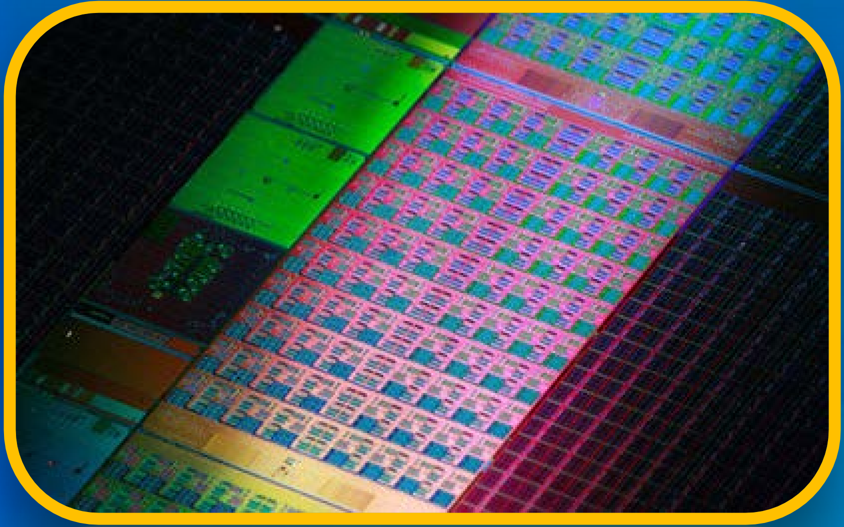
A Few Provocative Trends in Parallel Computing

Justin Rattner
Chief Technology Officer
Intel Corporation

This research was, in part, funded by the U.S. Government. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. Other names and brands may be claimed as the property of others.

Trends in Parallel Computing

Technology

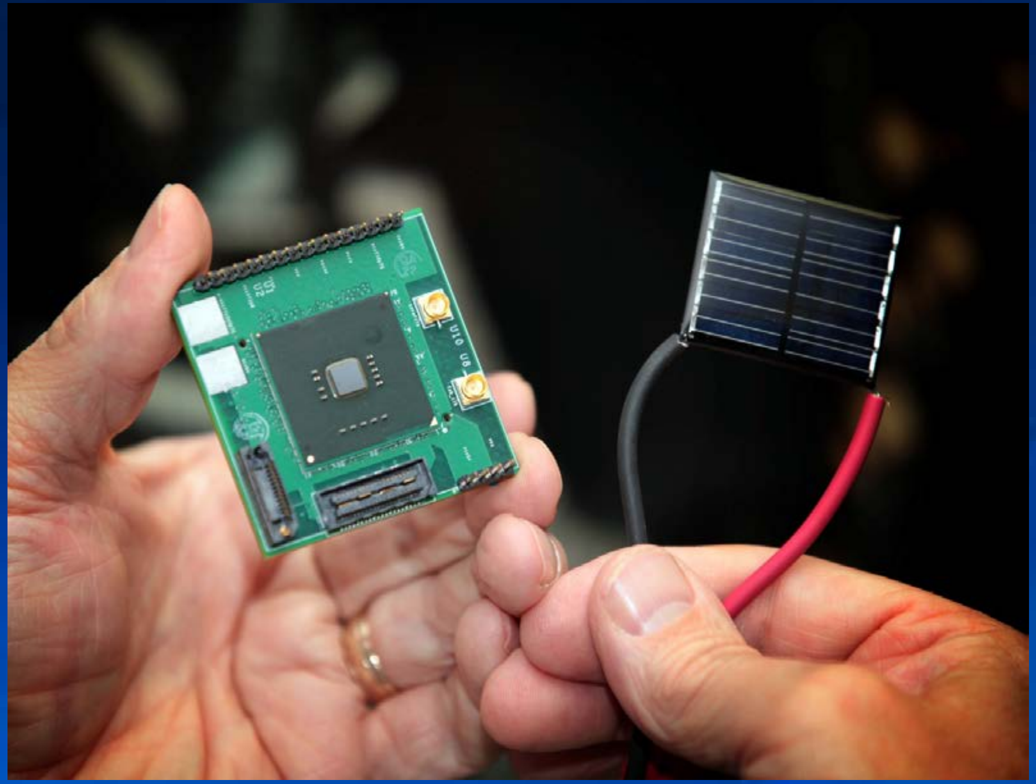


Tools

Business

Near Threshold Voltage Logic

- Claremont: an experimental NTV processor
  - An actual Pentium-era microarchitecture
  - Plugs into old Pentium motherboards
  - First demoed at Fall IDF 2011
- First processor to demonstrate the benefits of NTV circuits for *logic design*
- Key characteristics
  - Built in low-leakage 32nm SoC technology
  - Operates from 280mV@3MHz to 1.2V@915MHz
  - Runs both Linux and Windows

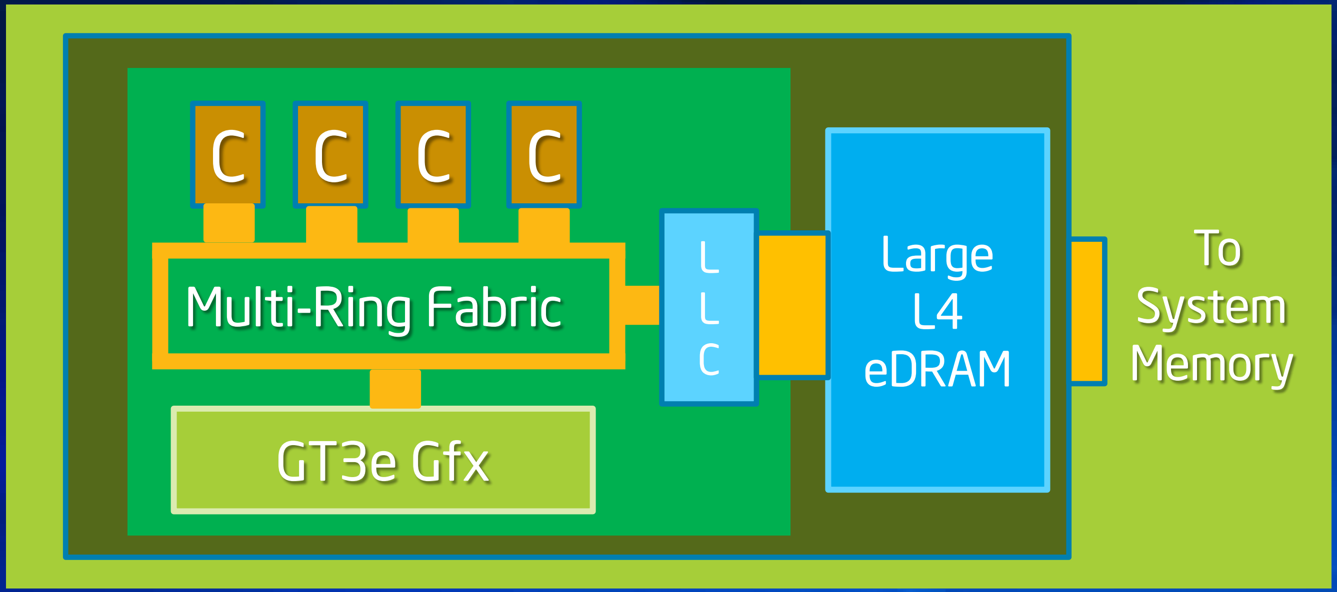


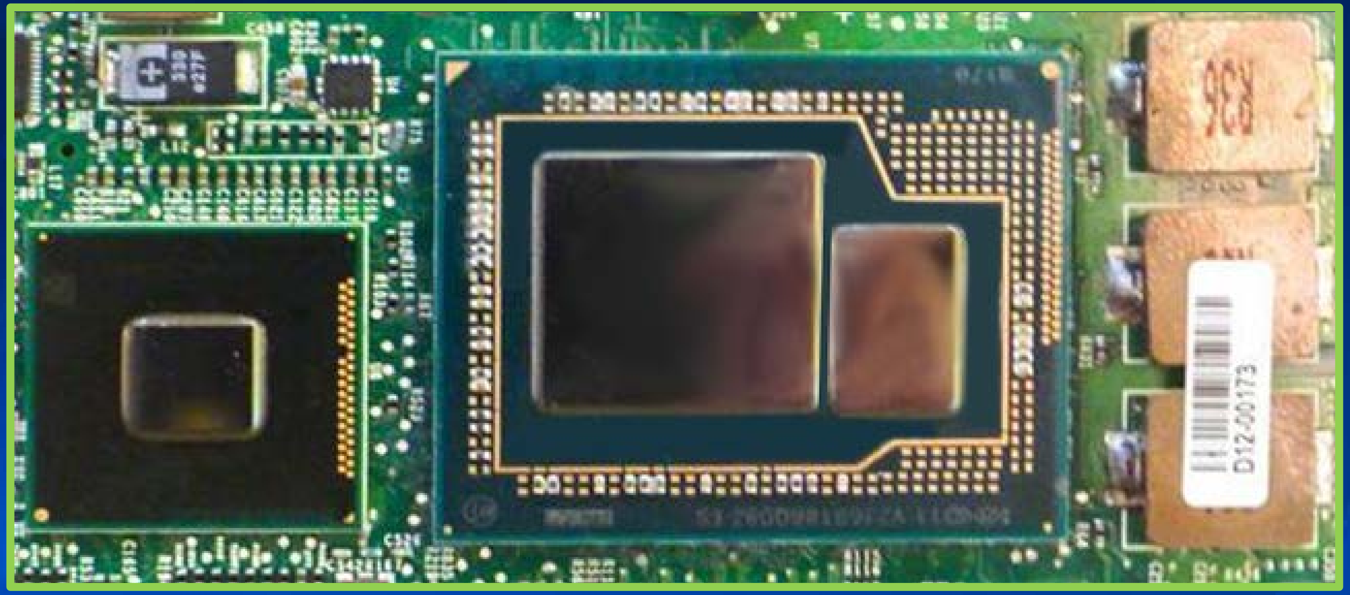
S. Vangal et al, A 280mV-to-1.2V Wide-Operating-Range IA-32 Processor in 32nm CMOS, ISSCC 2012

4.7x better energy efficiency in NTV mode

Co-Packaged Embedded DRAM

Intel® 4th Gen Core™ Processor Family with GT3e Graphics

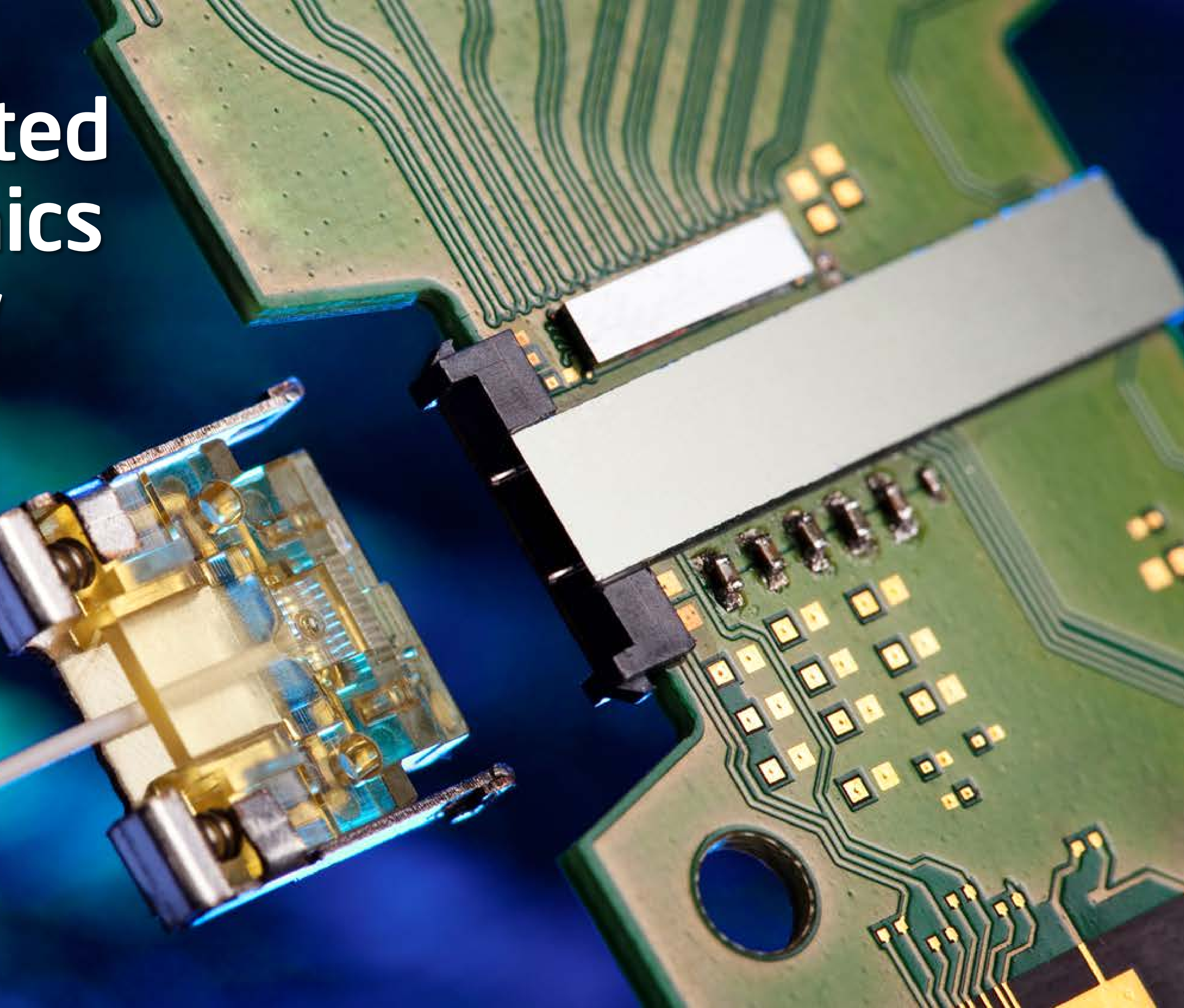




- Large L4 cache memory (size and speed TBA)
- Based on Intel's HiK 22 nm Tri-gate logic technology
- Enhanced to include a quality embedded DRAM (eDRAM) capability
- Designed to be built in high volume at low cost
- Future: other co-packaged processor/memory designs underway

Intel's Integrated Silicon Photonics Technology

- On-die lasers
- On-die detectors
- On-die modulators
- On-die multiplexers
- On-die demultiplexers
- Tested at the wafer level



Photonic Rack Architecture

Photonics in the Rack
Cable reduction
Bandwidth increase

Flexible Topologies
Resilient rings
Separable network & storage

Distributed Switching
Xeon & Atom Fabric
Aggregation

Better TCO
Memory, NICs and Storage span multiple CPU generations
Better serviceability

Thermal Efficiencies
Flexible component placement

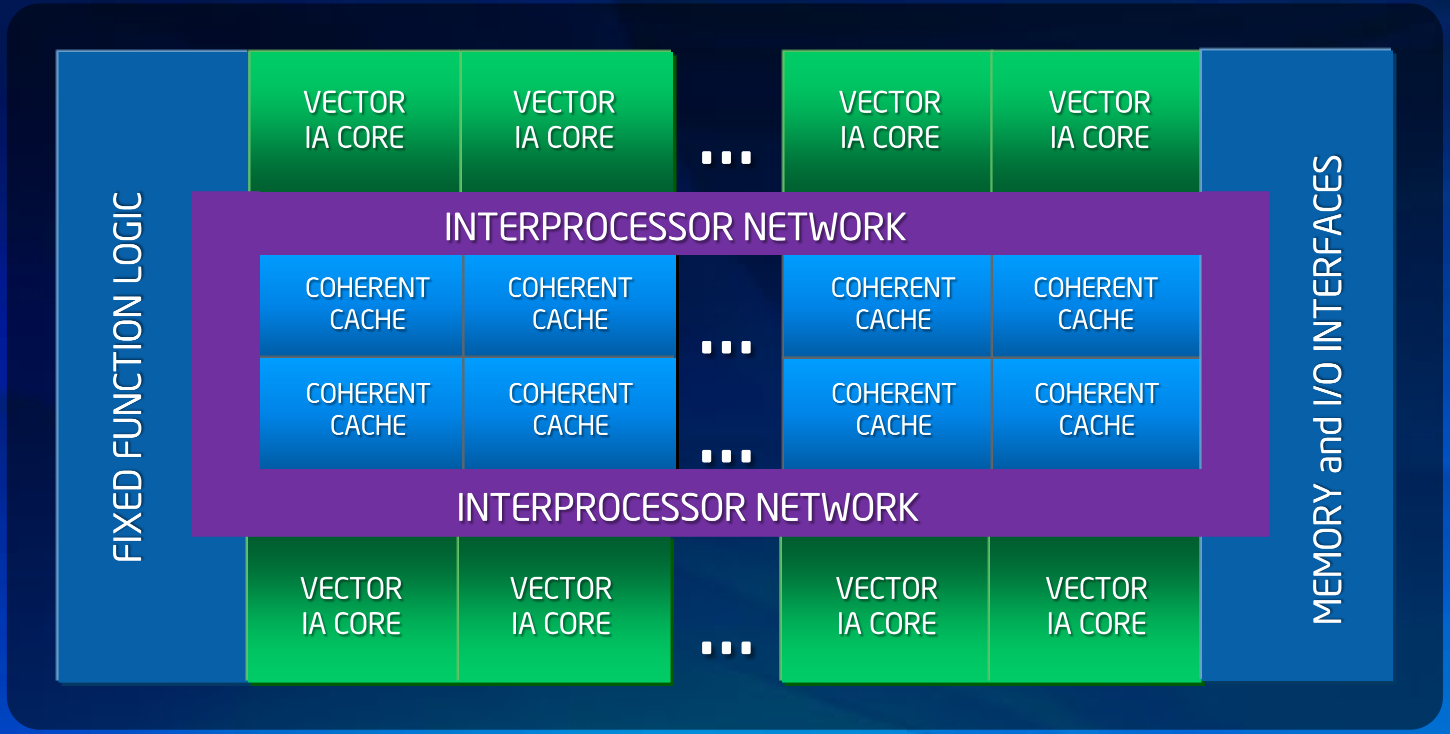
RAS
Redundant fabrics, storage, memory
Resources shared among CPUs

Photonic Rack Architectures: Accelerating Server Disaggregation

The Advent of HPC Optimized Processors

Intel's Xeon Phi Processor Featuring MIC Architecture

MIC Architecture



- Familiar multi-core + cache design
- Using Intel's HiK 22nm Tri-gate process
- 60 small, VPU-enhanced, IA cores
- Fully cache-coherent design
- >1 TFLOPS Linpack @ >3.0 GF/Watt
- >80% efficiency at 1.3GHz

Demonstrates a vector-enhanced, small-core, general-purpose architecture is HPC competitive

Tool Trends

Making Parallelism Accessible from the Browser

Parallel JavaScript

- Make multi-core, SIMD and GPU resources accessible
- Preserves the safety and security the Web requires
- Maintains web developers comfortable programming model
- Closely ties into the HTML5 ecosystem
- Targets web apps that feature computer vision, gaming, and image processing, and video editing

Parallel Languages and Frameworks

Which one is solving the parallel programming problem?

- Parallel languages are fascinating, but a small percentage ever get used by real applications programmers.
- What's working are parallel frameworks on top of today's most popular languages.

APIs tuned to needs of the application programmer while hiding the details of the parallelism

How Do We Generalize this Approach?

In collaboration with UC Berkeley ParLab

Define a Pattern Language of Parallel Application Programming

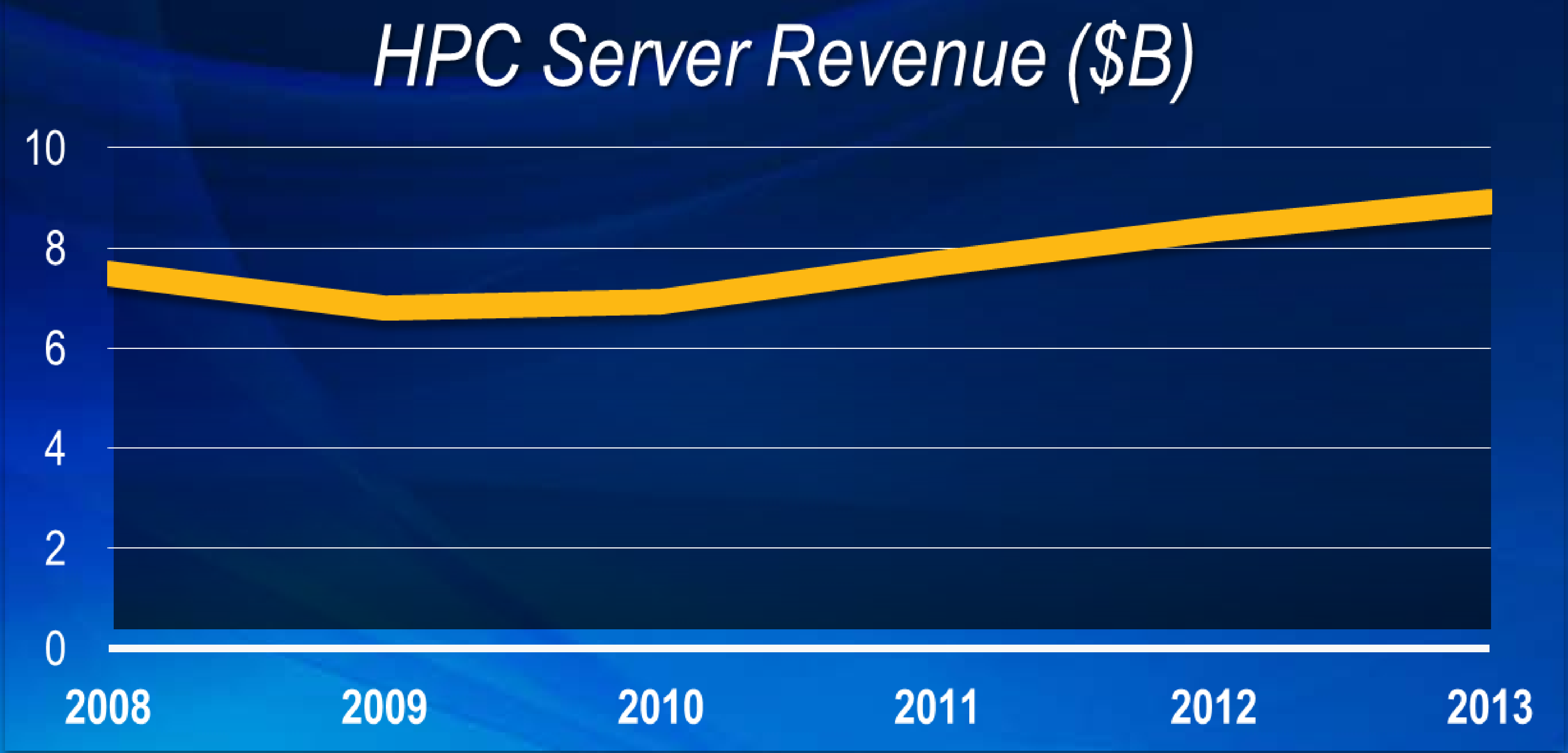
Software Architectures Expressed in a Framework

Framework Optimized with Software TransformationTools

Goal: Enable "domain experts" to write efficient parallel applications

Business Trends

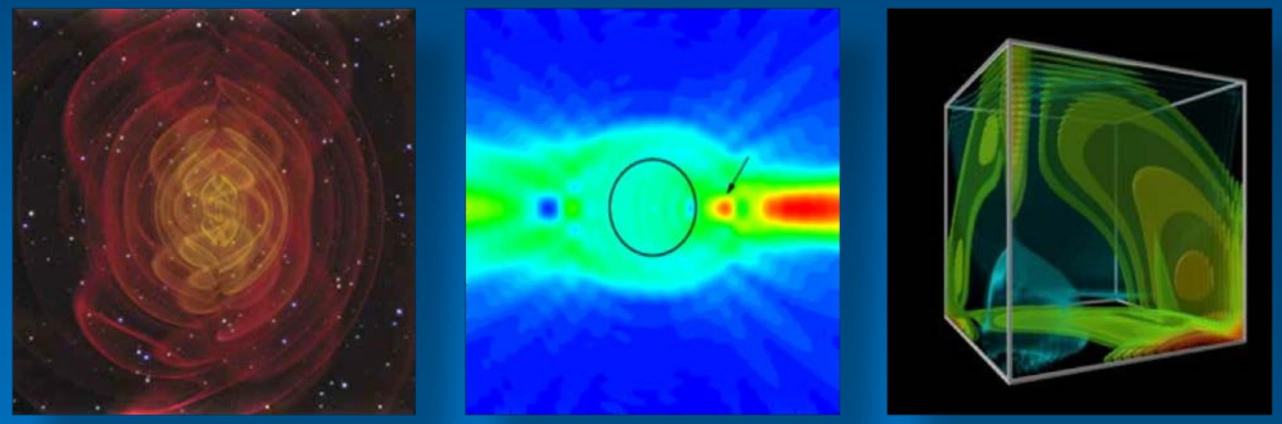
HPC Industry Growth Outlook

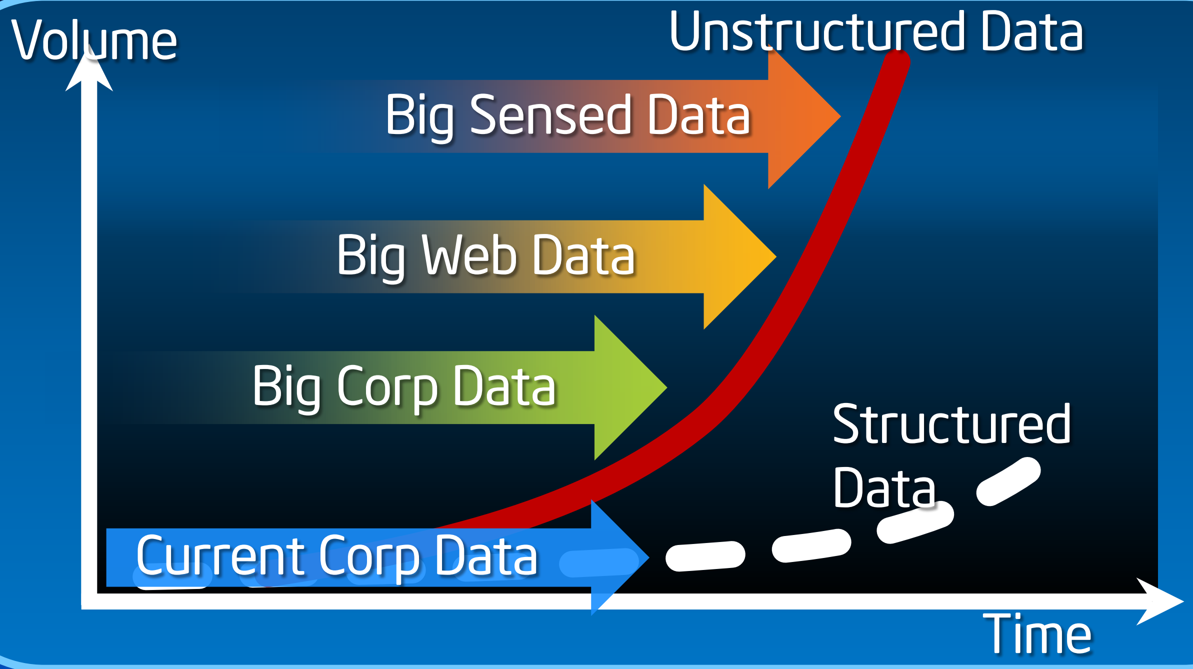


Source: InterSect 360 Research, 2009

What Will it Take to Significantly Accelerate Growth?

Is Big Data Good for Big Compute?





Big Compute

- Well-known applications in science and engineering with slow growth
- Computation on large matrices benefits from more concurrency

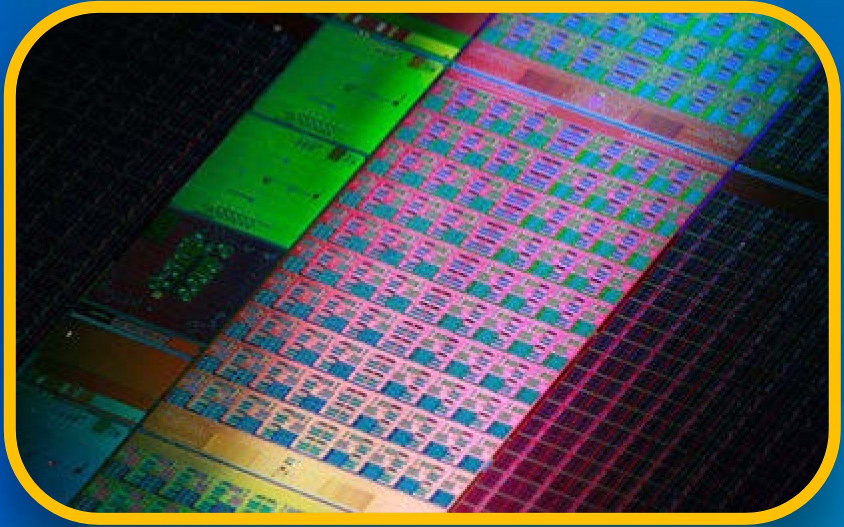
Big Data

- New high-growth applications expected to grow by ~50X this decade
- Machine learning algorithms benefit from increasing parallelism

Source: IDC Enterprise Predictions January2011.

Trends in Parallel Computing

Technology



Tools

Business