EVAL

Utilizing Processors with Variation-Induced Timing Errors

Smruti R. Sarangi

Brian Greskamp

Abhishek Tiwari

Josep Torrellas

University of Illinois

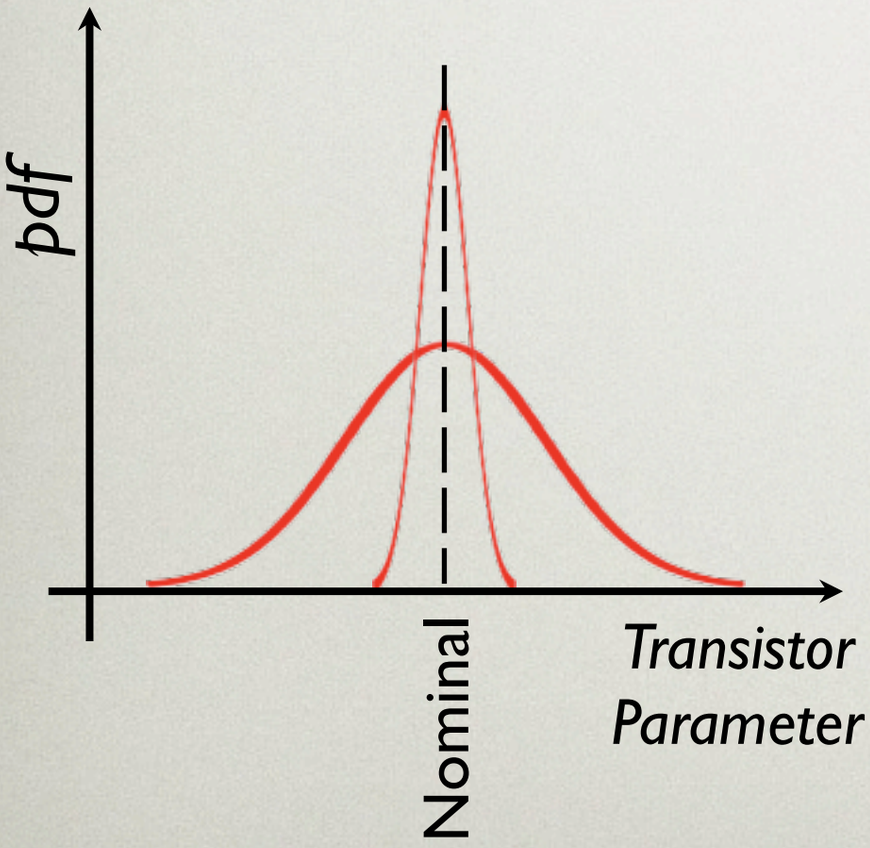
http://iacoma.cs.uiuc.edu

University of Illinois

i-acoma group

Problem: Process Variation

Transistor parameters (V_{th} and L_{eff}) vary across the processor die



Chip Frequency

Chip Leakage Power

Problem: Designing for the Worst Case

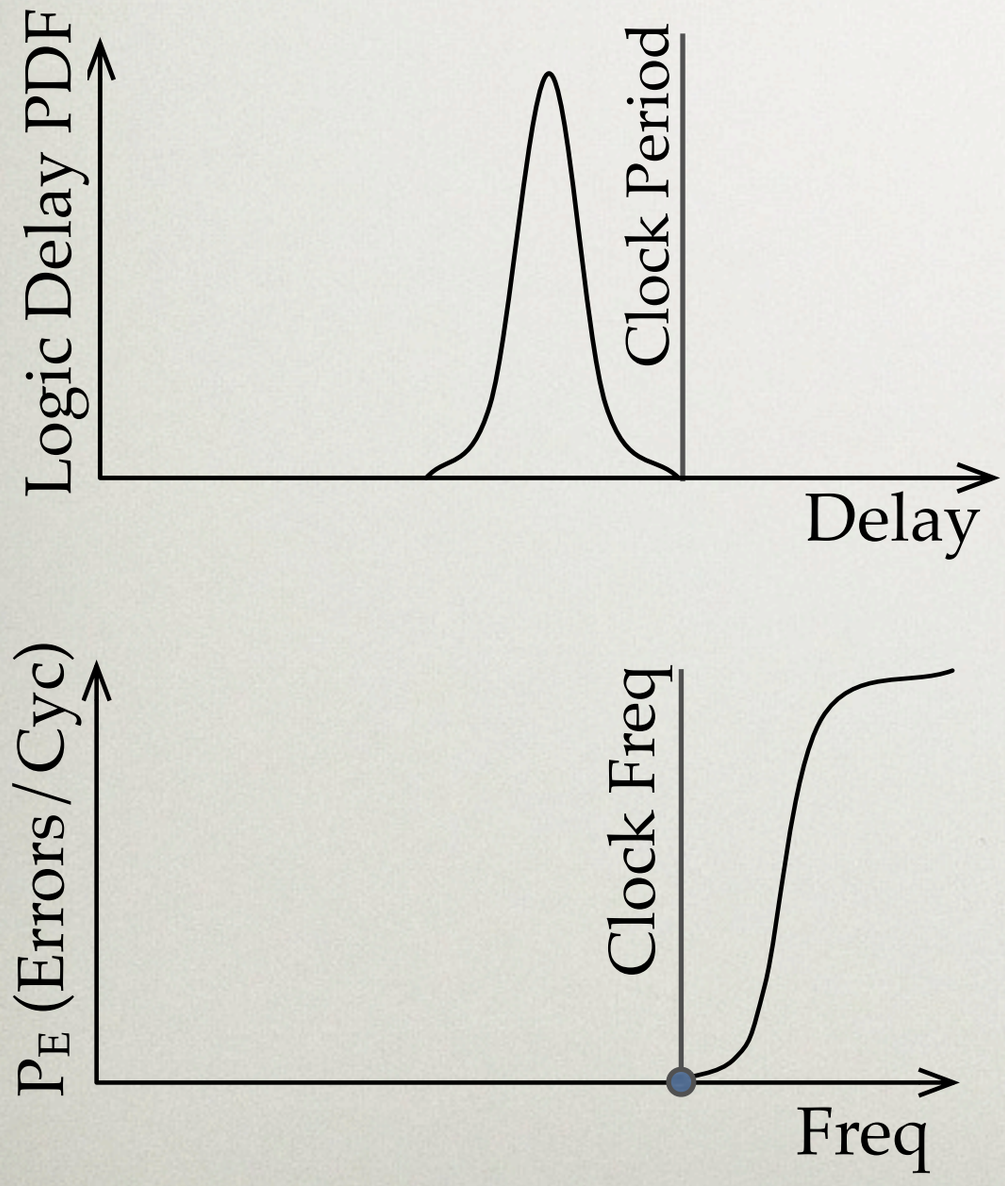
- Design for worst case gives poor performance
- Design for closer to nominal gives timing errors
  - Can we design the processor to tolerate errors and trade errors off for power and performance?
  - What is the area cost of an error-tolerant design?

Contributions

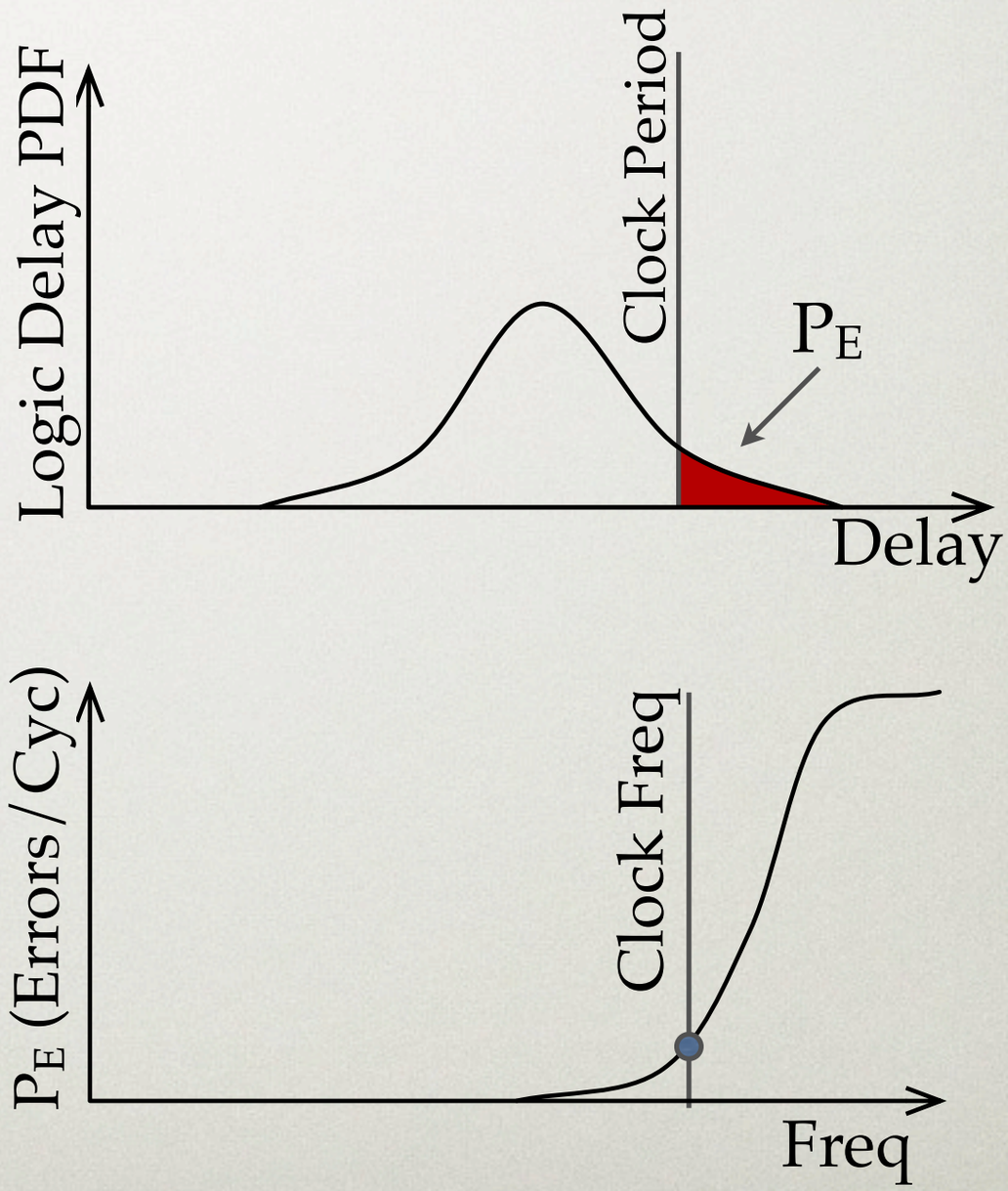
1. EVAL: Framework of how processors tolerate and mitigate variation-induced errors
   - Trade off error rate for power & frequency
2. High-dimensional dynamic adaptation to optimize power and performance in the presence of errors
   - Implementation based on machine learning
3. Example implementation improves performance by 40% under power & temp constraints
   - 11% core area cost

Variation-Induced Timing Errors

Before Variation



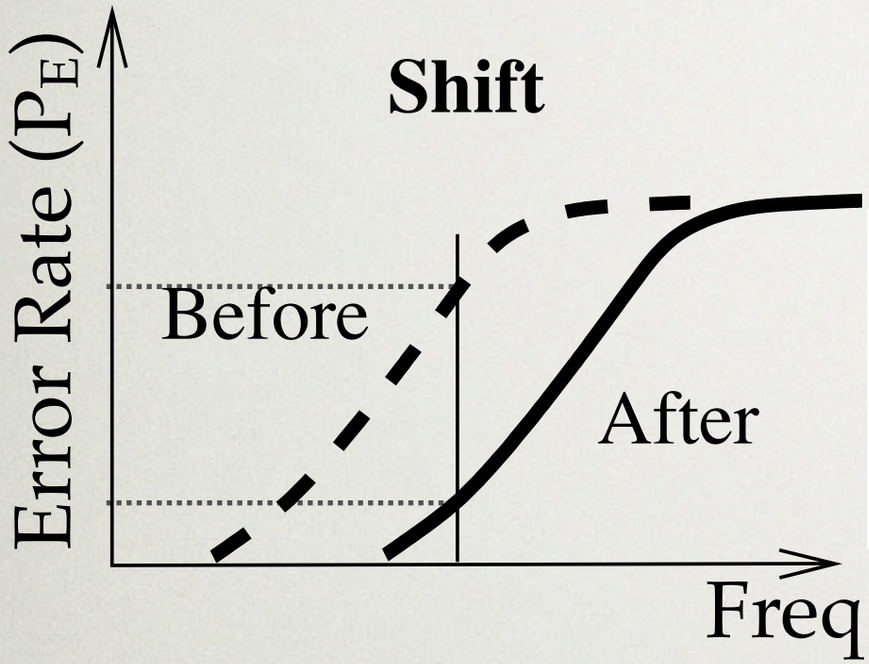
After Variation



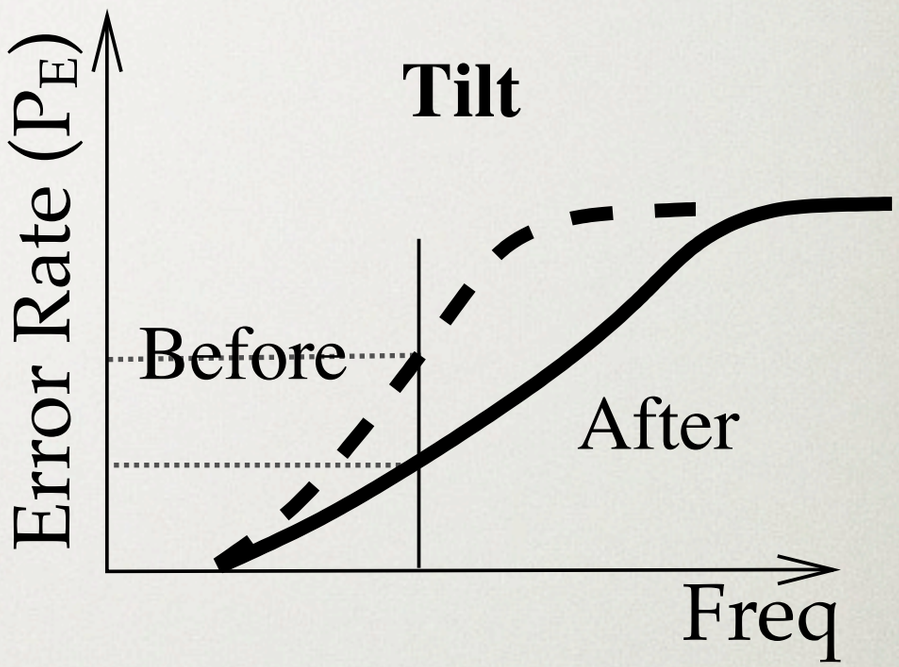
Eval Framework

- Run processor at high freq with nonzero timing error rate (P_E)
- Silently detect and correct errors
  - Examples: Razor, DIVA, Paceline, *etc.*
  - P_E determines performance, not correctness
- Modify P_E *vs* f curves so that variation-induced errors are less frequent
  - Transformations: Shift, Tilt, Reshape, Adapt

Shift and Tilt



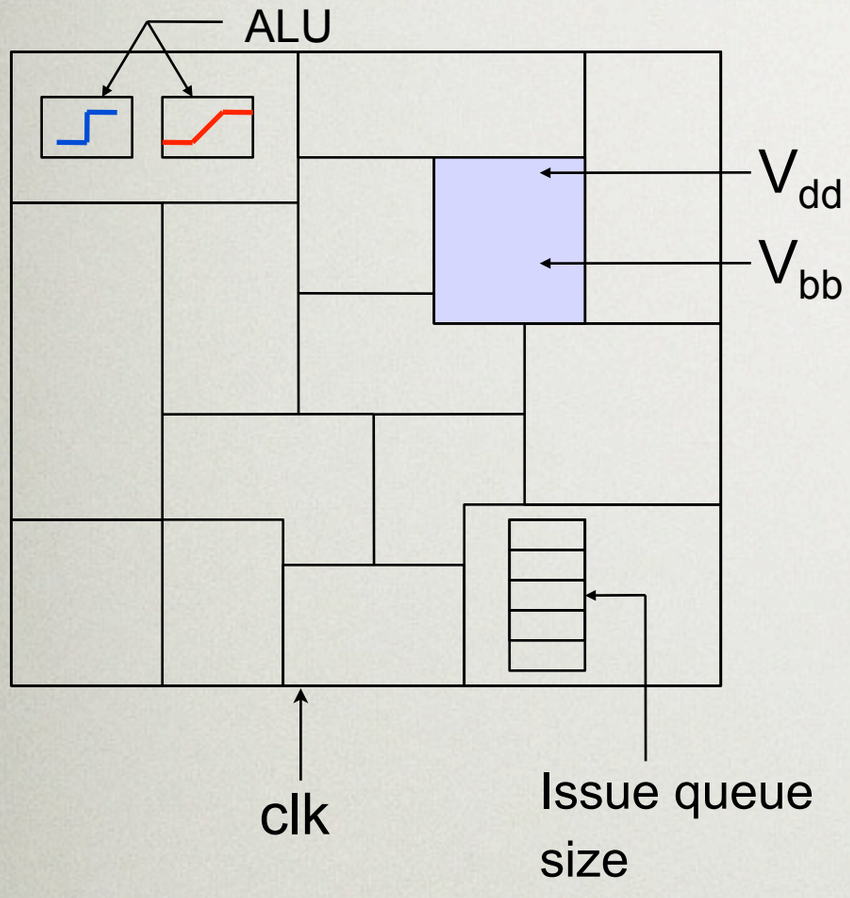
Power↑
P_E↓



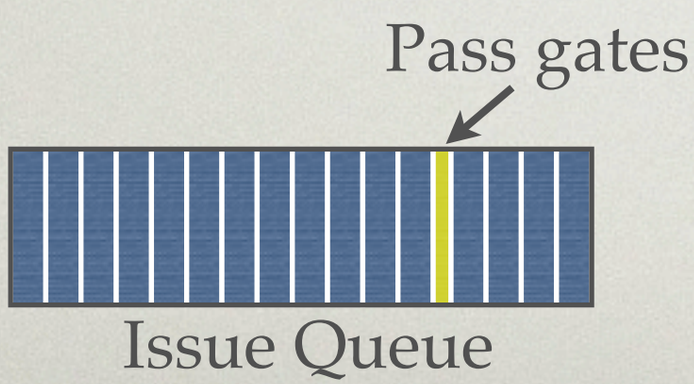
- Speed up all paths
- Methods:
  - ASV, Body Bias (ABB)
  - Structure Resizing

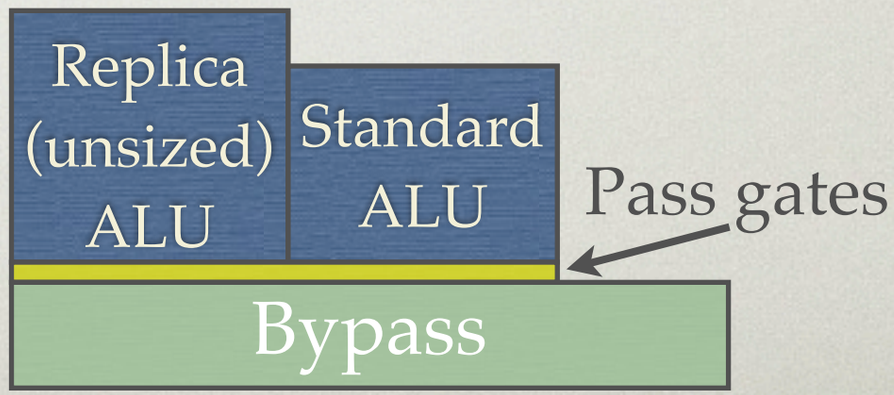
- Speed up some paths
- Methods
  - Design module without gate sizing

Shift and Tilt Implementation

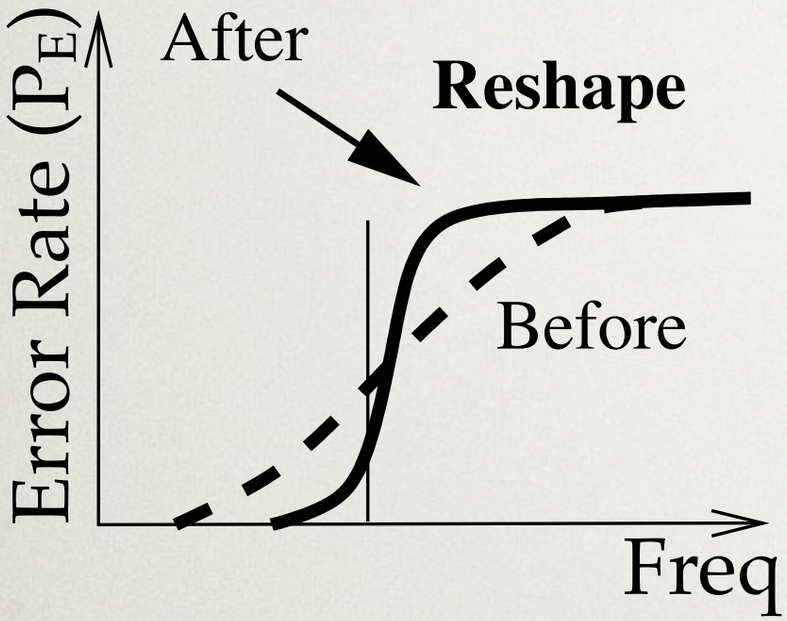


- Shift
  - Per-subsystem ASV, ABB
  - Issue Queue Resizing: Pass gates select full-size or 3/4 size
- Tilt: Include replica ALU designed without gate sizing

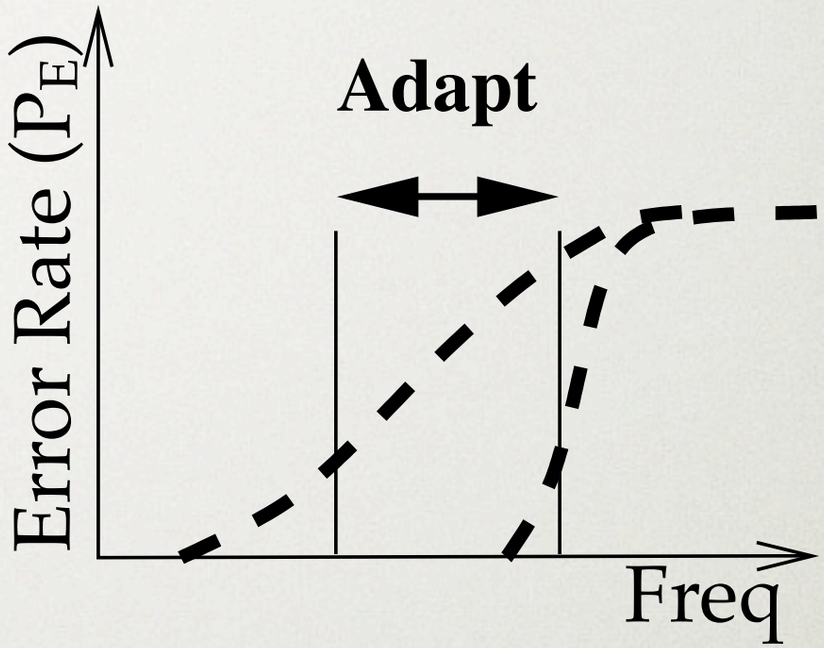




Reshape and Adapt

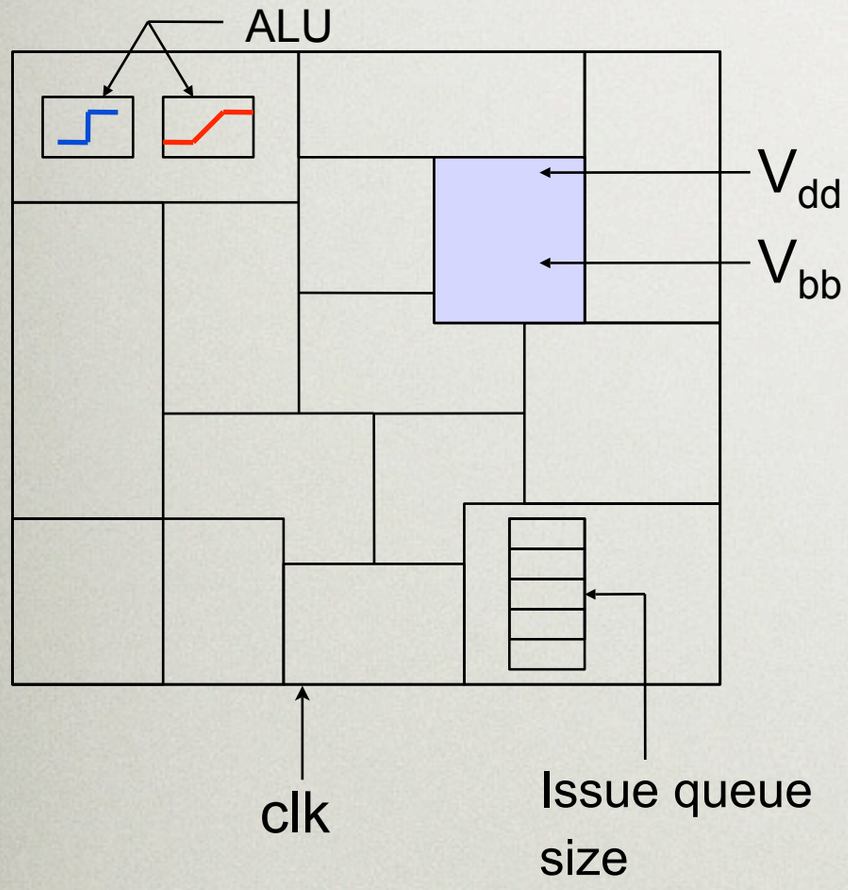


- Slow down fast pipe stages to save power
- Spend power to Shift and Tilt slow stages
- All stages have low P_E



- Apply Shift and Tilt differently for each application phase

High-Dimension Dynamic Adaptation: Controller

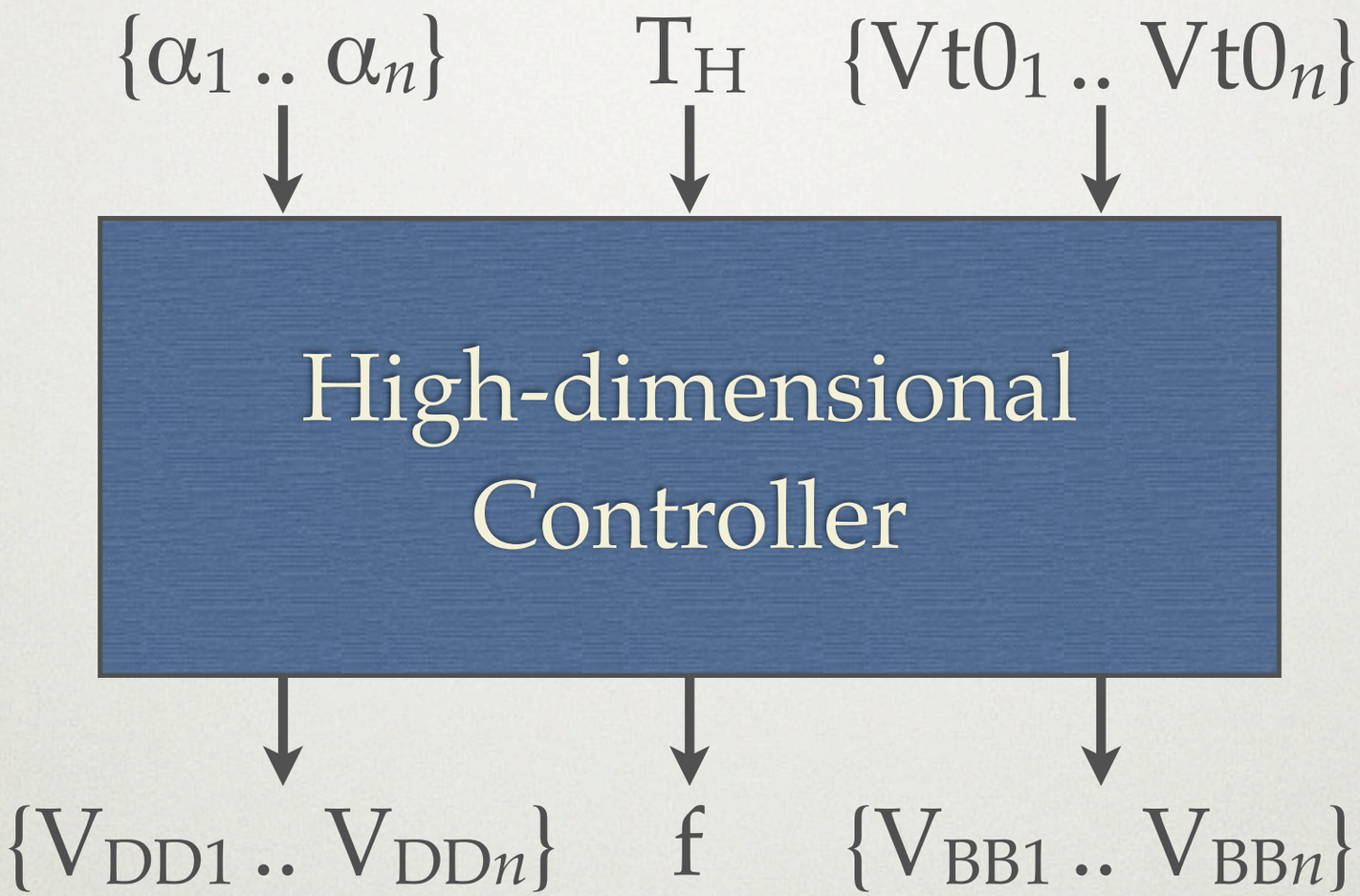


- Goal: Optimize Performance
- Constraints: Max power & temp, per-subsystem $P_E < 1/10K$ cycles
- Outputs: issue queue size, ALU impl, per-subsystem $\{V_{DD}, V_{BB}\}$

Controller Inputs

- Heatsink temperature sensor (T_H)
- Per-subsystem activity counters (α_i): Estimate dynamic power [Isci03]
- Per-subsystem variation parameters ($Vt0_i$) measured by manufacturer

Naive Controller

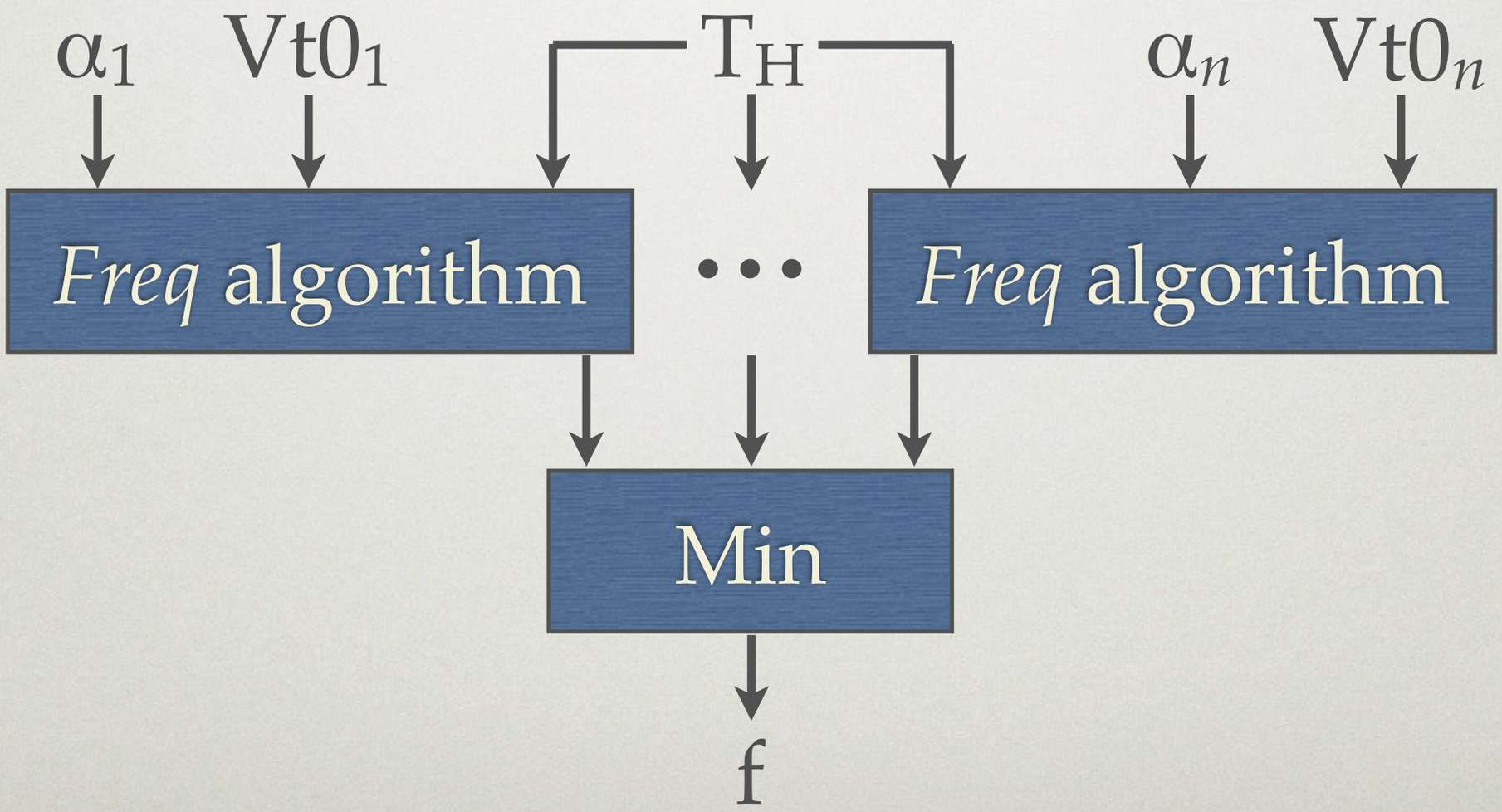


Too many dimensions to optimize simultaneously!

Reduced-Dimension Controller

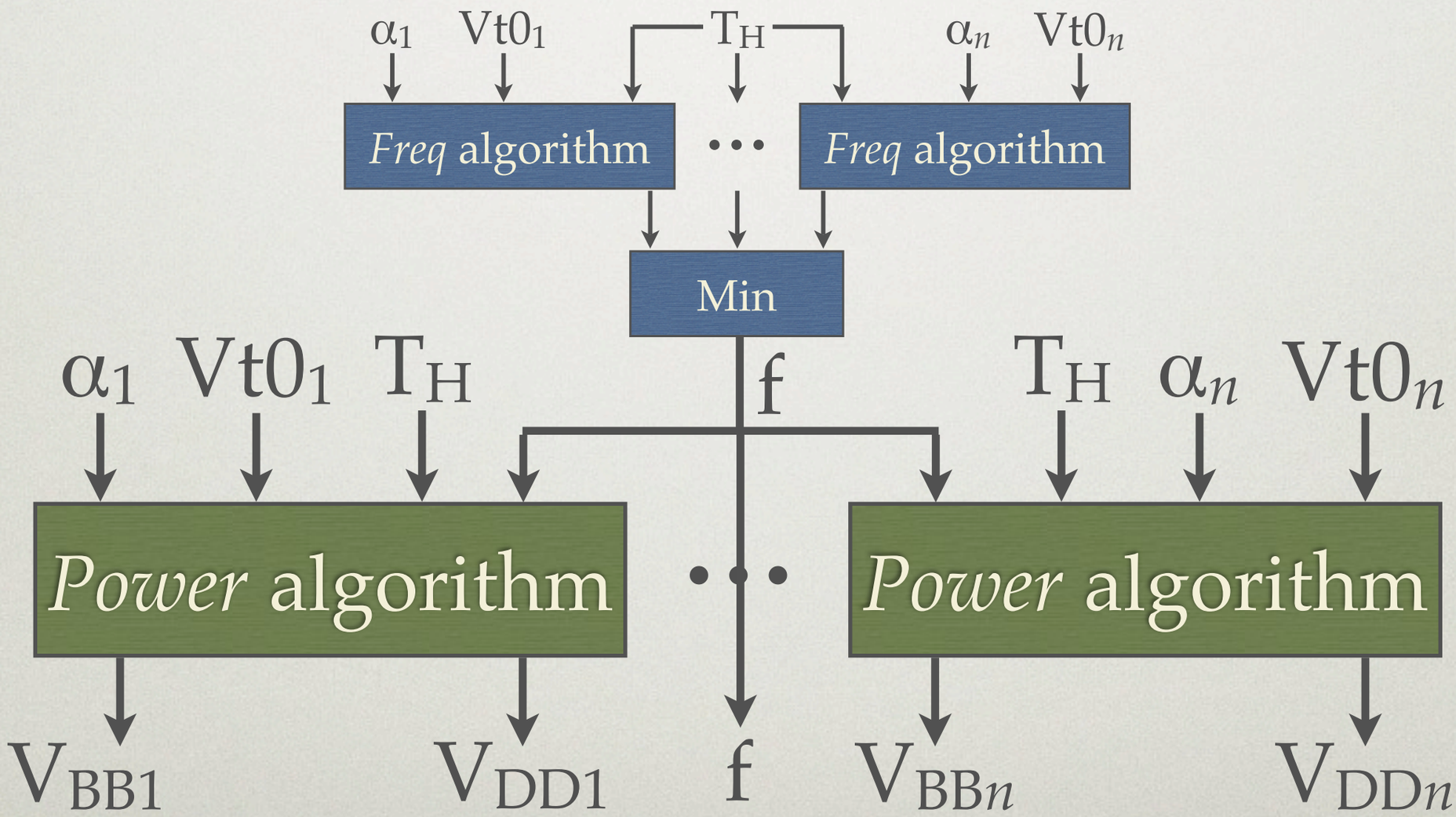
Dimension reduction: Decompose problem

Step 1: Determine what freq the core can support



Reduced-Dimension Controller

Step 2: Optimally configure voltages for each subsystem

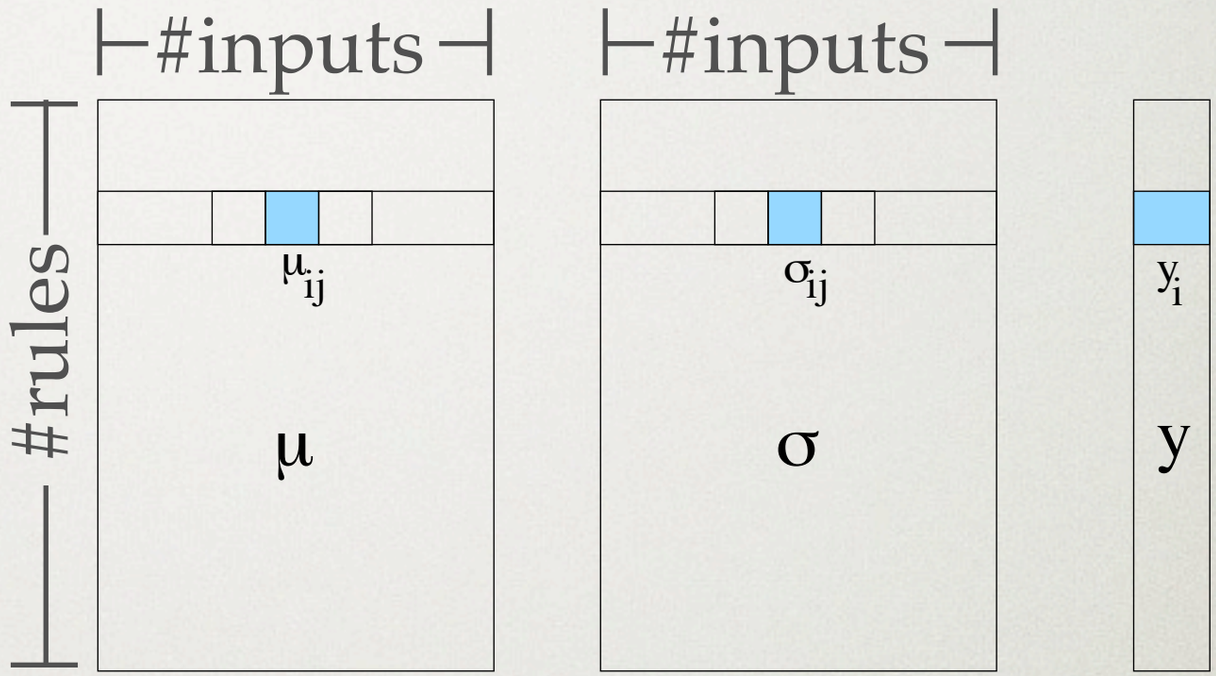


Controller Impl: Fuzzy Logic

Freq algorithm

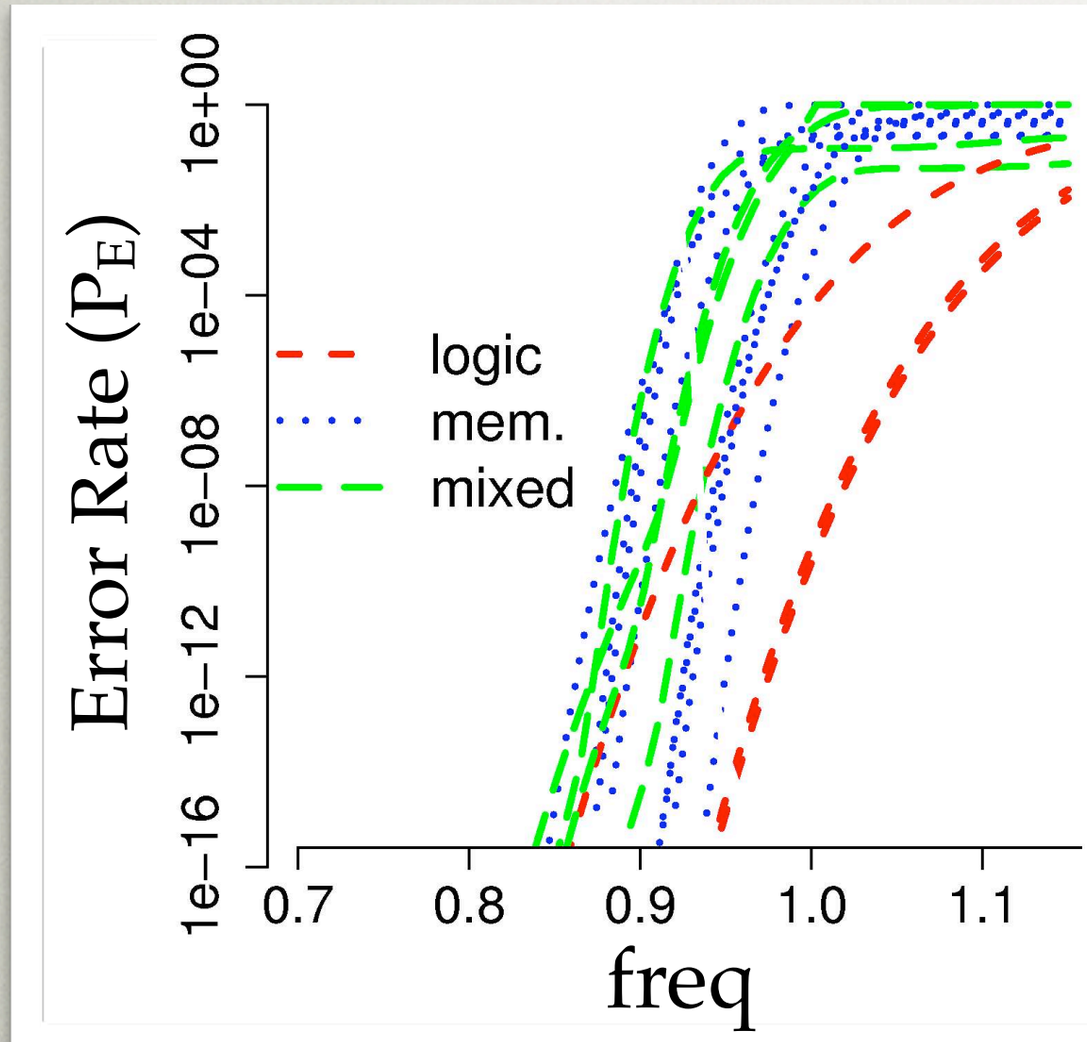
Power algorithm

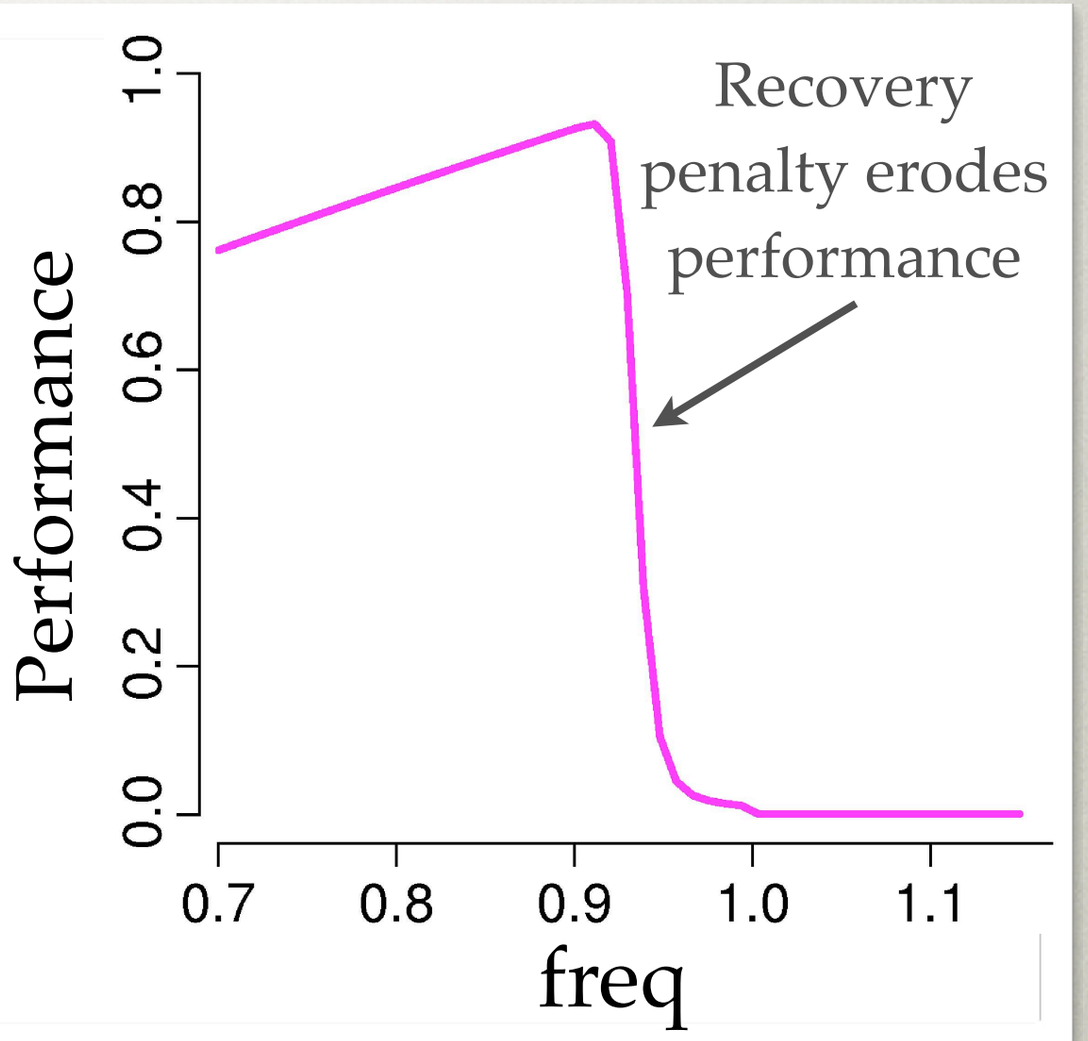
Subcontrollers built as fuzzy systems

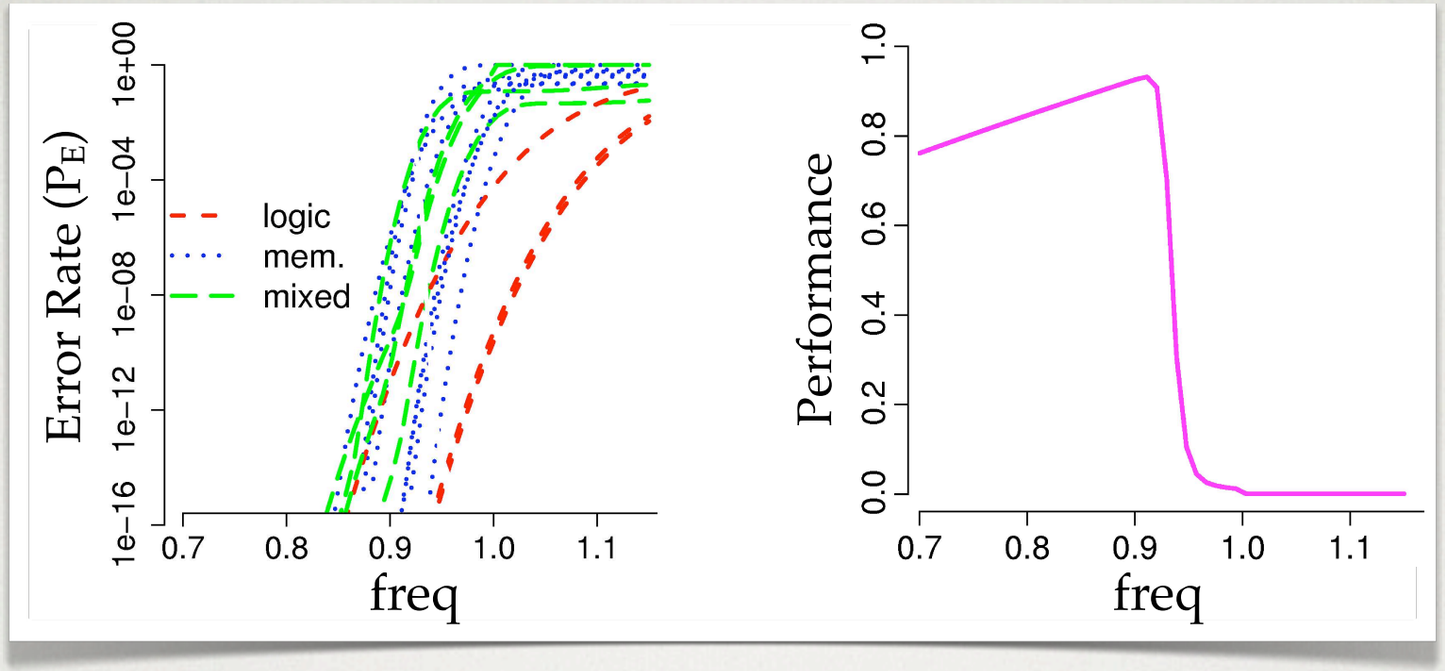


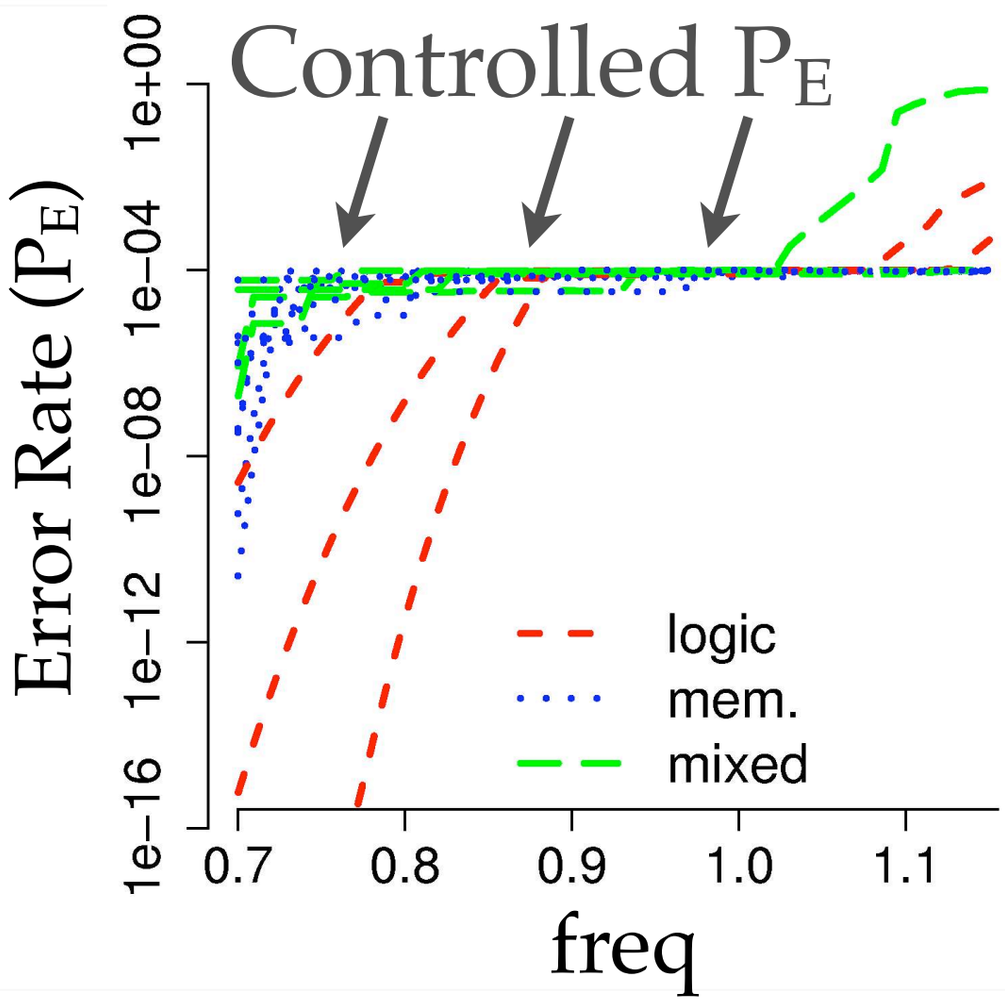
Advantages	Disadvantages
Very fast in software (~6μs total run time)	Must be trained before use
Very accurate in practice	Solutions are approximate

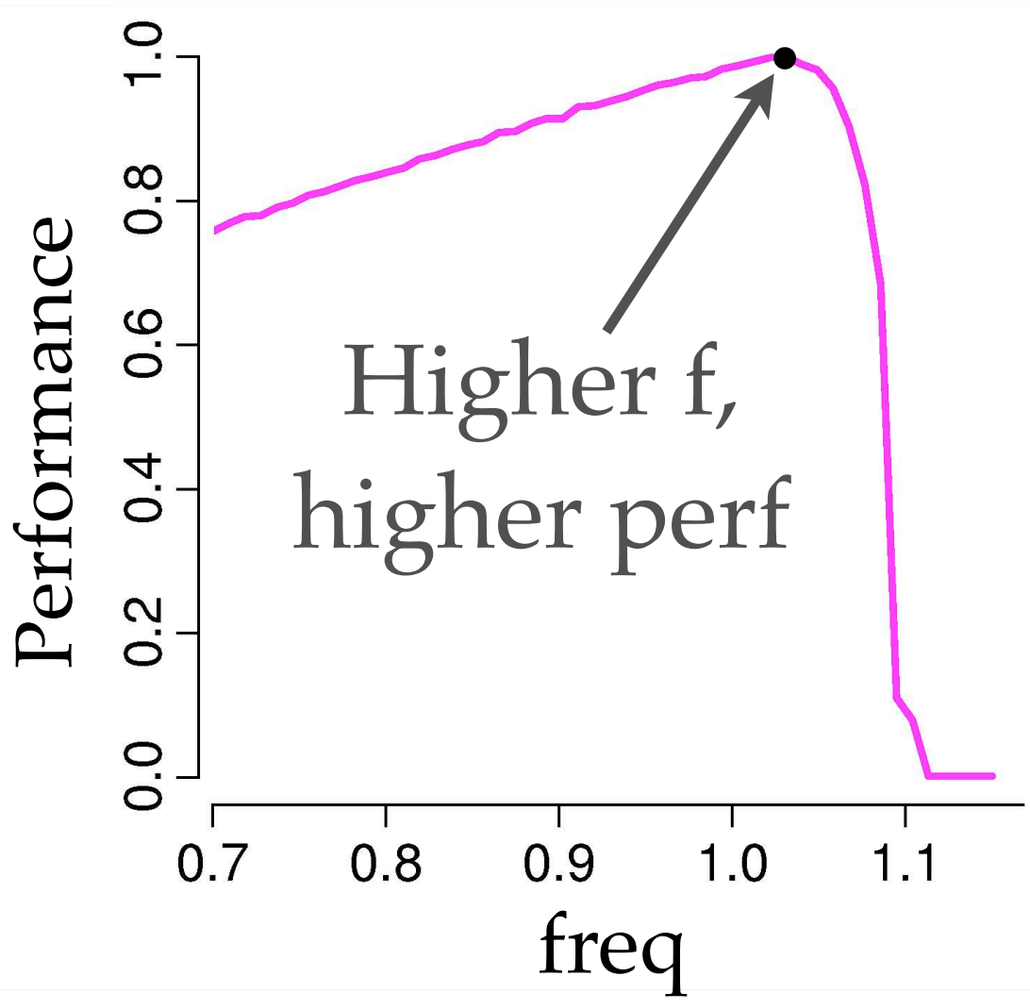
Trading Error Rate for Performance






Error Rate (P_E)
1e+00
1e−04
1e−08
1e−12
1e−16
logic
mem.
mixed
0.7
0.8
0.9
1.0
1.1
freq
Performance
0.0
0.2
0.4
0.6
0.8
1.0
0.7
0.8
0.9
1.0
1.1
freq


Controlled P_E
Error Rate (P_E)
1e+00
1e−04
1e−08
1e−12
1e−16
logic
mem.
mixed
0.7
0.8
0.9
1.0
1.1
freq

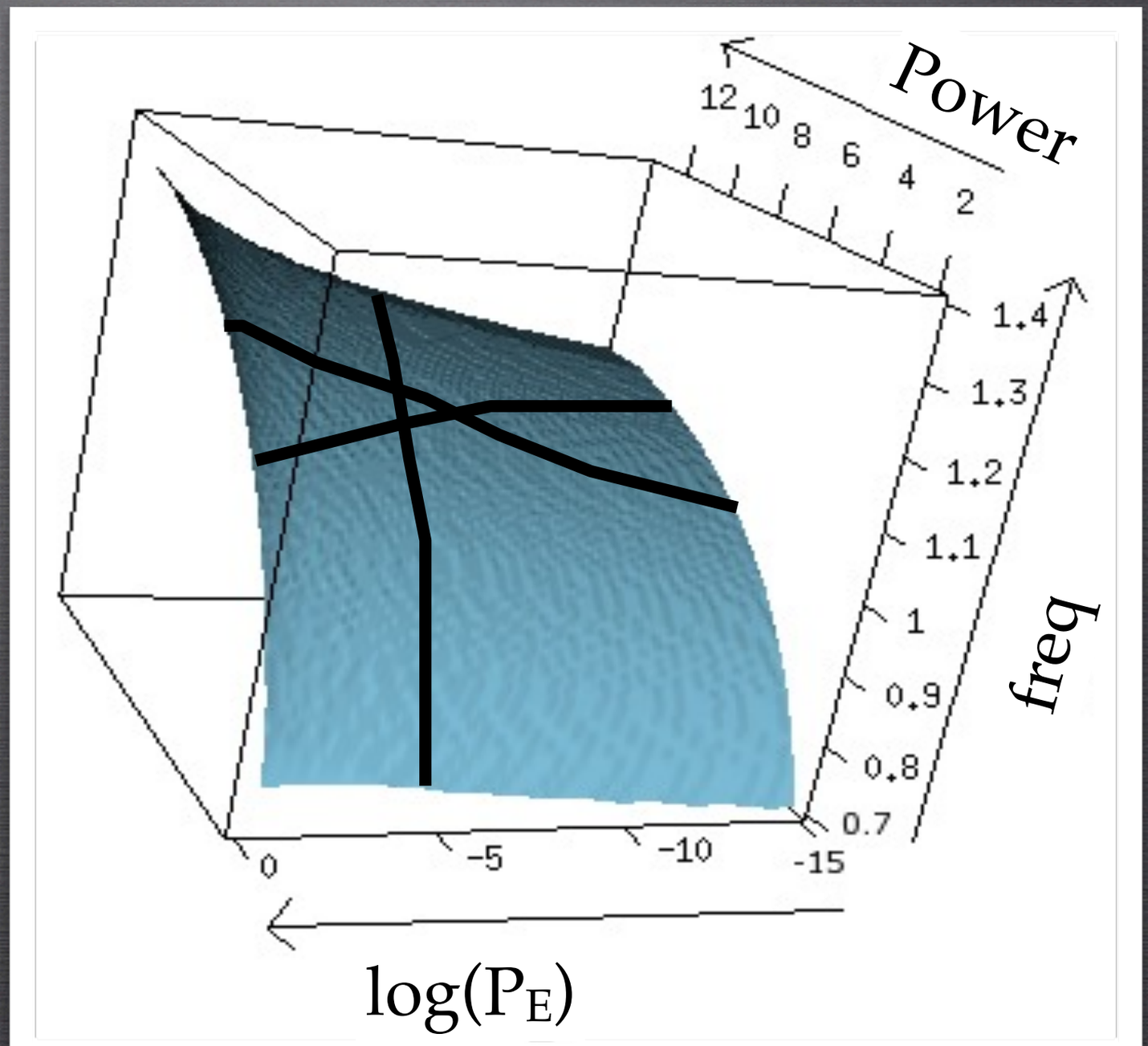

Performance
0.0
0.2
0.4
0.6
0.8
1.0
Higher f,
higher perf
0.7
0.8
0.9
1.0
1.1
freq

Error Rate: A Third Axis for Architects

FREQ ↑
POWER ↑

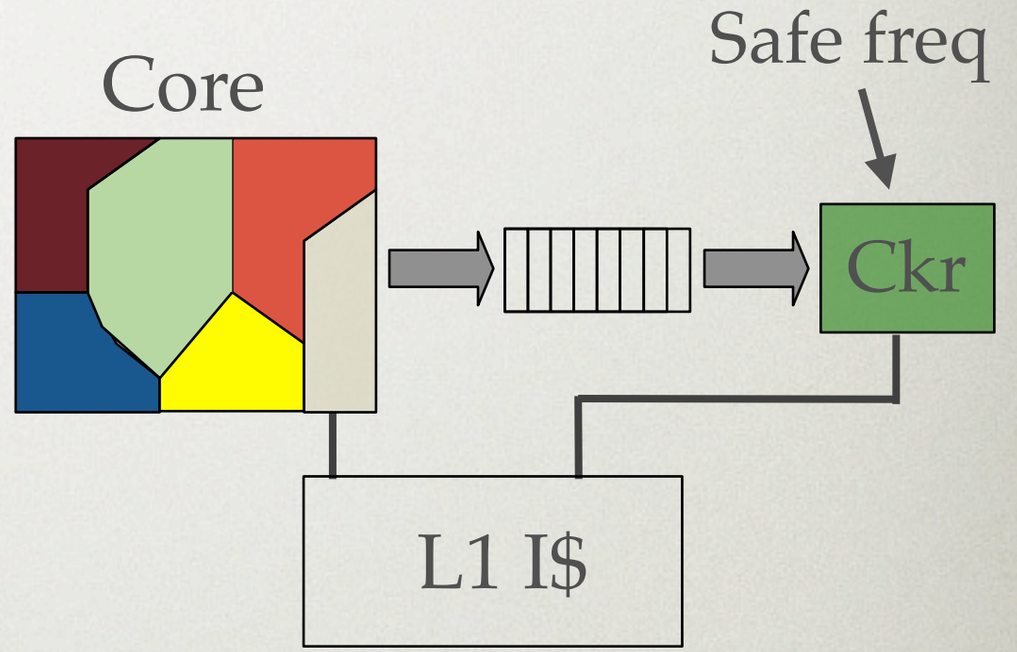
POWER ↓
ERRORS ↑

FREQ ↑
ERRORS ↑↑



Evaluation Setup

- Athlon-like 45nm Core (30W max)
- 15 domains for ASV, ABB
  - V_{DD}: [800mV, 1.2V]
    - V_{BB}: [-500mV, 500mV]
- Diva-like checker
- Variable issue queue (68 or 50 entries)
- Replicated int ALU and FPU add/multiply

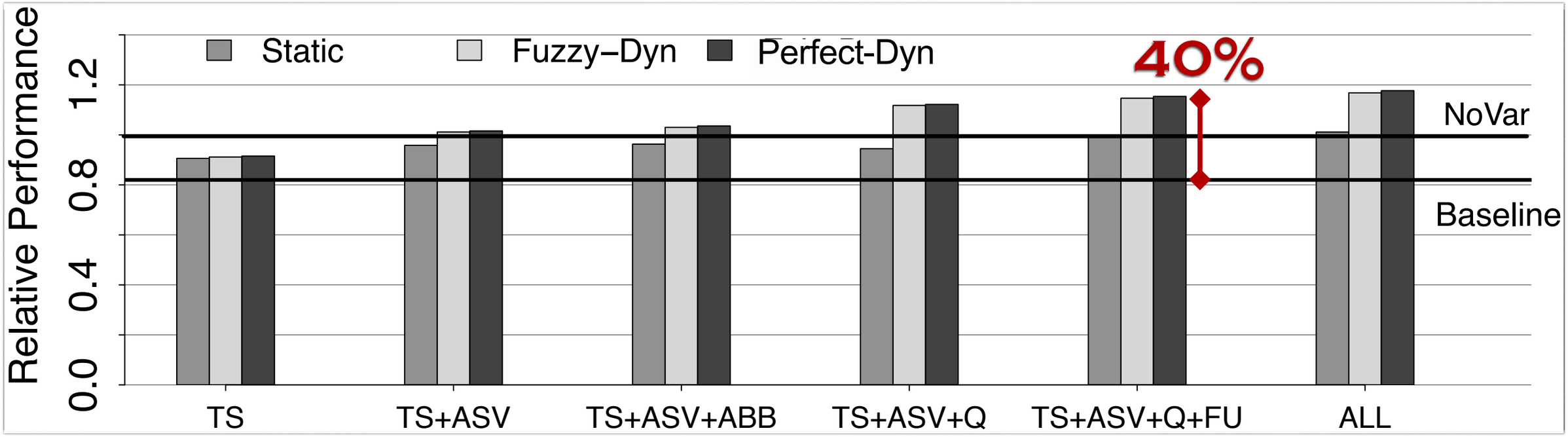


Environments Evaluated

	Timing Speculation (Diva)	ASV	ABB	Issue Q Resizing	FU Replication
Baseline					
TS	X				
TS+ASV	X	X			
TS+ASV+ABB	X	X	X		
TS+ASV+Q	X	X		X	
TS+ASV+Q+FU	X	X		X	X
ALL	X	X	X	X	X

i-acoma group

Performance



- 40% performance improvement in preferred environment (11% area overhead)
- Fuzzy controller as good as a perfect one

Conclusions

- Error rate as a *third axis* for µarch optimization beyond power and frequency
- High-dimensional dynamic adaptation optimizes power and performance in the presence of errors
  - Fuzzy controller is fast and accurate
  - 40% performance improvement under power & temp constraints, only 11% core area overhead

EVAL

Utilizing Processors with Variation-Induced Timing Errors

Smruti R. Sarangi

Brian Greskamp

Abhishek Tiwari

Josep Torrellas

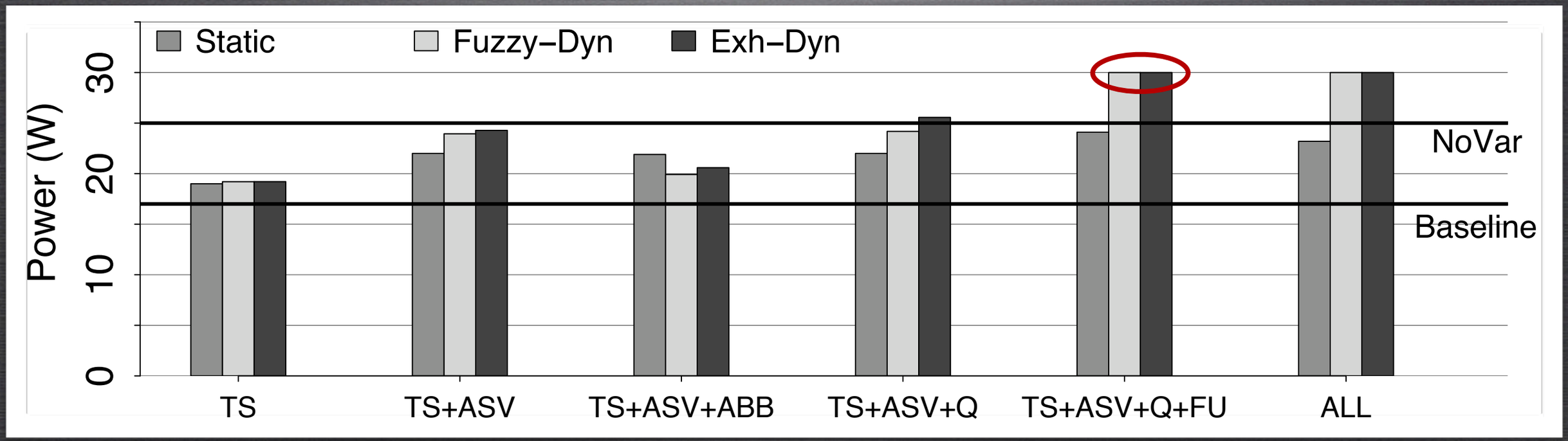
University of Illinois

http://iacoma.cs.uiuc.edu

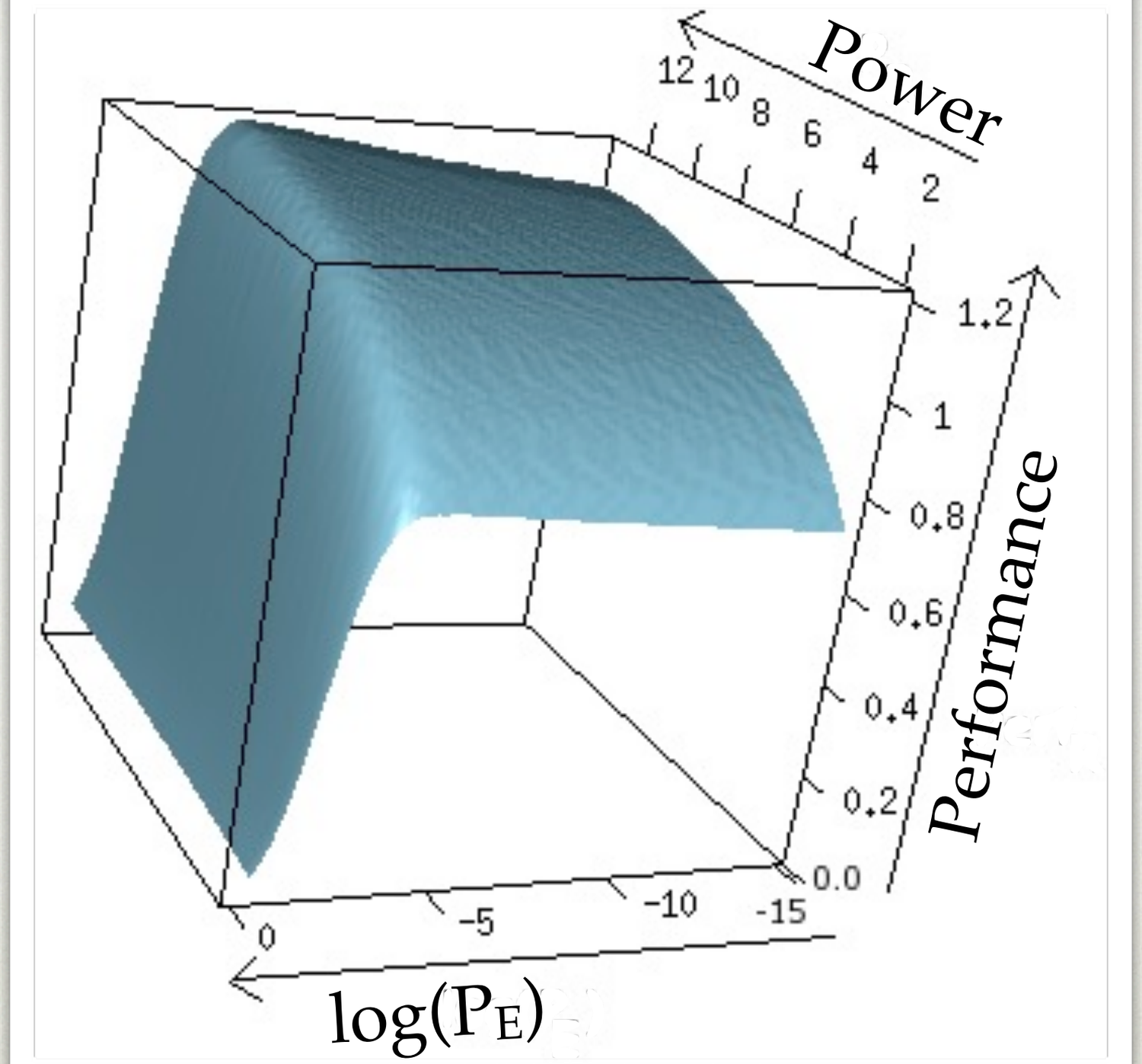
University of Illinois

i-acoma group

Power



- Successfully invests max allowable power to achieve performance gain
- Fuzzy control as good as exhaustive



Pareto-optimal surface for a single subsystem with ASV and Body Bias

Improving Controller Accuracy

- Dimension reduction ignores lateral heat conduction, global power constraint
- Control decision can be suboptimal or infeasible
- *Retuning cycles* repair or improve the solution
  - If initial soln is infeasible, increase freq exponentially until it is feasible
  - Increase freq linearly until a constraint is violated, back off by one step

i-acoma group

Controller Operation