Technology challenges and trends over the next decade (A look through a 2030 crystal ball)

Al Gara

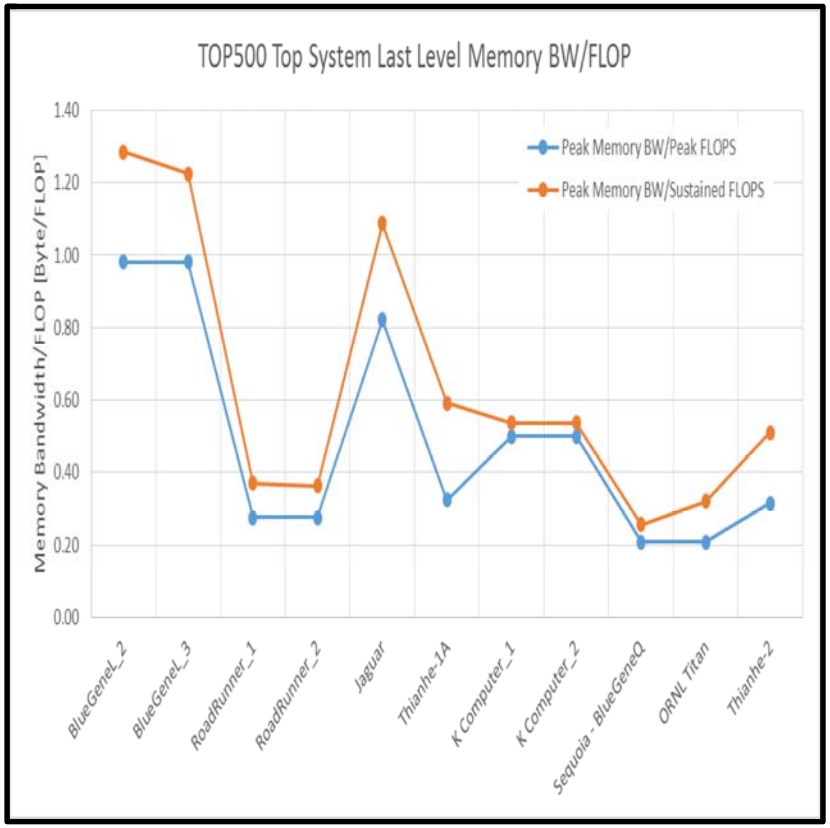
Intel Fellow & Chief HPC System Architect

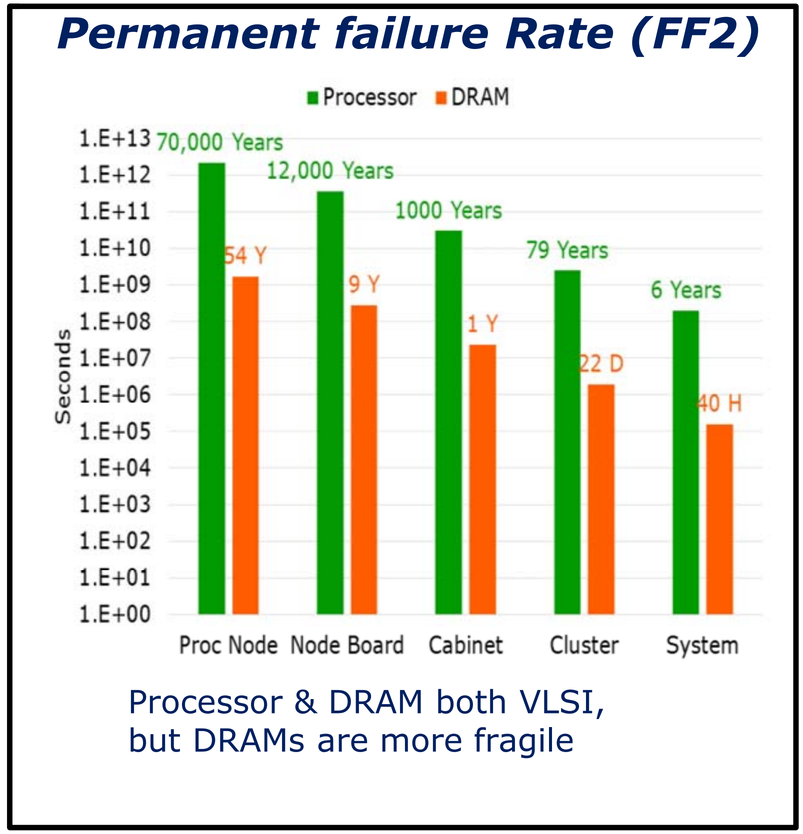
Today’s Focus Areas For Discussion

- Will look at various technologies and the challenges and make an educated guess as to how there will evolve over the next 10-15 years.
  - Memory, Fabric, Compute and Power

- How much did we know in 2000 as compared to today’s reality? (we grew ~ 50,000x in HPL from then till today)

Memory: where is it going…

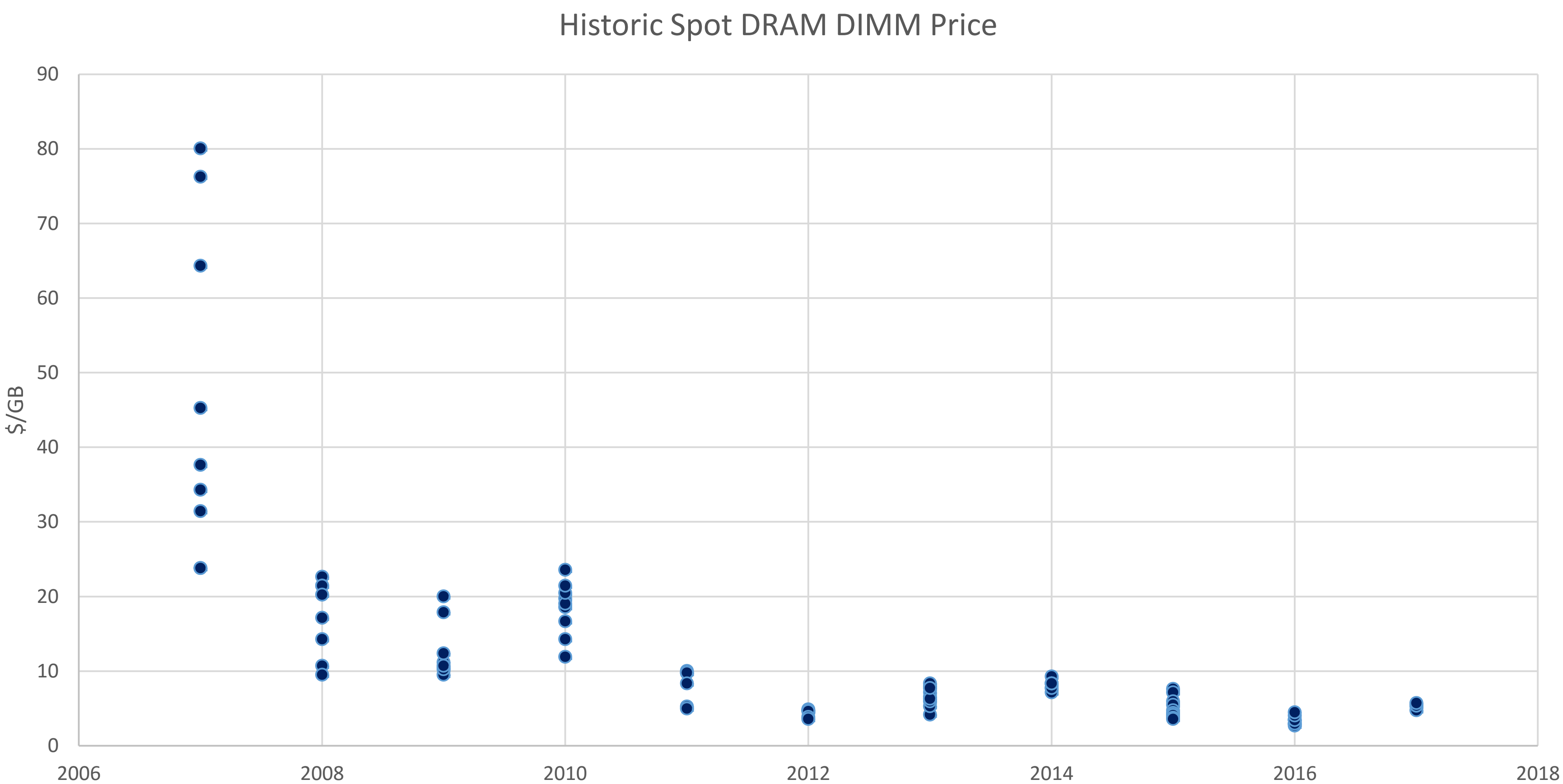




Memory bandwidth has been on a downward trend. Moving to alternatives to DRAM DIMMS is inevitable.

Memory dominates component failure now and into future. Need to carefully consider implications for in-package in this context

Memory: Always a moving target



Spot price now is higher than it was in 2012 for commodity DRAM

Memory: some good, some bad

The Good:

- New packaging and microarchitecture directions have allowed for the future in-package DRAM to keep up with compute
- This looks like a trend we can continue.
- The power improvements are also critical to the future.
- New memory technologies emerging

	DRAM DIMMS	High BW Memory
Cost/ Capacity	1x	1.2x to 1.5x (Short term view is ~ 3x)
Cost/BW	1x	1/10x to 1/40x (now 1/5 to 1/20)
Power/bit	1x	1/2 x to 1/10 x

The Bad:

- The capacity story far less compelling. Slowing improvements in capacity/$
- New emerging technologies are chasing a moving target in terms of BW. Expect a long time before they catch in-package DRAM
- Very good news for I/O in the future with NVMEM running at ~ 1/10 in terms of $/capacity.

The Prognosis :

- Will have a compelling BW story through 2030 but will struggle with slow capacity growth
- Will need to continue to get more performance out of same capacity gen over gen. (Threads etc)
- New memory technology will start further away and move closer but not likely to replace DRAM cells by 2030

Fabric

- Foundational to computing at all scale. Currently a cost problem. Power challenges and transients are difficult to handle quickly.

- Majority of the cost in the fabric is in the optics for long distance communications.
  - Large scale performance for some applications scales with bisection.
  - Historically bisection/perf has been falling
  - Topology choices will continue to trade off repeatable execution against bisection. Progress has been made in dragonfly-like topologies lately

- Silicon photonics will become prime time enabling a steeper drop in $/Gbps

Fabrics: Optics Remains The Challenge

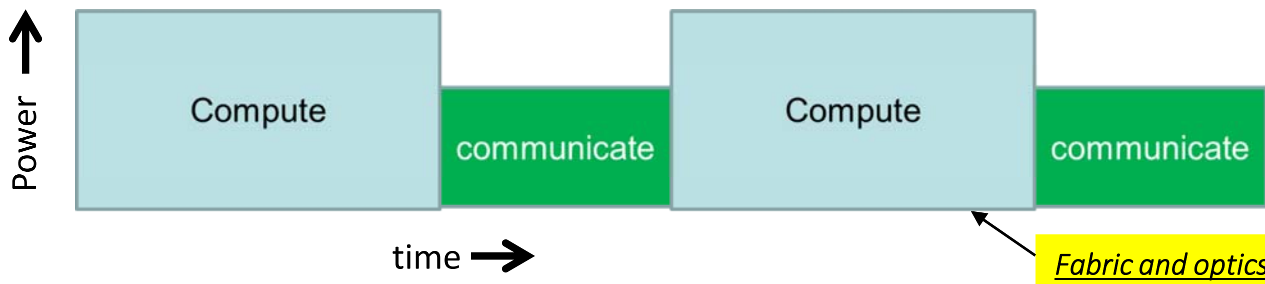
Exascale will not achieve this

Ideal Cost & Power (Requirements @ Exascale)

- 3 Exabits of total BW for ~3MW at $30M
- This translates to 3M 1Tbps links
  - 1W each (1mW/Gbps)
  - $10/link or $0.01/Gbps

20x improvement in power efficiency
50x improvement in cost per bit

Many/most high end applications are time step based



<u>Fabric and optics power most critical when it is NOT used.</u>

Power and performance trade off. Overlapping will drive up system maximum power or cause performance throttling. Timescale to idle power in fabric will determine whether this is hardware or software.

Fabrics: Topology and optical bundle size may matter

Topology	Number of unique Destinations from each rack (N= number of racks)
Dragonfly	~ N^2
Fat Tree	~ N
Torus	4 to 10

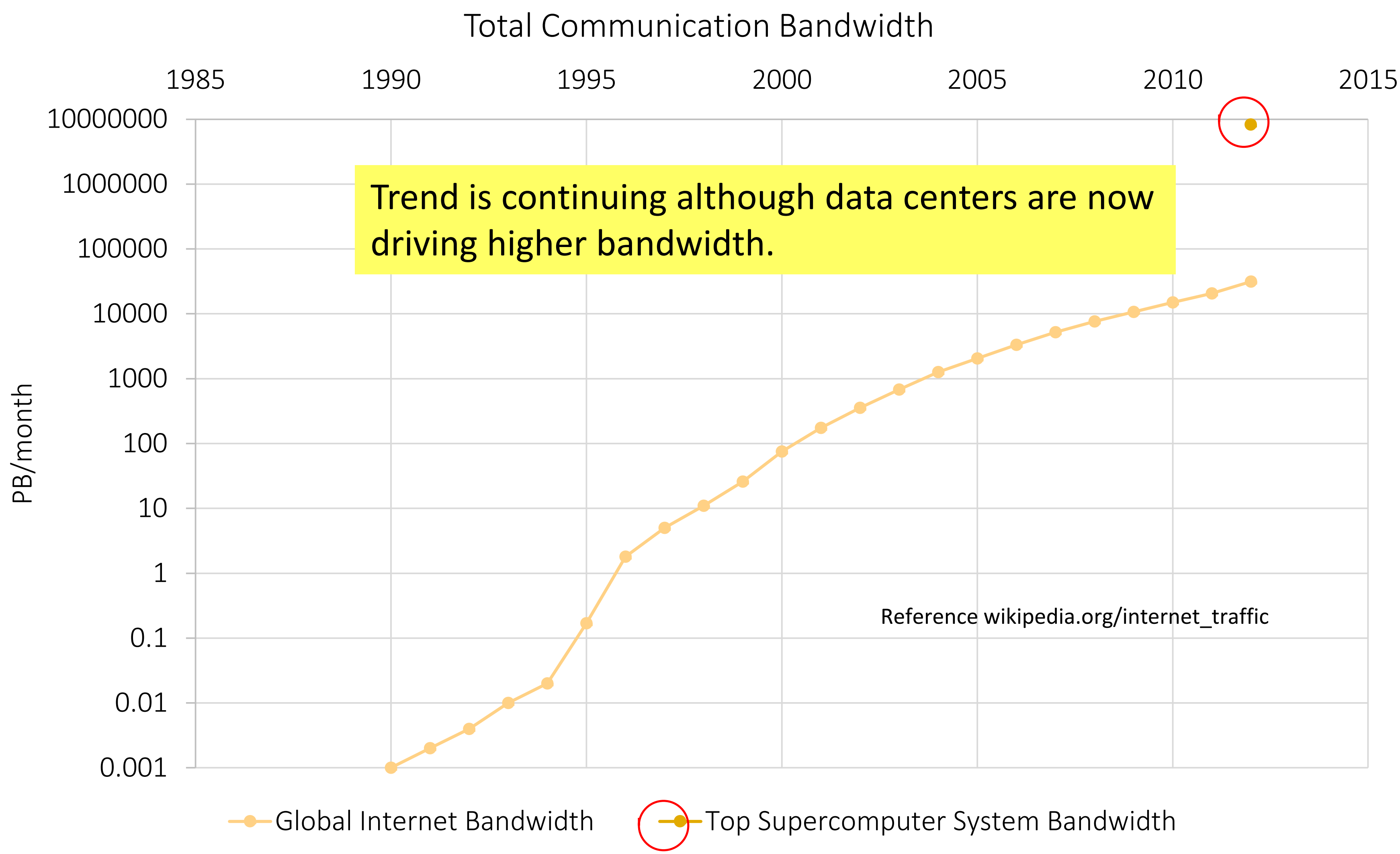
Getting to exceptional cost will require very high BW per fiber to amortize fixed costs connectors, fiber etc.

As BW/$ increases through WDM for example… if $ also raises as BW raises faster this can influence the choice of most cost effective topology.

We are already approaching small numbers of fabric ports in dragonfly topologies for the largest machines.

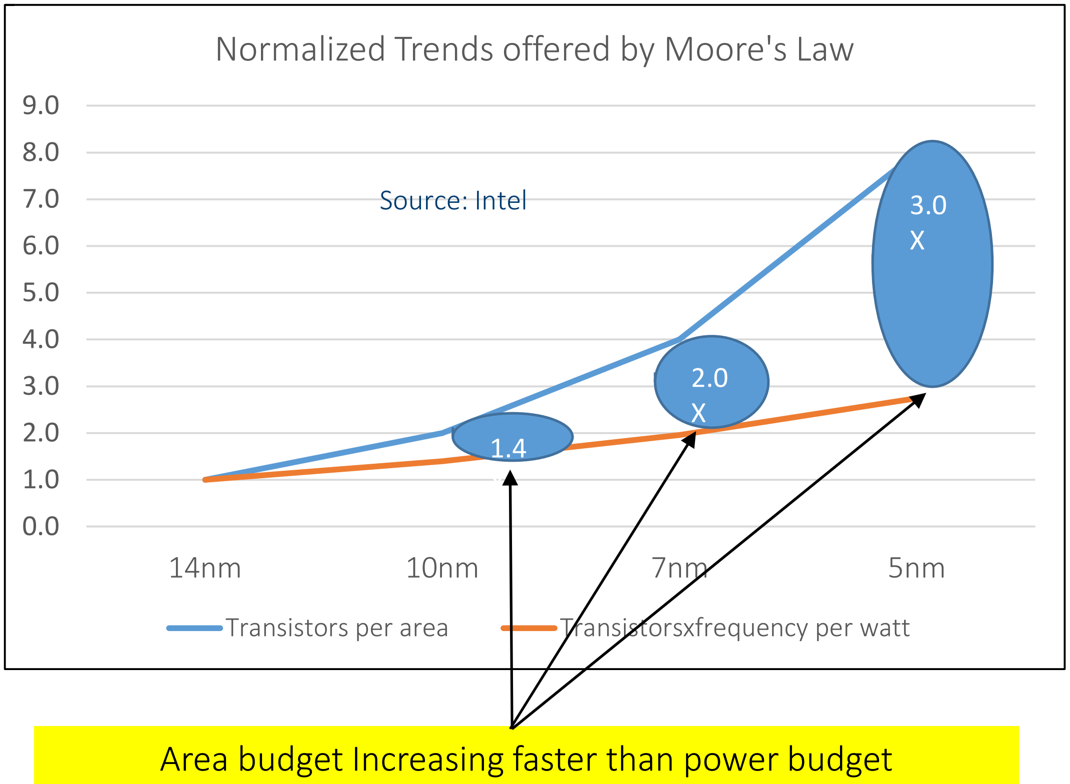
Prognosis: We will take a significant leap forward ~ 2020 (Now looking like 2022) in $/Gbps followed by strong price improvement through 2030. But 1000x would be a big stretch. 20x to 50x more likely.

Optical BW of Single Largest Supercomputer Swamps Total Worldwide Internet BW



As data center wakes up to BW needs, we will develop co-travelers in this.

Process Technology Trends Will Lead To On-chip Specialization



Process Technology Scaling Trends

- Area/Cost per transistor decreasing faster than Energy/transistor
- I/O requirements not improving as fast as we can support performance growth
- Implies that we will be able to build more transistors than we want to operate simultaneously. Address perf/W
- Architecture will become more specialized as a result → different functions or algorithms will use different transistors to operate most efficiently, and transistors not in use will be shut off

Processor: some good some bad, some disruptive

The Good:

- HPC is a fast growing business strategic to both commerce and governments.
- Investments in HPC including processor development have been strong and appear to be accelerating.
- This will enable innovation at the processor level in ways that we did not have on our radar screen 5 years ago.
- Silicon scaling is still moving albeit at a slowing pace. It always accounted for at most half the performance improvement over time.

The Bad:

- Getting 100x by 2030 more compute at same power and cost hard to see path to that with silicon scaling slowing.
- All paths lead to exploiting concurrency new ways. Frequencies not going up any time soon.

The Disruptive :

- New forms of compute could dramatically change the game for some algorithms.

Processor: the disruptive potential

Recent explosion in exploring new forms of compute:

- Neuromorphic is probably most promising but…
  -- still working to find broad class of algorithms.

- Activities into many other areas.. Quantum, low bit arithmetic, information encoding,

- Two paths for technology…

  1) Special purpose device
  2) Augmentation of "traditional compute"

- New forms of compute offer disruptions in both performance and energy efficiency.

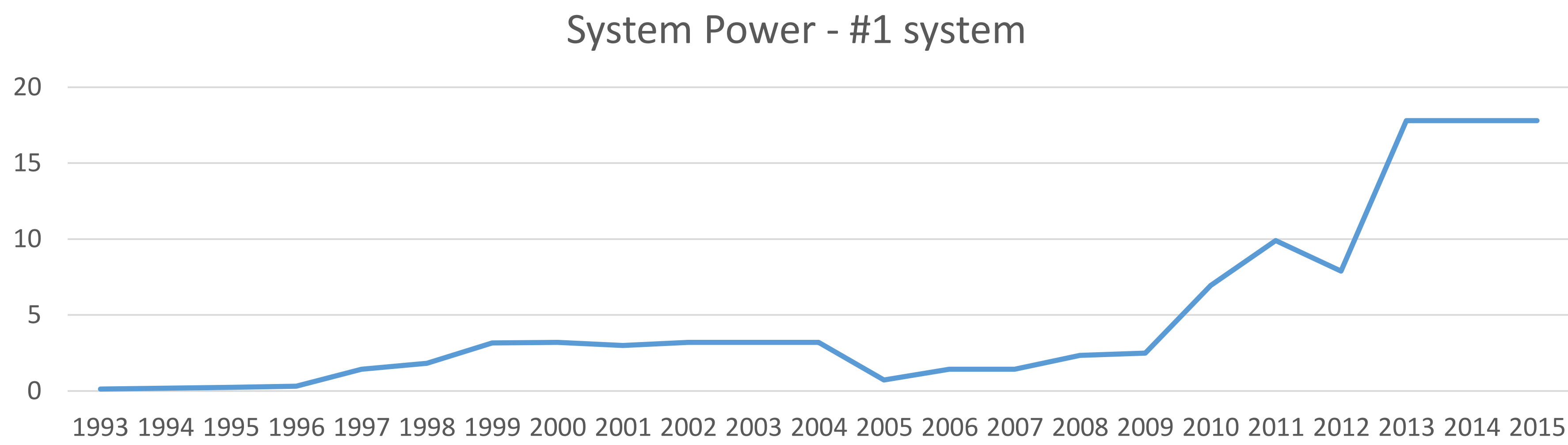
- How perf/W tracks will be critical to integration into conventional systems. Can't have an accelerator in a general purpose machine that drives 10x the power when it is used. That is a special purpose machine.

- Will start to see new technologies and techniques that will be focused on information processing per watt

The Prognosis :

- New forms of compute will dramatically change the game for some algorithms. Expect that integration into traditional systems will be fastest path to growth of new technologies. This will require some compromises and the merging of very different technologies can be difficult.
- Quantum and Neuromorphic will likely be used as special purpose accelerators before 2030.

Power: The long term challenge



We have had a recent growth in power. Where will this really go in 2030?

Looking at this from a simple TCO perspective….

Using $1M/MW-year current rule of thumb.

Assume this drops to $0.5M/MW-year in 2030

If we want to spend ~ 50% of the system total cost on power over 5 years then

~$200M machine cost = <u>80MW</u> * 5 years * $0.5M/MW-year

Evolution of Systems Over Time
Crystal ball version…

Power will REALLY be the challenge.

	Sequoia	2018	2022	2026	2030
Peak Flops	20 P	~ 200 P	2 EF	20 EF	100 EF
Perf (general)	1x	10X	100X	1000x	5,000x
Perf specialized	1x	10x	200x	5000x	50,000x
Memory cap (high performance)	1.6P PB	2.5 PB	8 PB	30 PB	60 PB
Mem BW	2.5 PB/s	30 PB/s	200 PB/s	1500 PB/s	5,000 PB/s
Fab Bisection	50 TB	1 PB/s	10 PB/s	50 PB/s	100 PB/s
Fab Injection	2 PB/s	3 PB/s	30 PB/s	150 PB/s	300 PB/s
Power	8 MW	18MW	25MW (Probably 40MW now)	50MW (65 MW?)	80MW (100MW?)

With growth in power we will see continued growth in systems for some time.
Expect a financial power cap to hit before 2030. Considerable slowing after that time.
1000x through 2030. (as compared to our previous 50,000x)

Legal Disclaimer

Today's presentations contain forward-looking statements. All statements made that are not historical facts are subject to a number of risks and uncertainties, and actual results may differ materially.

Intel does not control or audit the design or implementation of third party benchmarks or Web sites referenced in this document. Intel encourages all of its customers to visit the referenced Web sites or others where similar performance benchmarks are reported and confirm whether the referenced benchmarks are accurate and reflect performance of systems available for purchase.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See www.intel.com/products/processor_number for details.

Intel, processors, chipsets, and desktop boards may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Implications of Variations In Power