Every 100-150 Years We Hit an Accelerated Innovation Wave Which Drives Innovation

"If the Internet were a movie we'd still be in the opening credits"
– Gord Graylish

Internet Users in the World

3Bn
2Bn
1Bn

1951
First Commercial Computer (Ferranti Mark 1)

1959
Integrated Circuit is patented (Noyce/Kilby)

1969
ARPANET (internet forerunner)

1971
First microprocessor (Intel 4004)

1983
First IBM PC compatible laptops

1991
Tim Berners Lee publishes World Wide Web

Google
1997
Google.com registered

2003
Intel Centrino. WiFi Hot spots. Broadband

2004
Facebook launched

2007
iPhone launched

2010
iPad launched, other Android tablets follow

2012
Embedded Intelligence in WTC

2014 Netflix Google
Machine Learning

1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015

RISE OF THE CONNECTED, "AS A SERVICE" WORLD DISRUPTS ENTIRE INDUSTRIES ...YET, 60% OF THE WORLDS POPULATION ARE STILL UNCONNECTED

Today's Trading Landscape

Latency still matters…. A lot…

- Lower is still better, but ROI is smaller than before
- Can't be slower than competitors as the bar has been raised
- There's a hierarchy of needs either nanos or micros
- There's tension between algorithm sophistication and speed from a both a computational and programming perspective

Tech Trends to Consider in the Wake of the Wave

Moore's Law

Cloud

Heterogeneity

New materials

Are We There Yet?

"Are attempts to reduce...
latency becoming a case of ever
diminishing returns and ever
increasing investment?"

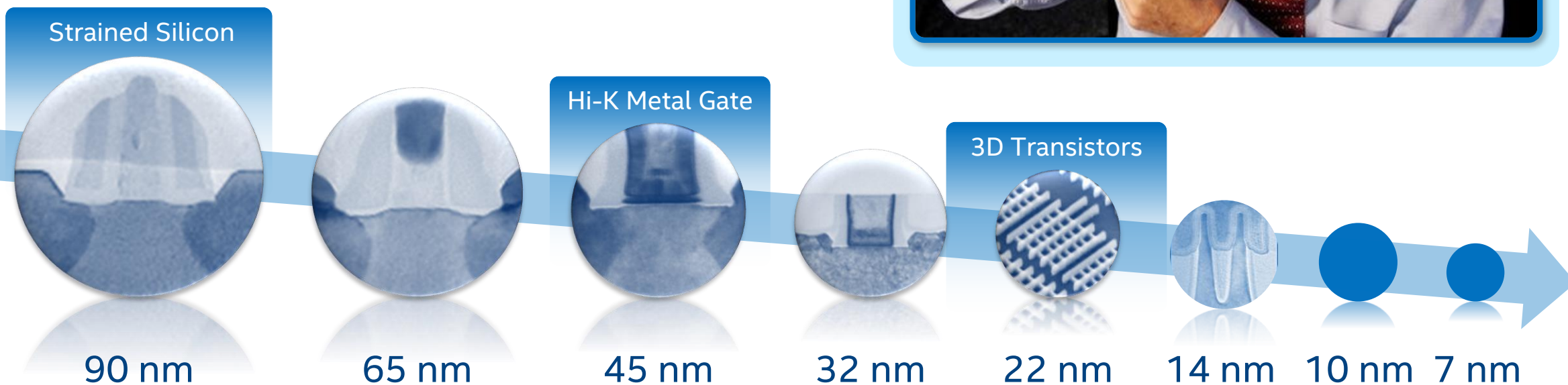- *Automated Trader April 5,*

Are We There Yet?

"Are attempts to reduce…
latency becoming a case of ever
diminishing returns and ever
increasing investment?"

- *Automated Trader April 5, 2007*

Tech Trend #1: Moore's Law

The key enabler of:
- *New Levels of Performance & Integration*
- *Higher functionality and complexity*
- *Control over power, cost, and size*

Strained Silicon

Hi-K Metal Gate

3D Transistors

90 nm   65 nm   45 nm   32 nm   22 nm   14 nm   10 nm   7 nm

(intel)

Recently Released Intel® Xeon® Processor E5-2600 v4

Broadwell microarchitecture, Built on 14nm process technology

➤ More Cores, More LL Cache, Larger Registers, Faster Memory

➤ Integrated: Memory Controller, I/O Controller, Voltage Regulator

➤ Enhanced ISA, AES-NI, Realtime Processor Trace

Feature	Xeon E5-2600 v3 (Haswell-EP)	Xeon E5-2600 v4 (Broadwell-EP)
Cores Per Socket	Up to 18	Up to 22
Threads Per Socket	Up to 36 threads	Up to 44 threads
Last-level Cache (LLC)	Up to 45 MB	Up to 55 MB
QPI Speed (GT/s)	2x QPI 1.1 channels 6.4, 8.0, 9.6 GT/s	
PCIe* Lanes/ Controllers/Speed(GT/s)	40 / 10 / PCIe* 3.0 (2.5, 5, 8 GT/s)	
Memory Population	4 channels of up to 3 RDIMMs or 3 LRDIMMs	+ 3DS LRDIMM&
Max Memory Speed	Up to 2133	Up to 2400
TDP (W)	160 (Workstation only), 145, 135, 120, 105, 90, 85, 65, 55	

4 Channels DDR4

DDR4

DDR4

DDR4

DDR4

BROADWELL-EP

Core Core

Core Core

Core Core

Shared Cache

2 QPI 1.1

QPI

QPI

DMI2

40 Lanes PCIe* 3.0

Requires BIOS and firmware update
& Depends on market availability

All products, computer systems, dates and figures specified are preliminary based on current expectations, and are subject to change without notice. Intel may make changes to specifications and product descriptions at any time, without notice
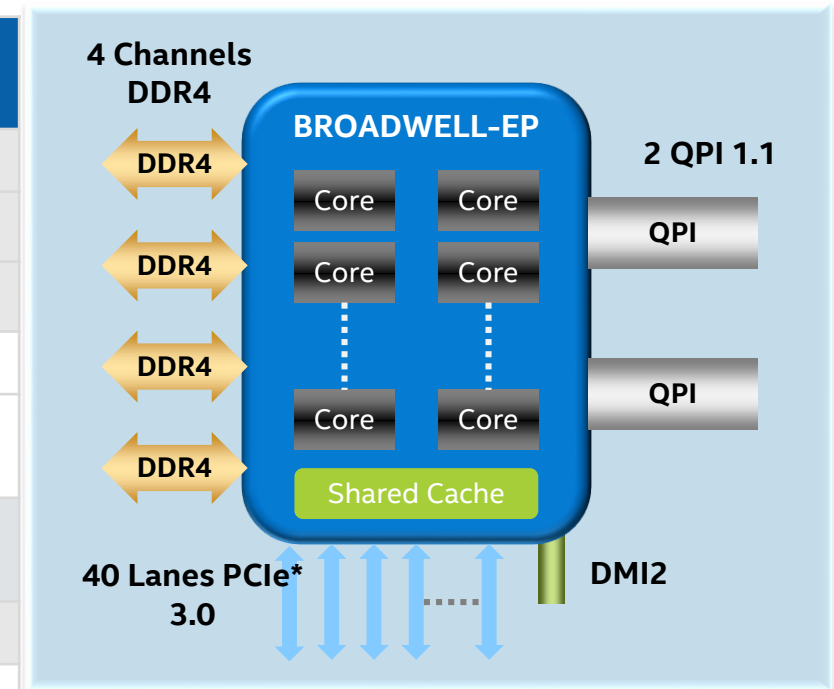
KNL Architecture Overview

ISA
Intel® Xeon® Processor Binary-Compatible (w/Broadwell)

On-package memory
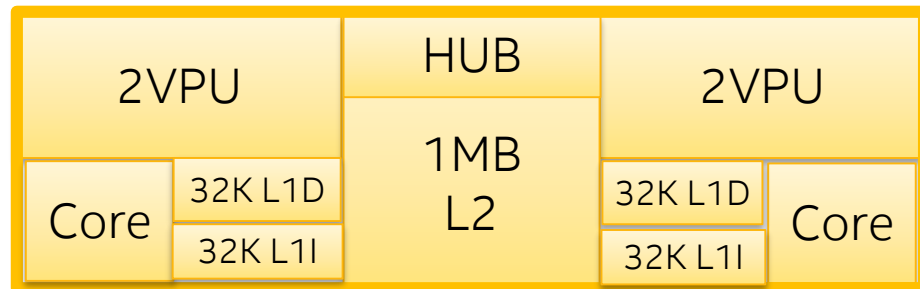Up to 16GB, ~460 GB/s STREAM at launch

Platform Memory
Up to 384GB (6ch DDR4-2400 MHz)
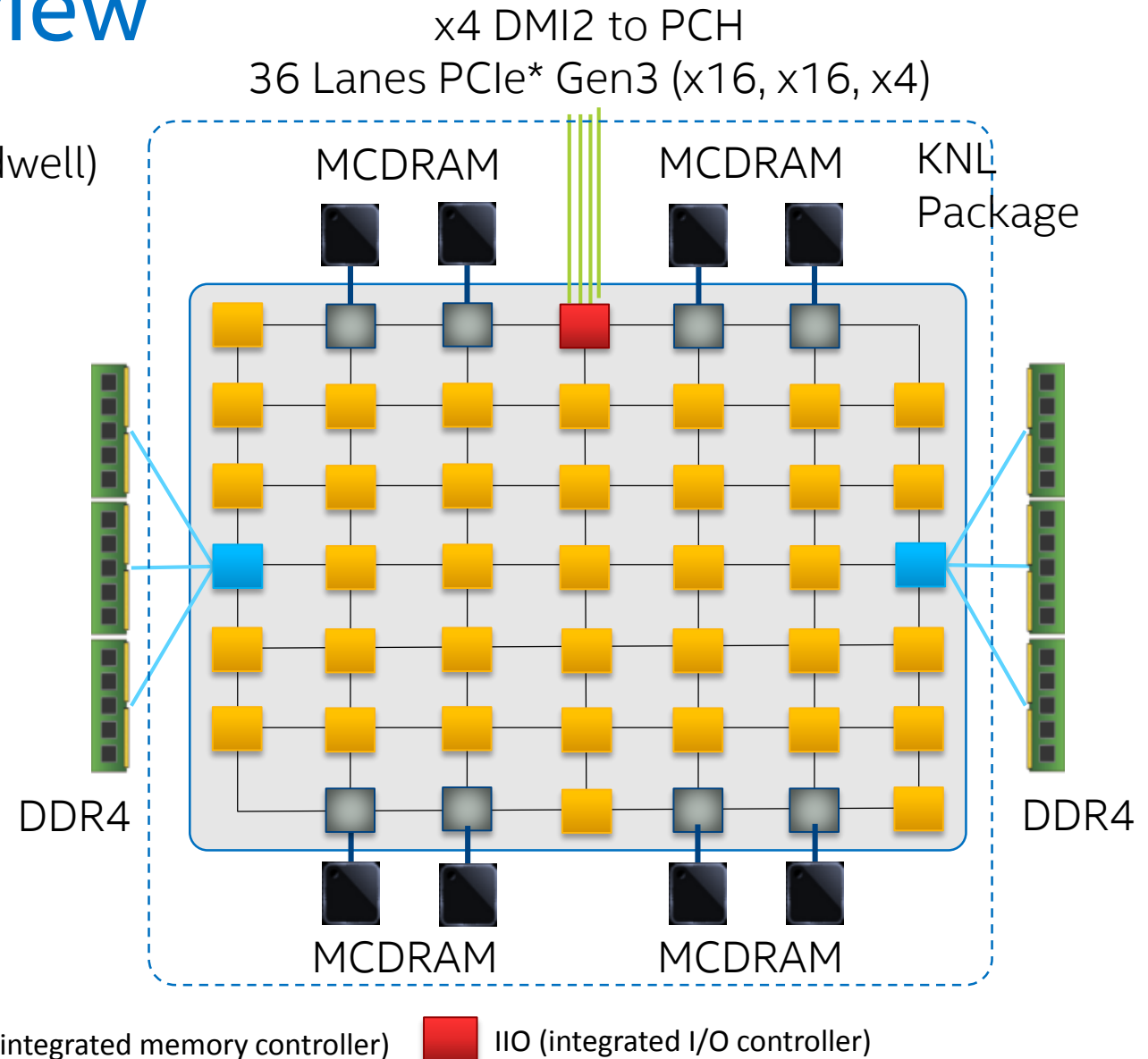
Fixed Bottlenecks
- ✓ 2D Mesh Architecture
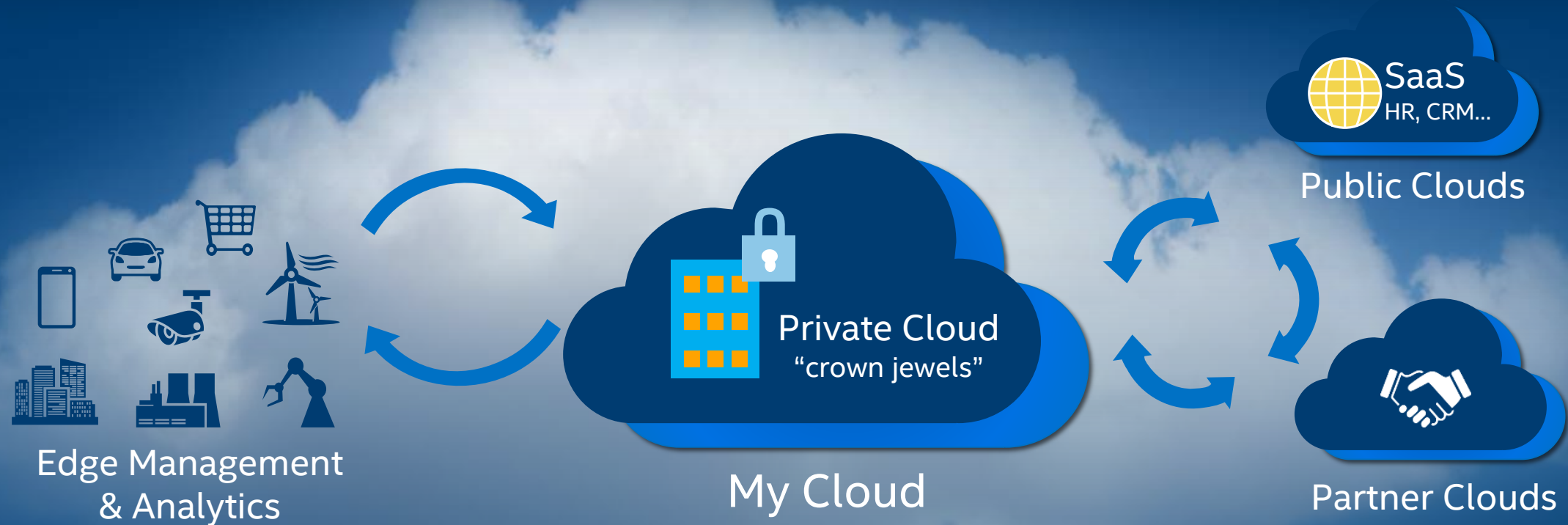- ✓ Out-of-Order Cores
- ✓ 3X single-thread vs. KNC

TILE:
(up to 36)

Enhanced Intel® Atom™ cores based on Silvermont Microarchitecture

x4 DMI2 to PCH
36 Lanes PCIe* Gen3 (x16, x16, x4)

🟧 Tile   ◼ EDC (embedded DRAM controller)   🟦 IMC (integrated memory controller)   🟥 IIO (integrated I/O controller)

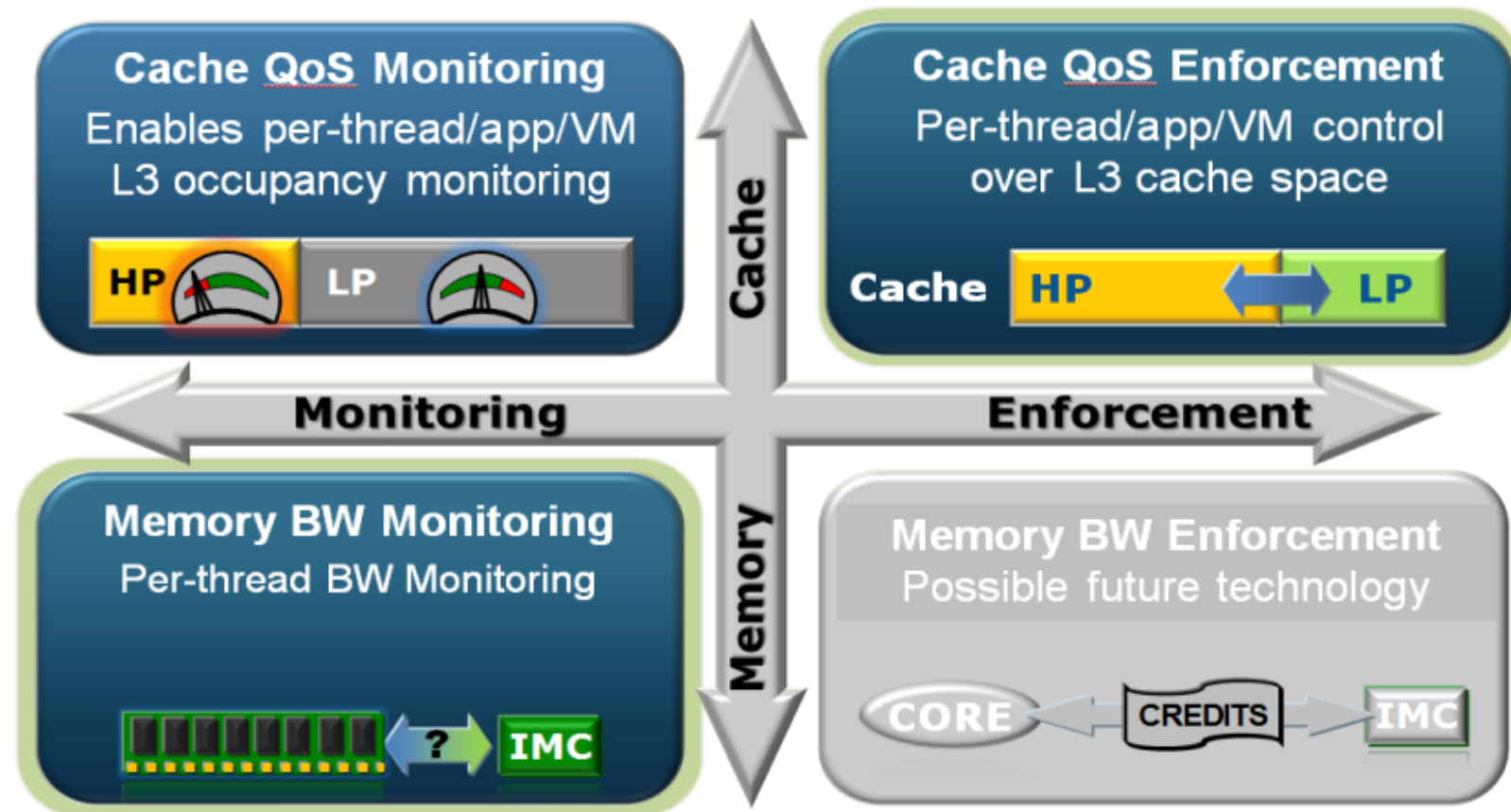# Tech Trend #2: The Journey to the Hybrid Cloud On-Demand

SaaS
HR, CRM...

Public Clouds

Private Cloud
"crown jewels"

My Cloud

Edge Management
& Analytics

Partner Clouds

HYBRID CLOUD – ENTERPRISES GETTING SERIOUS
50% OF ALL ENTERPRISES WILL HAVE HYBRID DEPLOYMENTS BY 2017 - GARTNER

Cache/Memory Quality of Service
Built for Virtualization and Cloud
Applicable for Trading

New Features for Broadwell Server

- Cache QoS Monitoring on L3
  - ❑ Double the RMIDs over Haswell Server for Class of Service Control over resources
- Cache QoS Enforcement on L3
- Memory BW Monitoring
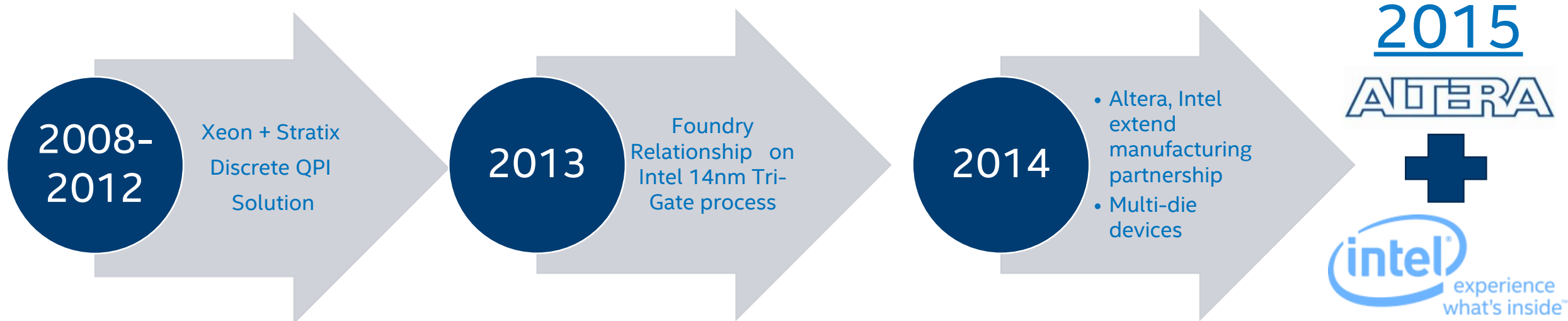 - ❑ Leverages same architecture as Cache QoS Monitoring

Using CQOS provides trading models another level of Control

Cache QoS Monitoring
Enables per-thread/app/VM L3 occupancy monitoring

HP   LP

Cache QoS Enforcement
Per-thread/app/VM control over L3 cache space

Cache   HP ⟷ LP

Monitoring   **Enforcement**

Cache

Memory

Memory BW Monitoring
Per-thread BW Monitoring

▯▯▯▯▯▯▯▯ ⟷ ? ⟷ IMC

Memory BW Enforcement
Possible future technology

CORE ⟷ CREDITS ⟷ IMC

"UP TO 1/3 OF CLOUD SERVICE PROVIDER NODES TO USE FPGAs BY 2020"

TECH TREND #3 HETEROGENEITY
ASSEMBLING THE RIGHT PARTS TO GET THE JOB DONE

2008-2012
Xeon + Stratix Discrete QPI Solution

2013
Foundry Relationship on Intel 14nm Tri-Gate process

2014
- Altera, Intel extend manufacturing partnership
- Multi-die devices

2015
ALTERA
+
intel
experience what's inside™

FGPA's have High Speed Connections to CPU's, Software CPU's and ARM CPU's. Integration with Xeon is a logical recognition of the value of FPGA's

INTEGRATION OF SILICON SOLUTIONS AND IP & TOOL-FLOW DRIVES PERFORMANCE

Manufacturing Is a Key Ingredient in FPGA

IP & tool-flow

Fully unified IP and development suite (languages, libraries, dev. environment)

Cross-platform dev. framework (e.g., OpenCL)

Independent development tools

Future

Enhanced productivity

Today

Optimized solutions for demanding workloads, e.g.,

Image ID   Security   Big data

Increased performance/watt

Discrete chips CPU & FPGA

Co-packaged CPU + FPGA

Integrated die CPU + FPGA

TECH TREND #4 NEW MEMORY ECOMONICS
Breaks the Memory Storage Barrier

Interfaces

Relative Latencies

| Processor |

On Core CPU | L1/2 Cache | ~1 ns

On Die | L3 Cache | ~10 ns

Direct Attach | Main Memory | ~100 ns

PCIe NVMe | NVMe | ~25,000 ns (25 us)

SAS, SATA | NAND SSD | ~100,000 ns (100 us)

SAS, SATA | Fast HDD | ~10,000,000 ns (10 ms)

Performance

Costs
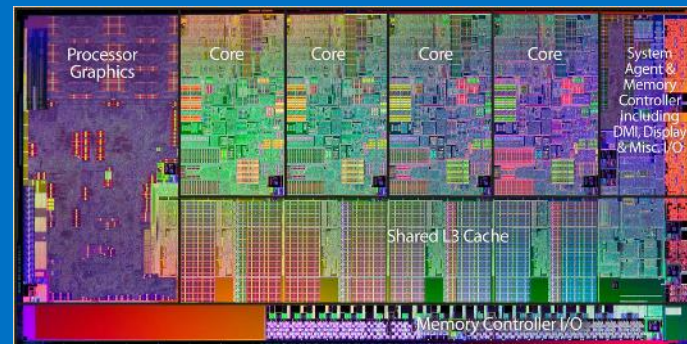
Latency is everywhere... in data, in calculations, in storage while the time to act is shrinking

Computer Architecture Reoptimization

Relative Bandwidth Improvement

Relative Latency Improvement

Microprocessor
Memory
Network
Storage
NAND

Computer Architecture Reoptimization

Computer Architecture Reoptimization
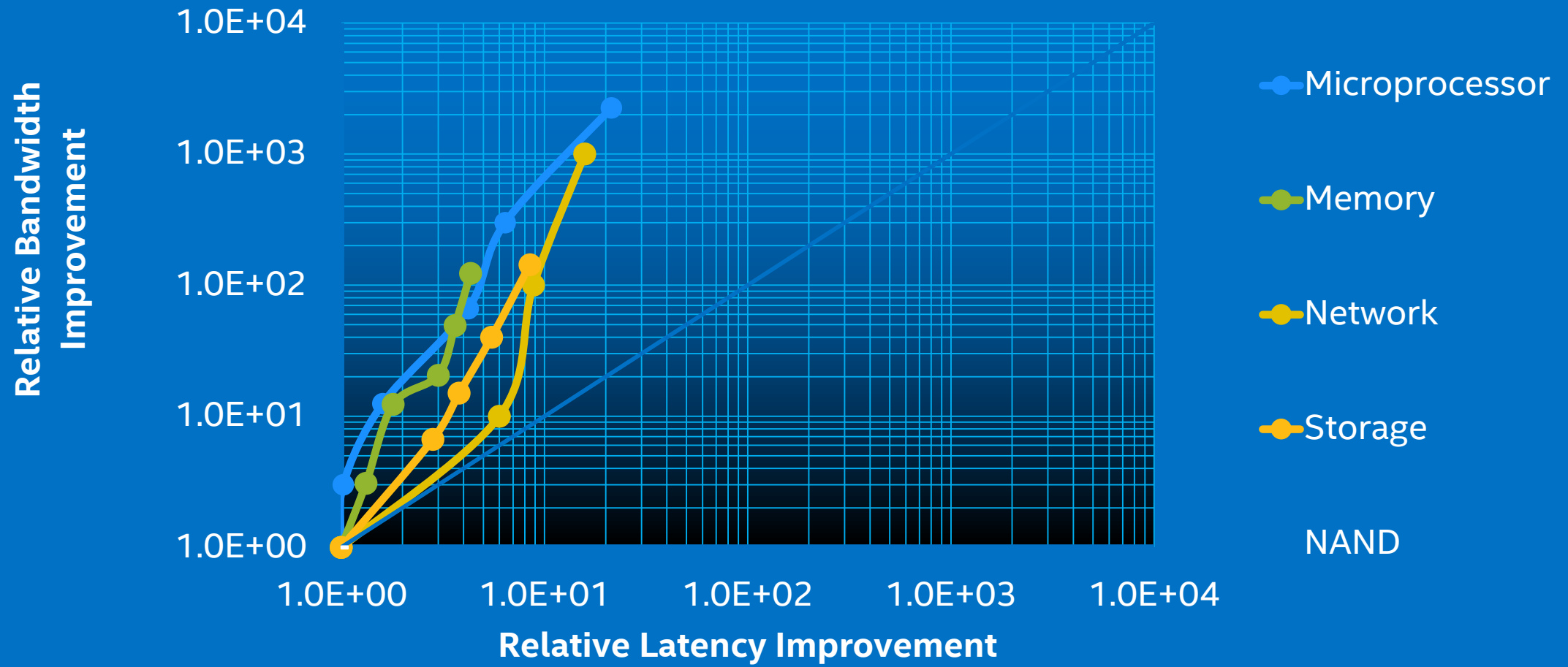
Source:
Intel 2011 & David Patterson "Latency Lags Bandwidth"
COMMUNICATIONS OF THE ACM October 2004/Vol. 47, No. 10
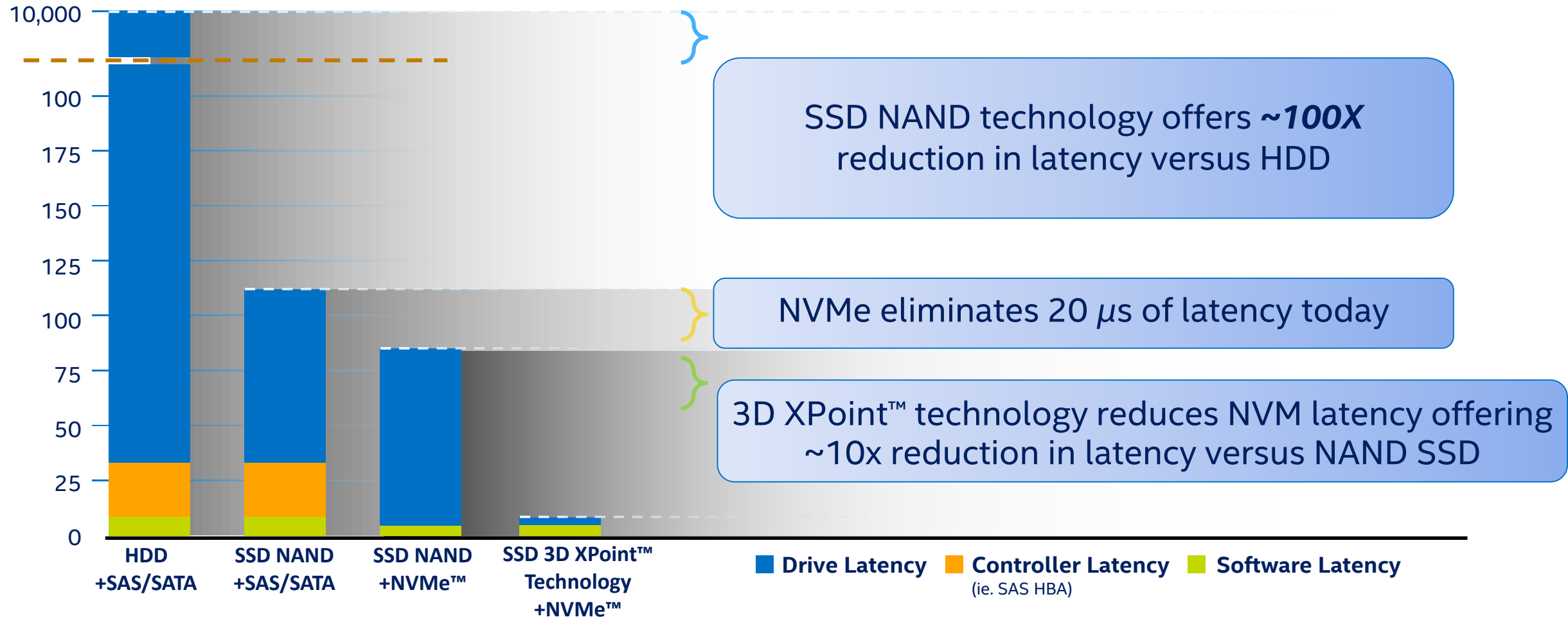
Kx Systems*

kdb+ 3.1* running the STAC-M3* workload (High-speed tick analysis)

"For the longest time, we were held back by slow discs and minimal amounts of memory . Kx customers require performance. As memory configurations have increased to six terabytes or larger, game-changing data strategies are becoming possible and changing how our customers do business when they can load their data into memory."[1]

Simon Garland **– Chief Strategist, Kx Systems**

- The STAC-M3* benchmark characterizes analysis of time-series data such as tick-by-tick quote and trade histories, which are crucial to many trading functions, from algorithm development to risk management. The key metric in STAC-M3* is responses times.
- SK Hynix* DDR4 memory, 60% more cores along with Intel® AVX2 instructions enabled significant growth over the past two generations in analytic capabilities using the same Brickland platform.

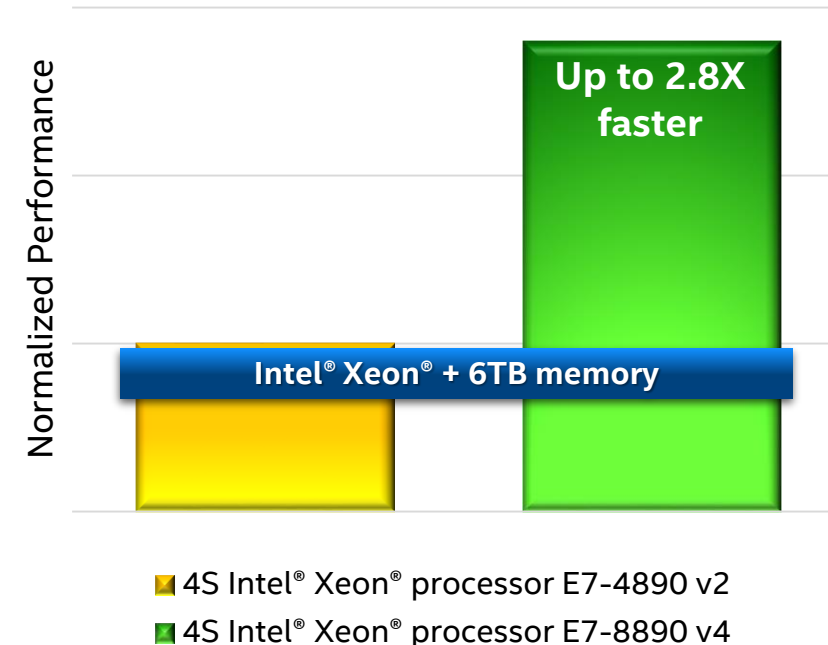**Improve your financial portfolio results by running more theoretical profit and loss analyses a day by up to 2.8x**
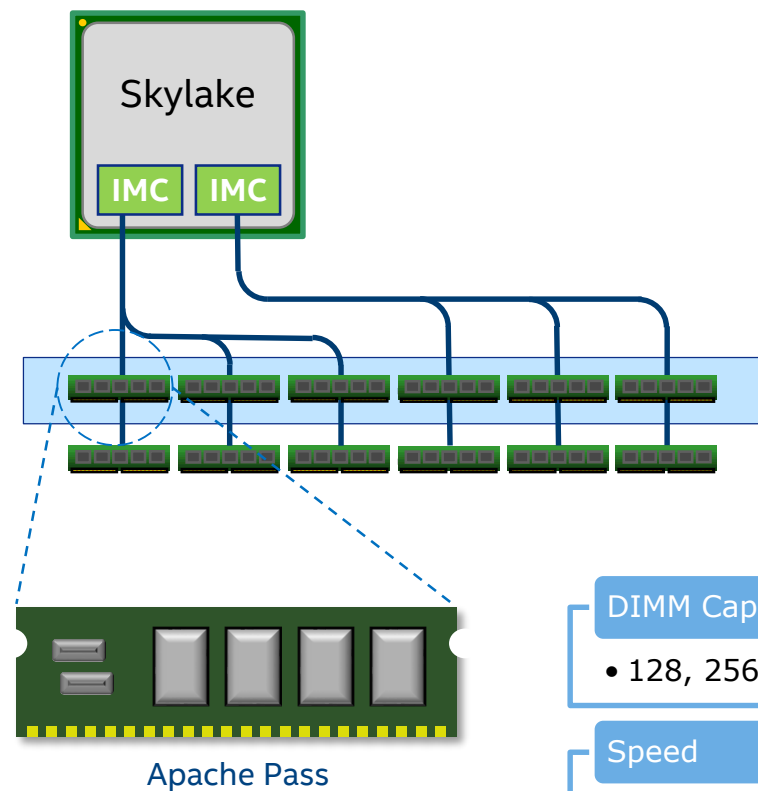
www.kx.com
www.STACresearch.com/kx

SK hynix

Financial Services

Improve financial portfolio analysis with the 4-Socket Servers using the Intel® Xeon® processor E7-8890 v4

Normalized Performance

Up to 2.8X faster

Intel® Xeon® + 6TB memory

4S Intel® Xeon® processor E7-4890 v2
4S Intel® Xeon® processor E7-8890 v4

Workload: STAC-M3* ß1.10T.THEOPL.TIME high-speed analytics on time series, tick-by-tick market data.

Intel NVM DIMM Technology

Skylake

IMC   IMC

Apache Pass

Apache Pass*

DDR4 DRAM*

- DDR4 electrical & physical
- Close to DRAM latency
- Cache line size access

DIMM Capacity
- 128, 256, 512GB

Speed
- 2400MT/sec

Platform Capacity
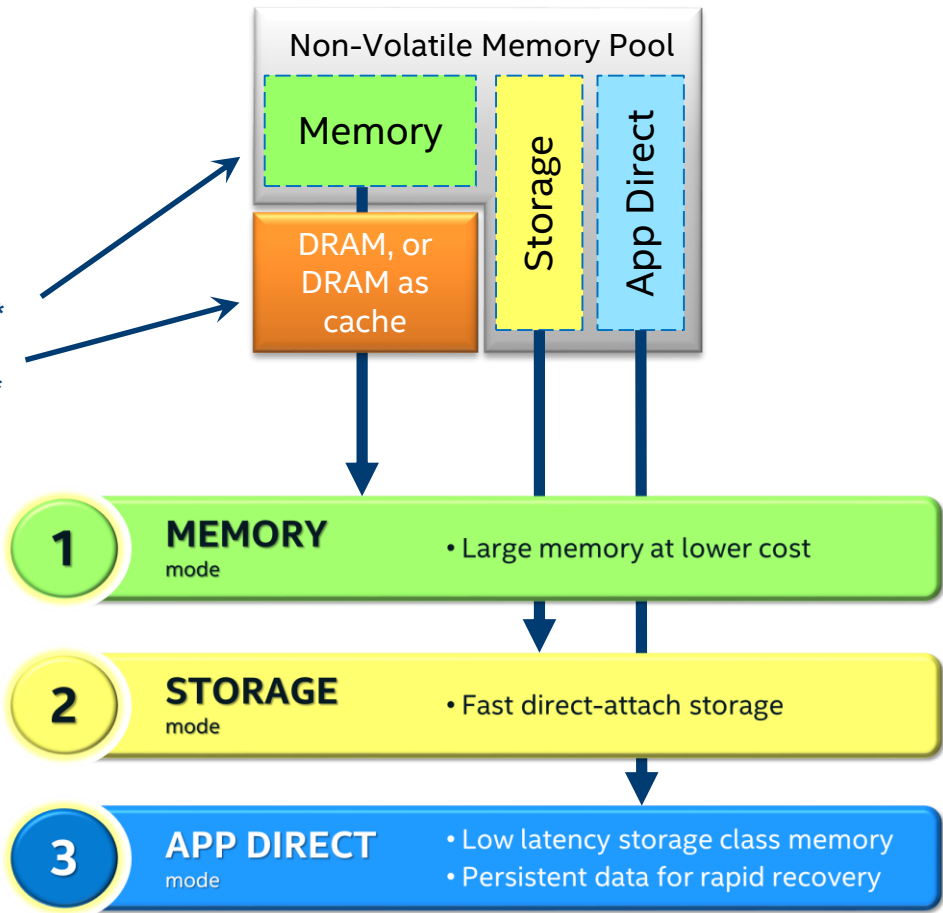- 6TB (3TB per CPU)

Flexible, Usage Specific Partitions

Non-Volatile Memory Pool

Memory

DRAM, or DRAM as cache

Storage

App Direct

1 MEMORY mode
- Large memory at lower cost

2 STORAGE mode
- Fast direct-attach storage

3 APP DIRECT mode
- Low latency storage class memory
- Persistent data for rapid recovery

* DIMM population shown as an example only.

(intel)

Conclusion

Trading will continue to benefit from technology trends driven by other industries

There will be first mover advantages for trading firms

- CPUs will provide not just more cores but also more control
- FPGAs will continue to get faster and more capable
- CPUs and FPGA will integrate
- New solid-state storage will shrink latencies

Trading firms will be able to compute ever larger problems

- CPU-FPGA integration, faster interconnects, faster solid state media will enable more sophisticated low-latency algorithms

What Can You Do?

- Be prepared to rethink your application architecture
- Experiment with these new technologies
- Use them to their fullest (e.g., leverage high-performance APIs)