Mitigating Parameter Variation with Dynamic Fine-Grain Body Biasing*

Radu Teodorescu, Jun Nakano, Abhishek Tiwari and Josep Torrellas

University of Illinois at Urbana-Champaign

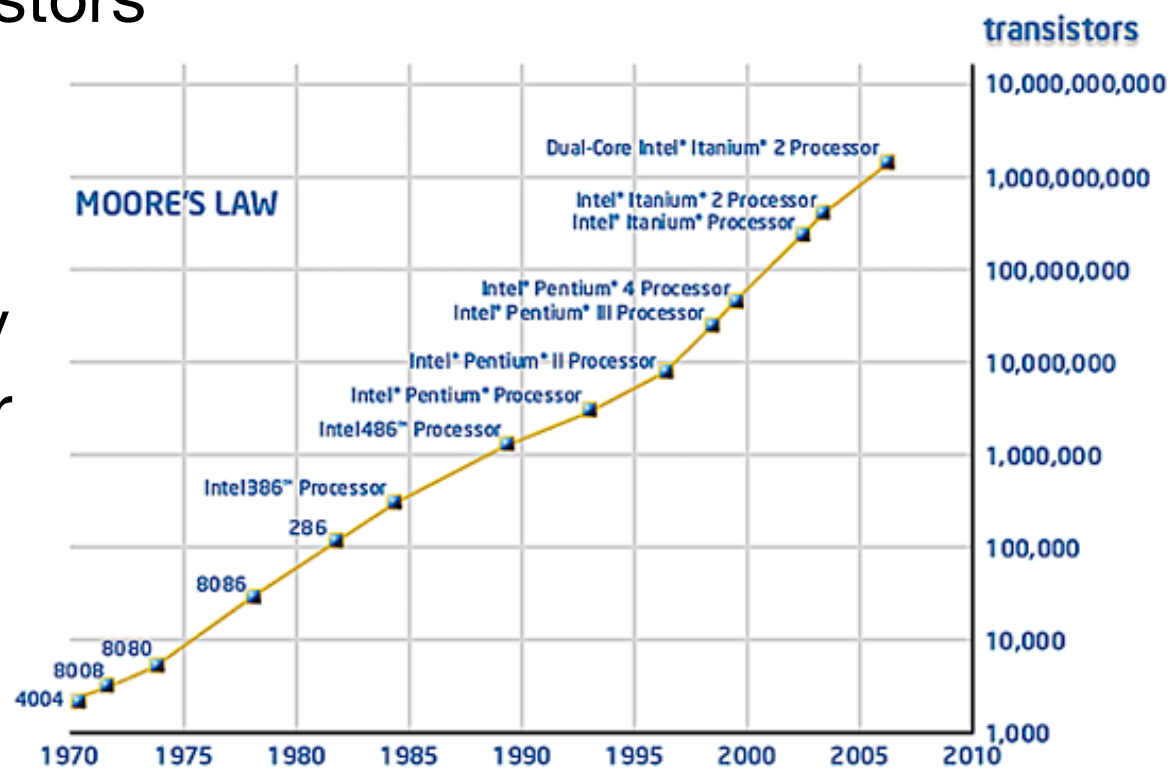
http://iacoma.cs.uiuc.edu

i-acoma group

*to appear in MICRO-40, December 2007

Motivation

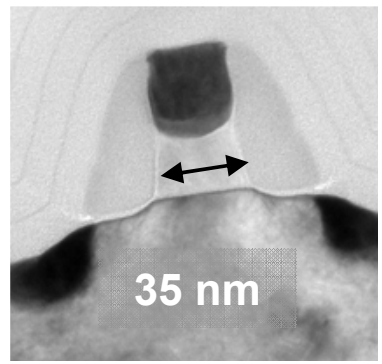
- Technology scaling continues
- More and more transistors every generation!
- However...
- Chips are increasingly affected by parameter variation

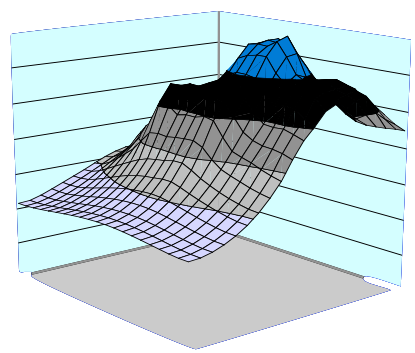


Intel Corp.

Parameter Variation

- Process variation
  - Manufacturing at low feature sizes
- Temperature variation
  - Uneven activity distribution
- Supply voltage variation
  - IR drop, di/dt noise





Intel Corp.

Effects of Parameter Variation

- Higher power consumption
- Lower frequency
- Uncertainty in the design process

Outline

- A Model of Process Variation
- Dynamic Fine-Grain Body Biasing
- Evaluation
- Conclusions

Outline

- A Model of Process Variation
- Dynamic Fine-Grain Body Biasing
- Evaluation
- Conclusions

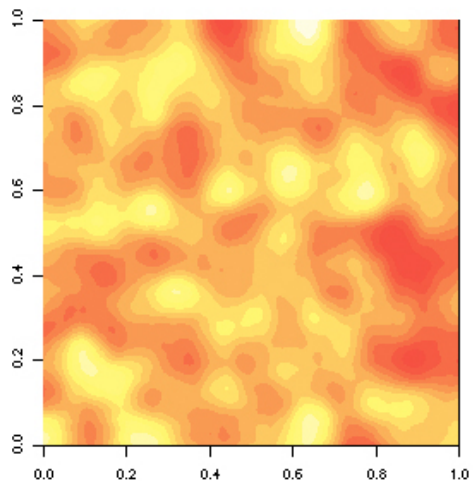
A Model For Process Variation

- Fast, simple and parameterizable model
- We model two key process parameters:
  - Transistor critical dimension (L_{eff}) and threshold voltage (V_{th})
- We also model temperature effects

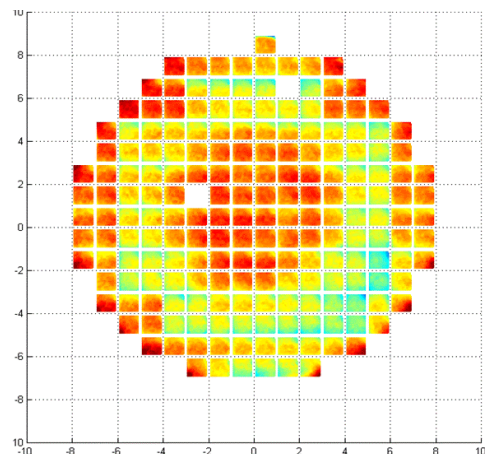
Variation Components

- Granularity:
  - Within die
  - Die-to-die

Within die



Die-to-die



- WID variation:
  - Systematic variation
  - Random variation

A Model For Process Variation

- Variation in any parameter P:

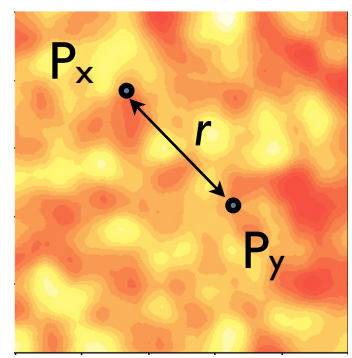
$$\Delta P = \Delta P_{D2D} + \Delta P_{WID} = \Delta P_{D2D} + \Delta P_{rand} + \Delta P_{sys}$$

- We focus on WID variation
  - D2D is a chip-wide offset to ΔP_{WID}
- Random and systematic components
  - Modeled as normal distributions
  - Treated separately - impact different levels of the microarchitecture

Systematic Variation

- We divide the chip into a grid of points
  - Each point has one random value of ΔP_{sys}
- Multivariate normal distribution ($\mu_{sys}=0$, σ_{sys})
  - Characterized by a correlation function:

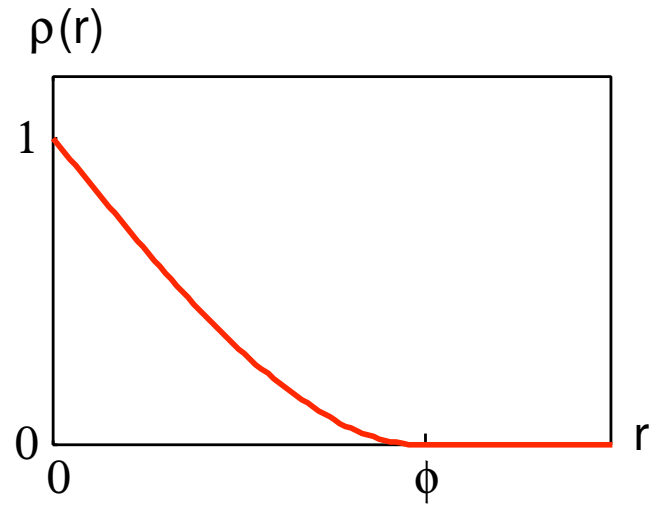
$$corr(P_{\vec{x}}, P_{\vec{y}}) = \rho(r)\ ;\ r = |\vec{x} - \vec{y}|$$



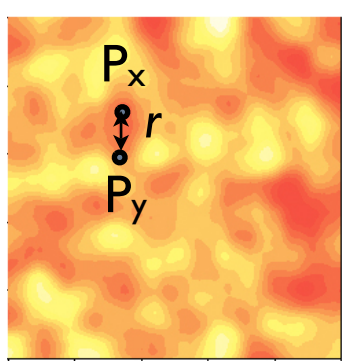
  - Correlation is position independent and isotropic
  - For *ρ(r)* we choose the spherical model

Spherical Model

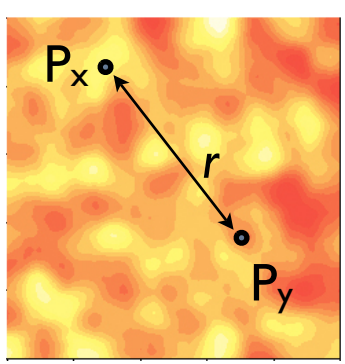
$$\rho(r) = \begin{cases} 1 - \frac{3r}{2\phi} + \frac{r^3}{2\phi^3} & : (r \leq \phi) \\ 0 & : \text{otherwise} \end{cases}$$



Stronger correlation



Weaker correlation



- Matches measured data [Friedberg et al. 05]

Random Variation

- Random variation - transistor level
- We model it analytically as a normal distribution
- Both ΔP_{rand} and ΔP_{sys} are normal and independent with σ_{rand} and σ_{sys}

$$\sigma_{total} = \sqrt{\sigma_{rand}^2 + \sigma_{sys}^2}$$

Outline

- A Model of Process Variation
- Dynamic Fine-Grain Body Biasing
- Evaluation
- Conclusions

Body Biasing

- Well known technique for V_{th} control
- A voltage is applied between source/drain and substrate of a transistor
  - Forward body bias: FBB - V_{th} ↓ - Freq ↑ - Leak ↑
  - Reverse body bias: RBB - V_{th} ↑ - Freq ↓ - Leak ↓
- Useful knob to control frequency and leakage

Body Bias Design Space

Space \ Time	Static BB fixed for chip lifetime	Simple adaptation FBB in active mode RBB in standby	Dynamic BB changes with T and workload
Chip-wide	• D2D variation [Intel Xscale]	• D2D variation, power, performance [Intel's 80-core chip]	
Fine-grain	• WID variation [Tschanz et al]	• WID variation, power, performance	• WID variation • T variation (space and time)

Body Bias Design Space

Space \ Time	Static BB fixed for chip lifetime	Simple adaptation FBB in active mode RBB in standby	Dynamic BB changes with T and workload
Chip-wide	• D2D variation [Intel Xscale]	• D2D variation, power, performance [Intel's 80-core chip]	
Fine-grain	• WI[S-FGBB [Tschanz et al]	• WID variation, power, performance	• WID variation • T variation (space and time)

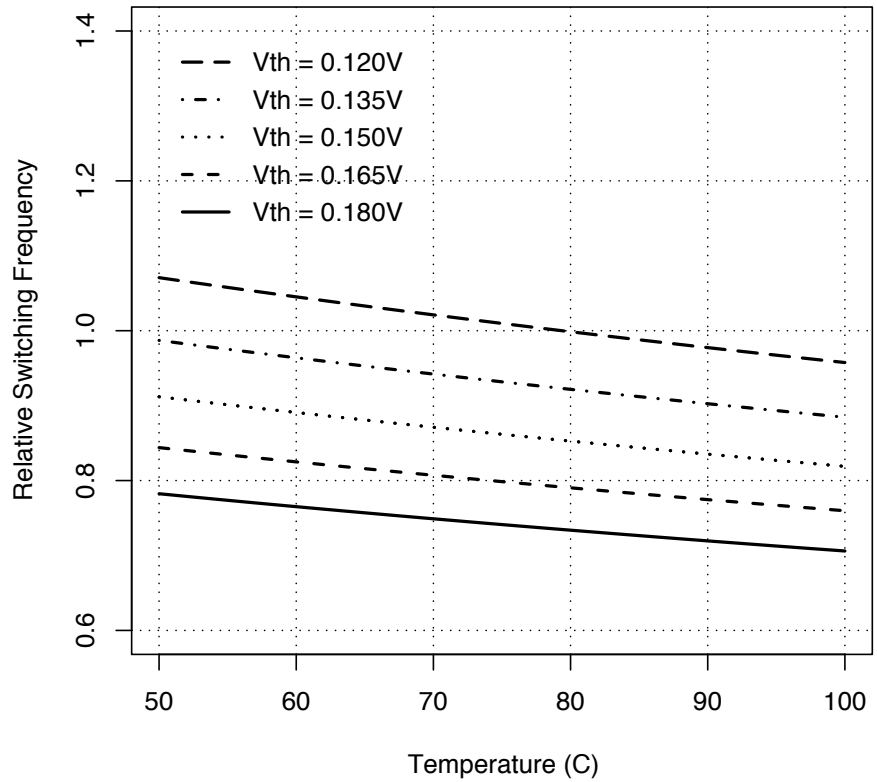
Body Bias Design Space

Space \ Time	Static BB fixed for chip lifetime	Simple adaptation FBB in active mode RBB in standby	Dynamic BB changes with T and workload
Chip-wide	• D2D variation [Intel Xscale]	• D2D variation, power, performance [Intel's 80-core chip]	
Fine-grain	• WI[S-FGBB [Tschanz et al]	• WID variation, power, performance	• WID variation • T D-FGBB (space and time)

Motivation for D-FGBB

- Body bias trades off frequency for leakage
- **Optimal** body bias:

  The **lowest** FBB or **highest** RBB s.t. circuit delay meets frequency target



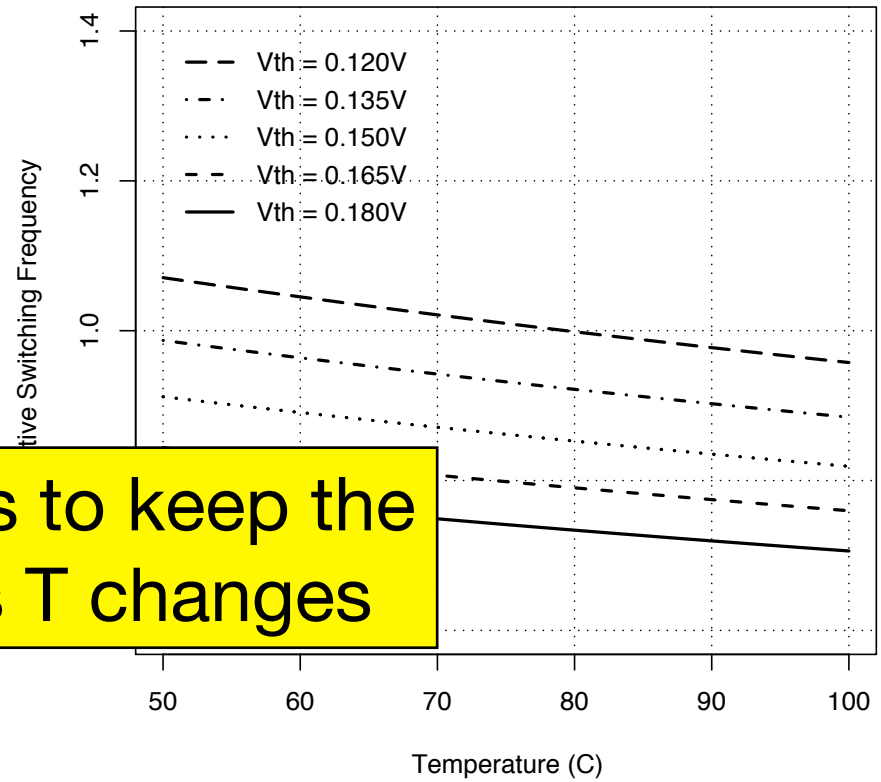
- Circuit delay changes with temperature
- Therefore optimal BB changes with temperature

Motivation for D-FGBB

- Body bias trades off frequency for leakage
- **Optimal** body bias:

The **lowest** FBB or **highest** RBB s.t. circuit d target

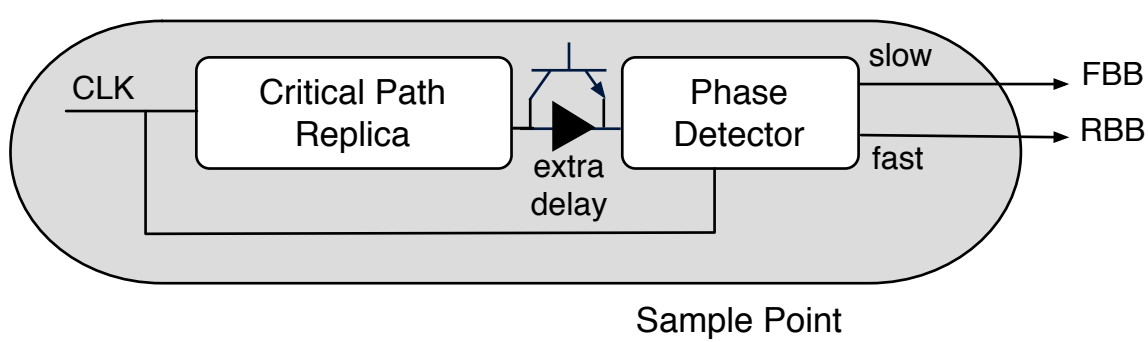
The goal of D-FGBB is to keep the body bias optimal as T changes



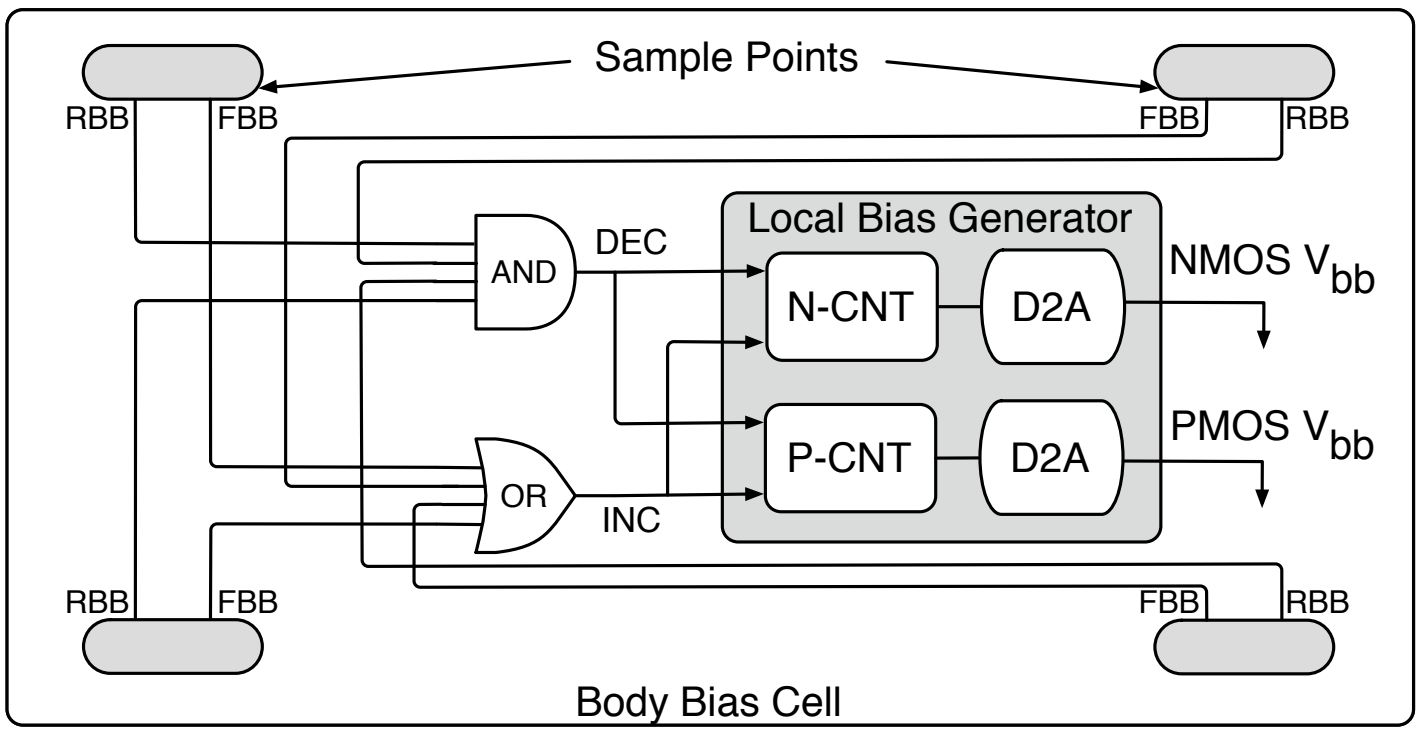
- Circuit delay changes with temperature
- Therefore optimal BB changes with temperature

Finding the Optimal BB

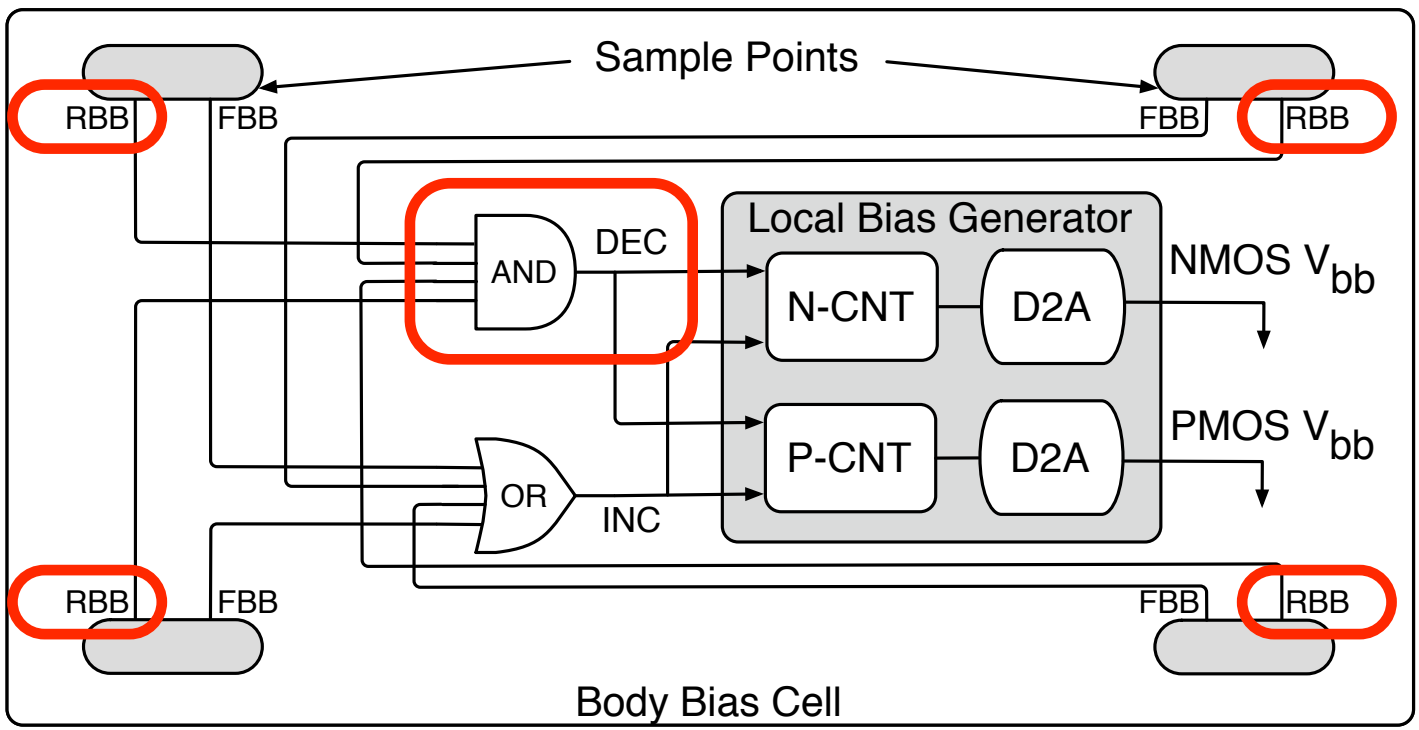
- Measure the delay of each BB cell
- Critical path replicas to sample cell delay
- Phase detector “times” the critical path replica
  - If slow - FBB signal raised
  - If fast - RBB signal raised



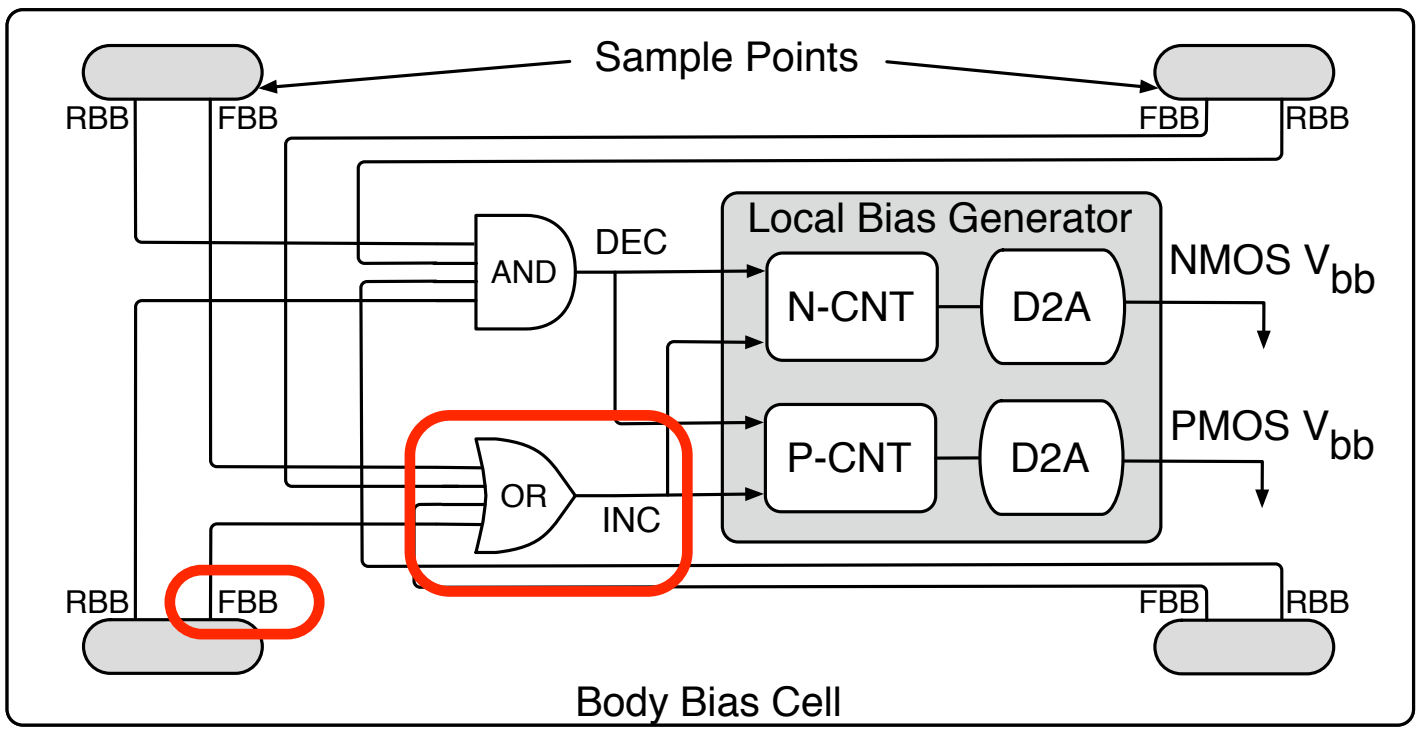
Applying Fine Grain BB



Applying Fine Grain BB



Applying Fine Grain BB



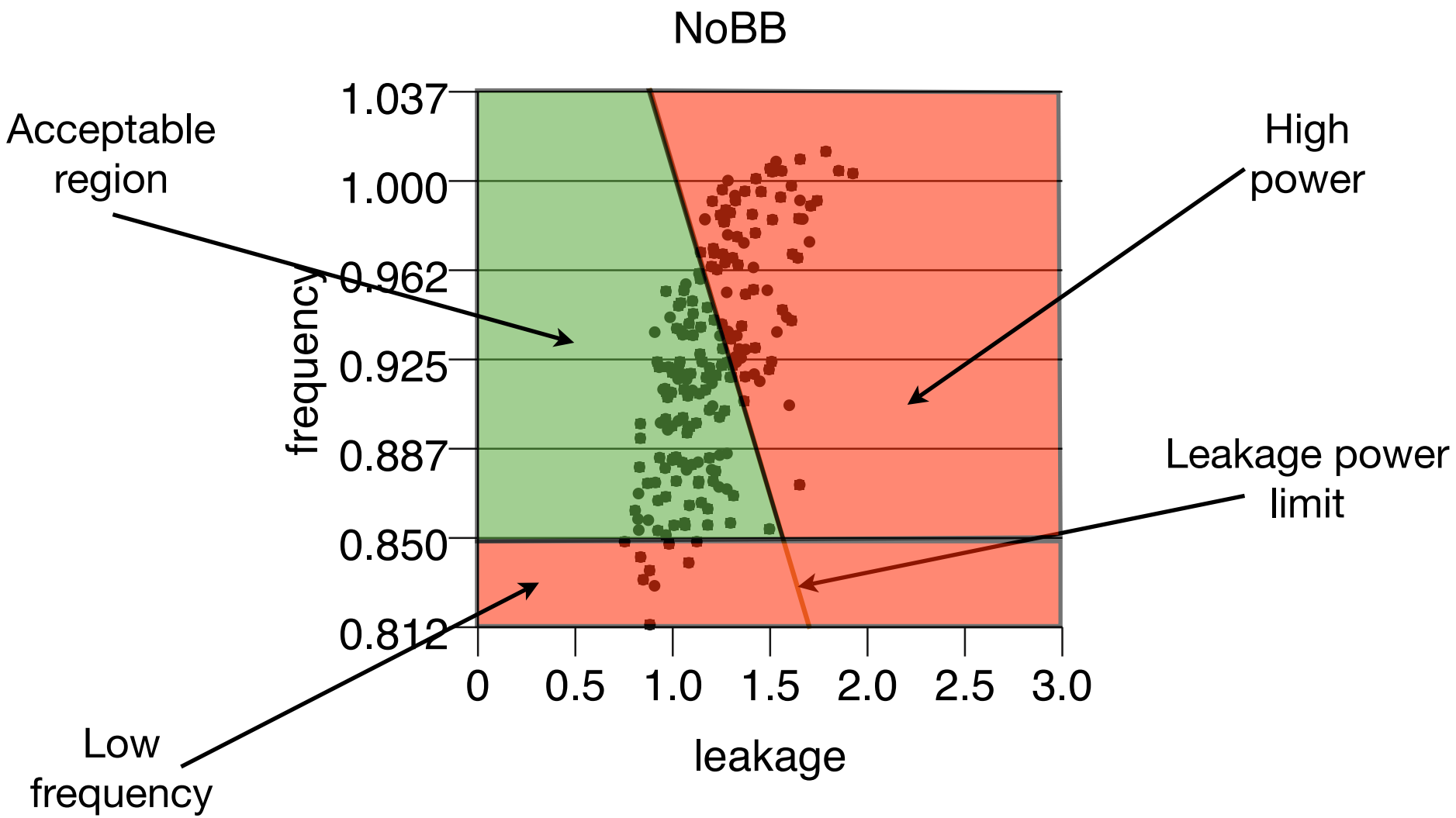
Applications of D-FGBB

Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

Applications of D-FGBB

Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

Improving a Chip's Operating Point

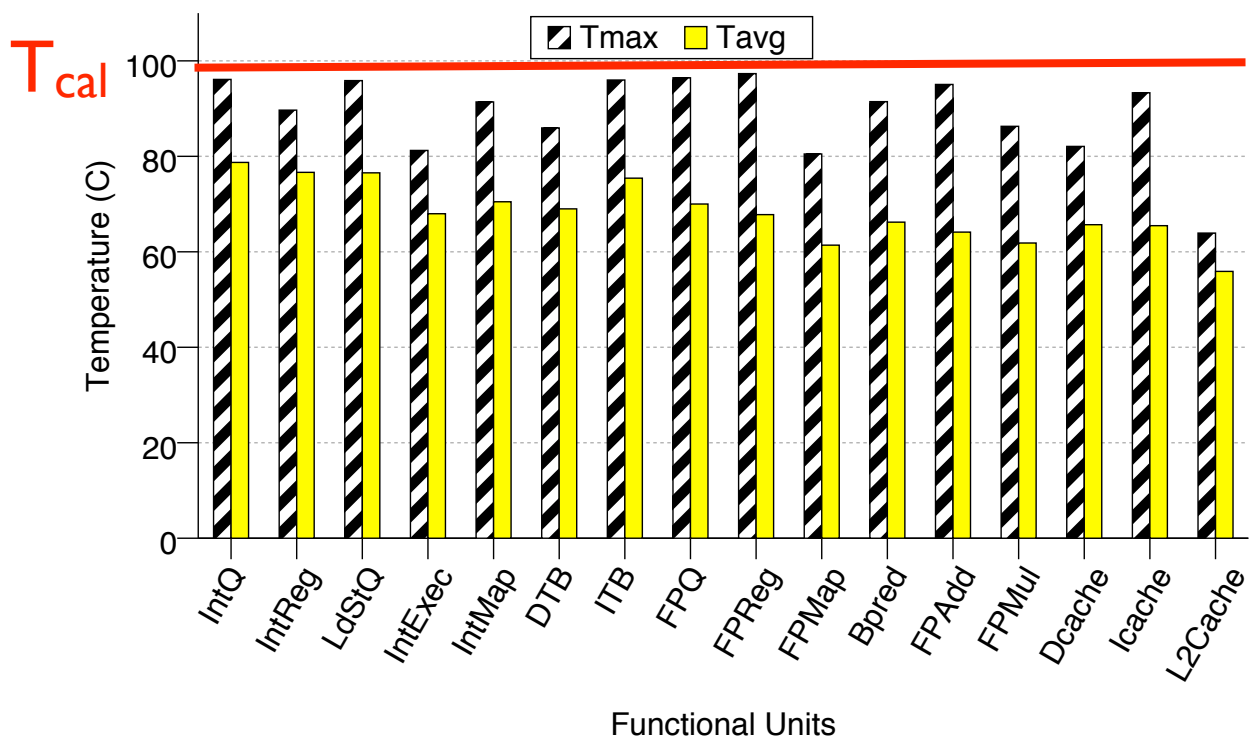


Improving a Chip's Operating Point

- Post-manufacturing calibration phase:
  1. Bring chip to T_{cal}
  2. Set target frequency F_{cal}^0, and run at full load
  3. BB is adjusted automatically
  4. Measure total power P_{cal}: if $P_{cal}<P_{target}$, $F_{cal}^1=F_{cal}^0$++, else $F_{cal}^1=F_{cal}^0$--
  5. Repeat if needed, until $P_{cal} \approx P_{target}$
- F_{cal}^i becomes the chip's frequency

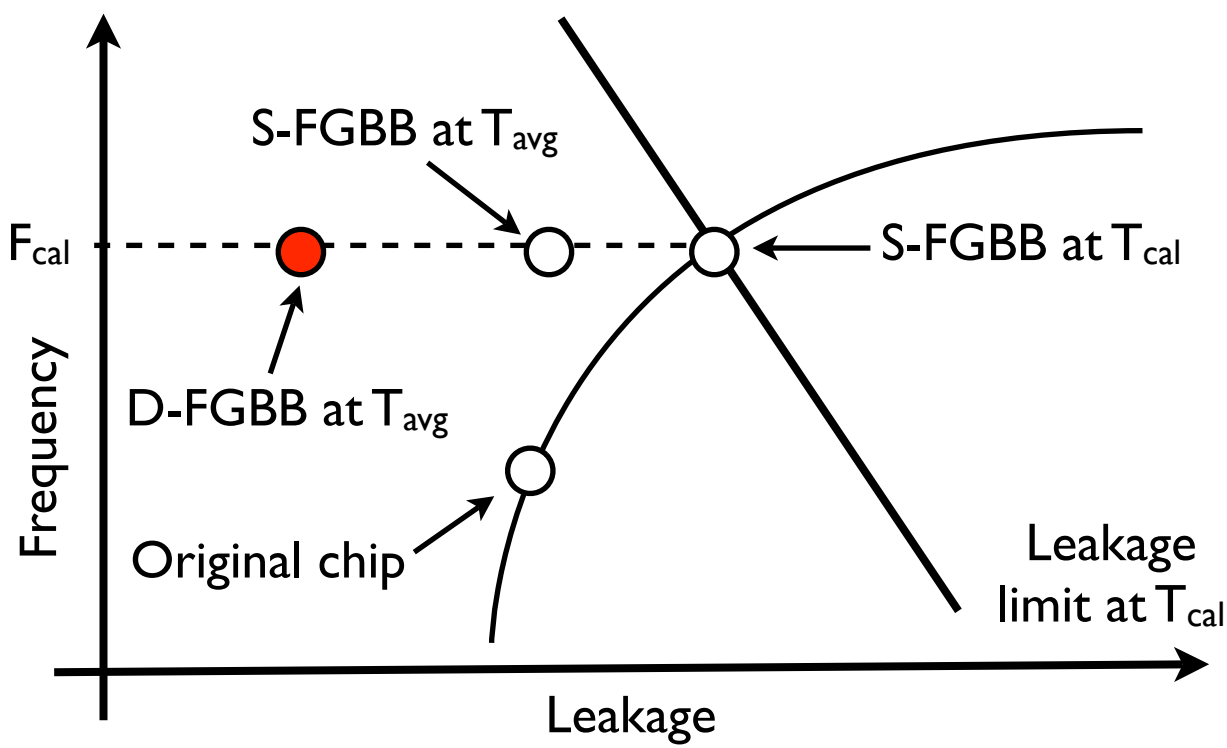
D-FGBB Adapts to Changes in T

- Calibration temperature T_{cal} is conservative
- Average T much lower:



D-FGBB Saves Leakage Power

- S-FGBB finds and sets F_{cal}
- D-FGBB adjusts dynamically to T changes to save power while running at F_{cal}

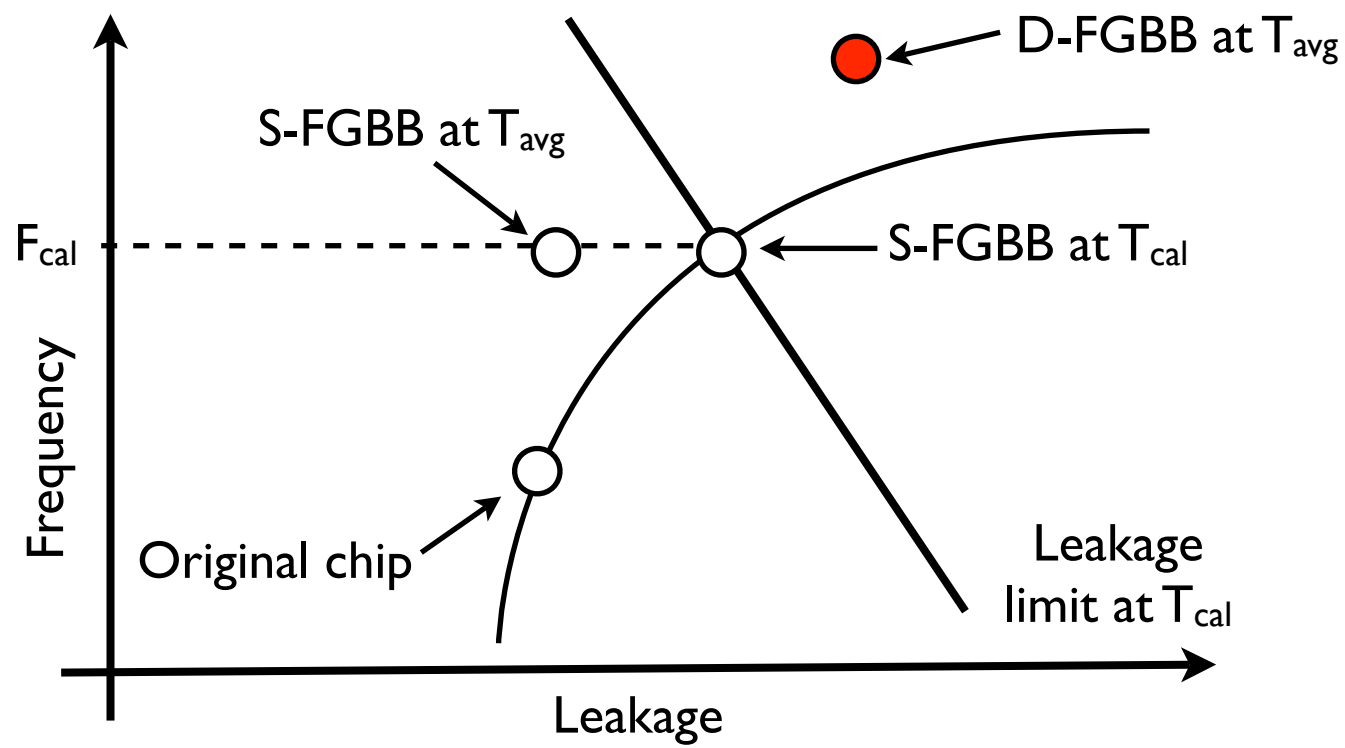


Applications of D-FGBB

Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

D-FGBB Improves Performance

- Average power $P_{avg}<P_{max}$
- D-FGBB is used to push the chip to $F_{avg}>F_{cal}$, as long as $P<P_{max}$

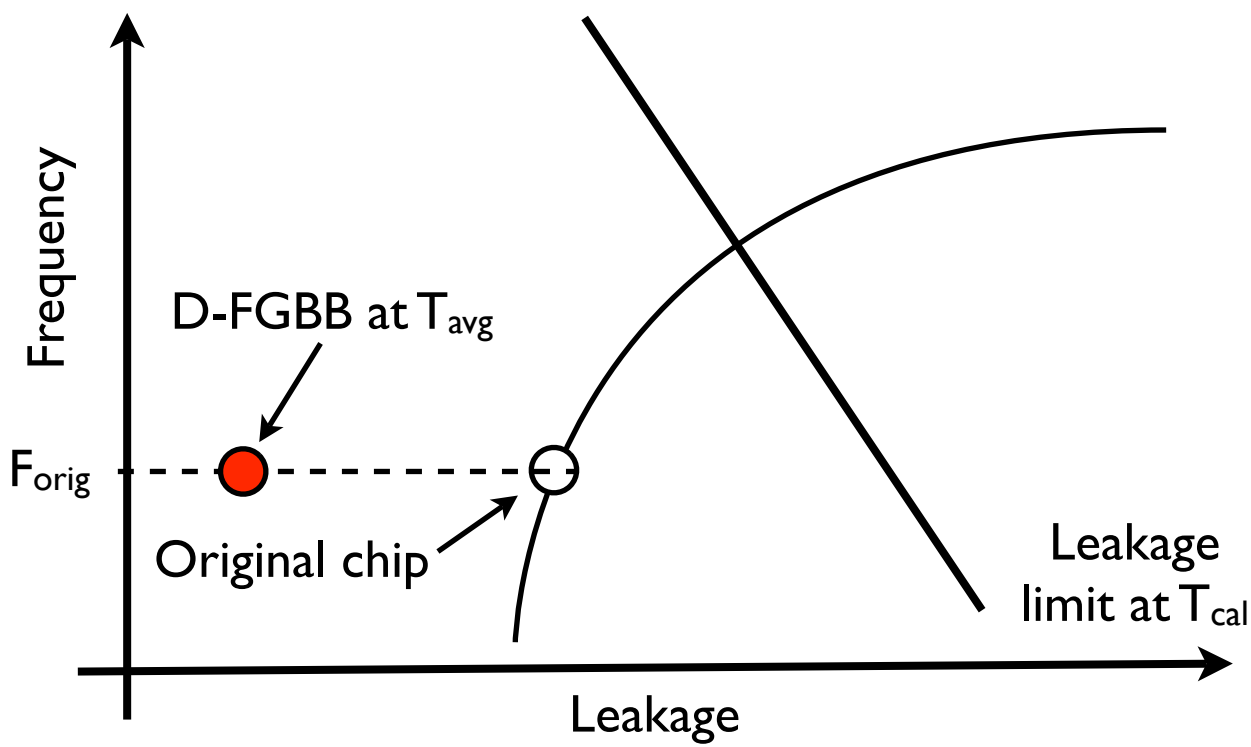


Applications of D-FGBB

Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

D-FGBB Saves Leakage Power

- The chip runs at its original F_{orig}
- D-FGBB adjusts dynamically to T changes to save power while running at F_{orig}



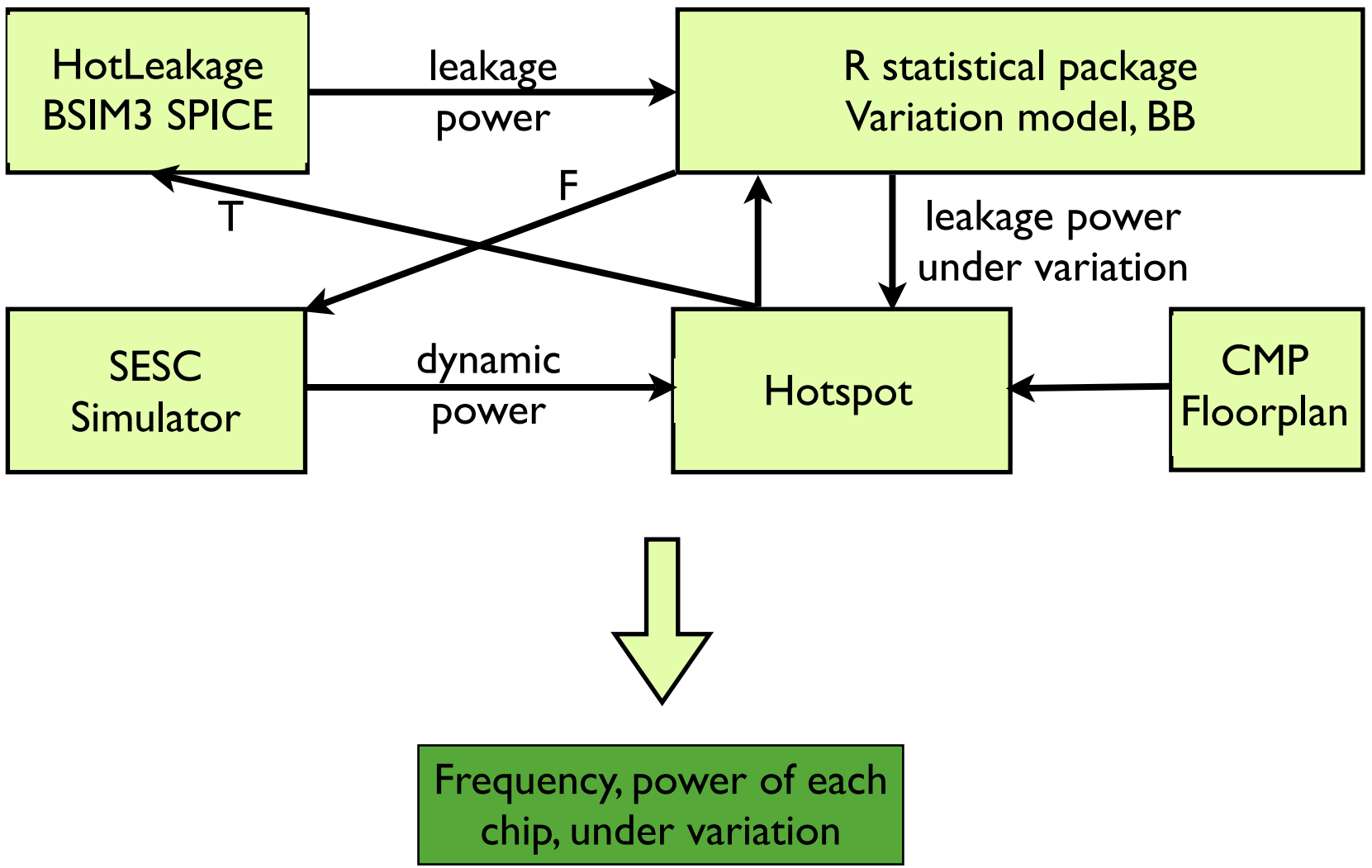
Outline

- A Model of Process Variation
- Dynamic Fine-Grain Body Biasing
- Evaluation
- Conclusions

Evaluation Infrastructure

- Statistical package R to generate variation maps for 200 chips
- SESC - cycle accurate microarchitectural simulator - execution time, dynamic power
  - Mix of SPECint and SPECfp benchmarks
- HotLeakage, SPICE model - leakage power
- Hotspot - temperature estimation

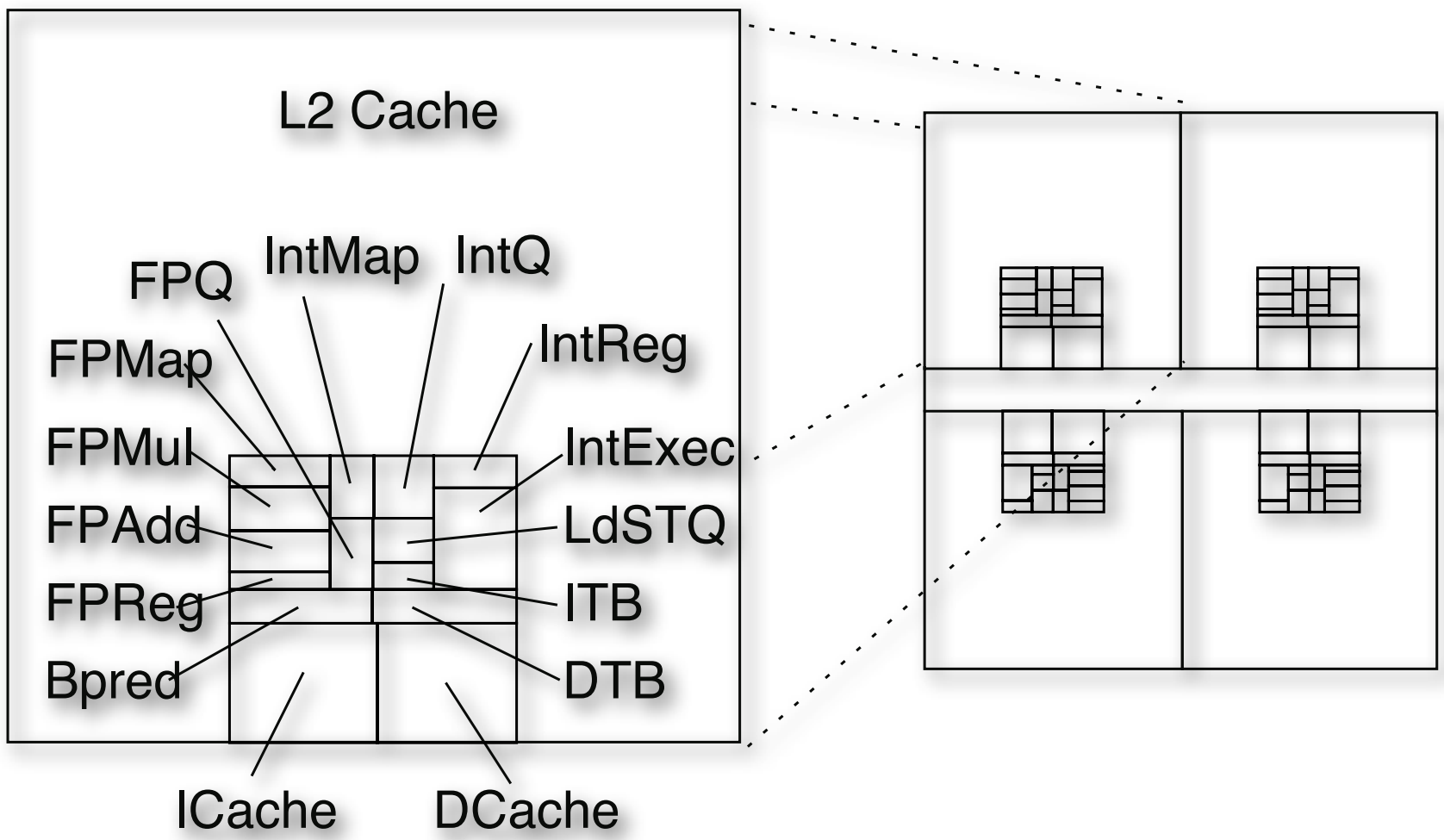
Evaluation Infrastructure



Evaluation Methodology

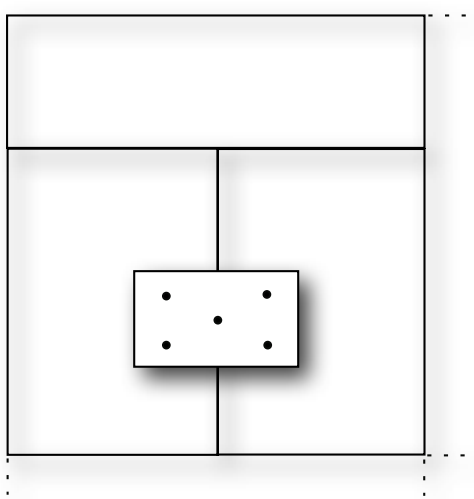
- 4-core CMP, based on Alpha 21364
- 45nm technology, 4GHz
- V_{th} variation: σ_{Vth}/μ_{Vth}=0.3-0.12, $\sigma_{sys}=\sigma_{rand}$
- L_{eff} variation $\sigma_{Leff}= \sigma_{Vth}/2$
- V_{dd}=1V, V_{th0}=150mV, $V_{bb}= \pm$500mV

CMP Architecture

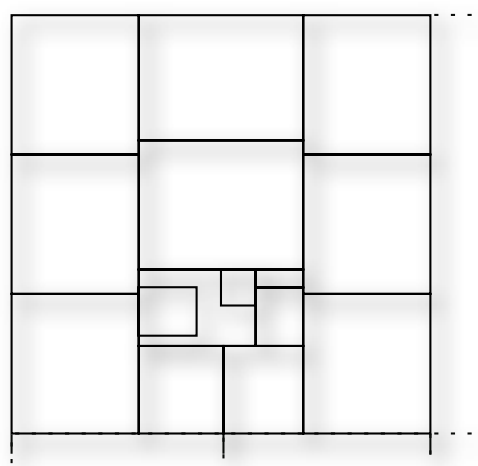


Body Bias Cells

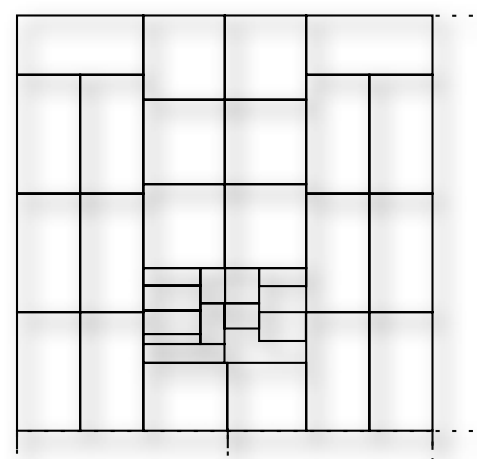
- We partition each core into BB cells
- Shapes and sizes follow functional units



FGBB16

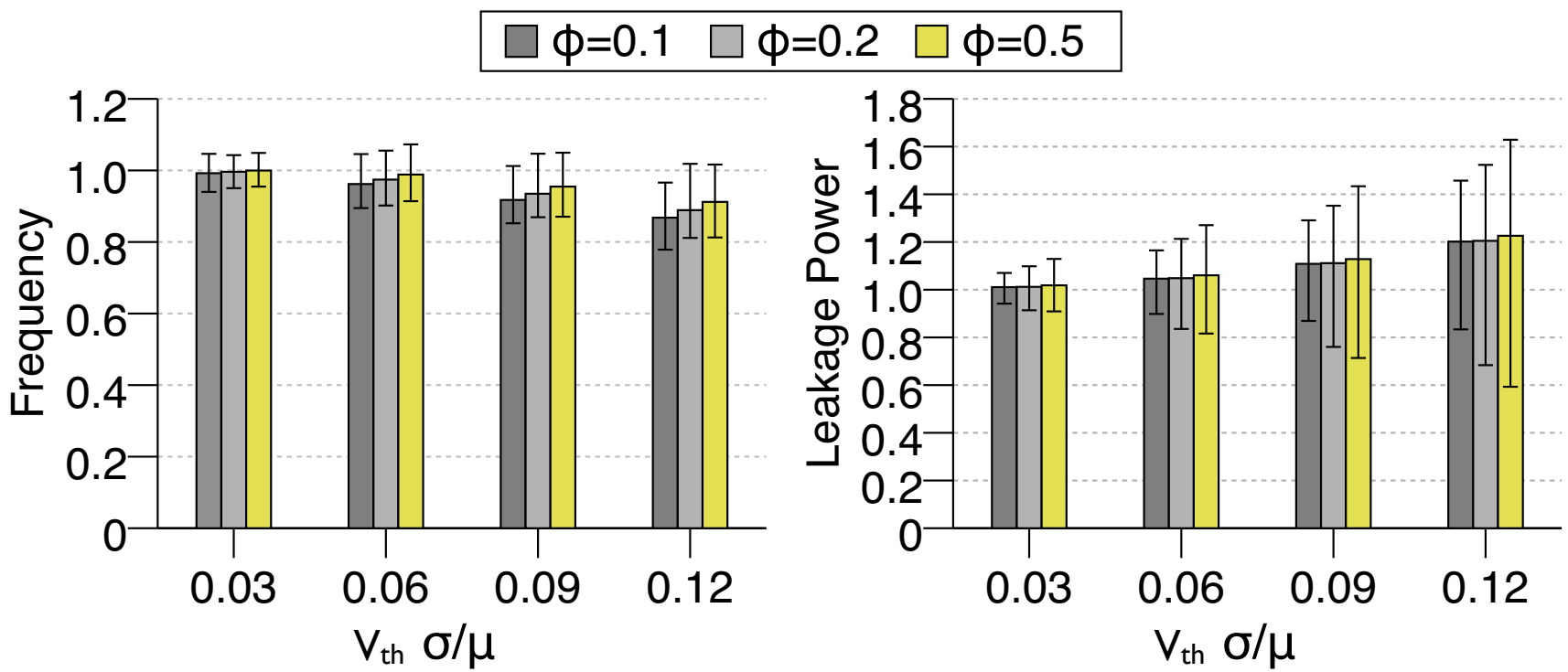


FGBB64



FGBB144

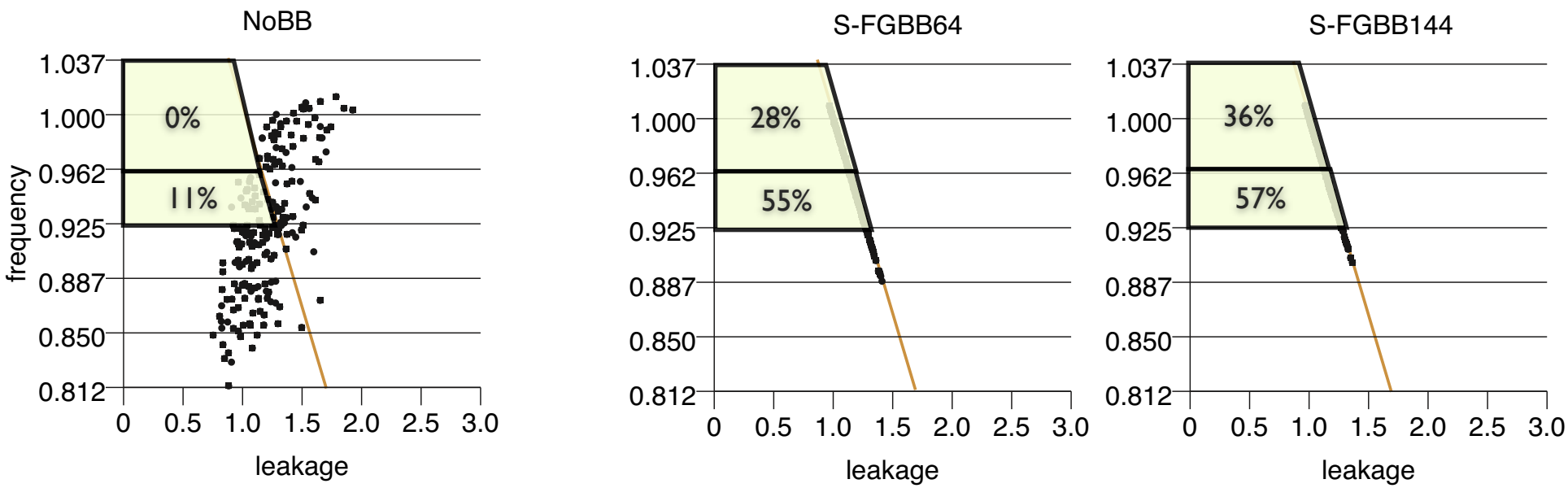
Variation Impact



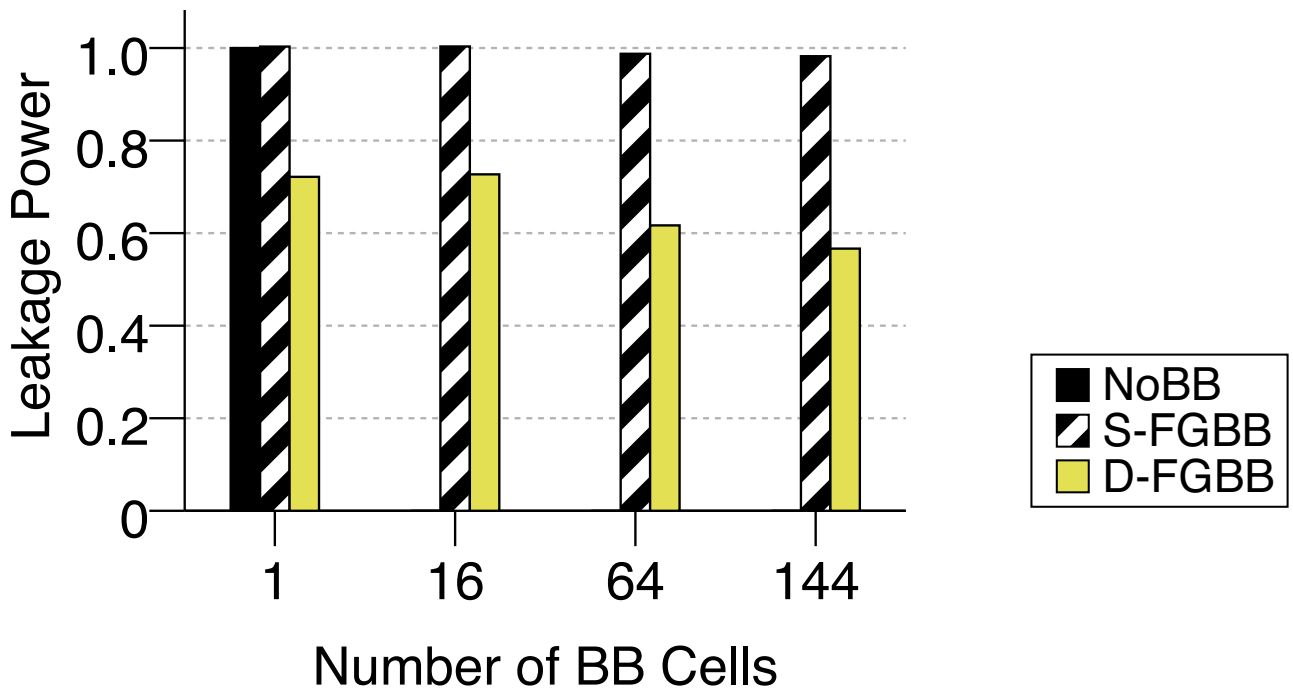
Applications of D-FGBB

Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

S-FGBB Improves the Chip's Operating Point



D-FGBB Reduces Leakage

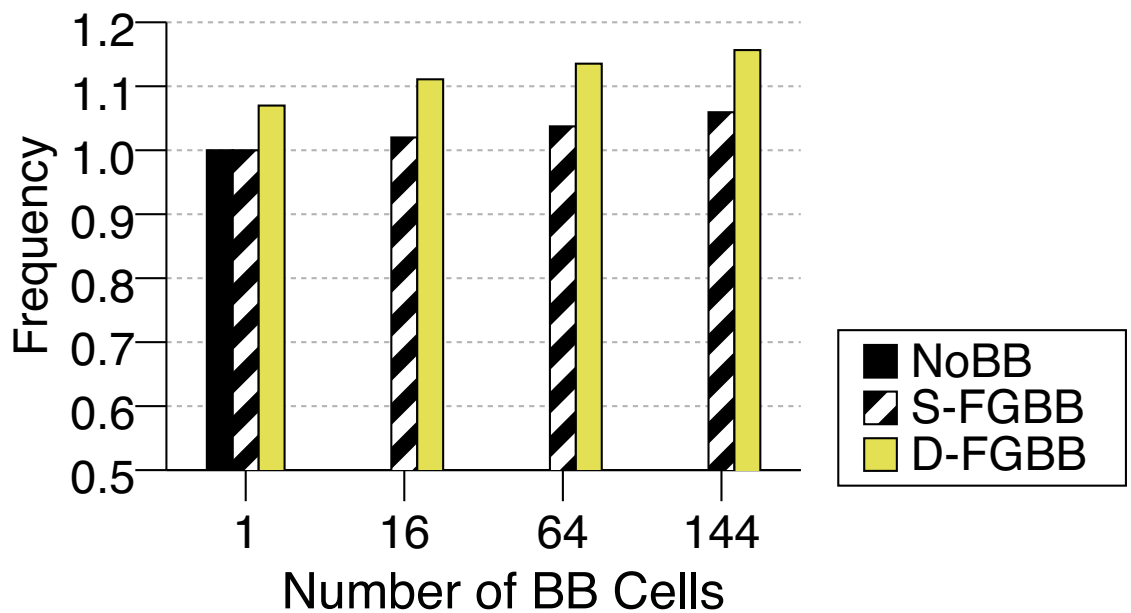


- Large leakage reduction after binning: 28-42%
- More BB cells result in higher savings

Applications of D-FGBB

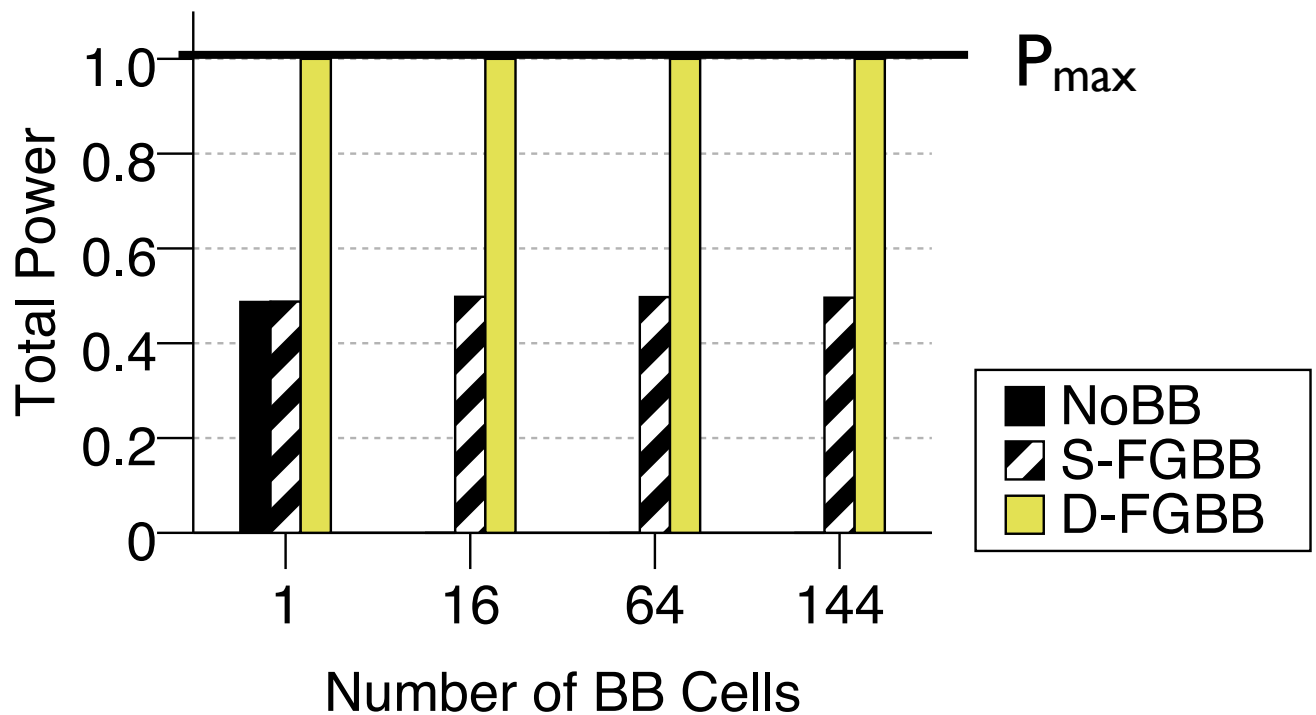
Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

D-FGBB Improves Frequency



- Average frequency improvement 7-9% over S-FGBB and 7-16% over NoBB
- More BB cells result in higher increase

Power Cost

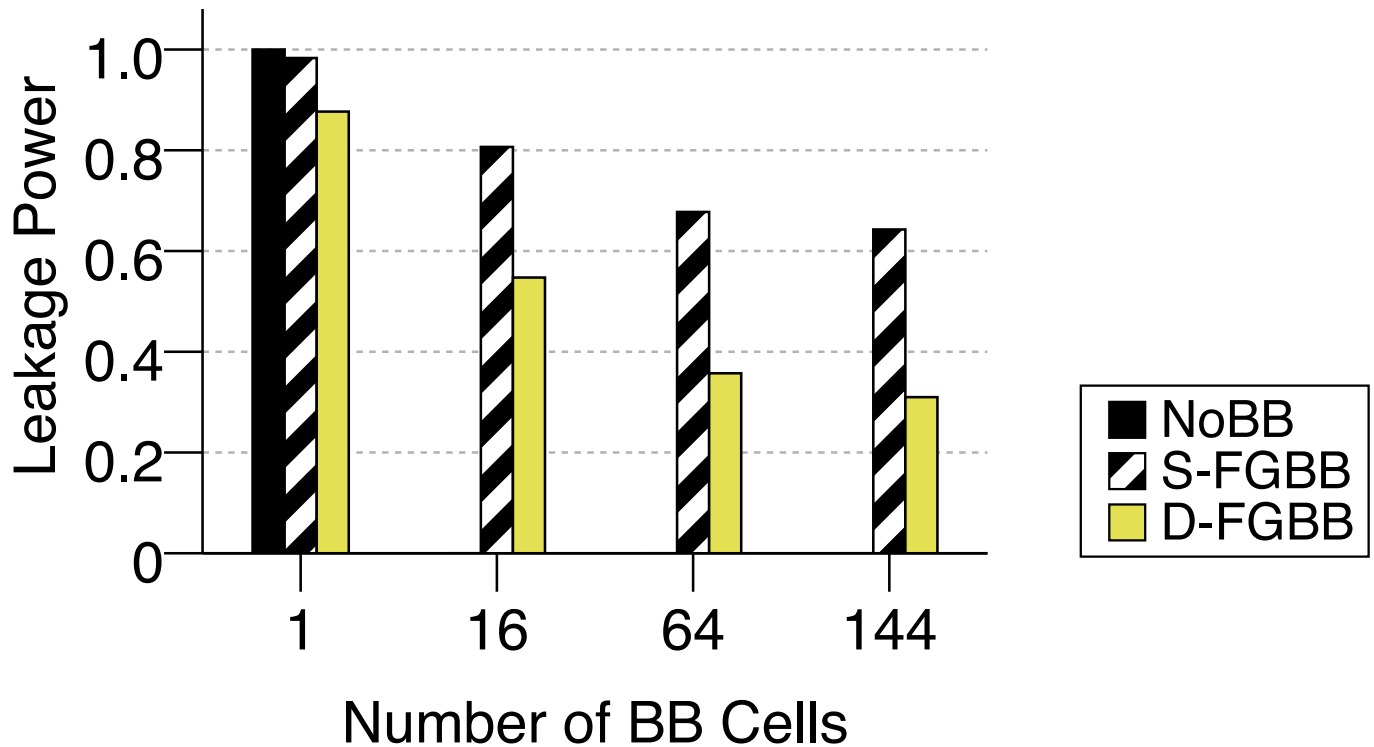


- Significant power cost, but still within the power budget

Applications of D-FGBB

Operating environments	S-FGBB	D-FGBB
Normal	Improve chip operating point	Save leakage power
High Performance	Improve chip operating point	Increase average frequency
Low Power	Save leakage power	Save leakage power

D-FGBB Reduces Leakage

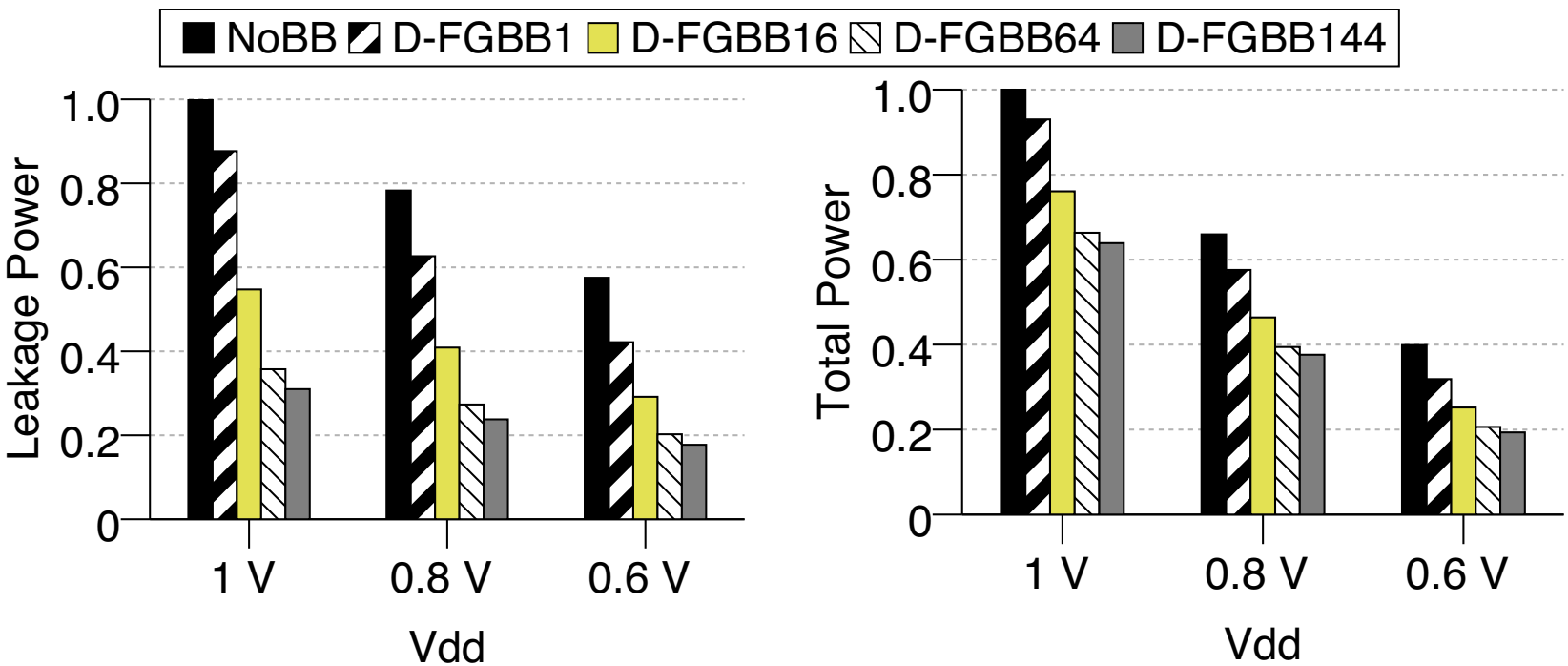


- Large leakage reduction at constant frequency: 10-51% vs. S-FGBB and 12-69% vs NoBB
- More BB cells result in higher savings

Combining D-FGBB with DVFS

- D-FGBB targets leakage power
- DVFS targets mostly dynamic power
- Can they be combined effectively?

Combining D-FGBB with DVFS



- D-FGBB scales well with DVFS
- S-FGBB does not scale unless calibrated at multiple voltages

Conclusions

- D-FGBB is an effective and versatile tool to address parameter variation
- We show three scenarios:
  - Normal: 28-42% leakage savings vs. S-FGBB
  - High performance: 7-9% frequency increase
  - Low power: 10-51% leakage reduction vs. S-FGBB
- Combines well with DVFS

More in our MICRO 2007 paper

http://iacoma.cs.uiuc.edu

- More details on the variation model
- A solution for combining D-FGBB with DVS
- Estimated overheads of D-FGBB
- More implementation details

Thank you! Questions?