Nehalem-EX: a 45nm, 8-core Enterprise Processor

Stefan Rusu

Intel Corporation

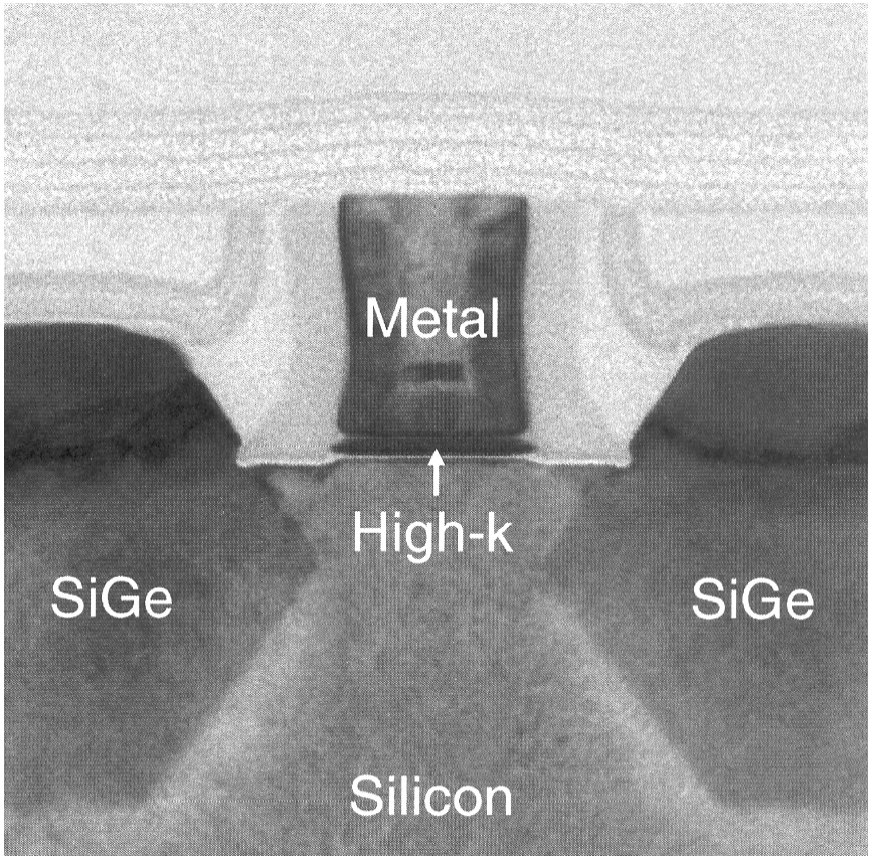
Santa Clara, CA

stefan.rusu@intel.com

Outline

- Process Technology
- Processor Scaling Trends
- Block Diagram and Die Photo
- Cache Design
- Core and Cache Recovery
- Clock and Voltage Domains
- Power, Package and Thermals
- I/O Links
- Idle Power Reduction
- Summary

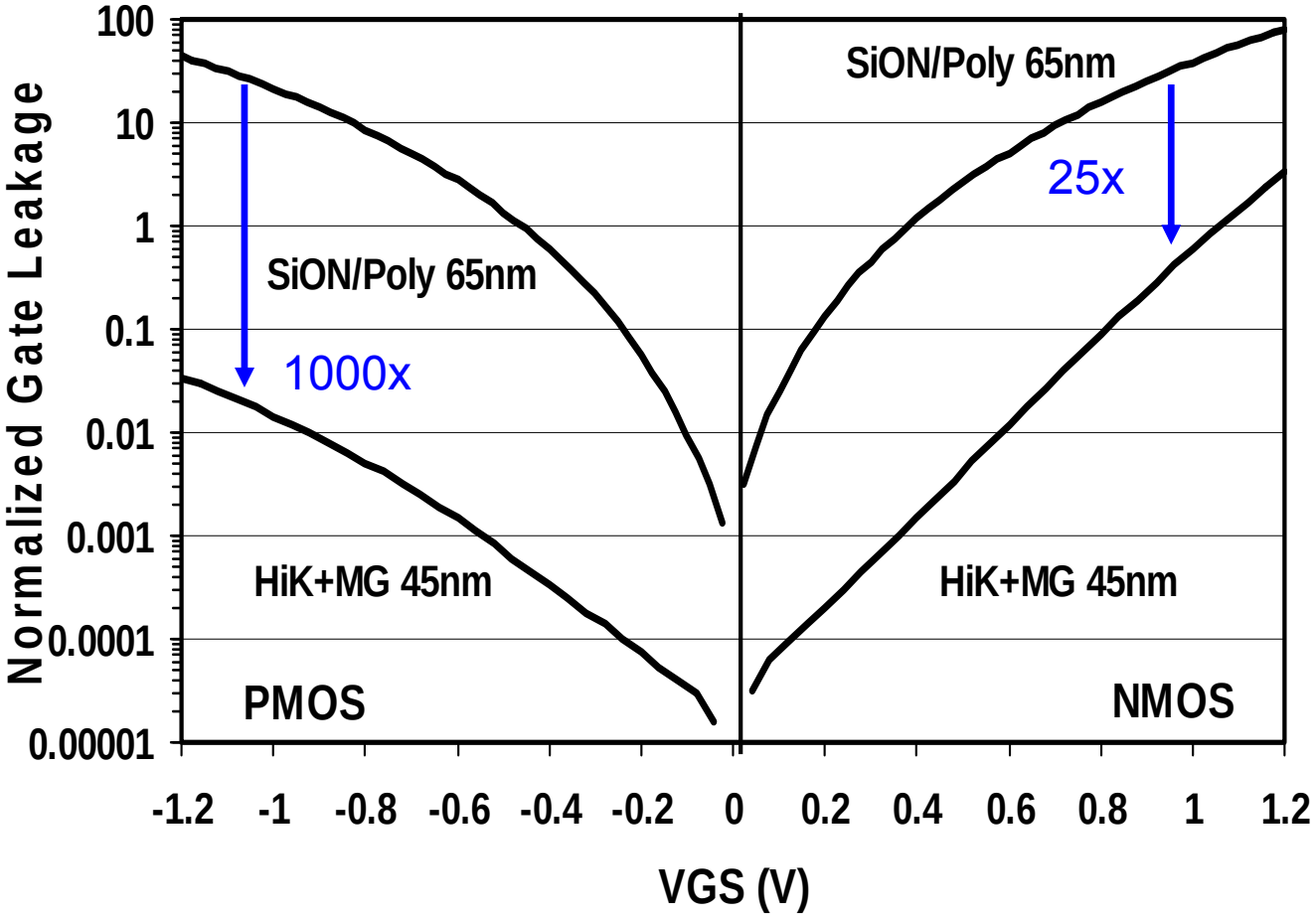
45nm Hi-k Metal Gate Technology



World's first Hi-k Metal Gate CMOS transistors integrated with 3rd generation strained silicon

K. Mistry, IEDM 2007

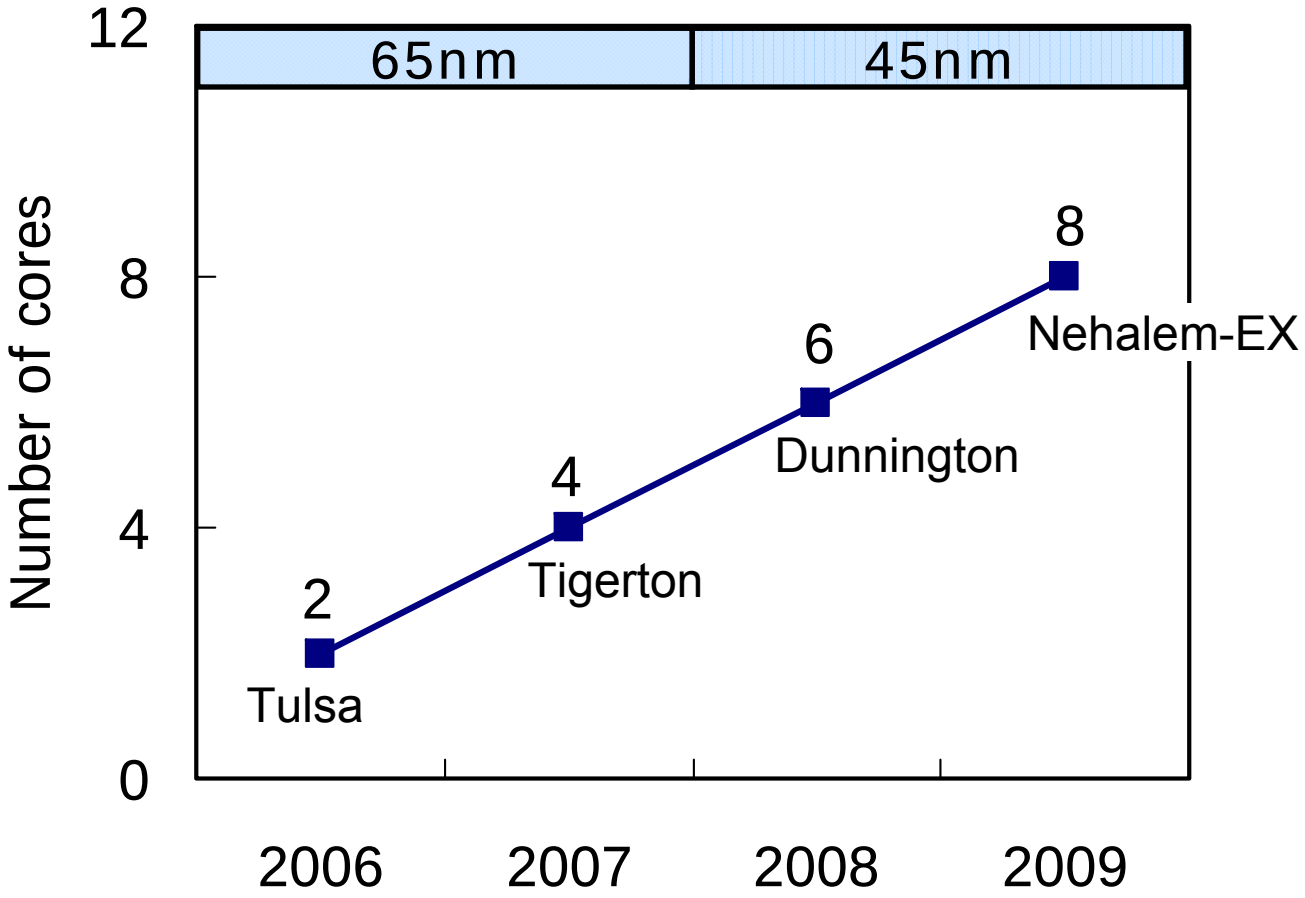
Gate Leakage Reduction



HK+MG significantly reduces gate leakage

K. Mistry, IEDM 2007

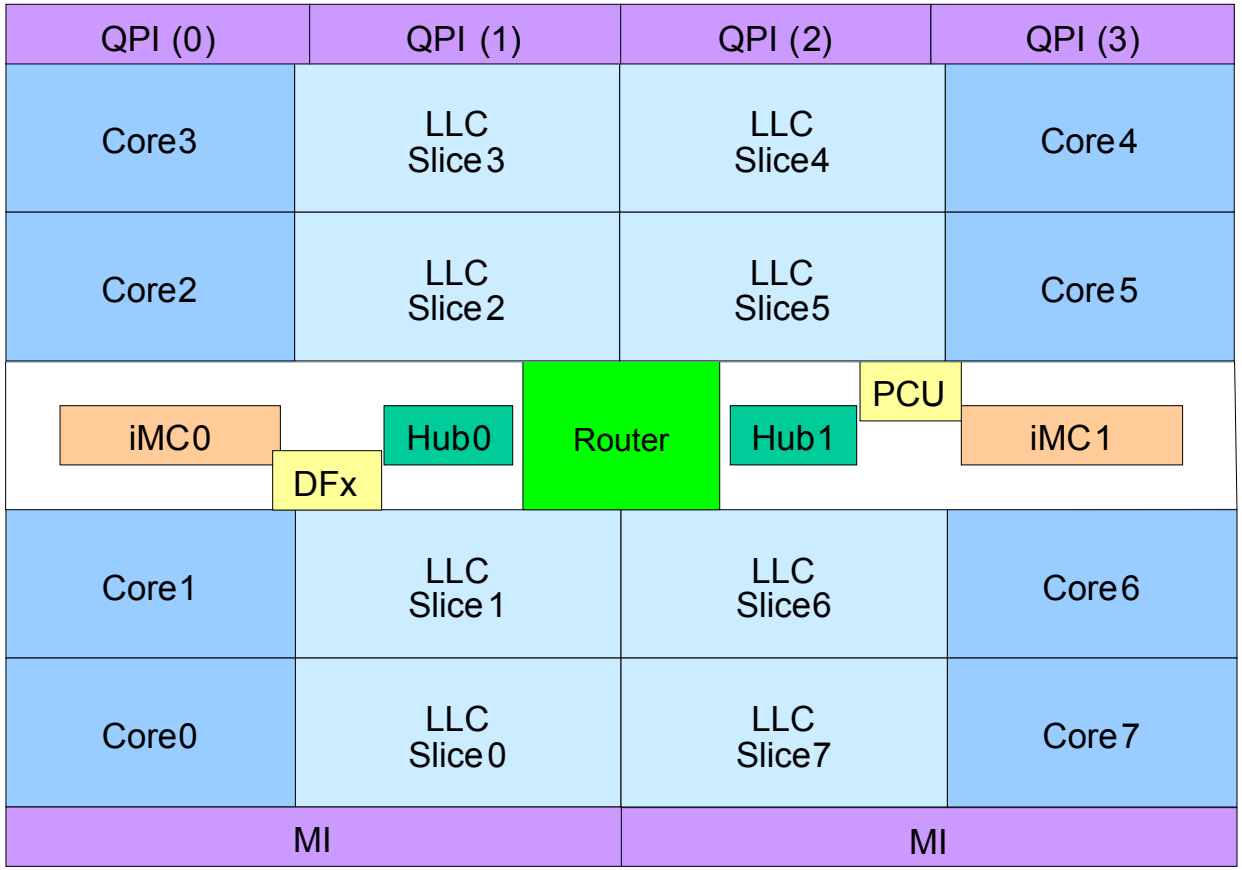
Xeon® EX Multi-Core Scaling Trend



Two additional cores every year

S. Rusu, ISSCC 2009

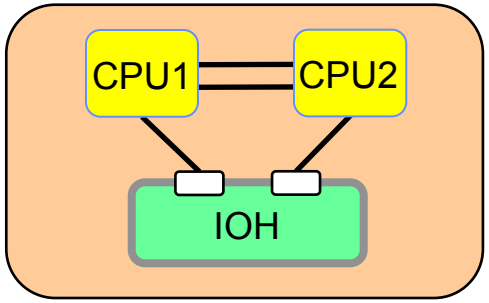
Processor Block Diagram

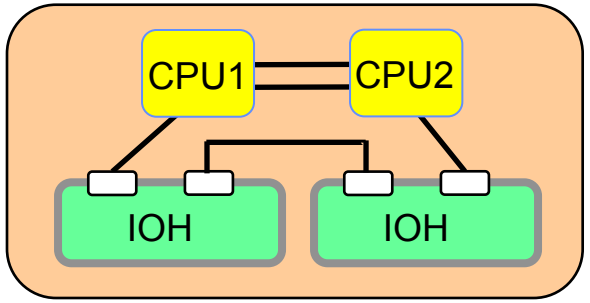


- 8 cores, 16 threads, 2 integrated memory controllers
- 4 point-to-point Quick Path Interconnect links
- Two Scalable Memory Interfaces per memory controller
- Two counter rotating rings to minimize latency

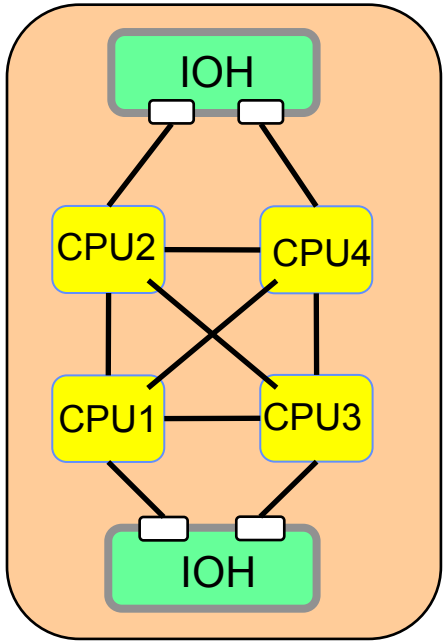
Platform Configuration Examples

2 Processors

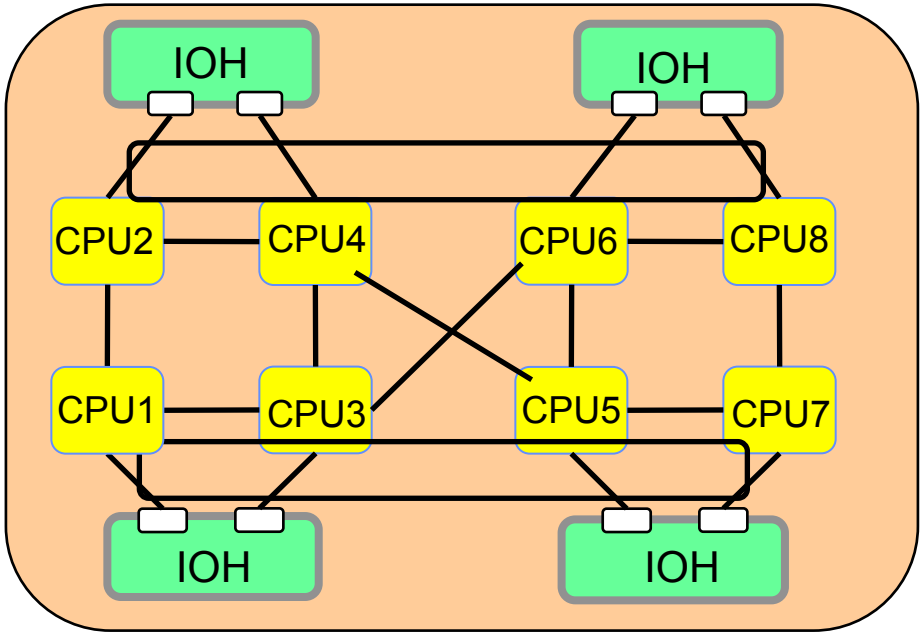




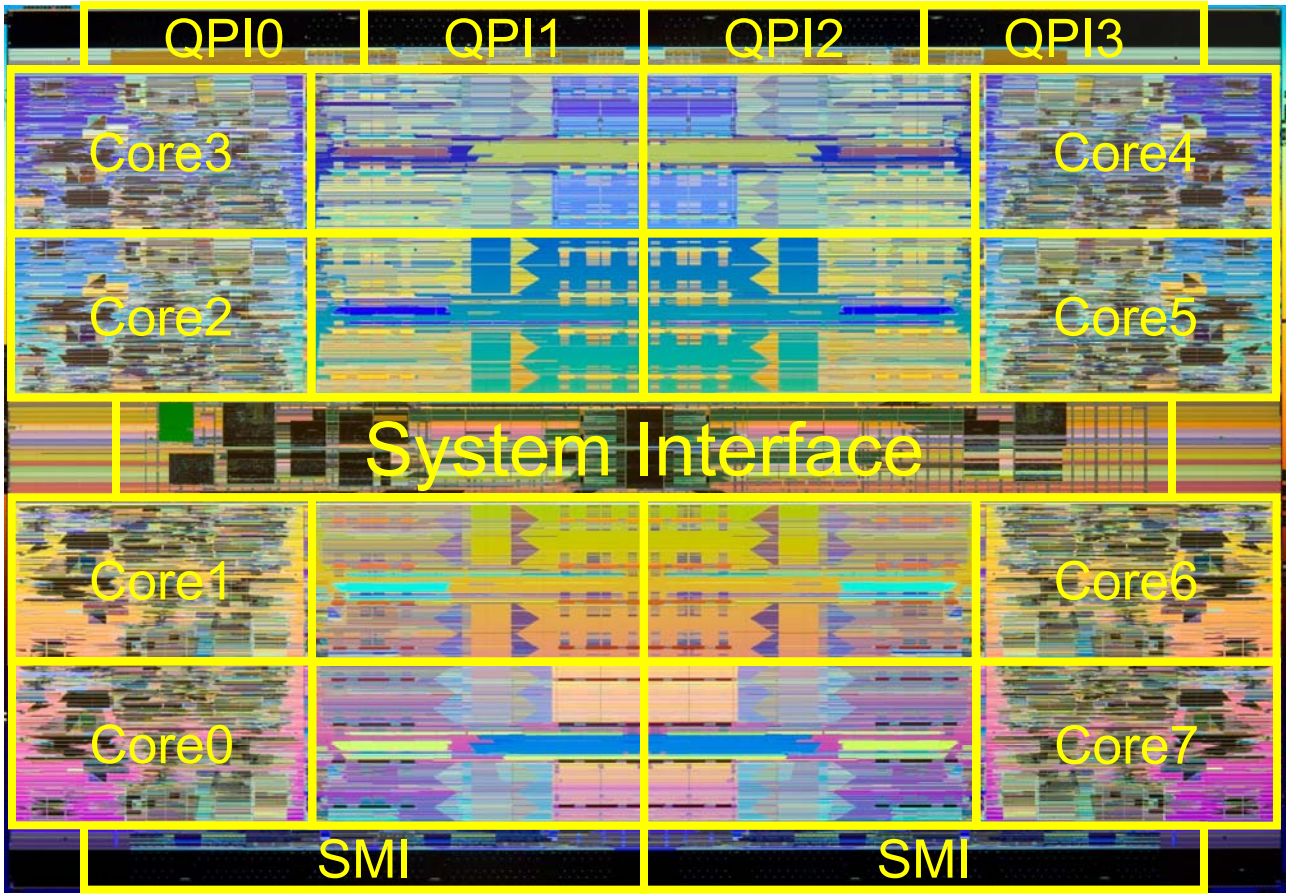
4 Processors



8 Processors



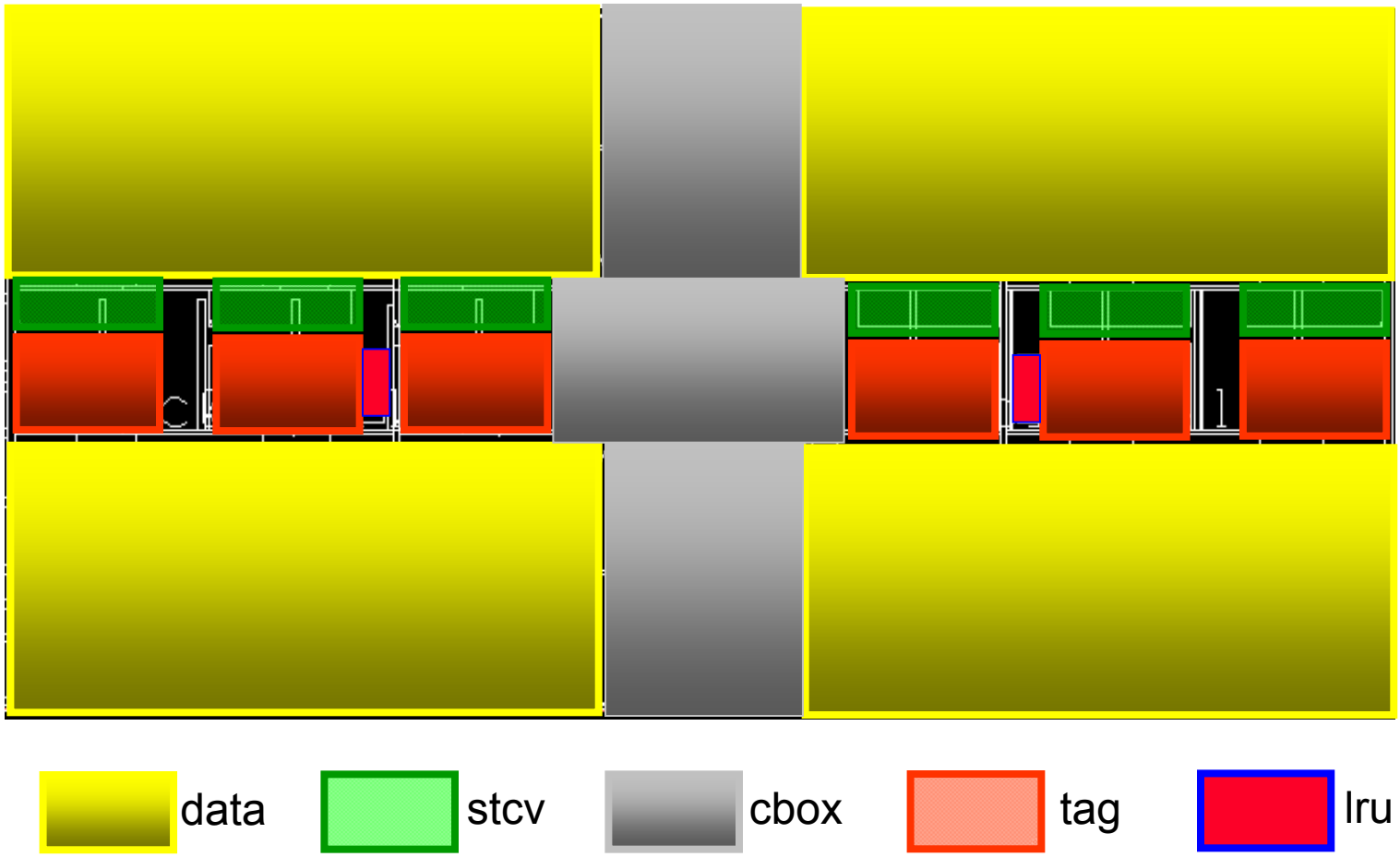
Die Photo



The largest device count reported for a microprocessor

2.3B transistors

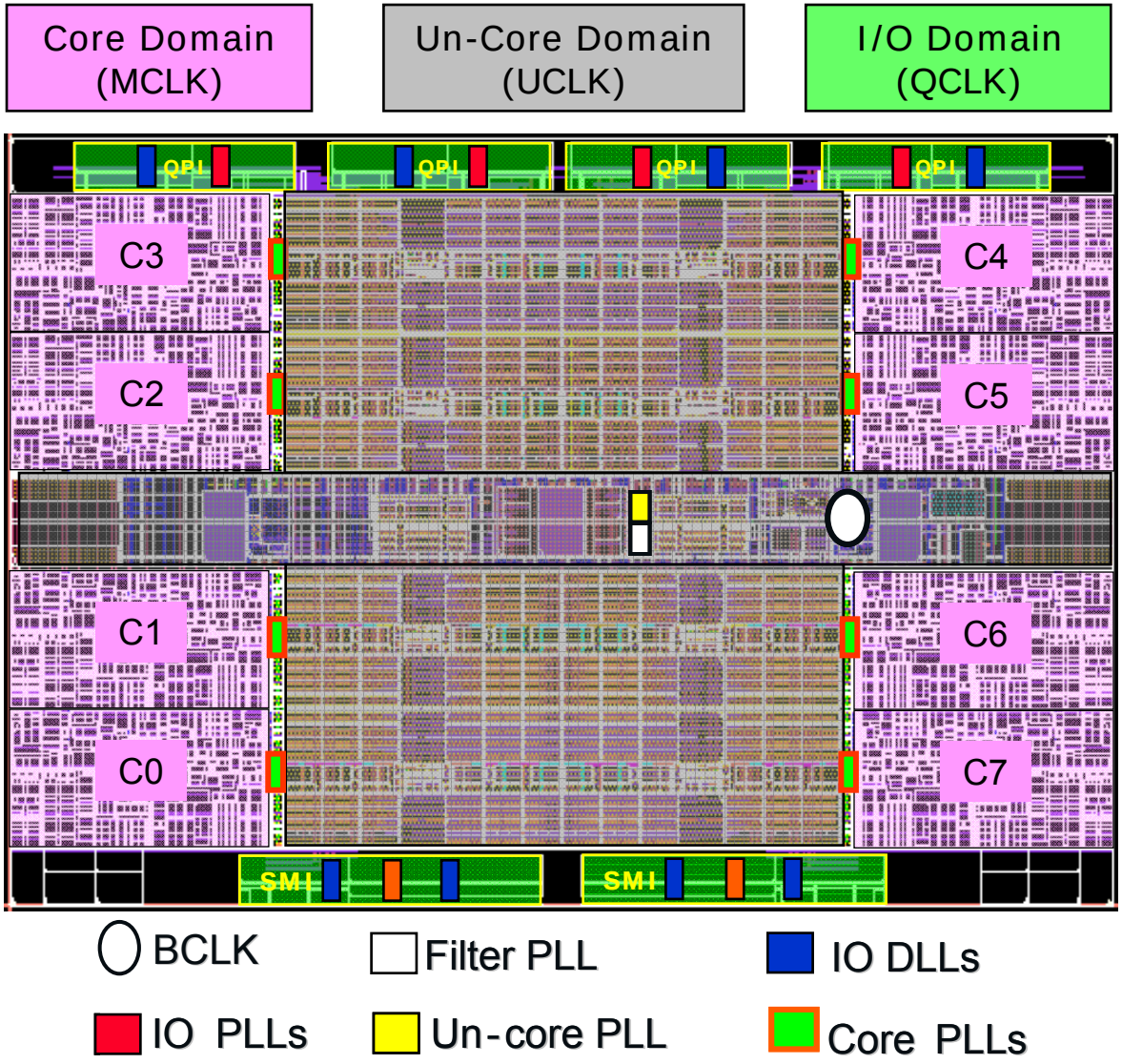
3MB L3 Cache Slice



24 ways, 64B line size, 48 sub-arrays per slice

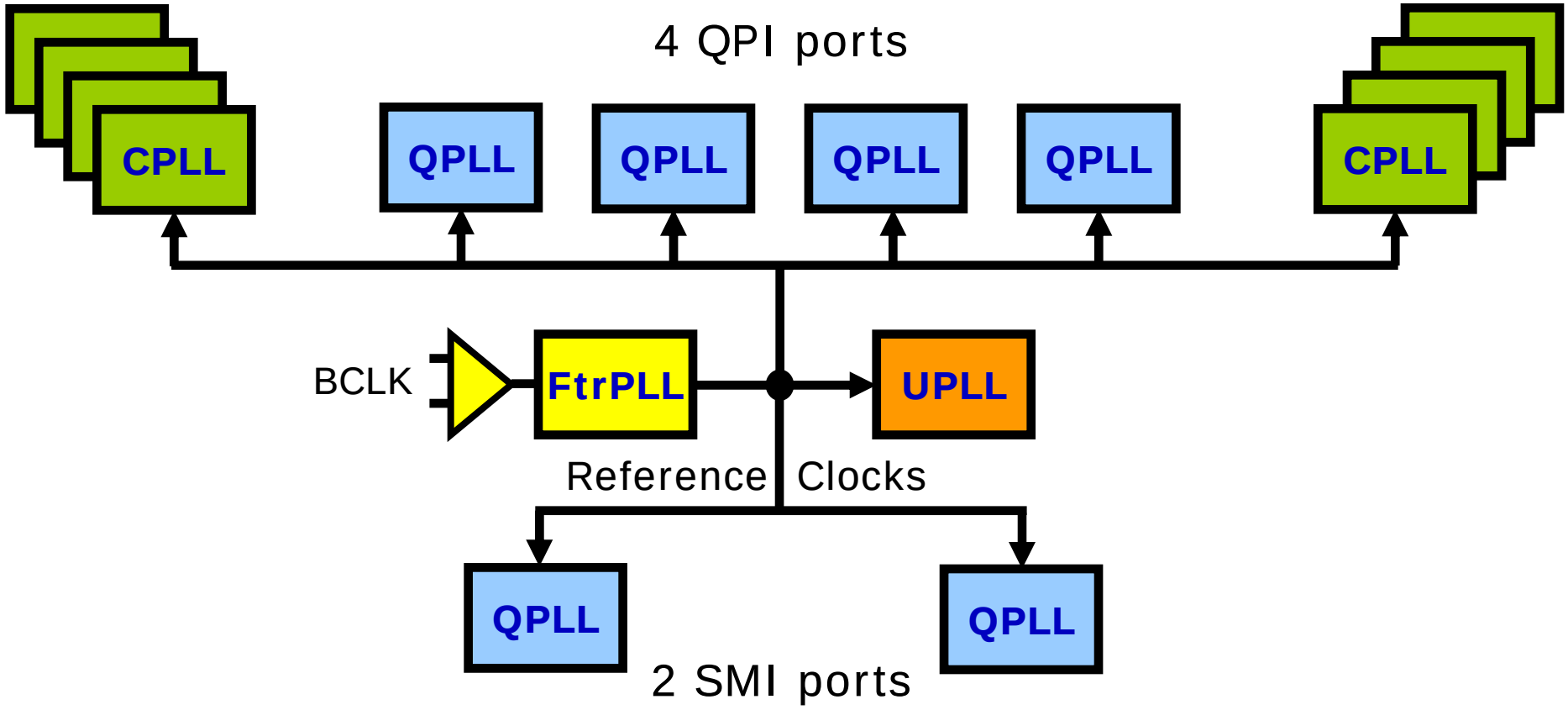
DECTED in data arrays, SECDED in tag arrays

Clock Domains



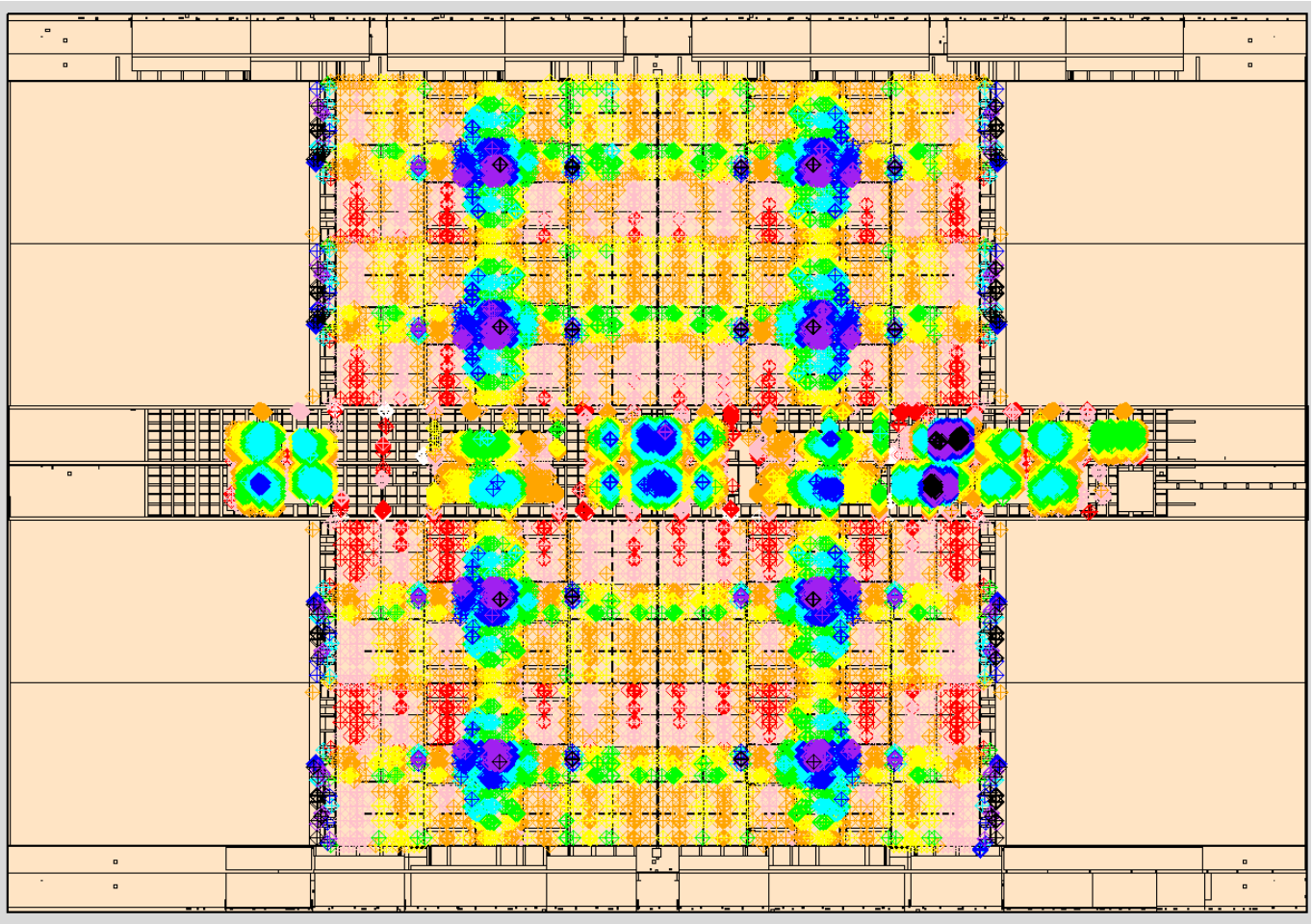
- 3 primary clock domains:
  - Core
  - Un-core
  - I/O
- 16 PLLs & 8 DLLs
  - Single system clock input (BCLK)

PLL Reference Clocks



Reference clocks are distributed to 15 destinations

Simulated Un-Core Clock Skew Profile

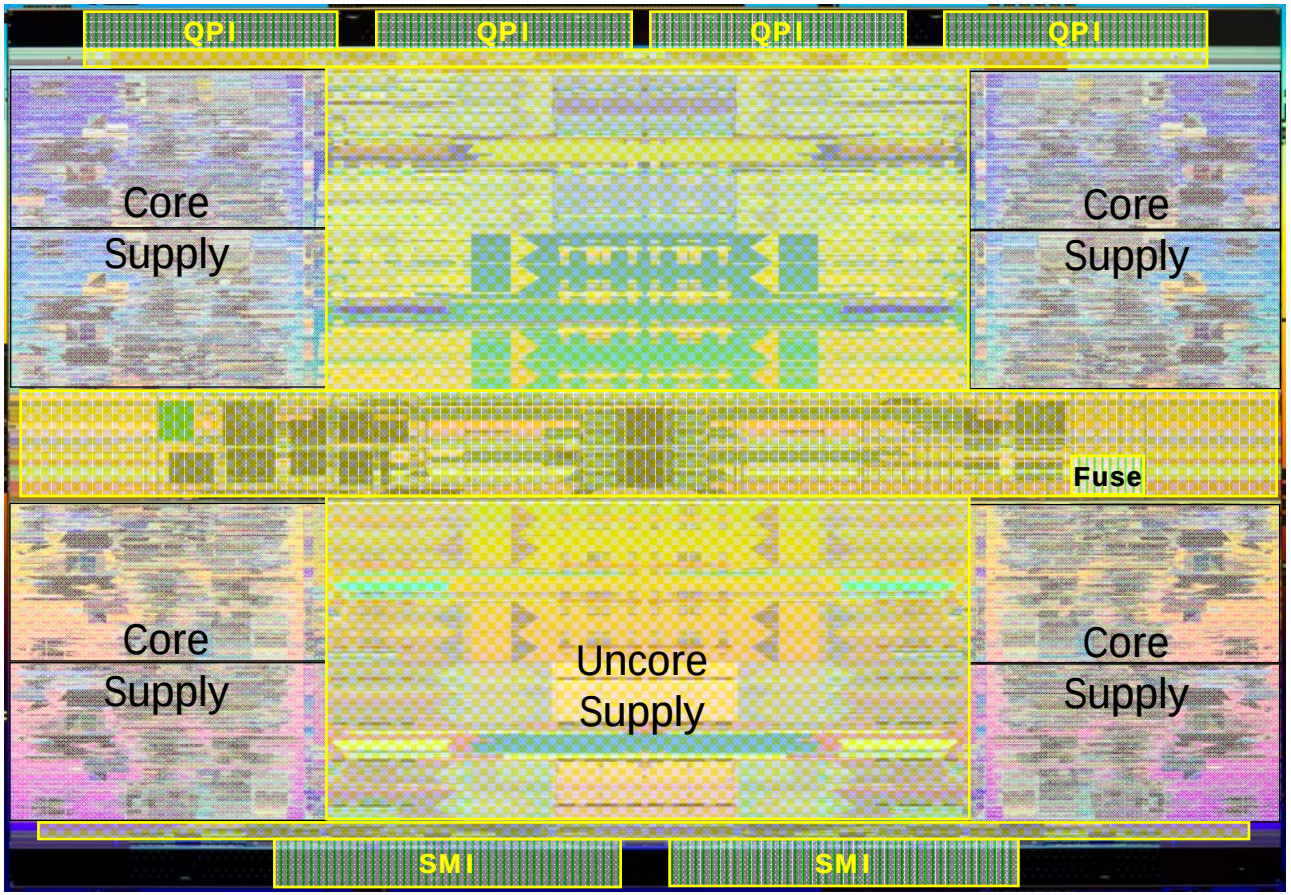


Color	Min	Max
White	0.001	2.09
Red	2.1	4.18
Pink	4.19	6.27
Orange	6.28	8.36
Yellow	8.37	10.45
Green	10.46	12.54
Cyan	12.55	14.63
Blue	14.64	16.72
Purple	16.73	18.81
Black	18.82	20.9

(ps)

- Relative un-core clock skew to ~34.8K regional clock buffer receivers
- Simulation based on 100% layout extracted model

Multiple Voltage Domains



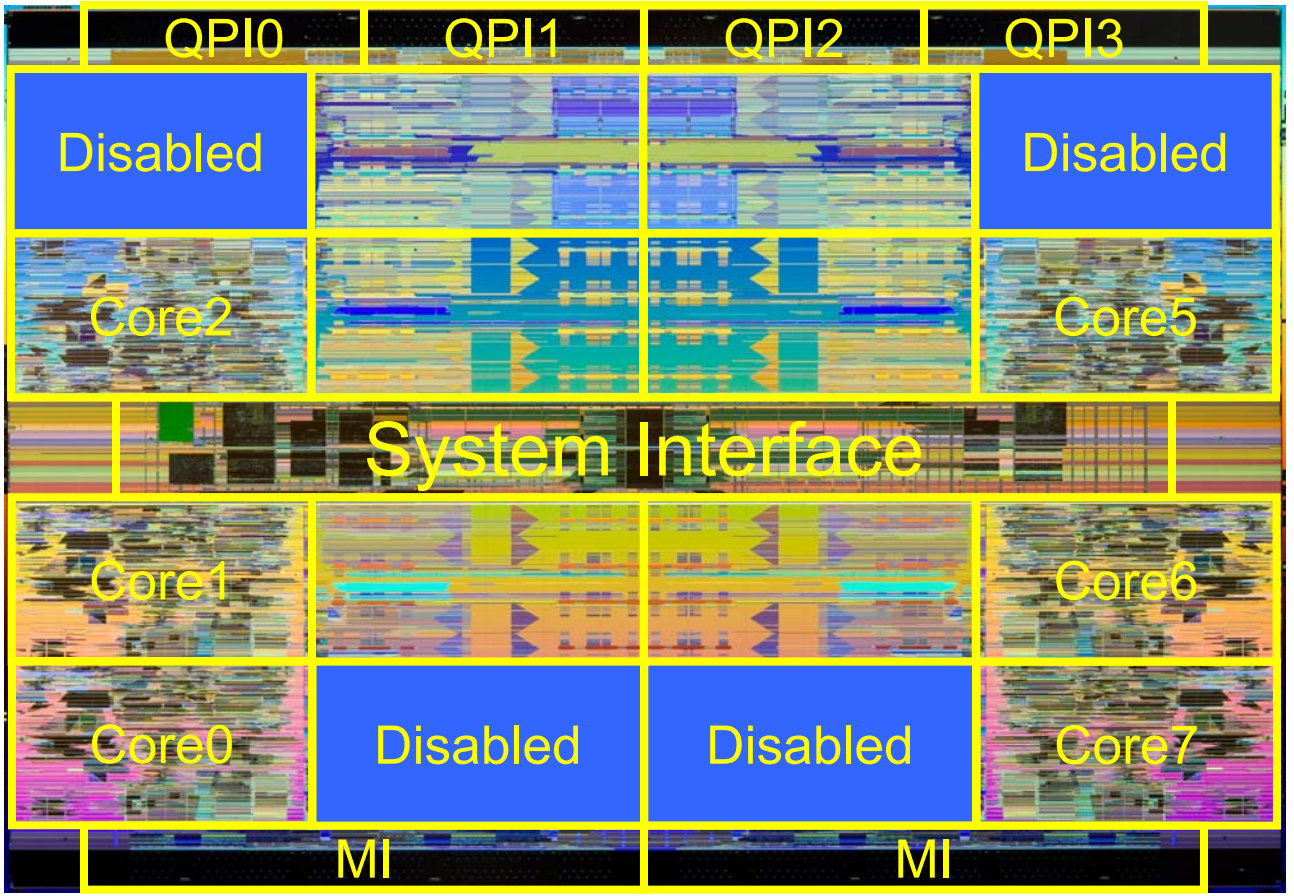
I/O Domain
1.1V
fixed

Un-Core Domain
0.9-1.1V
fixed

Core Domain
0.85-1.1V
variable

Multiple voltage domains minimize power consumption across the core and uncore areas

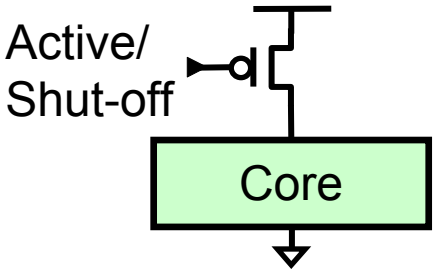
Core and Cache Recovery Example

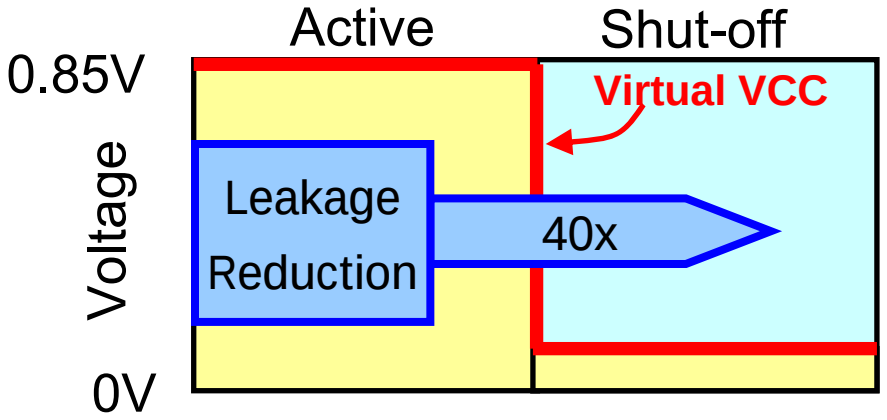


Disabled 2 cores and 2 cache slices

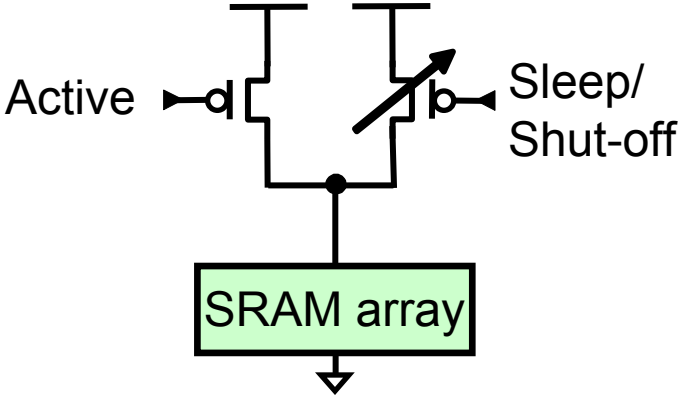
Minimize Power in Disabled Blocks

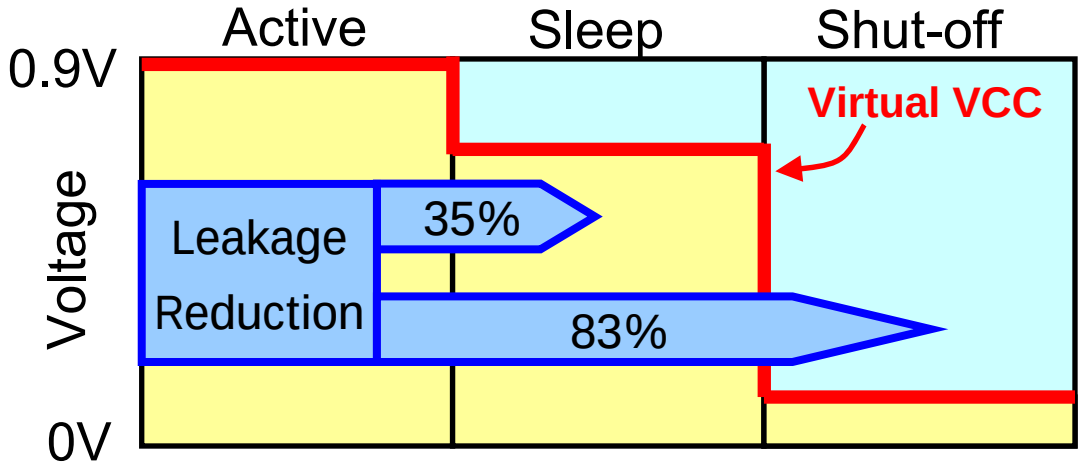
- Disabled cores ► Power gated



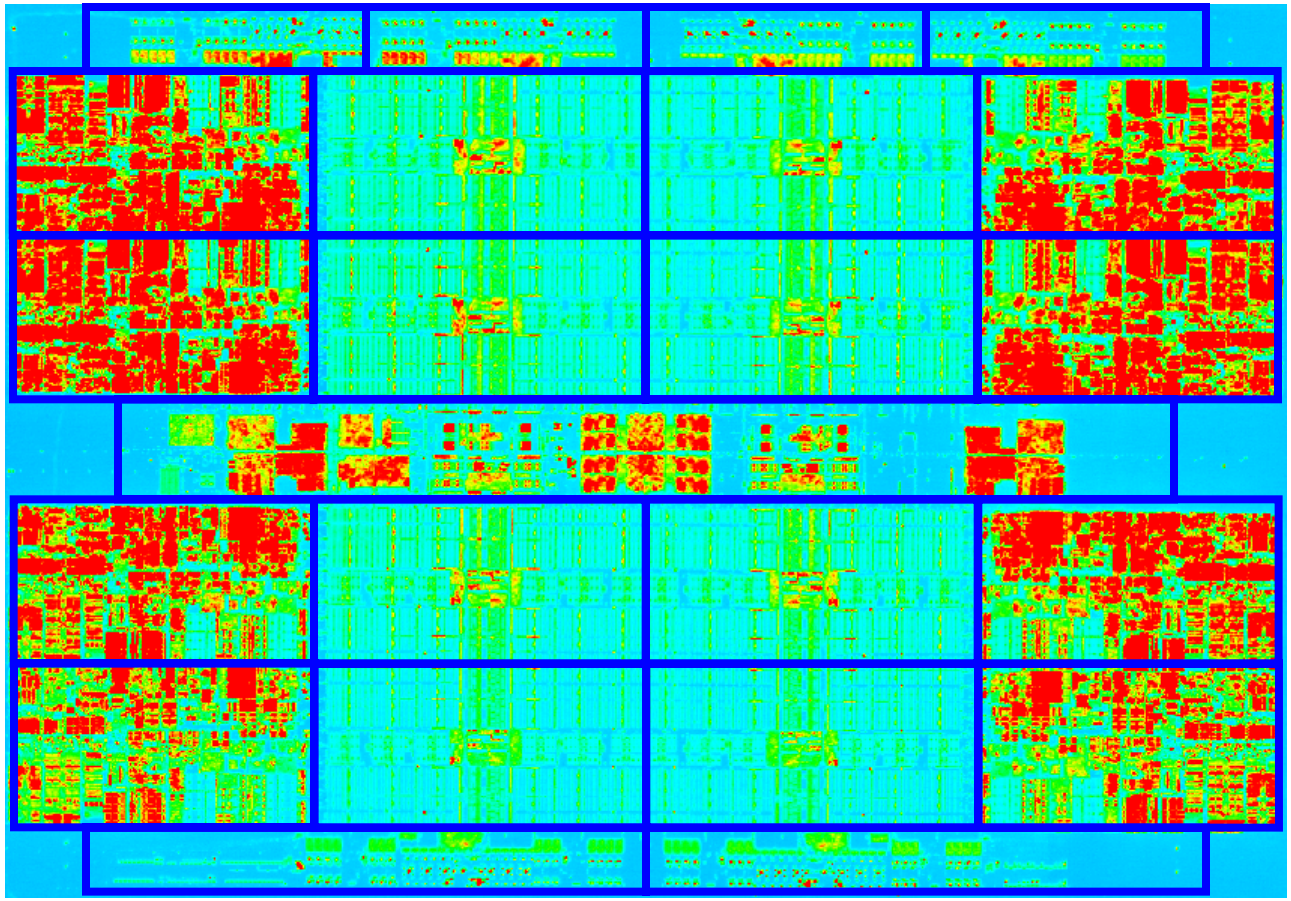


- Disabled cache slices ► All major arrays in shut-off



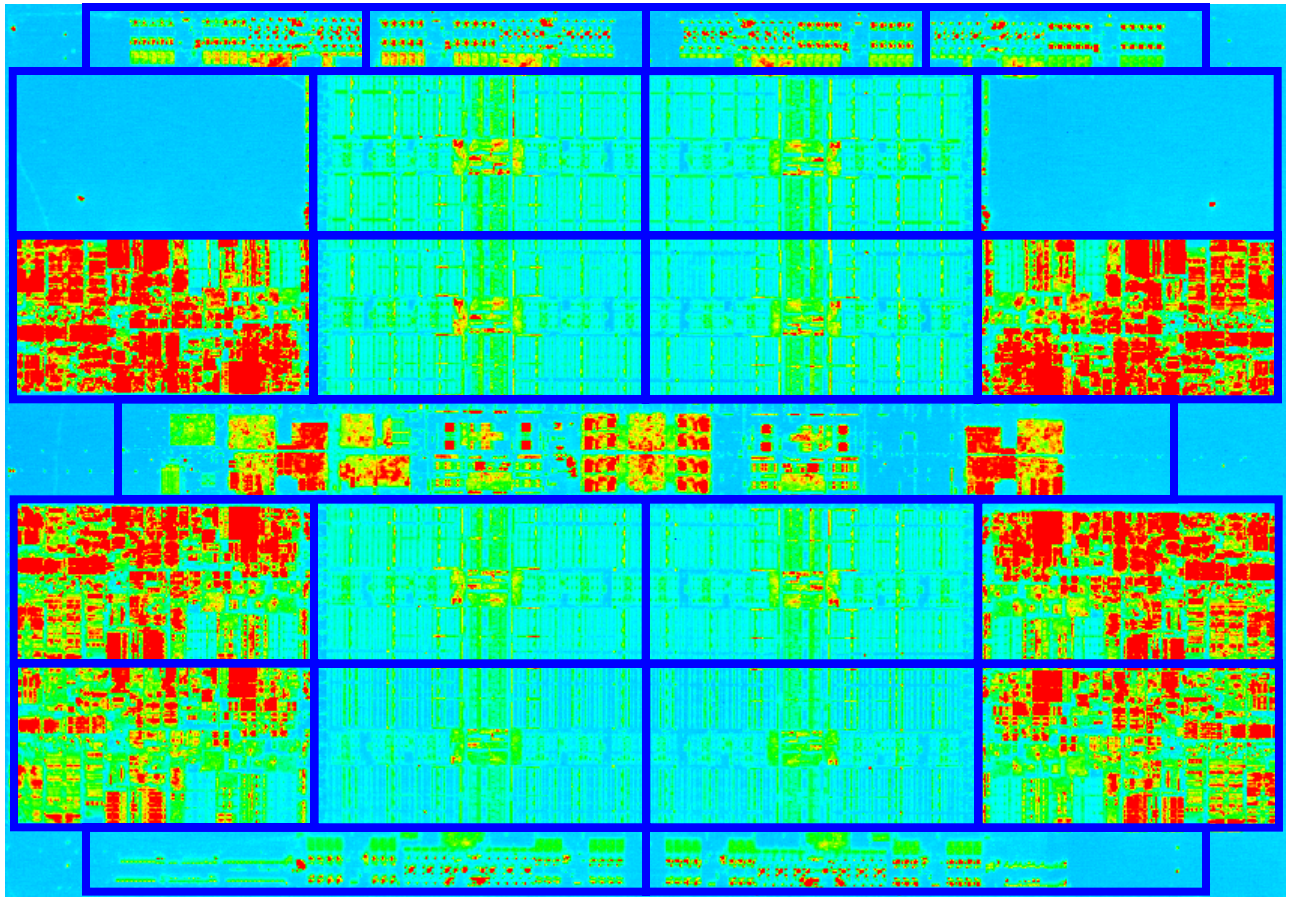


Core and Cache Recovery – Infrared Image



All cores and cache slices are enabled

Core and Cache Recovery – Infrared Image

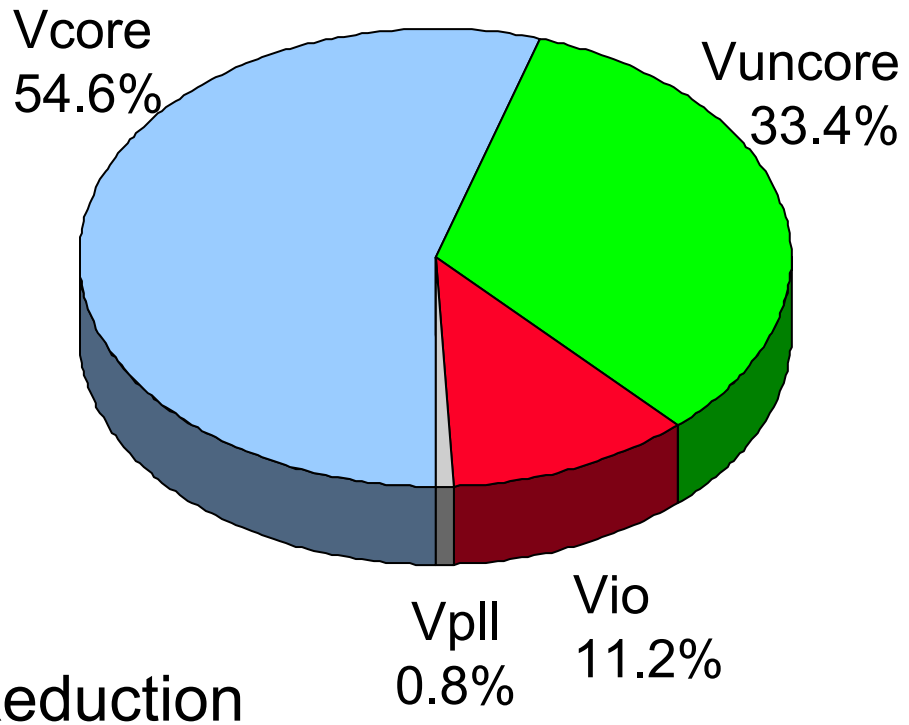


Shut-off 2 cores (top row) and 2 cache slices (bottom row)

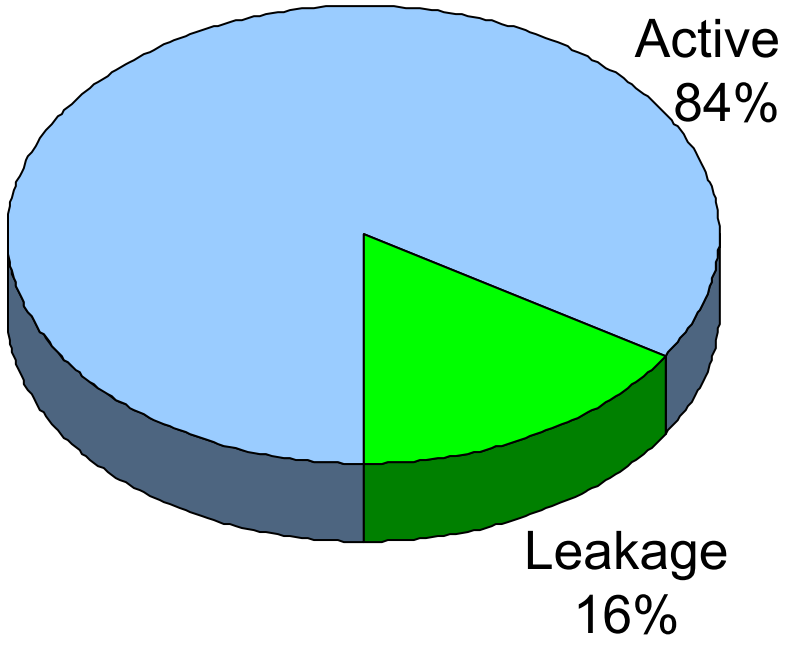
Disabled blocks are clock and power gated

Power and Leakage Breakdown

Power Breakdown



Leakage Breakdown

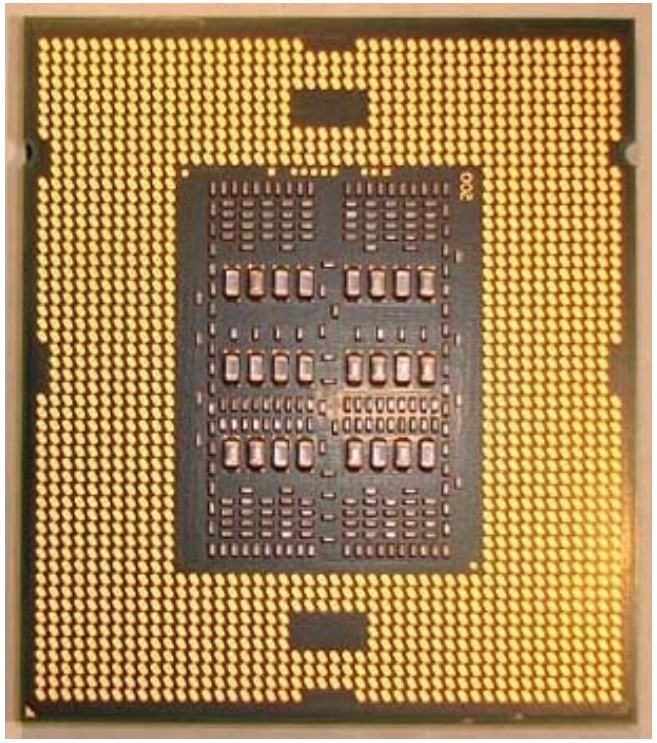


Reduction techniques

Clock gating
Run uncore at 0.9V

Long channel device usage:
58% cores, 85% uncore

Package Details



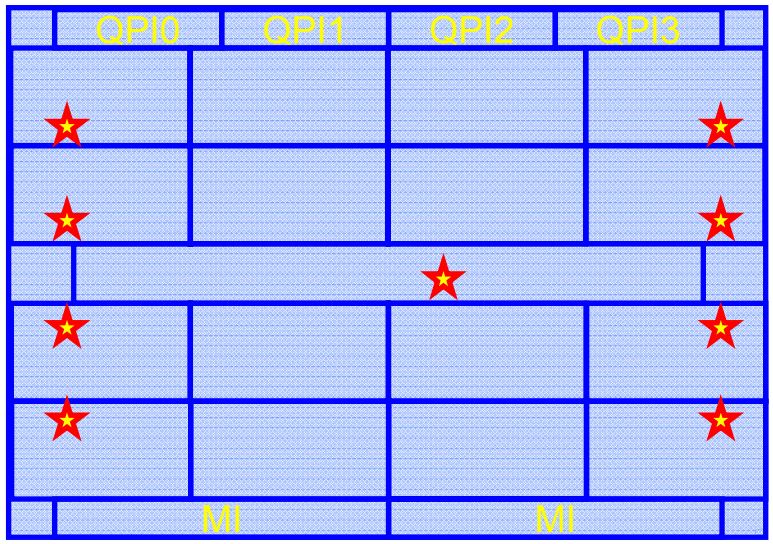
- 14 layer organic substrate
  - 49.1 x 56.4 mm
  - 5-4-5 layer stacking
- Integrated heat spreader
  - 35.5 x 43.1 mm
- System management components
  - ROM's for processor information
- 1567 total lands at 1.016 mm pitch
  - 717 signal IO's
- 32 x 24 mm cavity
  - Decoupling capacitors on package bottom directly opposite circuits
- 100% lead-free[1] and halogen-free[2]

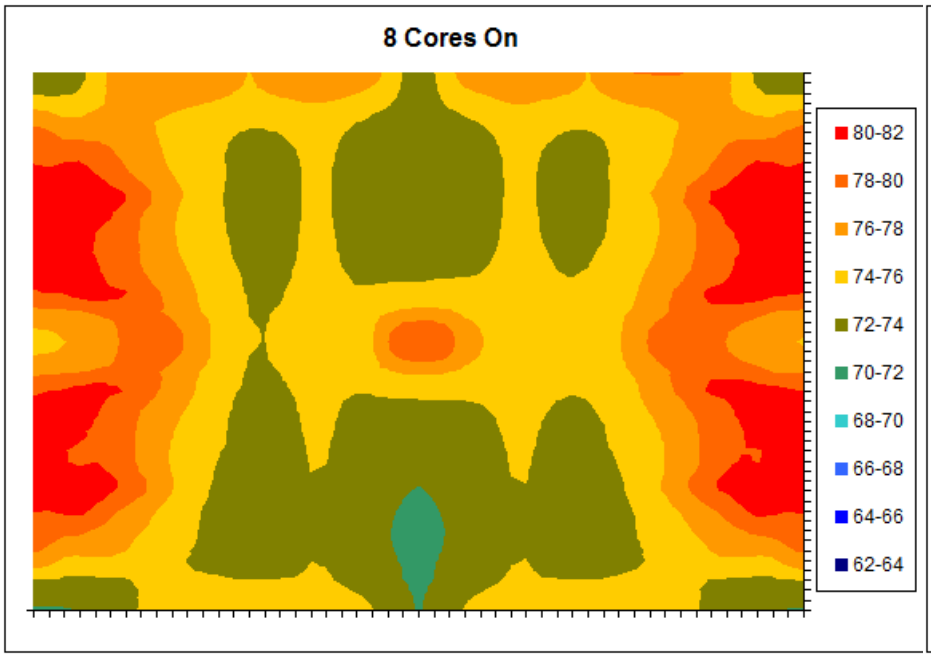
[1] 45nm product is manufactured on a lead-free process. Lead is below 1000 PPM per EU RoHS directive (2002/95/EC, Annex A). Some EU RoHS exemptions for lead may apply to other components used in the product package.

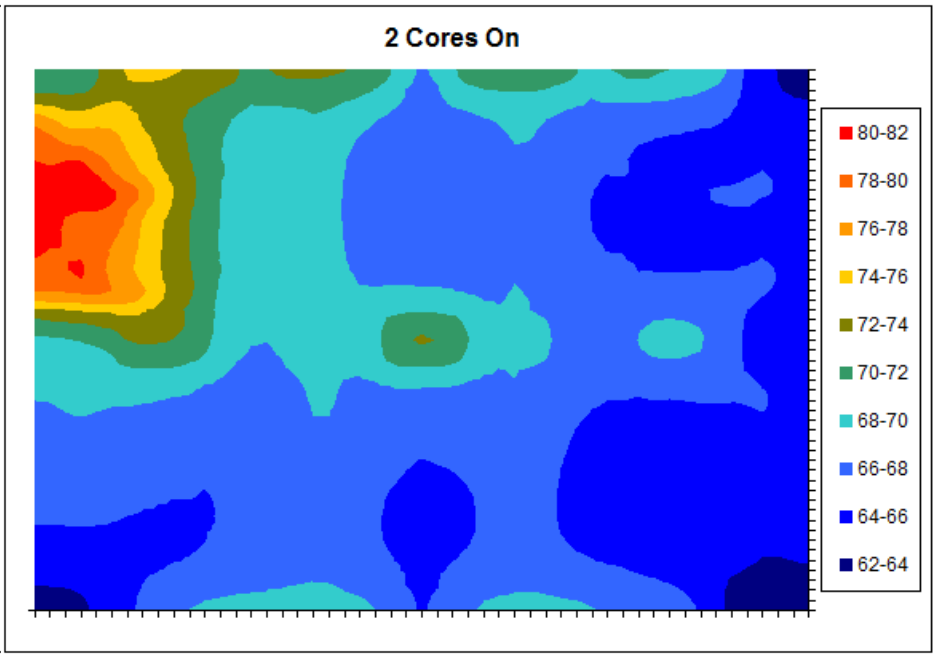
[2] Applies only to halogenated flame retardants and PVC in components. Halogens are below 900 PPM bromine and 900 PPM chlorine.

Thermal Sensors

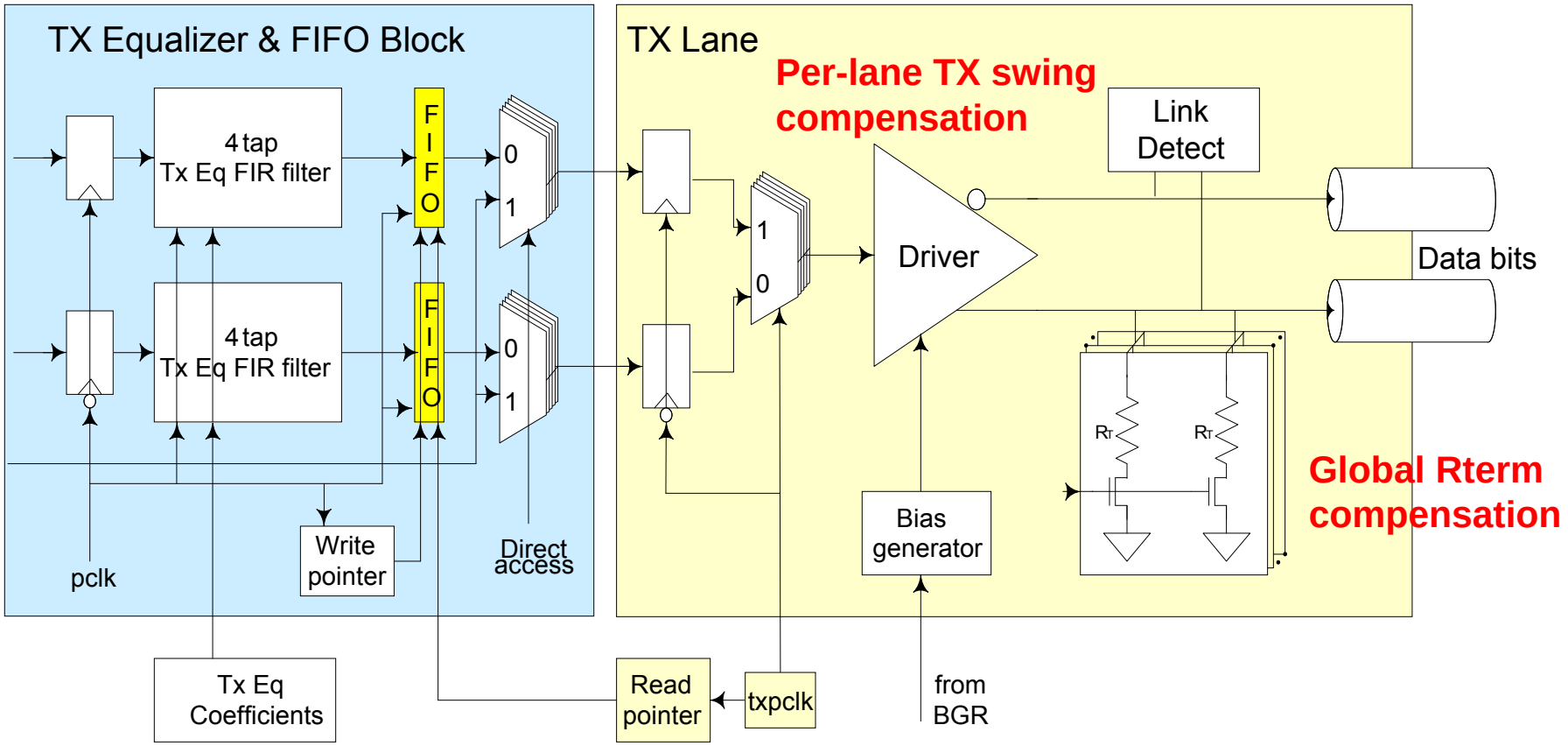
- 9 temperature sensors
  - One in each core hot spot
  - One in the die center
  - Temperature information is available through the PECI bus for system fan management
- Large die with cores spread apart is relatively easy to cool



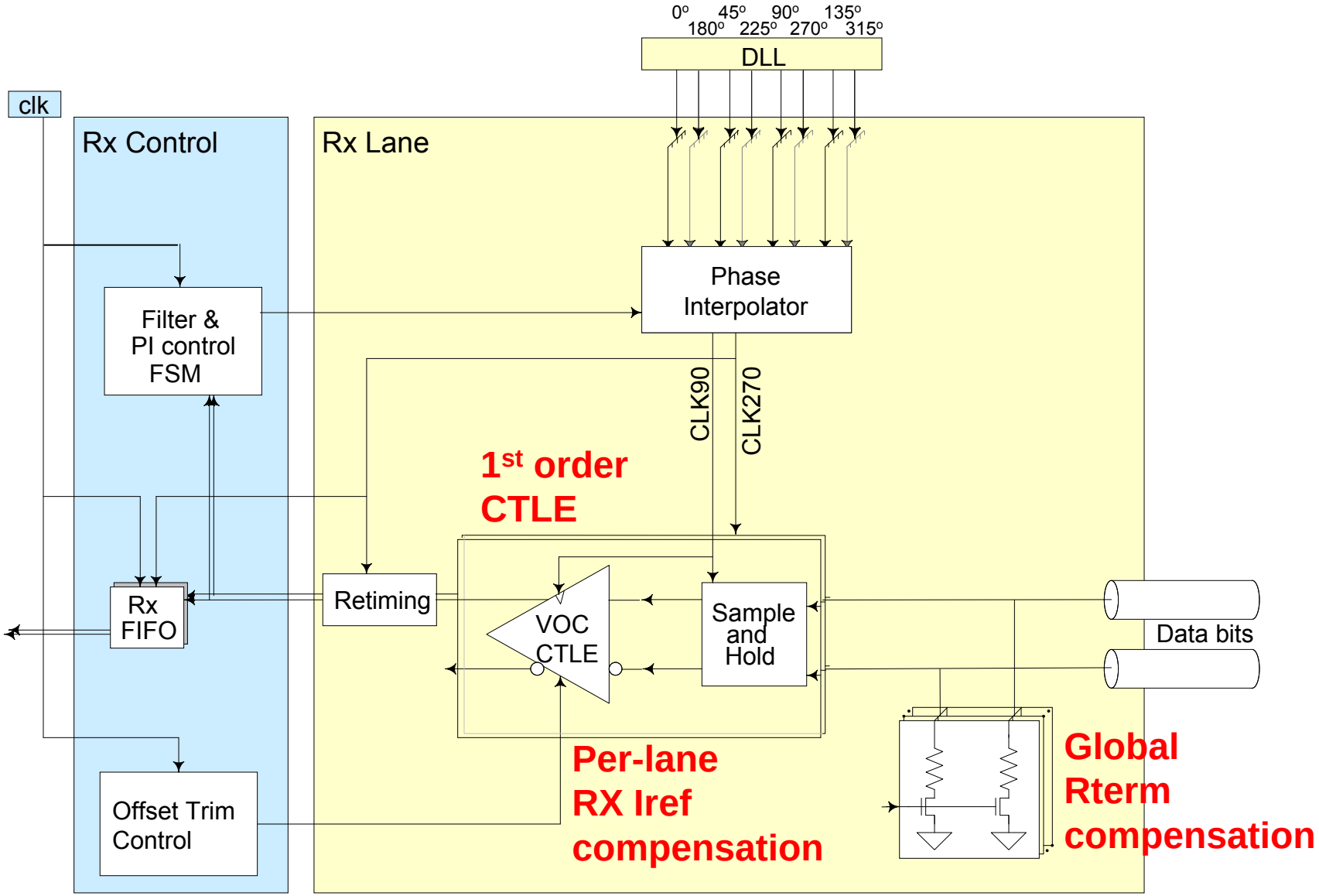




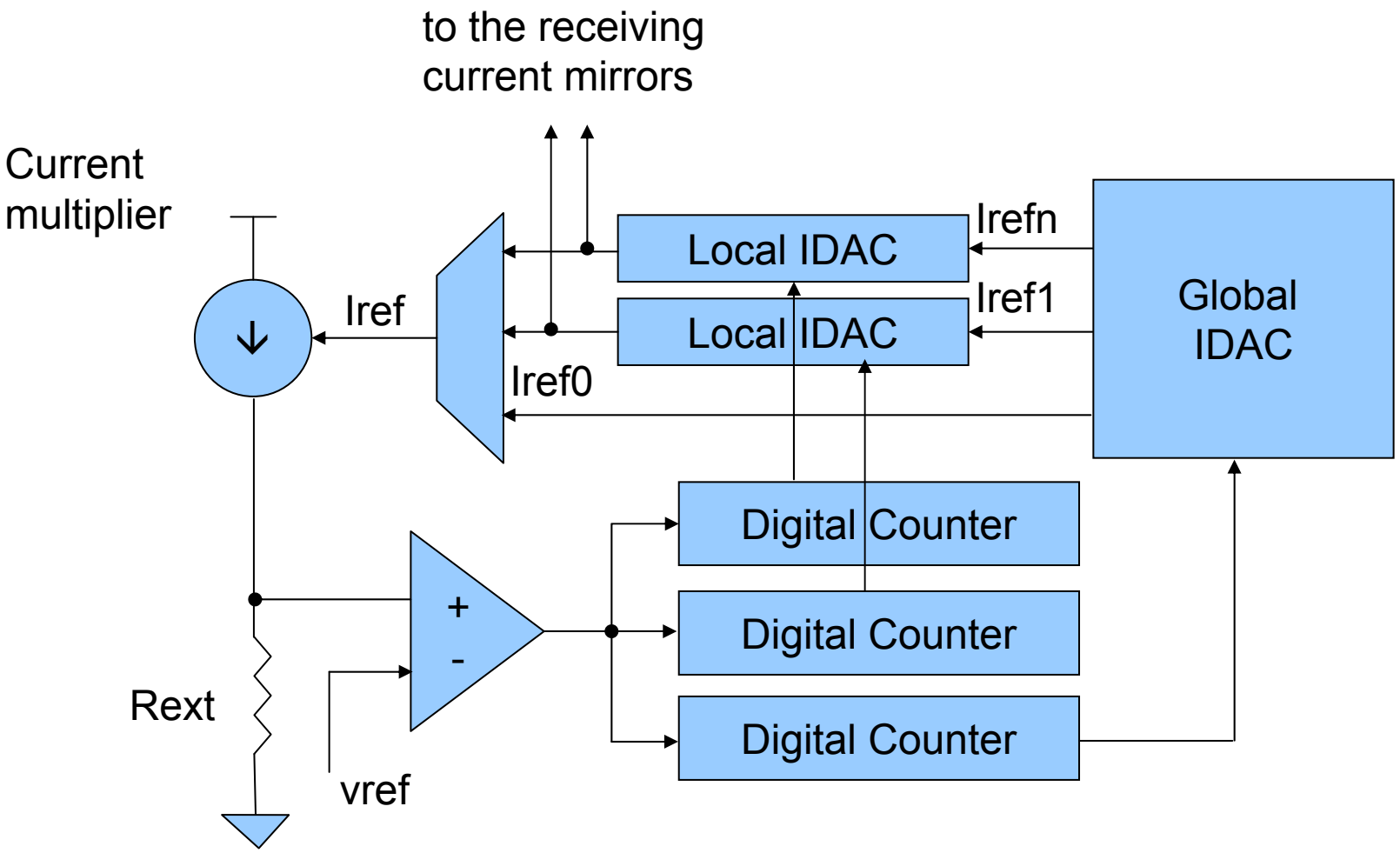
TX Block Diagram



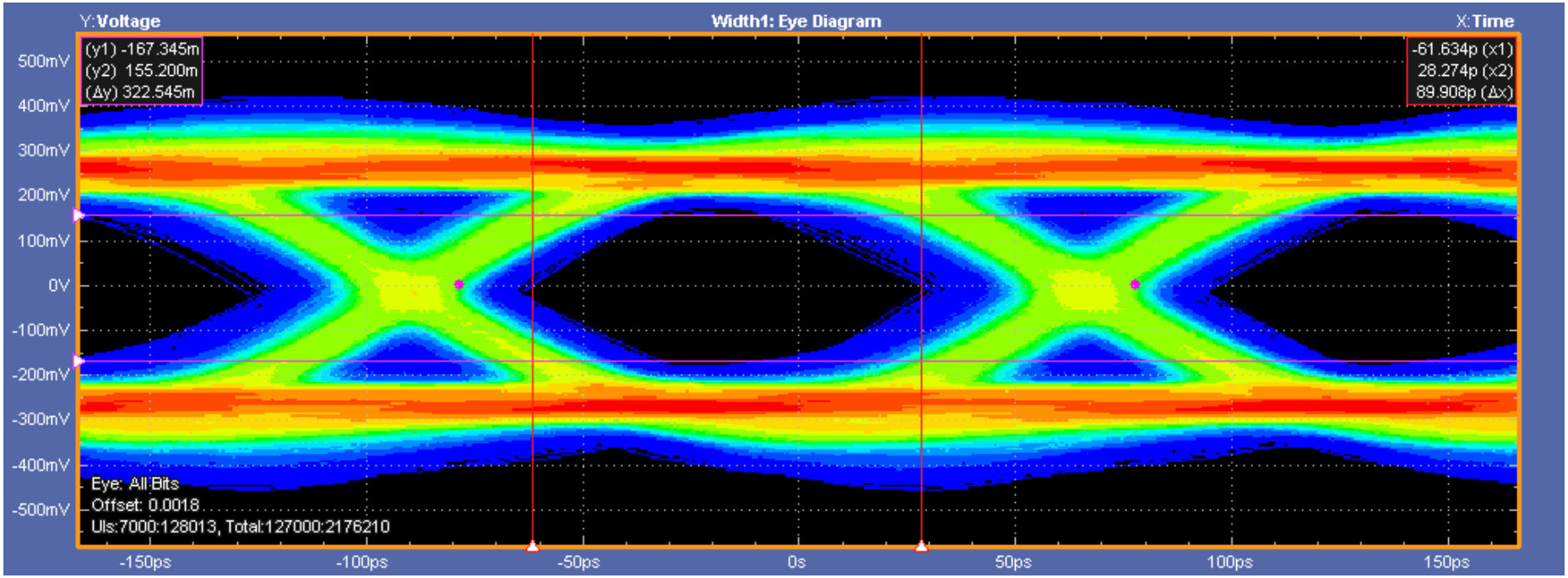
RX Block Diagram



I/O Reference Current Compensation

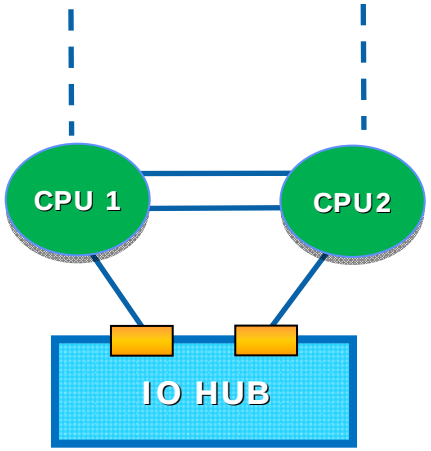


QPI Eye Diagram

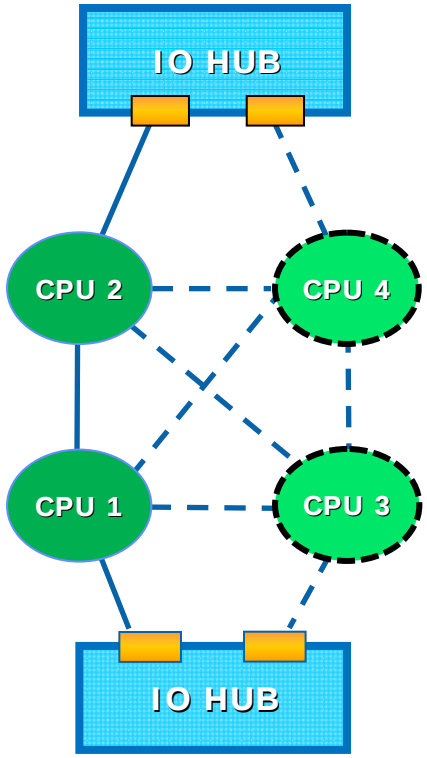


Eye diagram captured on 14”, 1-connector QPI link at 6.4GT/s

Disable Unused QPI ports

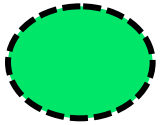


Dual processor
with one IOH



Partially populated
4 socket board

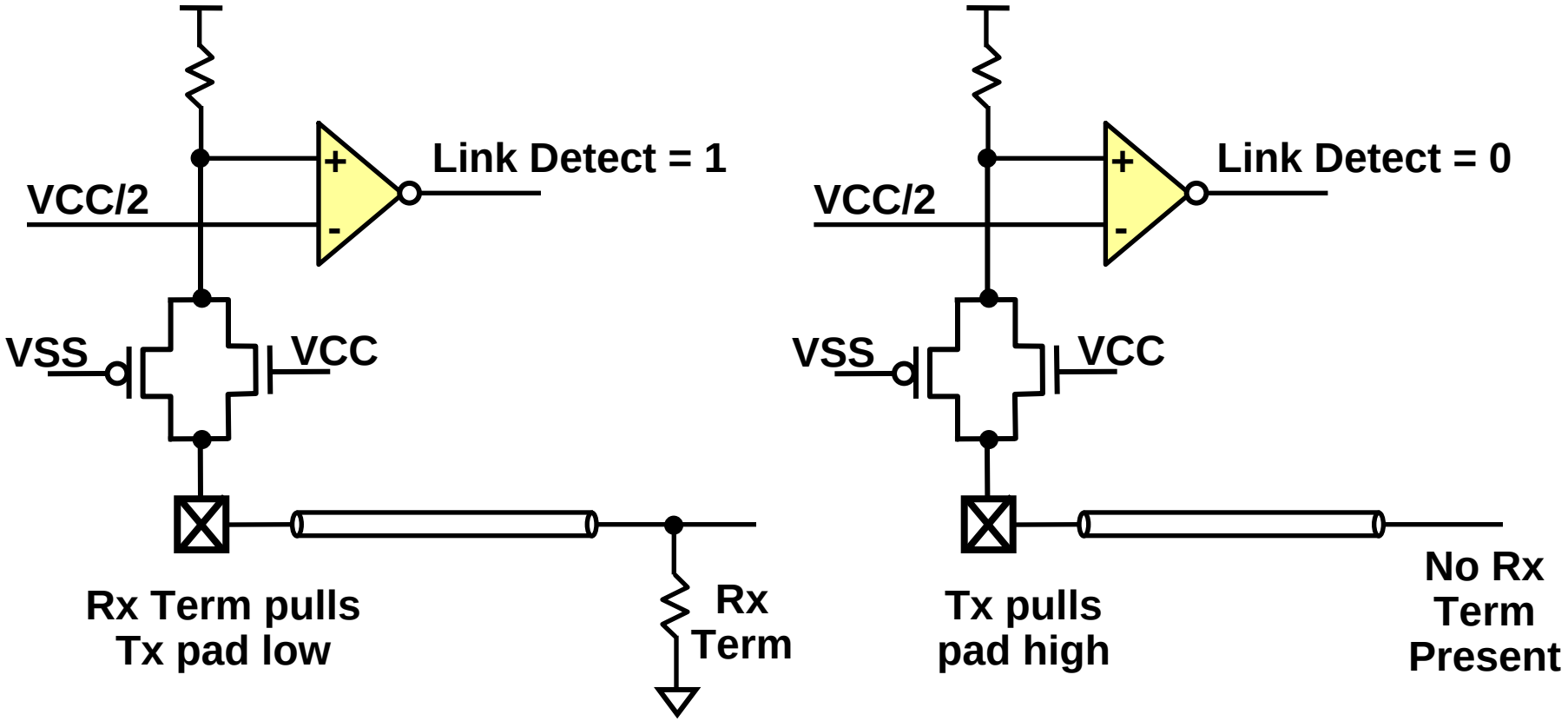
Legend:



Unpopulated
Socket

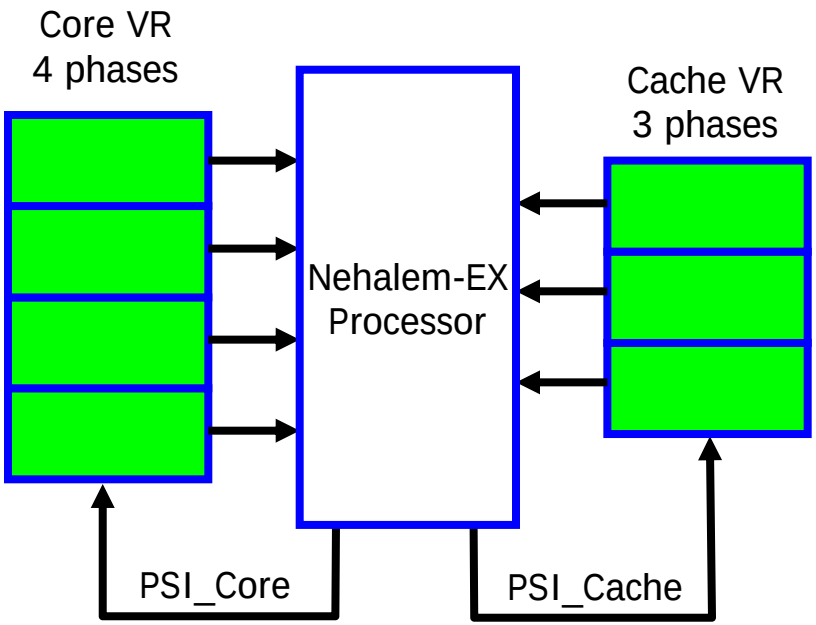
Unused
QPI links

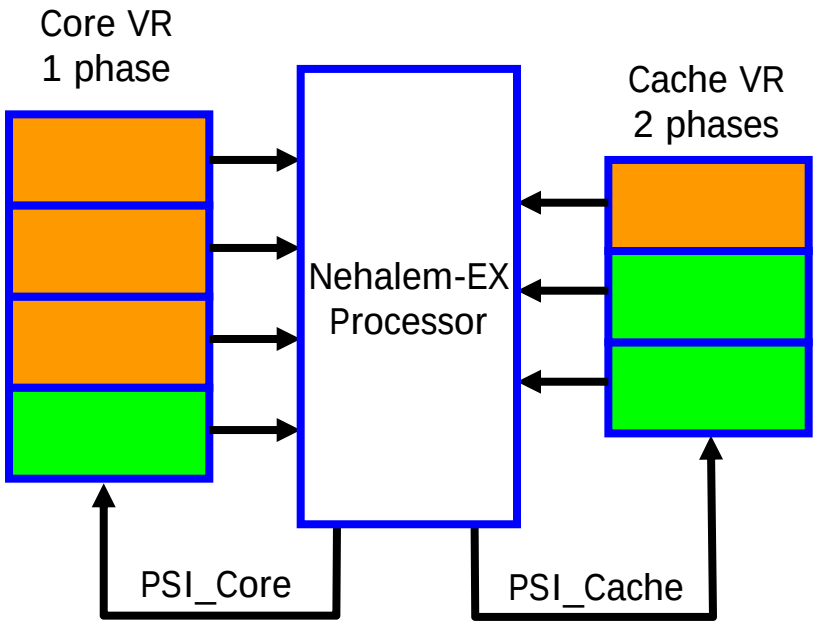
Link Detect Circuit



- Link detector senses unused links at power-on and disables them to save ~2W per port
  - Shut-off all driver bias currents
  - Turn-off the PLL to stop the clock

Load Adaptive Voltage Regulation





Full Load Mode

- All VR phases are enabled
- Maximum VR efficiency

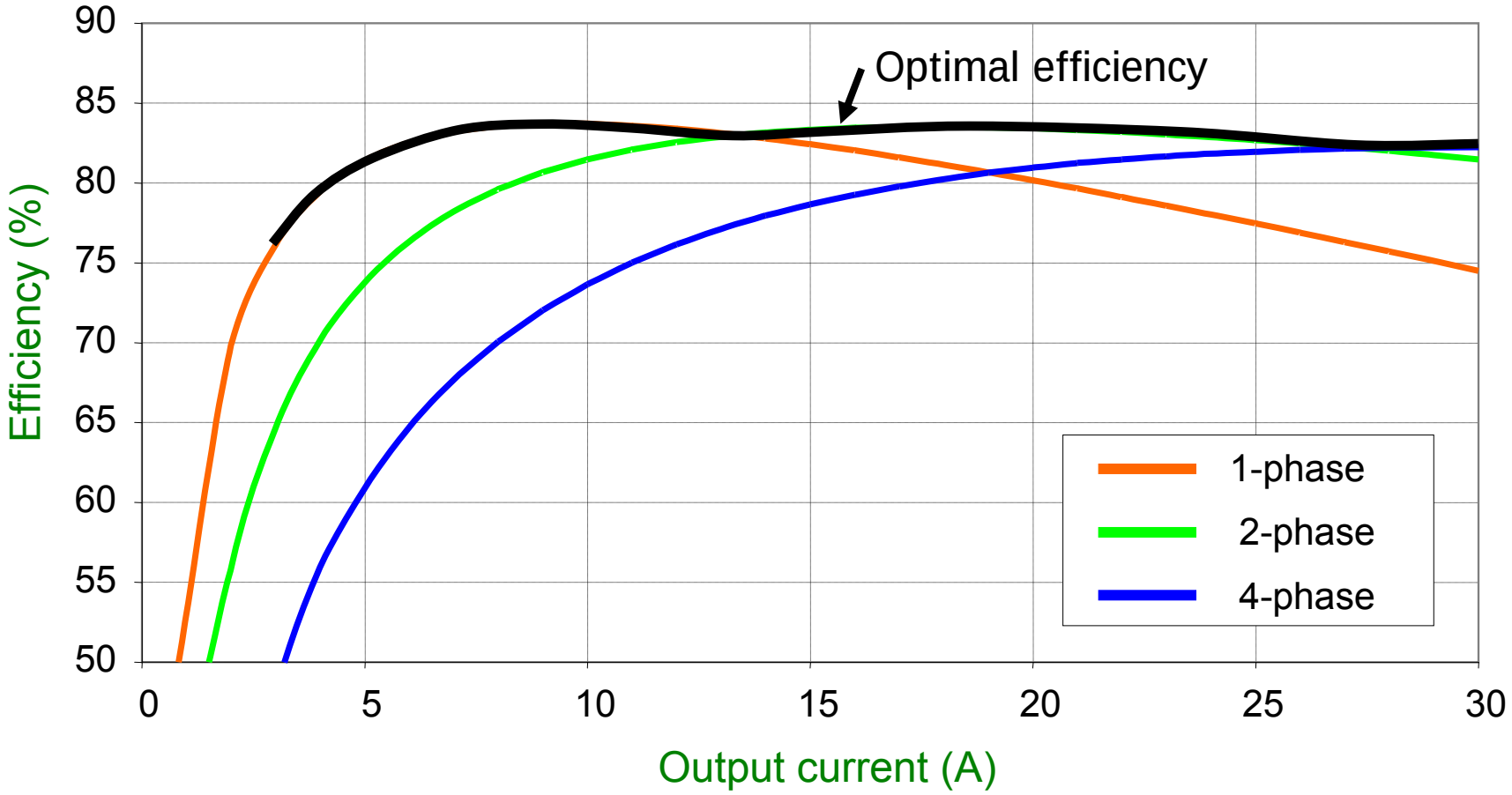
Idle Mode

- Turn off 3 core and 1 cache phases
- Maximum VR efficiency

Nehalem-EX extends the VR phase shut-off to the cache supply

About 2W power reduction per socket in idle mode

Load Efficiency of 4-phase Voltage Regulator



Shut-off VR phases in idle mode to improve efficiency

Summary

- Enterprise-class 45nm 8-cores, 16-threads Xeon® Processor with 24MB on-die shared L3 cache
  - Largest transistor count for a microprocessor
  - Two integrated memory controllers
  - Four point-to-point links at 6.4GT/s with per-lane compensation
- Active power and leakage reduction techniques
  - 45nm High-K Metal Gate process reduces leakage
  - Multiple voltage and clock domains minimize power consumption
  - Operate at the lowest possible voltage
  - Extensive use of long channel devices
  - Automatically disable unused QPI ports at power-on
  - Load adaptive voltage regulation reduces idle power
- Core and cache recovery enables multiple product options
  - Disabled cores and cache slices are clock and power gated