CASPER

CS773: Computer Architecture for Performance and Security

Lecture 7: 10K Feet View on the Power-efficient Architecture

https://www.cse.iitb.ac.in/~biswa/

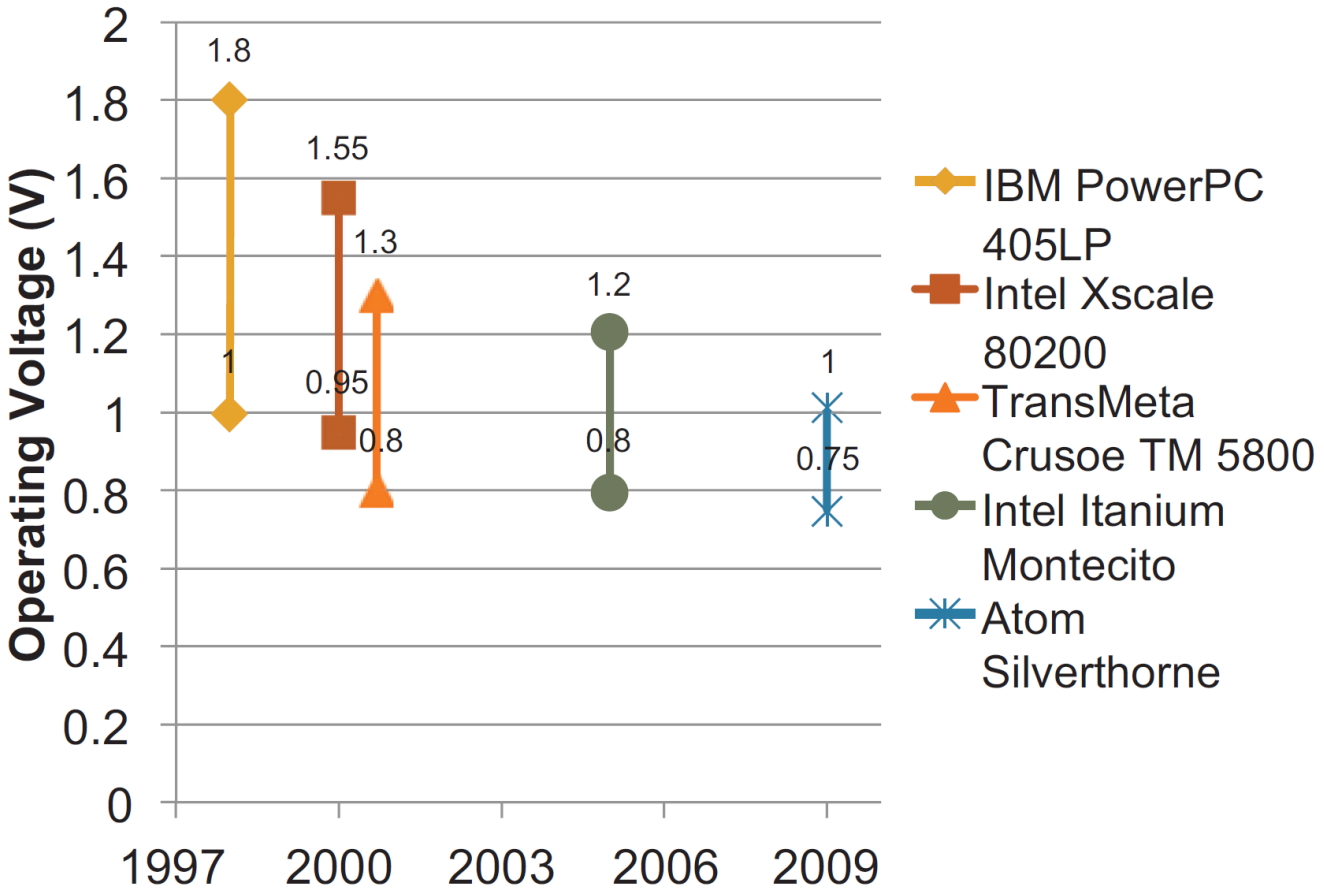
Power

Dynamic: CV^2f (frequency is also dependent on the voltage)

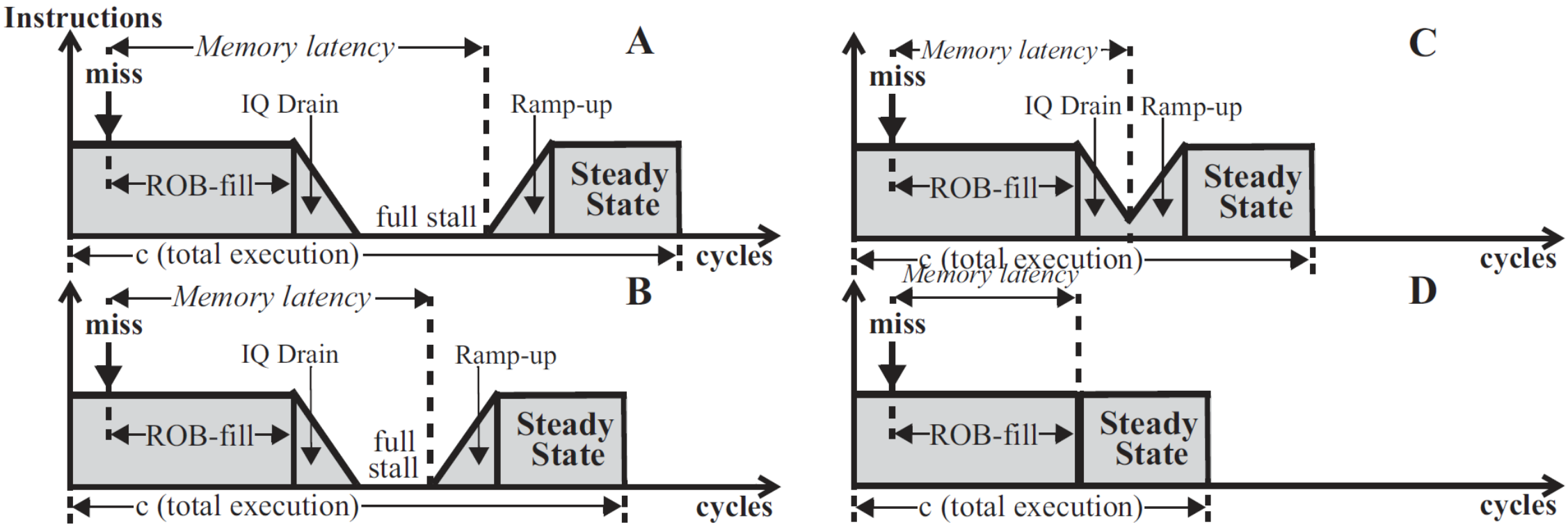
high frequency will also demand high voltage

Static: Leakage power

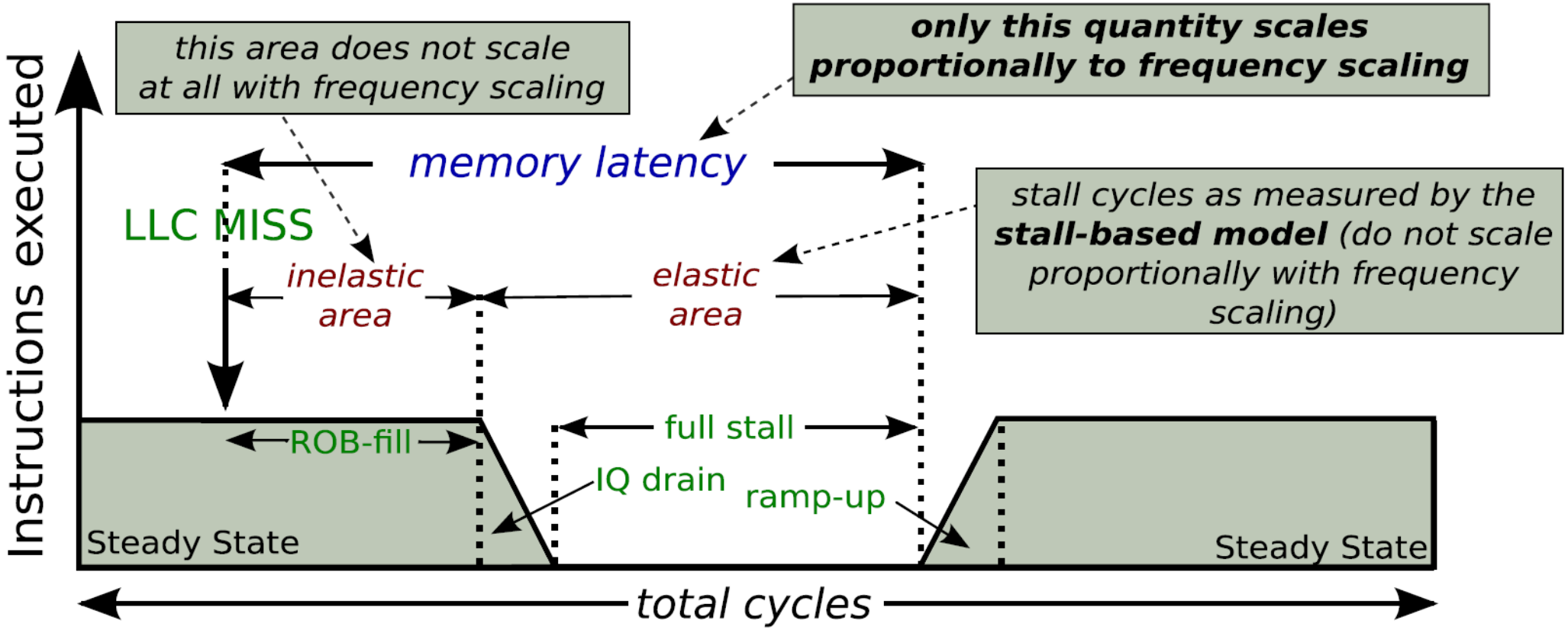
Dynamic Power



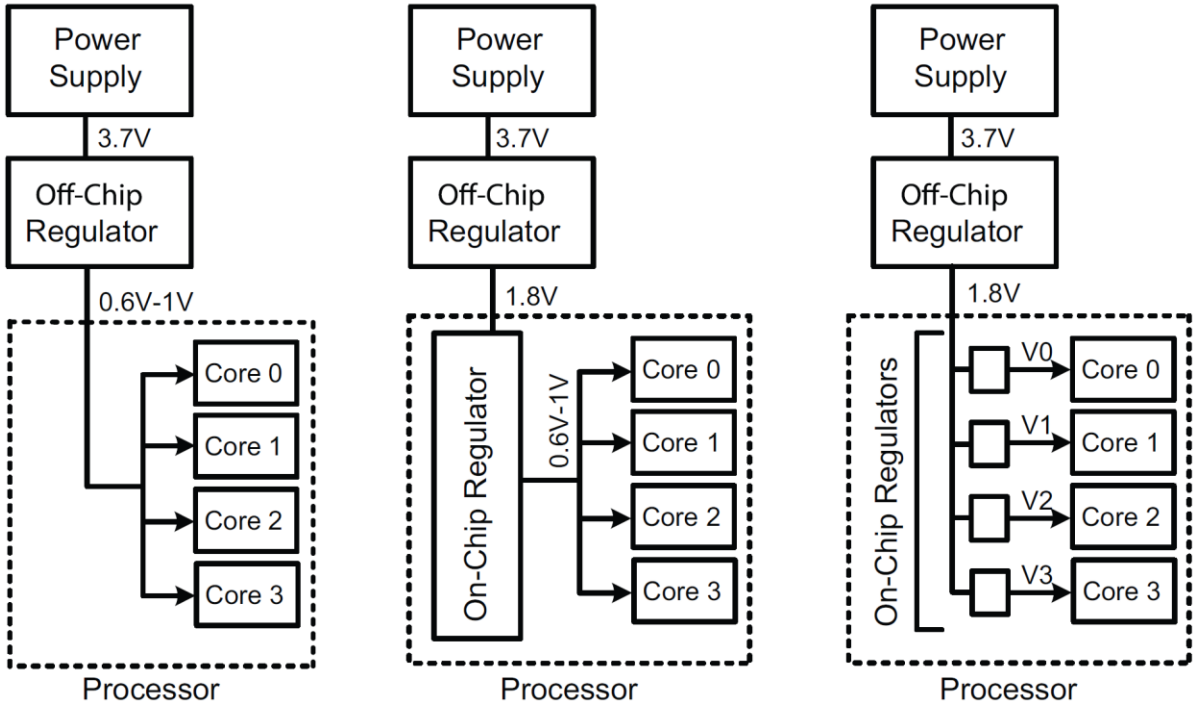
Effect of Frequency



A bit Deeper


this area does not scale at all with frequency scaling
only this quantity scales proportionally to frequency scaling
memory latency
LLC MISS
inelastic area
elastic area
stall cycles as measured by the stall-based model (do not scale proportionally with frequency scaling)
Instructions executed
ROB-fill
full stall
IQ drain
ramp-up
Steady State
Steady State
total cycles

DVS (Dynamic Voltage Scaling)



DFS Governors

Linux: cpufreq component (/sys/devices/system/cpu/): Performance, Powersave, Ondemand, Conservative, and Interactive

Hardware-managed P states (performance states)

Performance scaling ~ Frequency scaling

Per-core P states (PCPS)

Static Power (Decay Cache)

Put k-ways into the sleep mode when memory activity is low (power gating). Decay Cache, State-destroying

Example: 8MB 16-way cache becomes 512kB 2-way (intel IvyBridge)

Static Power (Drowsy Mode)

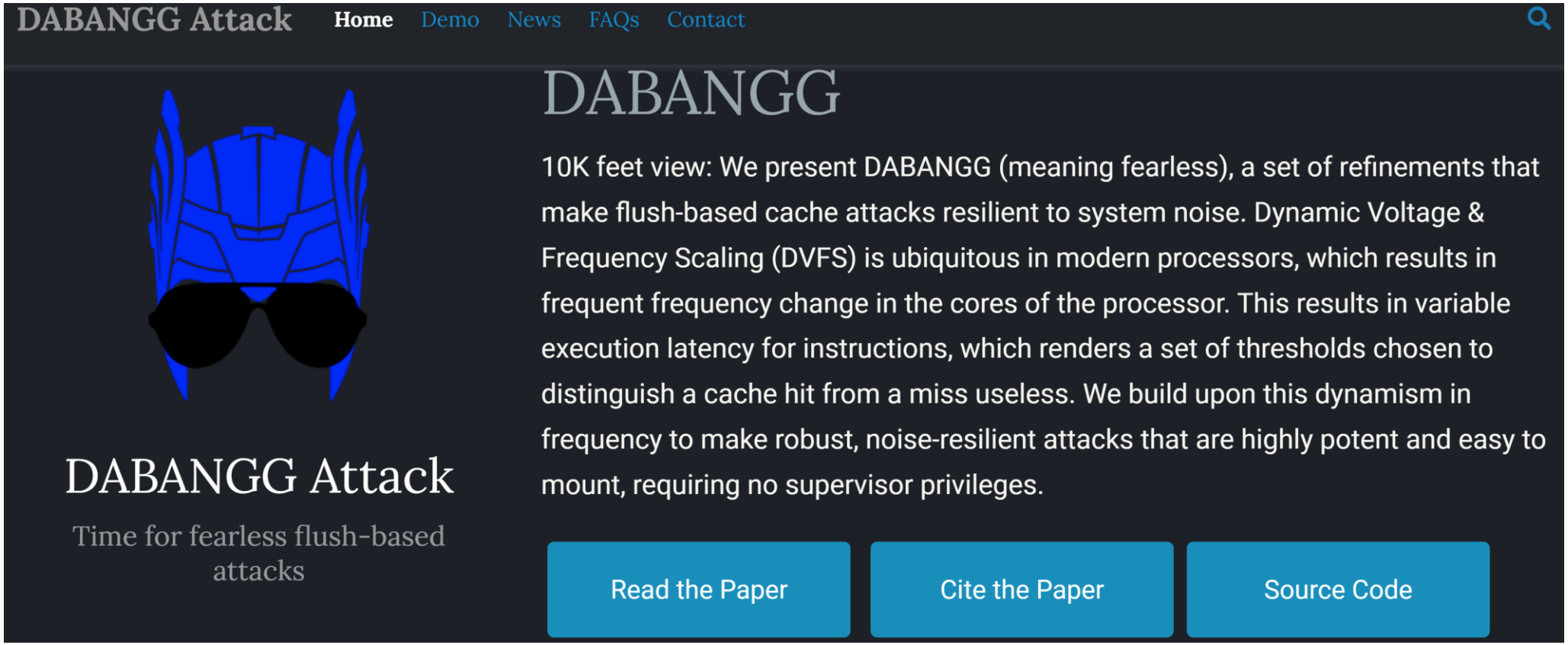
Drowsy mode: cache lines into low-power mode (line retains the data but can’t access it ☺, drowsy-bit per cache line), State-preserving

Switched to normal mode: If cache line accessed again

DVFS and Security

More papers in the coming weeks

DVFS and Cache Attacks



To appear in WOOT 2022

https://www.cse.iitb.ac.in/~biswa/DABANGG.pdf

Performance and Power

Delay = runtime

Energy = Watts * runtime

EDP (Energy Delay Product) = Watts * runtime * runtime

ED^2P and ED^3P

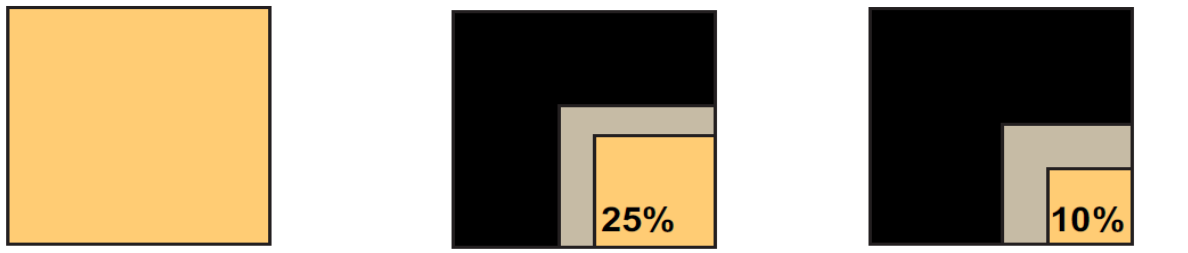
Dark Silicon

Dark Silicon

- Due to the breakdown of Dennardian scaling, the percentage of a silicon chip that can switch at full frequency is dropping exponentially with each process generation (45nm to 22nm to 11nm);larger fractions of a chip’s transistors become dark. RIP Multi-core systems ☹

Dark Silicon (End of Dennard Scaling)

Node	45nm	22nm	11nm
Year	2008	2014	2020
Area^{-1}	1	4	16
Peak freq	1	1.6	2.4
Power	1	1	0.6
		$(4 \times 1)^{-1} = 25\%$	$(16 \times 0.6)^{-1} = 10\%$



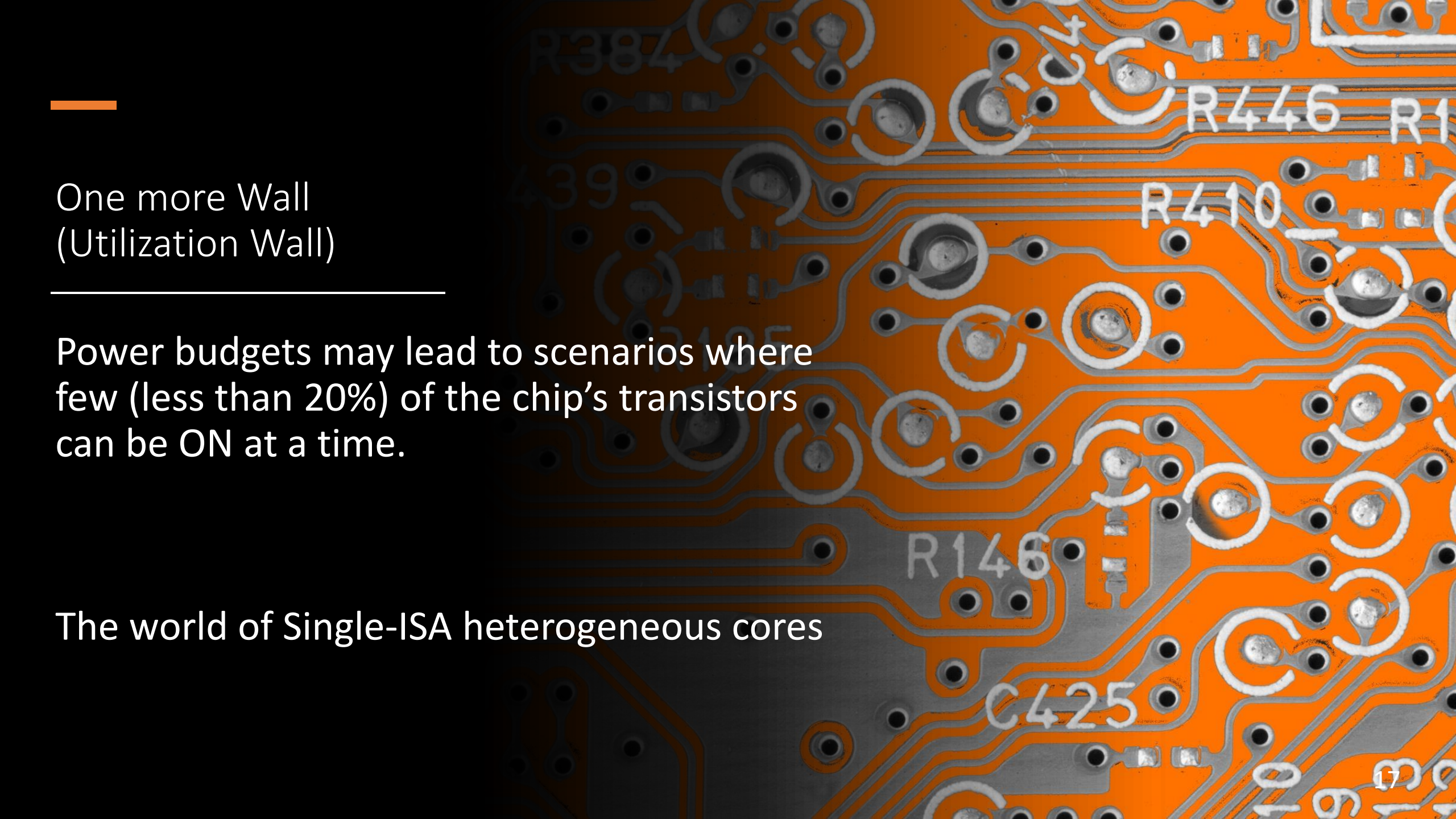
Source = ITRS 2008


Why did it happen? Piazza +1

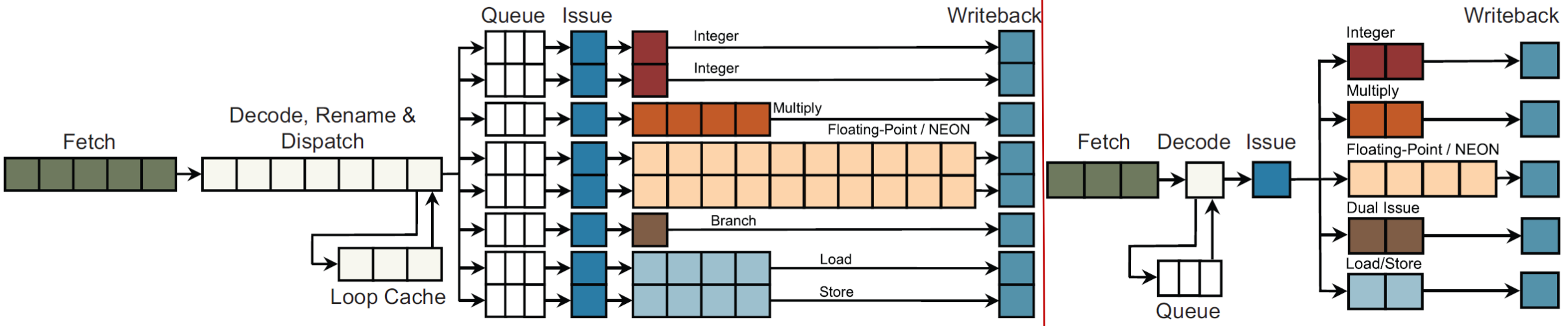
One more Wall (Utilization Wall)

Power budgets may lead to scenarios where few (less than 20%) of the chip’s transistors can be ON at a time.

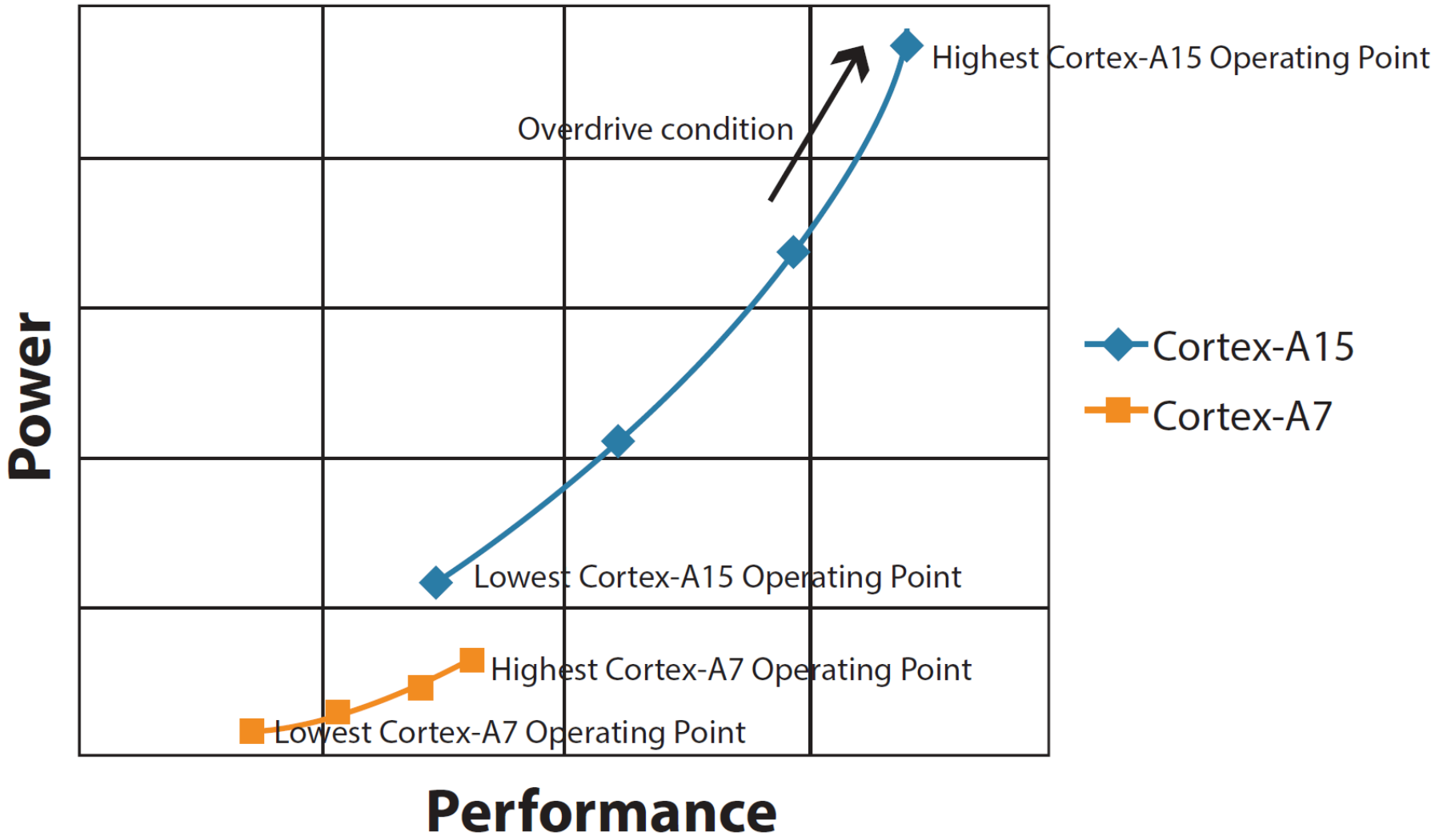
The world of Single-ISA heterogeneous cores



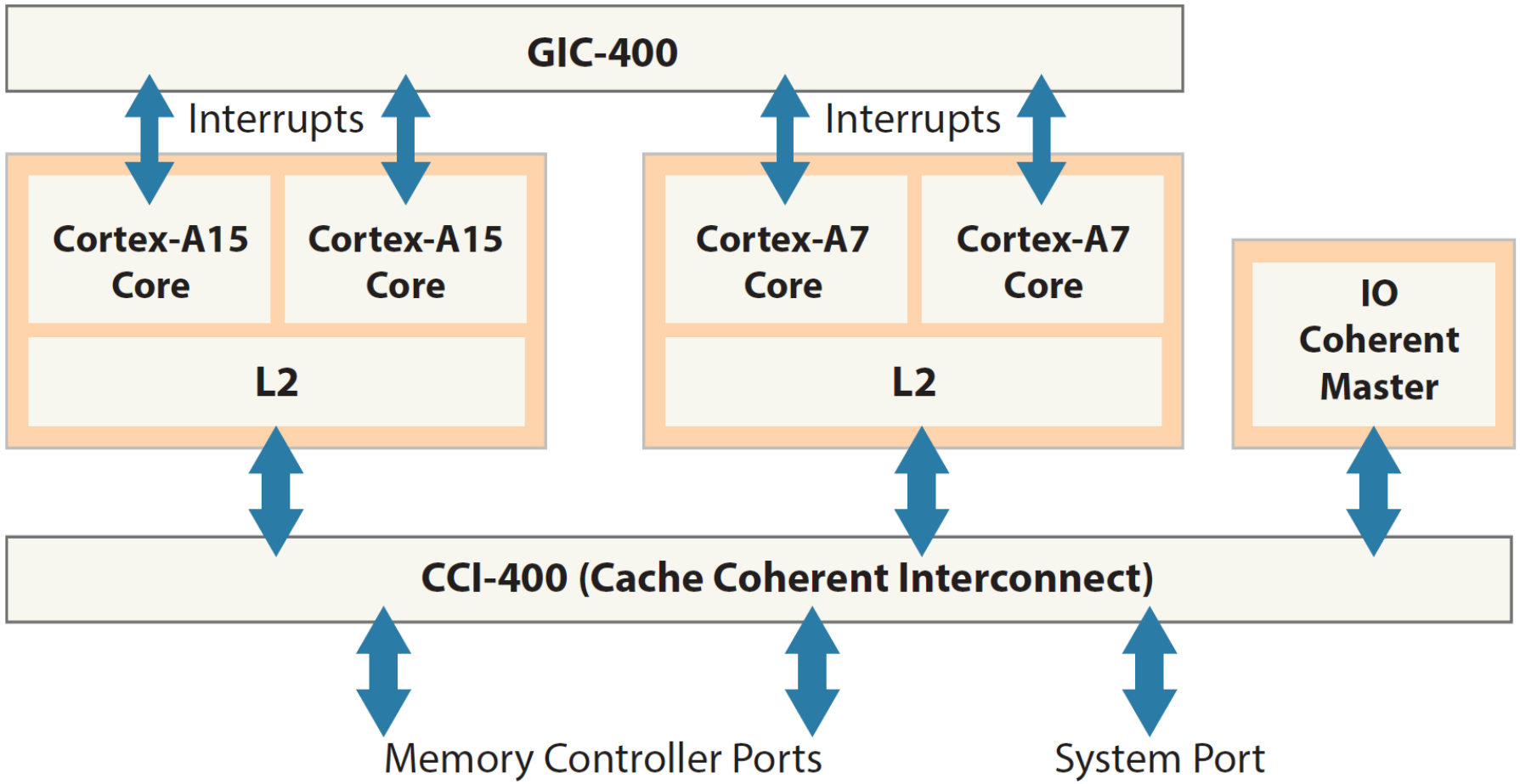
ARM Big/Little



Power/Performance Tradeoff



Big/Little Multicore (Based on P-state)



Thermal/Co2 emission ☹

Some other day!

Thanks