Race to Exascale: Opportunities and Challenges

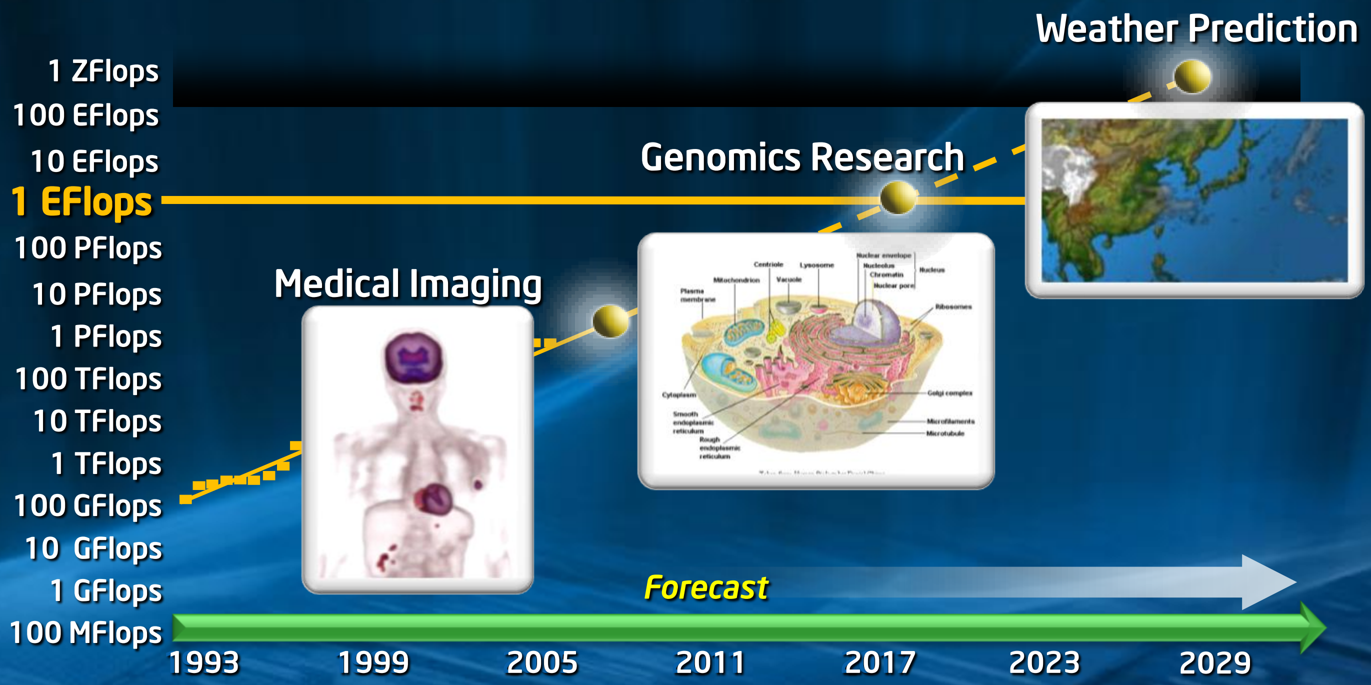
Avinash Sodani, Ph.D.

Chief Architect MIC Processor

Intel Corporation

Exascale

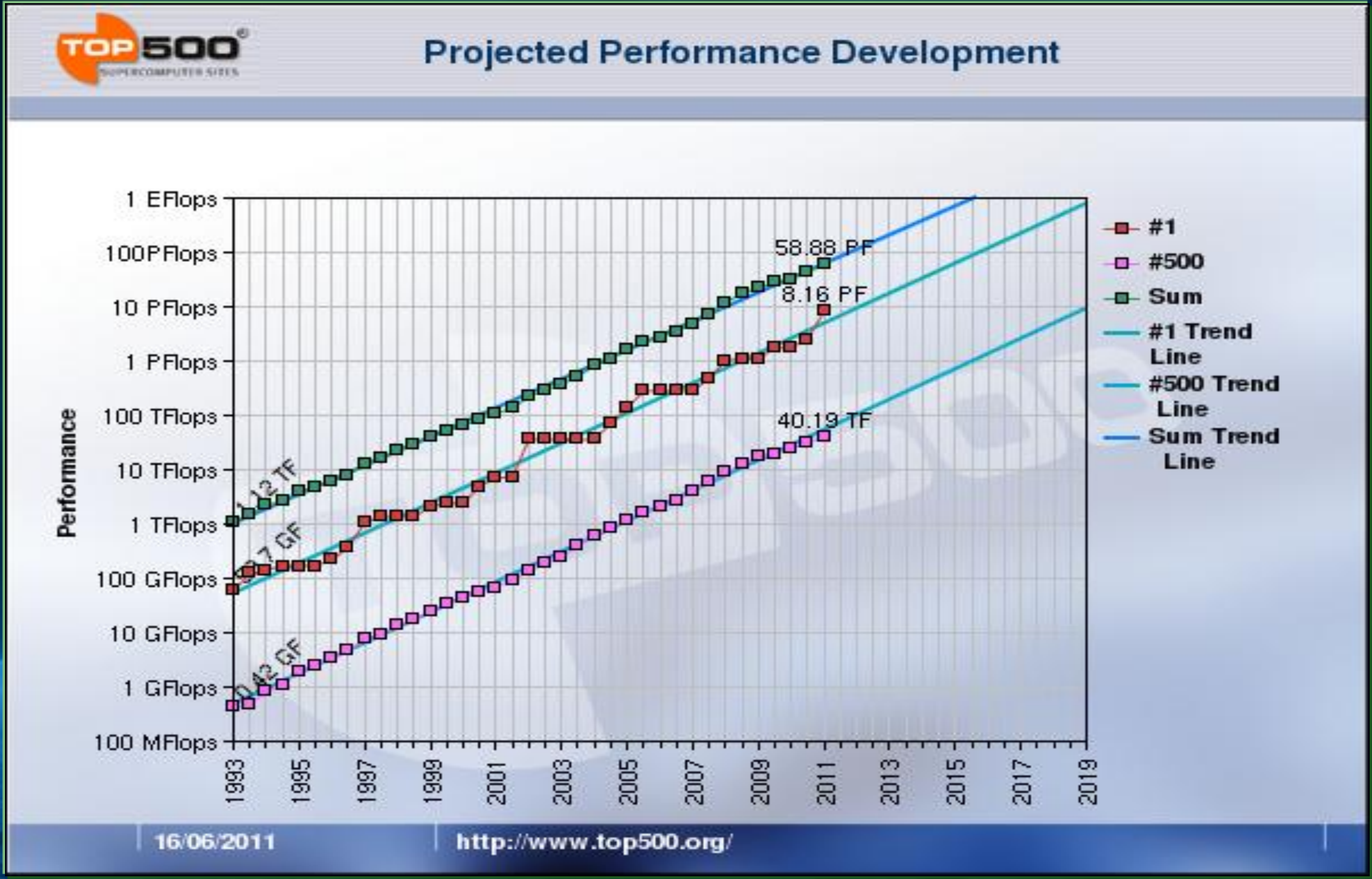
Goal: 1-ExaFlops (10^{18}) within 20 MW by 2018



Solve many yet impossible life changing problems

Make PFlop HPC computing affordable and ubiquitous

Trends to Exascale Performance



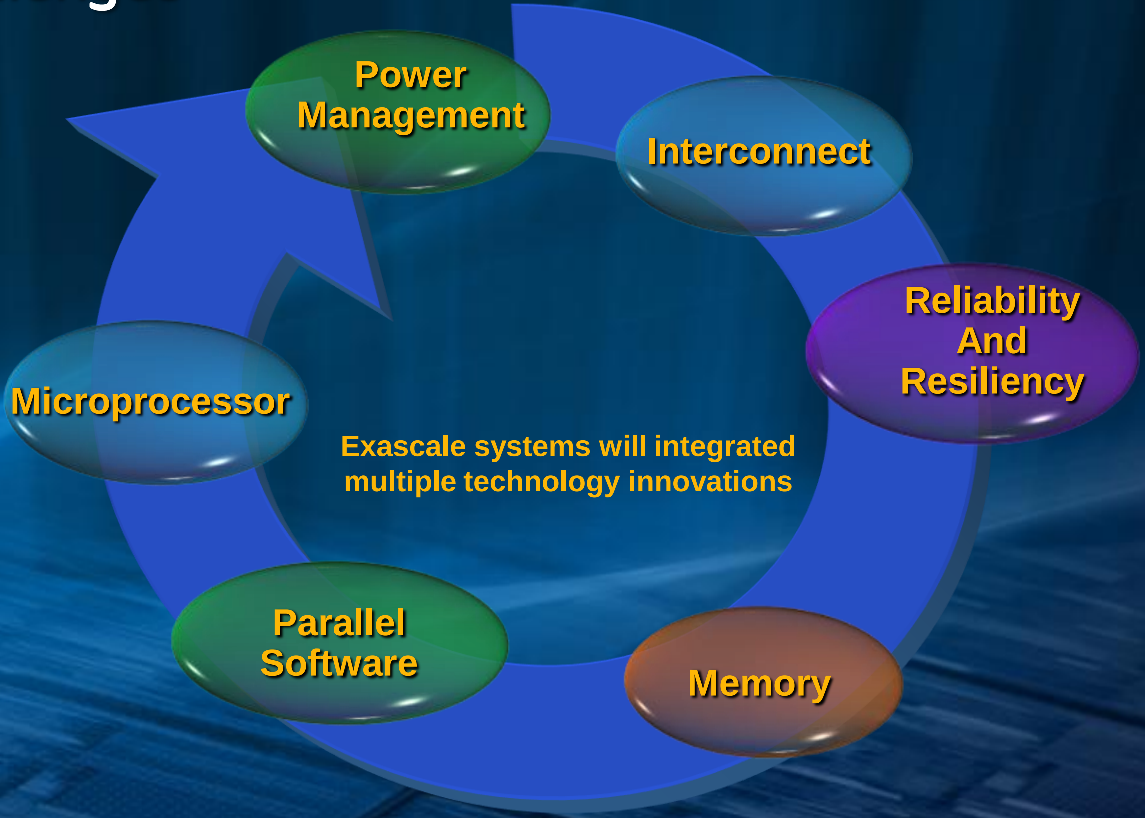
Roughly 10x performance every 4 years.

Predicts that we'll hit Exascale performance in 2018-19

But … that doesn't mean it is done!

Many system level challenges

- Power efficiency
- Compute density
- Memory technology
- Network technology
- Reliability
- Software



Exascale systems will integrated multiple technology innovations

… to name a few

Today we'll talk about

- Two Challenges: Power, Memory
- Commonly held myths about power consumption
- Approach forward

Challenge 1: Compute Power

At System Level:

Today: 10 PF, 12 MW → 1200 pJ/Op

Exaflop: 1000 PF, 20 MW → 20 pJ/Op

Needs improvements in all system components

Processor-subsystem needs to reduce to 10 pJ/Op

~60x improvement needed for Exascale

Challenge 2: Memory

Memory bandwidth fundamental to HPC performance

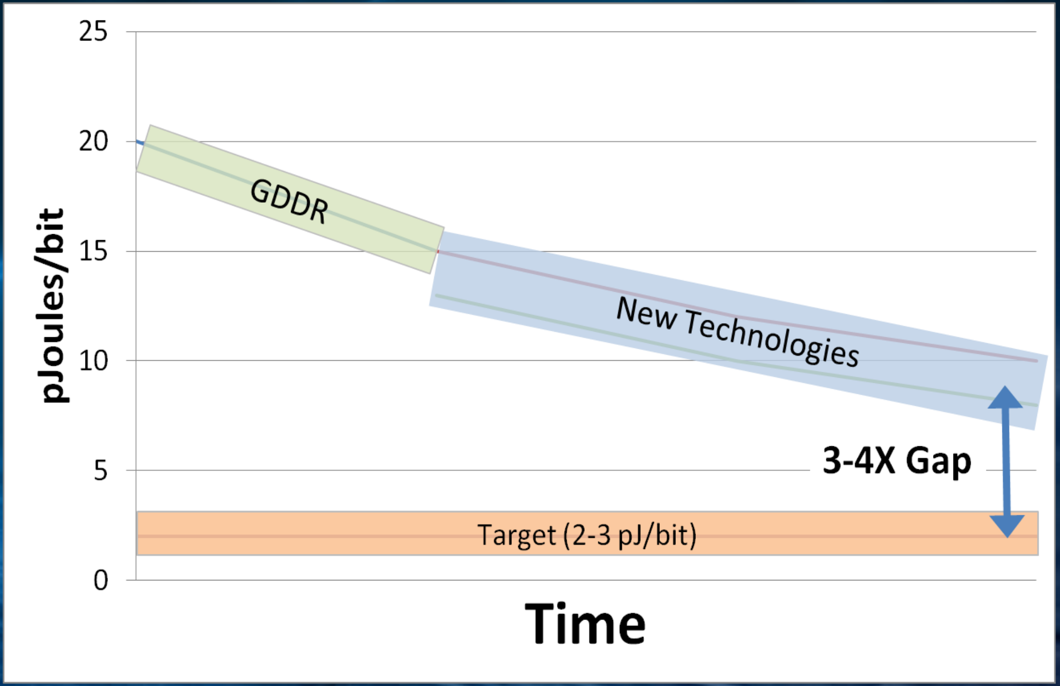
- Need to balance with capacity and power

1 ExaFlop Machine

→ ~200-300 PB/sec

→ ~2-3 pJ/bit

- GDDR will be out of steam
- Periphery connected solutions will run into pin count issues



Existing technology trends leave 3-4x gap on pJ/bit

Power: Commonly Held Myths

- IA cannot hit power efficiencies → Need major shift in programming paradigm
- OOO/Speculation and other ST features too power hungry
- IA memory model, Cache Coherence too power hungry
- Caches don't help for HPC → They waste power

Purpose of Exascale Pursuit

- Not just to achieve exascale flops
- Expectation is it will force efficiencies and innovations at multiple levels.
- Technologies invented in its pursuit will benefit all forms of computing.
- Hence, important that we don’t special case the technologies just for high performance Linpack score goals

My Musings

- Given aggressive goals → tempting to go for radical changes
  - New programming models
  - Big steps back on microarchitecture: removal of coherency, caches, speculation, etc.
  - Big shifts in HW/SW boundary

- Typically ubiquitous innovations
  - Do not rely on major behavior changes (they cause it)
  - Do not require major eco-system enablement as pre-requisite.
  - They simplify instead of complicate

- Meet targets while retaining features that make solution easy to use

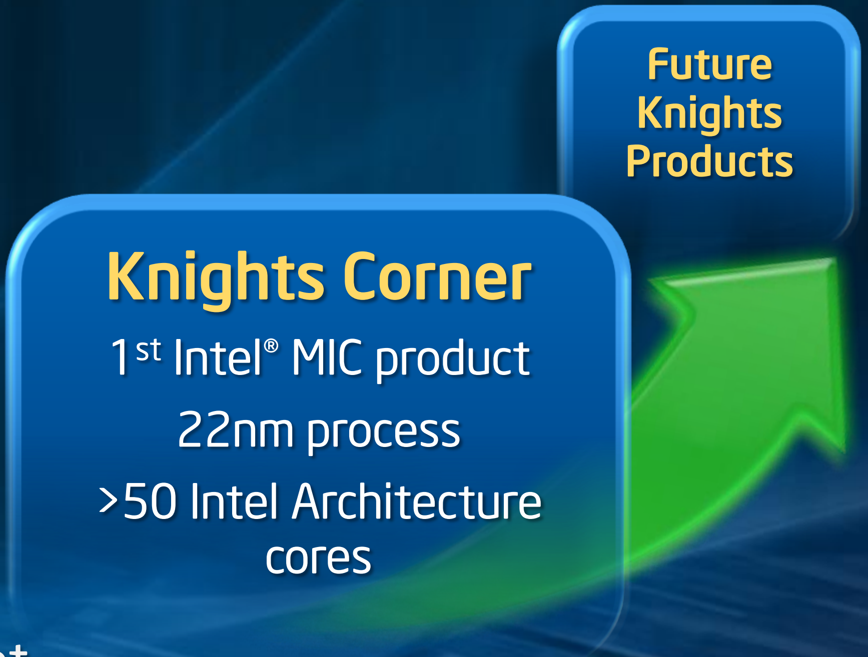
- That's the real Exascale challenge.

Many Integrated Core (MIC) Architecture

- Knights Line of products

- Key attributes
  - Many core and many threads
  - High power efficiency
  - Wide vectors and advance FP capability to provide FLOPs
  - Coherent caches
  - IA ISA and programming model
  - Runs standard programs and toolset

- Demo'ed Knights Corner in SC '11 running 1TF DP



Knights Corner

1st Intel® MIC product

22nm process

>50 Intel Architecture cores

Future Knights Products

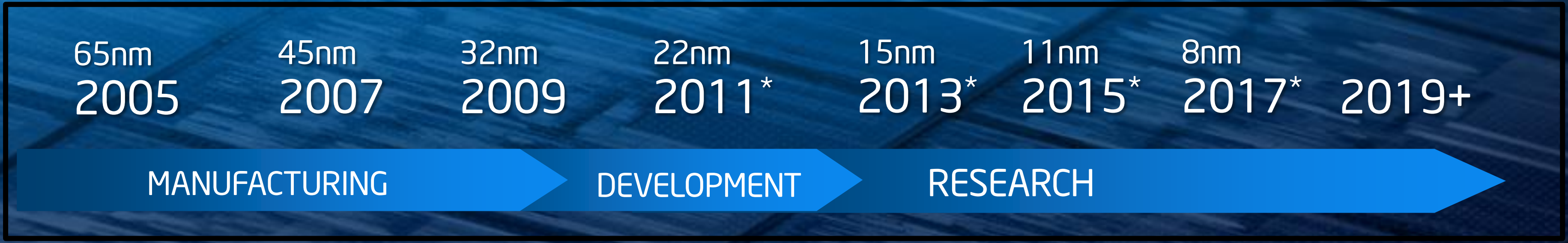
Myth 1:

IA cannot hit power efficiencies → Need major shift in programming paradigm

Performance/Power Progression

- Process:1.3x - 1.4x (per generation)
- Arch/Uarch: 1.1x - 2.0x (per generation)
- Multi core → Many core improvement: 4x (one time)

Recurring improvement: 1.4 – 3.0x every 2 years



65nm 2005

45nm 2007

32nm 2009

22nm 2011*

15nm 2013*

11nm 2015*

8nm 2017*

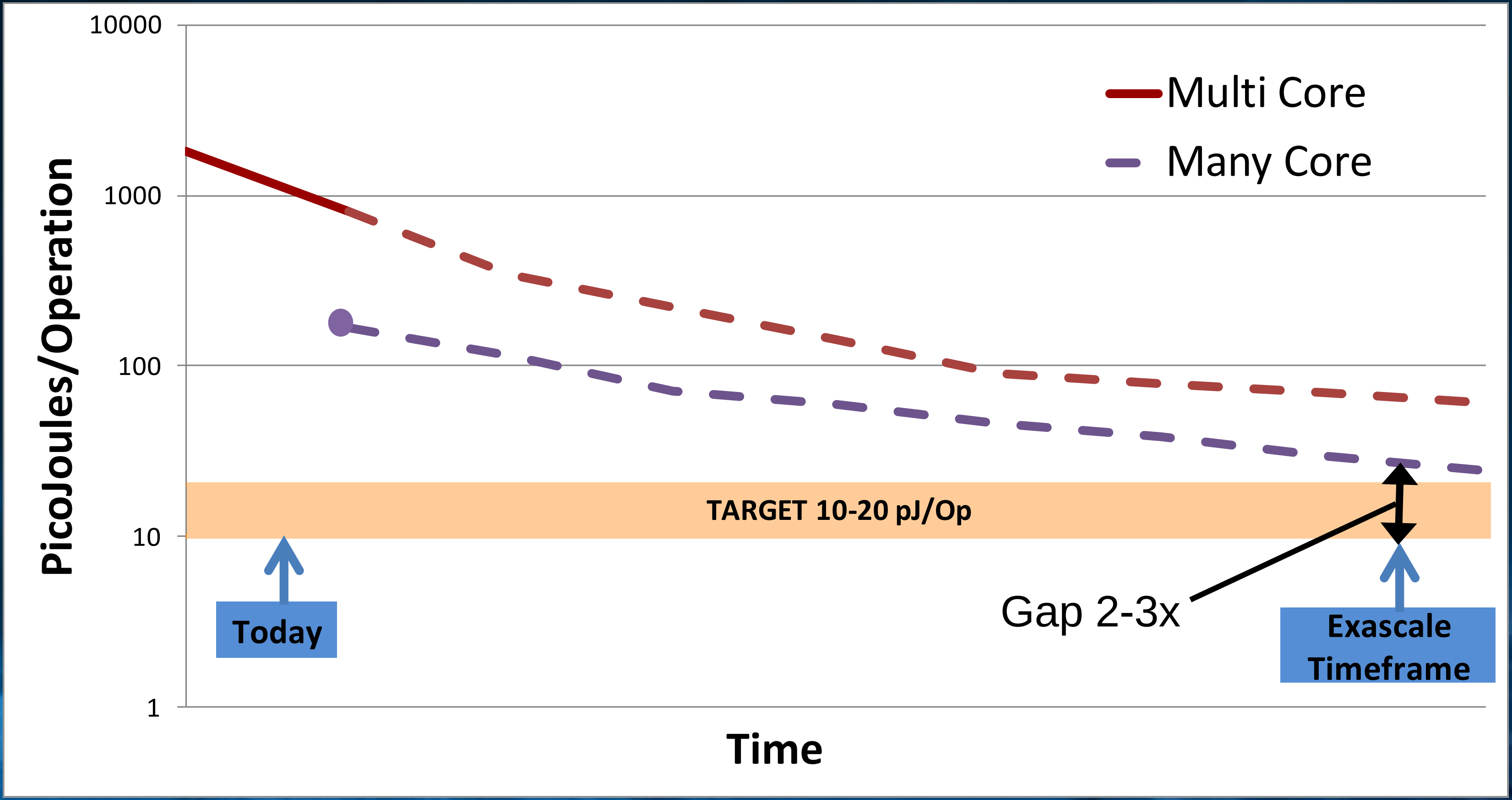
2019+

MANUFACTURING

DEVELOPMENT

RESEARCH

Power



Gap reduced to 2-3x from 50x with existing techniques!

Have 7-8 years to innovate to cover this gap

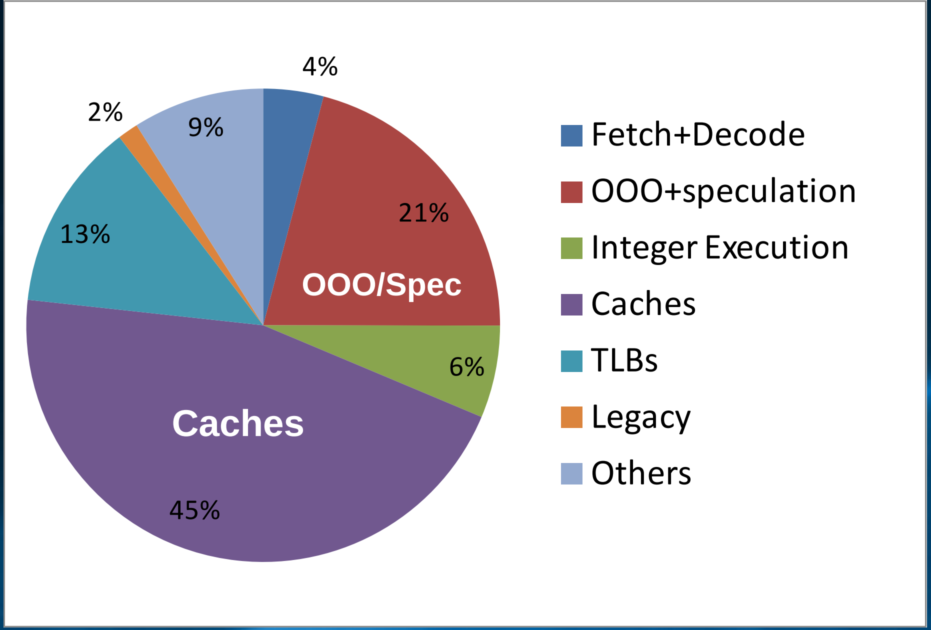
Do not need new programming paradigm to do that

Myth 2:

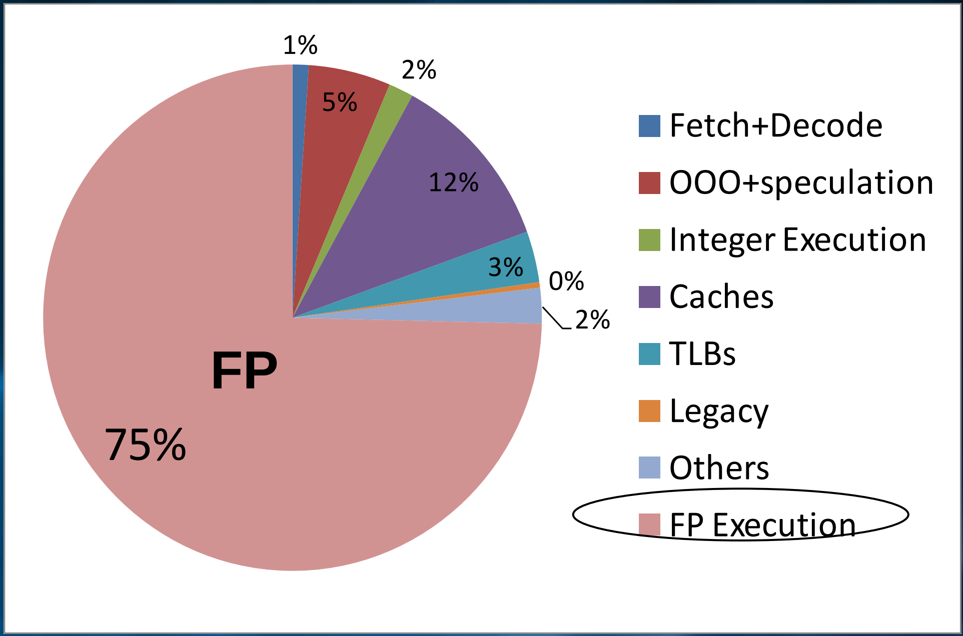
OOO/Speculation and other ST features too power hungry

Core Power Distribution

Non Compute-heavy Application

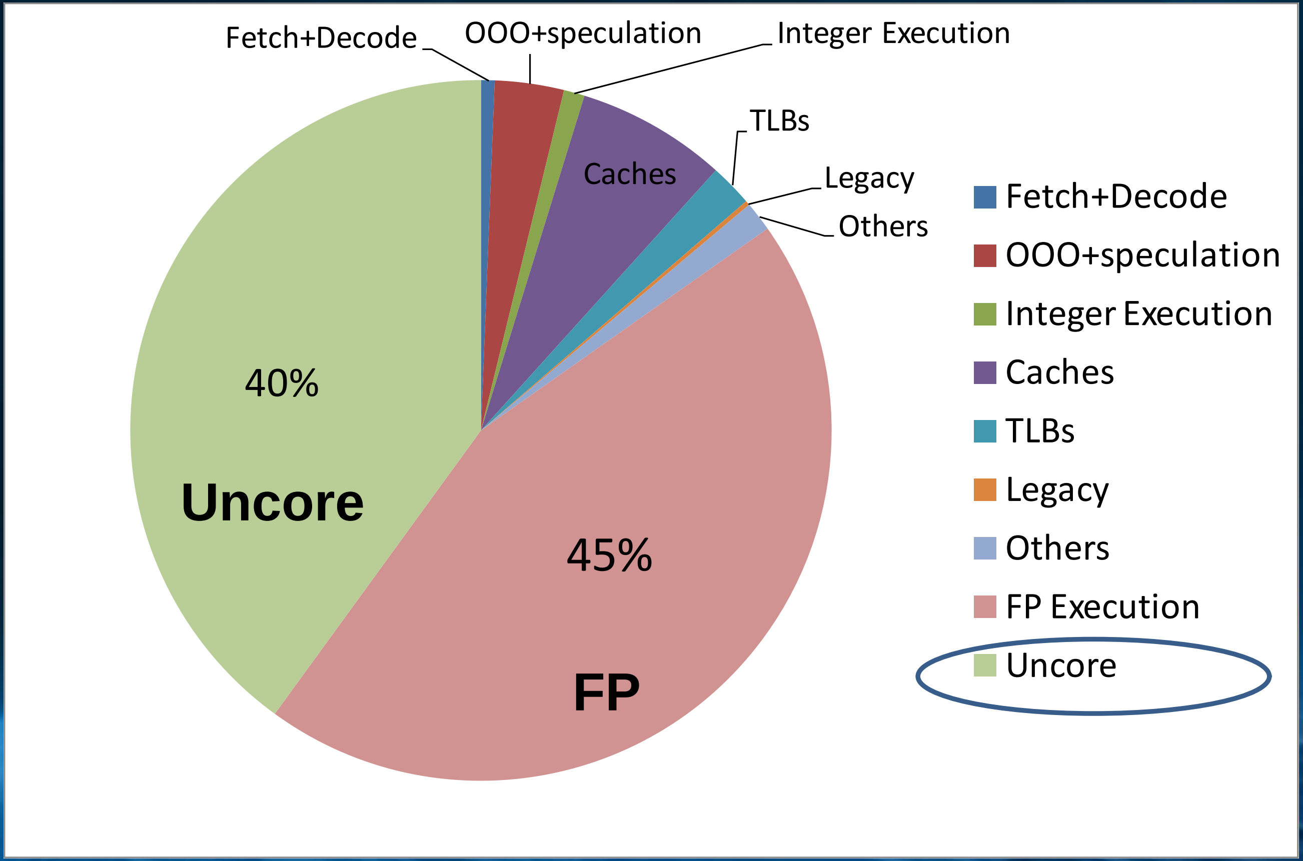


Compute-heavy Application



- Power dominated by compute – as should be the case
- OOO/Speculation/TLB: < 10%
- X86 Legacy+Decode = ~1%

Chip-level Power Distribution



At chip level core power is even smaller portion (~15%).

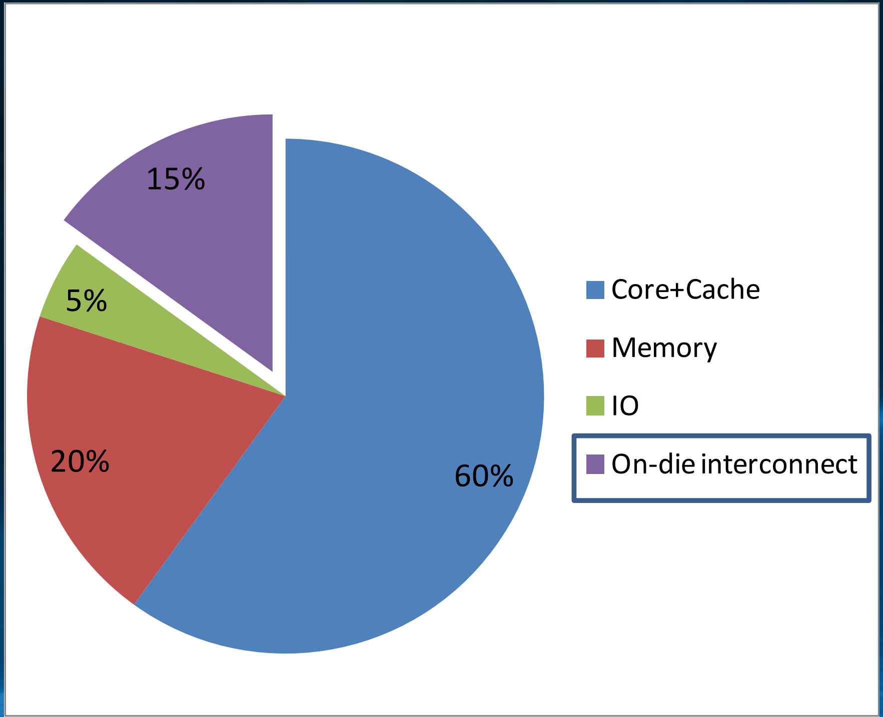
X86 support, OOO, TLBs ~6% of the chip power

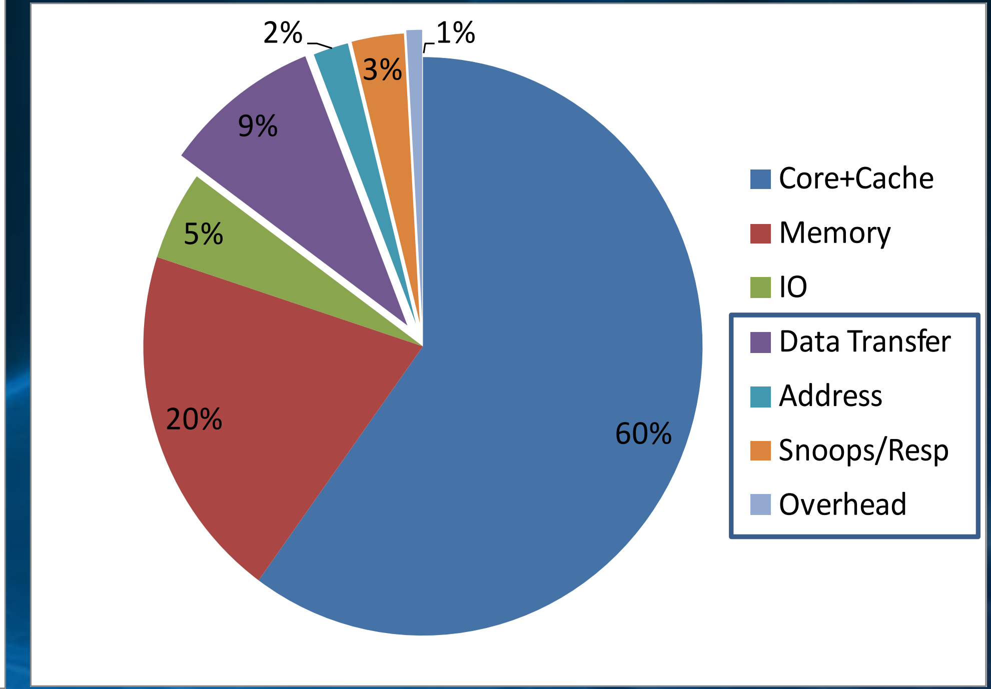
Benefits outweigh the gains from removing them

Myth 3:

IA memory model, Cache Coherence too power hungry

Coherency Power Distribution



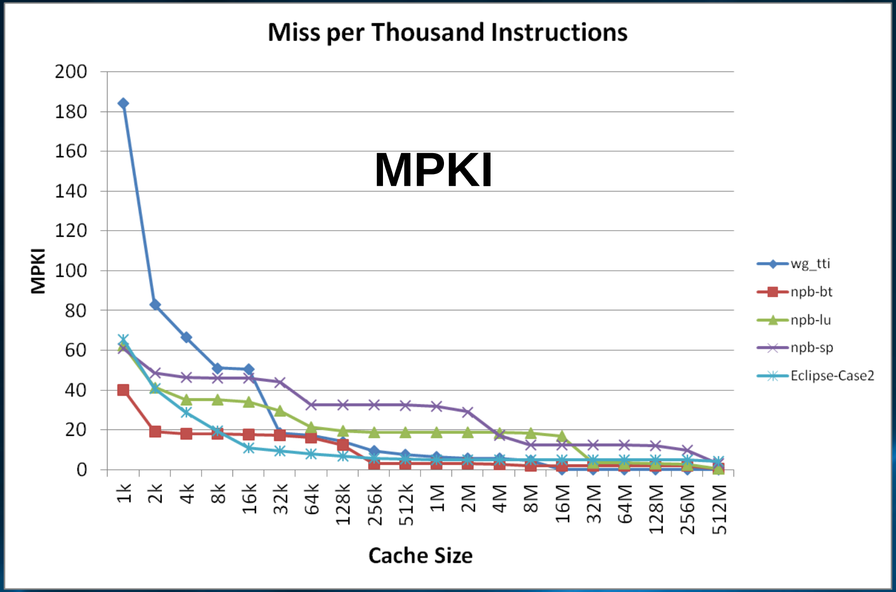


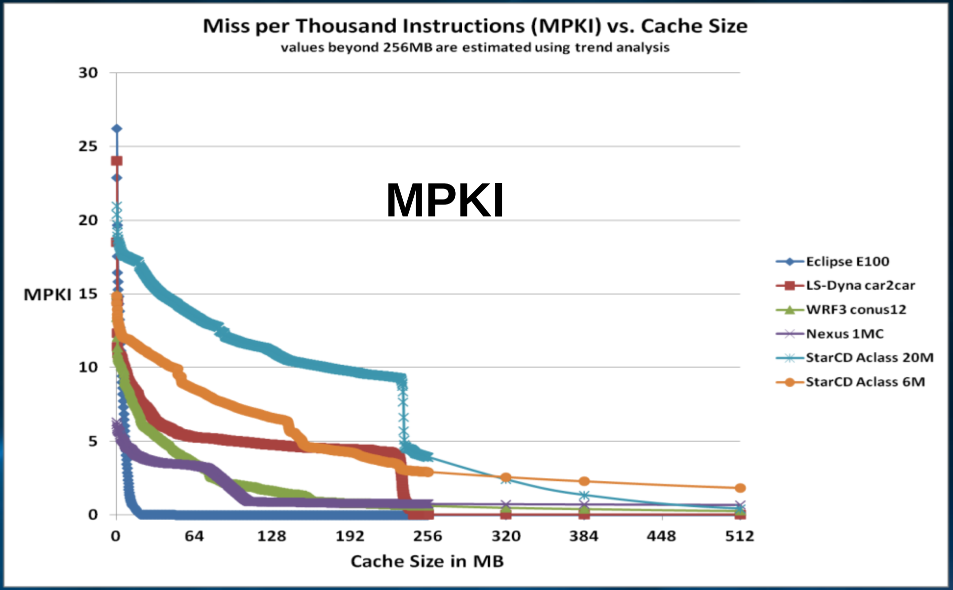
- Typically coherency traffic is 4% of total power
- Programming benefits outweigh the power cost

Myth 4:

Caches don’t help for HPC → They waste power

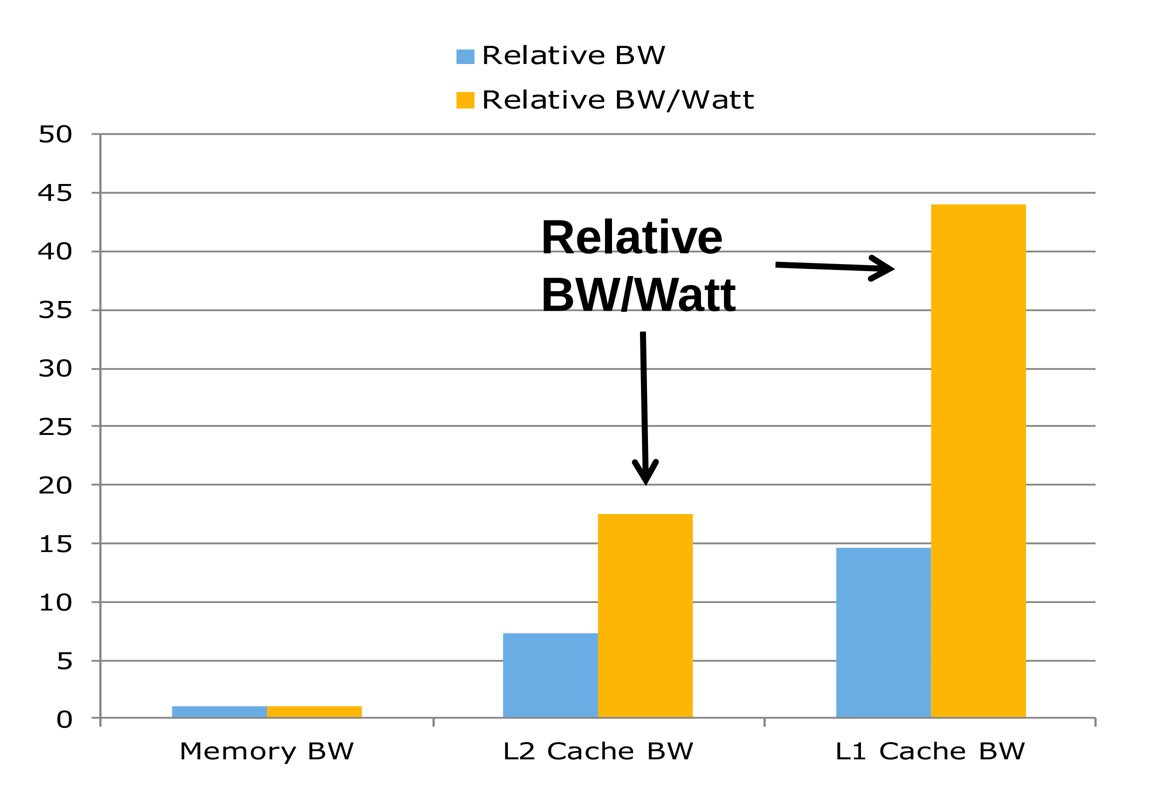
MPKI in HPC Workloads





- Most HPC workloads benefit from caches
- Less than 20 MPKI for 1M-4M caches

Caches save power



- Caches save power since memory communication avoided
- Caches 8x-45x better at BW/Watt compared to memory
- Power break-even point around 11% hit rate (L2 cache)

Challenge 1: Power: Approach Fwd

- Power/performance efficiency with general purpose core with existing programming model and ISA

- Support caches and coherence → Ease of programming.

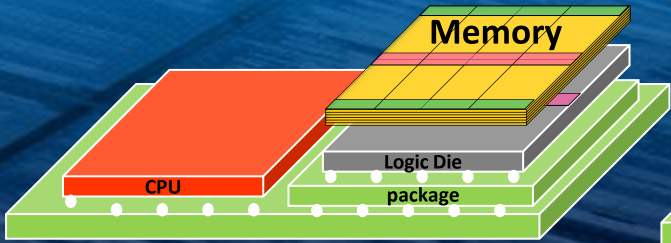
- Drive efficiencies in general purpose core
  - Continued reduction in power across the board
  - Reduce power in computation (execution units)
  - Reduce power in uncore and memory

Challenge 2: Memory: Approach Fwd

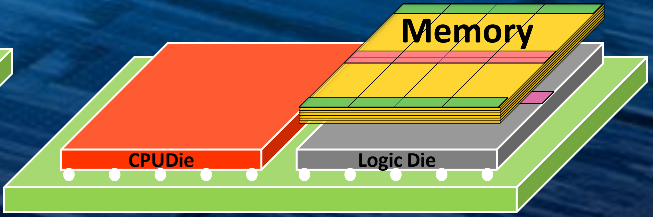
- Significant power consumed in Memory
  - – Need to drive 20 pj/bit to 2-3 pJ/bit
- Balancing BW, capacity and power is hard problem

- More hierarchical memories
- Progressively more integration

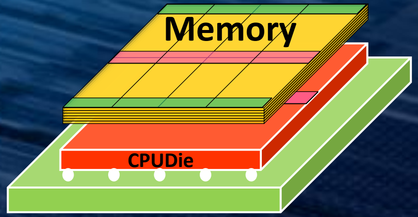
Multi-package Usage



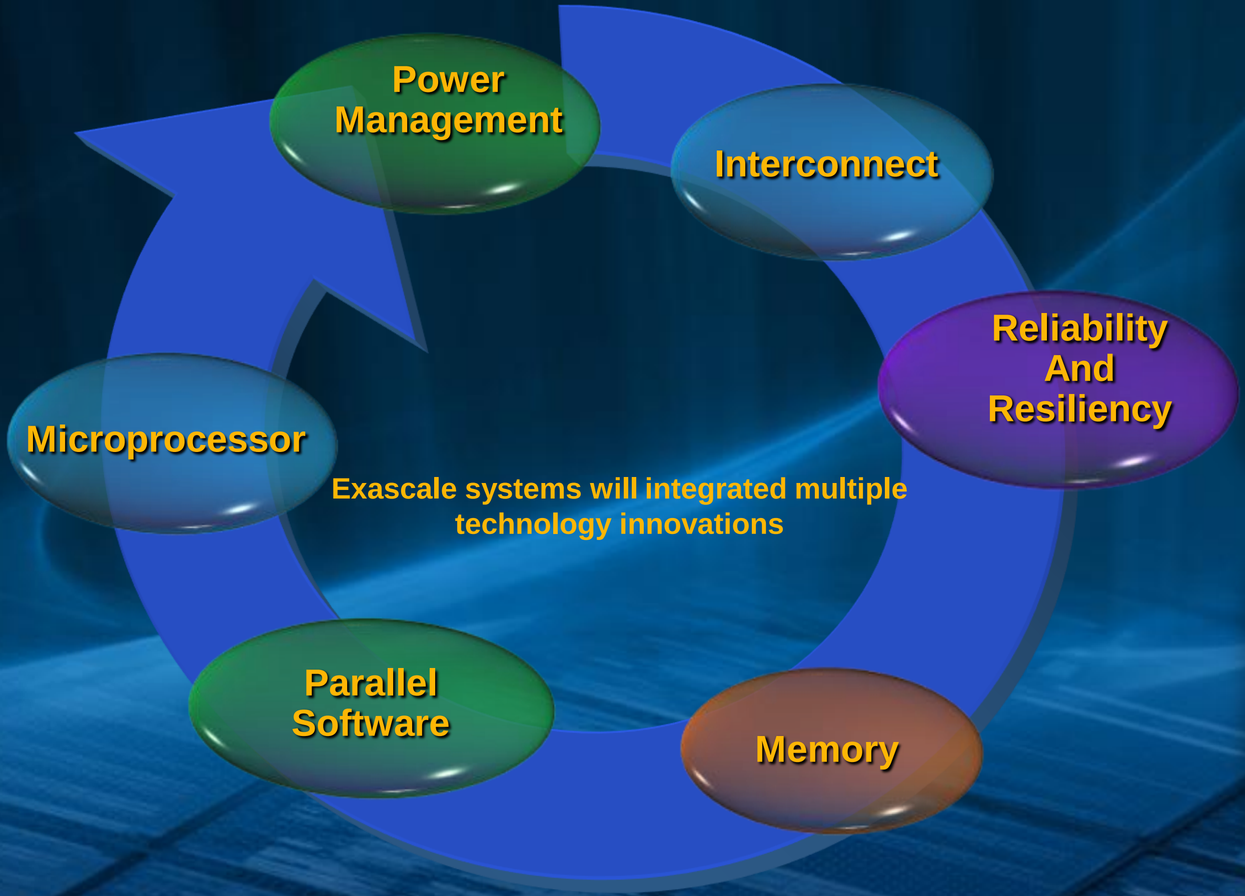
Multi-chip Package Usage



Direct Attach Usage



Exascale – System Level Challenge
A Multi-disciplinary Approach Is Necessary



Exascale systems will integrated multiple technology innovations

Silicon Photonics

Programmability

Summary

Many challenges to reach Exascale → Exciting times

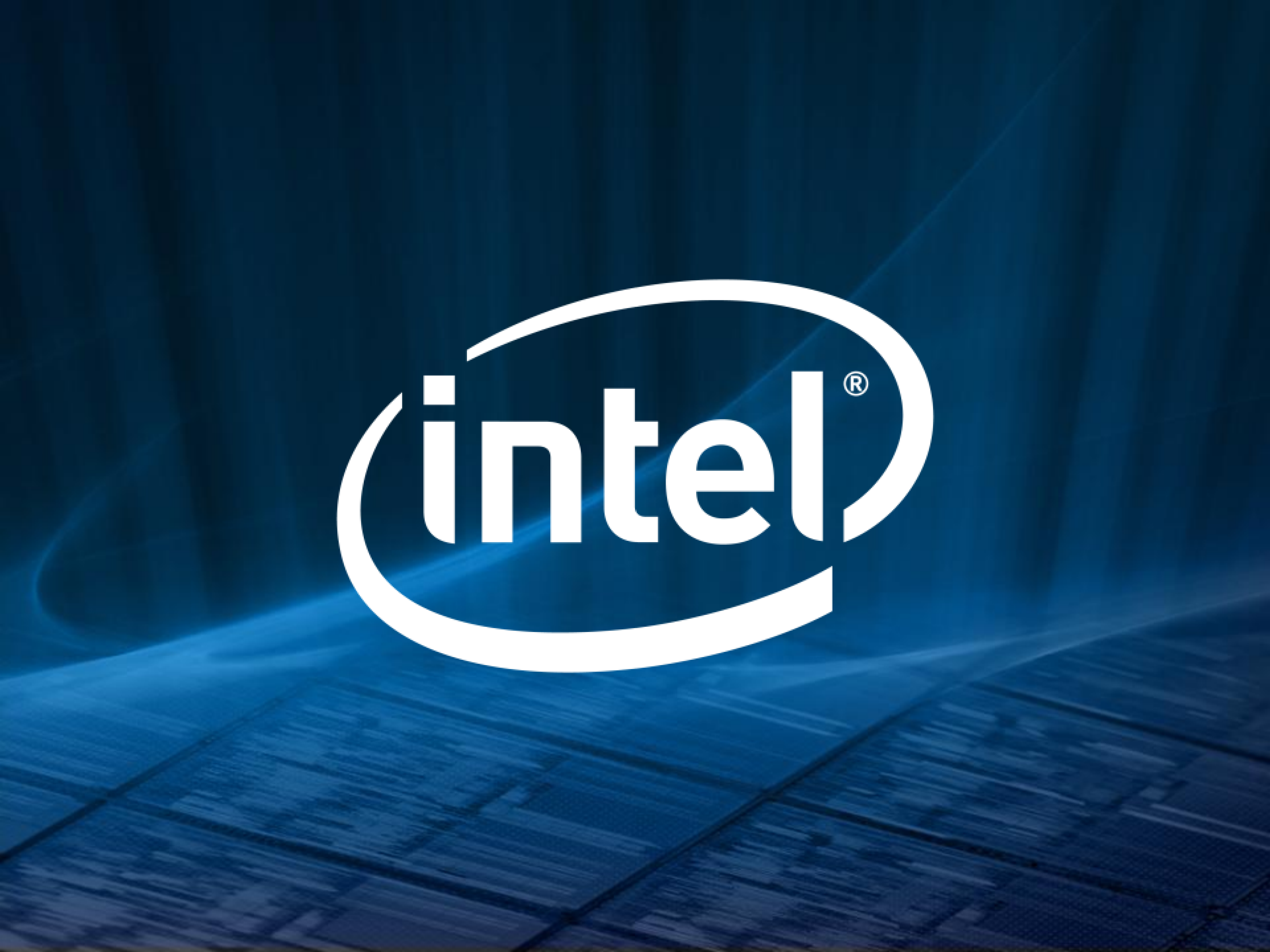
Important that ensuing innovations have broader applicability

Exascale efficiencies within reach w/ general purpose cores

- Without changing programming model or ISA
- Gap ~2x to Exascale pJ/op → 7-8 years to bridge that

More integration over time

- Reduces power, increases reliability, increases scalability


intel®

Backup