The Path To Exascale – Challenges and Opportunities

EMIT Workshop, Manchester, 30th July 2015

Gaurav Kaul – Solutions Architect

John Swinburne – HPC Architect

Legal Notices and Disclaimers

Legal Disclaimer & Optimization Notice

Optimization Notice

Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804

A bit of History ....

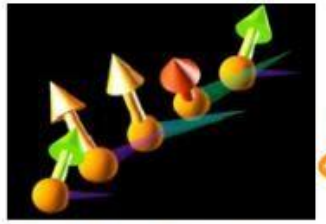
Computer performance and application performance increase ~10^3 every decade

~100 Kilowatts → ~5 Megawatts → 20-30 MW

~1 Exaflop/s

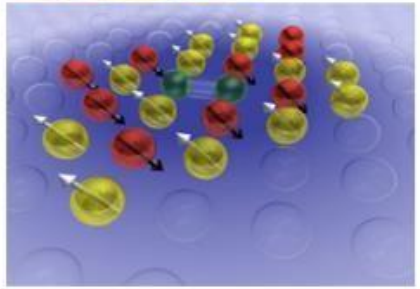
100 million or billion processing cores (!)

1 Gigaflop/s
Cray YMP
8 processors



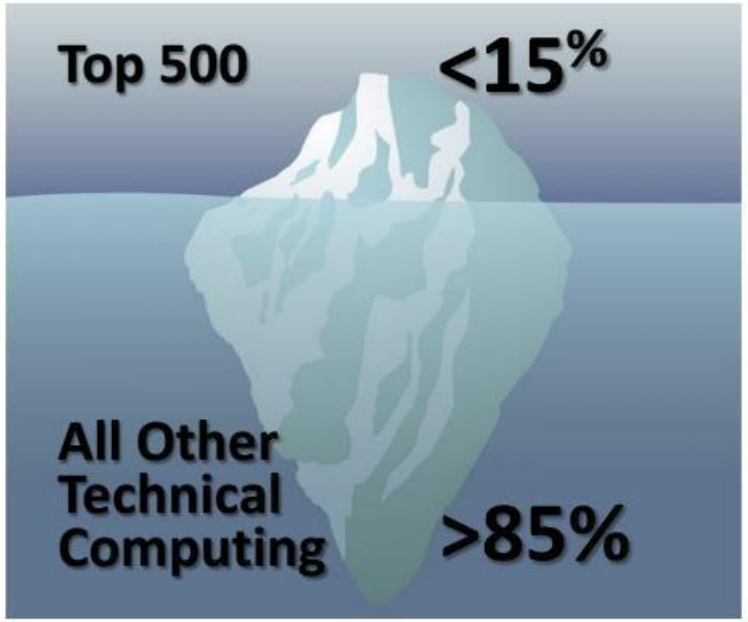
1.02 Teraflop/s
Cray T_{3E}
1'500 processors

1.35 Petaflop/s
Cray XT5
150'000 processors



1988	1998	2008	2018
First sustained GFlop/s Gordon Bell Prize 1988	First sustained TFlop/s Gordon Bell Prize 1998	First sustained PFlop/s Gordon Bell Prize 2008	Another 1,000x increase in sustained performance

The Top 500 Waterfall



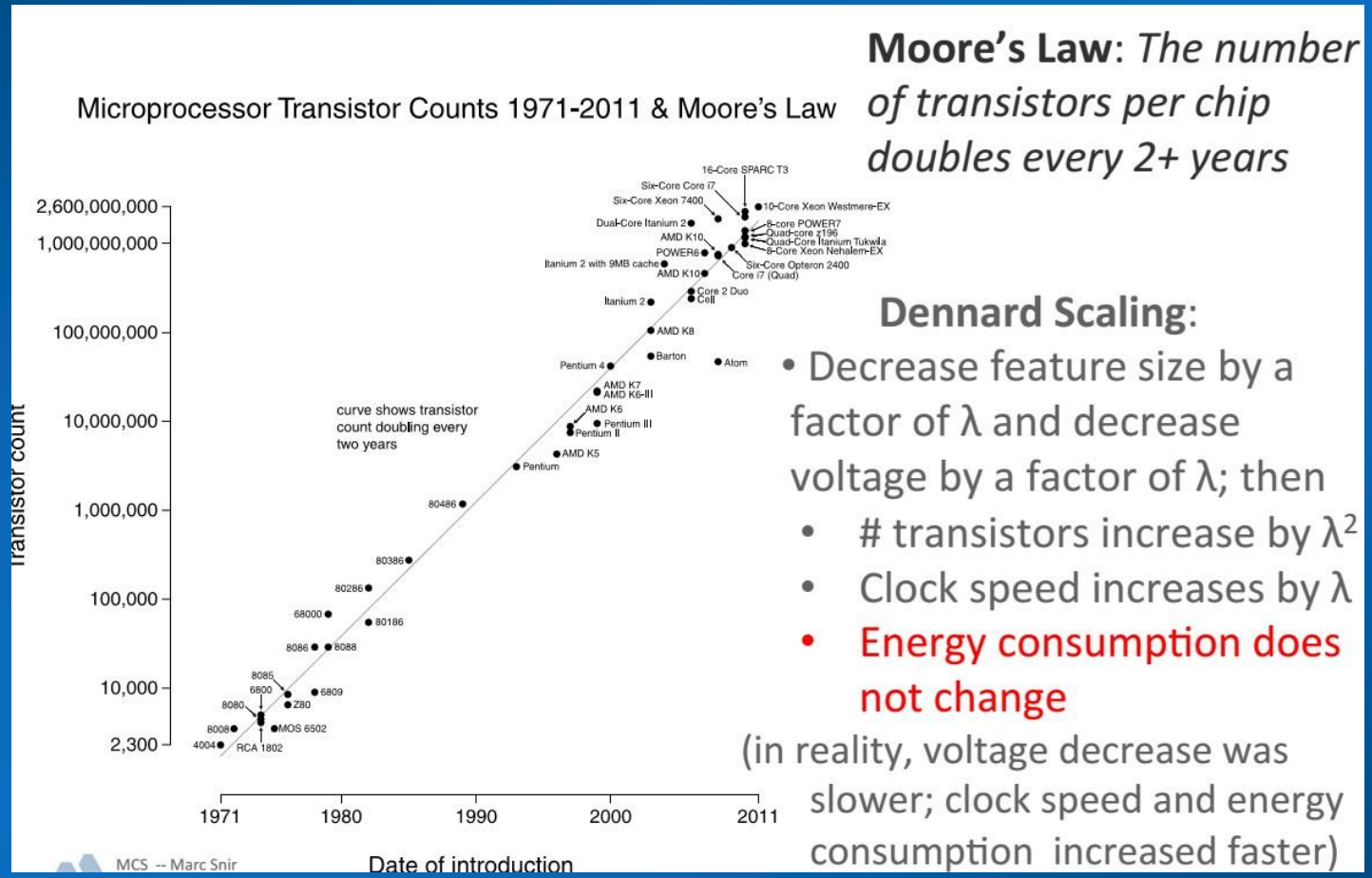
% of sockets sold

**plus.....similar waterfalls for other capabilities in areas like fabrics, storage, software, ...*

Source: Top500.org and Intel Estimate of Top500 sockets as % of sum of analysts reports of HPC and branded Workstations sockets. Performance waterfall timelines based on TOP500.org statistics (#1-#500) and Intel estimate (#500 to projected Intel Knights Landing)

50 years of Moore's Law

Moore and Dennard Scaling



Moore's Law: *The number of transistors per chip doubles every 2+ years*

Dennard Scaling:

- Decrease feature size by a factor of λ and decrease voltage by a factor of λ; then
  - # transistors increase by λ^2
  - Clock speed increases by λ
  - Energy consumption does not change

(in reality, voltage decrease was slower; clock speed and energy consumption increased faster)

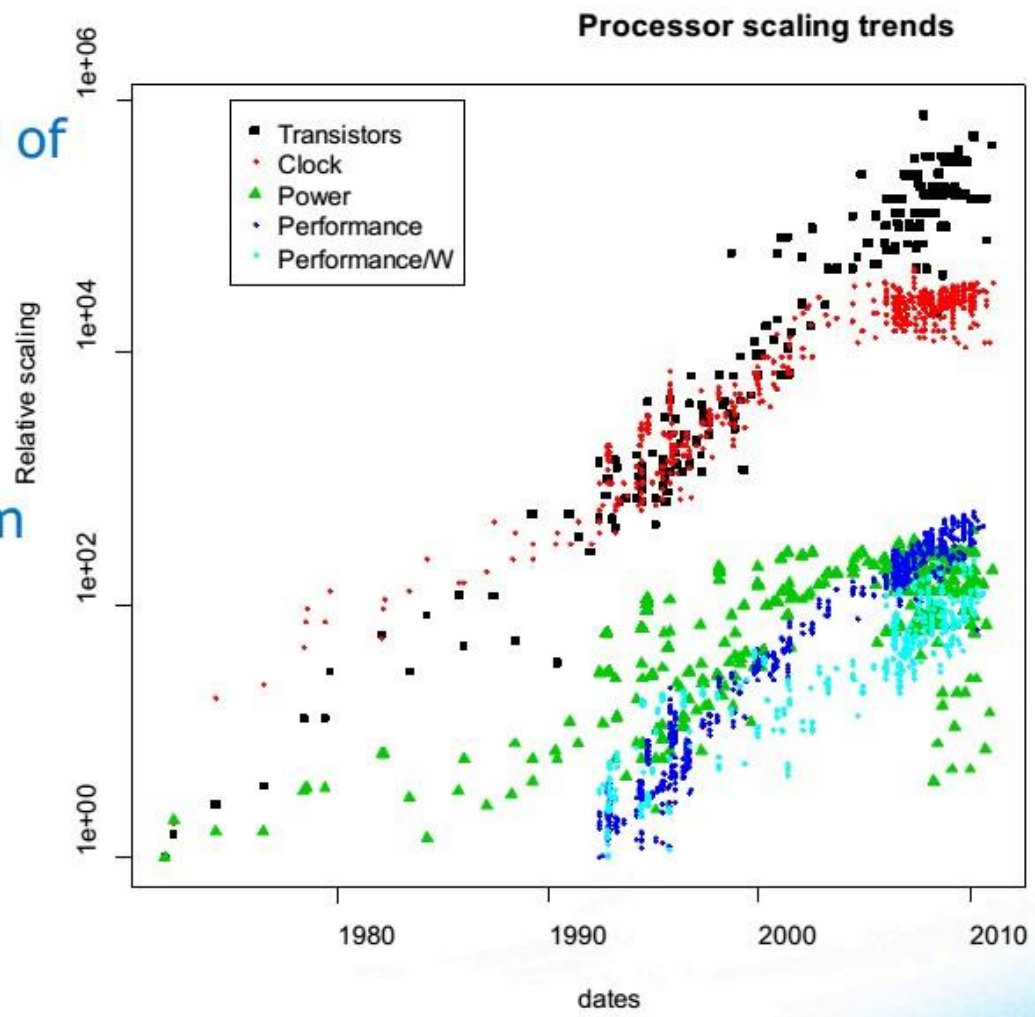
Current Processor Performance Trends

After ~2004 only the number of transistors continues to increase exponentially

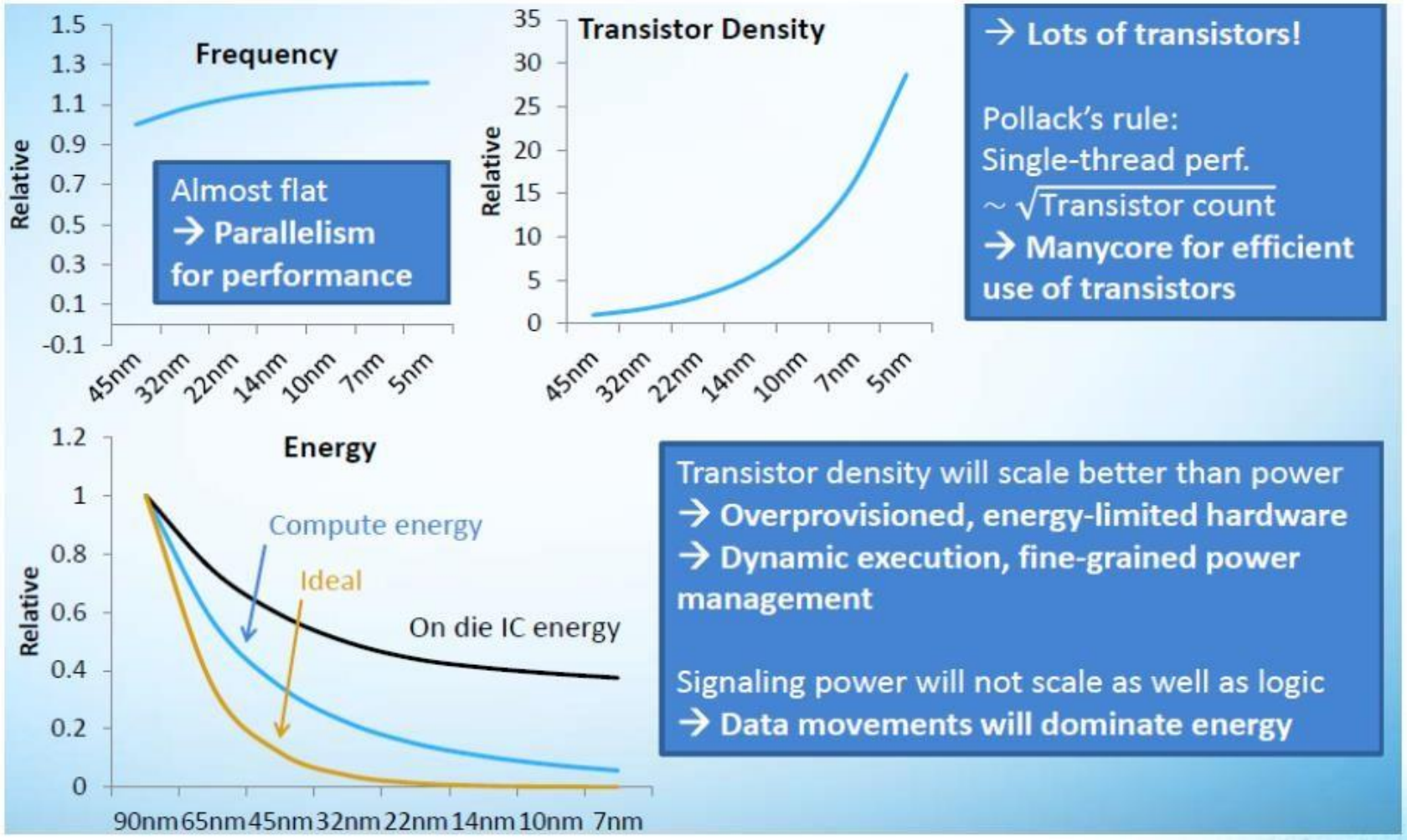
We have hit limits in

- Power
- Instruction level parallelism
- Clock speed

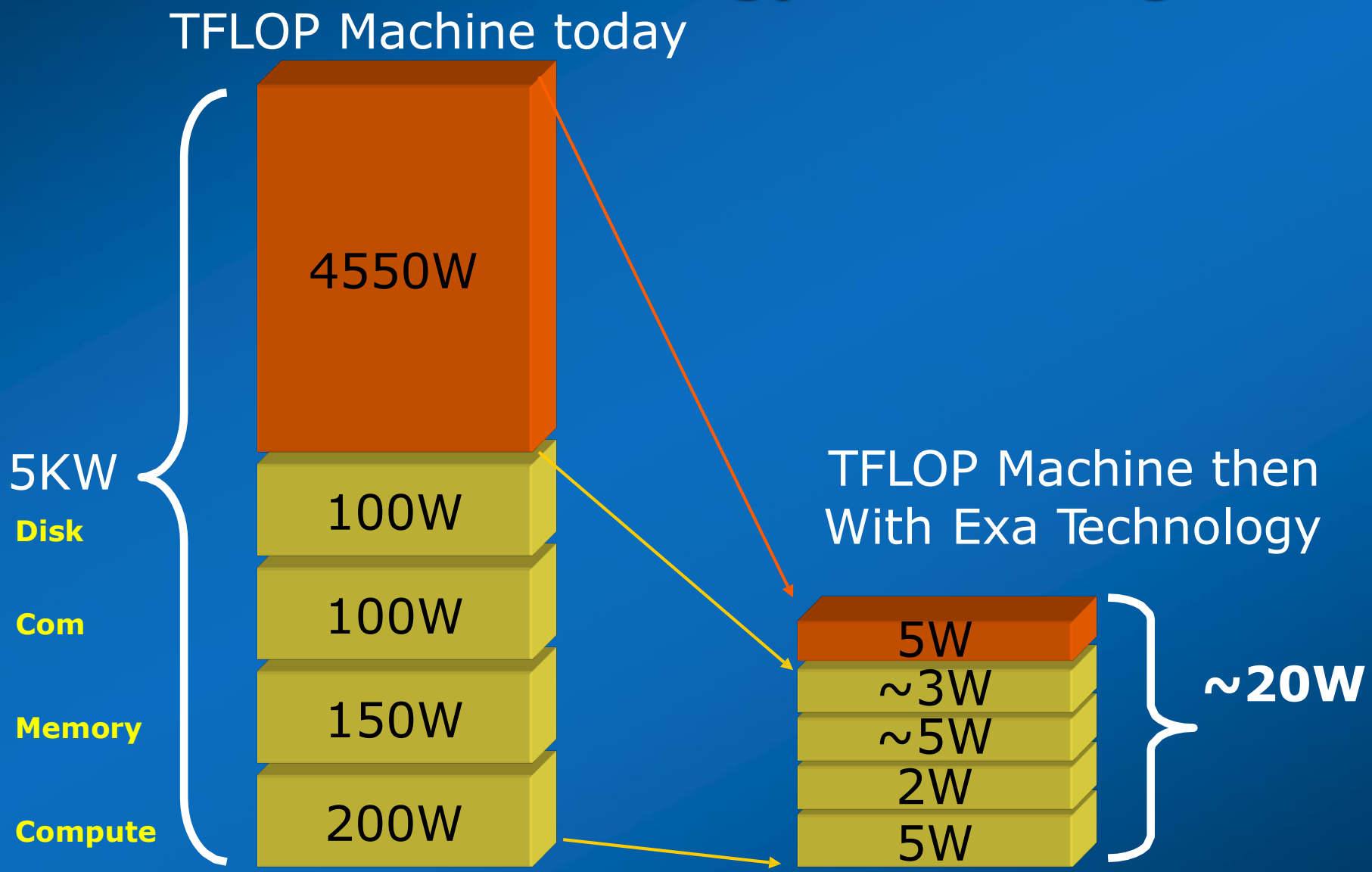
Single core scalar performance is now only growing slowly



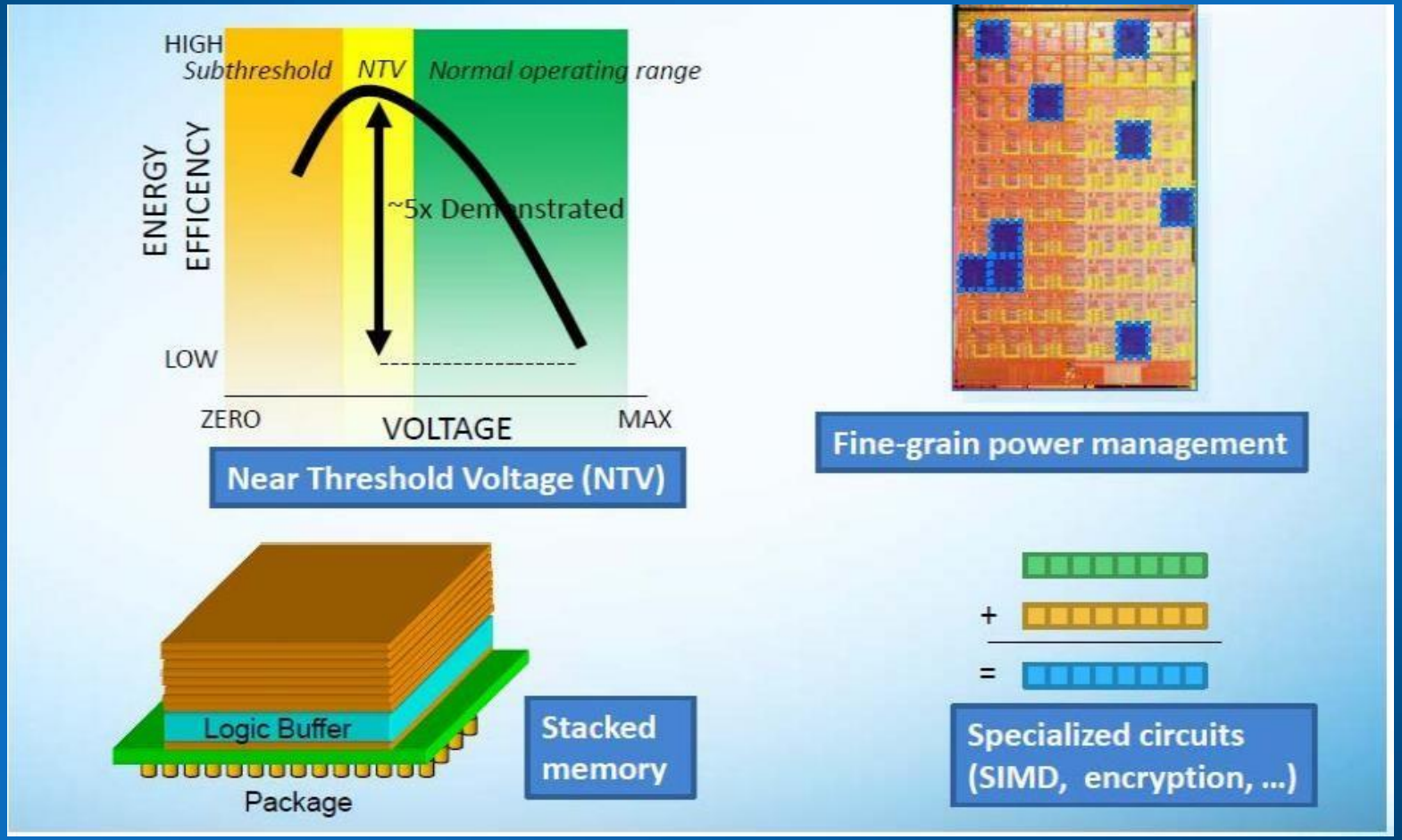
Technology Scaling Outlook



The Power & Energy Challenge



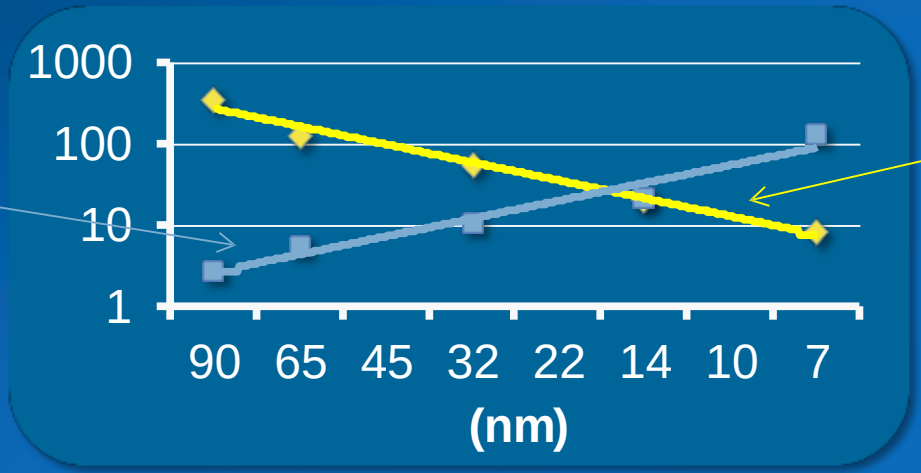
Promising Technologies



Rethink System Level Architecture

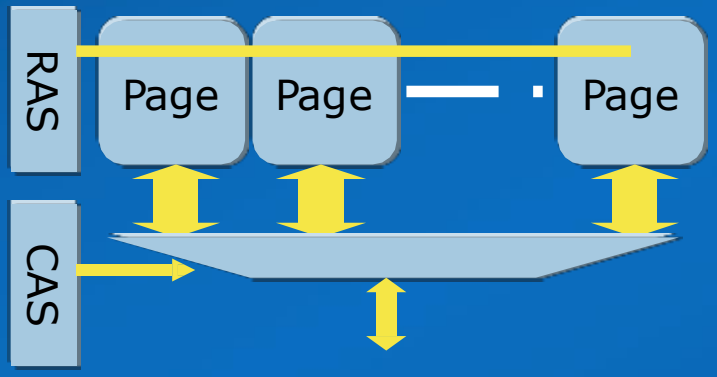
Revise DRAM Architecture

❶ Need exponentially increasing BW (GB/sec)



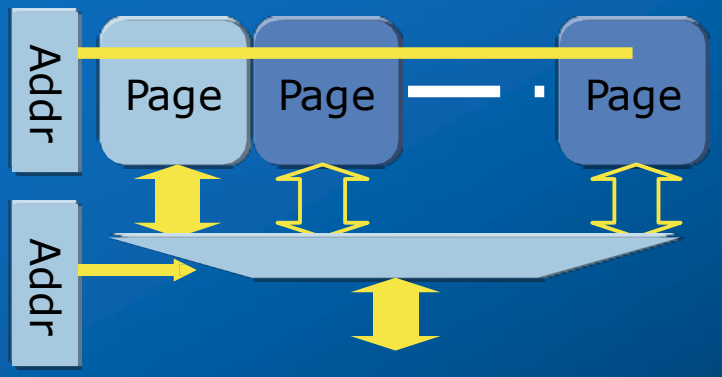
❷ Need exponentially decreasing energy (pJ/bit)

Traditional DRAM



Activates many pages
Lots of reads and writes (refresh)
Small amount of read data is used
Requires small number of pins

New DRAM architecture



Activates few pages
Read and write (refresh) what is needed
All read data is used
Requires large number of IO's (3D)

3D-Integration of DRAM and Logic

Logic Buffer Chip

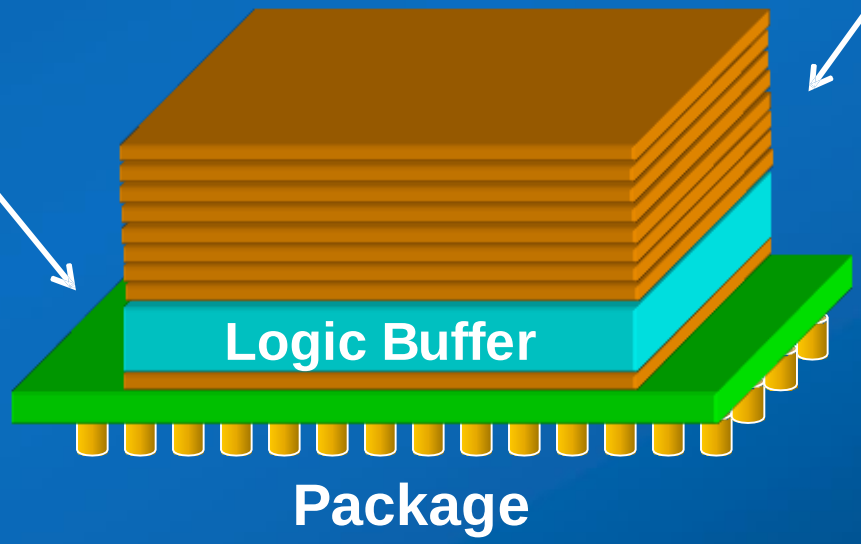
Technology optimized for:

- High speed signaling
- Energy efficient logic circuits
- Implement intelligence

DRAM Stack

Technology optimized for:

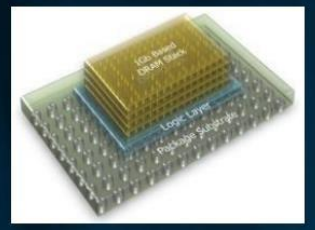
- Memory density
- Lower cost



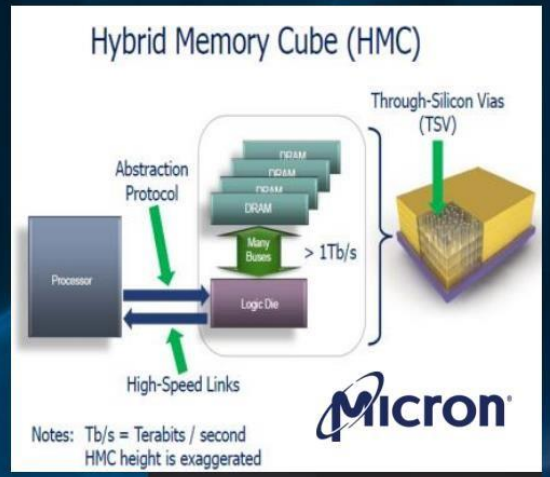
3D Integration provides best of both worlds

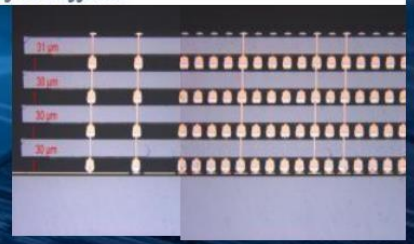
DRAM Scaling Using 3D Memory

1Tb/s HMC DRAM Prototype



- 3D integration technology
- 1Gb DRAM Array
- 512 MB total DRAM/cube
- 128GB/s Bandwidth
- <10 pj/bit energy





	Bandwidth	Energy Efficiency
DDR-3 (Today)	10.66 GB/Sec	50-75 pJ/bit
Hybrid Memory Cube	128 GB/Sec	**8 pJ/bit**

10X higher bandwidth, 10X lower energy

intel

Source: Micron

Needs a Paradigm Shift

Past and present priorities—

Single thread performance	Frequency
Programming productivity	Legacy, compatibility Architecture features for productivity
Constraints	(1) Cost (2) Reasonable Power/Energy

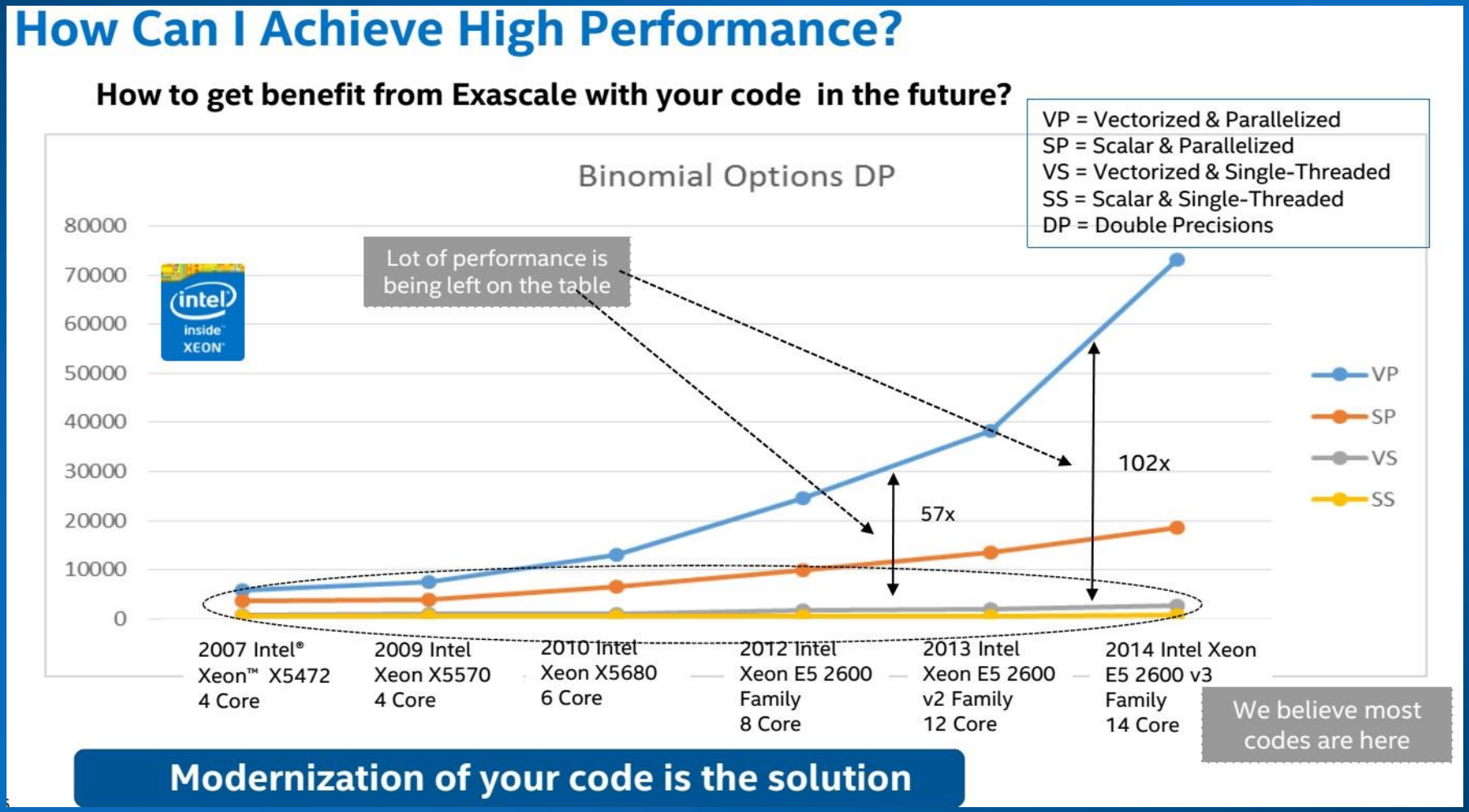
Future priorities—

Throughput performance	Parallelism
Power/Energy	Architecture features for energy Simplicity
Constraints	(1) Programming productivity (2) Cost

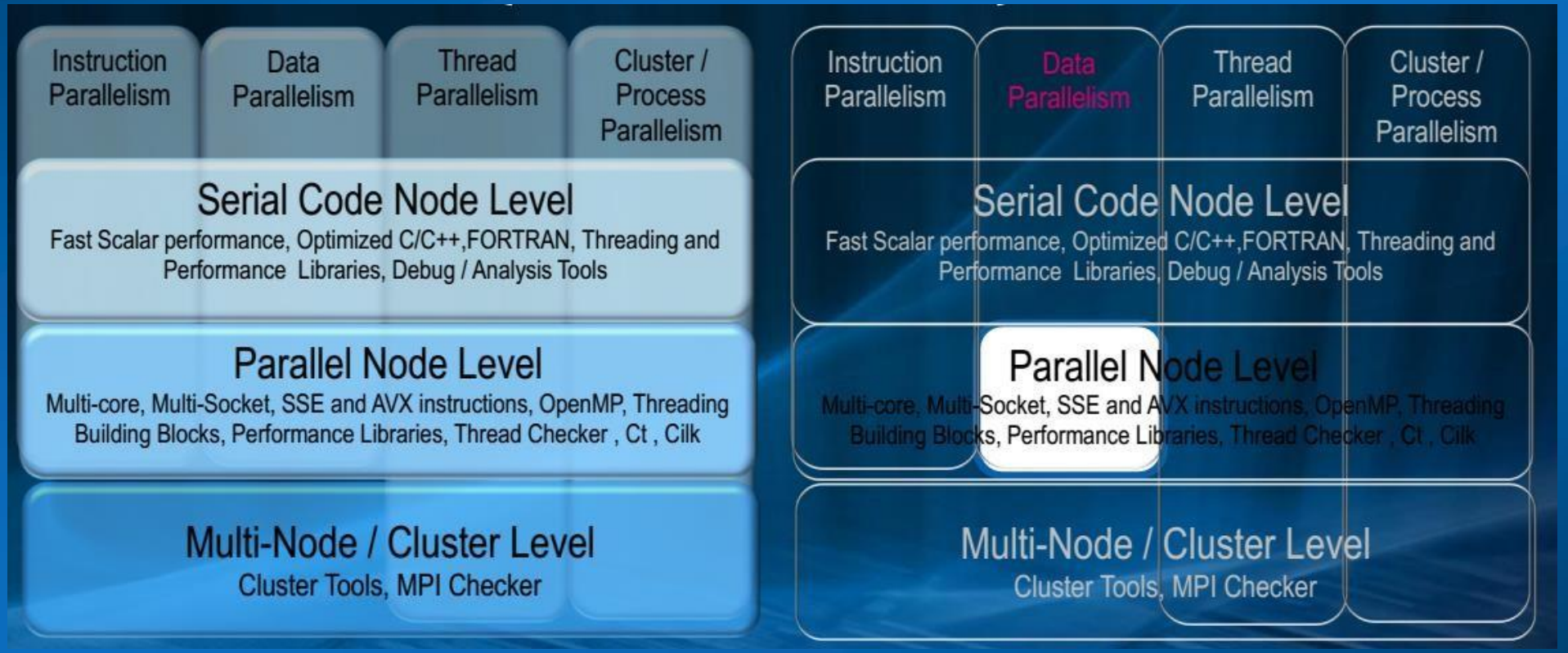
Evaluate each (old) architecture feature with new priorities

Intel: Investing to Remove 6 Bottlenecks

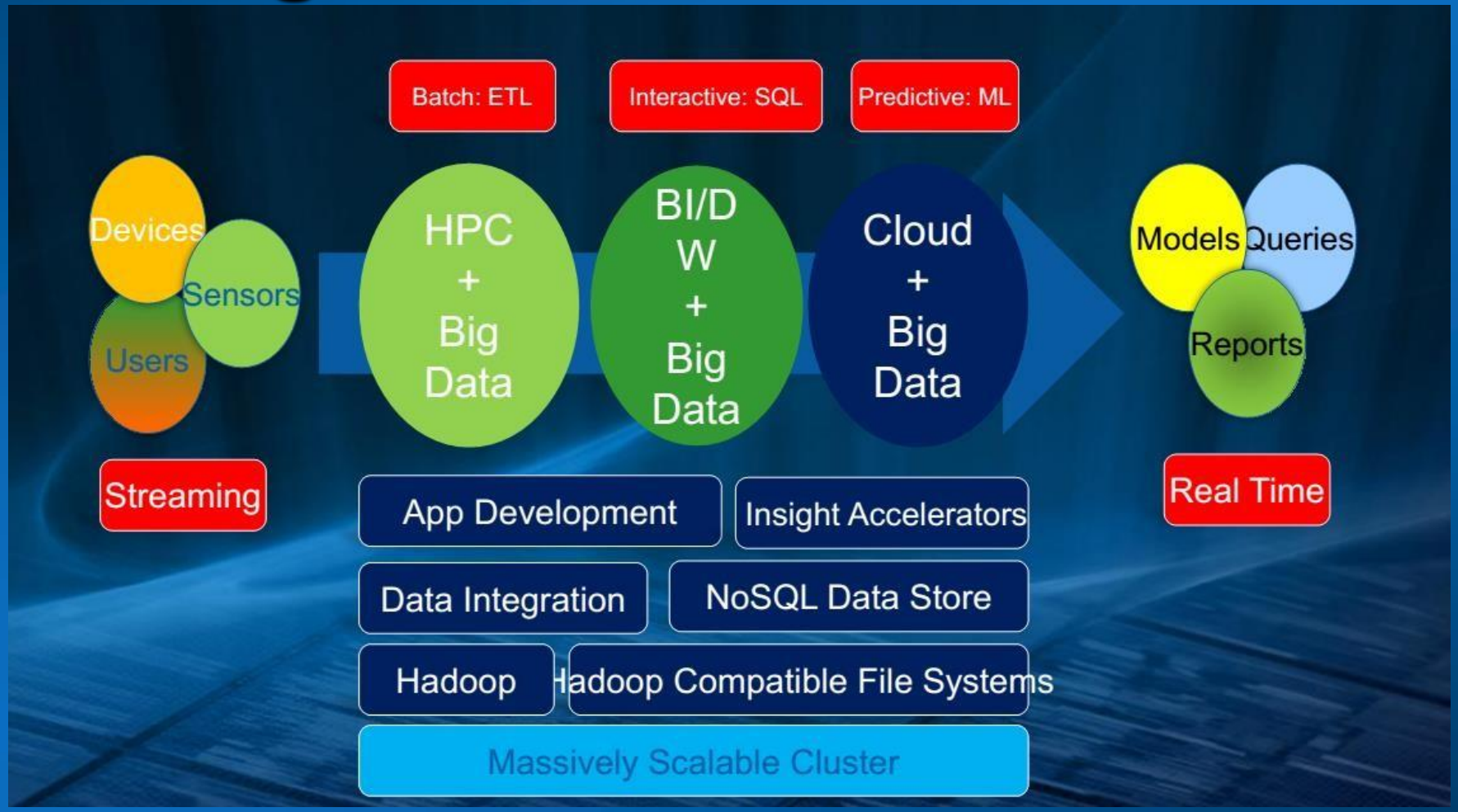
Impact on Applications



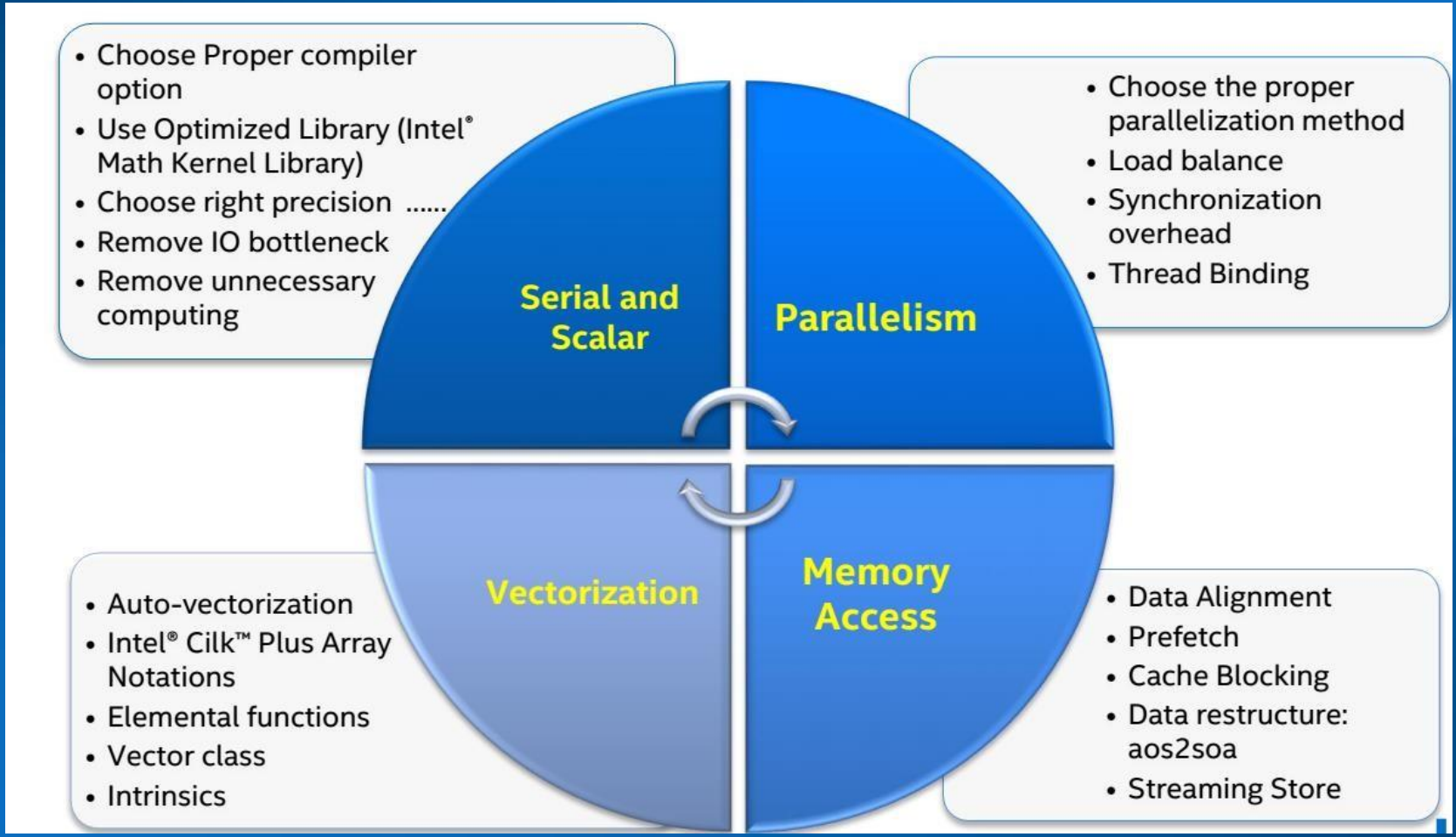
The Many Ways to Parallelism



And New Workloads will Emerge



Code Modernization – The 4D Approach



Intel® Xeon Phi™ Product Family

Based on Intel® Many Integrated Core (MIC) Architecture

"Meet **Knight's Landing: Intel's most powerful chip ever** is overflowing with cutting-edge technologies"

PCWorld – Jun 23, 2014

2013

Knights Corner

Intel® Xeon Phi™ x100 Product Family

22 nm process

Coprocessor

Over 1 TF DP Peak

Up to 61 Cores

Up to 16GB GDDR5

2016

Knights Landing

Intel® Xeon Phi™ x200 Product Family

14 nm process

Processor & Coprocessor

Over 3 TF DP Peak

Up to 72 Cores

On Package High-Bandwidth Memory

3X Single-Thread

Out-of-order core

Future

Knights Hill

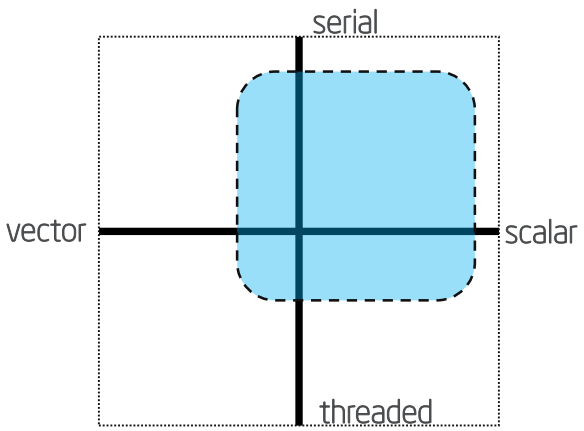
Next generation of the Intel® MIC Architecture Product Line

In planning

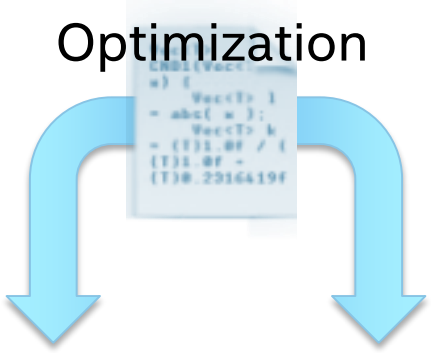
*Per Intel's announced products or planning process for future products

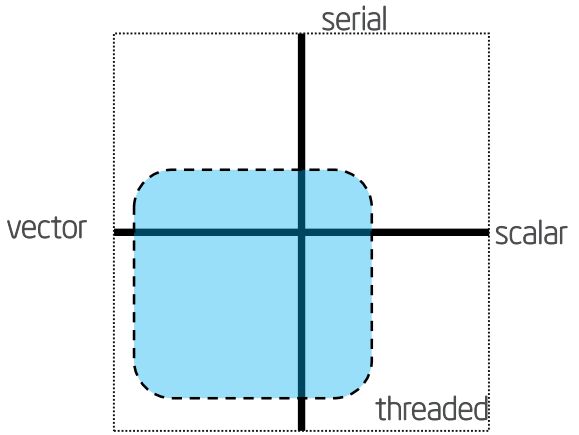
No Positioning Change

Knights Landing Targeted for Highly-Vectorizable, Parallel Apps



Single Source Code Optimization





Most Commonly Used Parallel Processor*

Parallel, Fast Serial

Multicore + Vector

Leadership Today and Tomorrow

intel inside XEON + intel inside XEON PHI

Optimized for Highly-Vectorizable Parallel Apps

Many Core

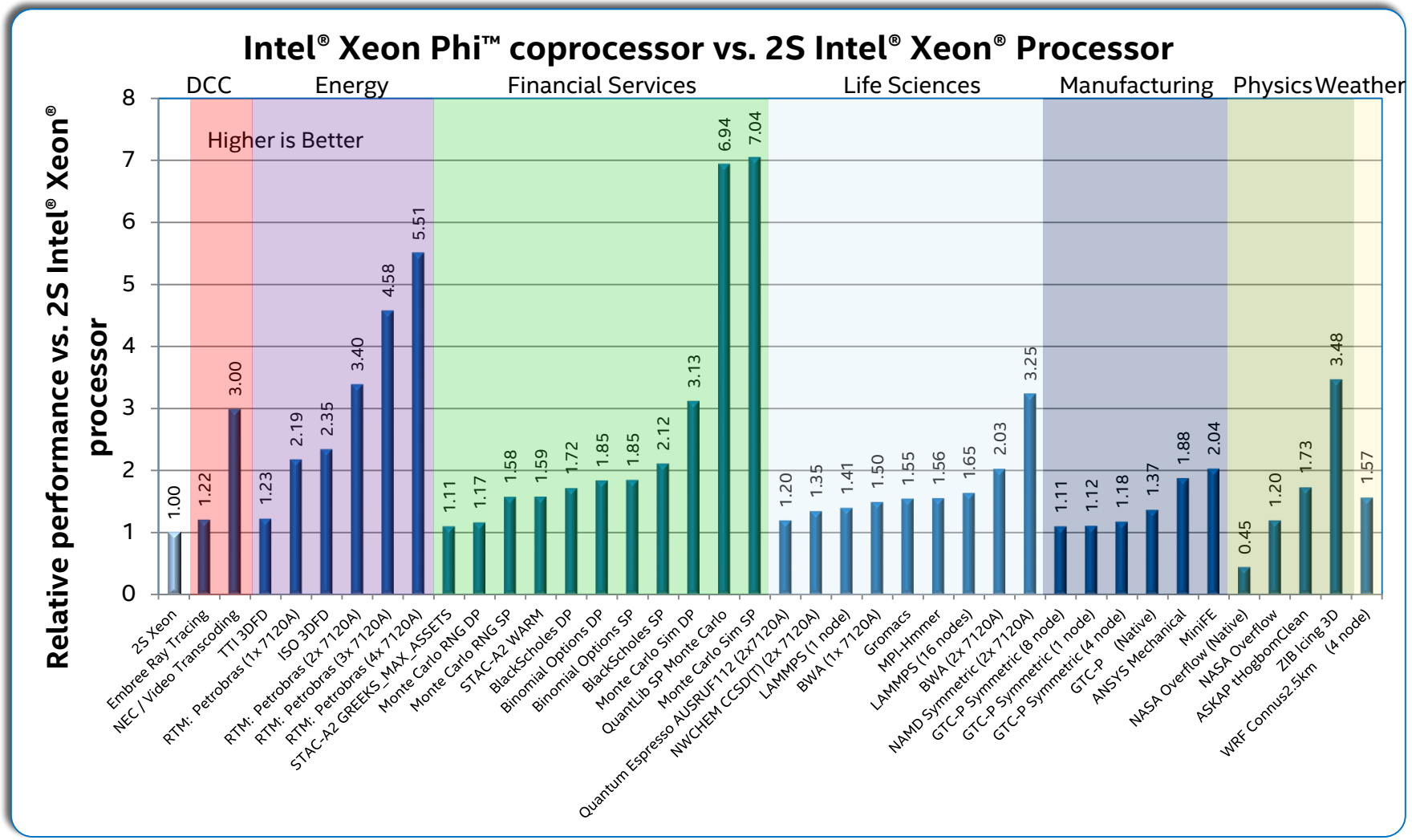
Support for 512 bit vectors

Higher memory bandwidth

Common SW programming

*Based on highest volume CPU in the IDC HPC Qview Q1'13

Is Xeon Phi compelling vs Xeon?



Intel Measured Results: Different hardware architectures may require different source code. Results are based on Intel's best efforts to use code optimized to run best on all architectures and perform the same work. Future code optimizations may result it different results.

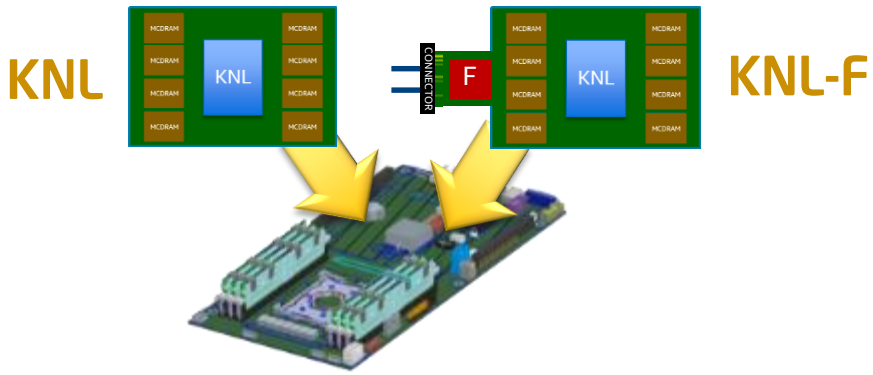
"Rifle shot" approach targeted with customers in FSI, Oil & Gas, and Life Sciences based on affinity/readiness

- Xeon = Intel® Xeon® processor
- Xeon Phi = Intel® Xeon Phi™ coprocessor

Three Knights Landing Products

Knights Landing Processor

"Self-boot" Intel® Xeon Phi™ processor platform



KNL and KNL-F Processors:

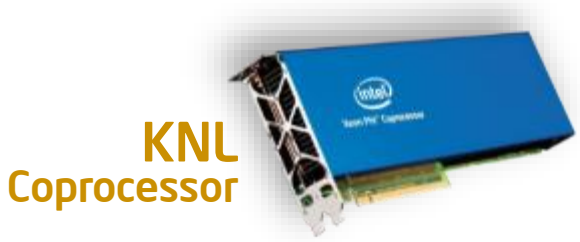
- Knights Landing IS the host processor
- Boots standard off-the-shelf OS's

Benefits:

- Higher performance density for highly parallel applications[2]
- Reduced system power consumption[2]
- Higher perf/Watt & perf/$$[3]

Knights Landing Coprocessor

Requires Intel® Xeon® processor host



Knights Landing Coprocessor:

- Solution for general purpose servers and workstations

Benefits:

- Targeted for applications with larger sections of serial work[1]
- Upgrade path from Knights Corner as PCIe card

1 Projections based on early product definition and as compared to prior generation Intel® Xeon Phi™ Coprocessors
2 Based on Intel internal analysis.. Lower power based on power consumption estimates between (2) HCAs compared to 15W additional power for KNL-F. Higher density based on removal of PCIe slots and associated HCAs populated in those slots.
3 Results have been estimated based on internal Intel analysis and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance. 2 Results based on internal Intel analysis using estimated theoretical Flops/s for KNL processors, along with estimated system power consumption and component pricing in the 2015 timeframe. See backup for complete system configurations.

A Paradigm Shift for Highly-Parallel

Server Processor and **Integration** are Keys to Future

Coprocessor

Fabric

Memory

Server Processor

Memory Bandwidth
~500 GB/s STREAM

Memory Capacity
Over 25x KNC*

Resiliency
Systems scalable to >100 PF

Power Efficiency
Over 25% better than card[1]

I/O
Up to 50 GB/s BW with int fabric

Cost
Less costly than discrete parts[1]

Flexibility
Limitless configurations

Density
3+ KNL with fabric in 1U[3]

**Comparison to 1st Generation Intel® Xeon Phi™ 7120P Coprocessor (formerly codenamed Knights Corner)*
[1]Results based on internal Intel analysis using estimated power consumption and projected component pricing in the 2015 timeframe. This analysis is provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.
[2]Comparison to a discrete Knights Landing processor and discrete fabric component.
[3]Theoretical density for air-cooled system; other cooling solutions and configurations will enable lower or higher density.

Knights Landing Architectural Diagram

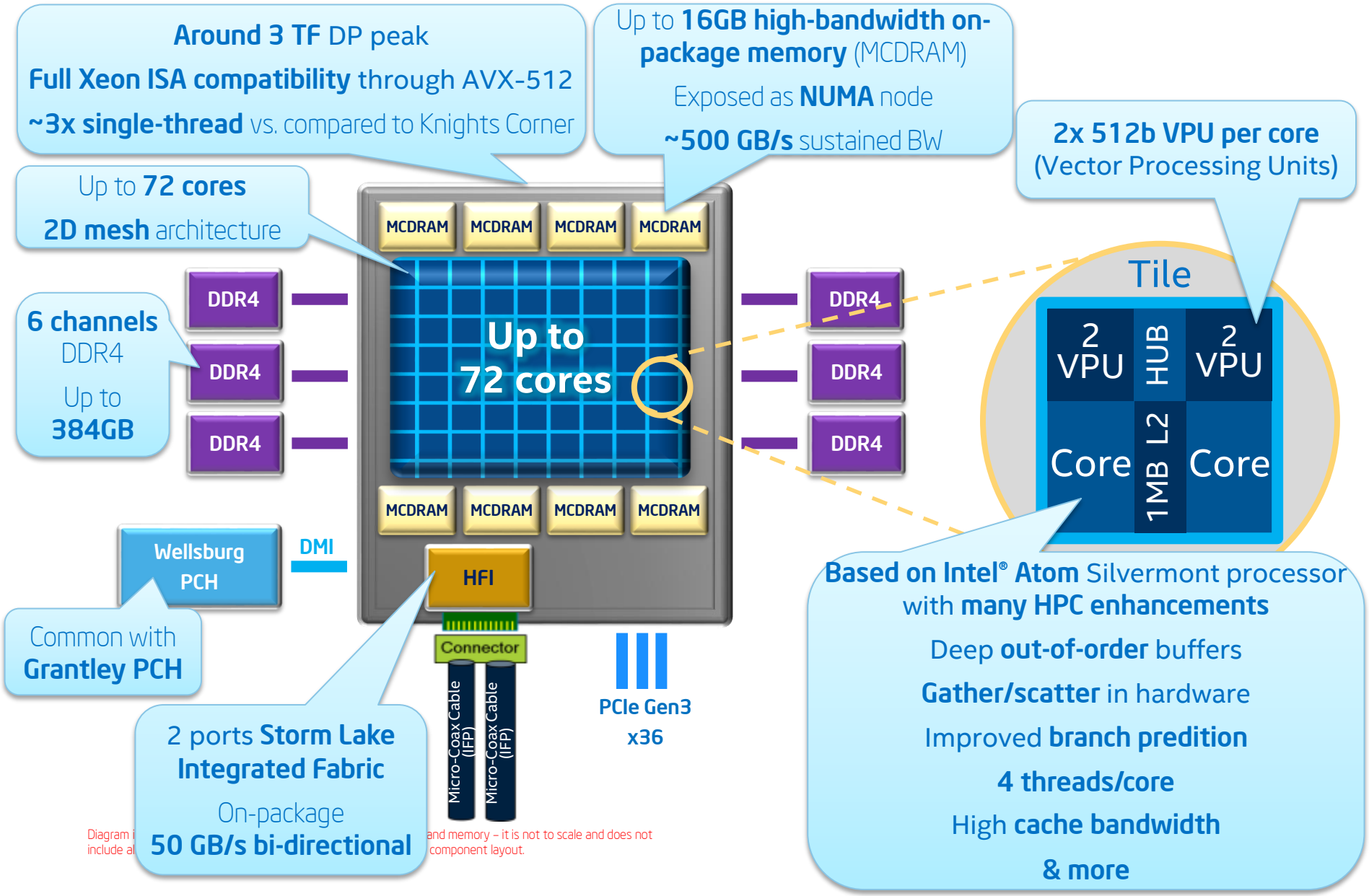


Diagram i include al and memory – it is not to scale and does not component layout.

Today's Parallel Investment Carries Forward

Sustained threading, vectorization, cache-blocking and more

MOST optimizations carry forward with a recompile

Incremental tuning gains

1st Generation Intel® Xeon Phi™ Coprocessor

Native or Symmetric or Offload

Recompile

MKL | MPI | TBB

OpenMP | Cilk Plus™ | OpenCL

KNL Enabled Performance Libraries & Runtimes

AVX-512

Cache Mode For High Bandwidth Memory

KNL Enabled Compilers

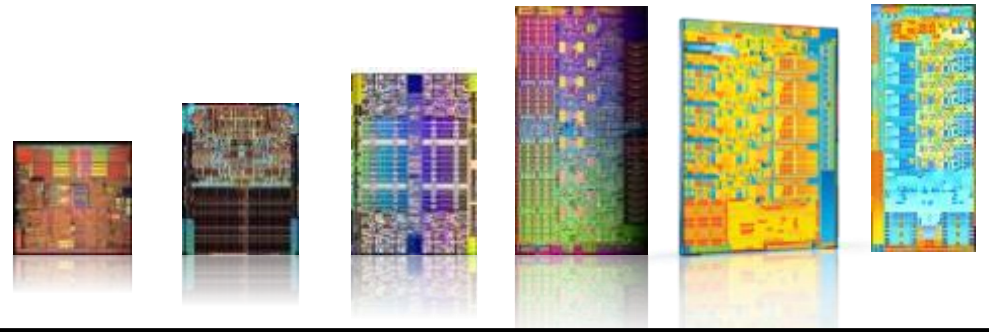
Tuning

KNL Enhance-ments

(memory, architecture, bandwidth, etc.)

Parallel is the Path Forward

Intel® Xeon® and Intel® Xeon Phi™ Product Families are both going parallel



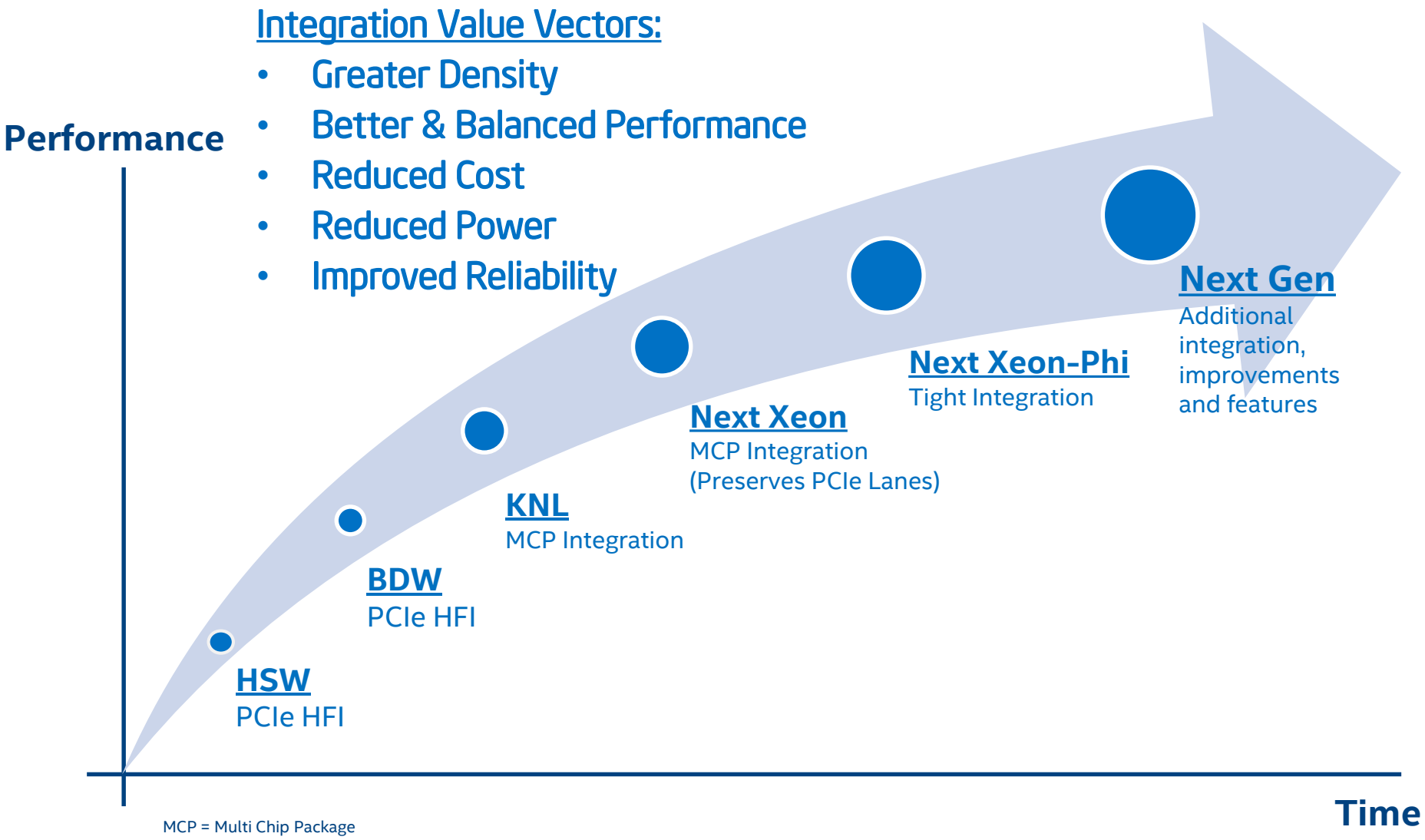
	Intel® Xeon® processor **64-bit**	Intel® Xeon® processor **5100 series**	Intel® Xeon® processor **5500 series**	Intel® Xeon® processor **5600 series**	Intel® Xeon® processor code-named **Sandy Bridge EP**	Intel® Xeon® processor code-named **Ivy Bridge EP 4S**[1]	**Future** Intel® Xeon® processor	Intel® Xeon Phi™ coprocessor **Knights Corner**	Intel® Xeon Phi™ processor & coprocessor **Knights Landing**[1]
Core(s)	1	2	4	6	8	12	*tbd*	57-61	Up to 72
Threads	2	2	8	12	16	24	*tbd*	228-244	Up to 288
SIMD Width	128	128	128	128	256	256	512	512	512

More cores → More Threads → Wider vectors

*Product specification for launched and shipped products available on ark.intel.com.

1. Not launched or in planning.

Intel® Omni-Path Fabric - CPU Integration

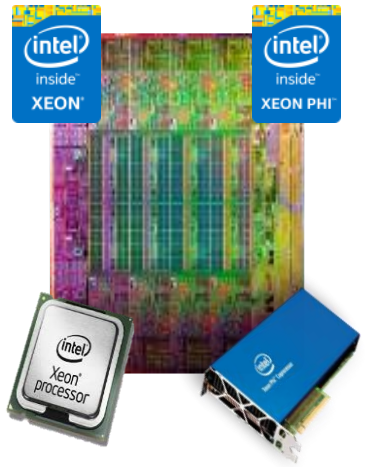


MCP = Multi Chip Package

Intel's Technical Computing Portfolio

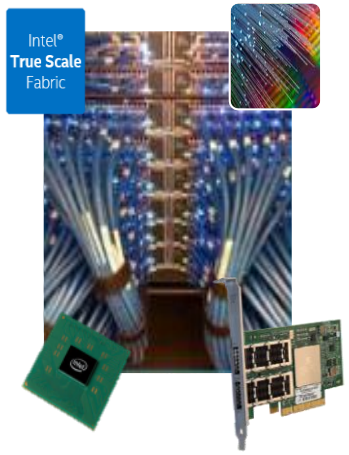
Technologies & Products

Compute Processing



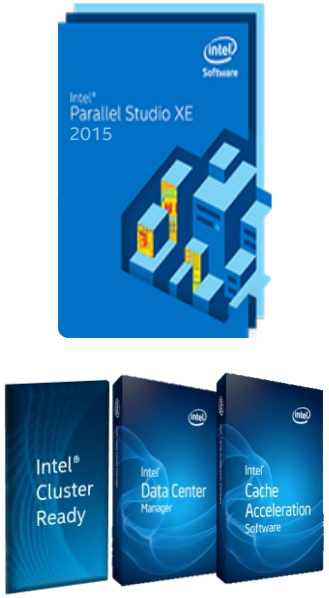
Systems & Boards

Network & Fabric



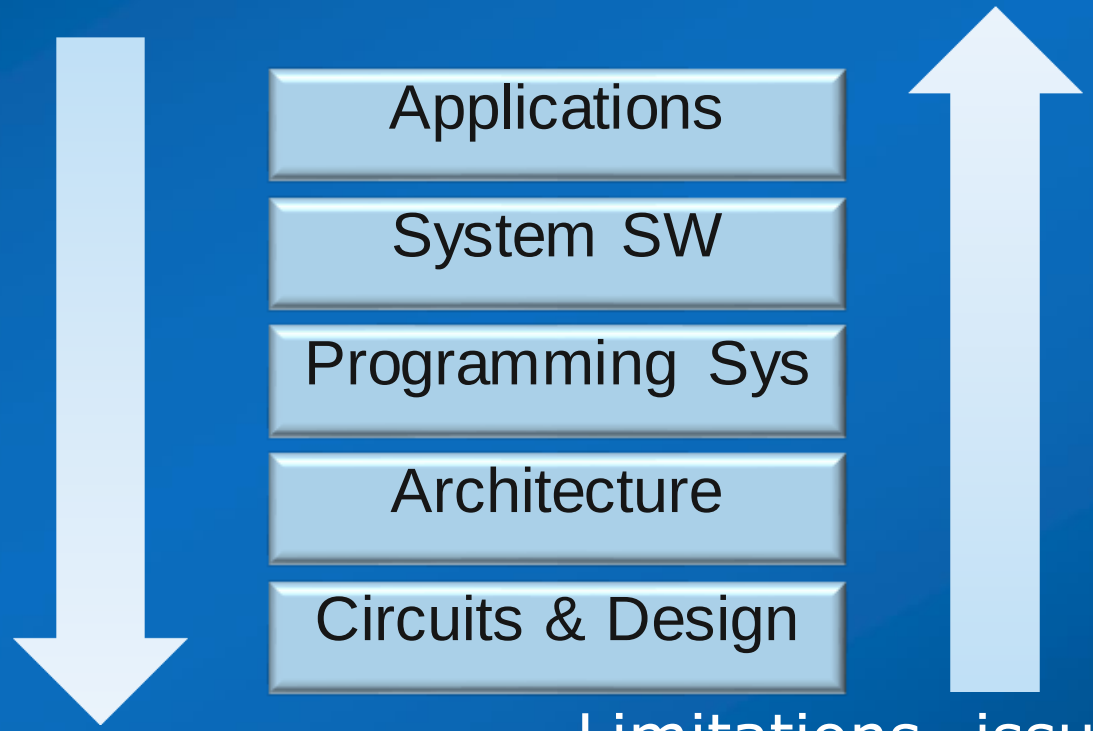
I/O & Storage

Software & Services



HW-SW Co-design

Applications and SW stack provide guidance for efficient system design



Limitations, issues and opportunities to exploit

What will matter in 10 years

	Now	2025
Perf/$	Linpack, Real Applications	Real Applications
Perf/Watt	Limited by worst case application	All applications will be able to run at chosen power level. Dynamic, optimal energy management.
Reliability	Use of file system checkpoint restart (spinning disks)	Transparent hardware and system software recovery. Checkpoints in non-mechanical media.
Big Data	Parallel IO	New storage paradigm

SW Challenges

Execution model

Programming model

1. Extreme parallelism (1000X due to Exa, additional 4X due to NTV)
2. Data locality—reduce data movement
3. Intelligent scheduling—move thread to data if necessary
4. Fine grain resource management (introspective)
5. Applications and algorithms incorporate paradigm change

Summary

- Exascale will be there by 2022 or so
- "Business as usual" (riding on Moore's Law and commodity technology) is becoming increasingly harder
- Supercomputers are becoming more "special purpose"
  - Expect most/all supercomputers to use floating point accelerators in a few years; more specialized accelerators to follow
- Can continue to push performance to zetascale
  - Will need to think of supercomputers as unique facilities, such as particle accelerators – not clusters of PCs
- *Supercomputing will become much more interesting*