Runnemede
(Intel, Reservoir, ETI, UIUC, UD, UCSD)

UHPC Research Technical Approach

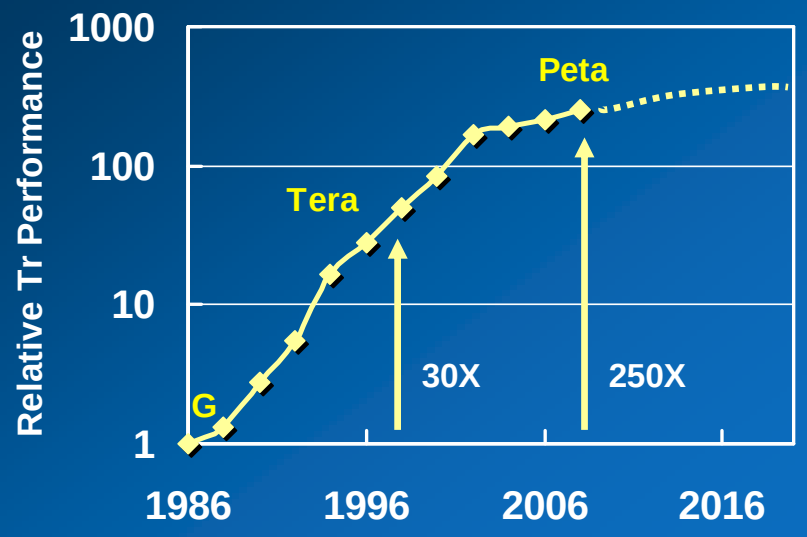
Roger Golliver
Dave Dunning

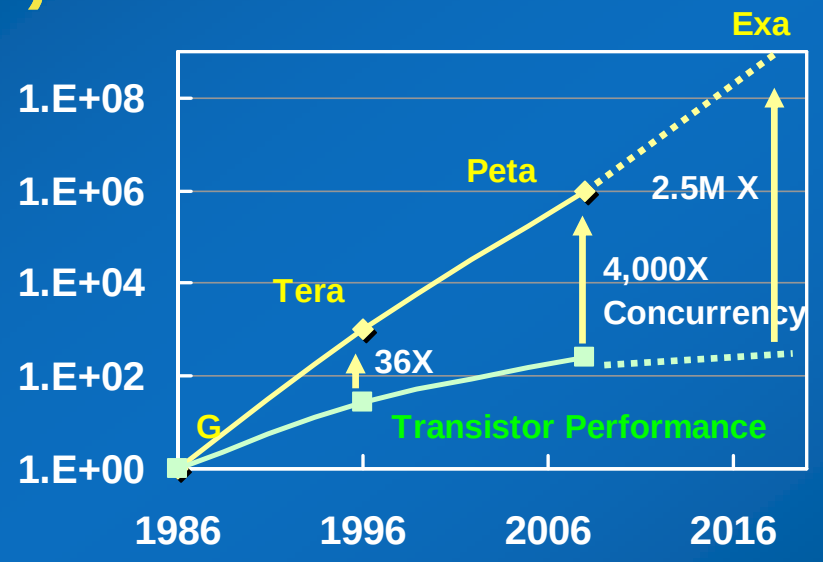
Intel Corp.

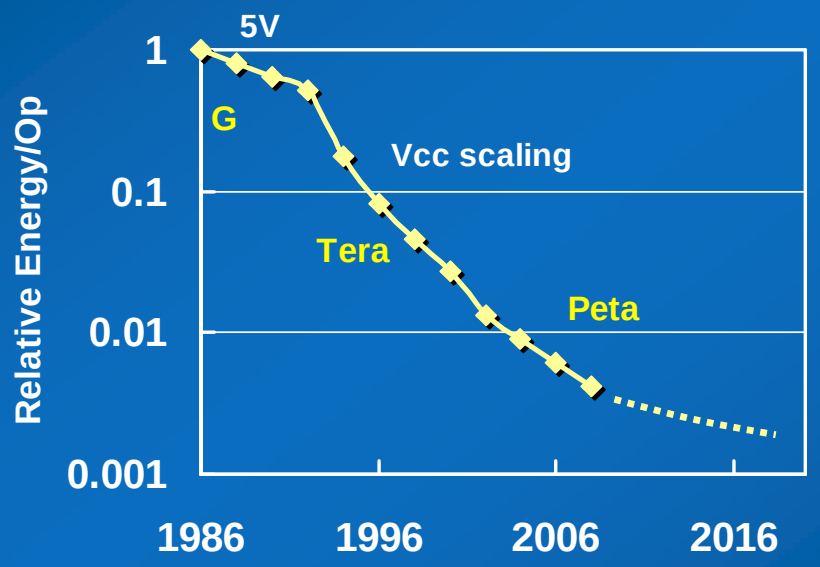
Oct 27, 2010

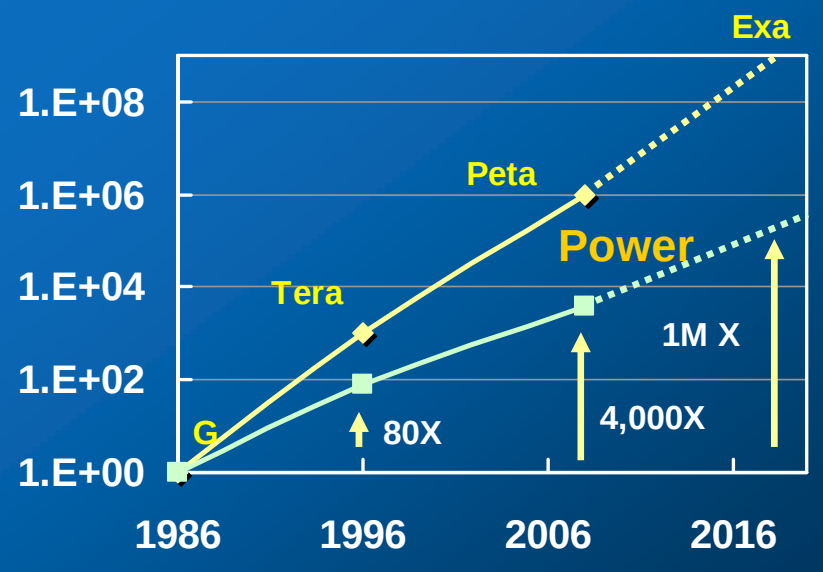
Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

From Giga to Exa, via Tera & Peta





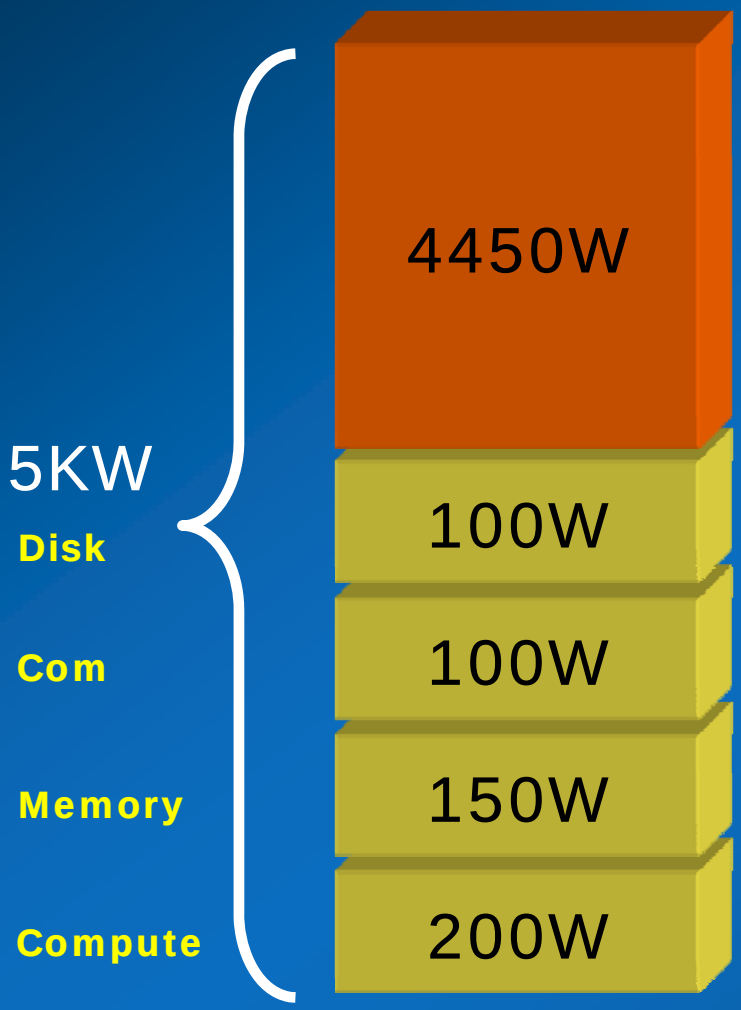




Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Building with Today's Technology

TFLOP Machine today



Decode and control
Translations
...etc
Power supply losses
Cooling...etc

10TB disk @ 1TB/disk @10W

100pJ com per FLOP

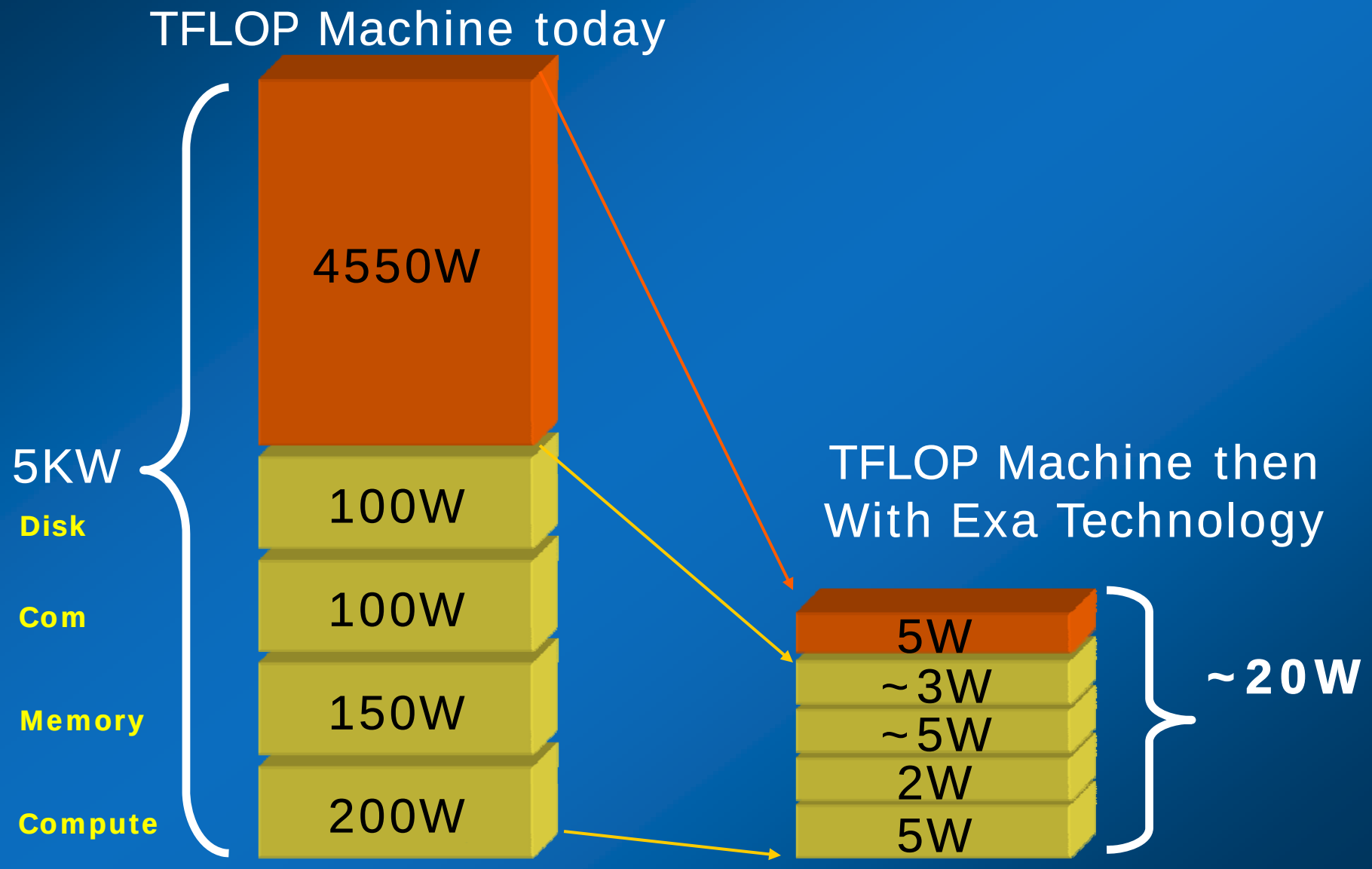
0.1B/FLOP @ 1.5nJ per Byte

200pJ per FLOP

KW Tera, MW Peta, GW Exa?

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

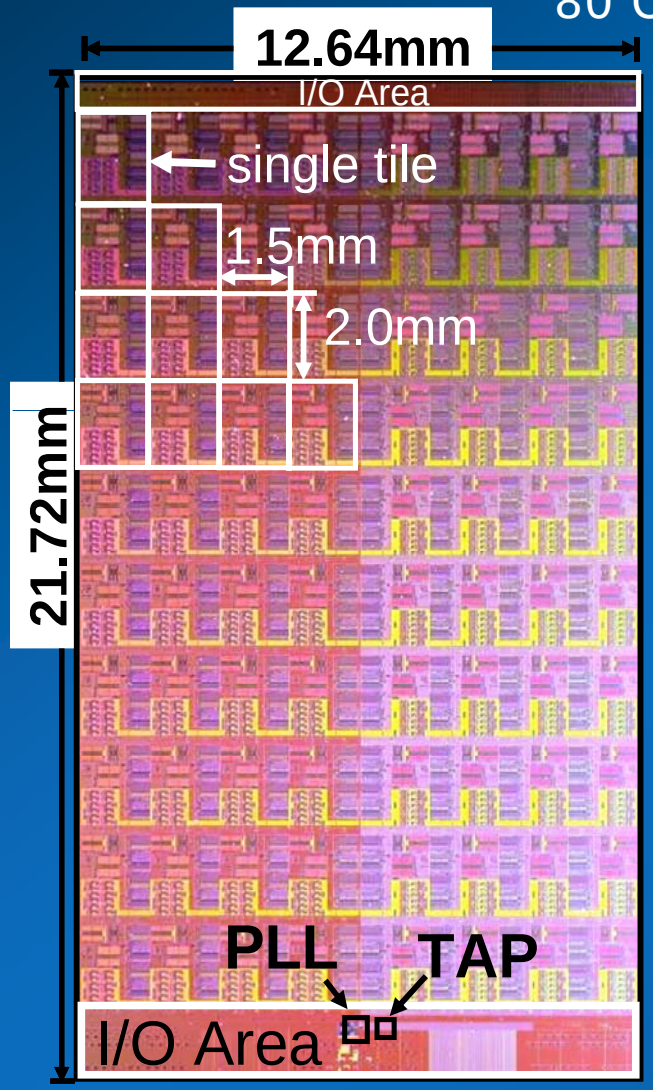
The Power & Energy Challenge

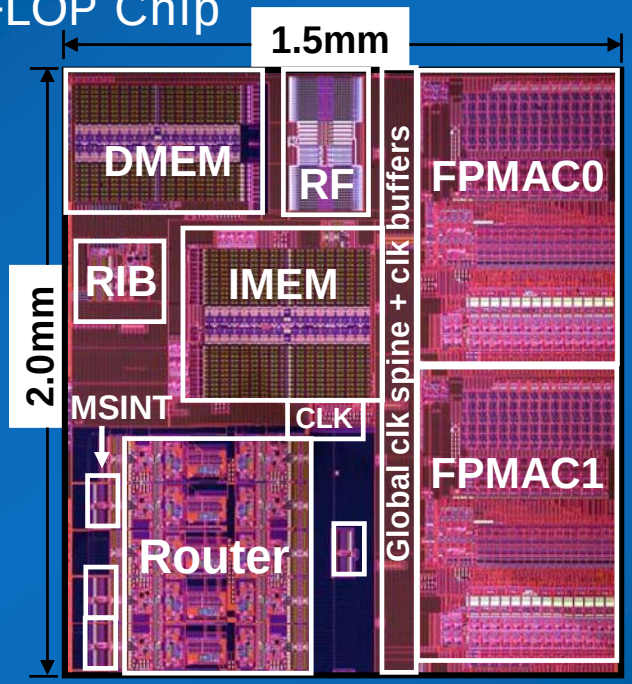


Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Starting Point: Optimistic yet Realistic

80 Core TFLOP Chip





Technology	65nm CMOS Process
Interconnect	1 poly, 8 metal (Cu)
Transistors	100 Million
Die Area	275mm^2
Tile area	3mm^2
Package	1248 pin LGA, 14 layers, 343 signal pins

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Scaling Assumptions

Technology (High Volume)	45nm (2008)	32nm (2010)	22nm (2012)	16nm (2014)	11nm (2016)	8nm (2018)	5nm (2020)
Transistor density	1.75	1.75	1.75	1.75	1.75	1.75	1.75
Frequency scaling	15%	10%	8%	5%	4%	3%	2%
Vdd scaling	-10%	-7.5%	-5%	-2.5%	-1.5%	-1%	-0.5%
Dimension & Capacitance	0.75	0.75	0.75	0.75	0.75	0.75	0.75
SD Leakage scaling/micron	1X Optimistic to 1.43X Pessimistic						

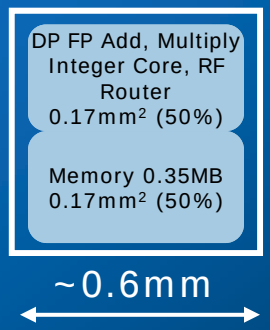
65nm Core + Local Memory

DP FP Add, Multiply Integer Core, RF Router
$5mm^2$ (50%)

Memory 0.35MB
$5mm^2$ (50%)

10mm2, 3GHz, 6GF, 1.8W

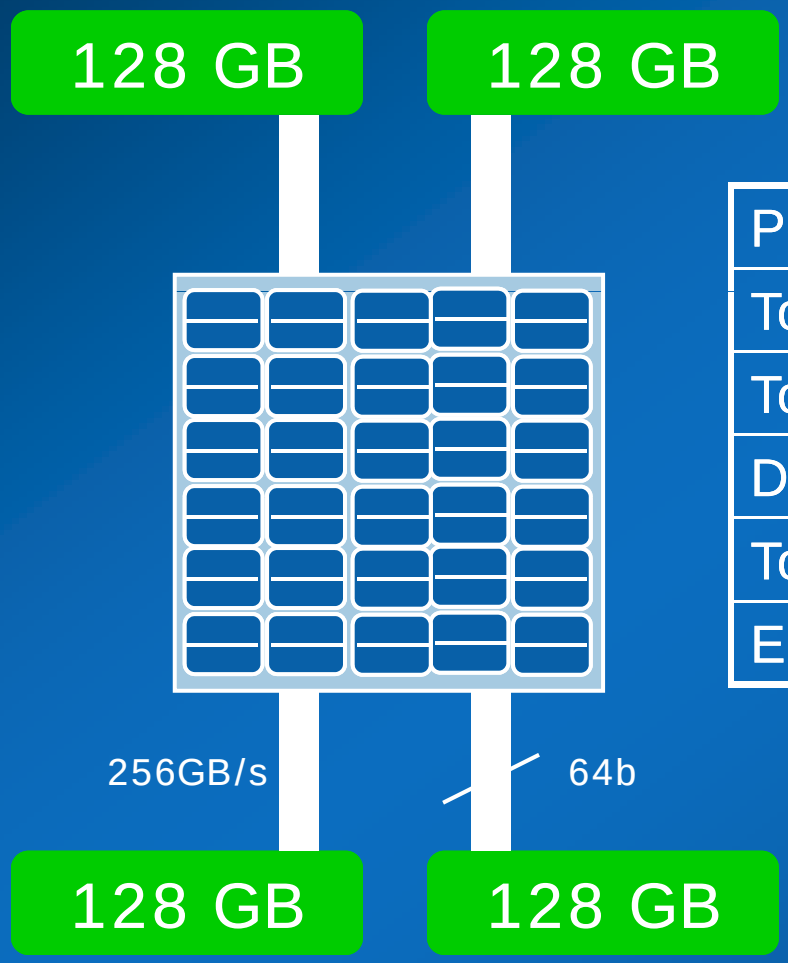
8nm Core + Local Memory



0.34mm2, 4.6GHz, 9.2GF, 0.24 to 0.46W

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Processor Node



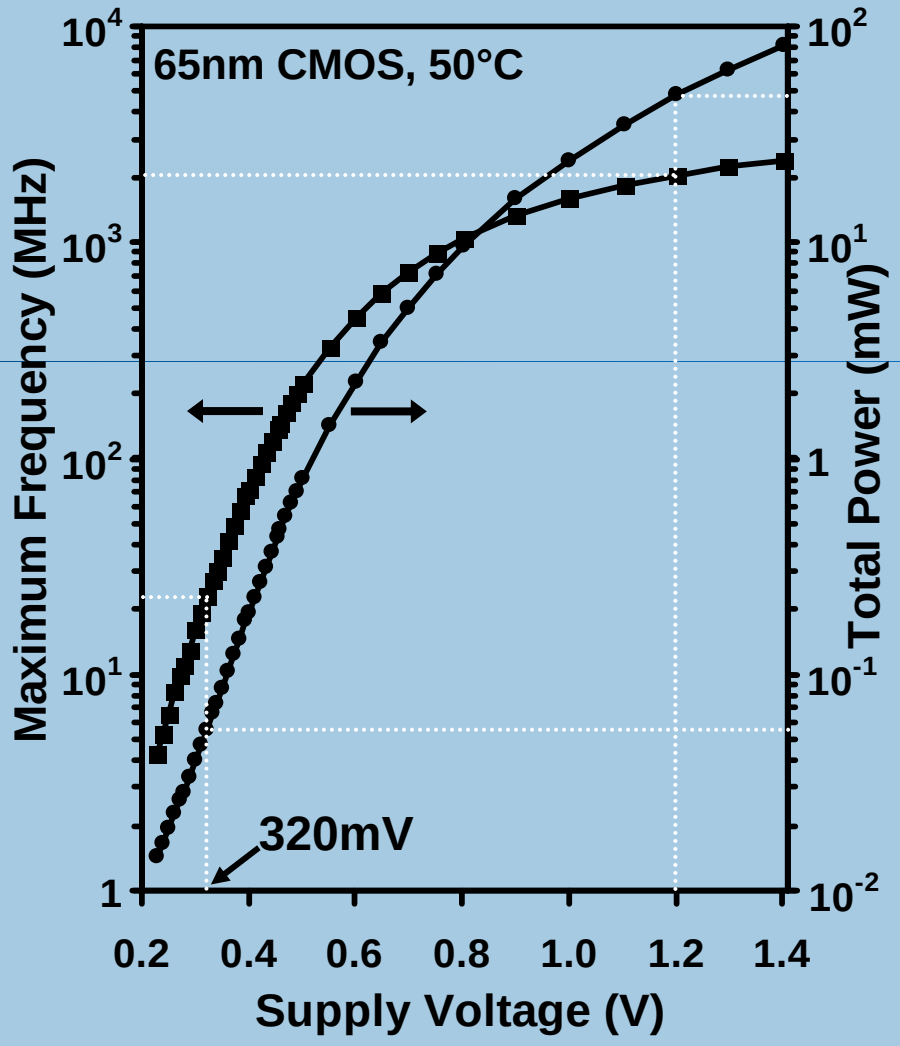
Peak performance	10.6 TF
Total DRAM Capacity	512GB
Total DRAM BW	1TB/s (0.1B/FLOP)
DRAM Power	800 W*
Total Power	1100 - 1400W
Energy efficiency	9.5 - 8 GF/Watt

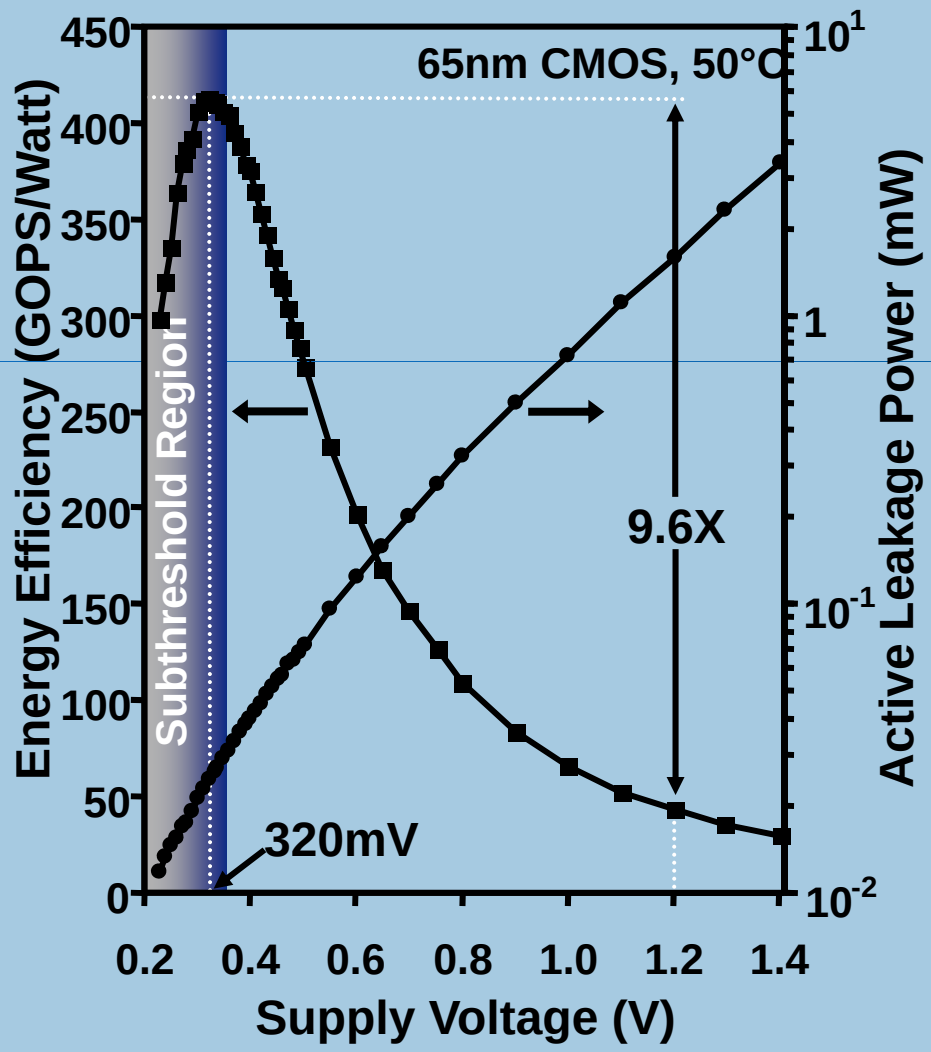
110-140 MW for Exascale

*Assumes 5% Vdd scaling each technology generation
140 pJ energy consumed per accessed bit

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

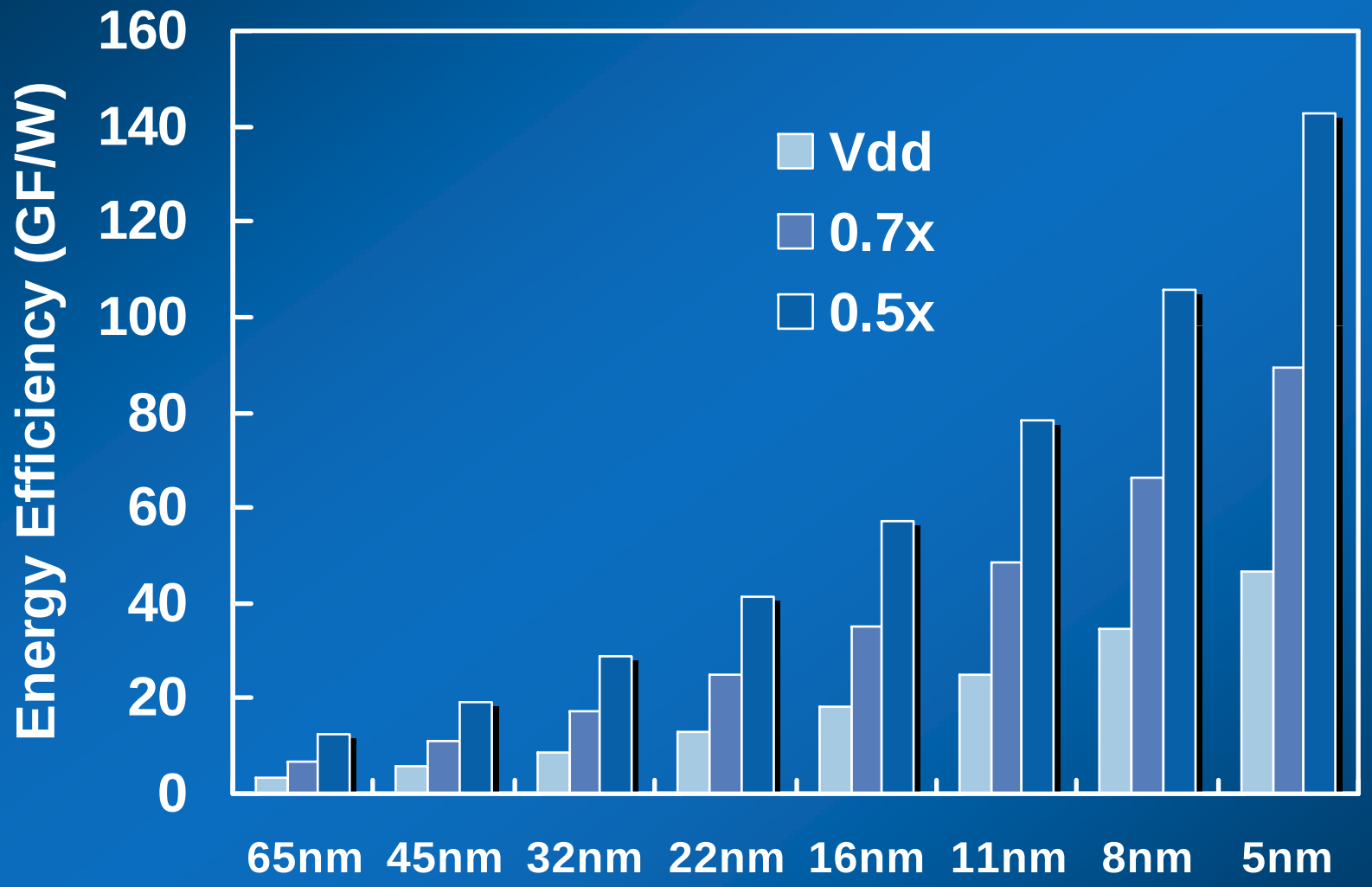
Near Threshold Logic





Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

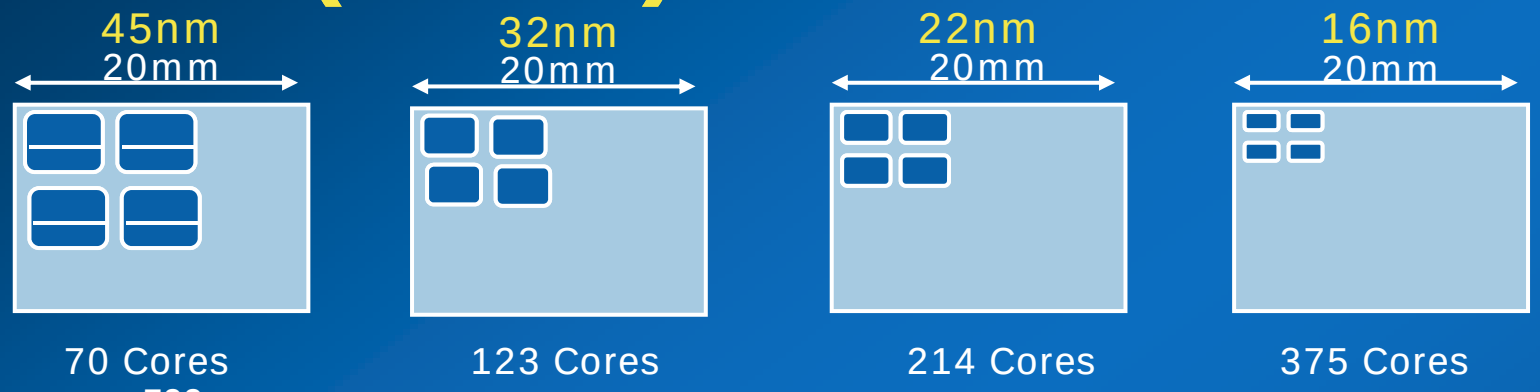
Energy Efficiency with Vdd Scaling

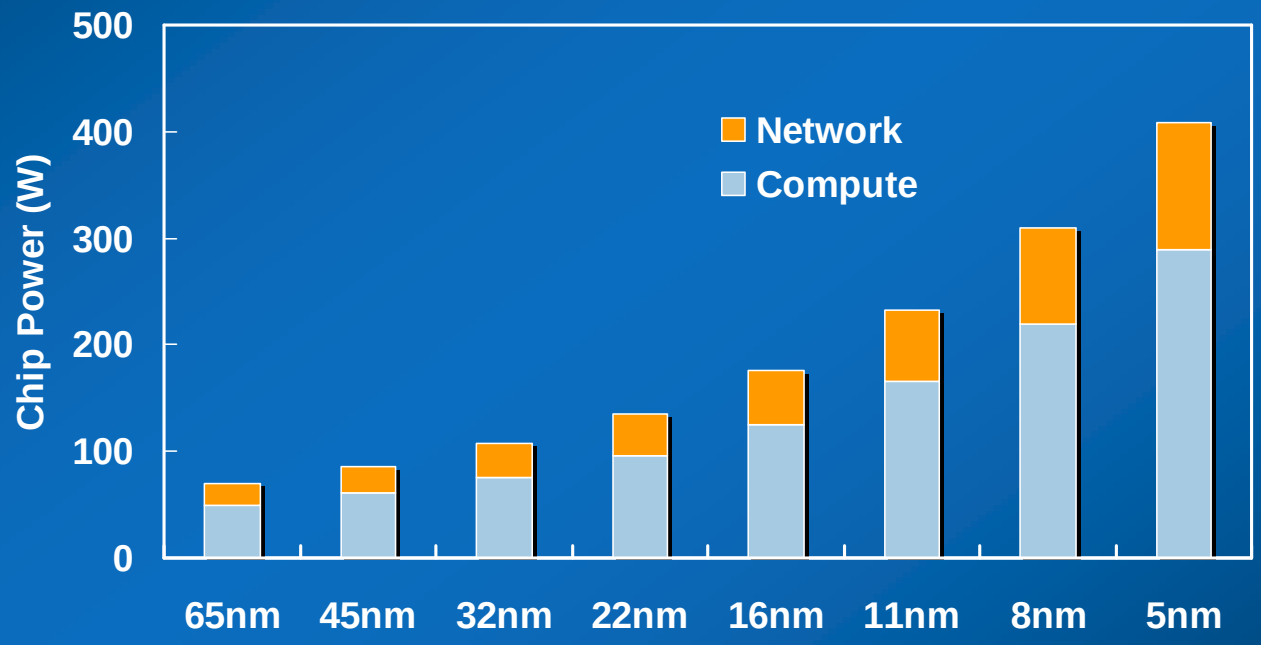


~3X Compute energy efficiency with Vdd Scaling

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

On-die (Mesh) Interconnect





On-die network (mesh) power is high

Worse if link width scales up each generation

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Mesh—Retrospective

Bus: Good at board level, does not extend well

- Transmission line issues: loss and signal integrity, limited frequency
- Width is limited by pins and board area
- Broadcast, simple to implement

Point to point busses: fast signaling over longer distance

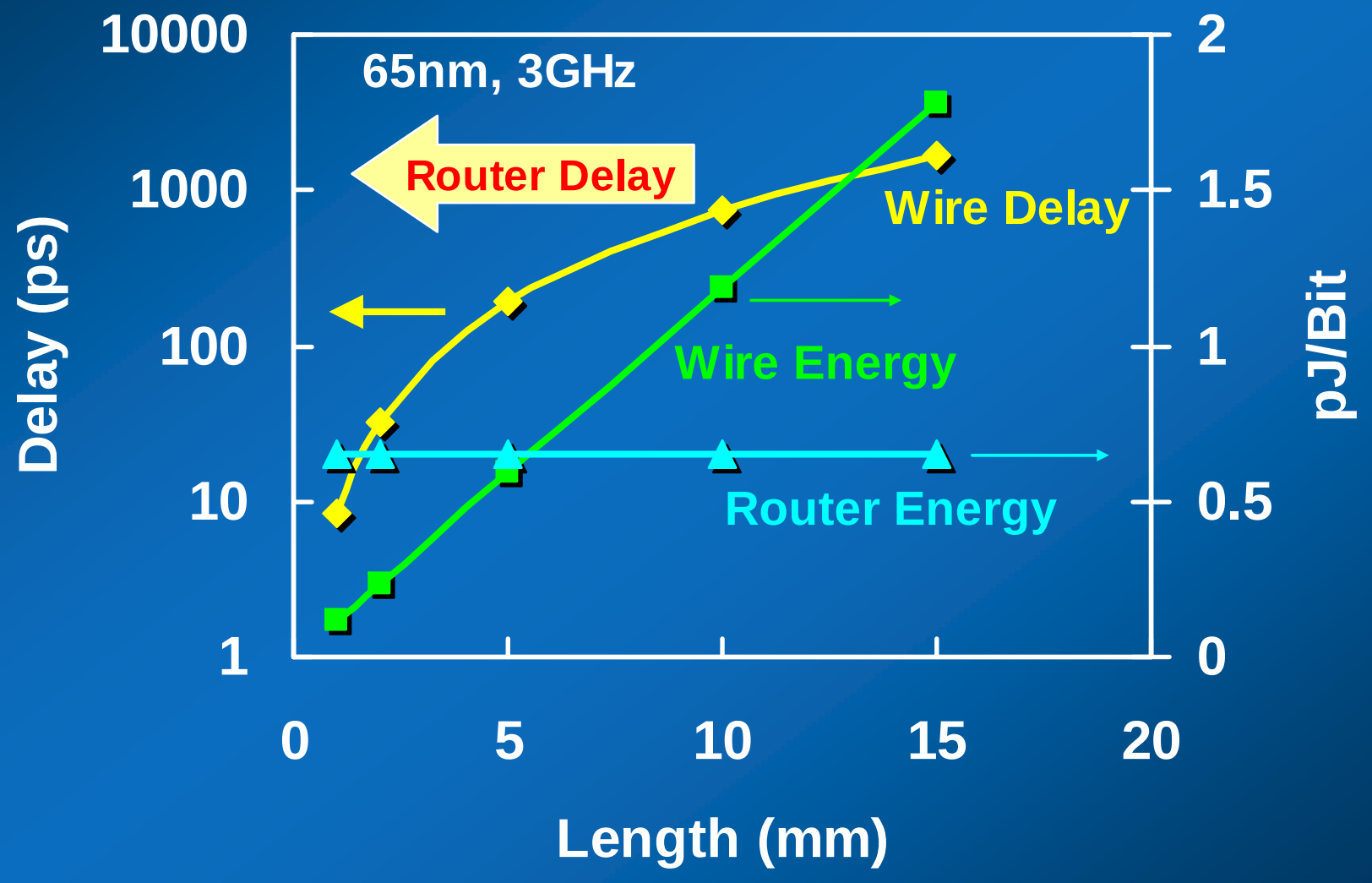
- Board level, between boards, and racks
- High frequency, narrow links
- 1D Ring, 2D Mesh and Torus to reduce latency
- Higher complexity and latency in each node

Hence, emergence of packet switched network

But, pt-to-pt packet switched network on a chip?

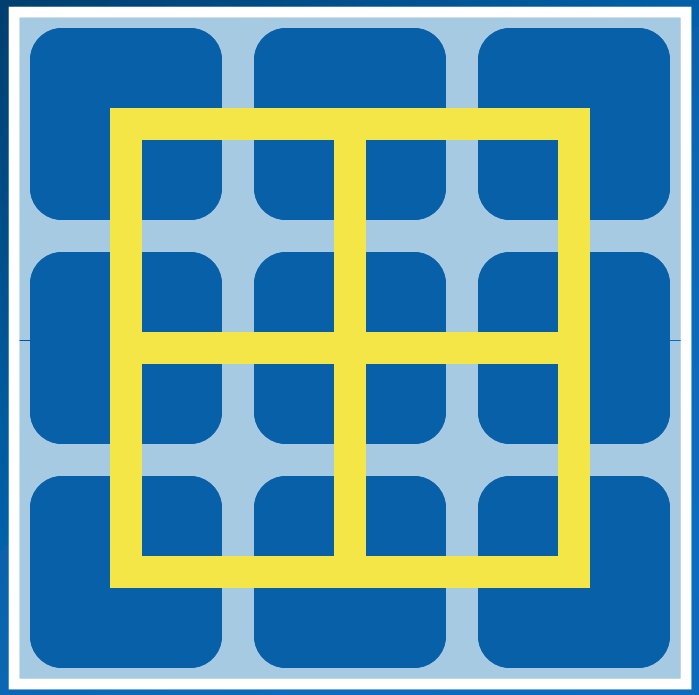
Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Interconnect Delay & Energy



Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Bus—The Other Extreme…



Issues:
Slow, < 300MHz
Shared, limited scalability?

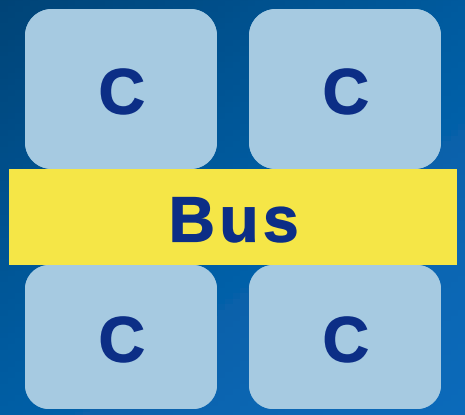
Solutions:
Repeaters to increase freq
Wide busses for bandwidth
Multiple busses for scalability

Benefits:
Power?
Simpler cache coherency

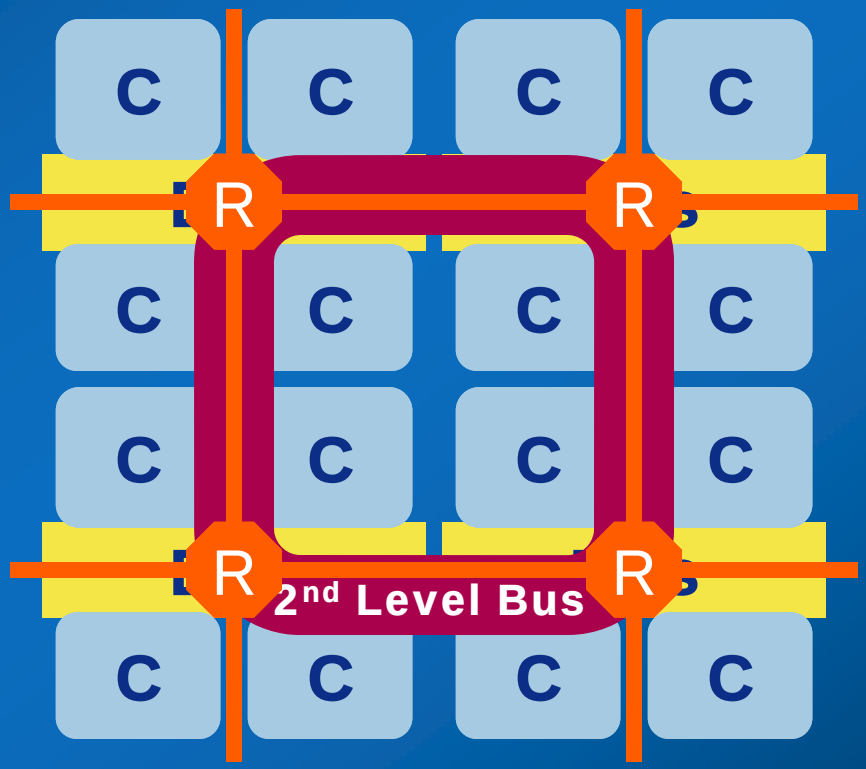
Move away from frequency, embrace parallelism

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Hierarchical & Heterogeneous



Bus to connect over short distances



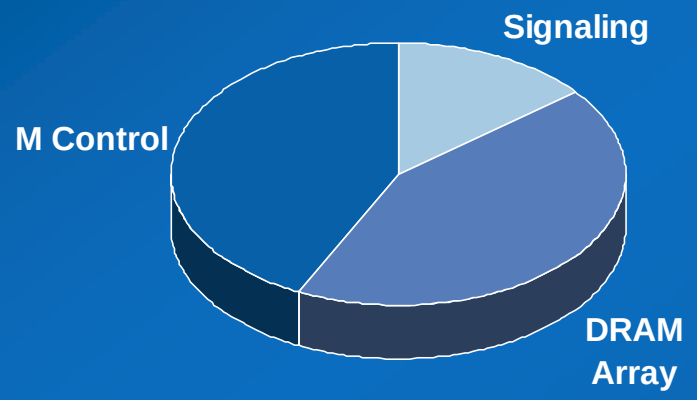
Hierarchy of Busses

Hierarchical circuit and packet switched networks

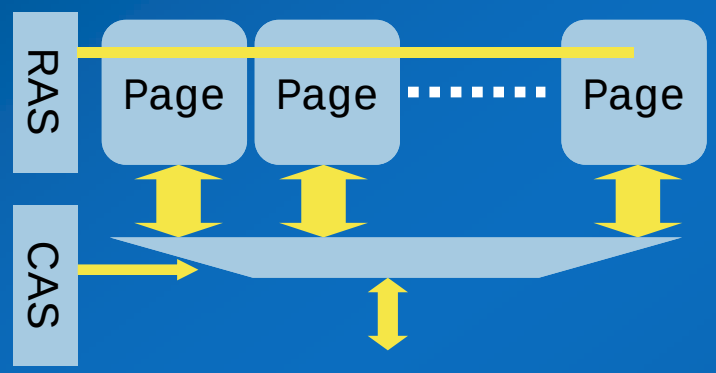
Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Revise DRAM Architecture

Energy cost today:
~150 pJ/bit

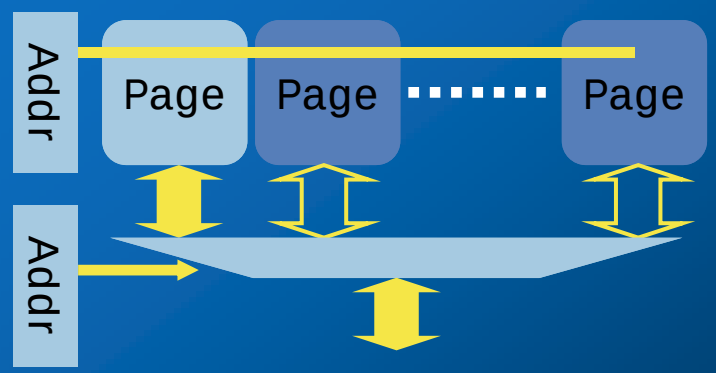


Traditional DRAM



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

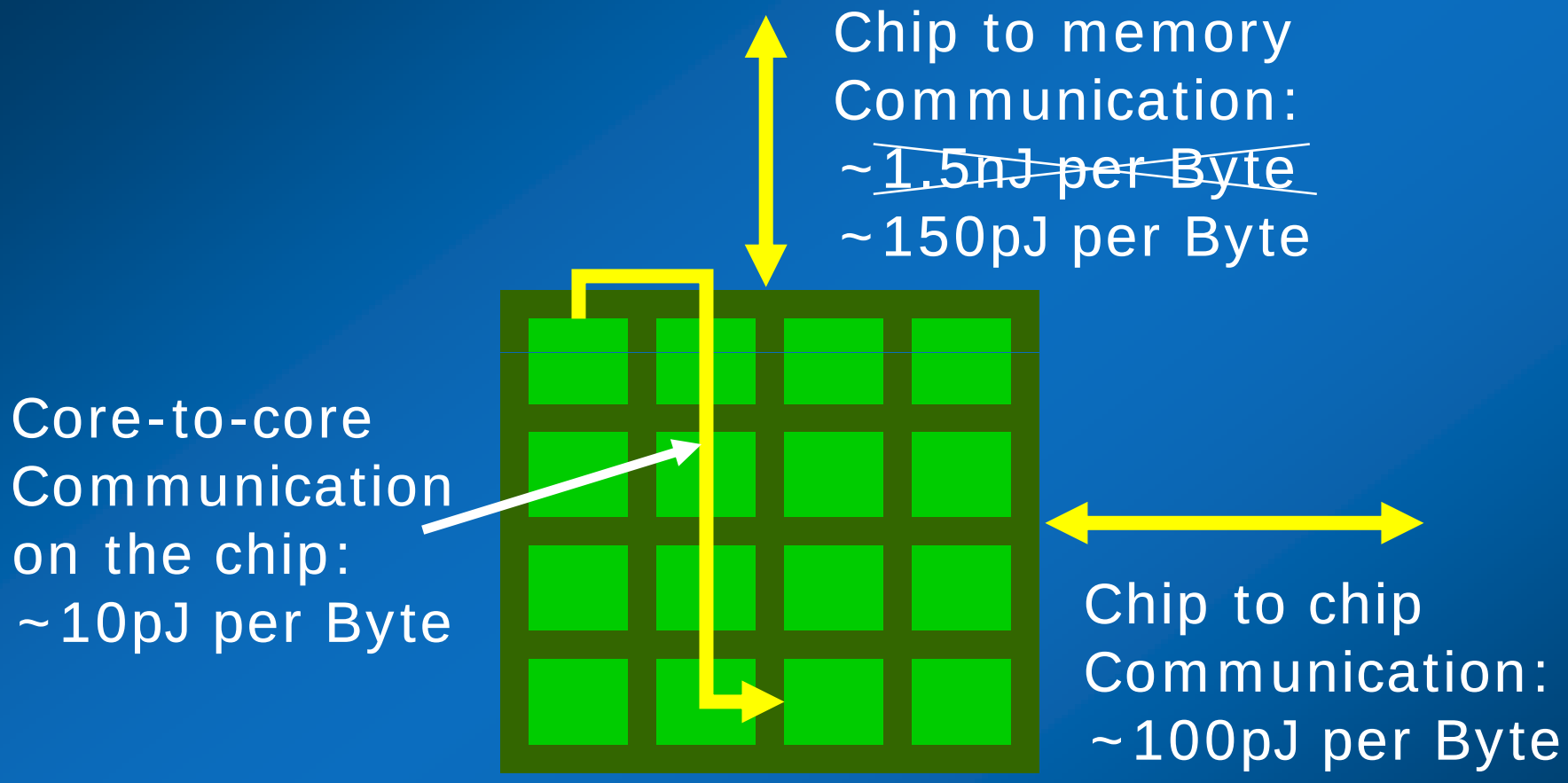
New DRAM architecture



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

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Data Locality



Data movement is expensive—keep it local
(1) Core to core, (2) Chip-to-chip, (3) Memory

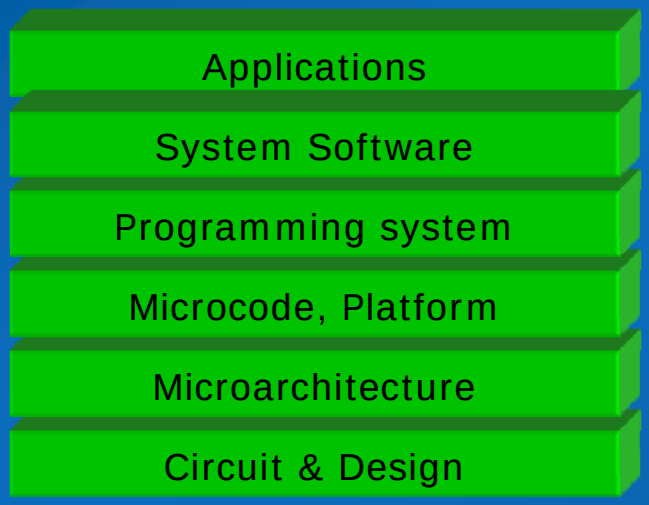
Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Resiliency

Faults	Example
Permanent faults	Stuck-at 0 & 1
Gradual faults	Variability Temperature
Intermittent faults	Soft errors Voltage droops
Aging faults	Degradation

Faults cause errors (data & control)	
Datapath errors	Detected by parity/ECC
Silent data corruption	Need HW hooks
Control errors	Control lost (Blue screen)

Minimal overhead for resiliency



Error detection
Fault isolation
Fault confinement
Reconfiguration
Recovery & Adapt

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

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

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Programming Objectives and Approaches

Programmability

High level algorithm expression

Gentle slope programming

Interactive visualization environment

Programming idioms / patterns

Hide hardware details

Hide resilience support

More Concurrency/Locality

Separation of semantics from mapping

Semantic annotations increase map choice

Extend scope of mining for concurrency

Geometric formulations of optimizations

Algorithm variant exploration

Analytic hierarchical parallelism / locality opts

Energy Reduction

Support explicit memories, comm, sync, tasks

Aggressive producer-consumer fusion

Automatic transforms for arithmetic intensity

Hierarchical and heterogeneous mapping

Support for new energy efficient algorithms

Support for new hardware energy features

Runtime Support

Granularity control for optimal codelets

Generation of models to drive runtime opt

Generation of parametric mappings

Generation of dynamic mappings

Callbacks and monitoring feedback

Produce runtime/hardware execution idioms

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Programming Technical Foundation and Philosophy

Existing
"Background IP"

Concurrent Collections (CnC)
Maximal concurrency program expression and separates algorithm from mapping

Hierarchically Tiled Arrays (HTA)
Annotation system for succinct algorithm data distribution relevant to UHPC algorithms

R-Stream Compiler
Analytical formulations and solvers of complex parallelism / locality constraints objectives
Explicitly managed architecture support

New!

Extended Generalized Dependence Graph IR

New!

- Denotations Annotations Language
- Analysis Transformation Optimization
- Interactive System
- Model / Code Generation

Approach is semantics based:
Algorithm representations, transformations and optimization repertoire *before* designing a new language

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Execution Model Design Topics & Approaches

Guiding Characteristics:

- Provides governing principles
- Impacts design across all layers
- Provides conceptual framework for co-design
- Supports operation "decision chain"
- Addresses critical efficiency factors

Topics

Energy Efficiency

Massive concurrency

Performance scalability

Data locality/movement

Resiliency

Security

Programmability

HW/SW Co-Design

Approach

Fine-grained/Event-driven

Appropriate regulation structure

Regulation/computation sharing same mechanism

Self-aware via model-based control architecture

Tightly coupled, goal-driven

Energy proportional

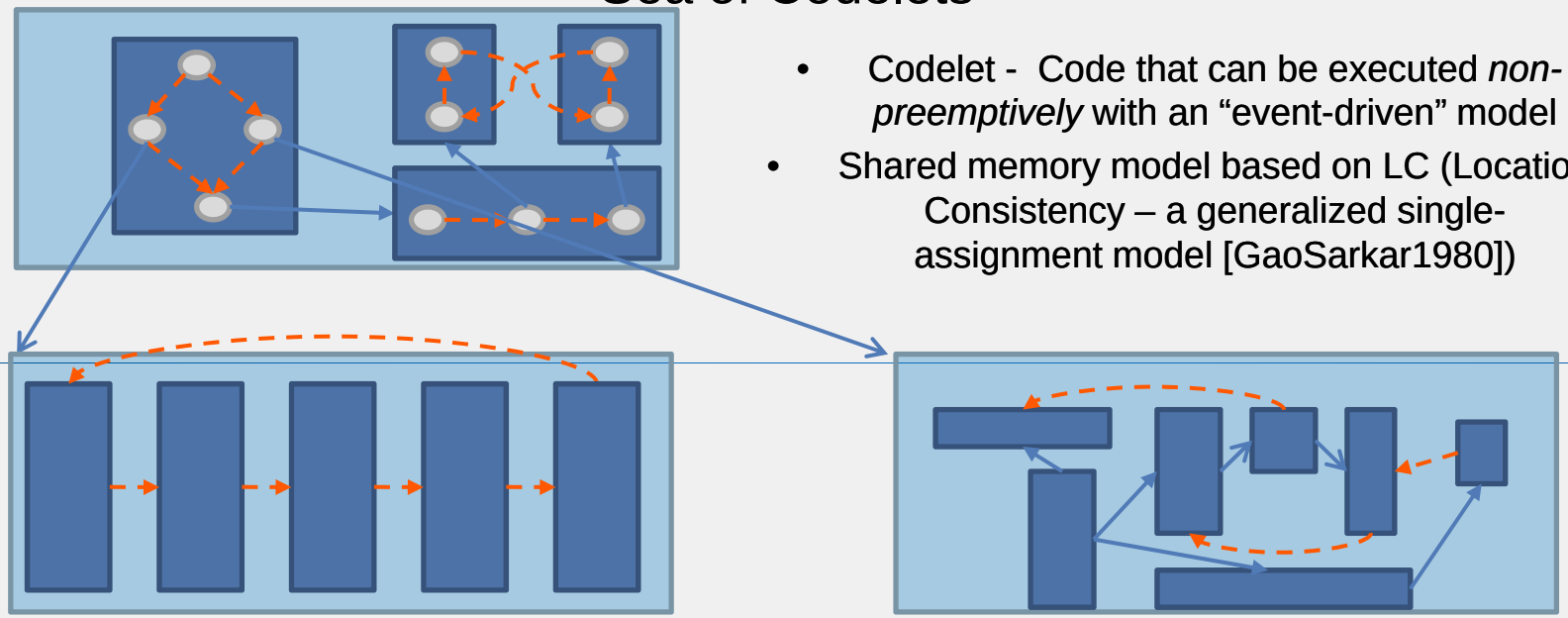
Security- and resilience-symmetric

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Execution Model and Codelets

Programming Models/Systems

Sea of Codelets



- Codelet - Code that can be executed *non-preemptively* with an "event-driven" model
- Shared memory model based on LC (Location Consistency – a generalized single-assignment model [GaoSarkar1980])

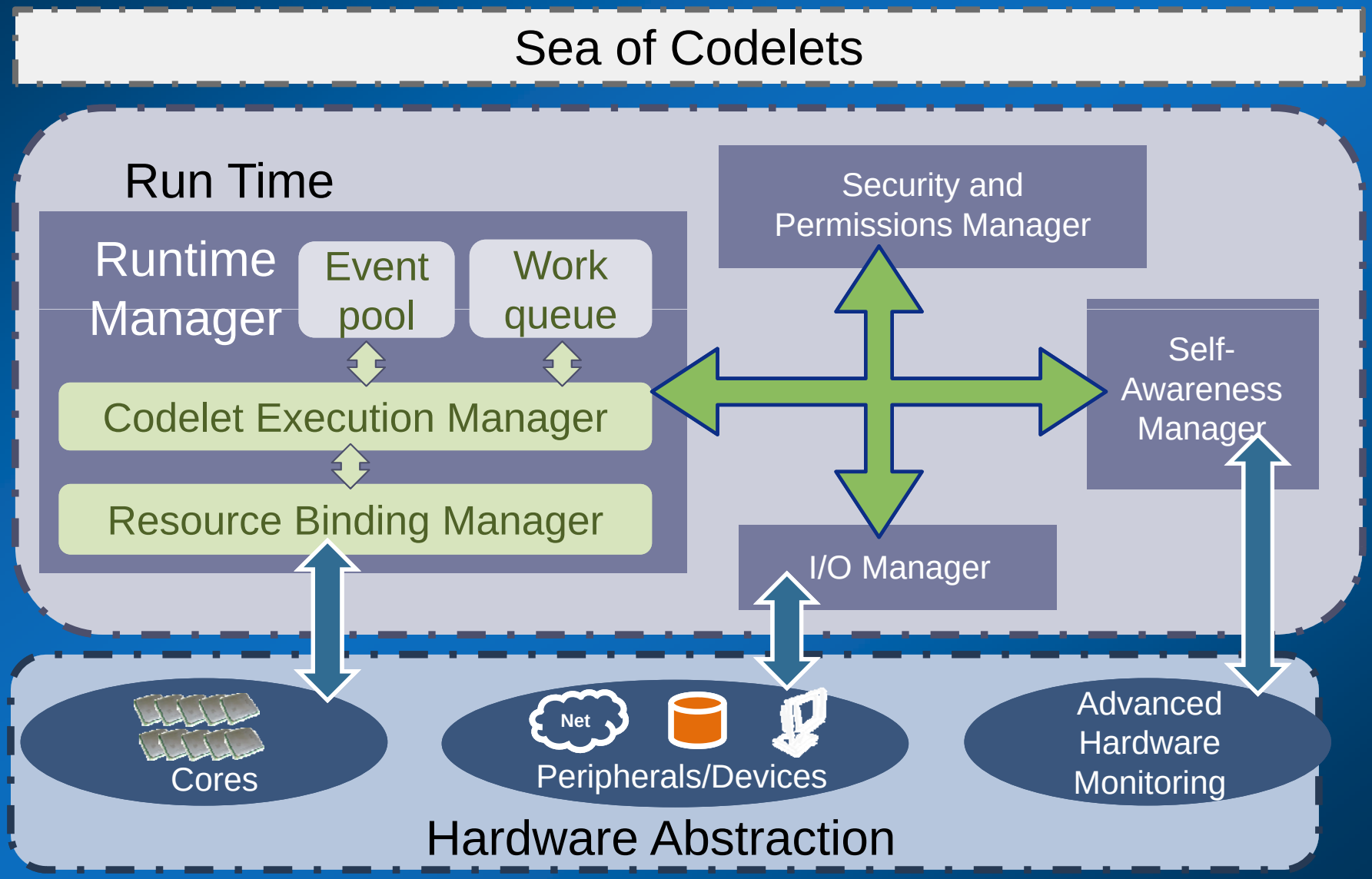
Run Time System

Hardware Abstraction

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Execution Model and Runtime

Programming Models/Systems



Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Applications Objectives and Approaches

Co-Design

Derive design choices from requirements

Characterize Challenge Problems

Derive basic idioms

Define hardware counters and hooks

Emphasize app areas not well-served

Energy Reduction

Define “balanced machine” ops/joule

Pragmas for user-level throttling

Data layout based on locality models

Integrated Hardware/Software

Control frequencies from user-space

CANAL reports from compiler

Dynamic analysis reports for user

Self-Aware System

Learn from previous runs

Learn locality patterns

Anticipate program phase changes

Generation of dynamic mappings

Callbacks and monitoring feedback

Produce runtime/hardware execution idioms

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Applications Technical Foundation and Philosophy

Existing

Static analysis
Break application into constituent idioms

Dynamic analysis
World's fastest binary instrumentation and tracer

Performance and Power Models
Predict performance and energy based on machine settings

New!

Integrated Hardware / Software Stack

New!

Pragmas for power throttling

Fine-grained energy counters

Feedback and control

Learn from run to run

Approach is idiom and phase based: depending on pattern of computation, data access, and communication set voltages and gates accordingly. Tap into all sources of info including past history

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Summary

Von-Neumann computing & CMOS technology

Voltage scaling to reduce power and energy

- Explodes parallelism
- Cost of communication vs computation—critical balance
- Resiliency to combat side-effects and unreliability

Programming system for extreme parallelism

Application driven, HW/SW co-design approach

Self awareness & execution model to harmonize

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007

Thank you

Copyright © 2010, Intel Corporation (Copyright applies to commercial uses only)
Notwithstanding the above copyright/restriction, the Government possess
Unlimited Rights in accordance with Article VIII of Agreement HR0011-10-3-0007