



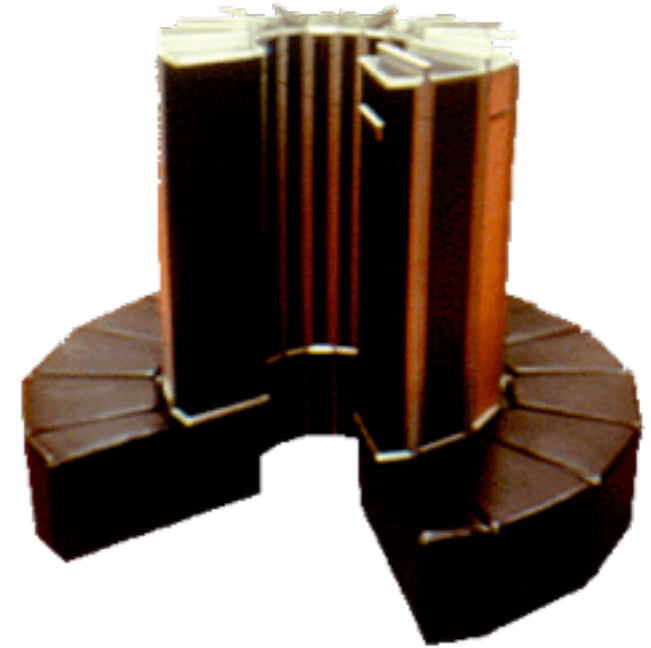
Mitigating the Effects of Process Variation in Ultra-low Voltage Chip Multiprocessors using Dual Supply Voltages and Half-Speed Stages

Timothy N. Miller, Renji Thomas, Radu Teodorescu
*Department of Computer Science and Engineering
The Ohio State University*



<http://arch.cse.ohio-state.edu>

Universal Demand for Low Power



Apple, Inc.

<http://keetsa.com/blog/wp-content/uploads/2008/03/eco-friendly.jpg>

http://laptoping.com/wp-content/OLPC_XO_Laptop.jpg

<http://logic.stanford.edu/talks/Web2.0/Cray-1-Supercomputer-1976.gif>

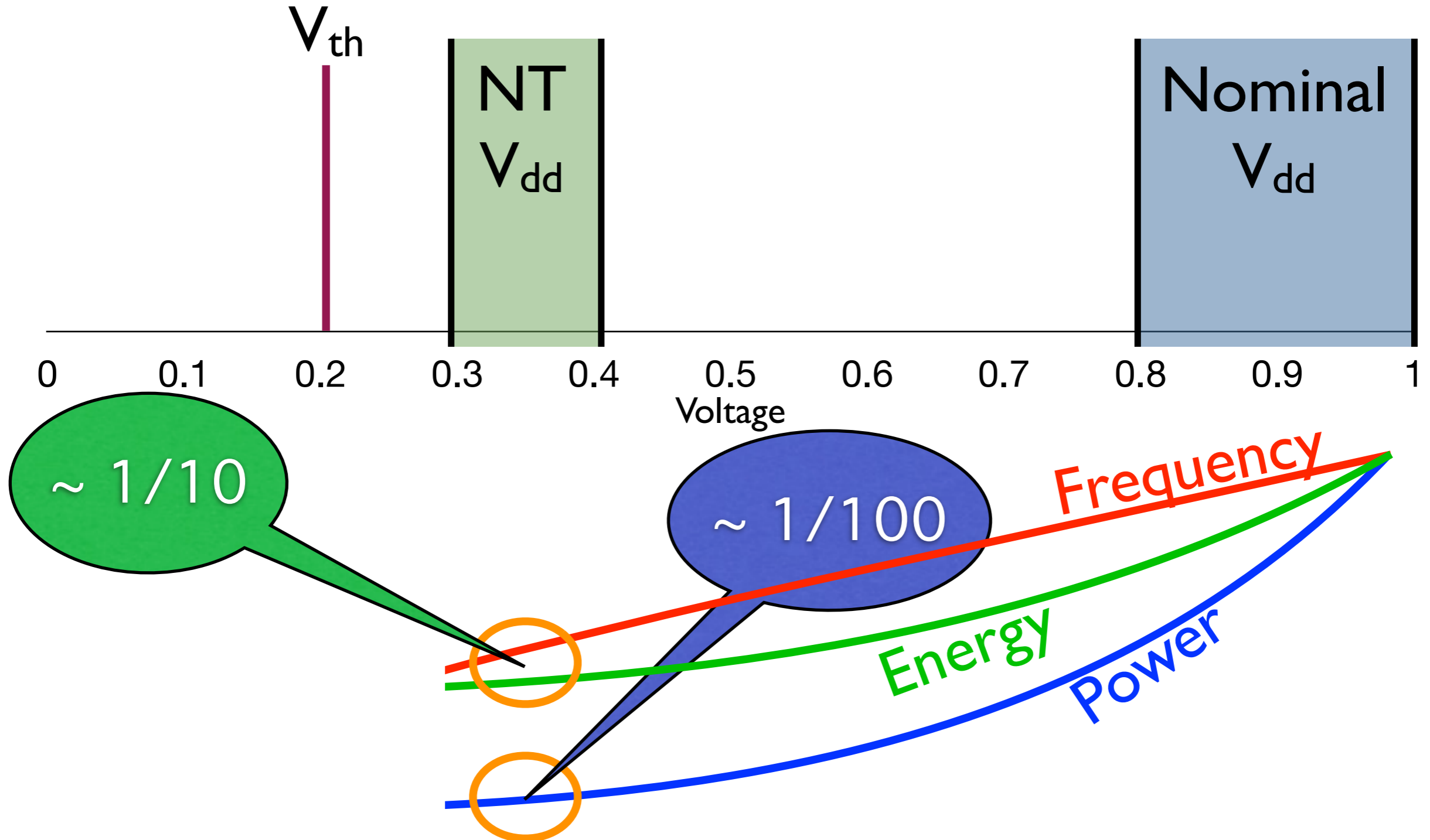
- Increasing demand for low-power devices
 - Smart phones, laptops - battery life
 - Desktops - electricity costs
 - Servers - power & cooling constraints



The New Power Wall

- Power density and cooling challenges
- Today: Core i7
 - 32nm, 750M transistors, 3GHz, 4 cores
 - Power ~ 100W (40W/cm²)
- Soon: Core YN128
 - 11nm, ~16B transistors, ~4GHz, 128 cores
 - Power ~ >> 1000W (400W/cm²)

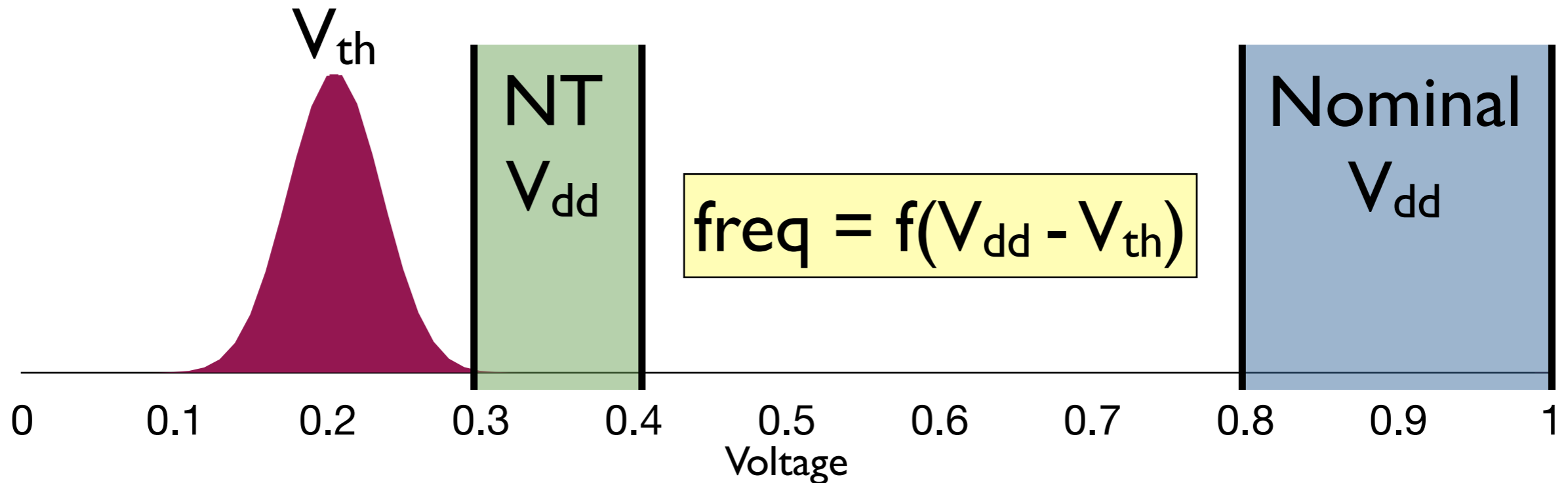
Ultra-low Voltage Operation



Pros and Cons of Ultra-low Voltage

- Power budget: 100 Watts
 - $4 \times 3\text{GHz}$ → 12 Gflops (nominal)
 - $400 \times 300\text{MHz}$ → 120 Gflops (NTC)
- 10x frequency reduction
- Much higher sensitivity to process variation

Effects of Process Variation



- High degree of variation
- Amplified effects at near-threshold

Variation effect on max delay at nominal

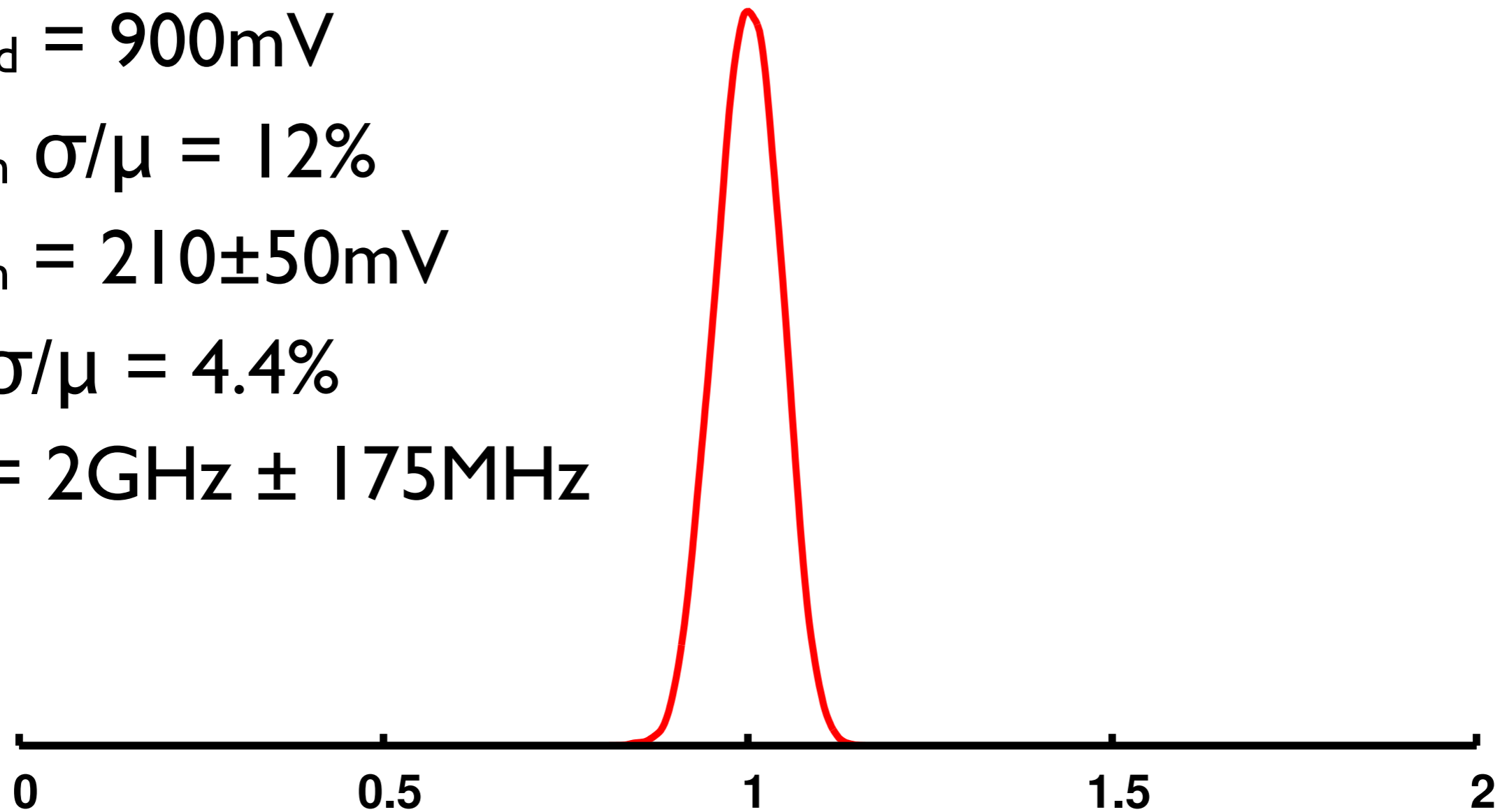
$$V_{dd} = 900\text{mV}$$

$$V_{th} \sigma/\mu = 12\%$$

$$V_{th} = 210 \pm 50\text{mV}$$

$$F \sigma/\mu = 4.4\%$$

$$F = 2\text{GHz} \pm 175\text{MHz}$$



Frequency Distribution

Variation effect on max delay at NT

$V_{dd} = 900\text{mV}$

$V_{th} \sigma/\mu = 12\%$

$F \sigma/\mu = 4.4\%$

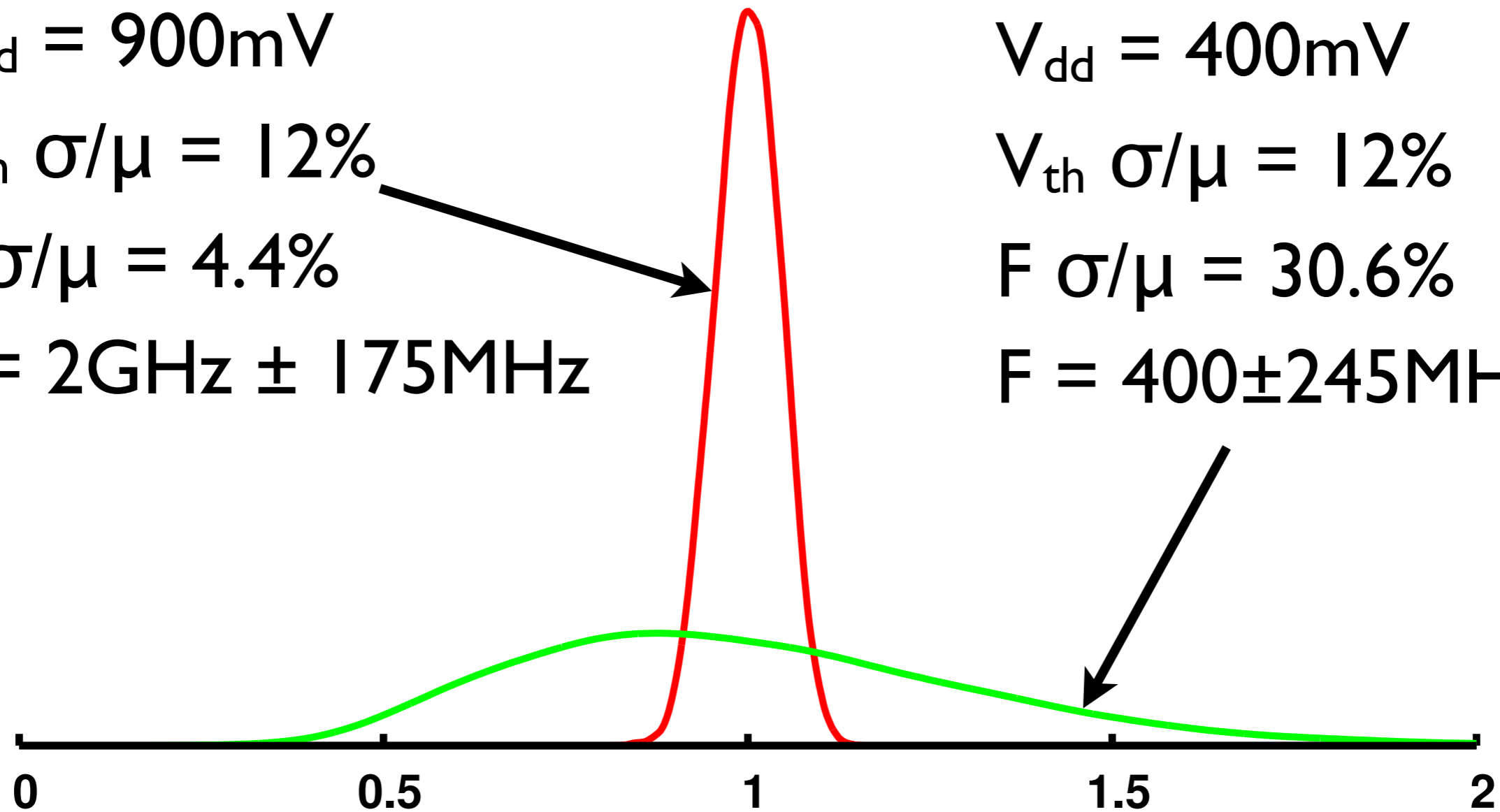
$F = 2\text{GHz} \pm 175\text{MHz}$

$V_{dd} = 400\text{mV}$

$V_{th} \sigma/\mu = 12\%$

$F \sigma/\mu = 30.6\%$

$F = 400 \pm 245\text{MHz}$



Frequency Distribution

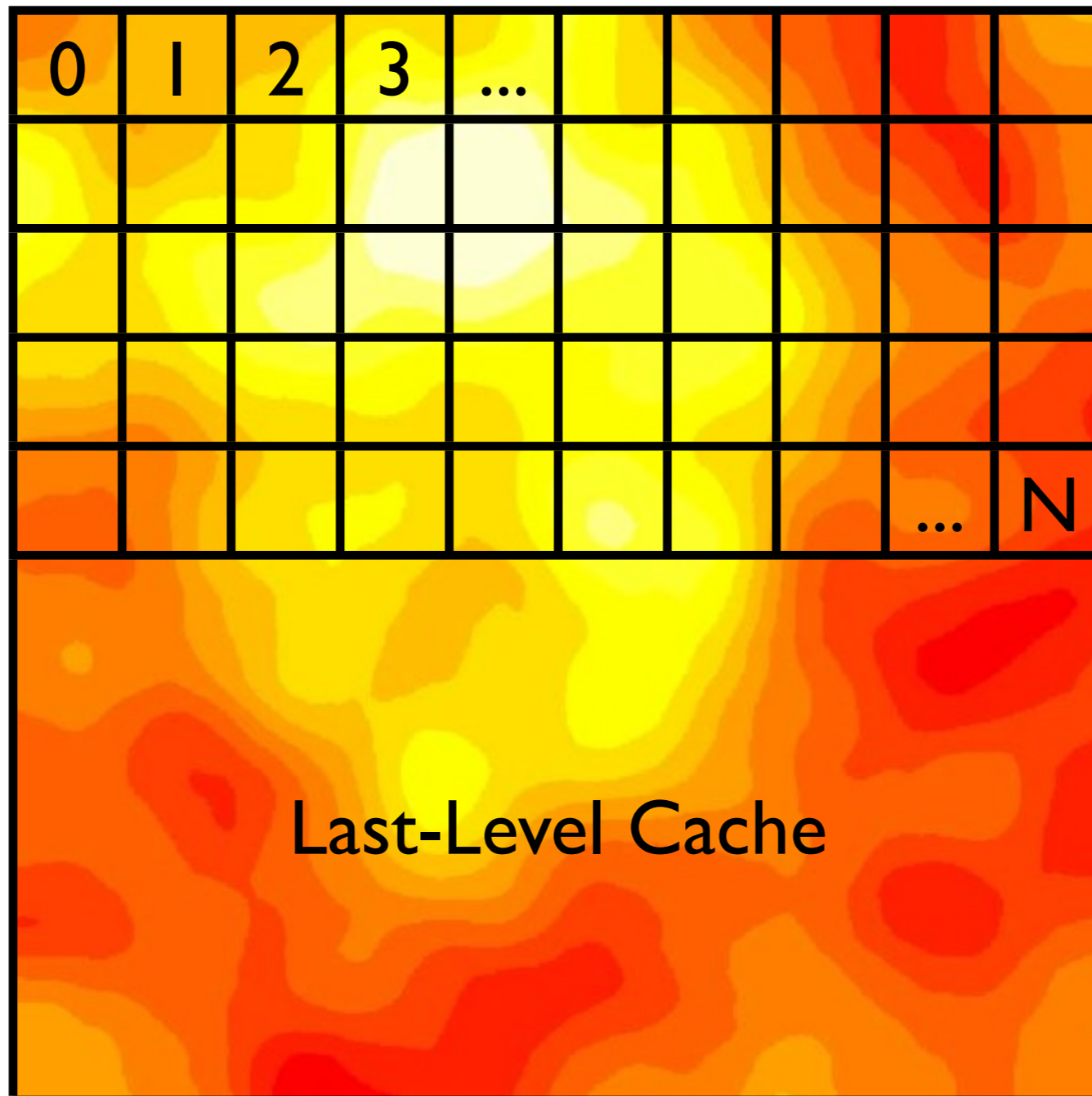
Consequences of Variation

- High frequency variation from core to core
 - (1) All cores at max $F \rightarrow$ Heterogeneous
 - (2) All cores at same $F \rightarrow$ Inefficient
- Frequency of CMP dictated by slowest core
- Faster cores could run at a lower voltage, to achieve the same frequency, saving power.

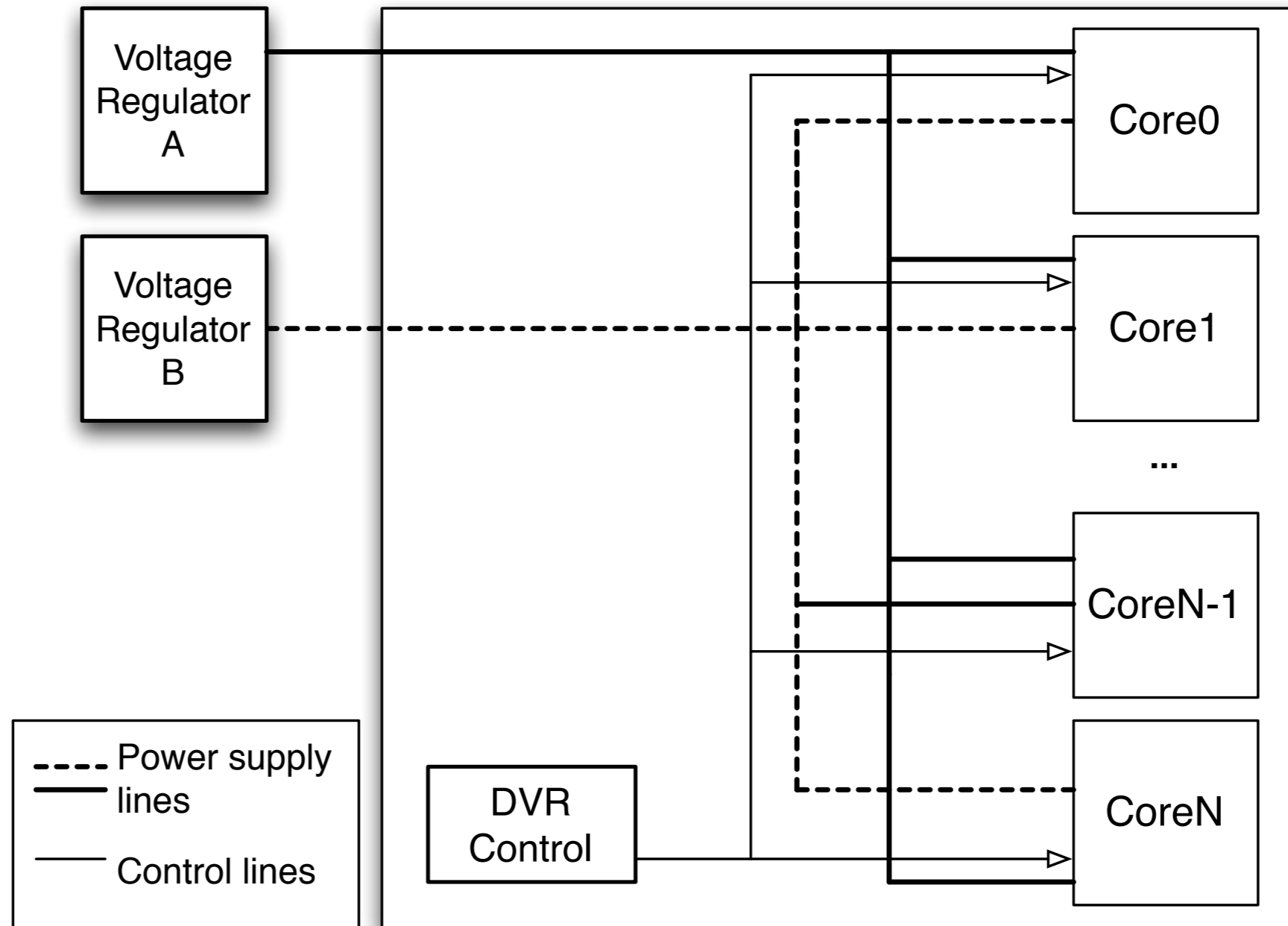
Proposed Solutions

- Dual Voltage Rail (DVR)
- Half Speed Unit (HSU)

Core-to-core variation



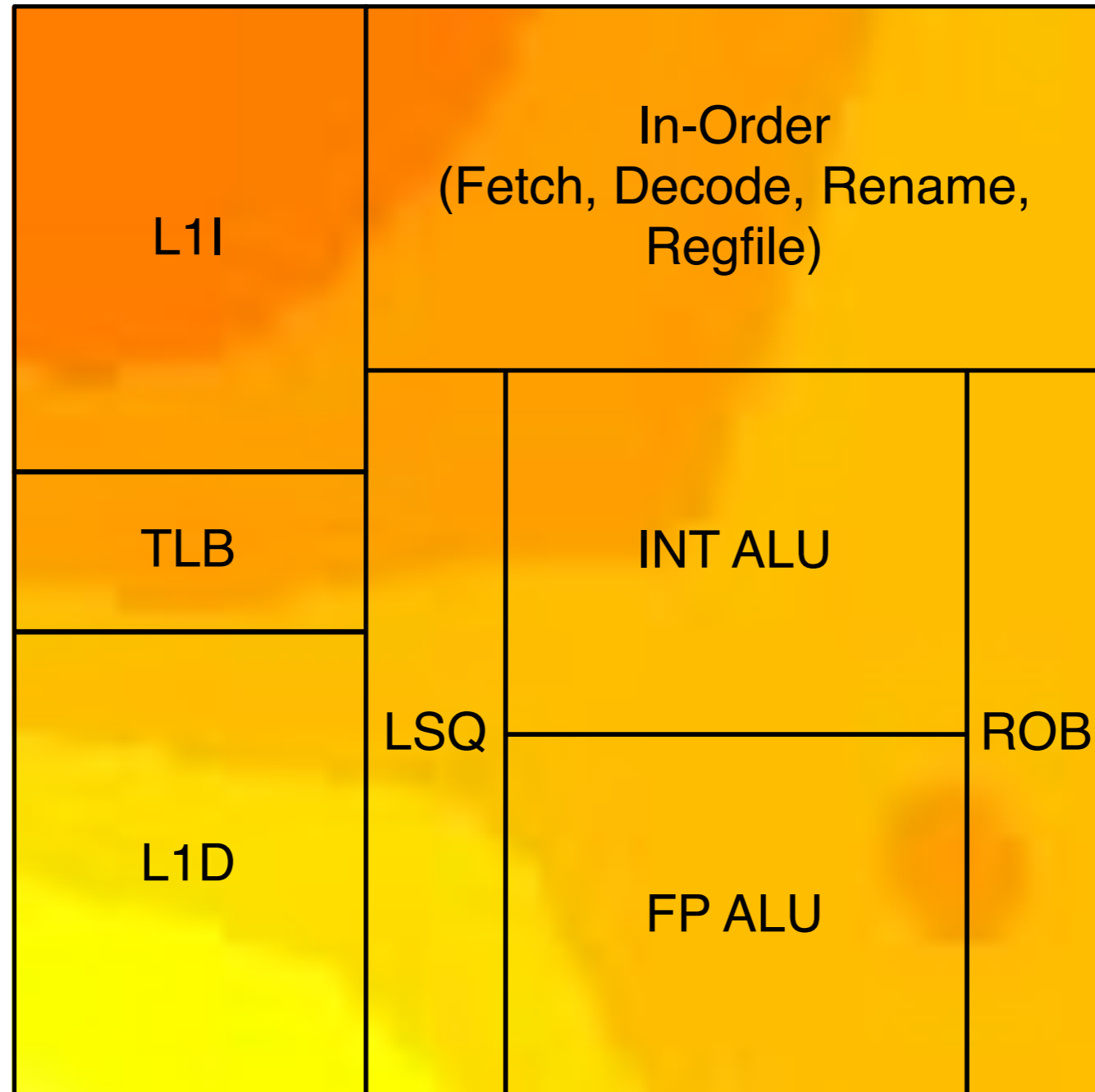
Dual voltage rails (DVR)



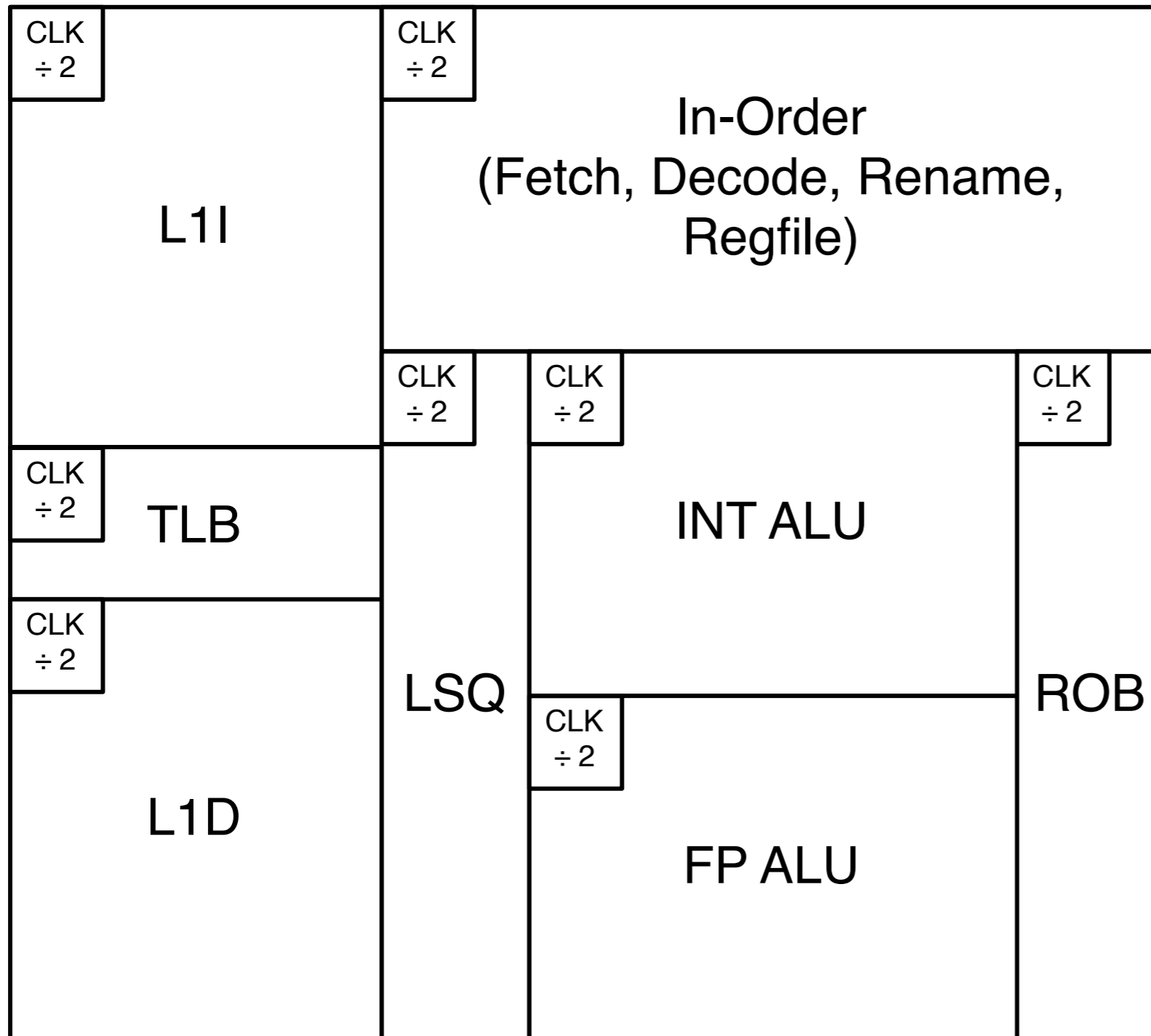
DVR: Main Idea

- Slow cores on high rail, for speed
- Fast cores on low rail, to save power
- Reduced core-to-core variation
- Increased frequency at constant power

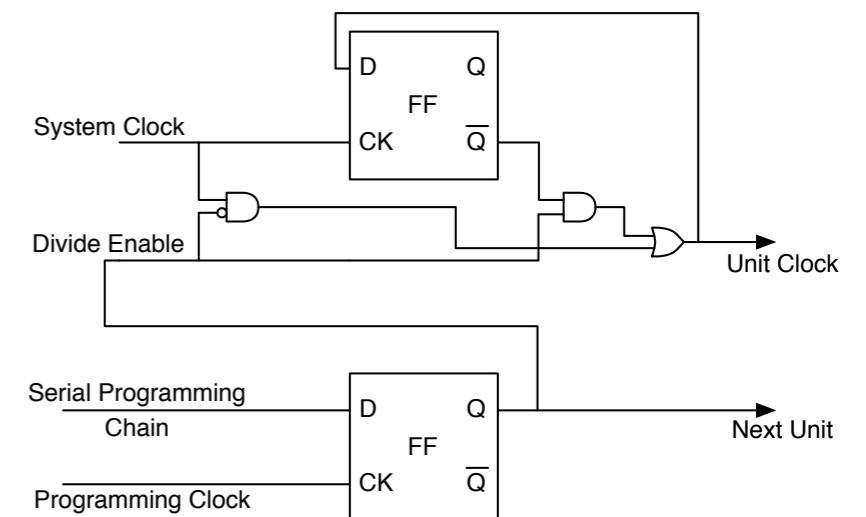
Within-core variation



Half-Speed Unit



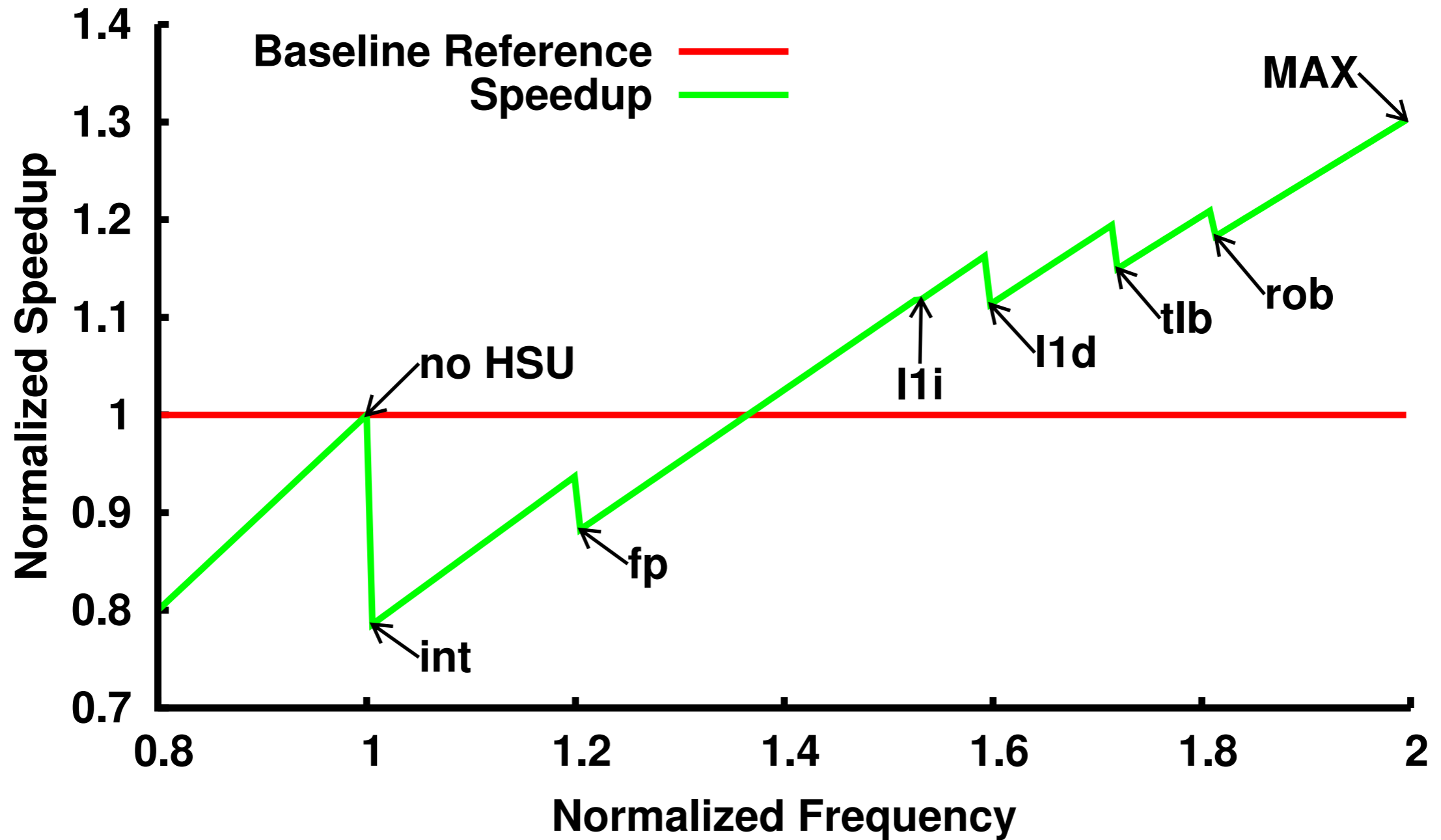
Functional Blocks can run at system frequency or half system frequency



HSU: Main Idea

- Slowest blocks have divider enabled, removing them from the critical path.
- Allows frequency to be raised for faster blocks.
- System frequency no longer limited by slowest core.

Effect of HSU on Performance



DVR & HSU Calibration

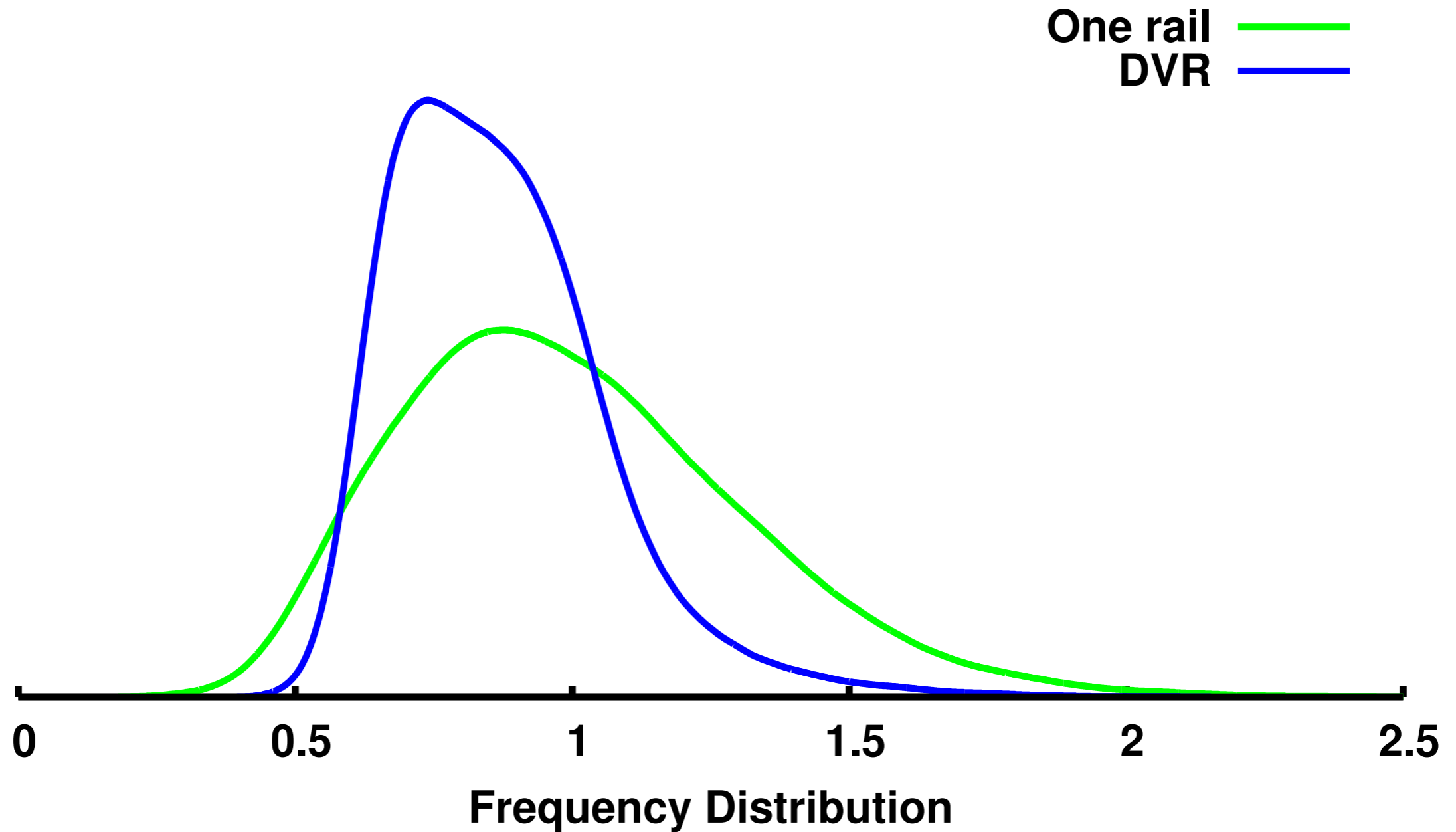
- Optimization goal:
 - Identify voltage levels and core configurations that maximize Frequency for a given Power budget.
- Requirements:
 - Functions to estimate a core's:
 - maximum Frequency from Voltage
 - Power from Frequency and Voltage

Experimental Setup

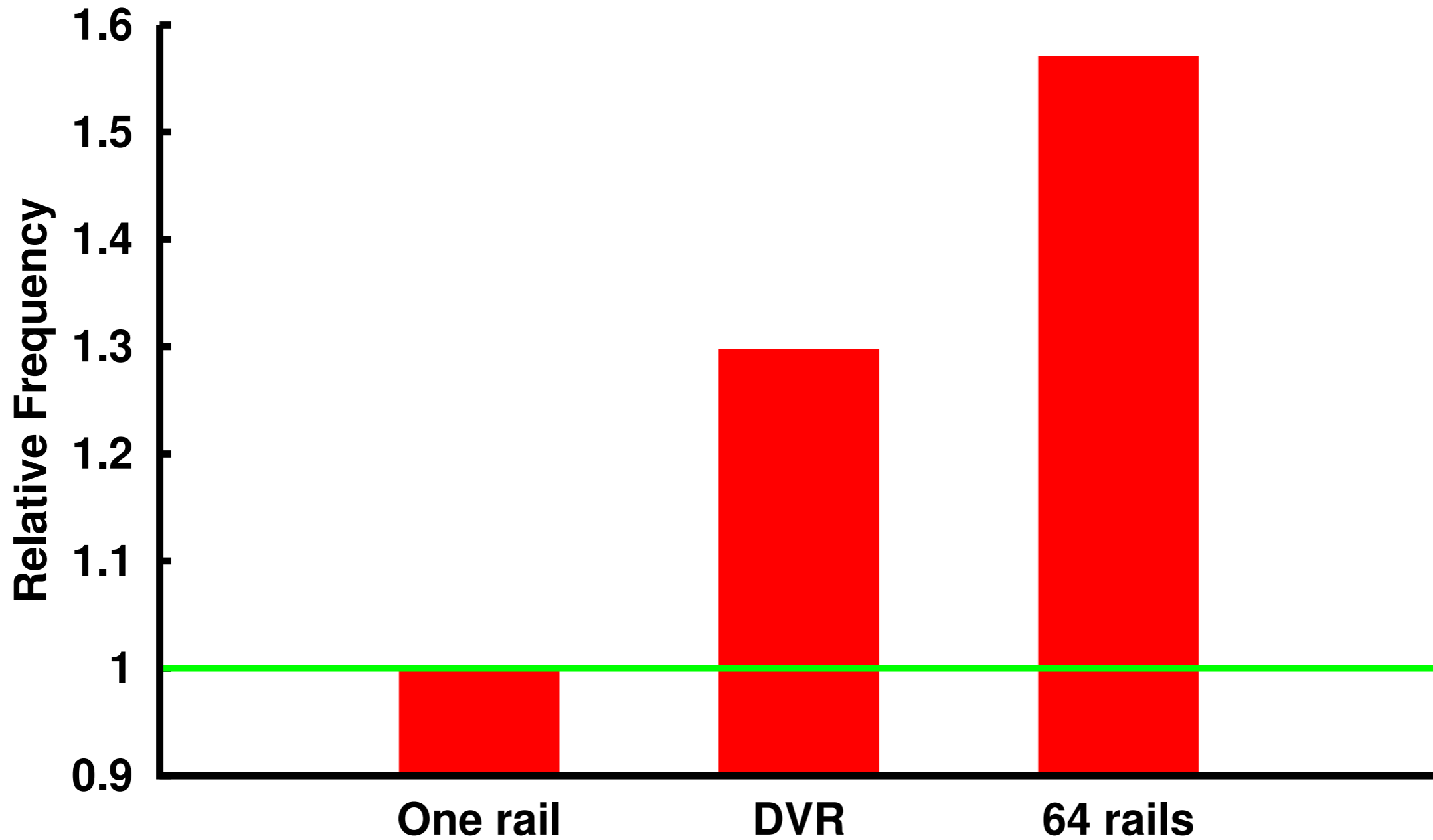
- Processor
 - Modeled by SESC
 - Dual-issue OOO
 - 32nm, 64 cores
 - 2GHz at 900mV
 - 200-800MHz at NT
 - NT at 300-500mV
- Benchmarks
 - SPEC2000
 - 4112 simulations
- Circuit modeling
 - SPICE
 - Marković, et al *
- Variation modeling
 - VARIUS

* D. Markovic, C. Wang, L. Alarcon, T.-T. Liu, and J. Rabaey, "Ultralow-power design in near-threshold region," Proceedings of the IEEE, vol. 98, no. 2, pp. 237–252, Feb. 2010.

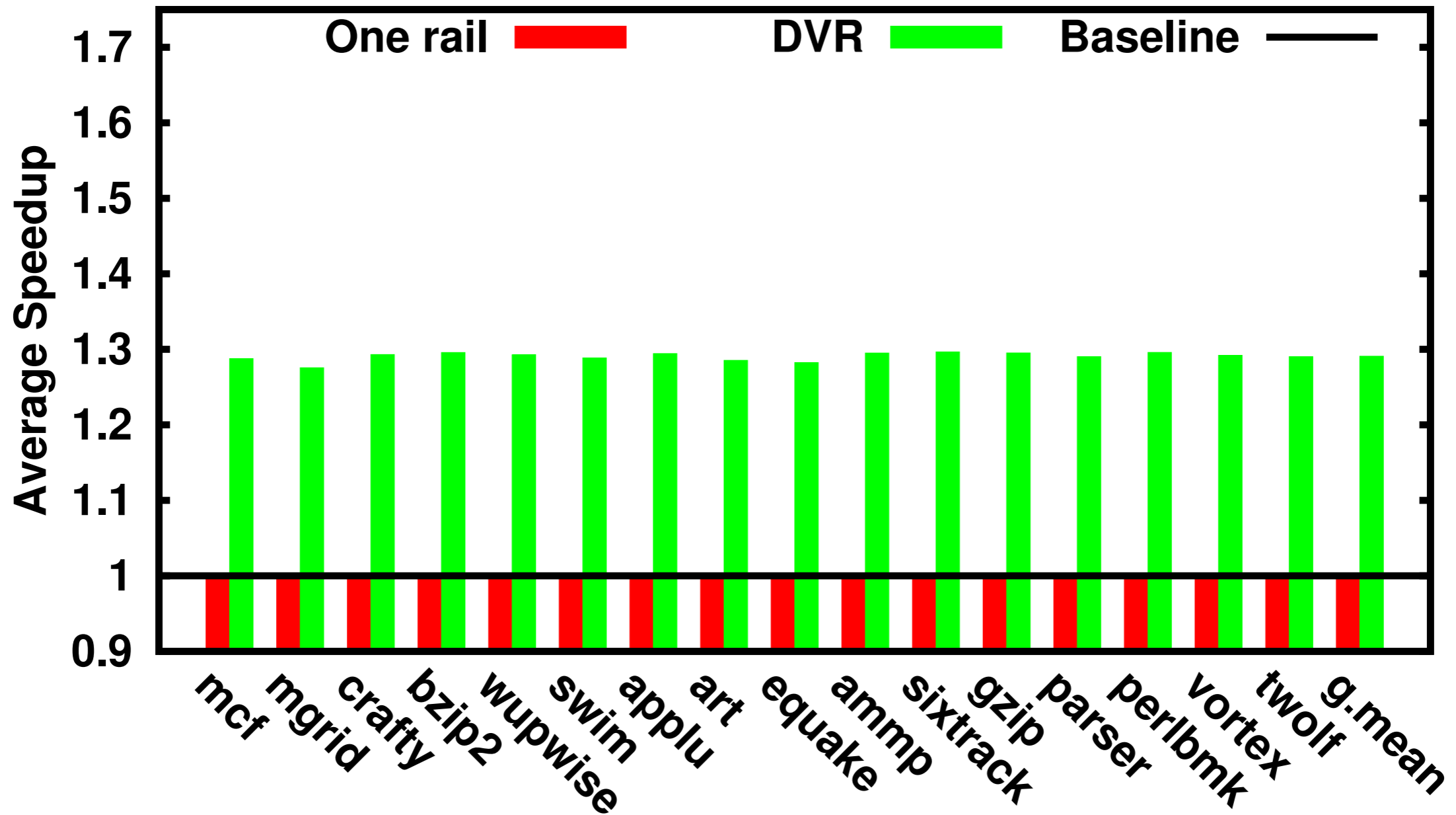
DVR: Reduced Core-to-core Variation



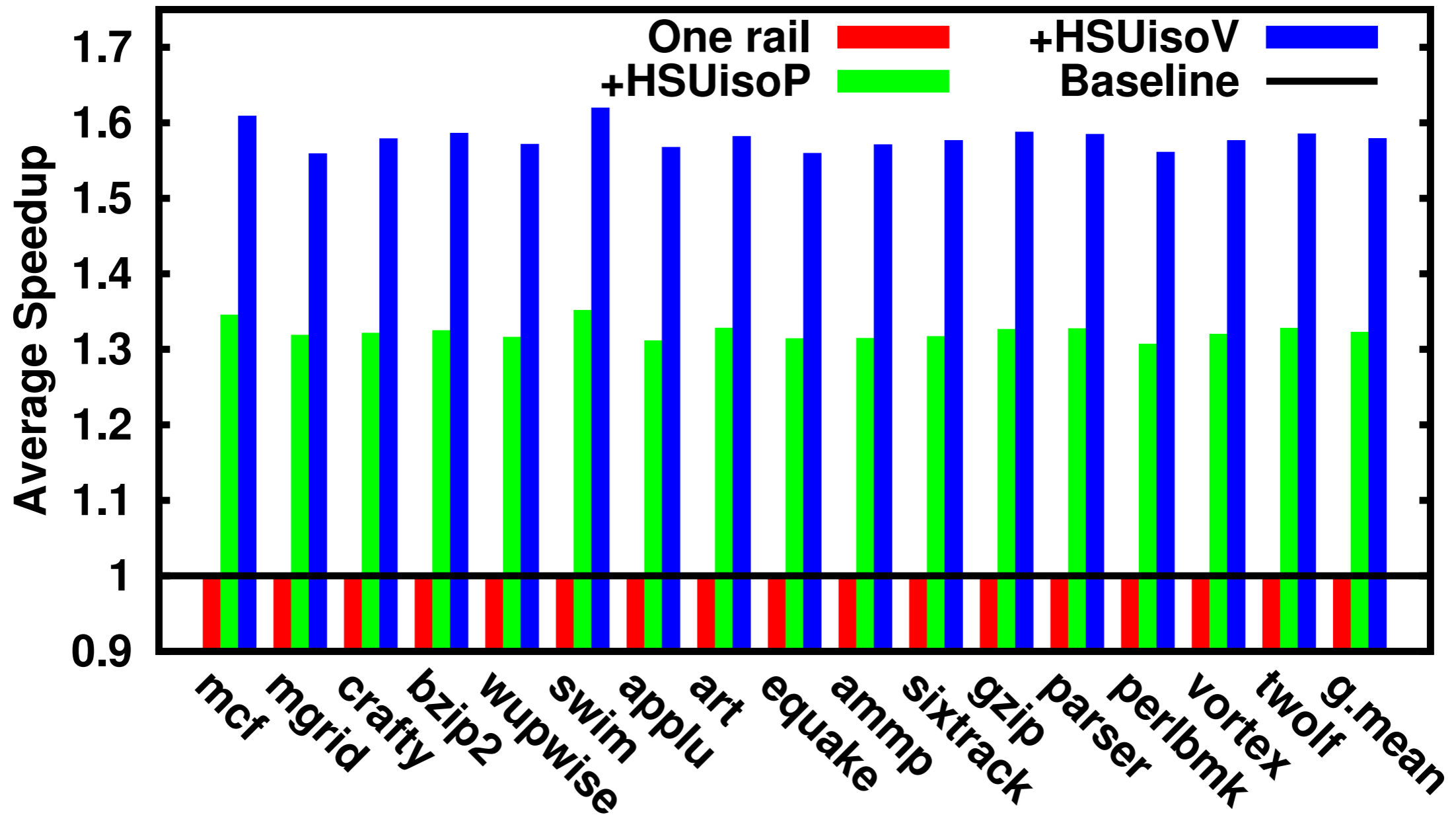
DVR: Performance Improvement



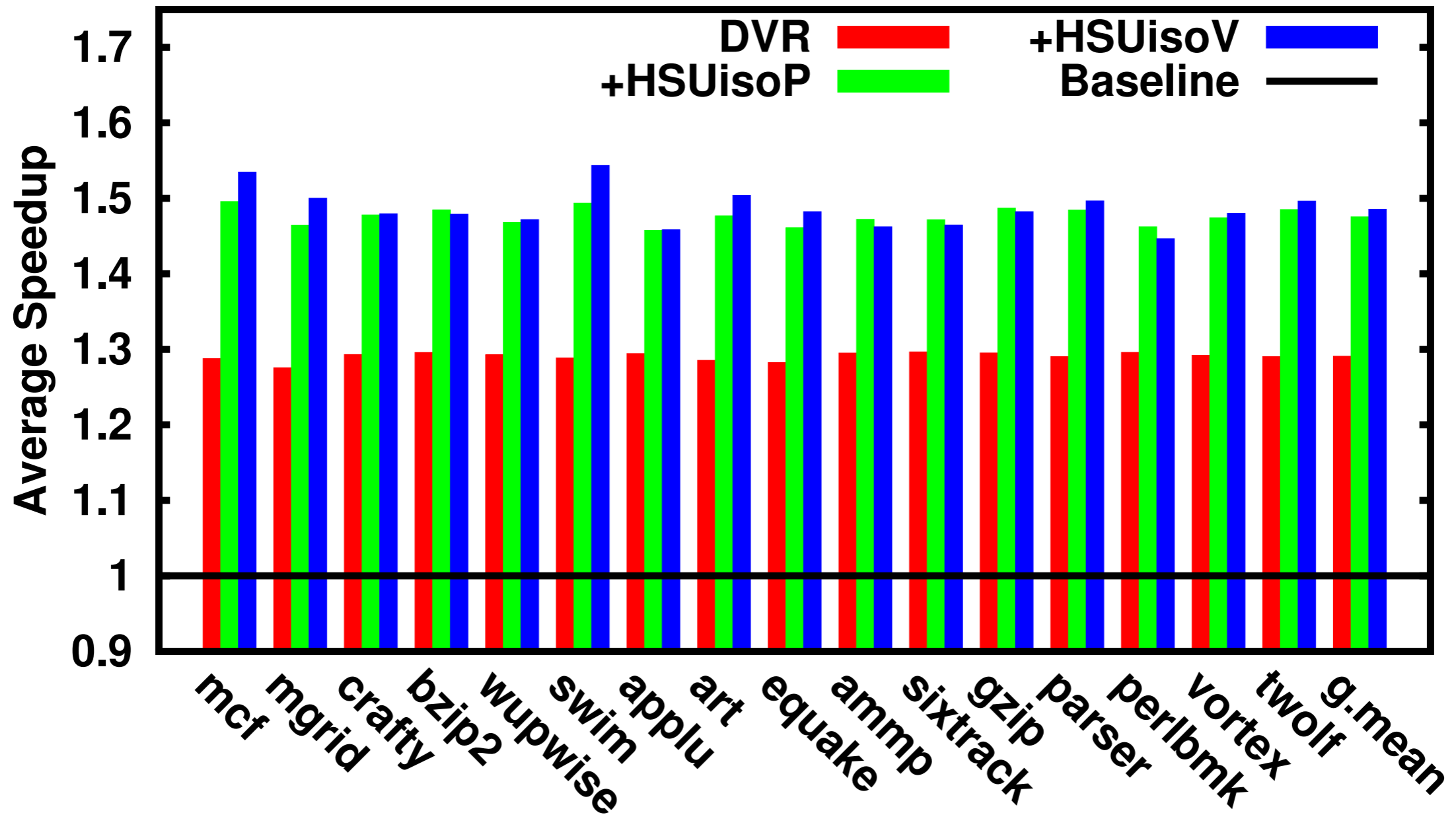
DVR Performance



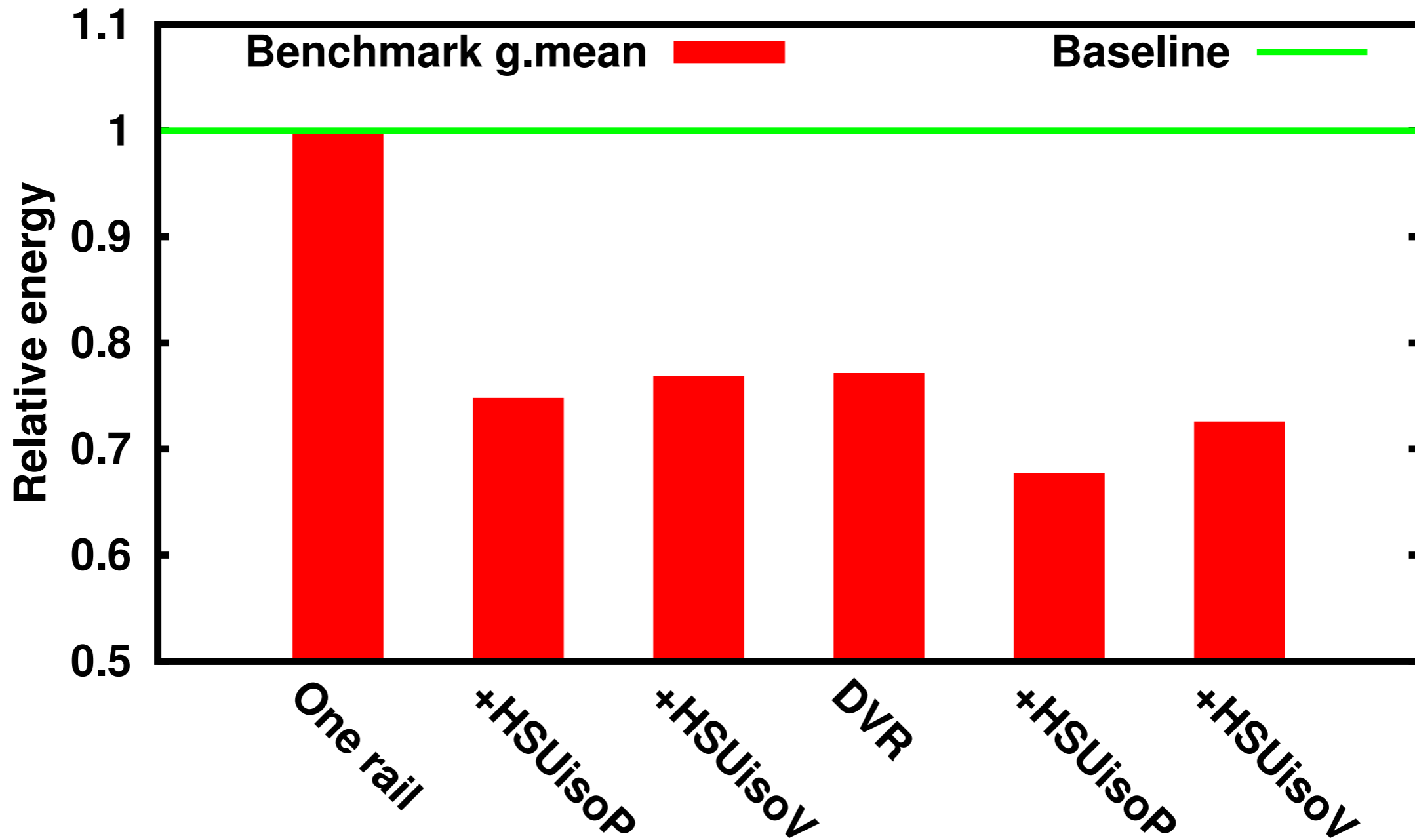
HSU Performance



DVR + HSU Performance



Energy



Related Work

- Global Asynchronous Locally Synchronous (GALS)
 - Complex to implement
- Other dual-voltage work
 - Finer granularity, higher overhead
 - For variation or design-time optimization

Conclusions

- Substantial effects of variation at NTC.
- Simple techniques for reducing those effects.
- Able to improve performance by 50%
energy efficiency by 32%.

Questions?

HSU Speedup PDF

