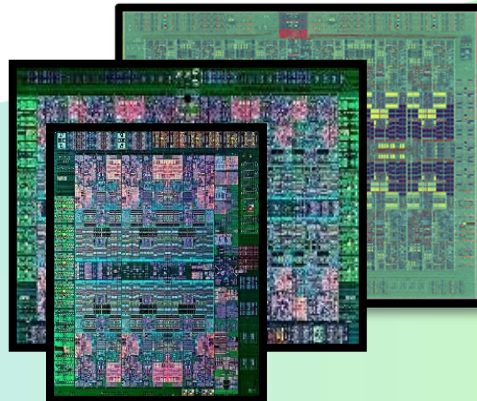


**POWER7+**  
32 nm

**Enterprise**

- 2.5x Larger L3 cache
- On-die acceleration
- Zero-power core idle state

**2H12**

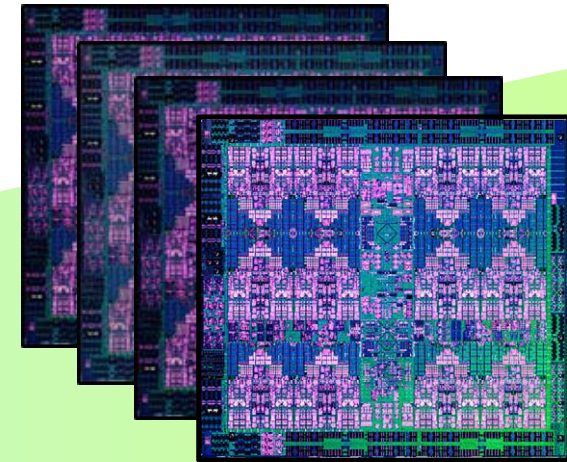


**POWER8 Family**  
22nm

**Enterprise &  
Big Data Optimized**

- Up to 12 Cores
- SMT8
- CAPI Acceleration
- High Bandwidth GPU Attach

**1H14 – 2H16**



**POWER9 Family**  
14nm

**Built for the Cognitive Era**

- Enhanced Core and Chip Architecture Optimized for Emerging Workloads
- Processor Family with Scale-Up and Scale-Out Optimized Silicon
- Premier Platform for Accelerated Computing

**2H17 – 2H18+**

## Emerging Analytics, AI, Cognitive

- New core for stronger thread performance
- Delivers 2x compute resource per socket
- Built for acceleration – OpenPOWER solution enablement



## Technical / HPC

- Highest bandwidth GPU attach
- Advanced GPU/CPU interaction and memory sharing
- High bandwidth direct attach memory



## Cloud / HSDC

- Power / Packaging / Cost optimizations for a range of platforms
- Superior virtualization features: security, power management, QoS, interrupt
- State of the art IO technology for network and storage performance



## Enterprise

- Large, flat, Scale-Up Systems
- Buffered memory for maximum capacity
- Leading RAS
- Improved caching





## New Core Microarchitecture

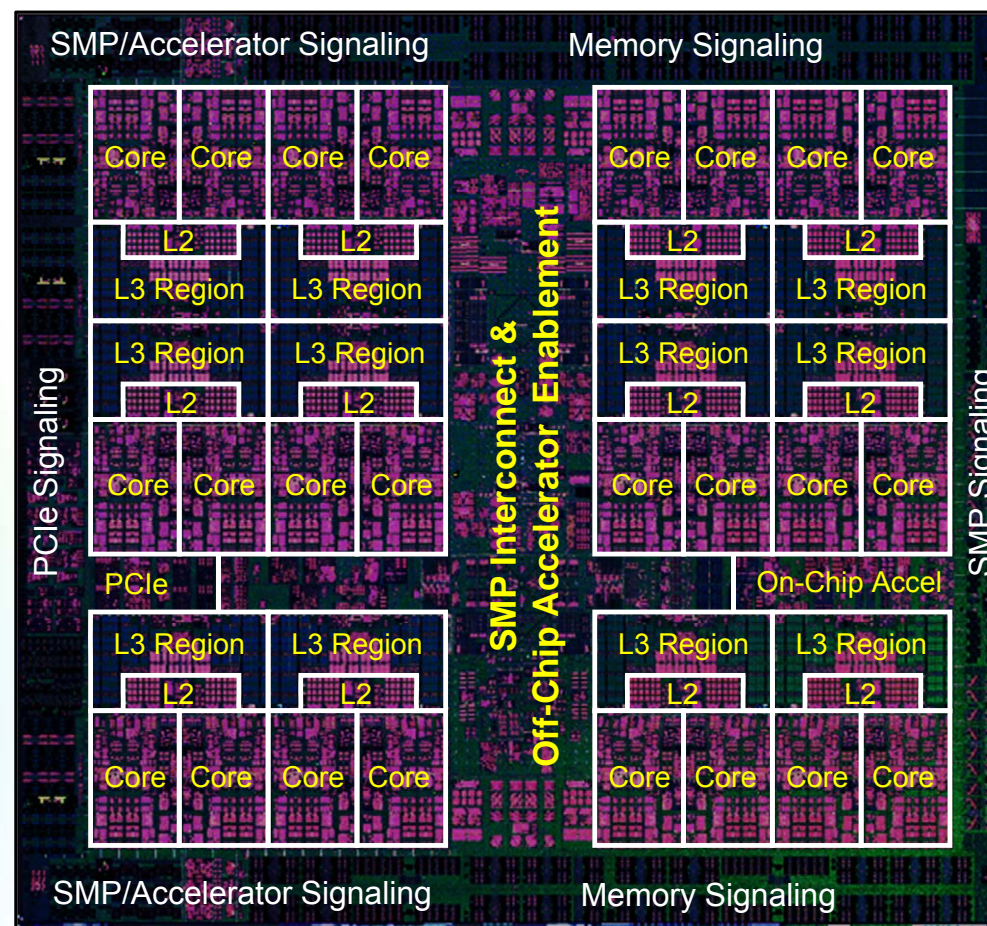
- Stronger thread performance
- Efficient agile pipeline
- POWER ISA v3.0

## Enhanced Cache Hierarchy

- 120MB NUCA L3 architecture
- 12 x 20-way associative regions
- Advanced replacement policies
- Fed by 7 TB/s on-chip bandwidth

## Cloud + Virtualization Innovation

- Quality of service assists
- New interrupt architecture
- Workload optimized frequency
- Hardware enforced trusted execution



## 14nm finFET Semiconductor Process

- Improved device performance and reduced energy
- 17 layer metal stack and eDRAM
- 8.0 billion transistors

## Leadership

### Hardware Acceleration Platform

- Enhanced on-chip acceleration
- Nvidia NVLink 2.0: High bandwidth, advanced new features
- CAPI 2.0: Coherent accelerator and storage attach (PCIe G4)
- New CAPI: Improved latency and bandwidth, open interface

### State of the Art I/O Subsystem

- PCIe Gen4 – 48 lanes

### High Bandwidth Signaling Technology

- 16 Gb/s interface
  - Local SMP
- 25 Gb/s interface – 25G Link
  - Accelerator, remote SMP

## Four targeted implementations

### SMP scalability / Memory subsystem

#### Scale-Out – 2 Socket Optimized

#### Robust 2 socket SMP system

#### Direct Memory Attach

- Up to 8 DDR4 ports
- Commodity packaging form factor

#### Scale-Up – Multi-Socket Optimized

#### Scalable System Topology / Capacity

- Large multi-socket

#### Buffered Memory Attach

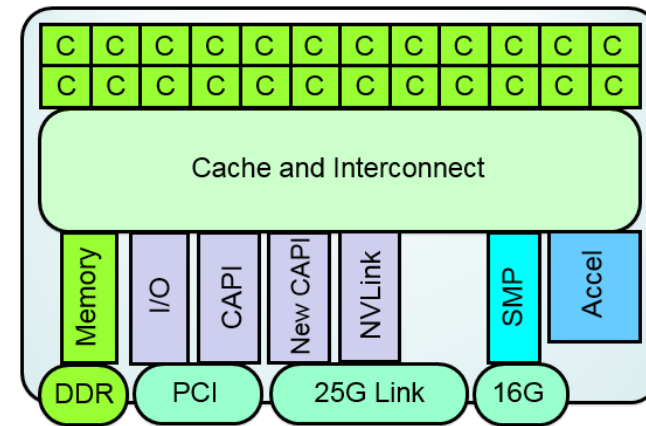
- 8 Buffered channels

## Core Count / Size

### SMT4 Core

24 SMT4 Cores / Chip

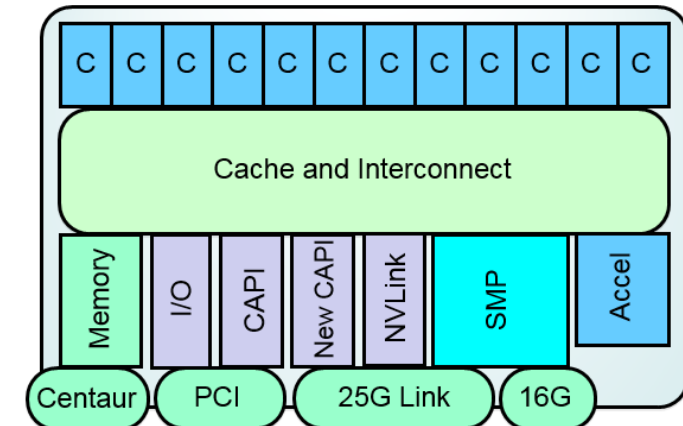
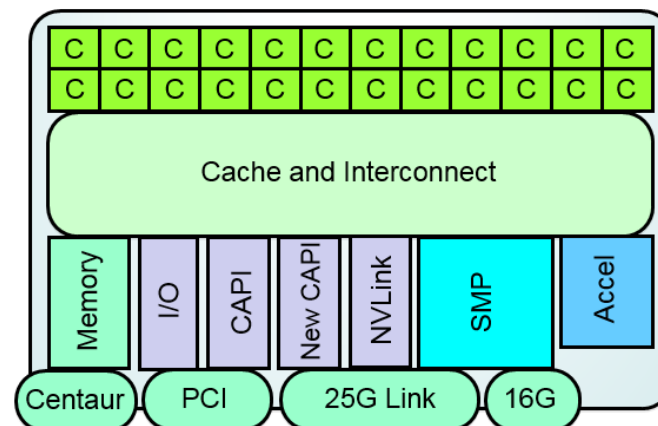
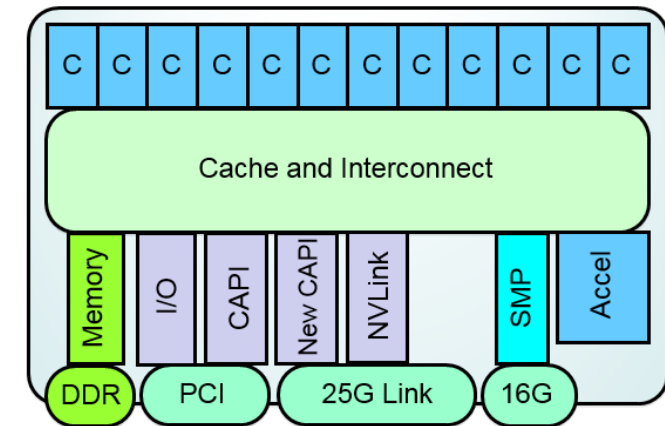
Linux Ecosystem Optimized



### SMT8 Core

12 SMT8 Cores / Chip

PowerVM Ecosystem Continuity



## Optimized for Stronger Thread Performance and Efficiency

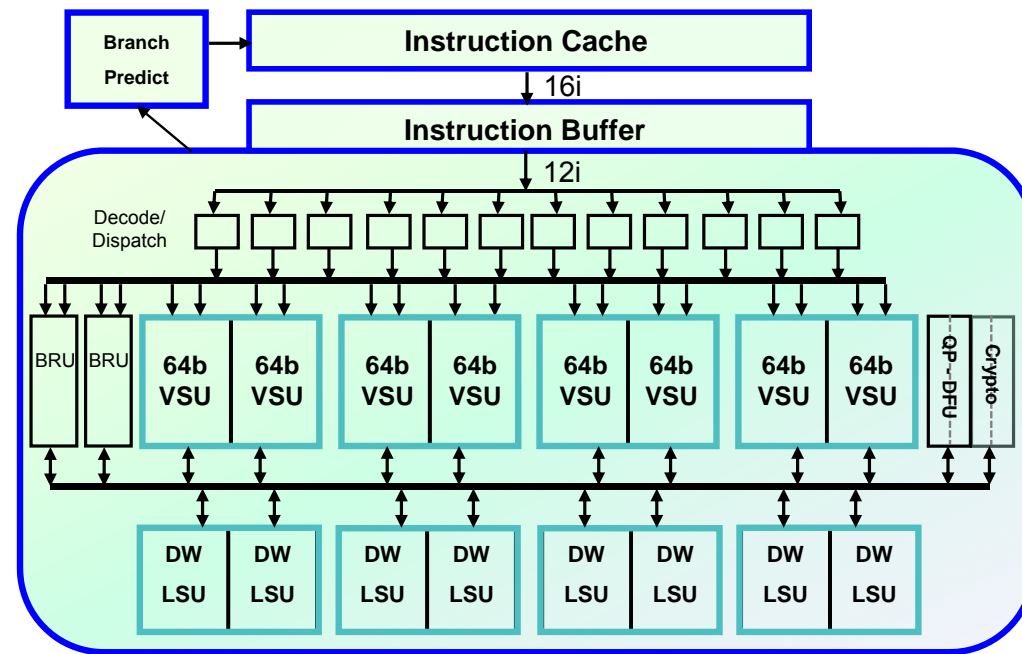
- Increased execution bandwidth efficiency for a range of workloads including commercial, cognitive and analytics
- Sophisticated instruction scheduling and branch prediction for unoptimized applications and interpretive languages
- Adaptive features for improved efficiency and performance especially in lower memory bandwidth systems

## Available with SMT8 or SMT4 Cores

8 or 4 threaded core built from modular execution slices

### POWER9 SMT8 Core

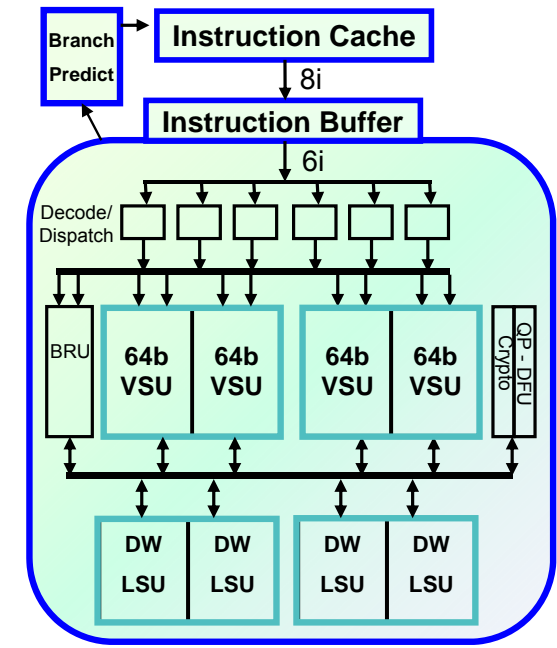
- PowerVM Ecosystem Continuity
- Strongest Thread
- Optimized for Large Partitions



**SMT8 Core**

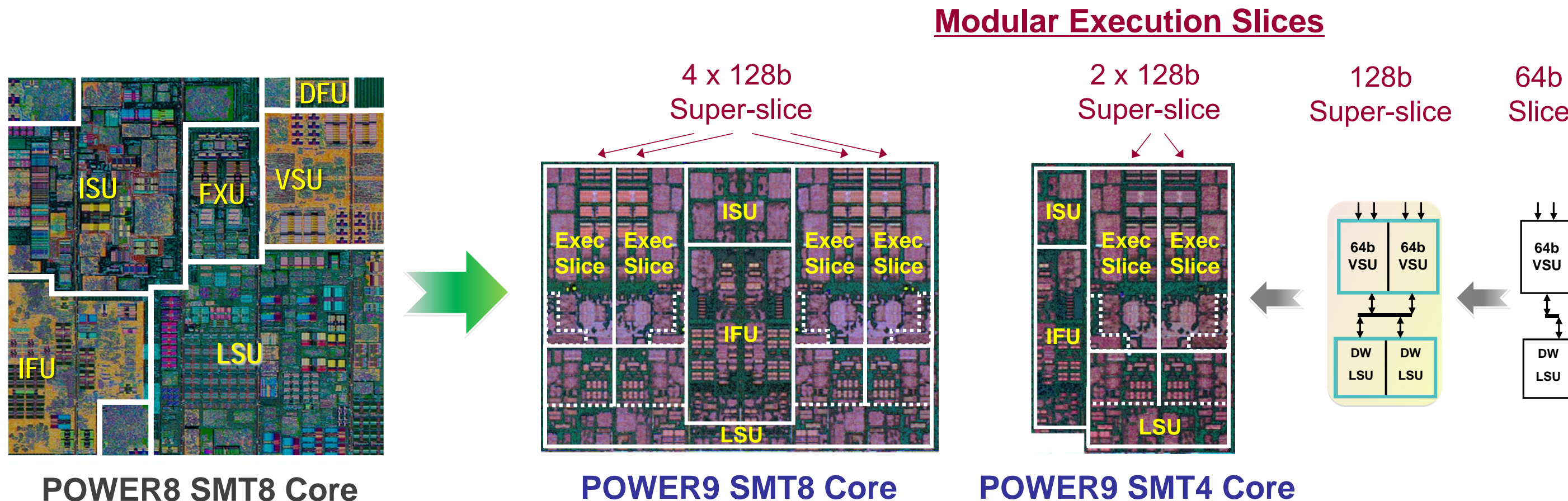
### POWER9 SMT4 Core

- Linux Ecosystem Focus
- Core Count / Socket
- Virtualization Granularity



**SMT4 Core**





## Re-factored Core Provides Improved Efficiency & Workload Alignment

- Enhanced pipeline efficiency with modular execution and intelligent pipeline control
- Increased pipeline utilization with symmetric data-type engines: Fixed, Float, 128b, SIMD
- Shared compute resource optimizes data-type interchange

## Shorter Pipelines with Reduced Disruption

### Improved application performance for modern codes

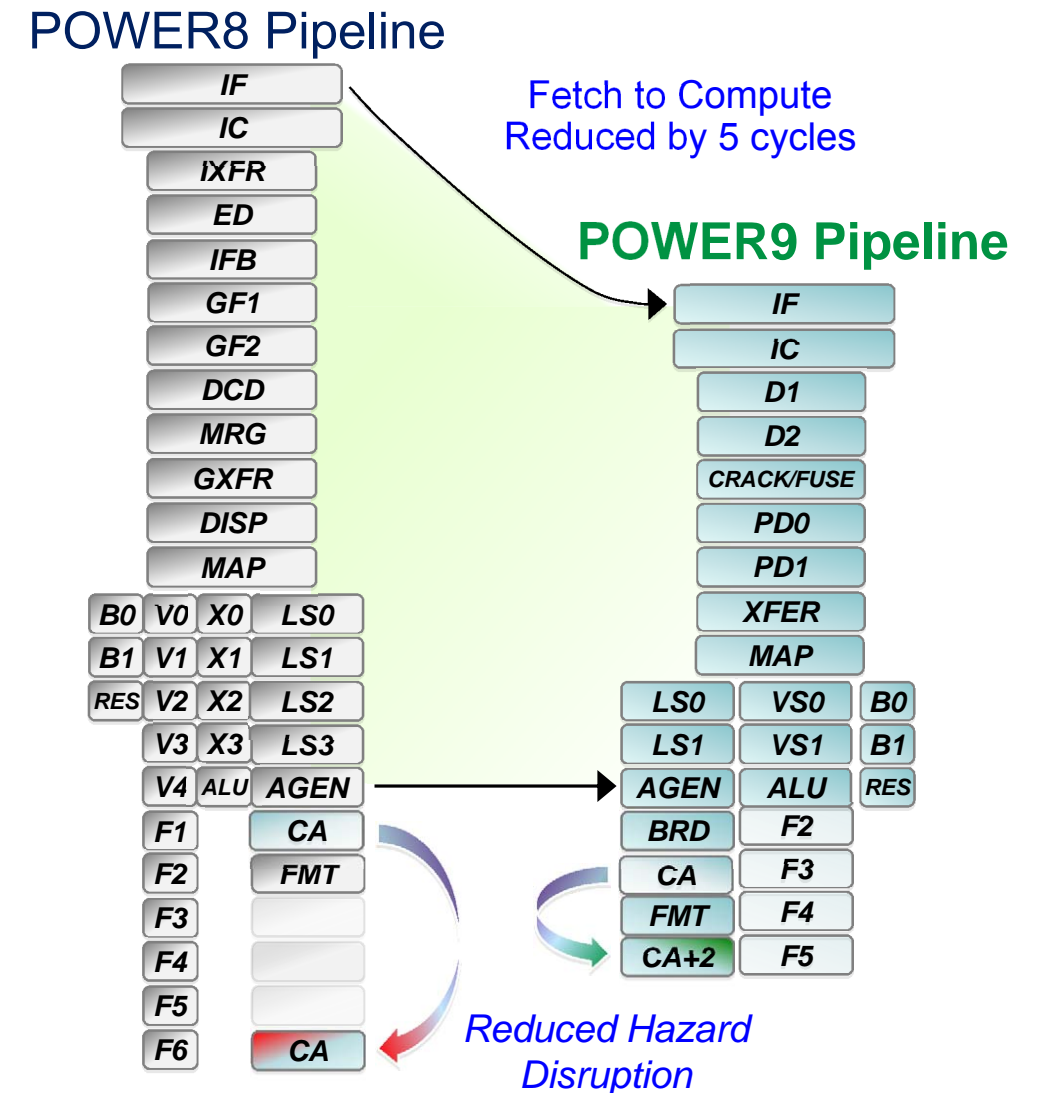
- Shorten fetch to compute by 5 cycles
- Advanced branch prediction

### Higher performance and pipeline utilization

- Improved instruction management
  - Removed instruction grouping and reduced cracking
  - Enhanced instruction fusion
  - Complete up to 128 (64 – SMT4 Core) instructions per cycle

### Reduced latency and improved scalability

- Local pipe control of load/store operations
  - Improved hazard avoidance
  - Local recycles – reduced hazard disruption
  - Improved lock management





## SMT4 Core Resources

### Fetch / Branch

- 32kB, 8-way Instruction Cache
- 8 fetch, 6 decode
- 1x branch execution

### Slices issue VSU and AGEN

- 4x scalar-64b / 2x vector-128b
- 4x load/store AGEN

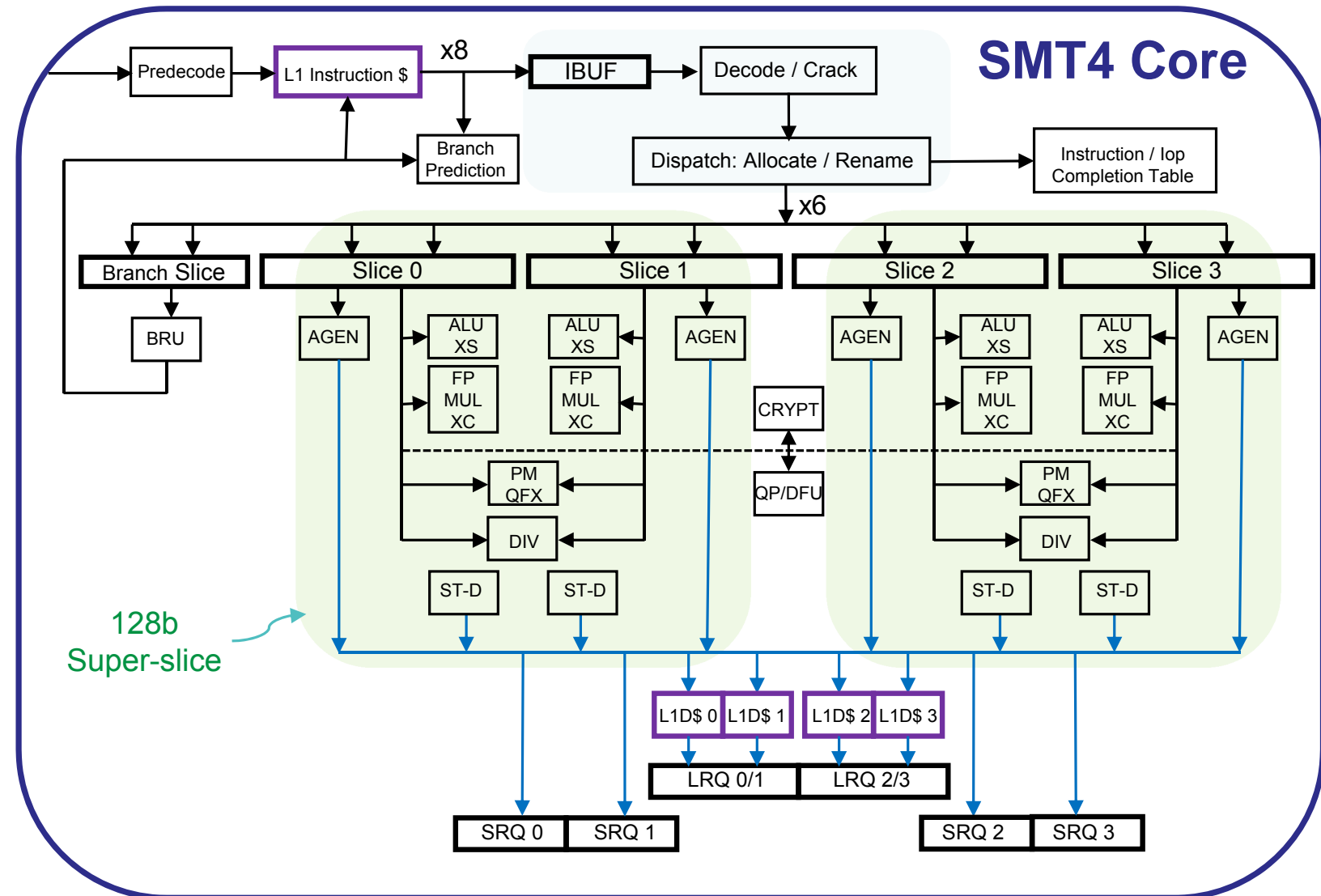
### Vector Scalar Unit (VSU) Pipes

- 4x ALU + Simple (64b)
- 4x FP + FX-MUL + Complex (64b)
- 2x Permute (128b)
- 2x Quad Fixed (128b)
- 2x Fixed Divide (64b)
- 1x Quad FP & Decimal FP
- 1x Cryptography

### Load Store Unit (LSU) Slices

- 32kB, 8-way Data Cache
- Up to 4 DW load or store

## Symmetric Engines Per Data-Type for Higher Performance on Diverse Workloads



**Efficient Cores Deliver 2x Compute Resource per Socket**

## New Instruction Set Architecture Implemented on POWER9

### Broader data type support

- 128-bit IEEE 754 Quad-Precision Float – Full width quad-precision for financial and security applications
- Expanded BCD and 128b Decimal Integer – For database and native analytics
- Half-Precision Float Conversion – Optimized for accelerator bandwidth and data exchange

### Support Emerging Algorithms

- Enhanced Arithmetic and SIMD
- Random Number Generation Instruction

### Accelerate Emerging Workloads

- Memory Atomics – For high scale data-centric applications
- Hardware Assisted Garbage Collection – Optimize response time of interpretive languages

### Cloud Optimization

- Enhanced Translation Architecture – Optimized for Linux
- New Interrupt Architecture – Automated partition routing for extreme virtualization
- Enhanced Accelerator Virtualization
- Hardware Enforced Trusted Execution

### Energy & Frequency Management

- POWER9 Workload Optimized Frequency – Manage energy between threads and cores with reduced wakeup latency



## Big Caches for Massively Parallel Compute and Heterogeneous Interaction

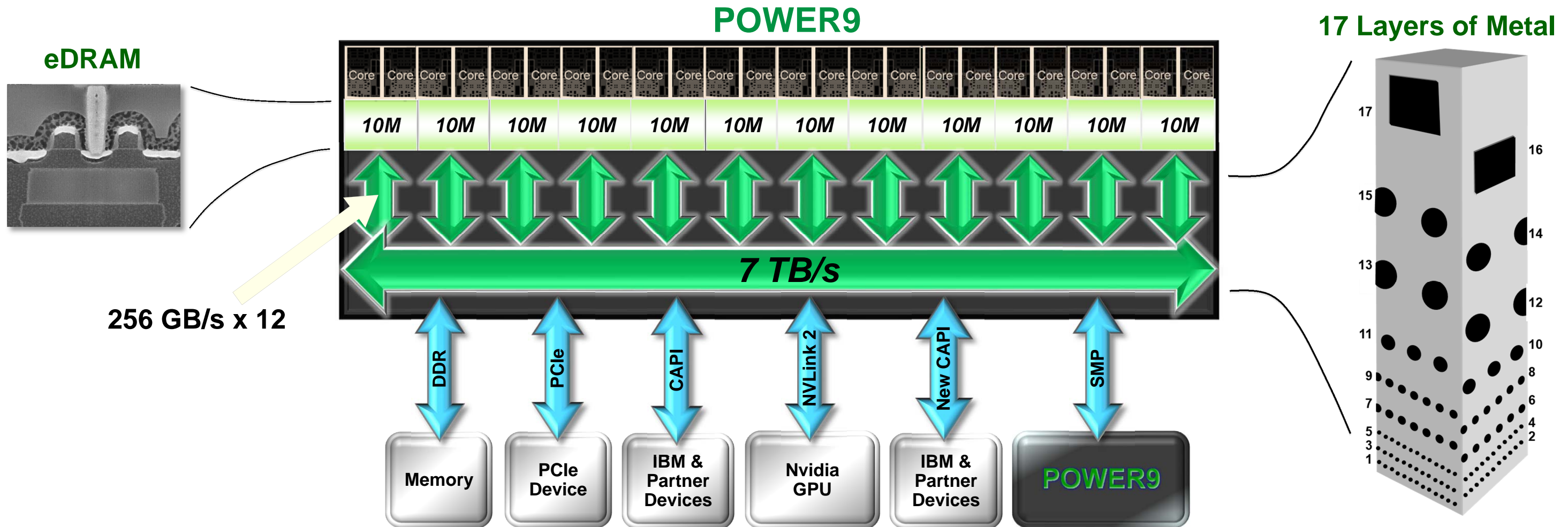
### L3 Cache: 120 MB Shared Capacity NUCA Cache

- 10 MB Capacity + 512k L2 per SMT8 Core
- Enhanced Replacement with Reuse & Data-Type Awareness  
12 x 20 way associativity

## Extreme Switching Bandwidth for the Most Demanding Compute and Accelerated Workloads

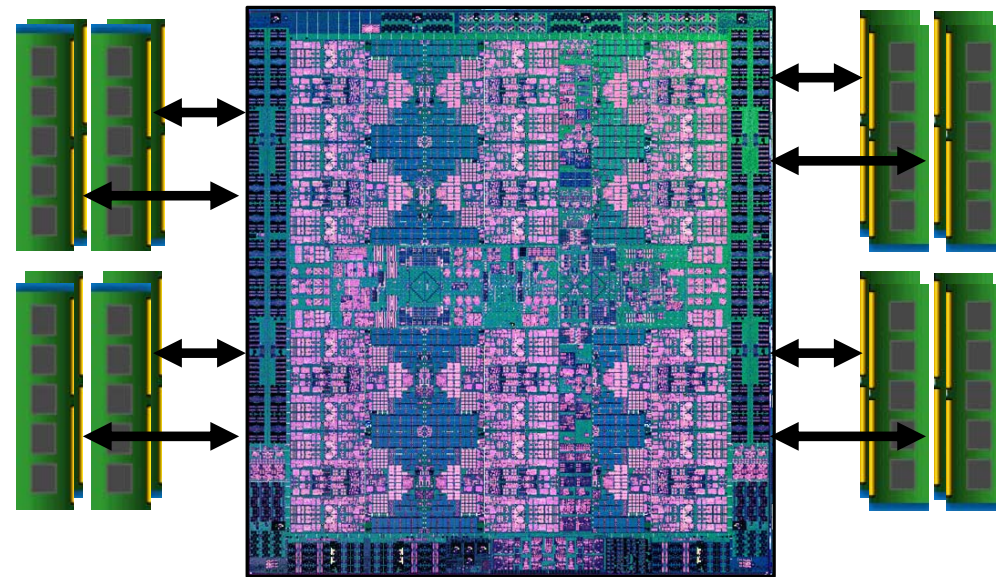
### High-Throughput On-Chip Fabric

- Over 7 TB/s On-chip Switch
- Move Data in/out at 256 GB/s per SMT8 Core





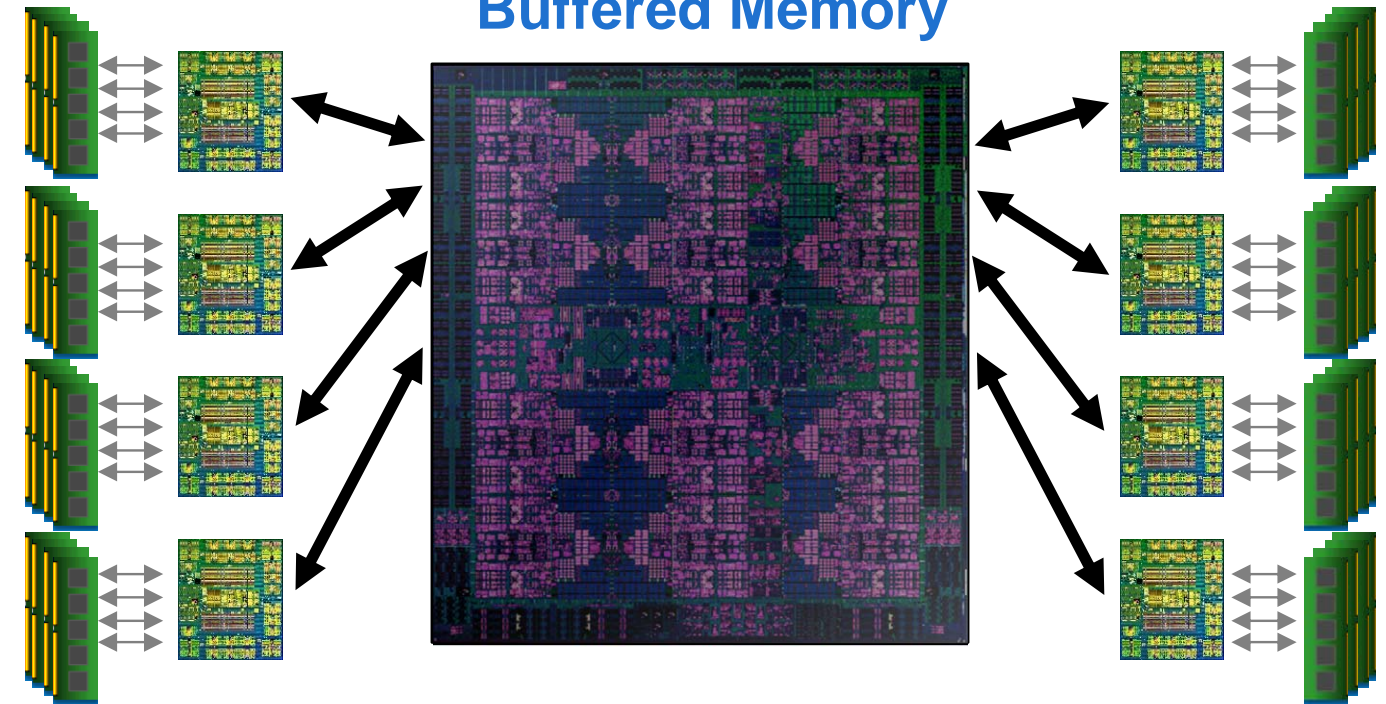
## Scale Out Direct Attach Memory



### 8 Direct DDR4 Ports

- Up to 120 GB/s of sustained bandwidth
- Low latency access
- Commodity packaging form factor
- Adaptive 64B / 128B reads

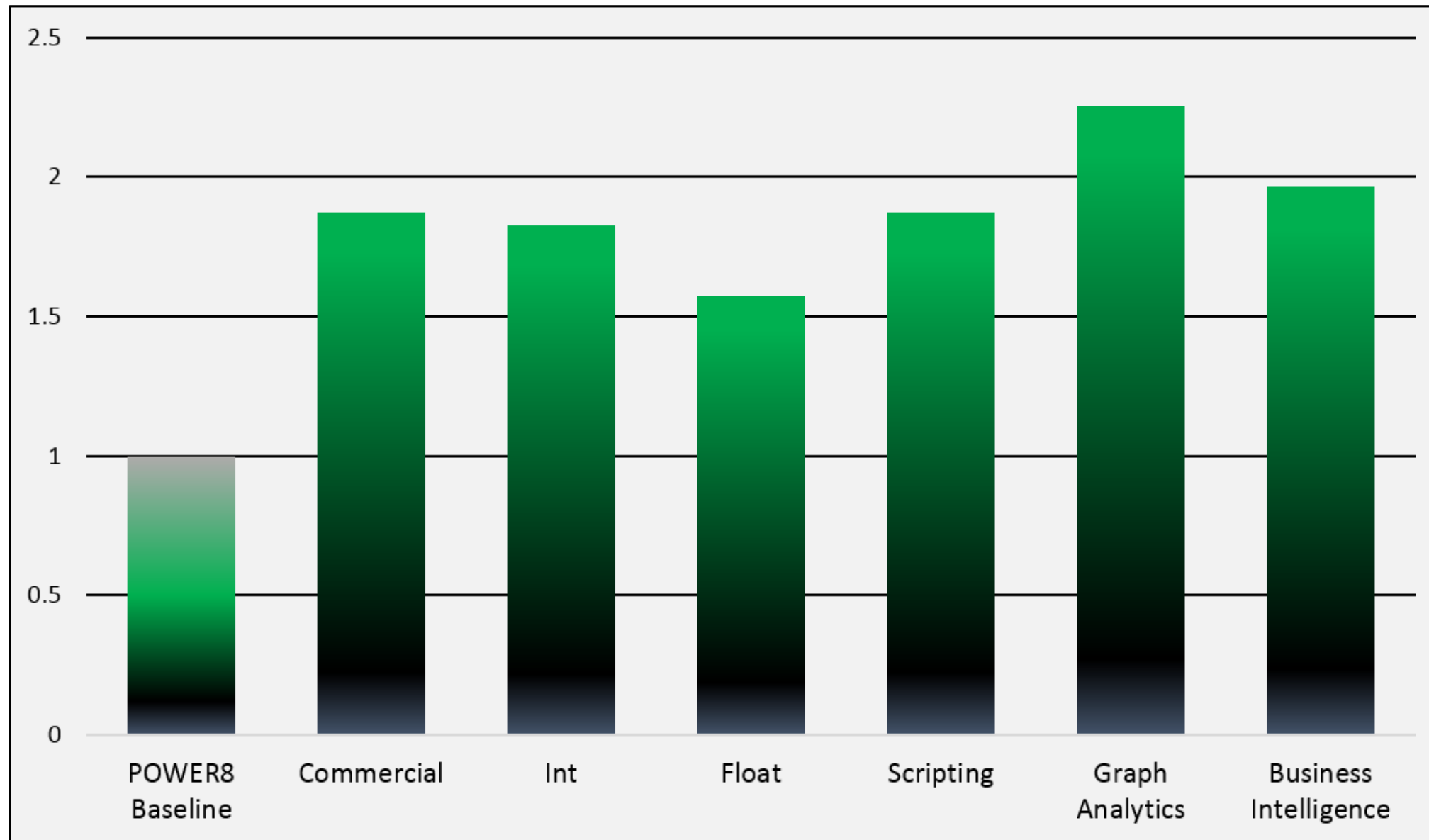
## Scale Up Buffered Memory



### 8 Buffered Channels

- Up to 230GB/s of sustained bandwidth
- Extreme capacity – up to 8TB / socket
- Superior RAS with chip kill and lane sparing
- Compatible with POWER8 system memory
- Agnostic interface for alternate memory innovations

## Socket Performance



Scale-Out configuration @ constant frequency

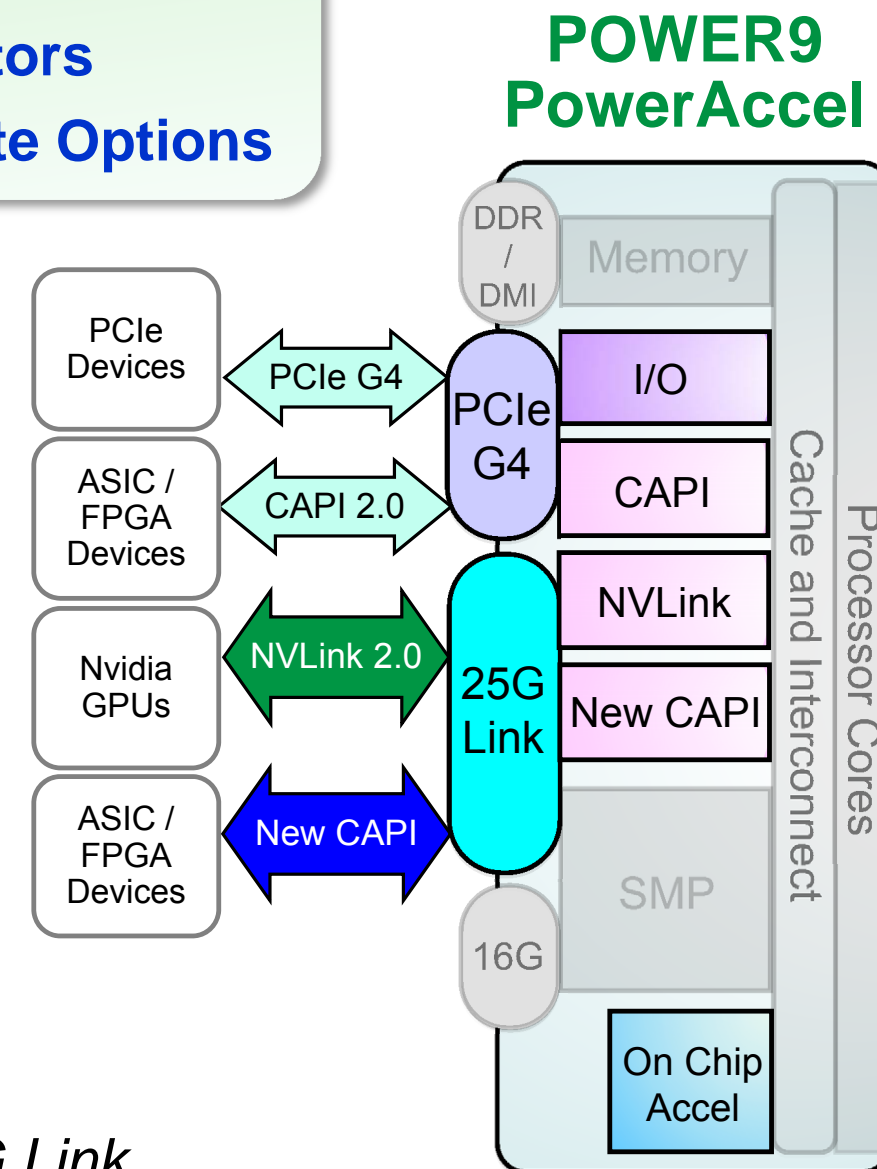
- Extreme Processor / Accelerator Bandwidth and Reduced Latency
- Coherent Memory and Virtual Addressing Capability for all Accelerators
- OpenPOWER Community Enablement – Robust Accelerated Compute Options

- State of the Art I/O and Acceleration Attachment Signaling

- PCIe Gen 4 x 48 lanes – 192 GB/s duplex bandwidth
- 25G Link x 48 lanes – 300 GB/s duplex bandwidth

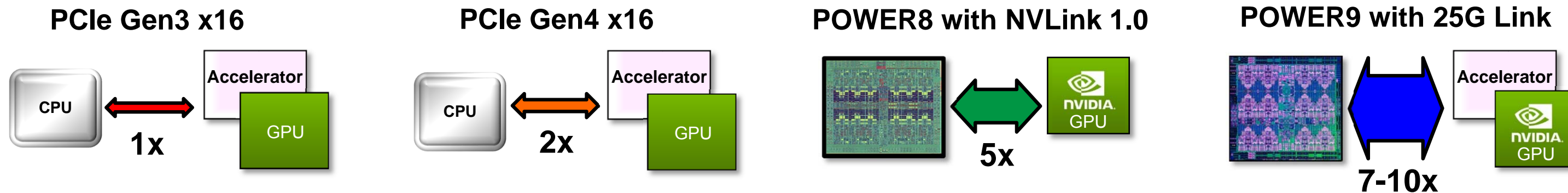
- Robust Accelerated Compute Options with OPEN standards

- On-Chip Acceleration – Gzip x1, 842 Compression x2, AES/SHA x2
- CAPI 2.0 – 4x bandwidth of POWER8 using *PCIe Gen 4*
- NVLink 2.0 – Next generation of GPU/CPU bandwidth and integration
- New CAPI – High bandwidth, low latency and open interface using *25G Link*





## Extreme CPU/Accelerator Bandwidth



*Increased Performance / Features / Acceleration Opportunity*

### Seamless CPU/Accelerator Interaction

- Coherent memory sharing
- Enhanced virtual address translation
- Data interaction with reduced SW & HW overhead

### Broader Application of Heterogeneous Compute

- Designed for efficient programming models
- Accelerate complex analytic / cognitive applications

## OpenPOWER™ Foundation

- Accelerating Open Innovation
- Grown from 5 to over 200 members in less than 3 years

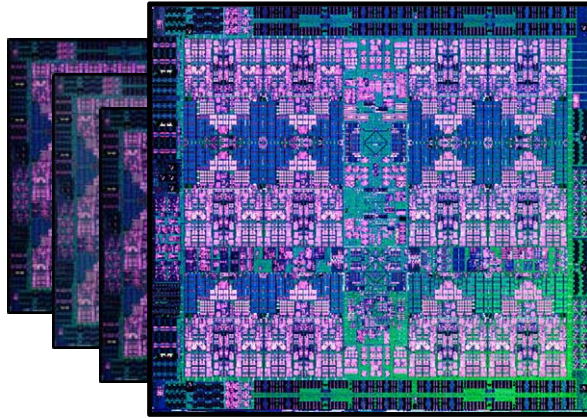
## POWER9: Engineered for OpenPOWER Application

- Built for a Broad Range of Deployments and Platforms
- Open and Flexible Solutions
- Ideal for Developers



The image displays a comprehensive grid of logos for the OpenPOWER ecosystem, organized into six horizontal categories:

- Implementation / HPC / Research:** Includes logos for ASU, ASTRI, Agency for Science, Technology and Research, CAMBRIDGE CARES, Carnegie Mellon University, cfms, CINECA, Hartree Centre, mtb, icm, JÜLICH, Lawrence Livermore National Laboratory, LSU, NUS, NANYANG TECHNOLOGICAL UNIVERSITY, OAK RIDGE National Laboratory, OSU, OVH.COM, RACKSPACE, SANDIA NATIONAL LABORATORIES, SDSC, SYMBIOSIS INSTITUTE OF COMPUTER STUDIES & RESEARCH, TACC, RICE, rzg, SASTRA UNIVERSITY, USC, UNIVERSITY OF ARKANSAS, UNIVERSITY OF OREGON, UF FLORIDA, and MAISON DE LA SIMULATION.
- Software:** Includes American Megatrends, BYOSOFT, FIXSTARS, FreeBSD, Google, GPUDB, 红旗 Linux, REDIS, SYNERSCOPE, T2i, and ubuntu.
- System / Integration:** Includes ARROW ELECTRONICS, INC., AVNET, BULL, CIARA, Cirrascale, CSPI, DRC, E4, IBM, inspur, Mark III Systems, Microway, NEC, OCF, ONE STOP SYSTEMS, PENGUIN COMPUTING, RAPTOR ENGINEERING, rikor, WRTDS Technologies, STACK VELOCITY, TEXHOI POM, UNISOURCE, and YA DRO.
- I/O / Storage / Acceleration:** Includes ALGO-LOGIC, ALPHA DATA, AVAGO TECHNOLOGIES, BittWare, blaeograph, BLUEBEE, BROCADE, Chelsio Communications, CONVEY COMPUTES, DataDirect NETWORKS, edico, genome, EVERSPIN TECHNOLOGIES, FUSION, HGST, HITACHI, Inphi, Interface Masters, MAXELLER Technologies, Mellanox TECHNOLOGIES, MEMBLAZE, MICRON, Microsemi, MYRICOM Network Products, Nallatech, and NVIDIA.
- Boards / Systems:** Includes Celestica, 创和通讯, IBM, Inventec, MSI, 柏飞电子, TYAN, and wistron.
- Chip / SOC:** Includes IBM, IDT, infineon, POWERCORE, SYNAPSIS design, and VeriSilicon.



**Built for the Cognitive Era**



## Enhanced Core and Chip Architecture for Emerging Workloads

- New Core Optimized for Emerging Algorithms to Interpret and Reason
- Bandwidth, Scale, and Capacity, to Ingest and Analyze

## Processor Family with Scale-Out and Scale-Up Optimized Silicon

- Enabling a Range of Platform Optimizations – from HSDC Clusters to Enterprise Class Systems
- Extreme Virtualization Capabilities for the Cloud

## Premier Acceleration Platform

- Heterogeneous Compute Options to Enable New Application Paradigms
- State of the Art I/O
- Engineered to be Open



## 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 inquiries, 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.

Revised September 26, 2006

## Special notices (continued)

IBM, the IBM logo, ibm.com AIX, AIX (logo), IBM Watson, DB2 Universal Database, POWER, PowerLinux, PowerVM, PowerVM (logo), PowerHA, Power Architecture, Power Family, POWER Hypervisor, Power Systems, Power Systems (logo), POWER2, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, POWER7, POWER7+, and POWER8 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>.

NVIDIA, the NVIDIA logo, and NVLink are trademarks or registered trademarks of NVIDIA Corporation in the United States 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 world-wide basis.

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.

The OpenPOWER word mark and the OpenPOWER Logo mark, and related marks, are trademarks and service marks licensed by OpenPOWER.

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