Future Architectural Directions
EE HPC WG Workshop

Al Gara
Intel Fellow & Intel® Xeon Phi™ Chief Architect

System Scaling

- Power in fabric is mostly an issue when fabric is idle.
  - must be able to move idle-> active in nsecs to enable hardware management.
- There is no substitute for bisection bandwidth.
  - FFT, random access, all-to-all
  - "Big" nodes, "small" nodes…. Doesn't matter.
  - Optics costs still dominate cost and are the limiting factor currently.
- Heavyweight fabrics have become the standard.
  - This will likely present a significant problem ala POSIX with NVM storage..
  - If optics costs plummet… we may have a hard time exploiting

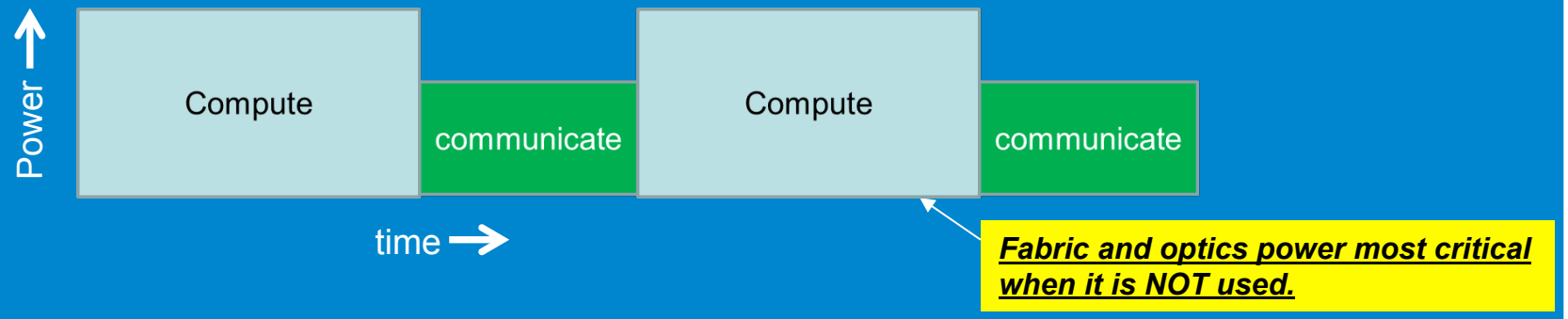
Fabrics: Optics Remains The Challenge

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



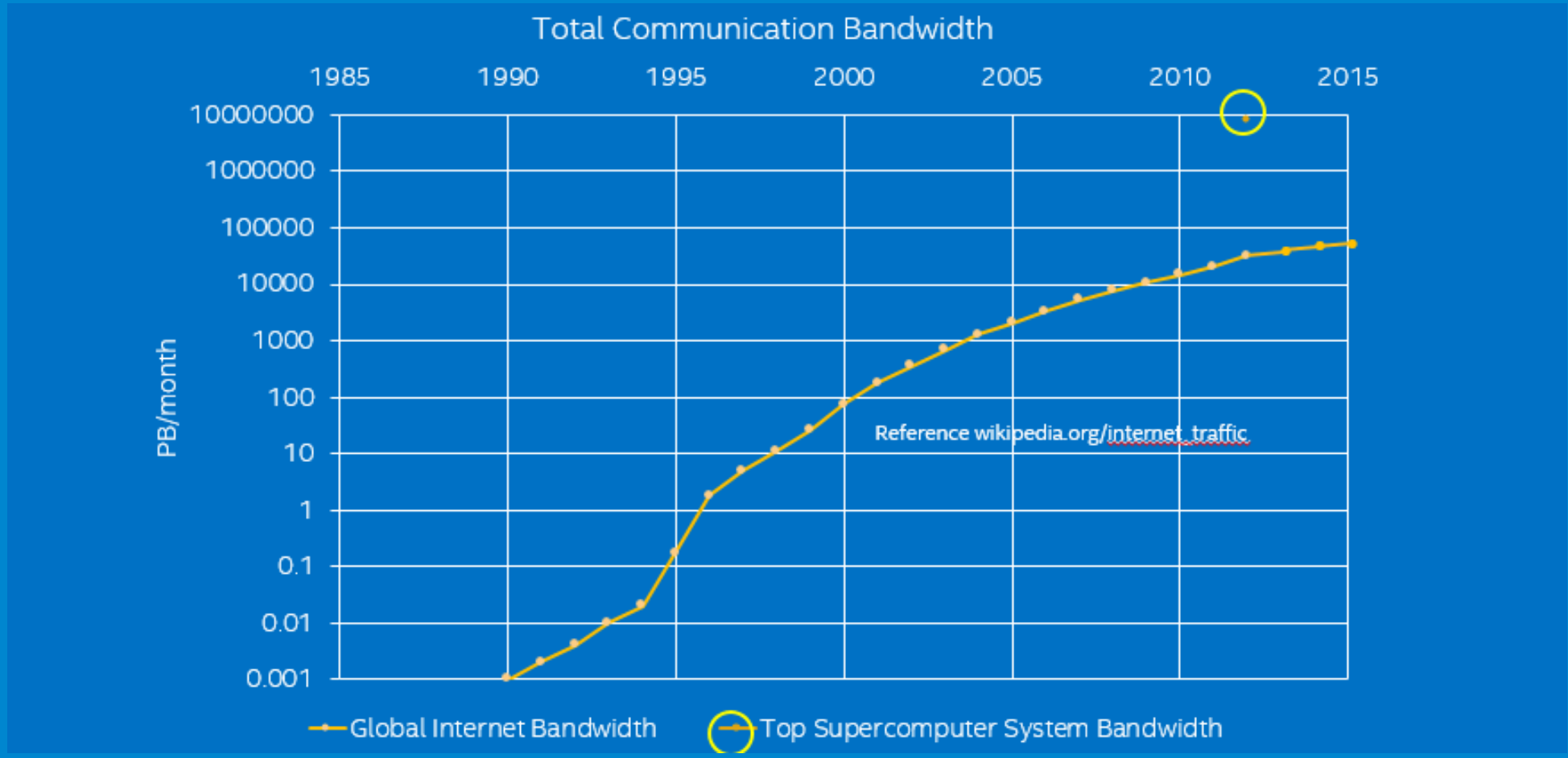
Power and performance trade off. Overlapping will drive up system maximum power or cause performance throttling

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 minimum size limits in dragonfly topologies for the largest machines.

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.

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.

NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL® PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. INTEL PRODUCTS ARE NOT INTENDED FOR USE IN MEDICAL, LIFE SAVING, OR LIFE SUSTAINING APPLICATIONS.

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.

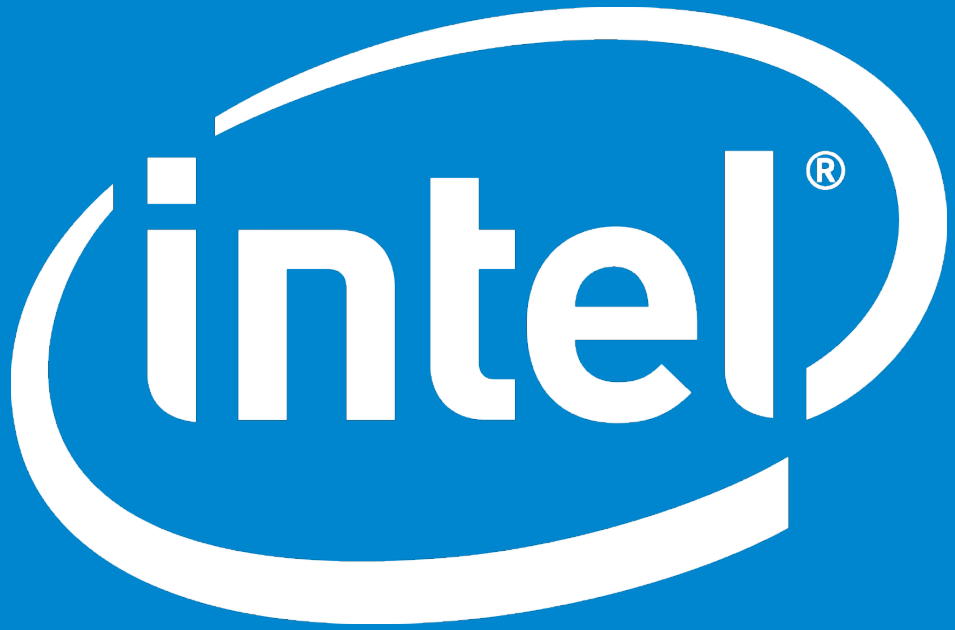
Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See www.intel.com/products/processor_number for details.

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.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit http://www.intel.com/performance.

Configurations: {describe config + what test used + who did the testing, if other than Intel}.


intel