VARIUS-NTV: A Model of Process Variations at Near-Threshold Voltages

Ulya R. Karpuzcu❖, Krishna Kolluru*, Nam Sung Kim*, Josep Torrellas❖

❖ University of Illinois

* University of Wisconsin

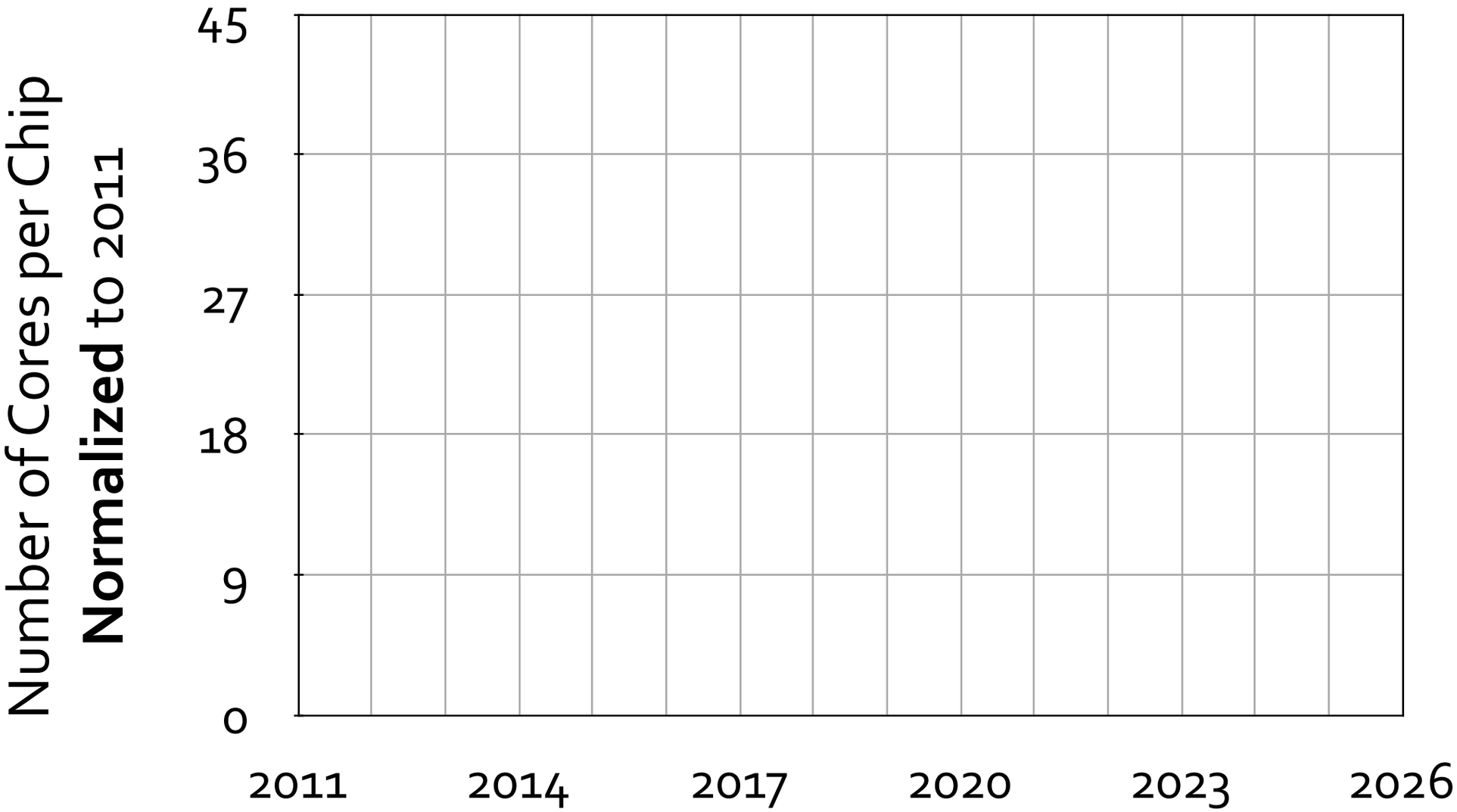
http://iacoma.cs.uiuc.edu/varius/ntv

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

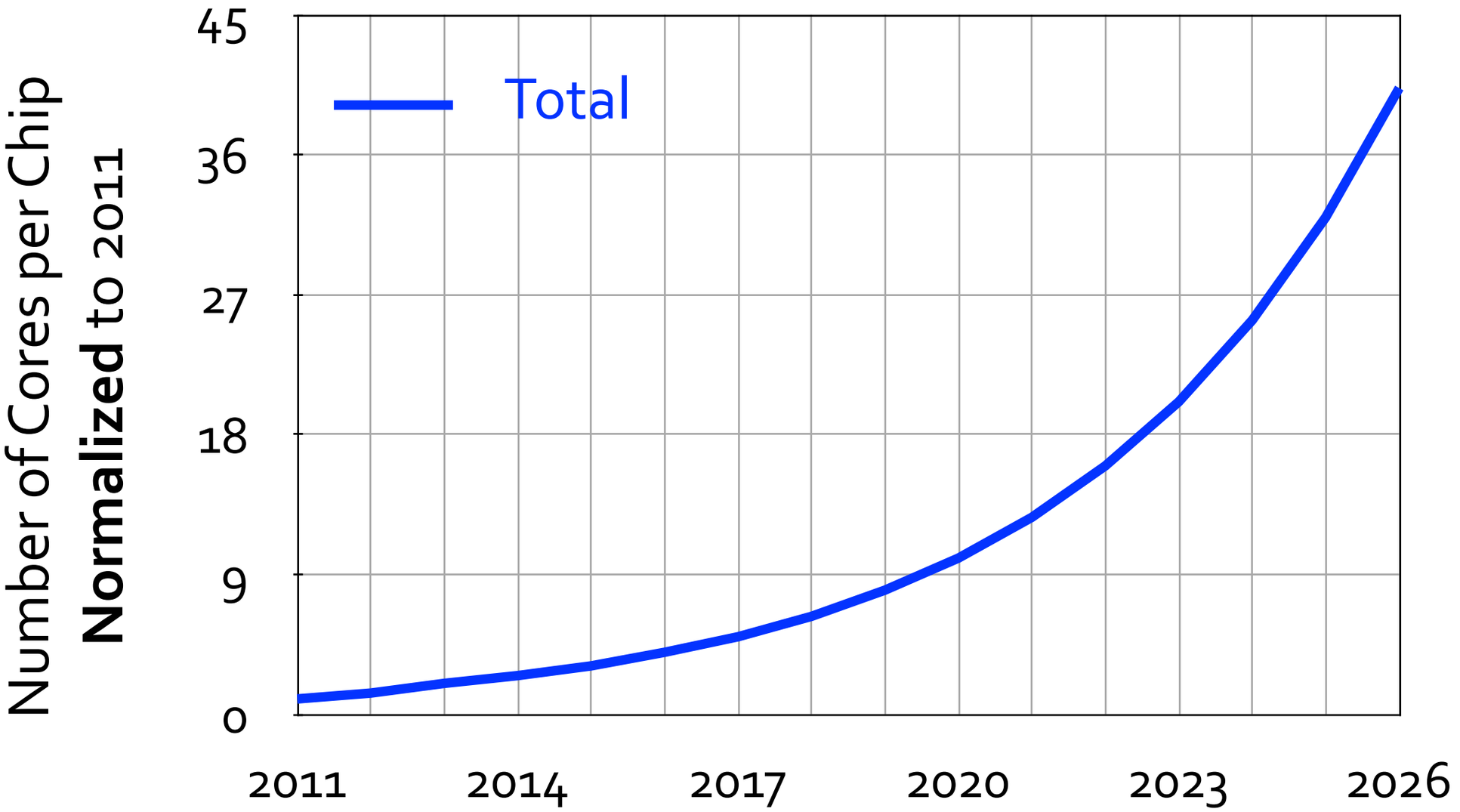
WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Problem: The Many-Core Power Wall

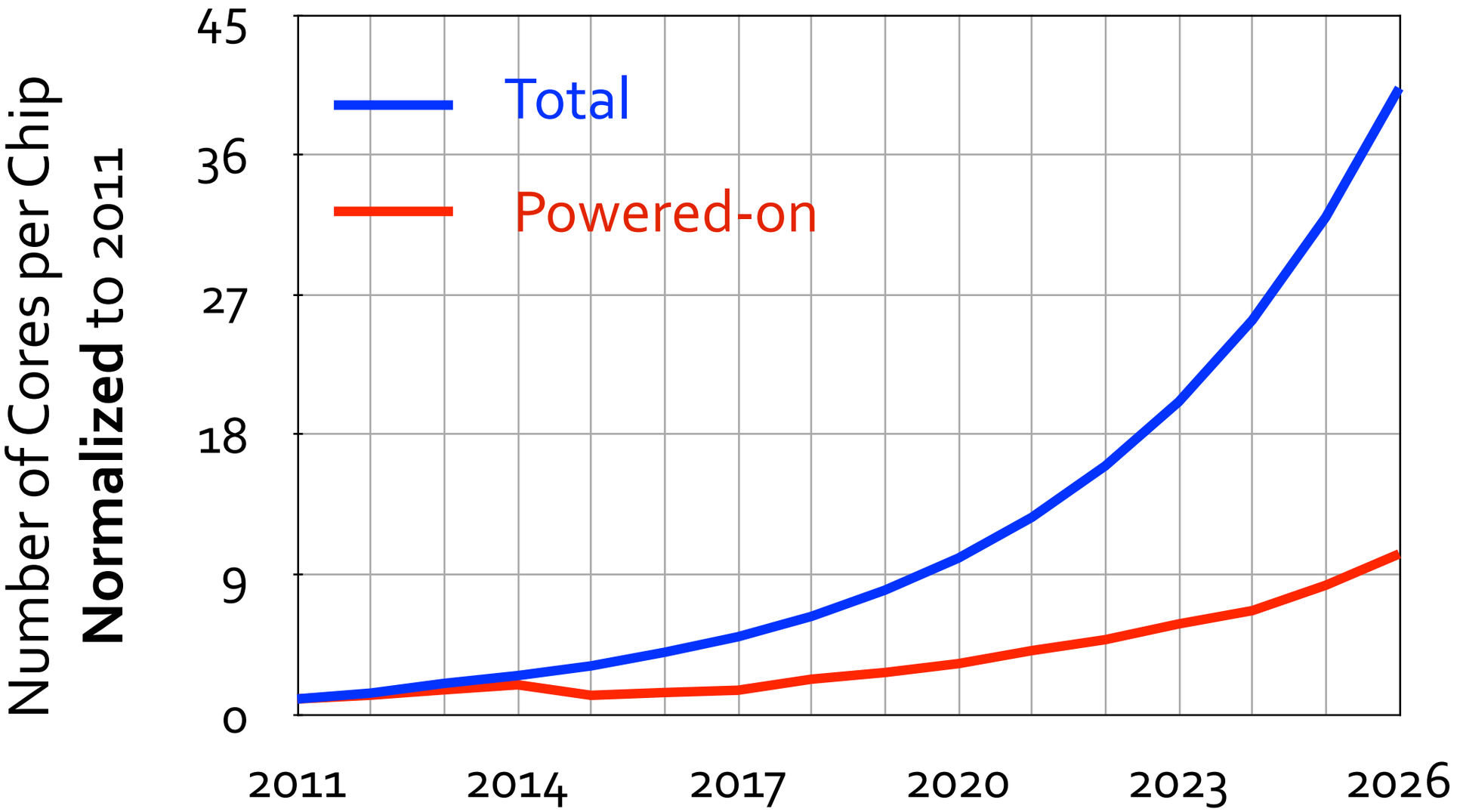
Problem: The Many-Core Power Wall



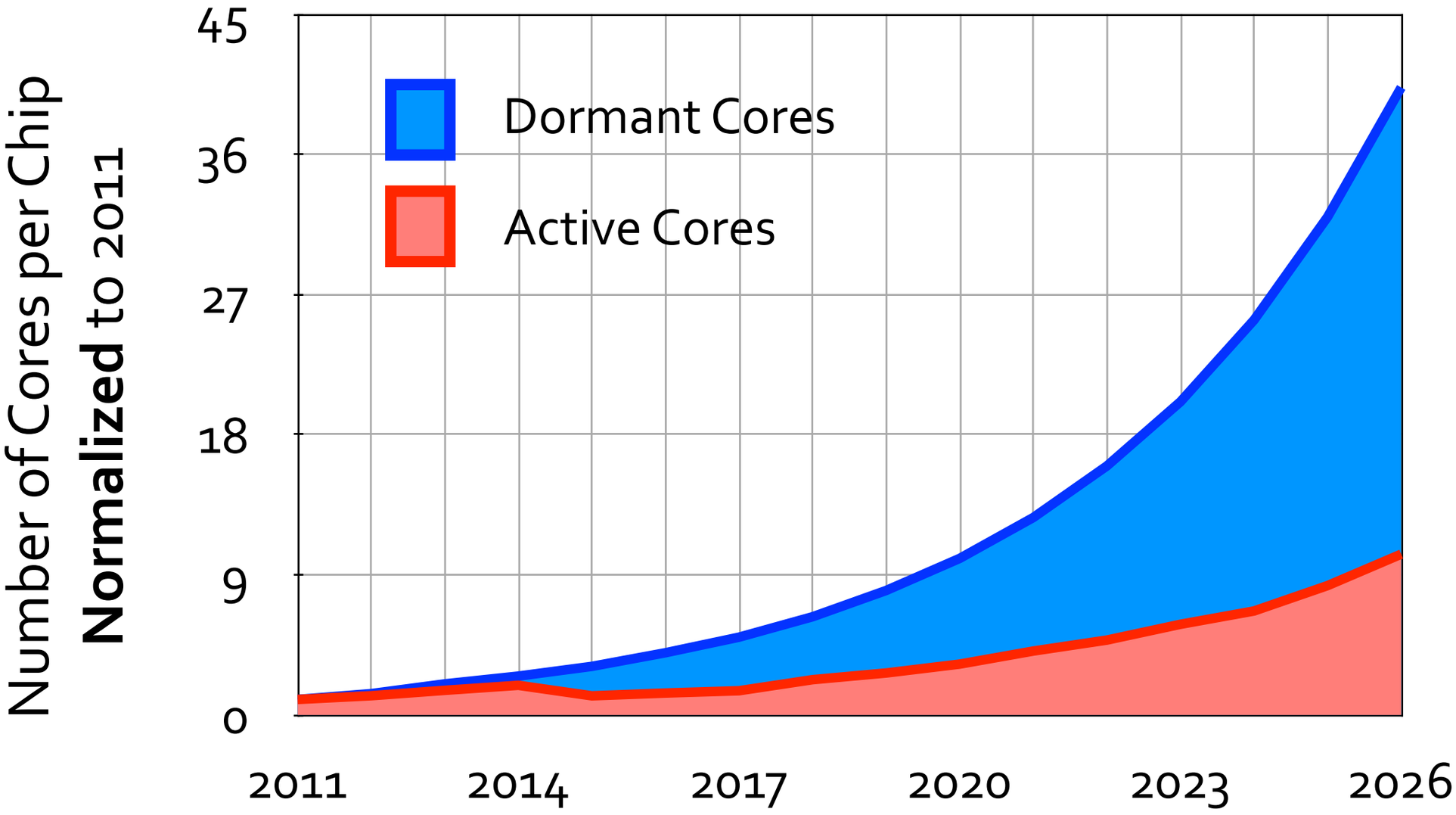
Problem: The Many-Core Power Wall



Problem: The Many-Core Power Wall



Problem: The Many-Core Power Wall



Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

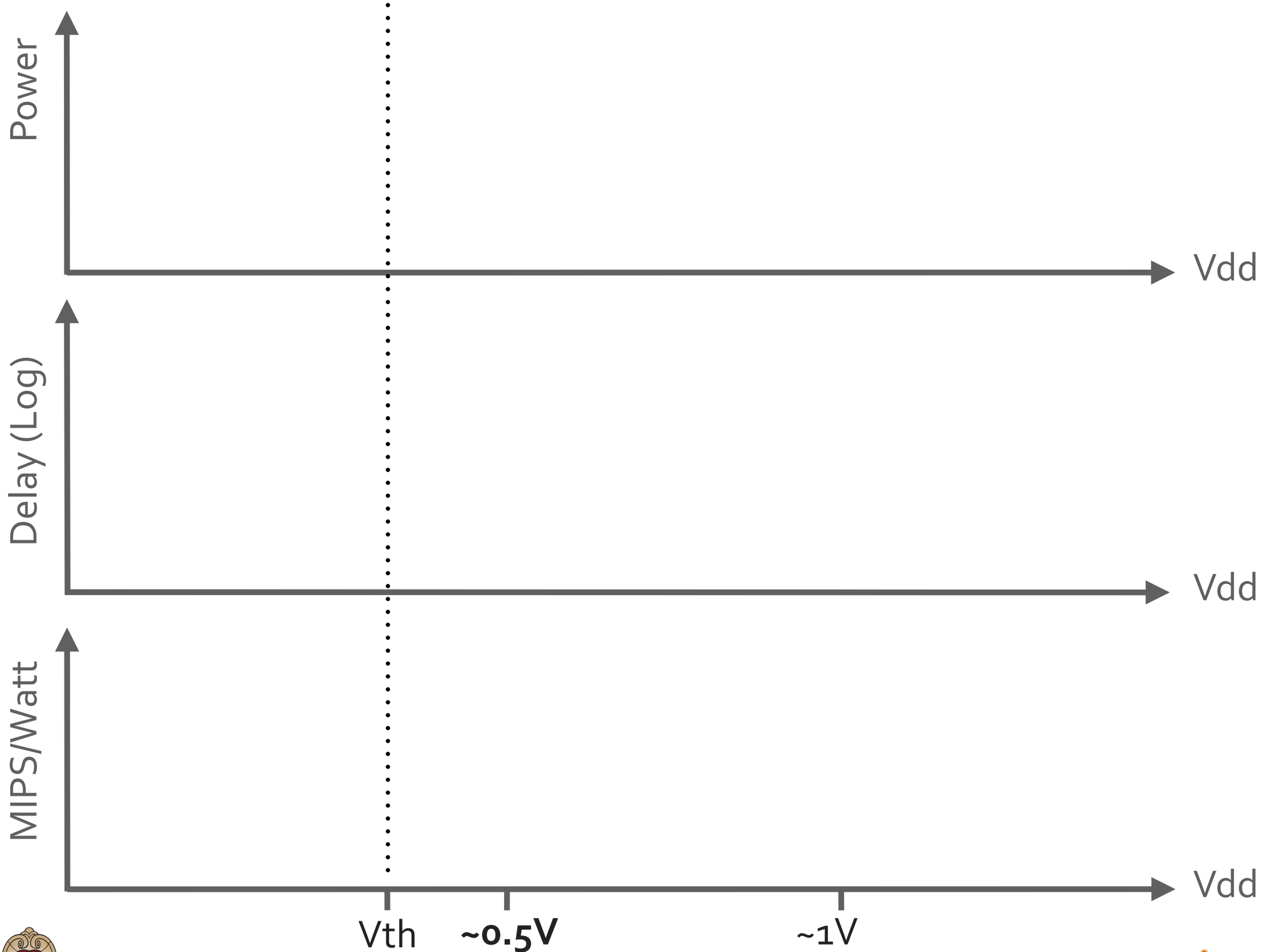
- Vdd remains slightly above Vth

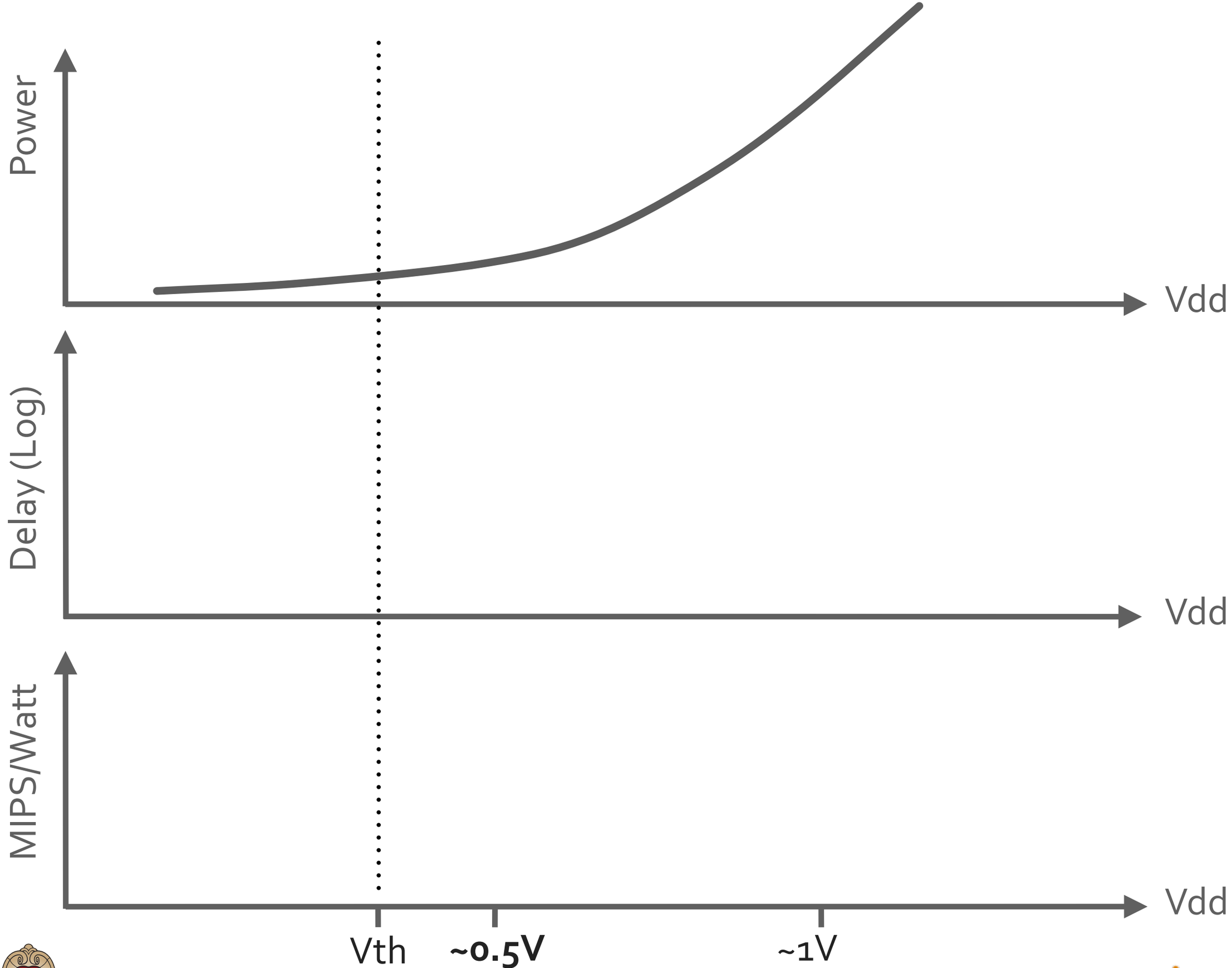
Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

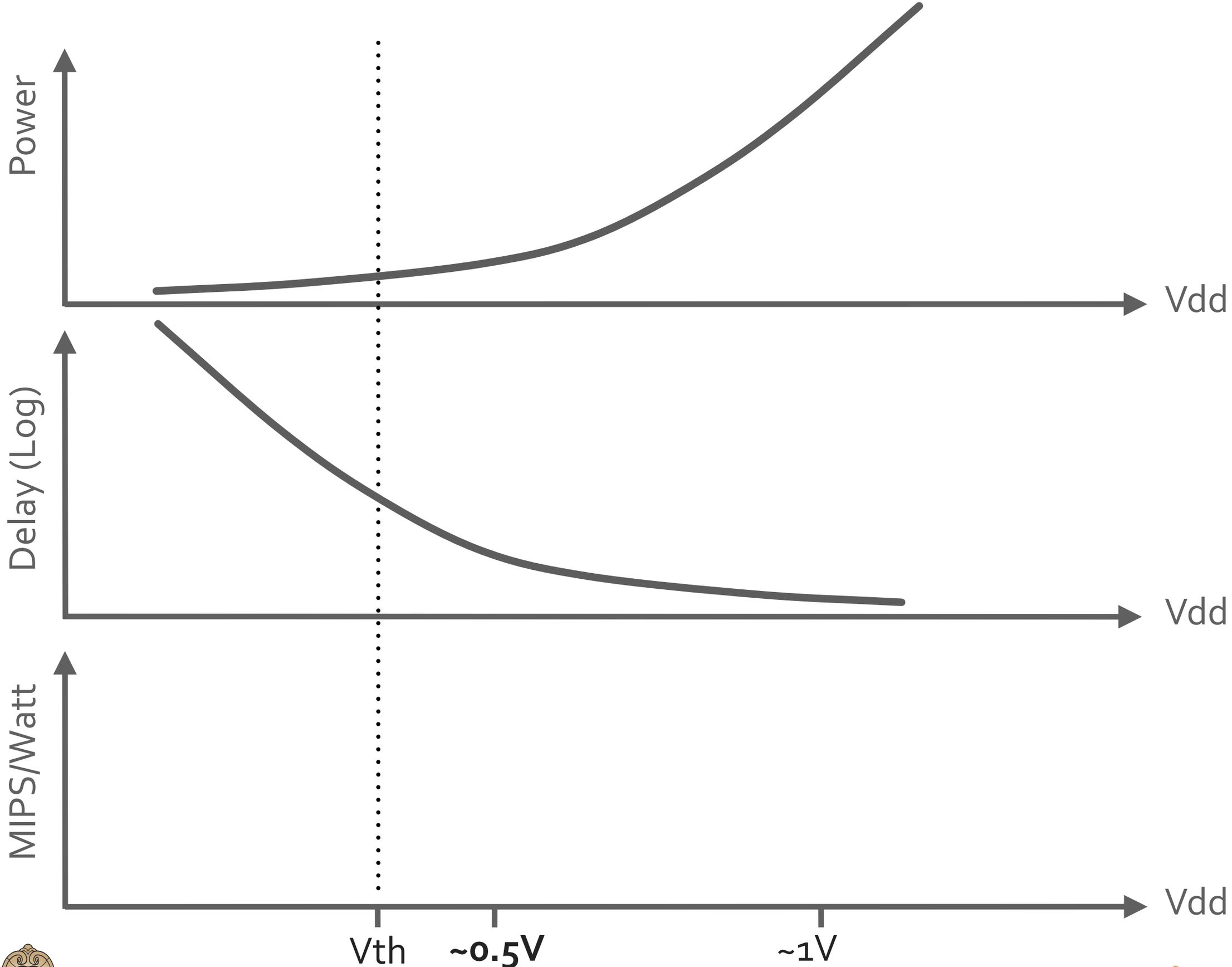
- Vdd remains slightly above Vth
  - Near-threshold Voltage (NTV): ~0.5V
  - Conventional = Super-threshold Voltage (STV): ~1V

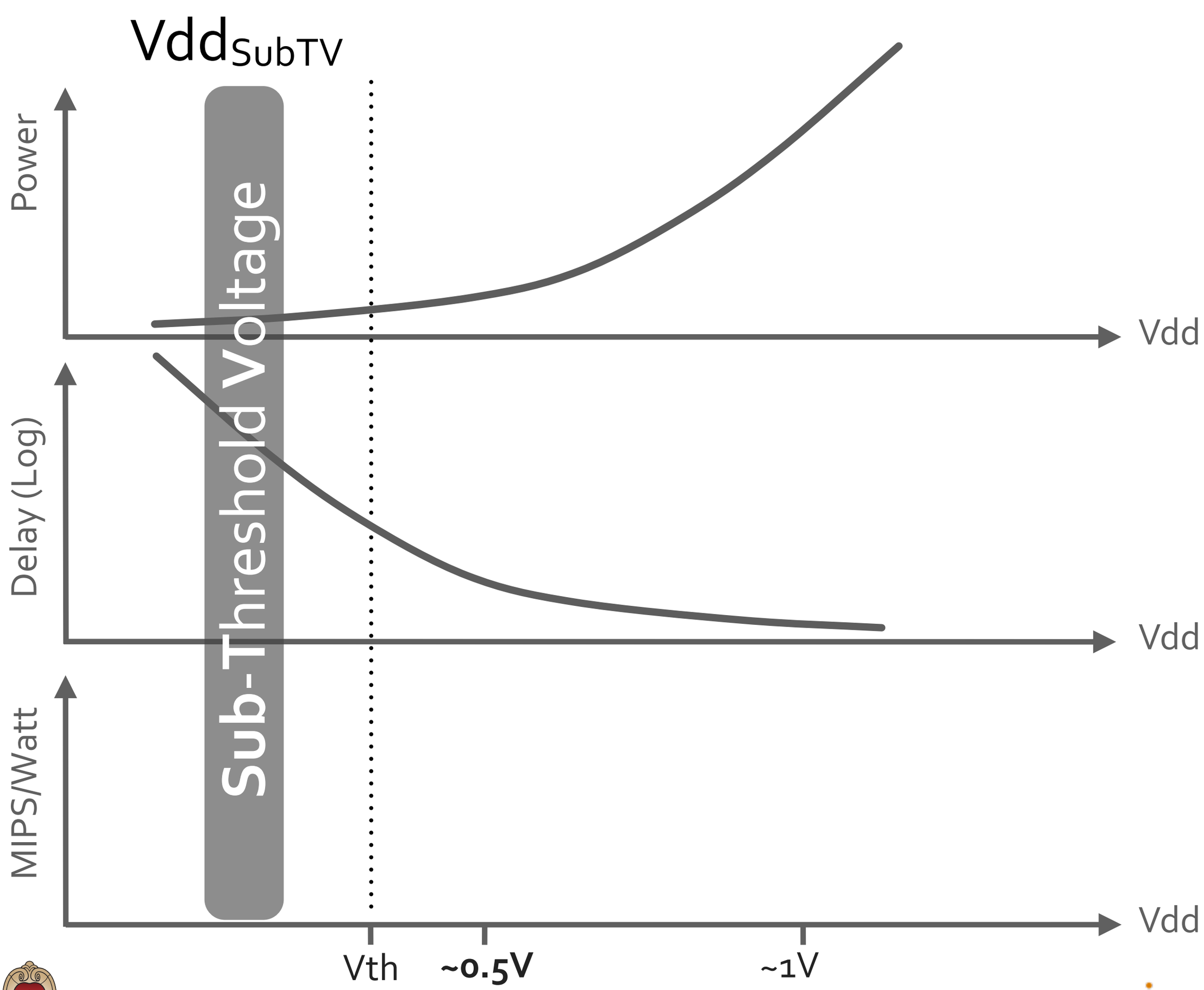
Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

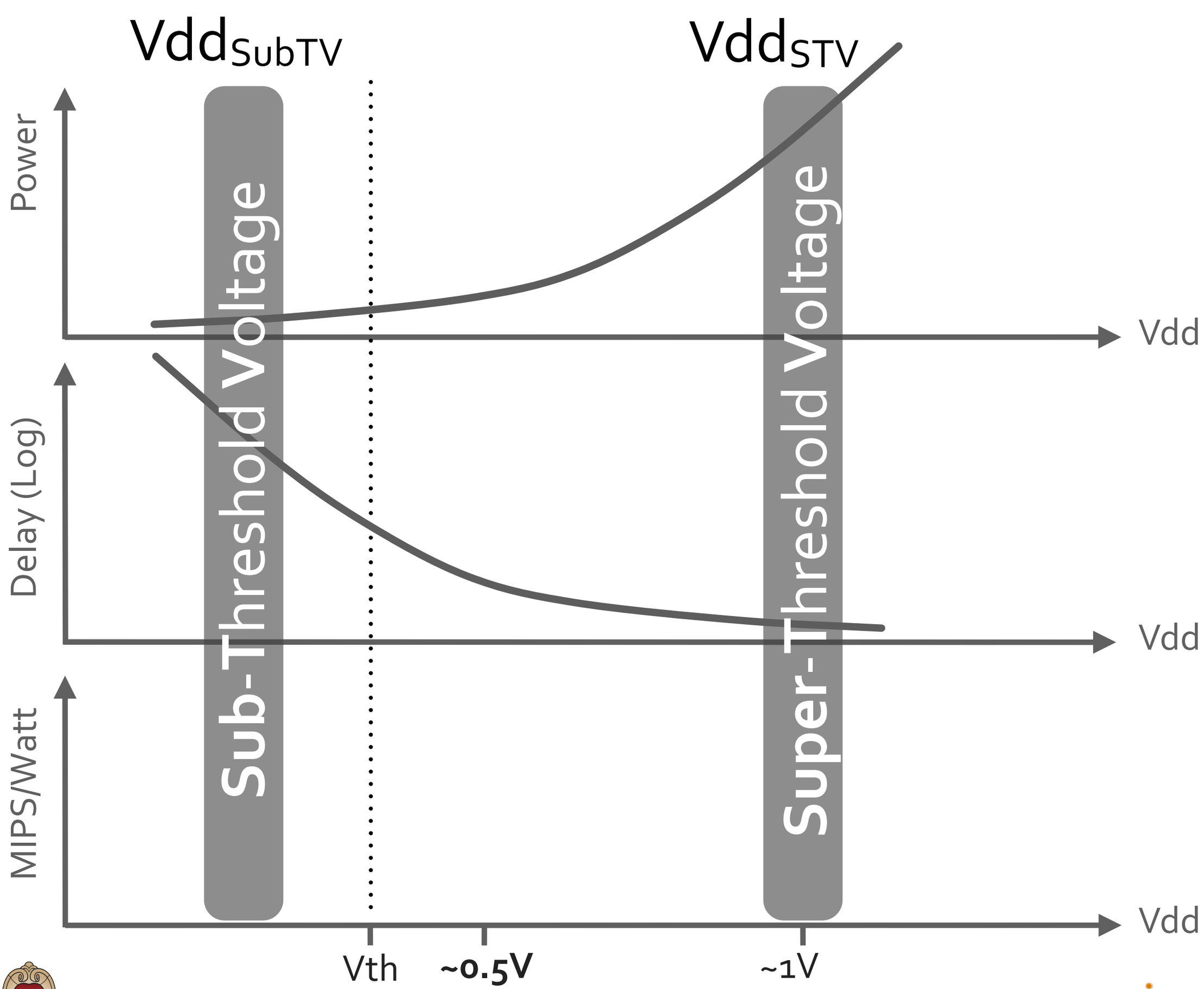
- Vdd remains slightly above Vth
  - Near-threshold Voltage (NTV): ~0.5V
  - Conventional = Super-threshold Voltage (STV): ~1V
- Operation with much higher energy-efficiency
  - P_{DYN} and P_{STA} are super-linear functions of Vdd

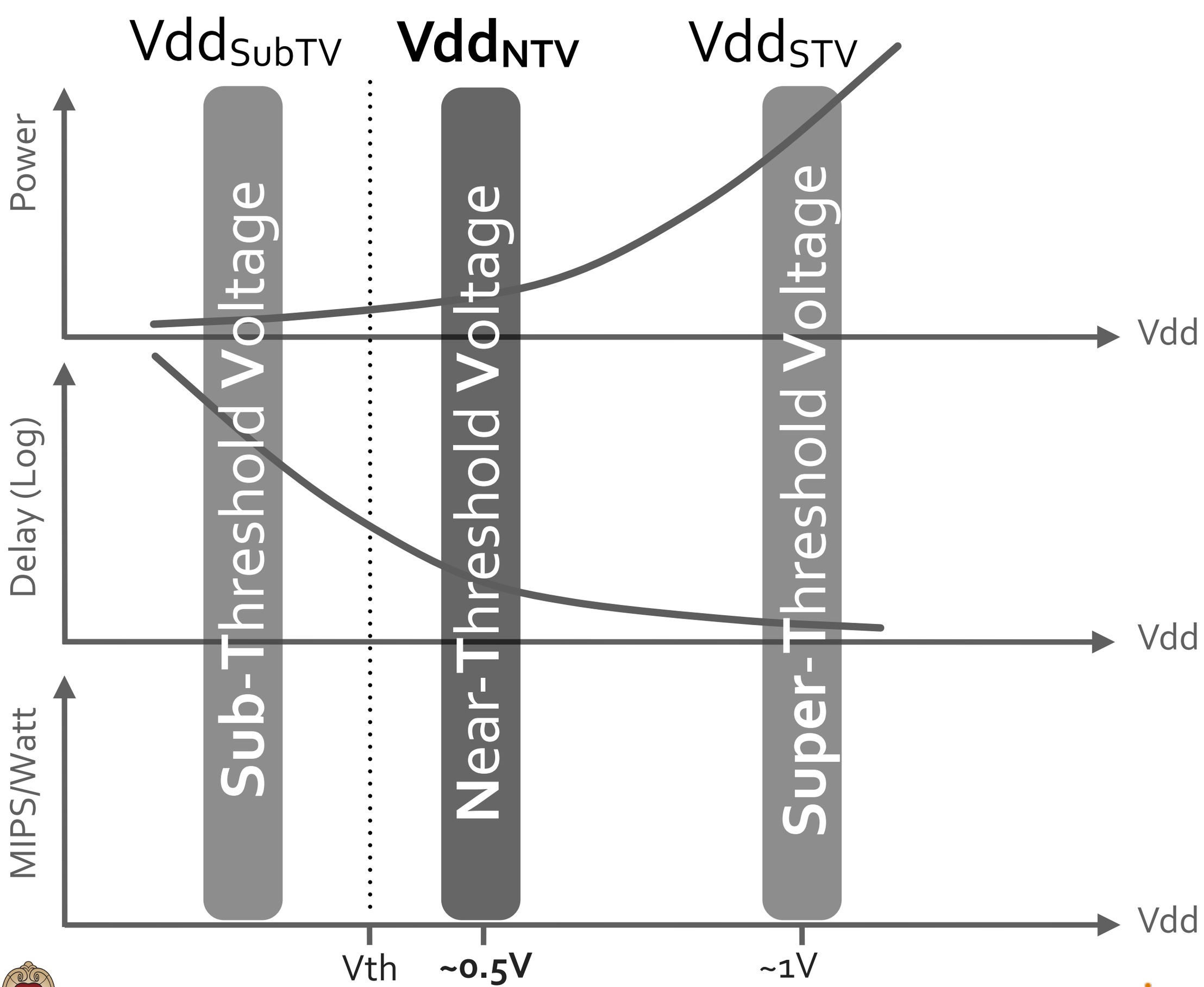

Power
Vdd
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V

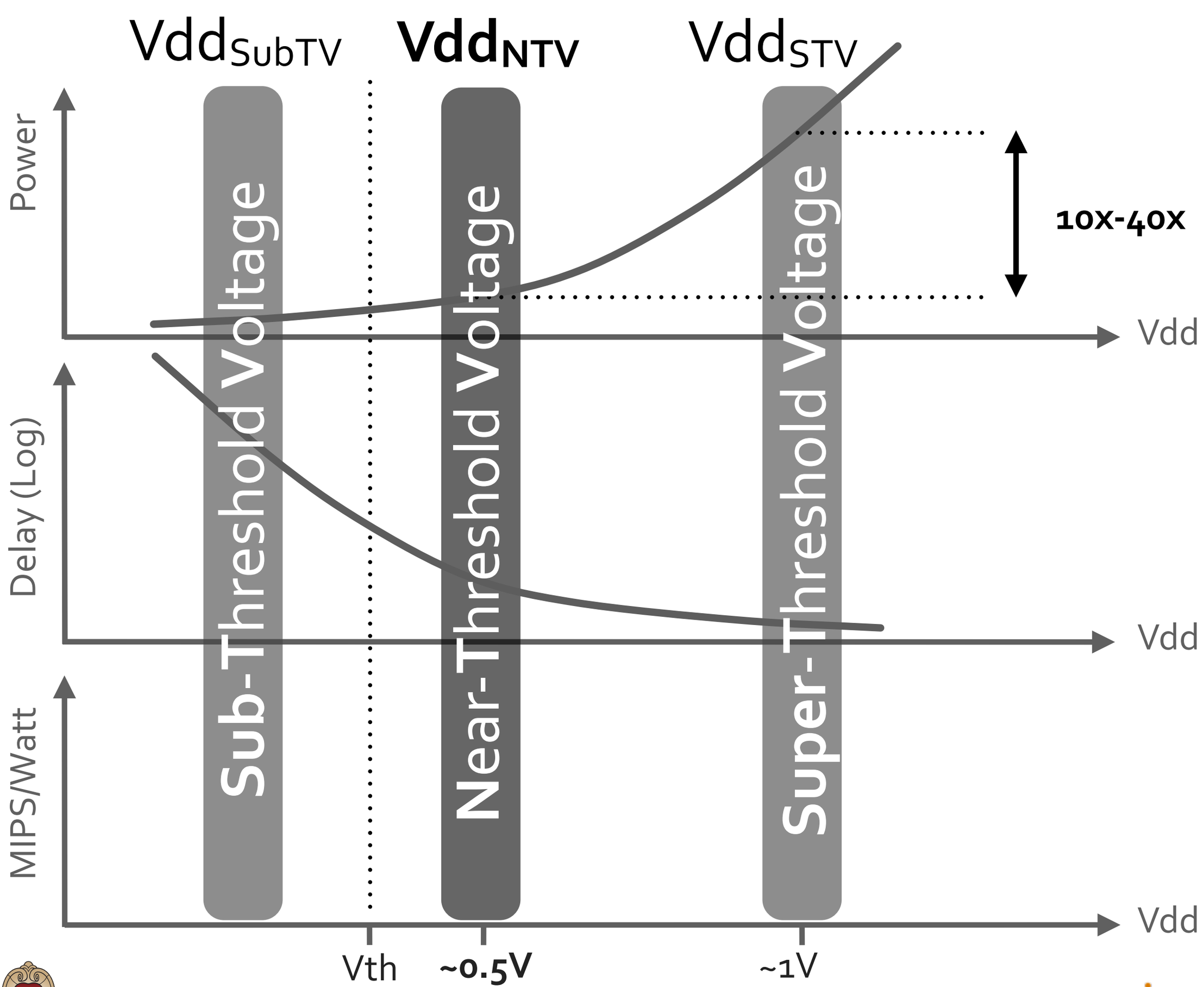

Power
Vdd
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V

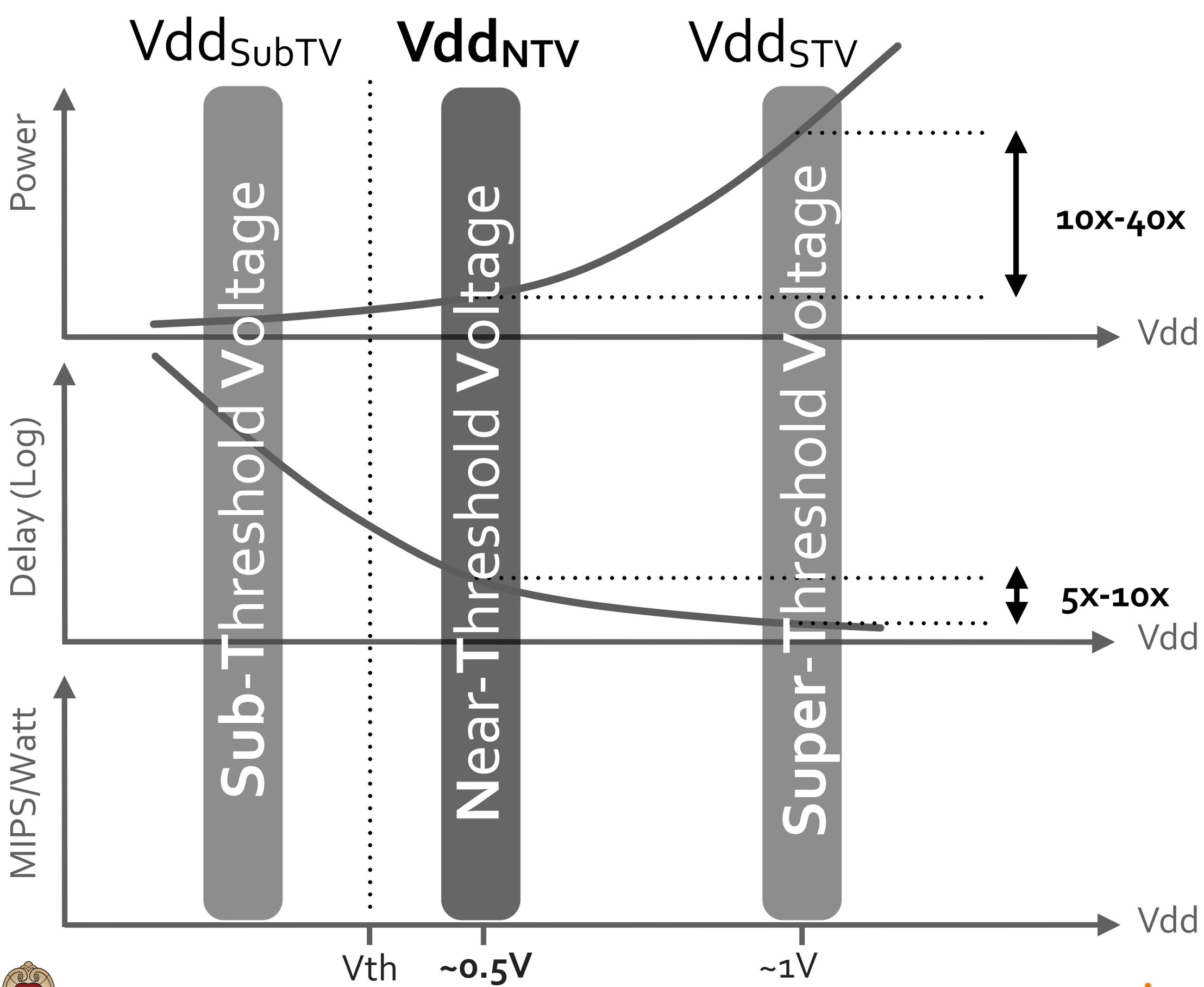

Power
Vdd
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V

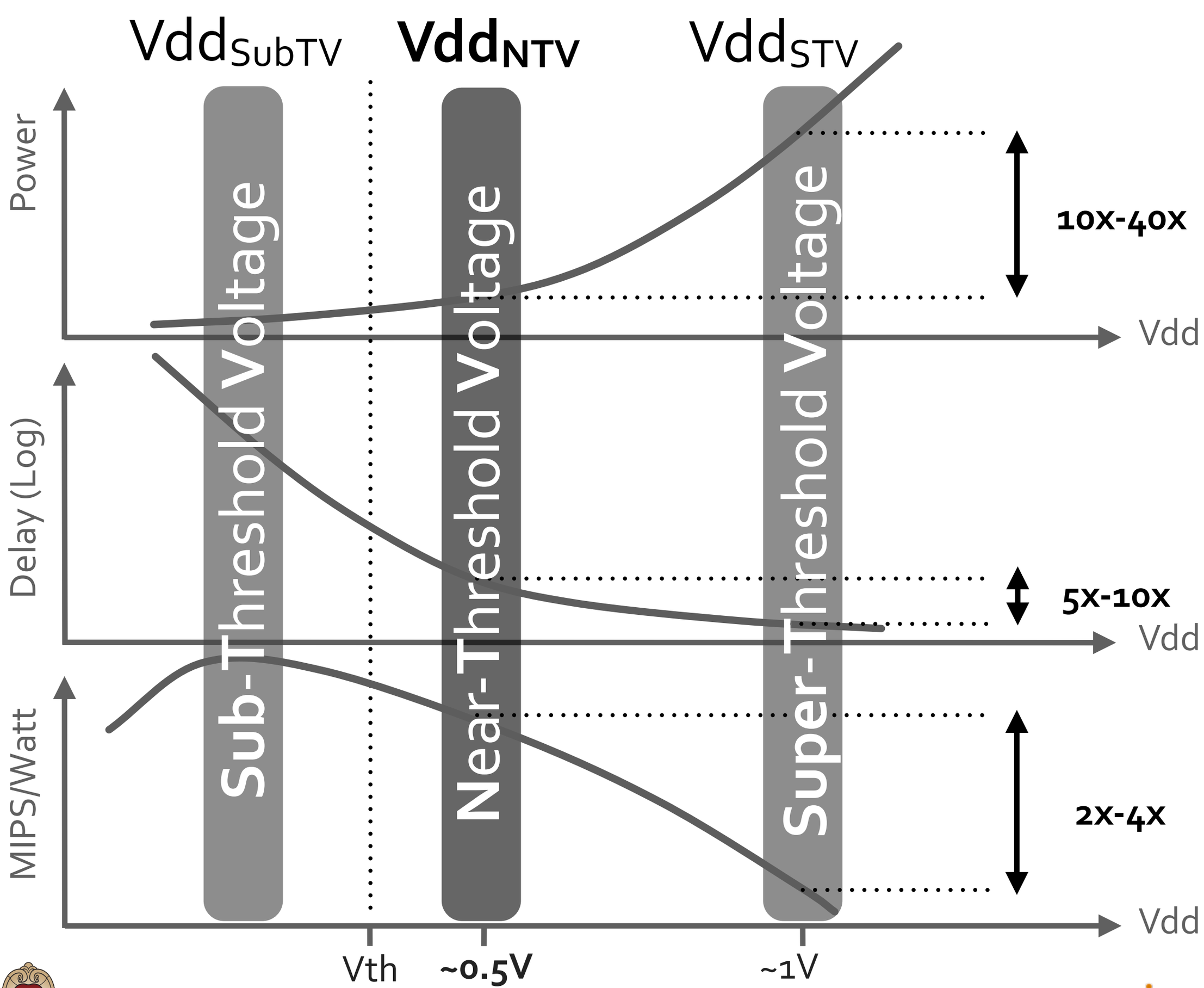

Vdd_{SubTV}
Power
Vdd
Sub-Threshold Voltage
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V

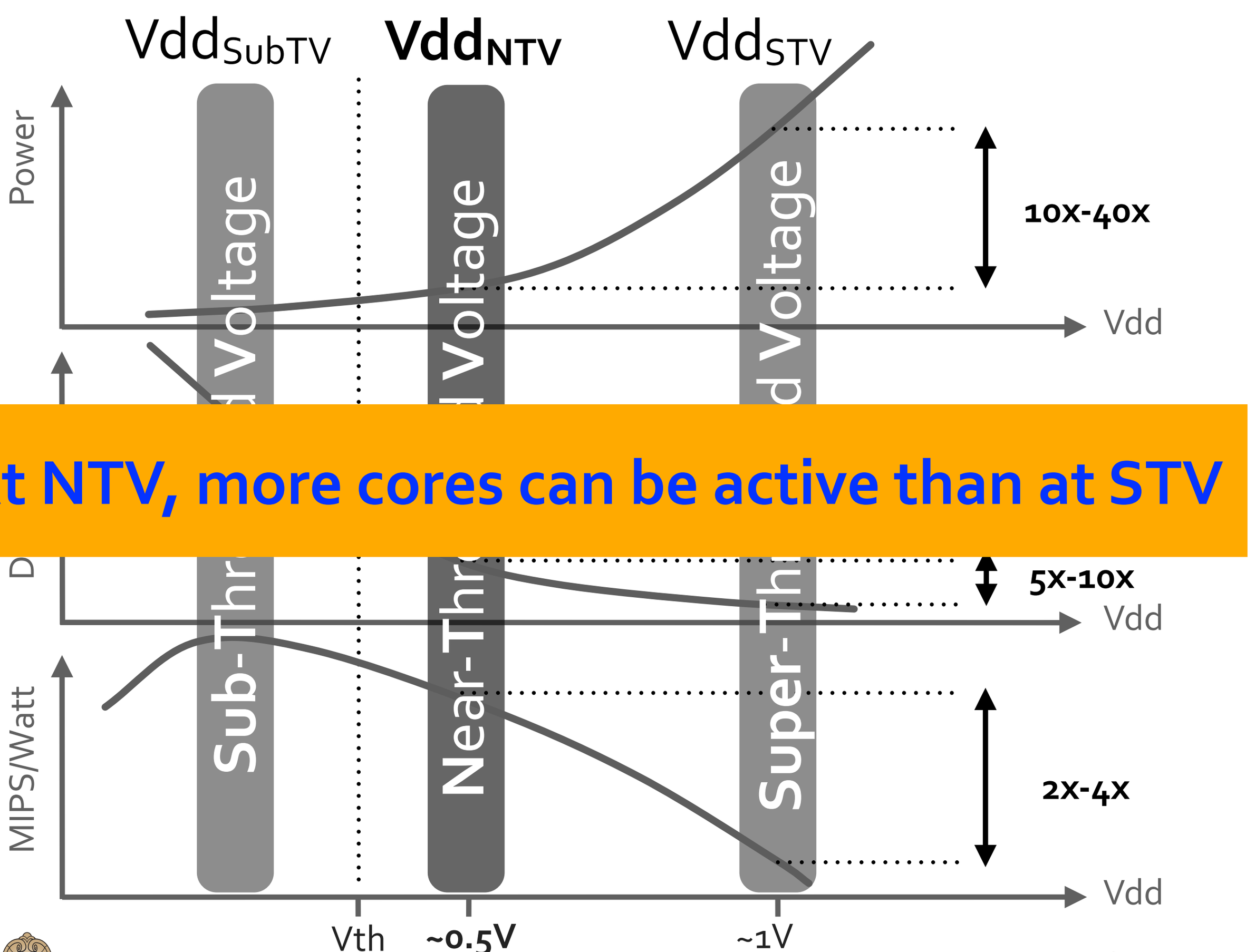

Vdd_SubTV
Vdd_STV
Power
Vdd
Sub-Threshold Voltage
Super-Threshold Voltage
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V


Vdd_SubTV
Vdd_NTV
Vdd_STV
Power
Vdd
Sub-Threshold Voltage
Near-Threshold Voltage
Super-Threshold Voltage
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V


Vdd_SubTV
Vdd_NTV
Vdd_STV
Power
Vdd
10X-40X
Sub-Threshold Voltage
Near-Threshold Voltage
Super-Threshold Voltage
Delay (Log)
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V


Vdd_SubTV
Vdd_NTV
Vdd_STV
Power
Vdd
10X-40X
Sub-Threshold Voltage
Near-Threshold Voltage
Super-Threshold Voltage
Delay (Log)
5X-10X
Vdd
MIPS/Watt
Vdd
Vth
~0.5V
~1V


Vdd_SubTV
Vdd_NTV
Vdd_STV
Power
Vdd
10X-40X
Delay (Log)
Vdd
5X-10X
MIPS/Watt
Vdd
2X-4X
Sub-Threshold Voltage
Near-Threshold Voltage
Super-Threshold Voltage
Vth
~0.5V
~1V



At NTV, more cores can be active than at STV

Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

- Key NTV barrier : Increased sensitivity to parametric variation

Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

- Key NTV barrier : Increased sensitivity to parametric variation
- How to cope with increased sensitivity to variations at NTV?

Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

- Key NTV barrier : Increased sensitivity to parametric variation
- How to cope with increased sensitivity to variations at NTV?
  - Holistic approach involving all levels of system stack

Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

- Key NTV barrier : Increased sensitivity to parametric variation
- How to cope with increased sensitivity to variations at NTV?
  - Holistic approach involving all levels of system stack
  - First step: Characterize the impact of variation


Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

- Key NTV barrier : Increased sensitivity to parametric variation
- How to cope with increased sensitivity to variations at NTV?
  - Holistic approach involving all levels of system stack
  - First step: Characterize the impact of variation
    - **Contribution: VARIUS-NTV**


Pushing Back the Many-Core Power Wall: Near-threshold Voltage Operation

- Key NTV barrier : Increased sensitivity to parametric variation
- How to cope with increased sensitivity to variations at NTV?
  - Holistic approach involving all levels of system stack
  - First step: Characterize the impact of variation
    - **Contribution: VARIUS-NTV**
    - A (μ)architectural model of parametric variation for NTV


Parametric Variation: Basics

Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff

Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?

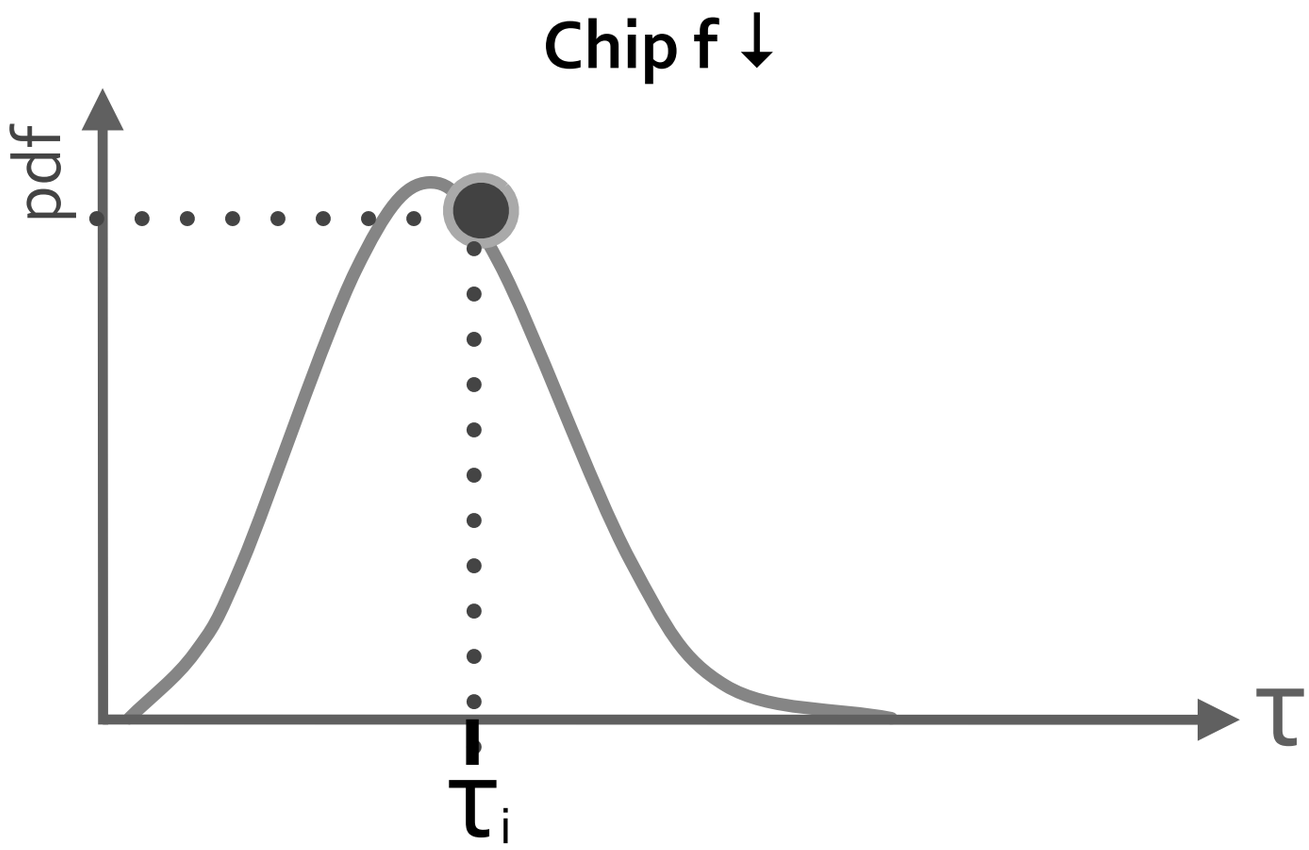
Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?

Chip f ↓

Parametric Variation: Basics

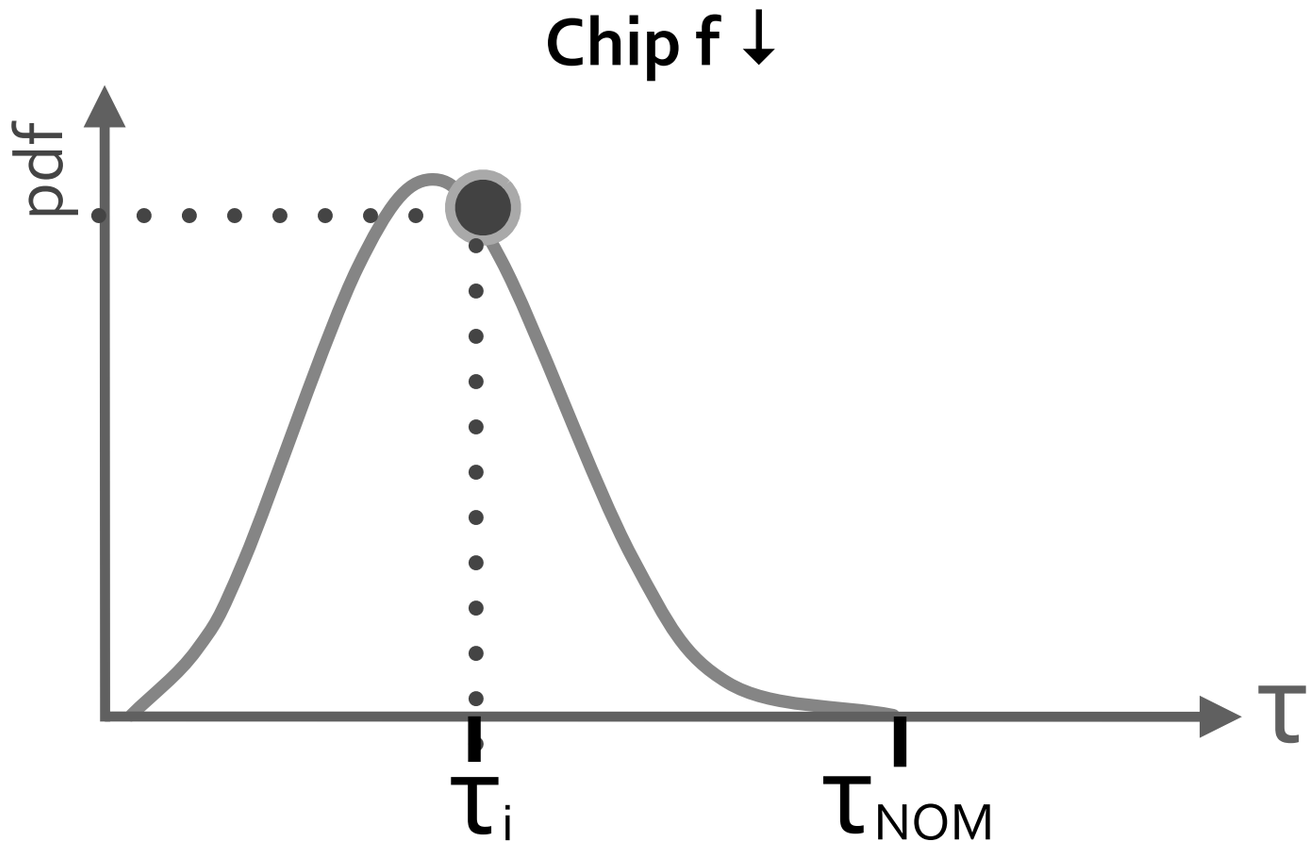
- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?



- Probability of a path with $\tau = \tau_i$ being exercised

Parametric Variation: Basics

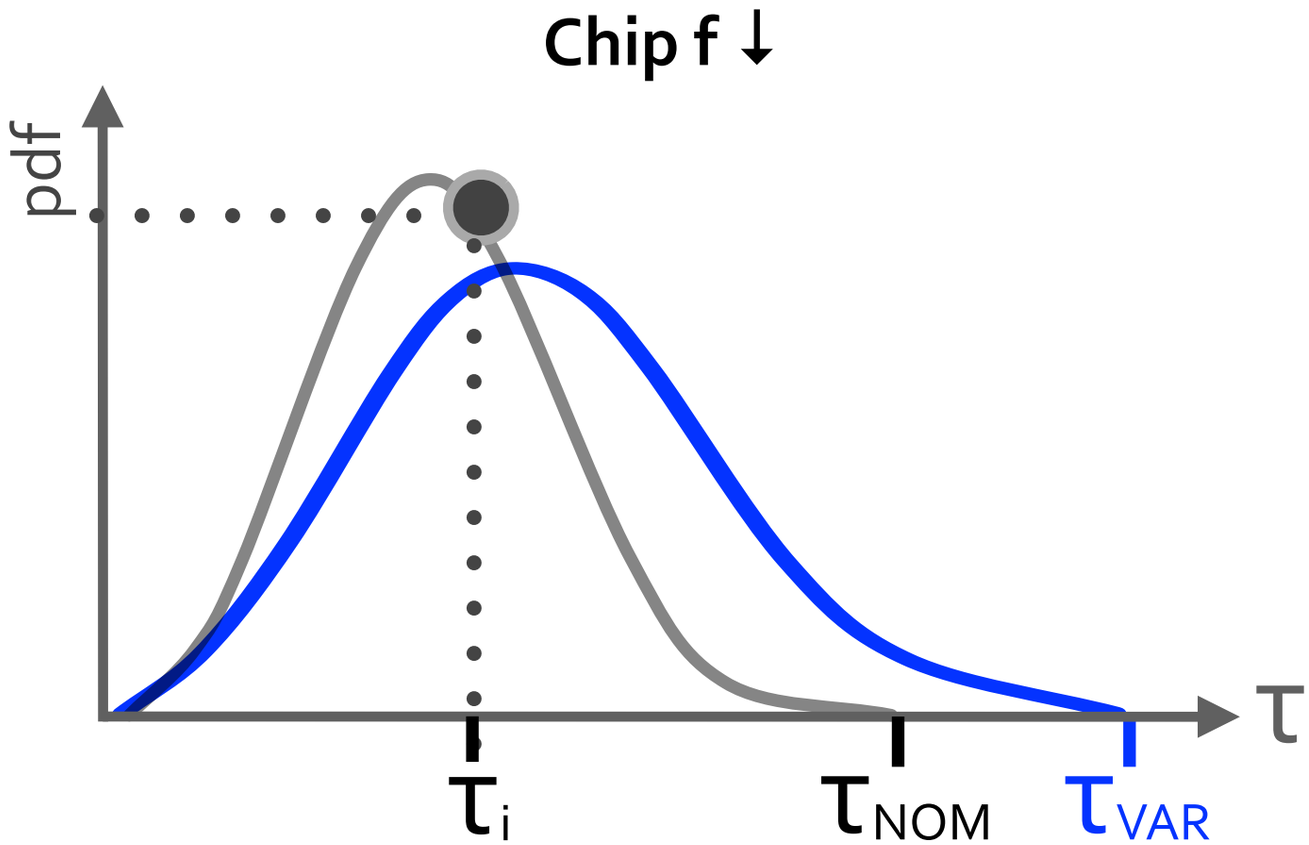
- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?



- Probability of a path with $\tau = \tau_i$ being exercised

Parametric Variation: Basics

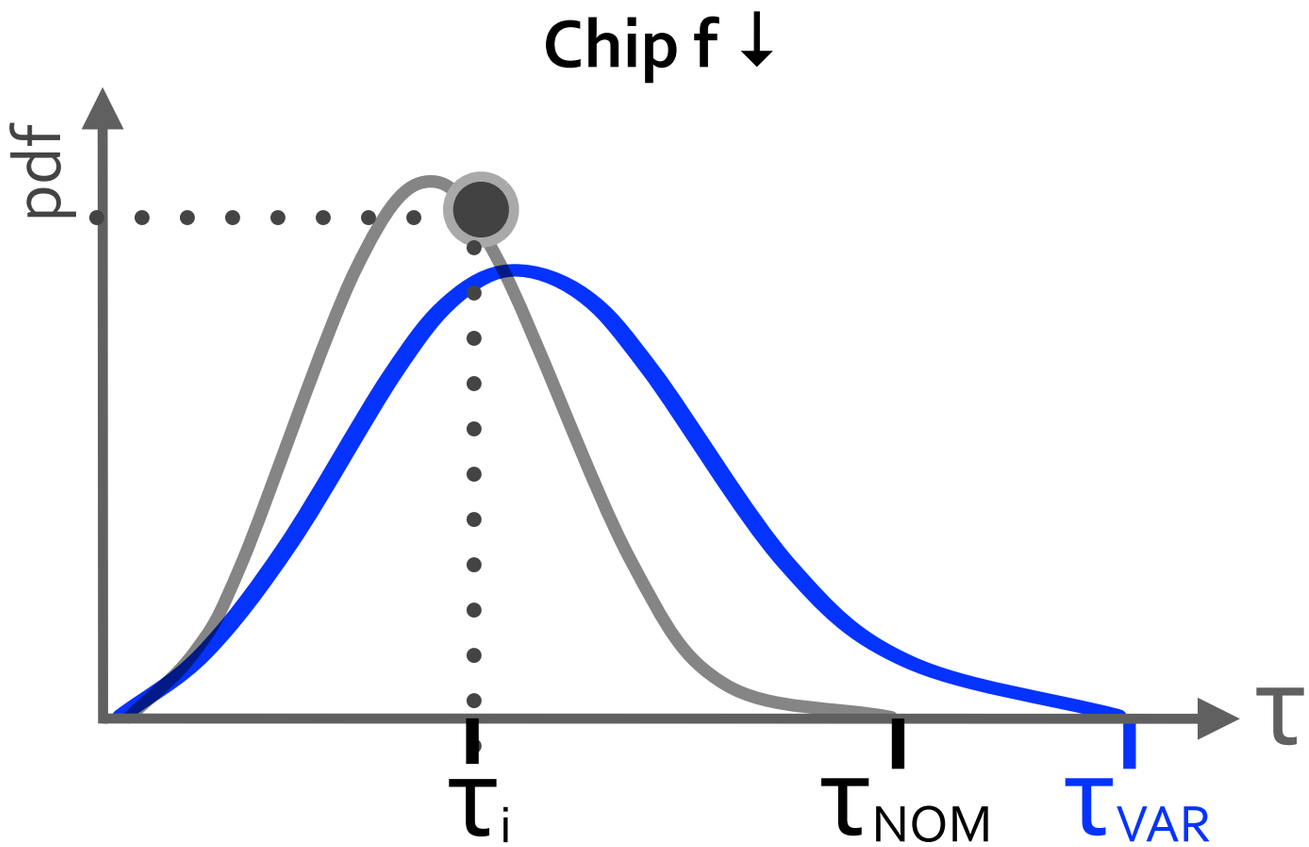
- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?



- Probability of a path with $\tau = \tau_i$ being exercised

Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?

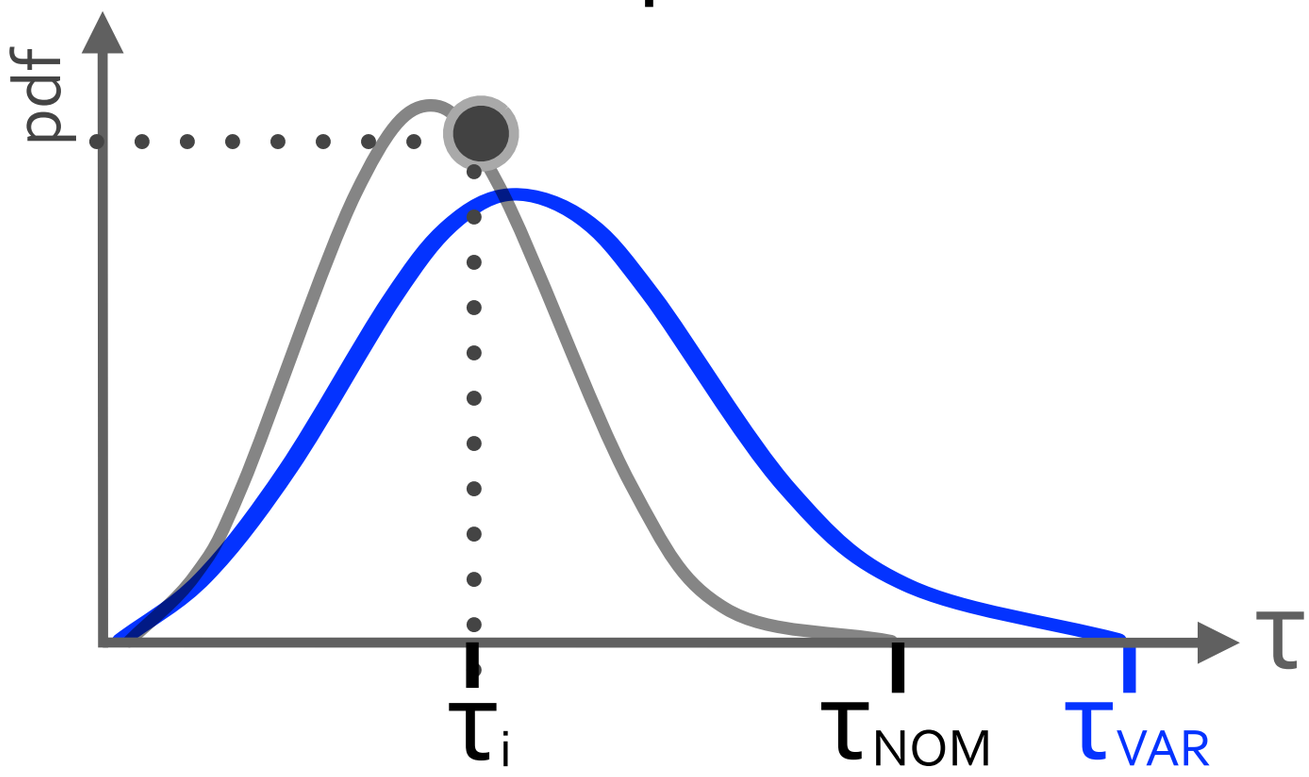


- Probability of a path with $\tau = \tau_i$ being exercised

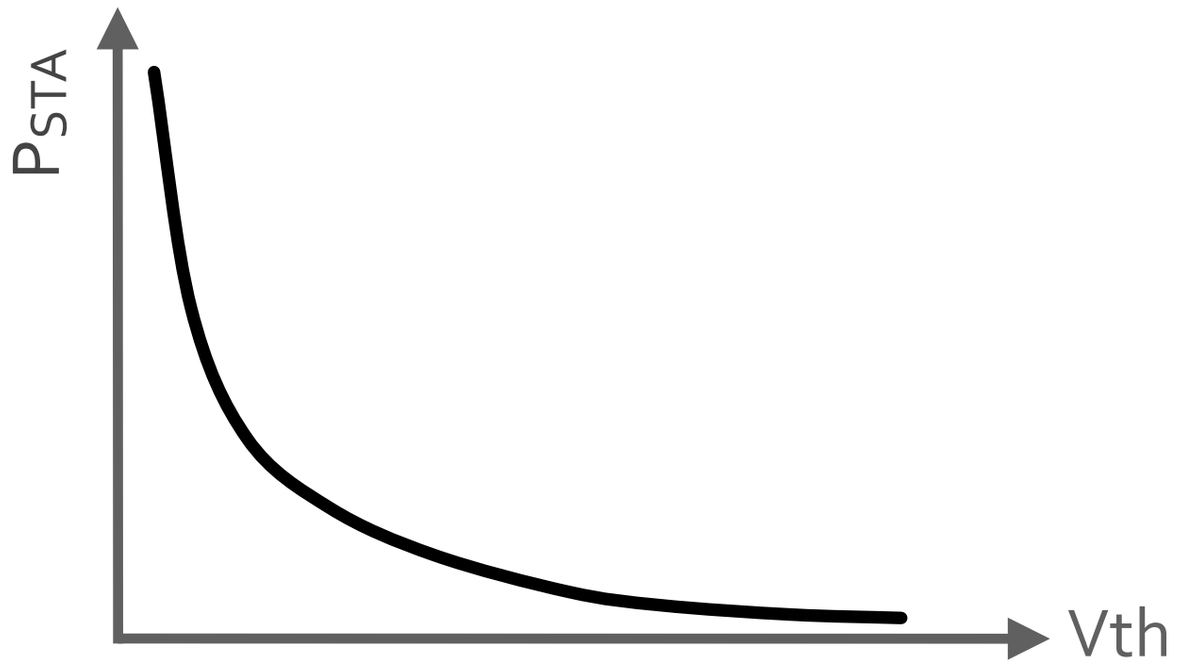
Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?

Chip f ↓



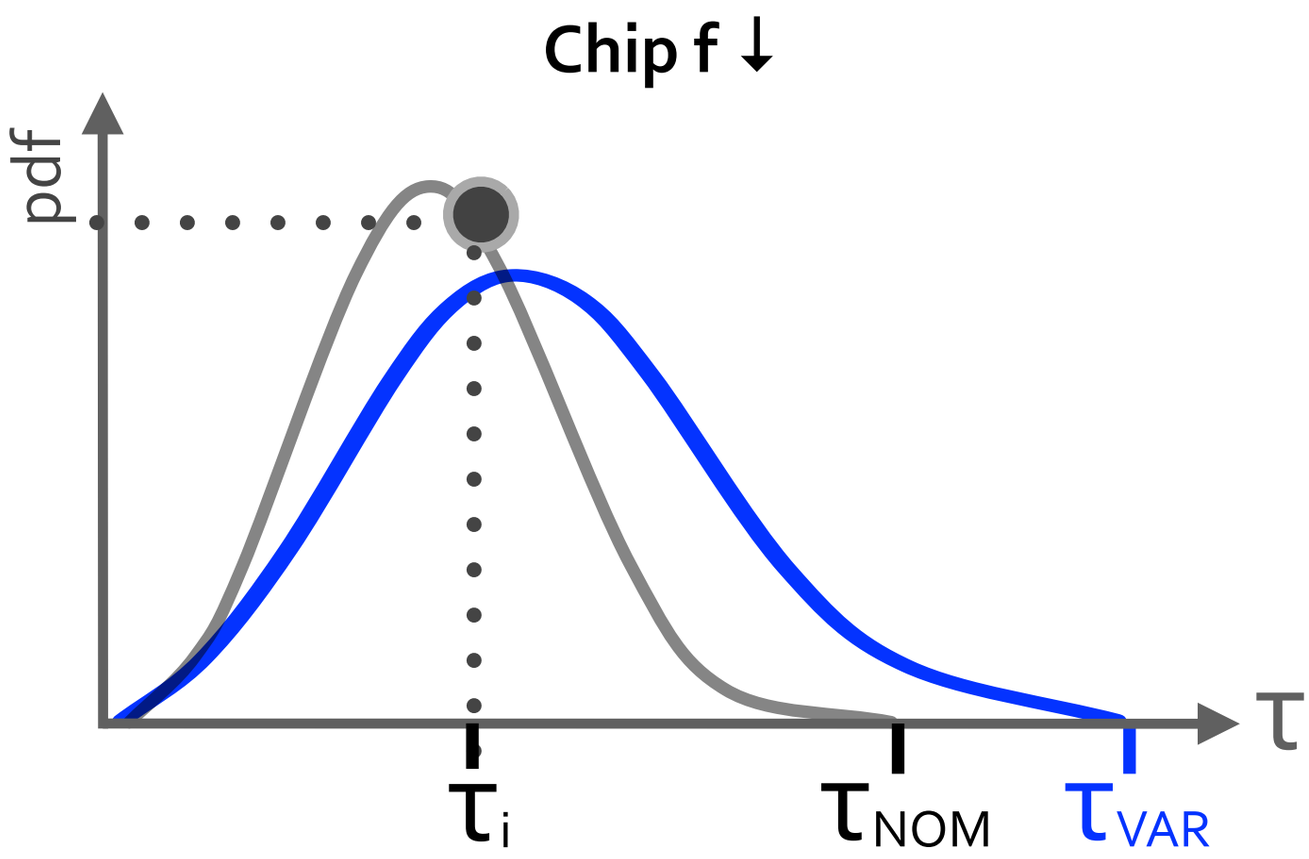
Chip P_{STA} ↑

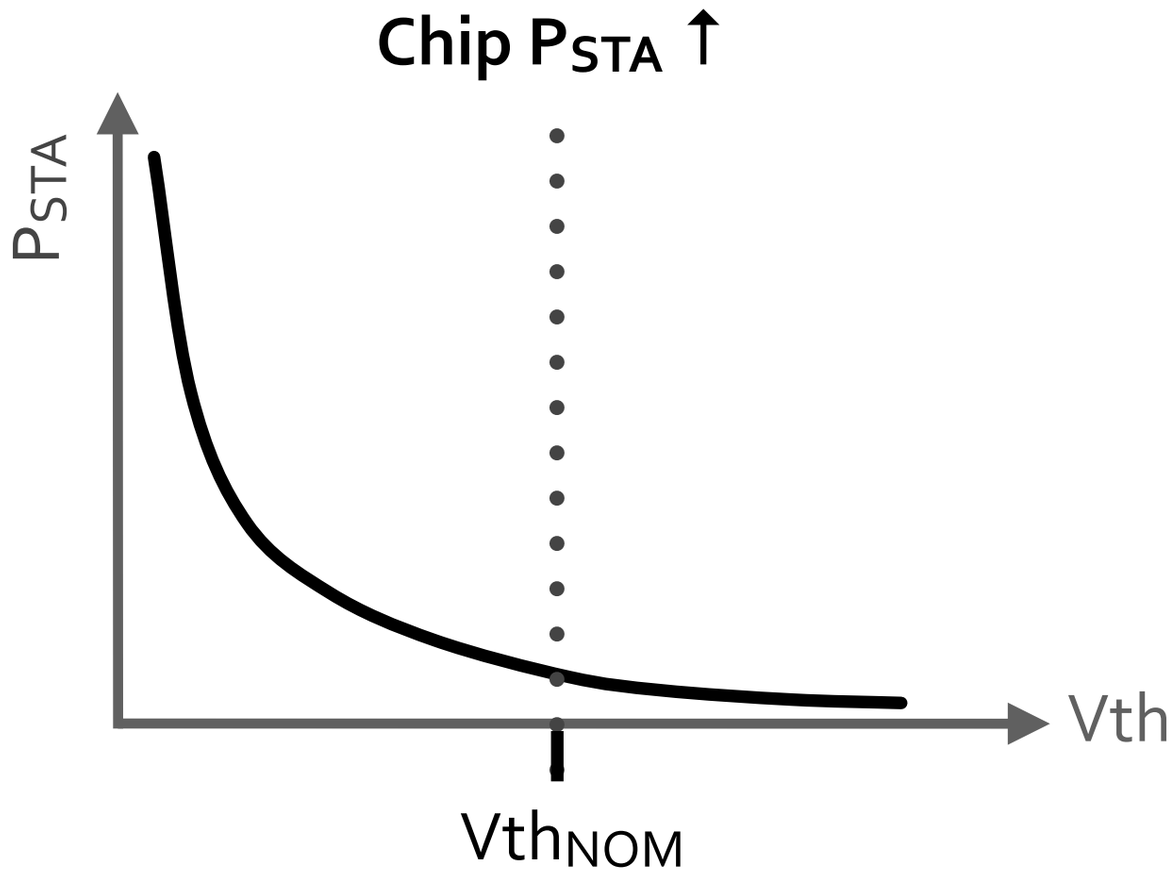


- Probability of a path with $\tau = \tau_i$ being exercised

Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?

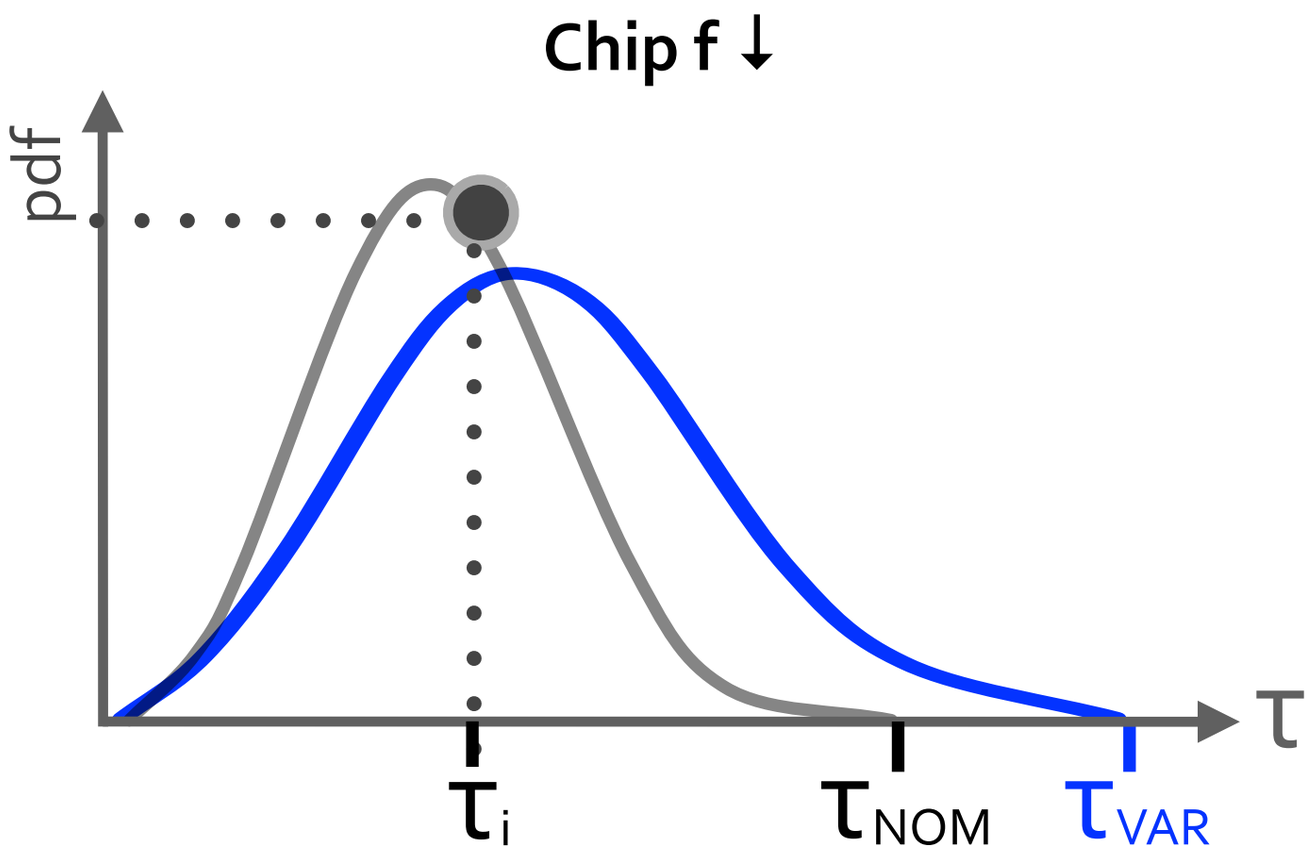


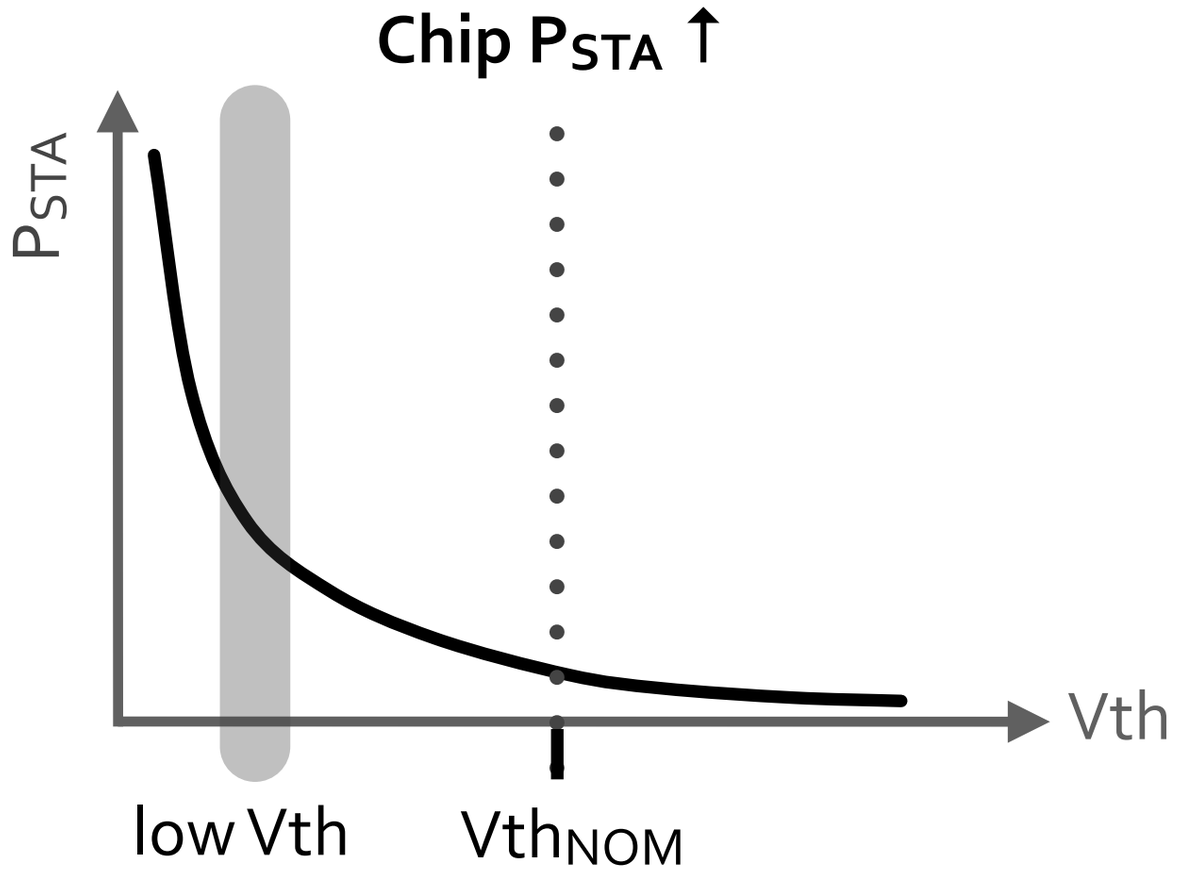


- Probability of a path with $\tau = \tau_i$ being exercised

Parametric Variation: Basics

- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?

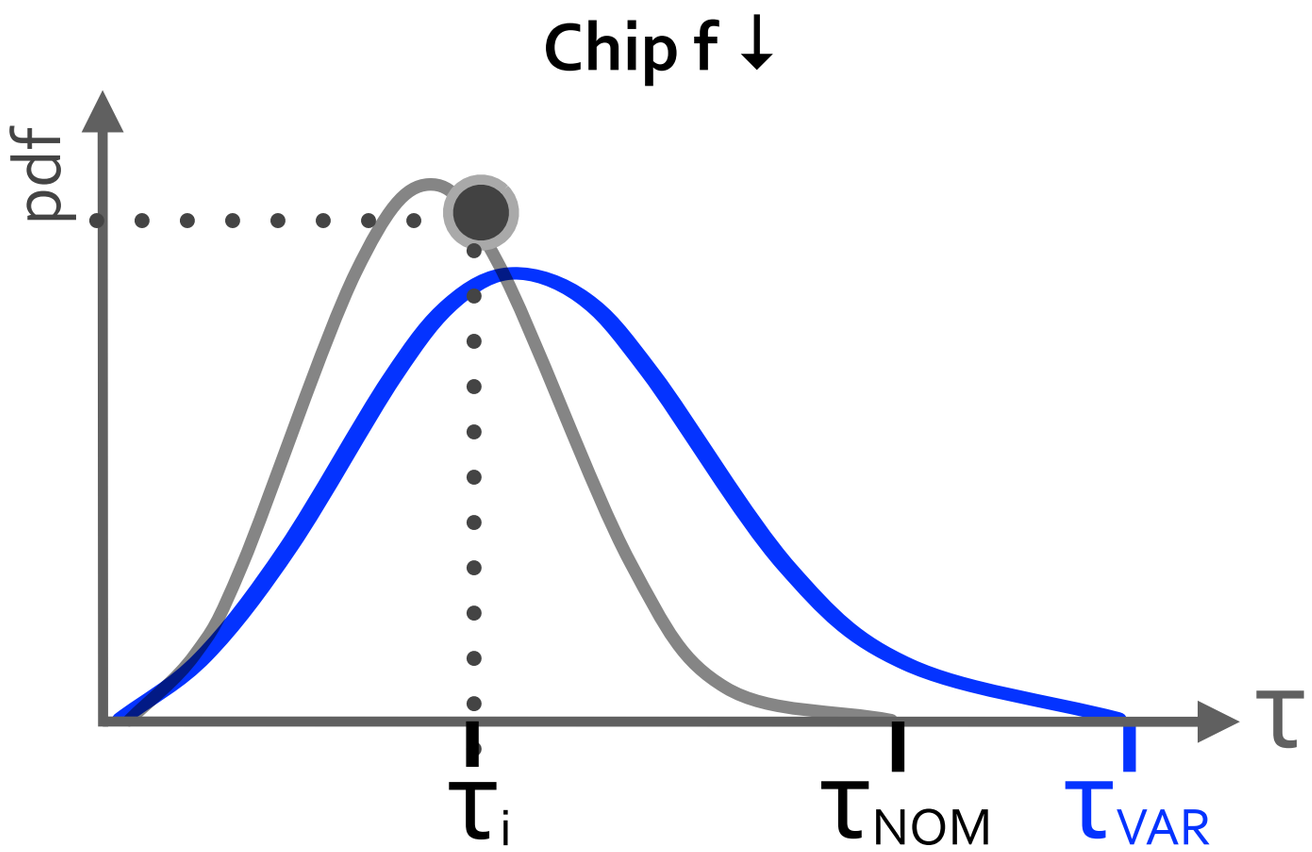


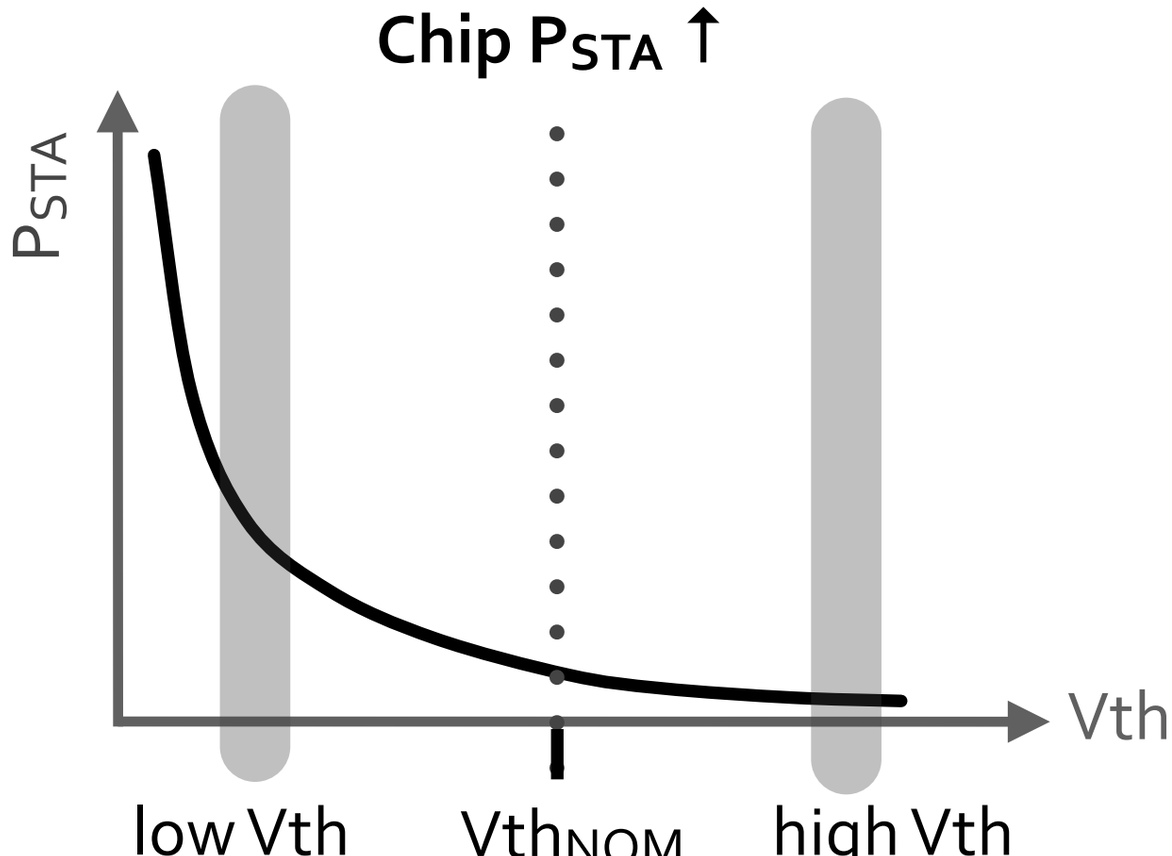


- Probability of a path with $\tau = \tau_i$ being exercised

Parametric Variation: Basics

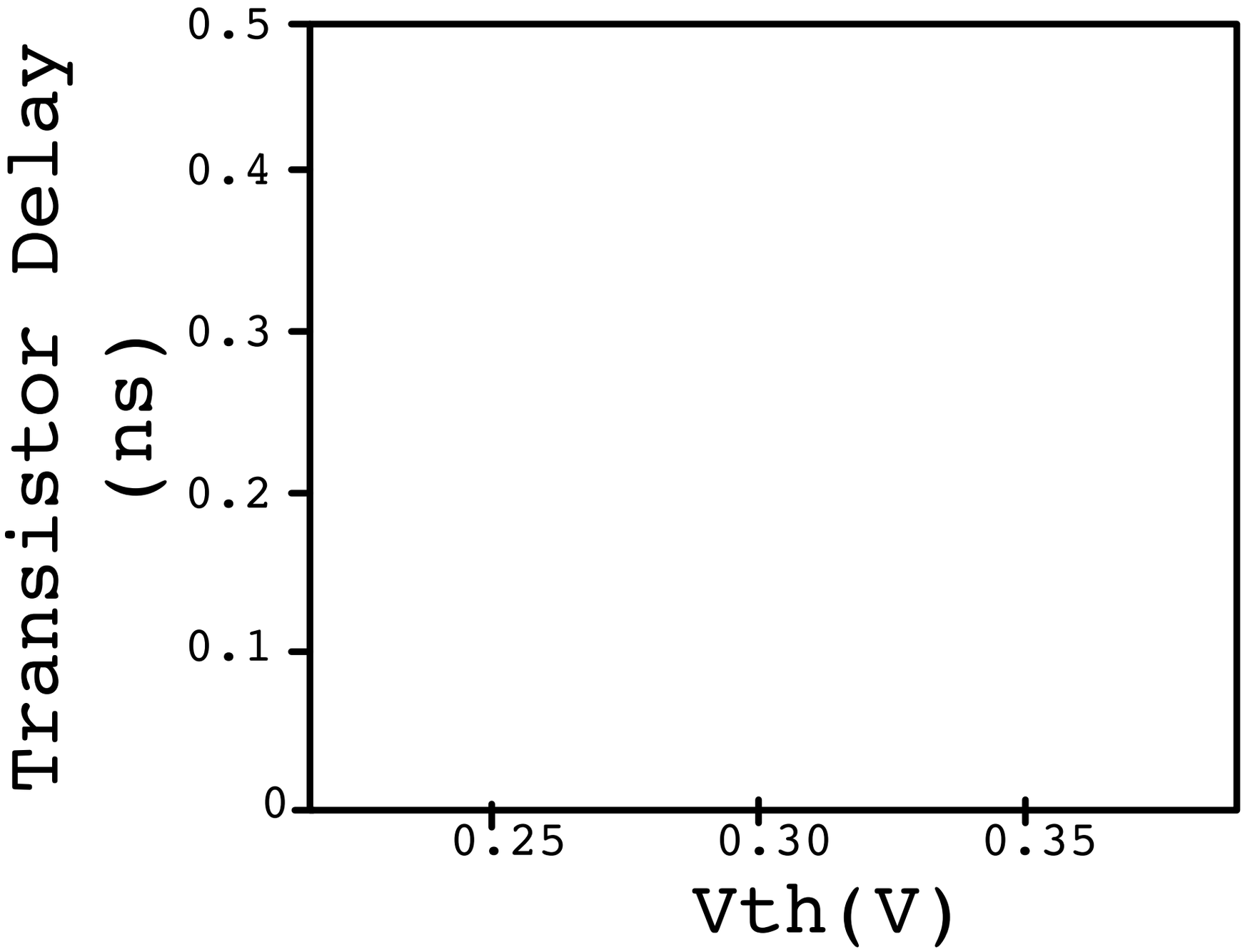
- Deviation of device parameters from nominal: Vth, Leff
- **Impact of variation**?



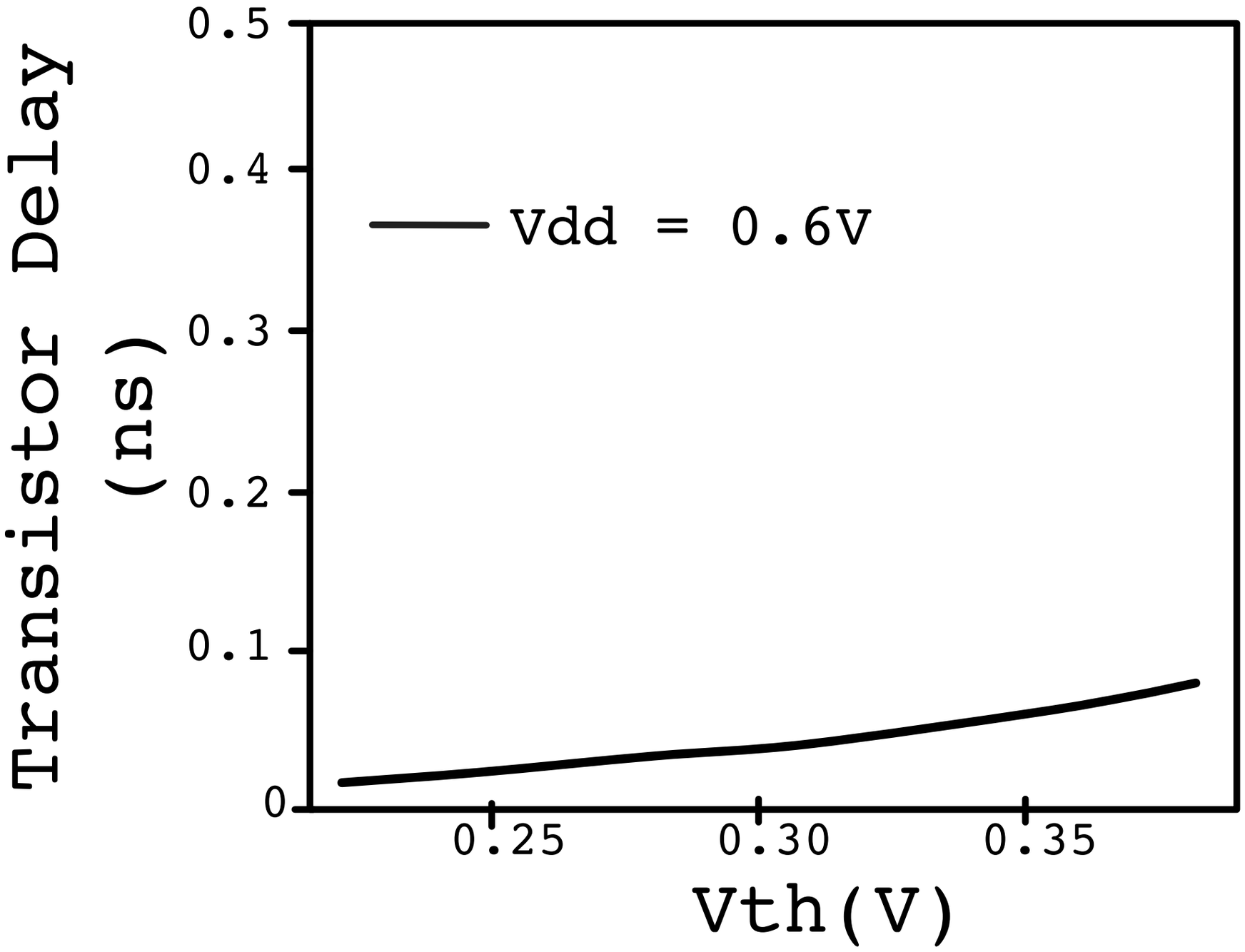


- Probability of a path with $\tau = \tau_i$ being exercised

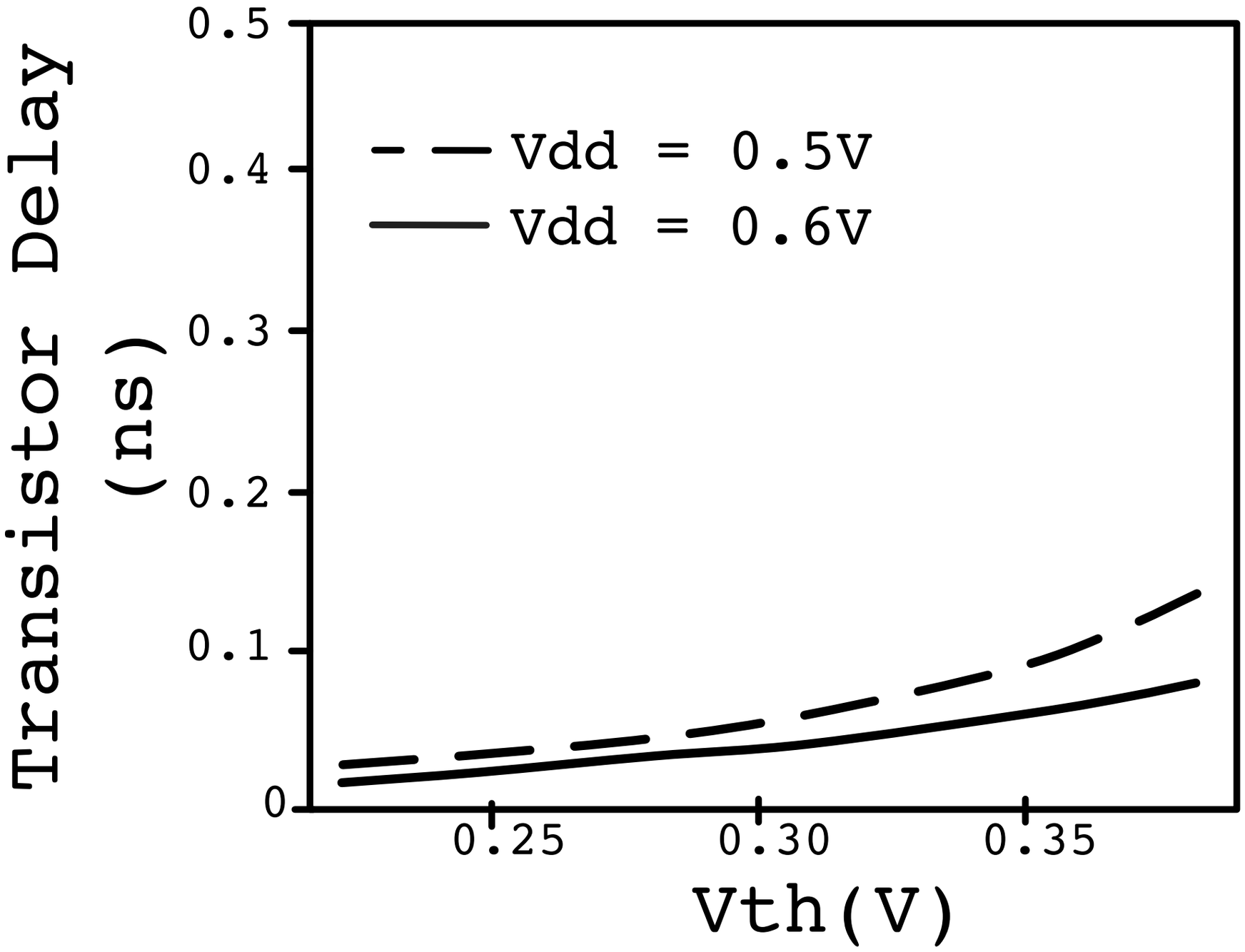
Parametric Variation at NTV



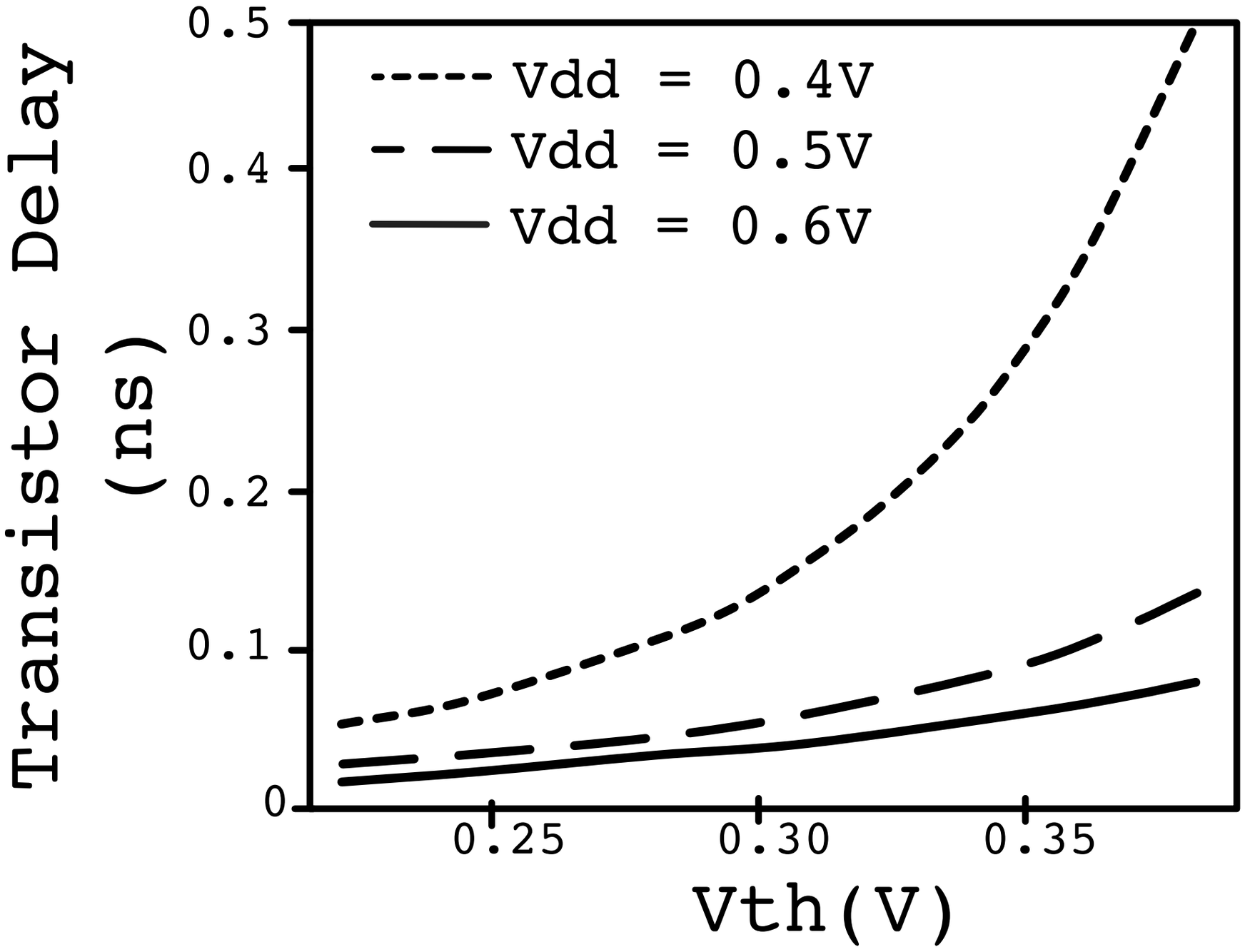
Parametric Variation at NTV



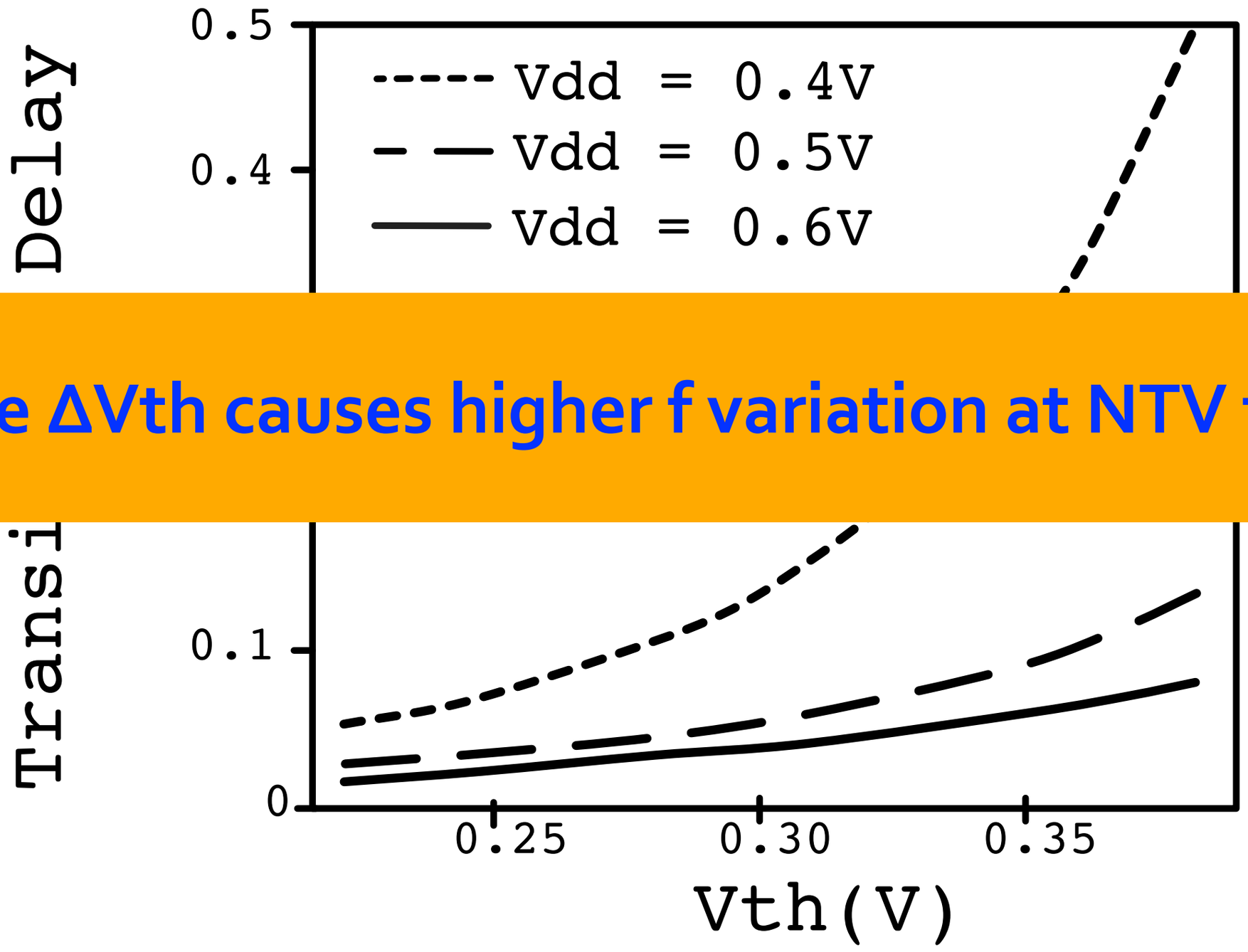
Parametric Variation at NTV



Parametric Variation at NTV



Parametric Variation at NTV



Same ΔVth causes higher f variation at NTV than at STV

VARIUS-NTV: Overview

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

  Systematic variation

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

Systematic variation + Random variation

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

Systematic variation + Random variation

Multi-variate Gaussian Distribution

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

Systematic variation + Random variation

Multi-variate Gaussian Distribution

Spatial Correlation

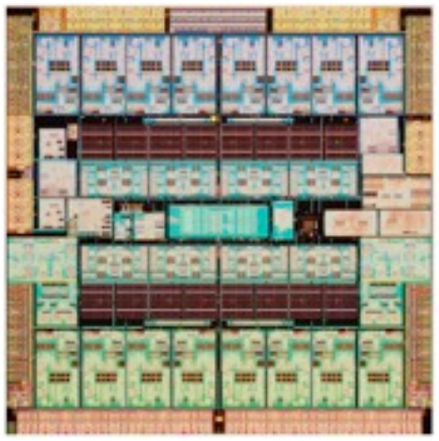
VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

Systematic variation + Random variation

Multi-variate Gaussian Distribution

Spatial Correlation



Floorplan

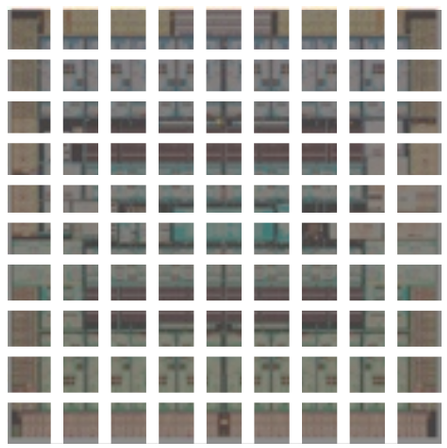
VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

Systematic variation + Random variation

Multi-variate Gaussian Distribution

Spatial Correlation



Floorplan

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

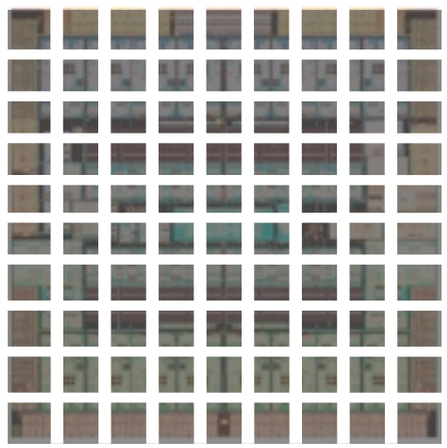
Systematic variation + Random variation

Multi-variate Gaussian Distribution

Spatial Correlation

No Correlation

Model at transistor granularity



Floorplan

VARIUS-NTV: Overview

- Extend VARIUS [Sarangi'08]

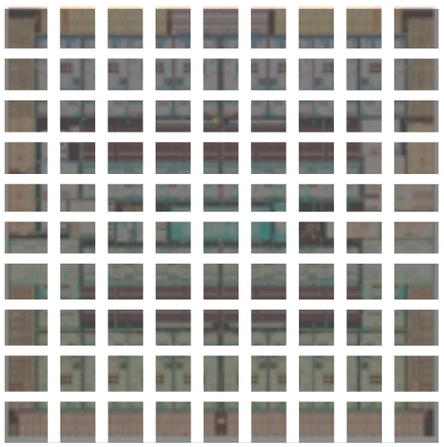
Systematic variation + Random variation

Multi-variate Gaussian Distribution

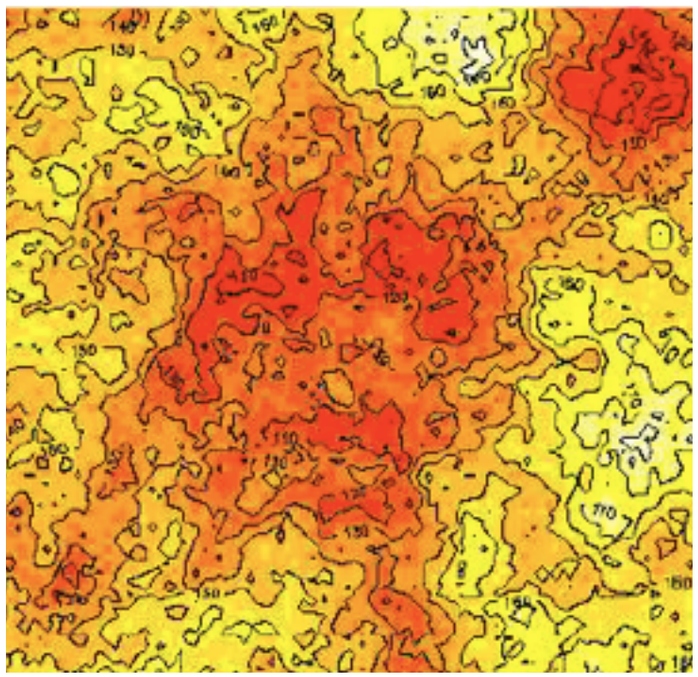
Spatial Correlation

No Correlation

Model at transistor granularity

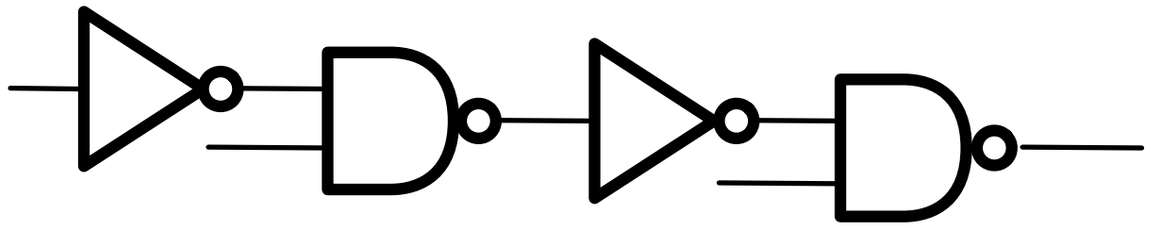


Floorplan

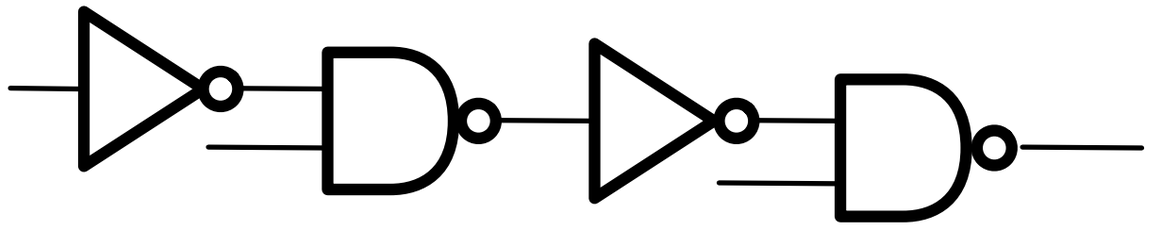


Vth, Leff maps

VARIUS-NTV: Logic Timing Model



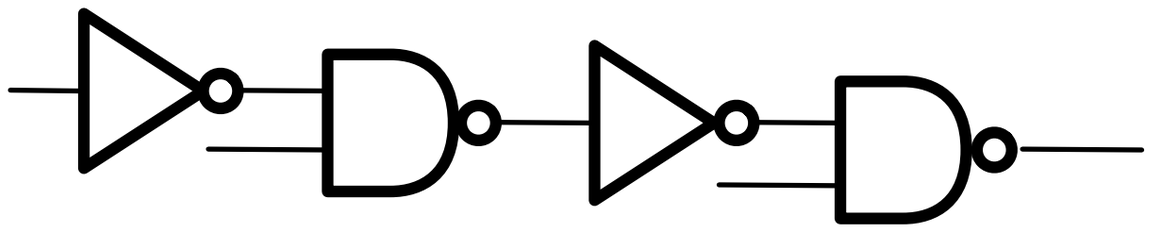
VARIUS-NTV: Logic Timing Model



Gate delay

VARIUS-NTV: EKV-based $\tau \propto \frac{\mathrm{Vdd} \times \mathrm{L_{eff}}}{\ln^2(\mathrm{e}^{(\frac{\mathrm{Vdd}-\mathrm{Vth}}{2\times \mathrm{n}\times \mathrm{v_t}})}+1)}$

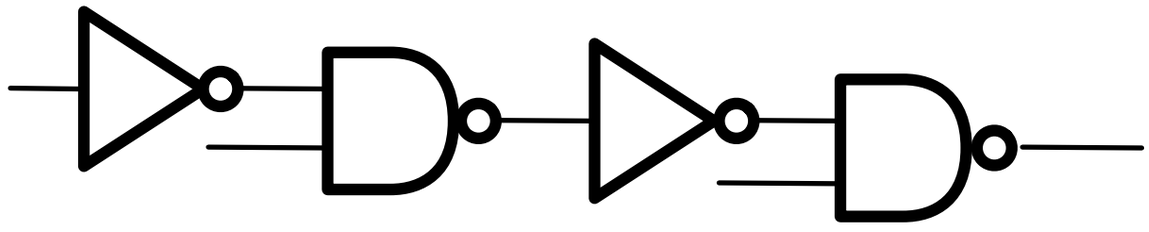
VARIUS-NTV: Logic Timing Model



Gate delay

VARIUS-NTV: EKV-based $\tau \propto \frac{\mathrm{Vdd} \times \mathrm{L_{eff}}}{\ln^2(e^{(\frac{\mathrm{Vdd}-\mathrm{Vth}}{2 \times n \times \mu_t})} + 1)}$

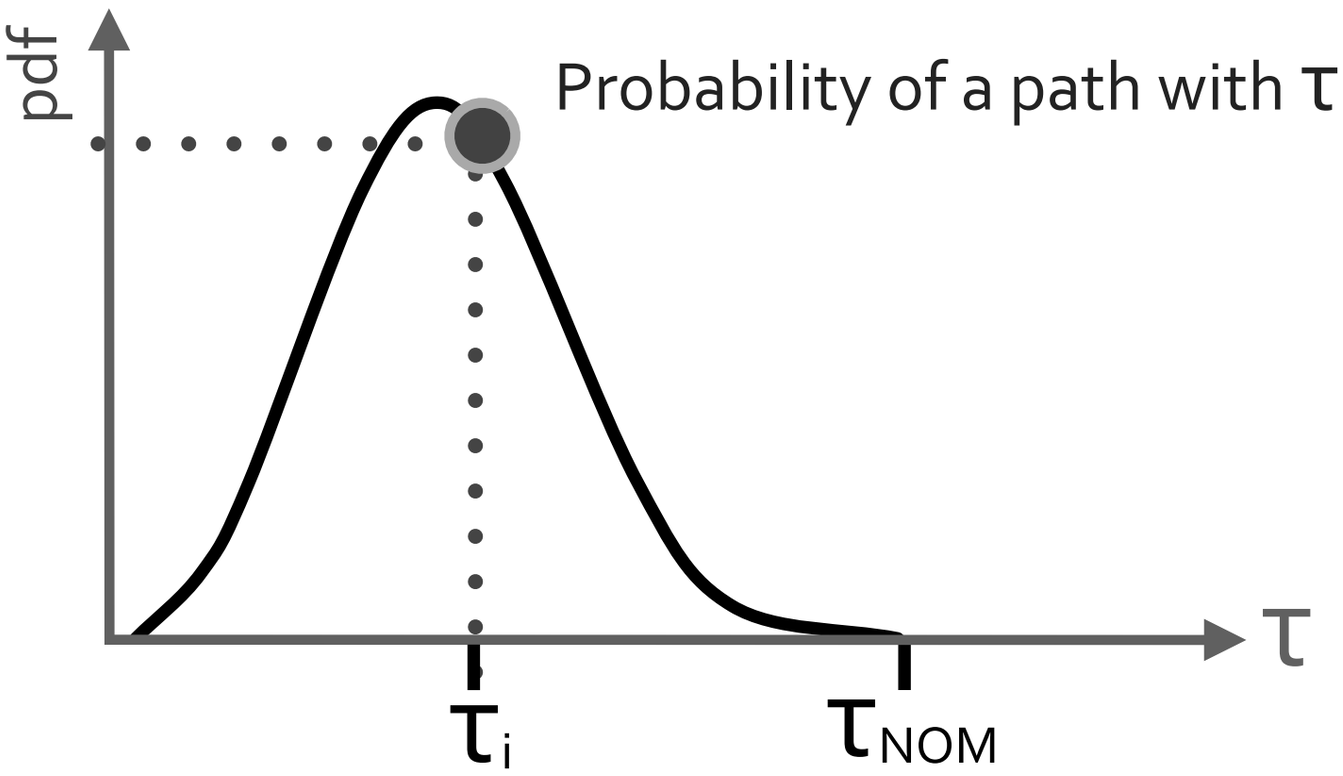
VARIUS-NTV: Logic Timing Model



Gate delay

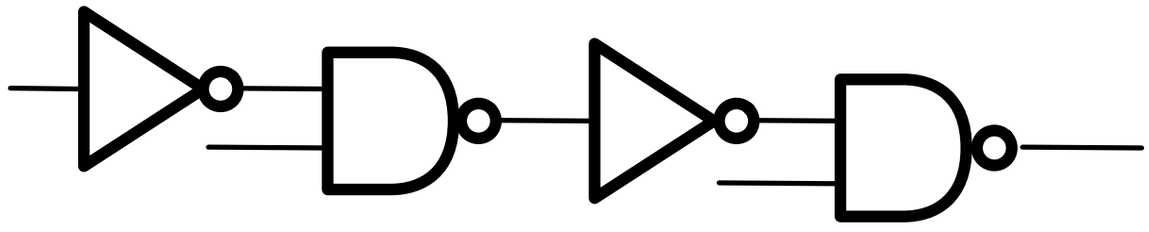
VARIUS-NTV: EKV-based $\tau \propto \frac{\mathrm{Vdd} \times \mathrm{L_{eff}}}{\ln^2(e^{(\frac{\mathrm{Vdd}-\mathrm{Vth}}{2 \times n \times \mu_t})} + 1)}$

Path delay



Probability of a path with $\tau = \tau_i$ being exercised

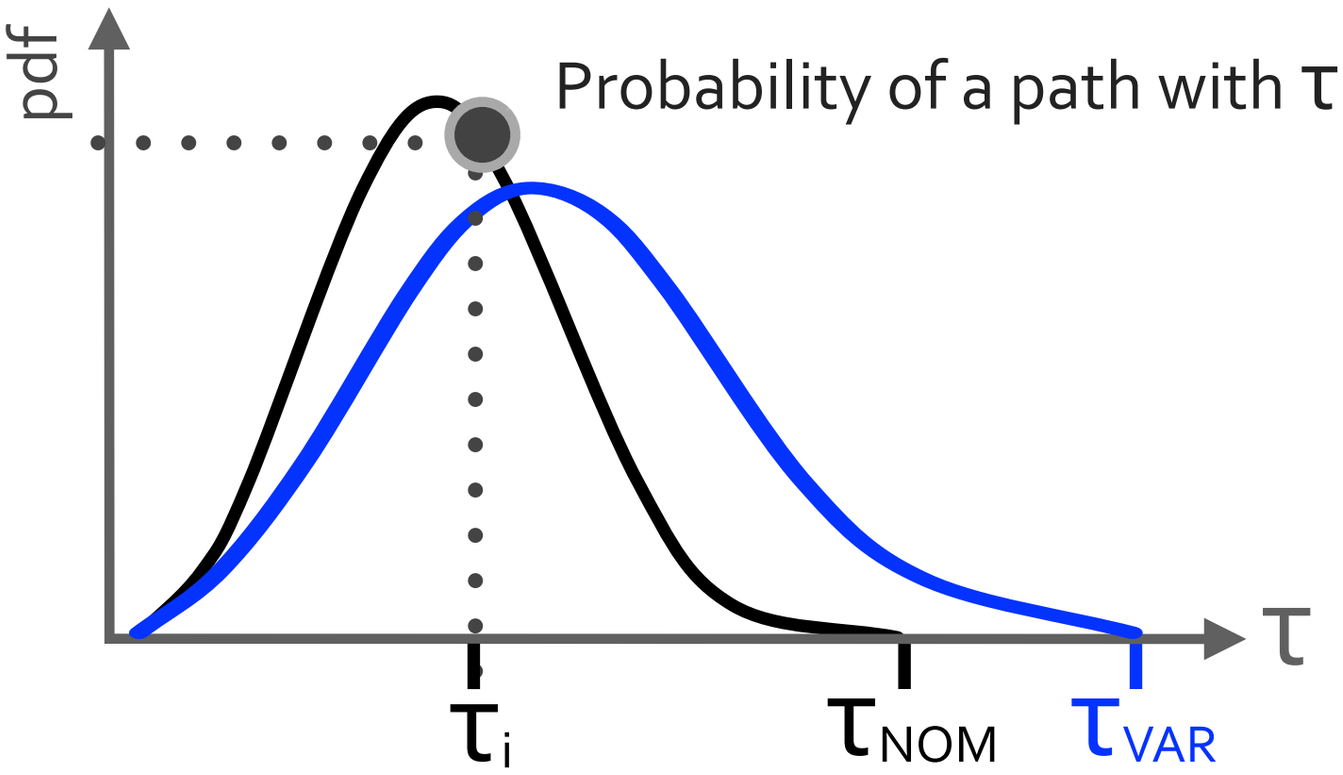
VARIUS-NTV: Logic Timing Model



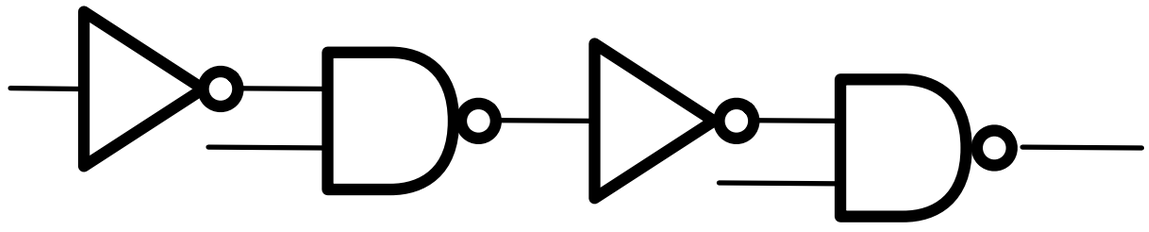
Gate delay

VARIUS-NTV: EKV-based $\tau \propto \frac{\mathrm{Vdd} \times \mathrm{L_{eff}}}{\ln^2(e^{(\frac{\mathrm{Vdd}-\mathrm{Vth}}{2 \times n \times \mu_t})} + 1)}$

Path delay



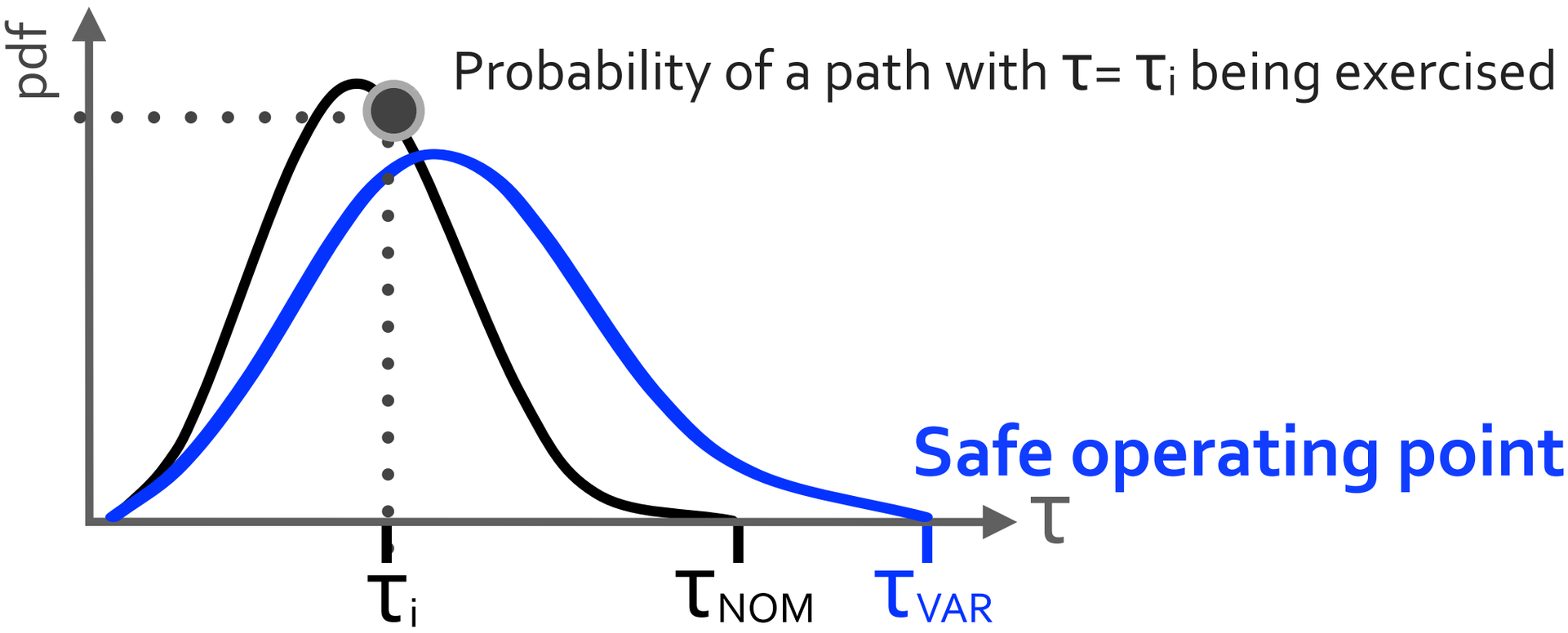
VARIUS-NTV: Logic Timing Model



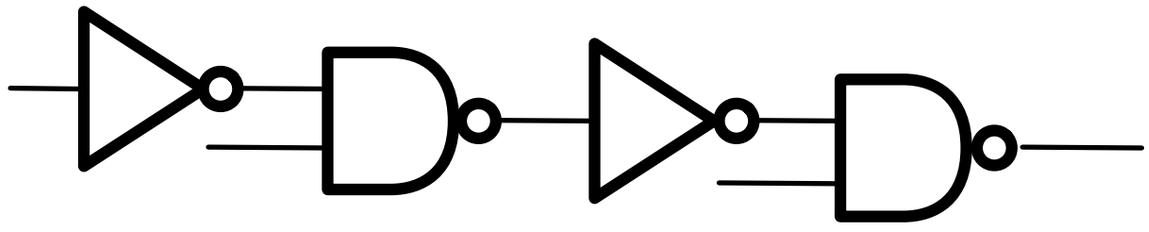
Gate delay

VARIUS-NTV: EKV-based $\tau \propto \frac{\mathrm{Vdd} \times \mathrm{L_{eff}}}{\ln^2(e^{(\frac{\mathrm{Vdd}-\mathrm{Vth}}{2 \times n \times \mu_t})} + 1)}$

Path delay



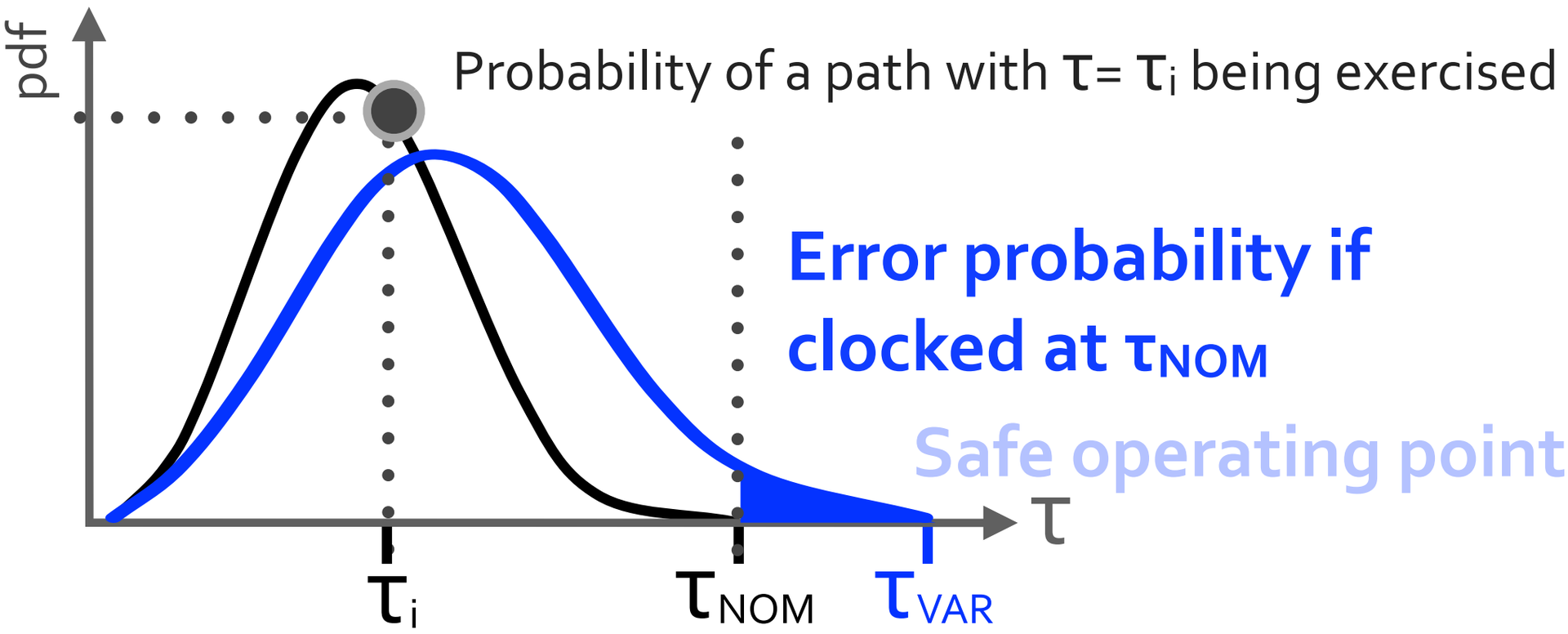
VARIUS-NTV: Logic Timing Model



Gate delay

VARIUS-NTV: EKV-based $\tau \propto \frac{\mathrm{Vdd} \times \mathrm{L_{eff}}}{\ln^2(e^{(\frac{\mathrm{Vdd}-\mathrm{Vth}}{2 \times n \times \mu_t})} + 1)}$

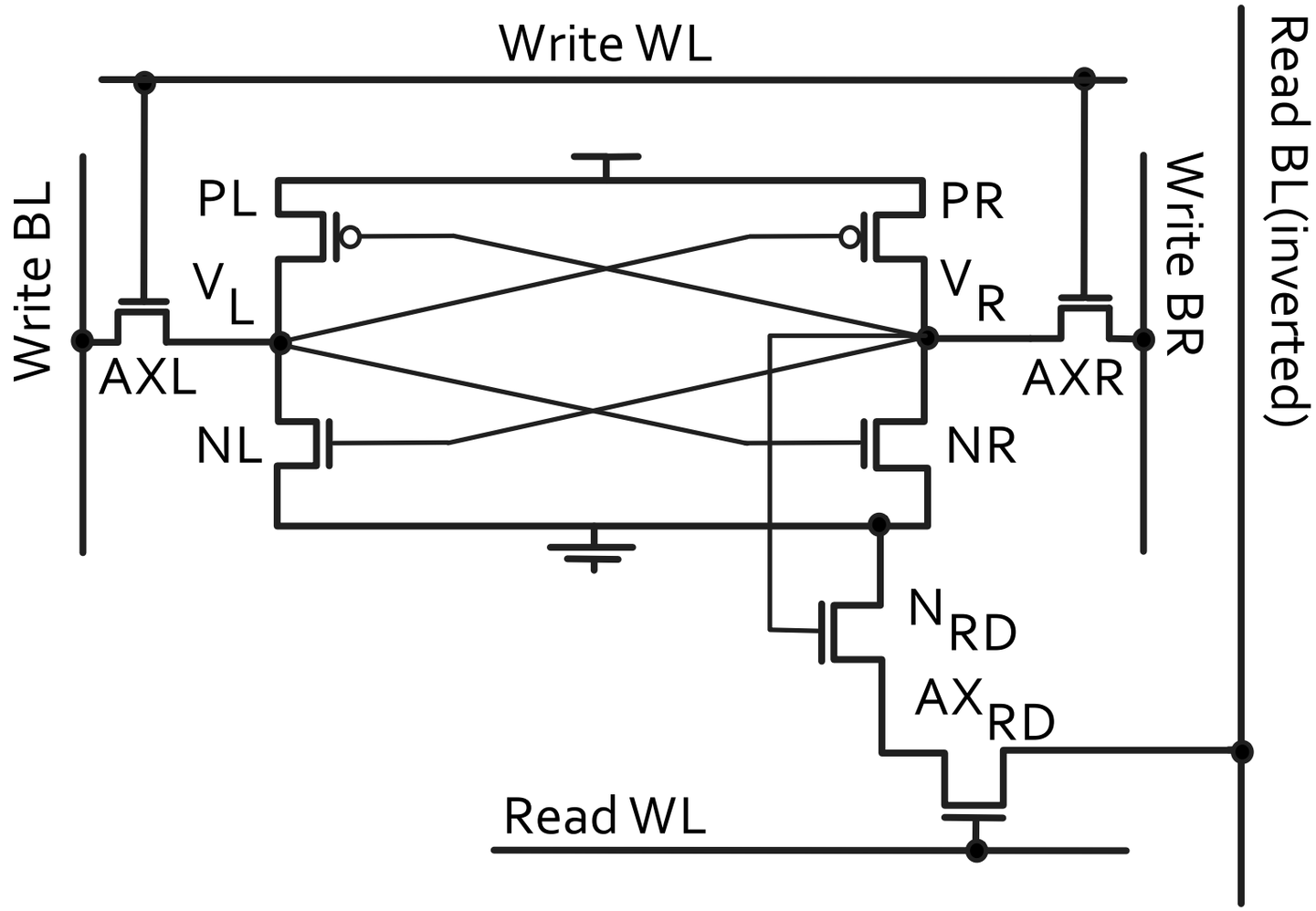
Path delay



VARIUS-NTV: Memory Model

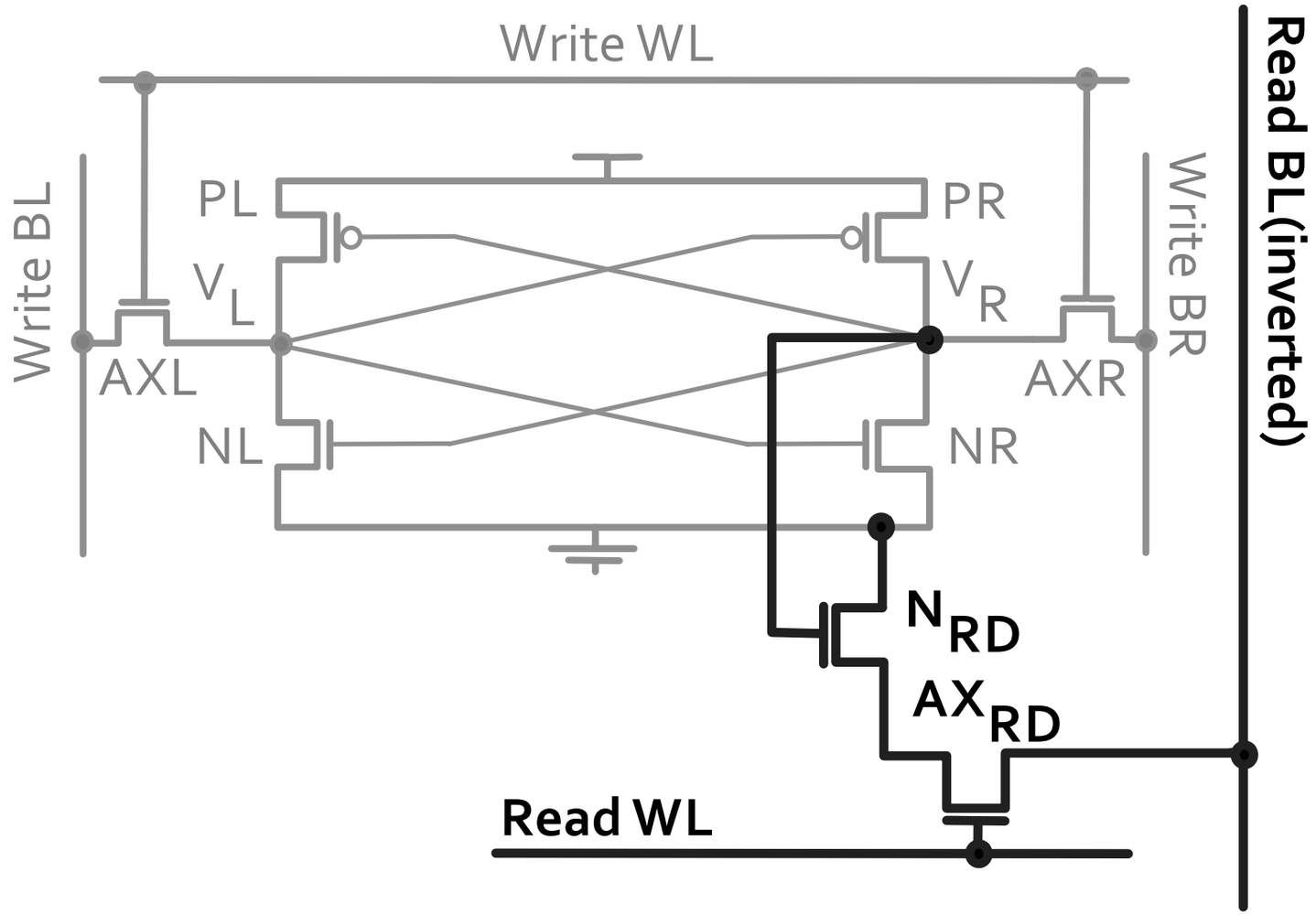
VARIUS-NTV: Memory Model

- 8T SRAM Cell



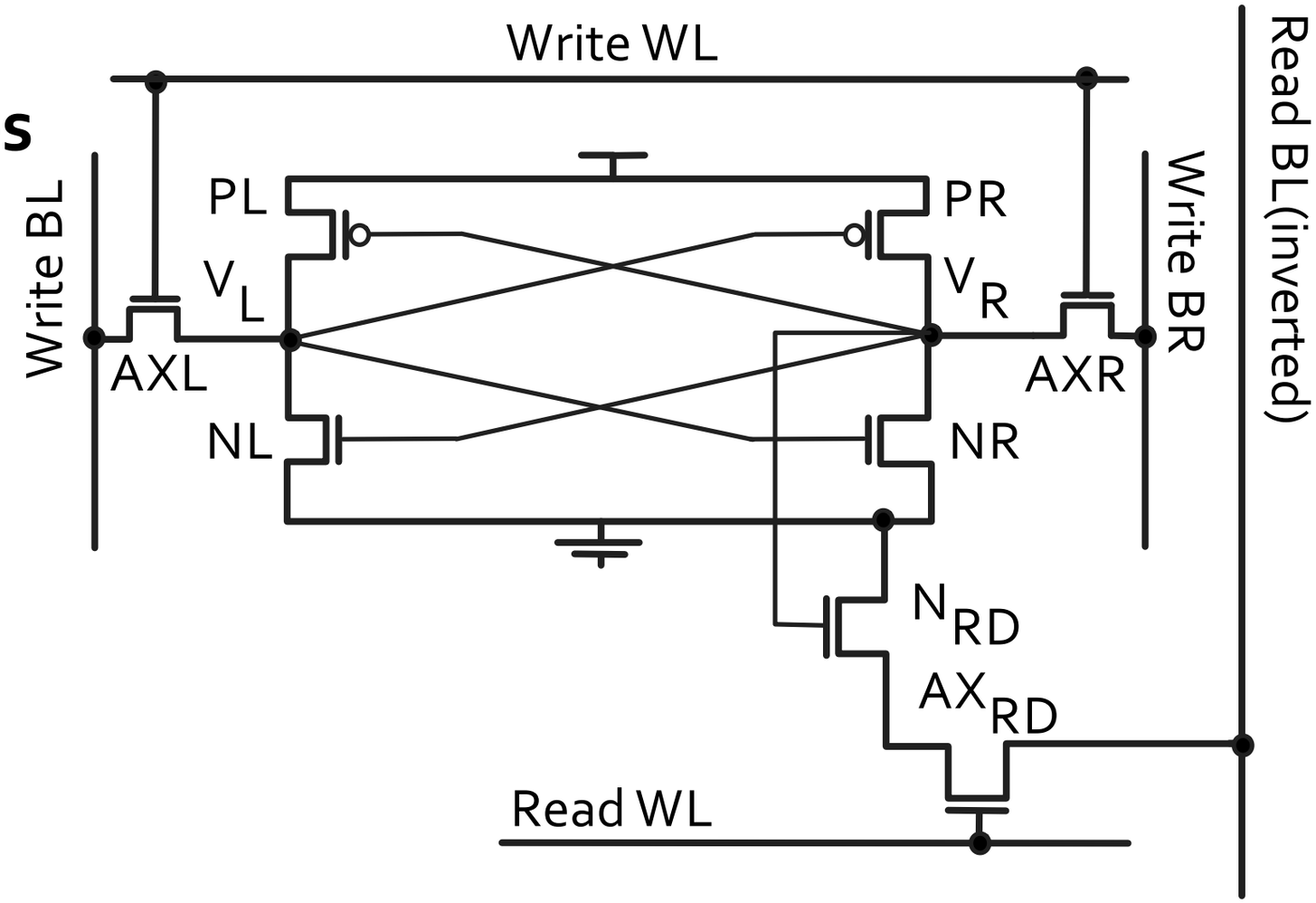
VARIUS-NTV: Memory Model

- 8T SRAM Cell



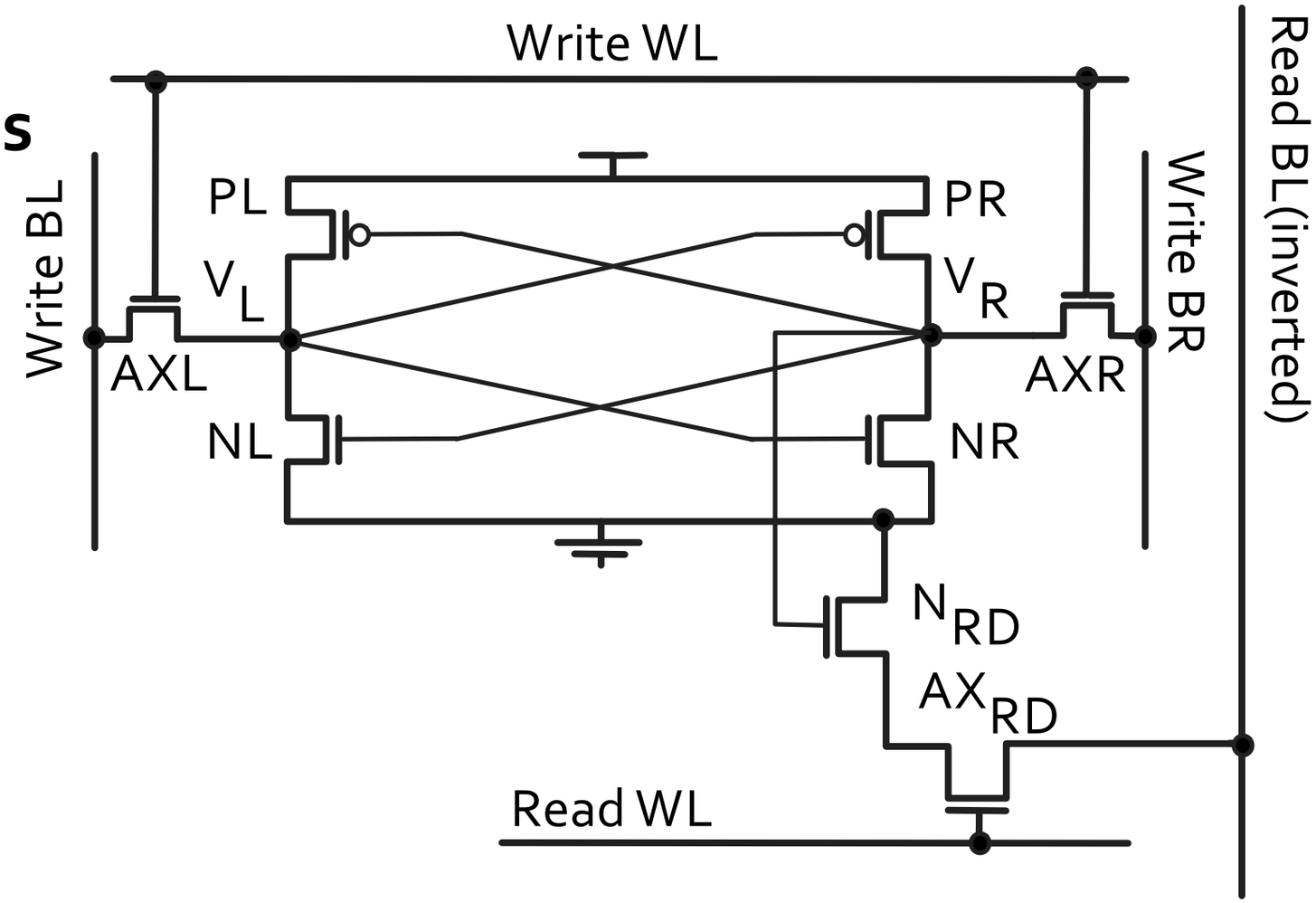
VARIUS-NTV: Memory Model

- 8T SRAM Cell
- Model various failure types



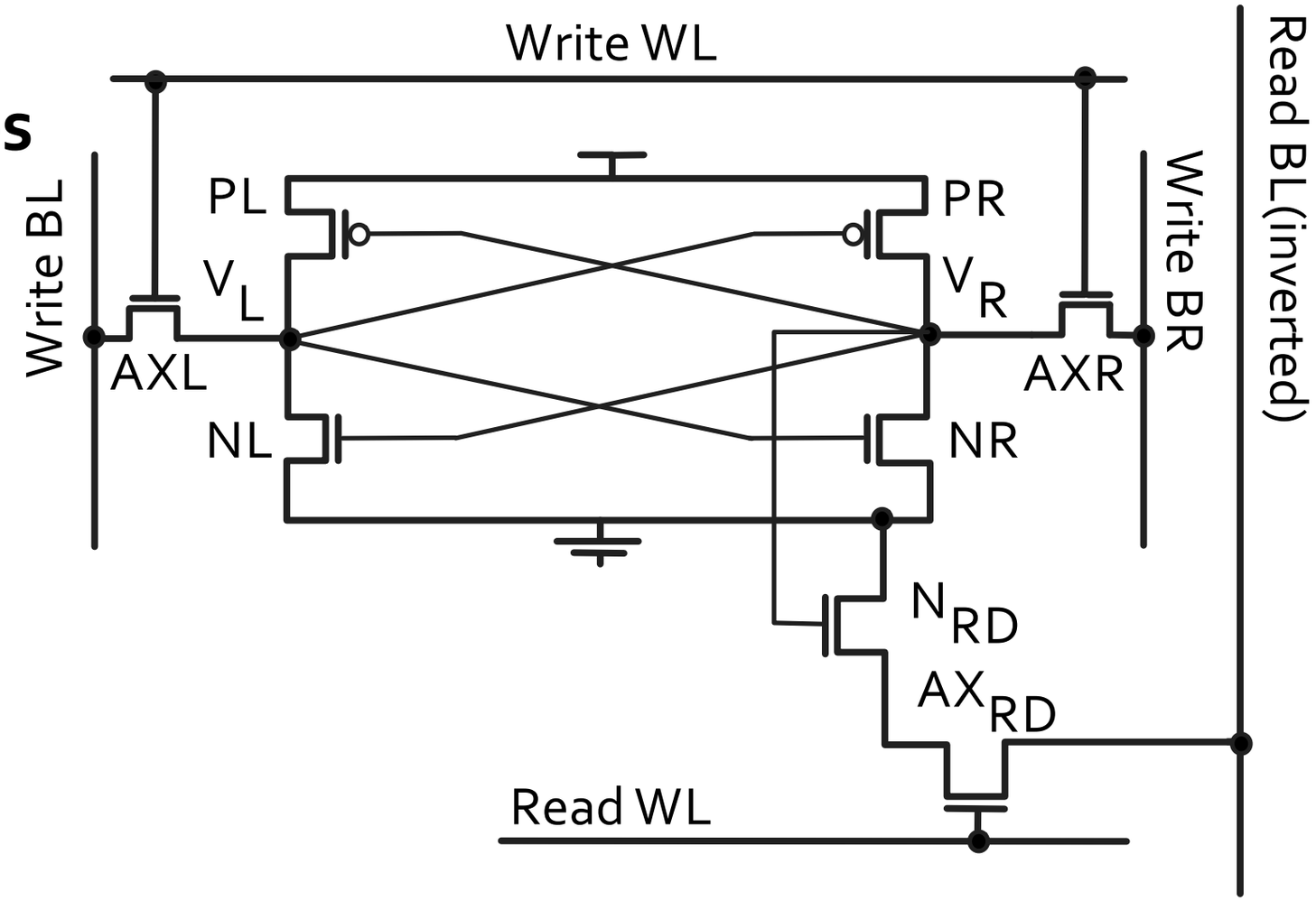
VARIUS-NTV: Memory Model

- 8T SRAM Cell
- Model various failure types
  - Hold



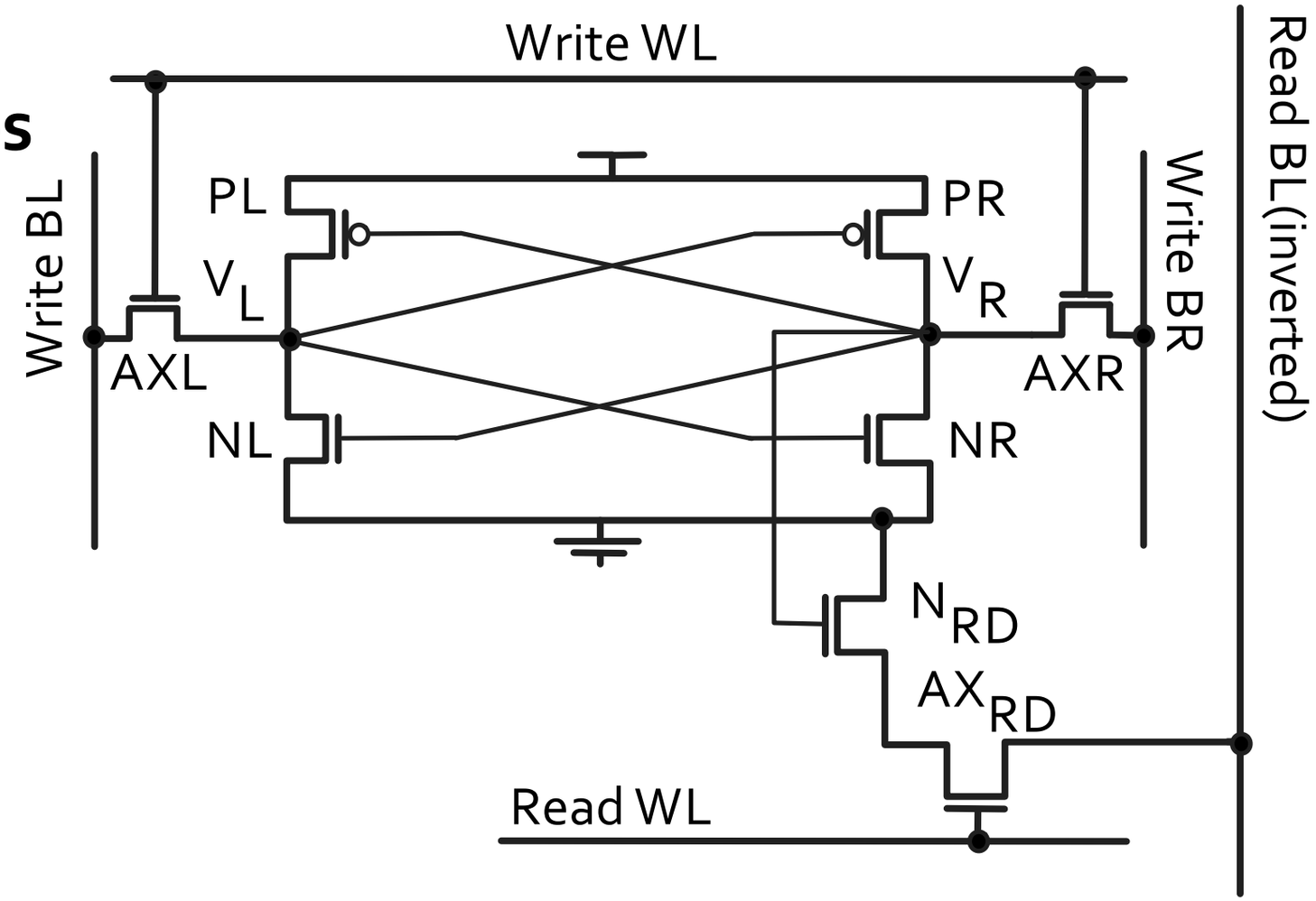
VARIUS-NTV: Memory Model

- 8T SRAM Cell
- Model various failure types
  - Hold
  - WR Stability



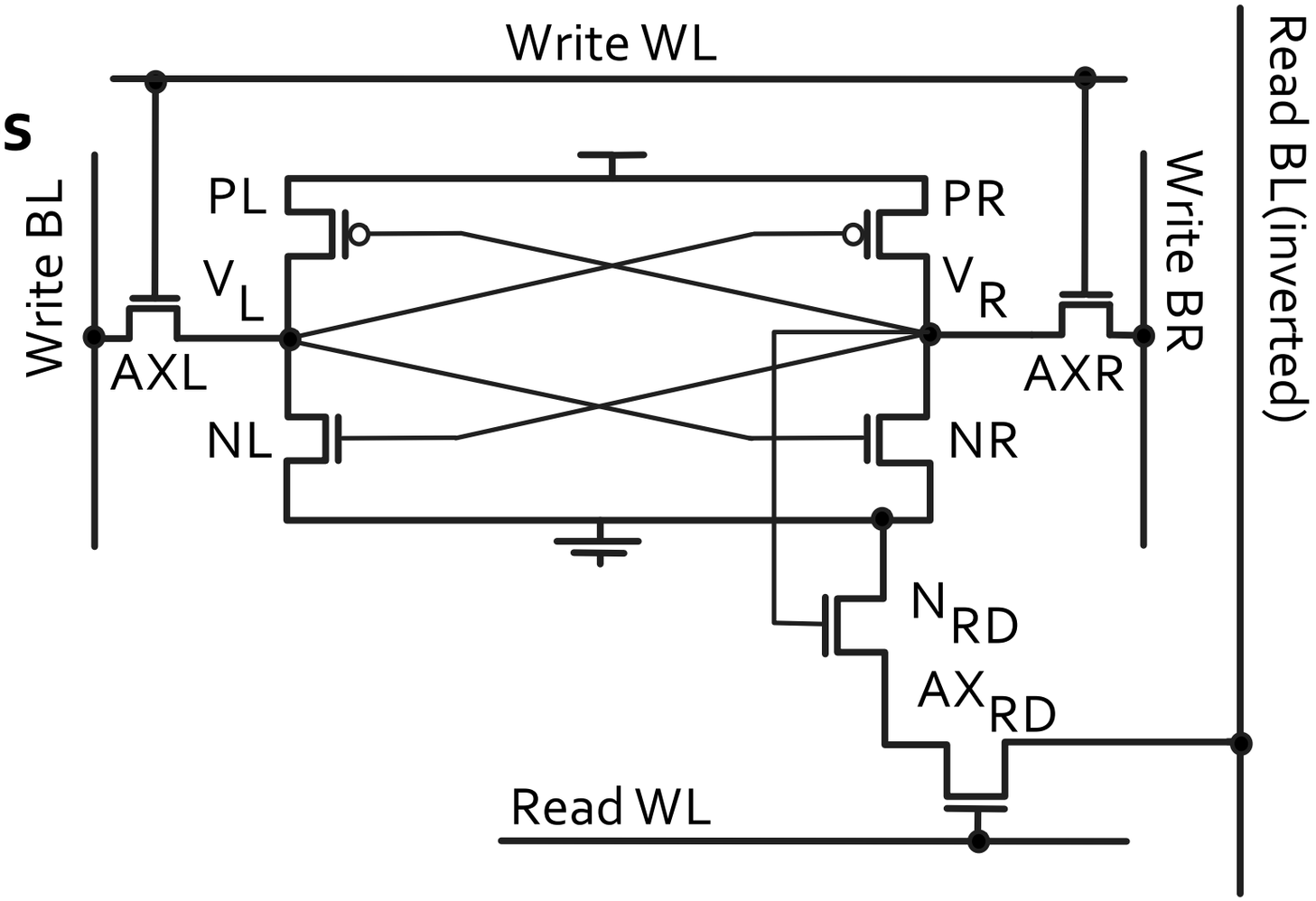
VARIUS-NTV: Memory Model

- 8T SRAM Cell
- Model various failure types
  - Hold
  - WR Stability
  - WR Timing



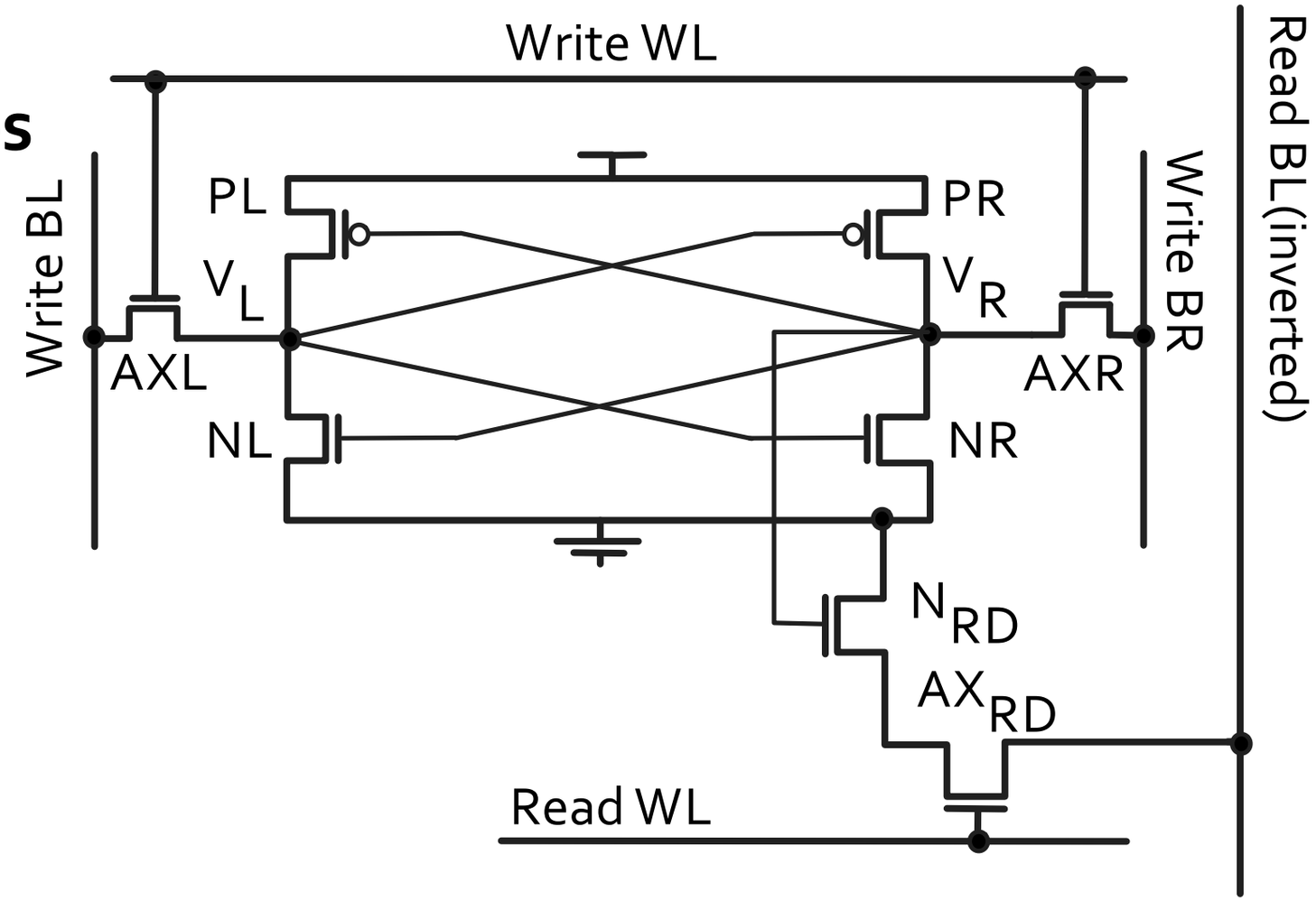
VARIUS-NTV: Memory Model

- 8T SRAM Cell
- Model various failure types
  - Hold
  - WR Stability
  - WR Timing
  - Rd Timing

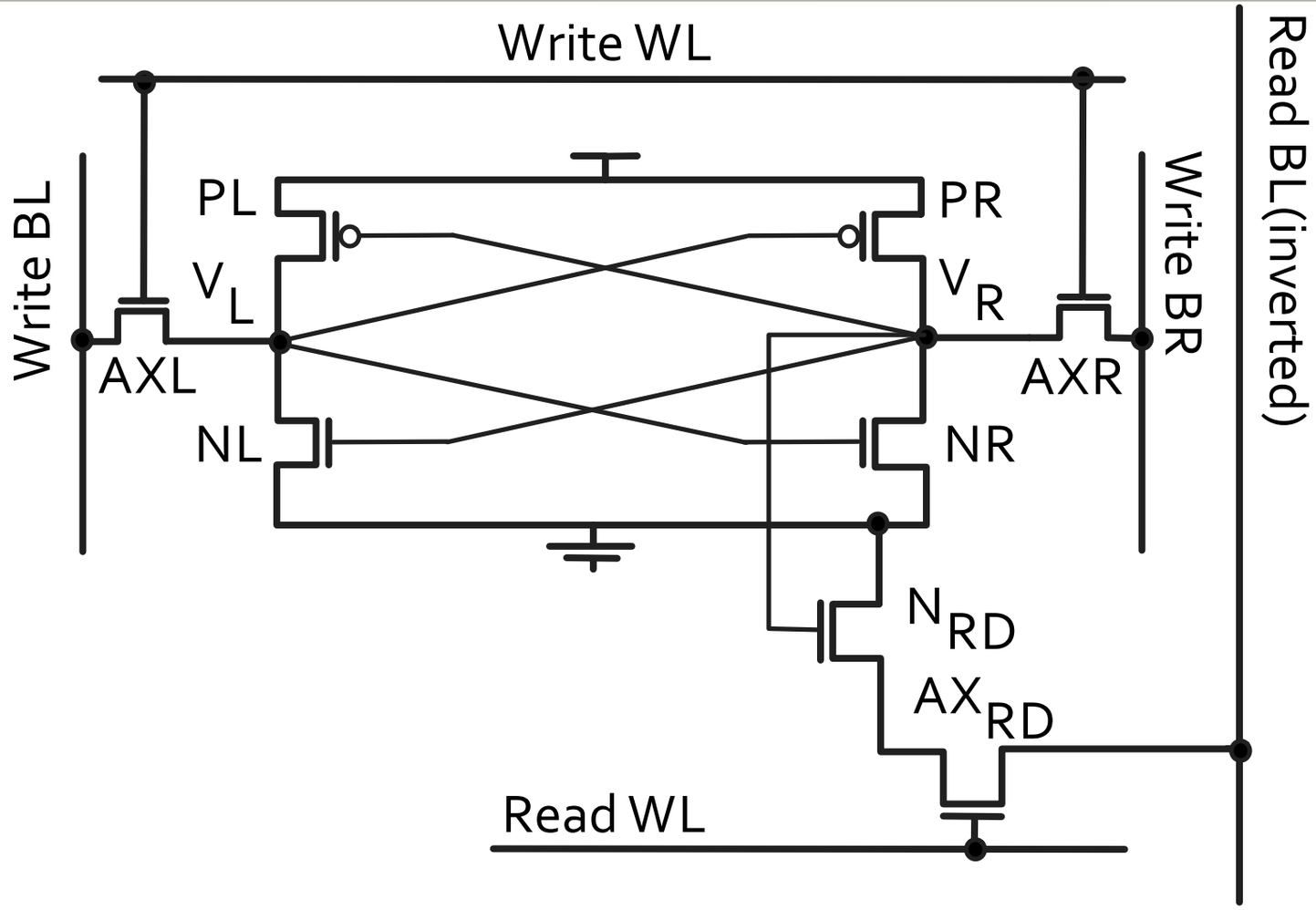


VARIUS-NTV: Memory Model

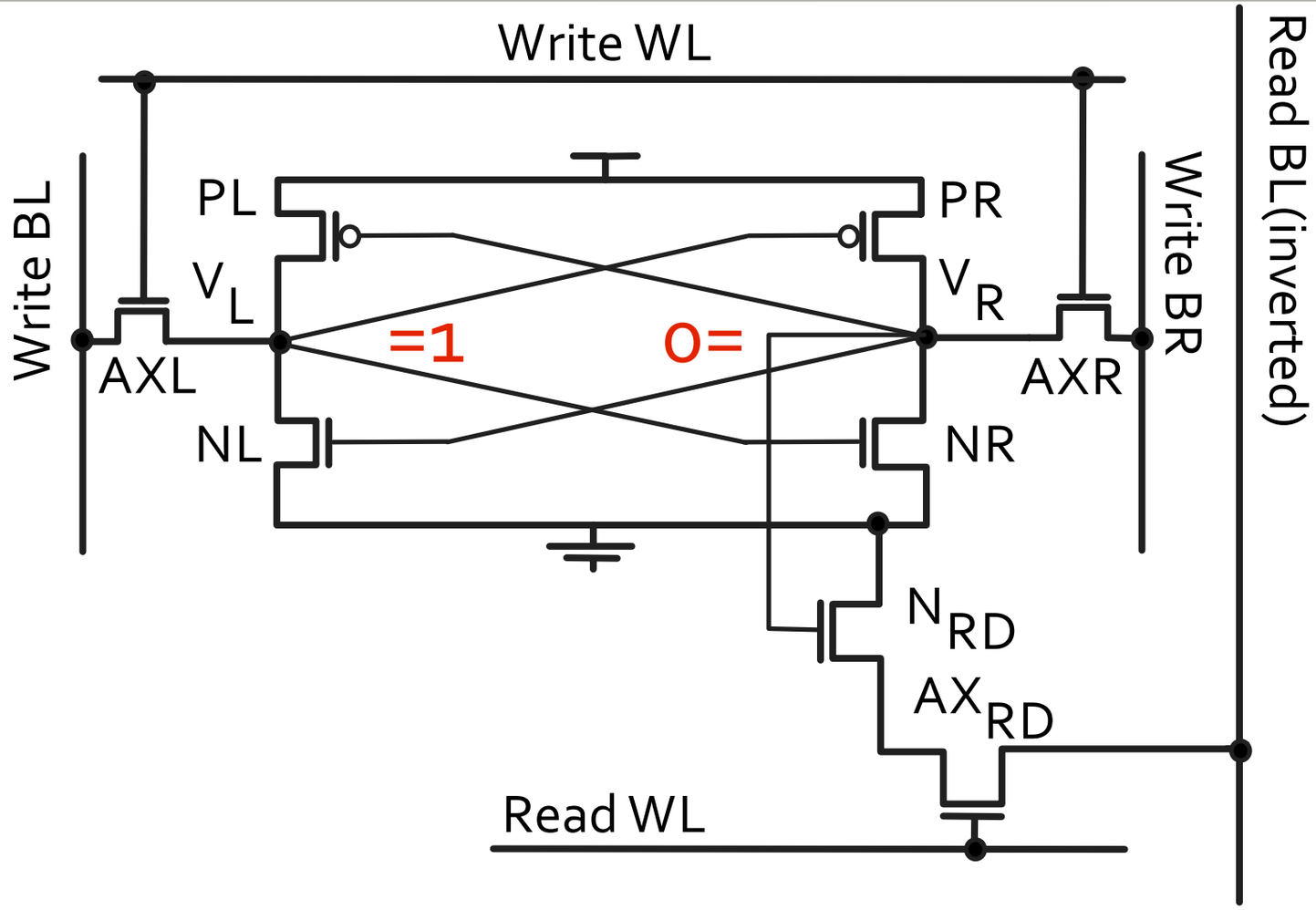
- **8T SRAM Cell**
- **Model various failure types**
  - **Hold**
  - **WR Stability**
  - **WR Timing**
  - Rd Timing
- **Account for leakage**



Hold Analysis

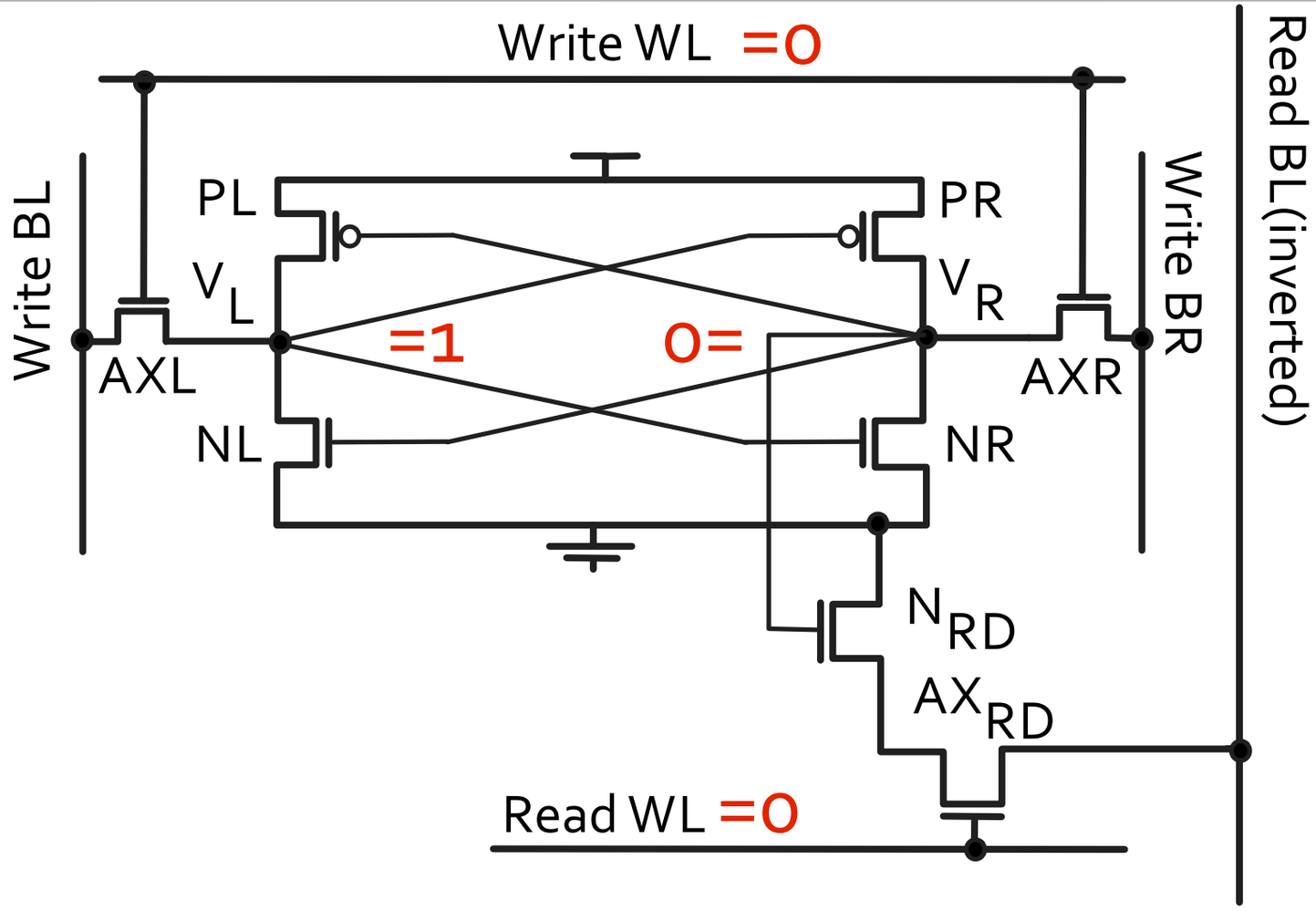


Hold Analysis



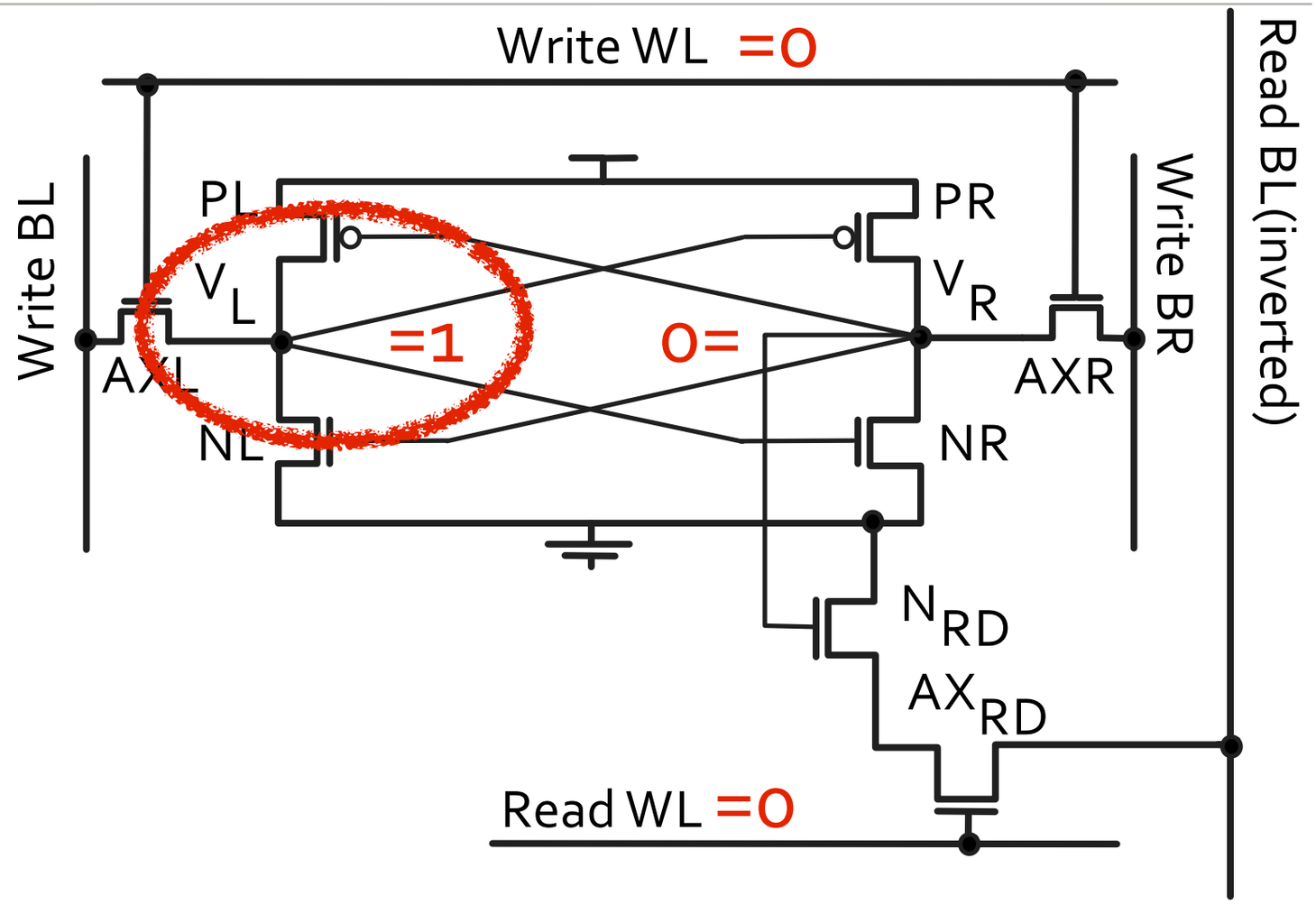
Hold Analysis

- The cell is not accessed



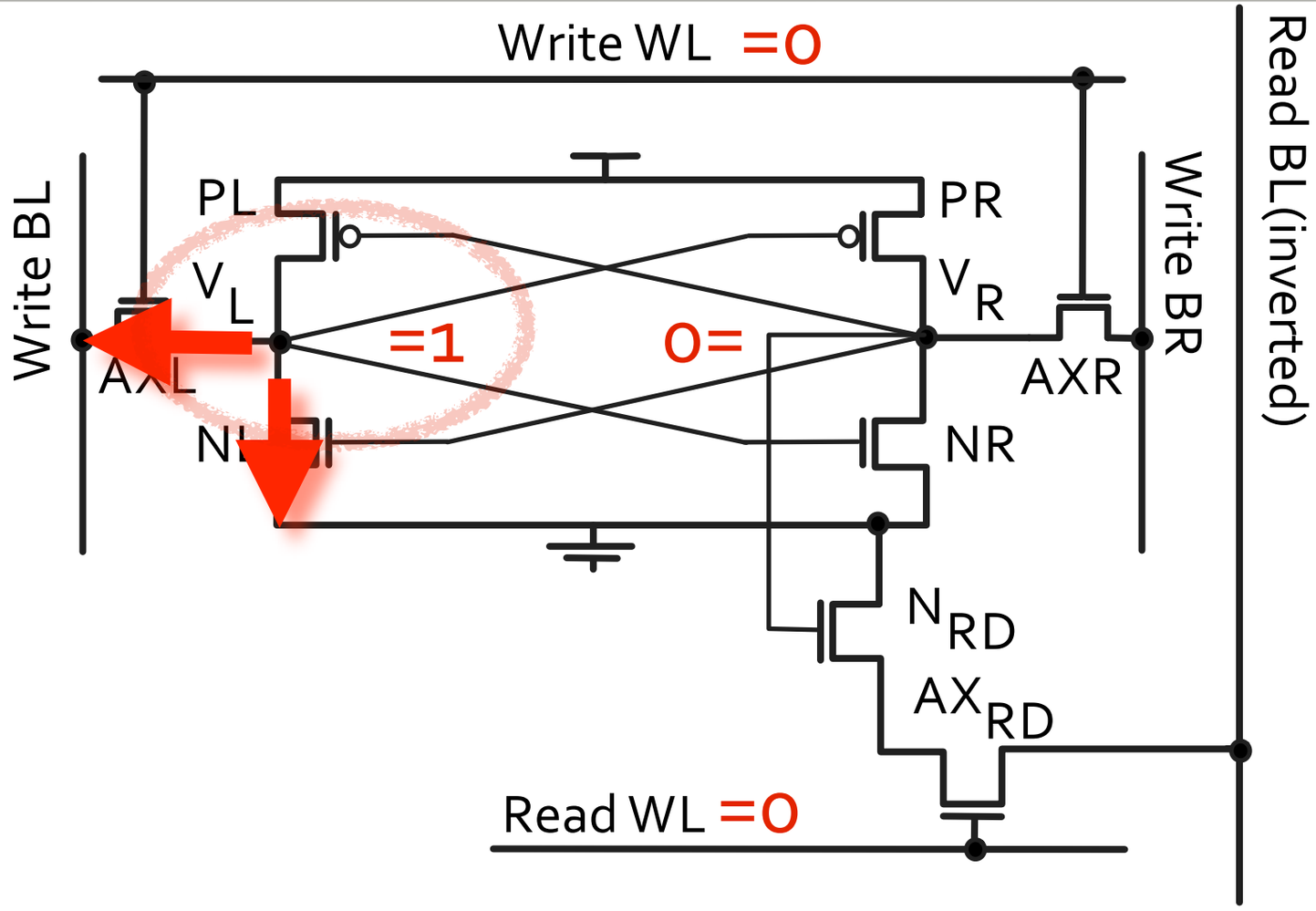
Hold Analysis

- The cell is not accessed
- Node L looses its state



Hold Analysis

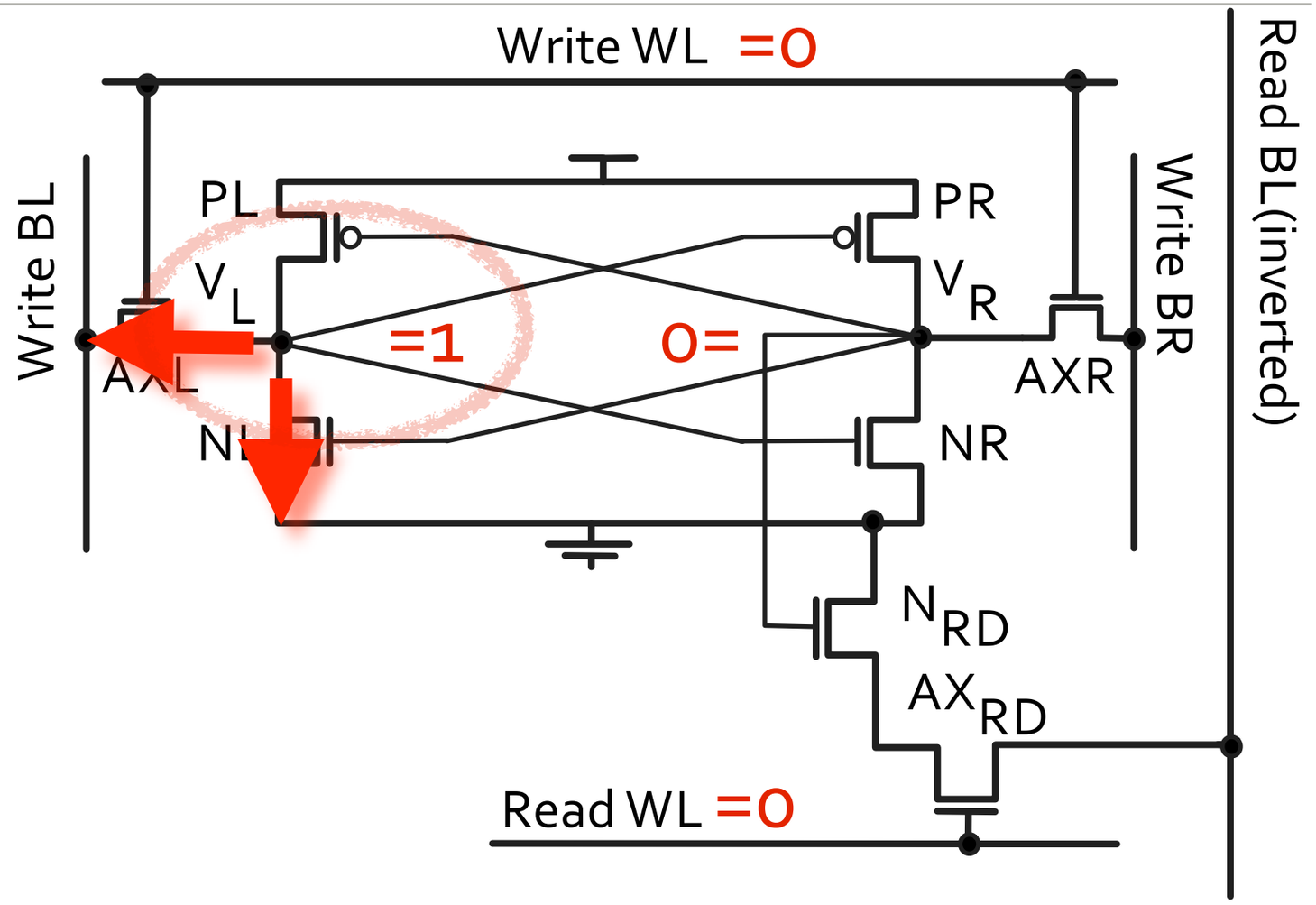
- The cell is not accessed
- Node L looses its state
  - Excessive leakage



Hold Analysis

- The cell is not accessed
- Node L looses its state
  - Excessive leakage

VARIUS-NTV

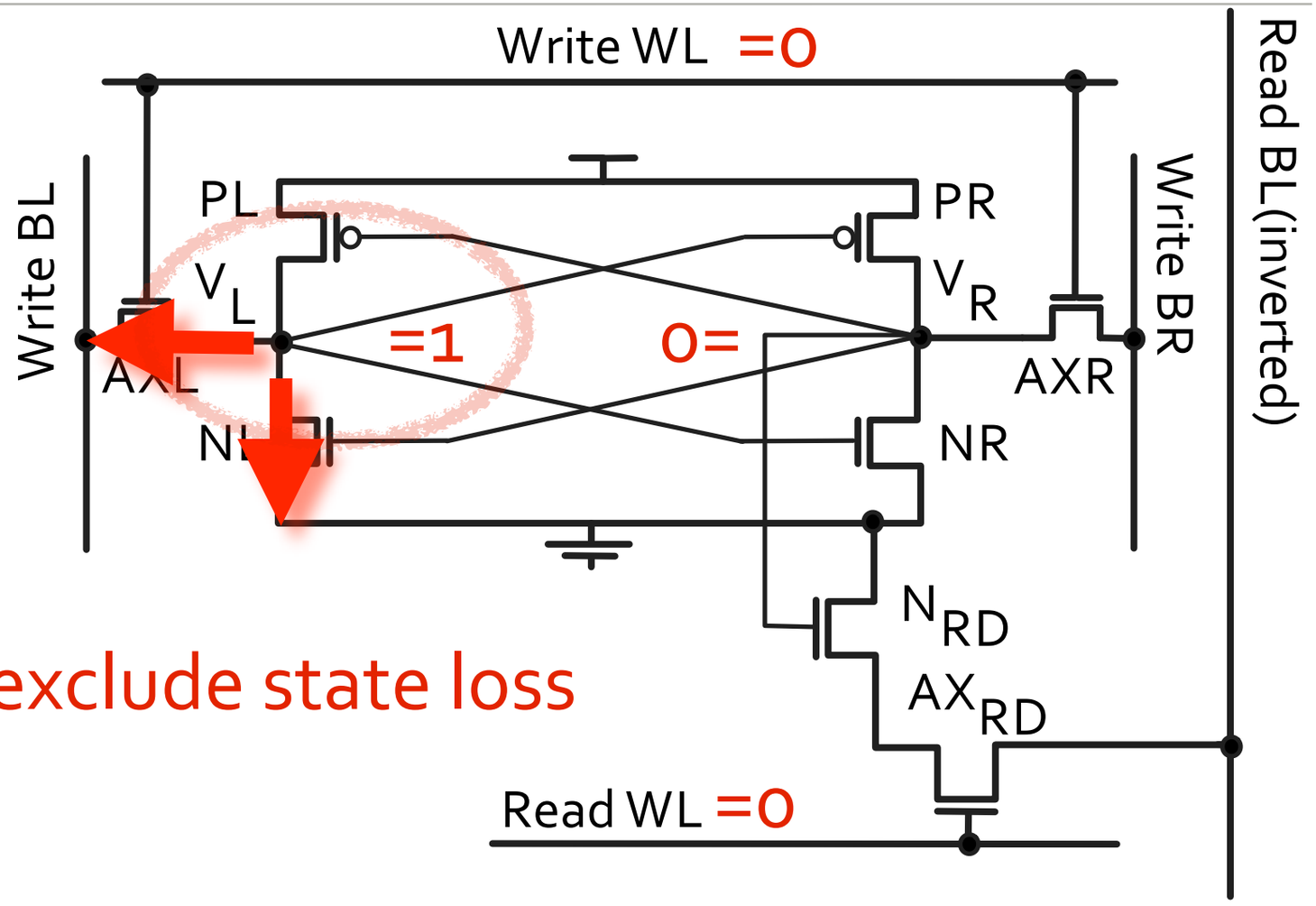


Hold Analysis

- The cell is not accessed
- Node L looses its state
  - Excessive leakage

VARIUS-NTV

- Minimum V_{dd} (V_{ddMIN}) to exclude state loss

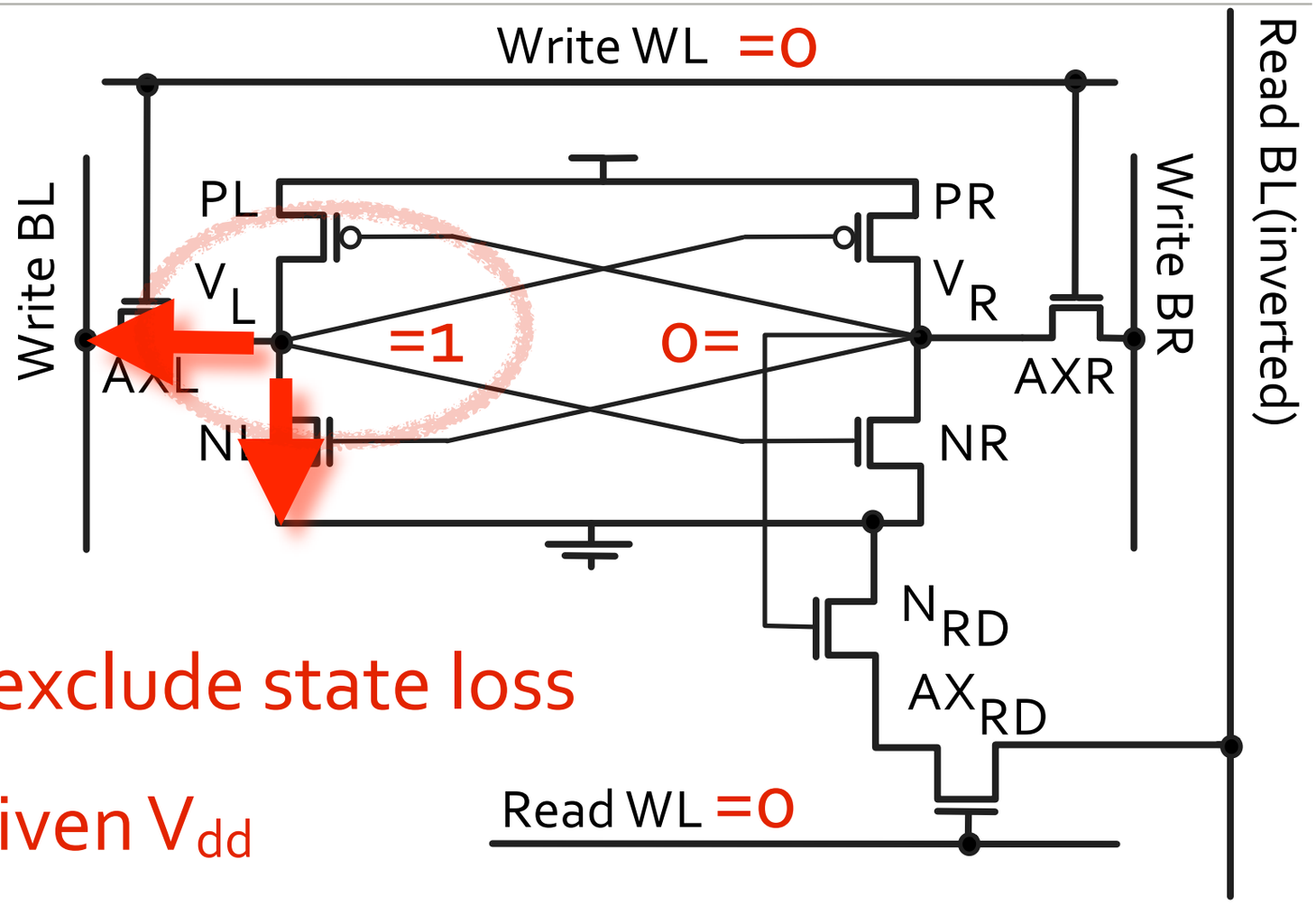


Hold Analysis

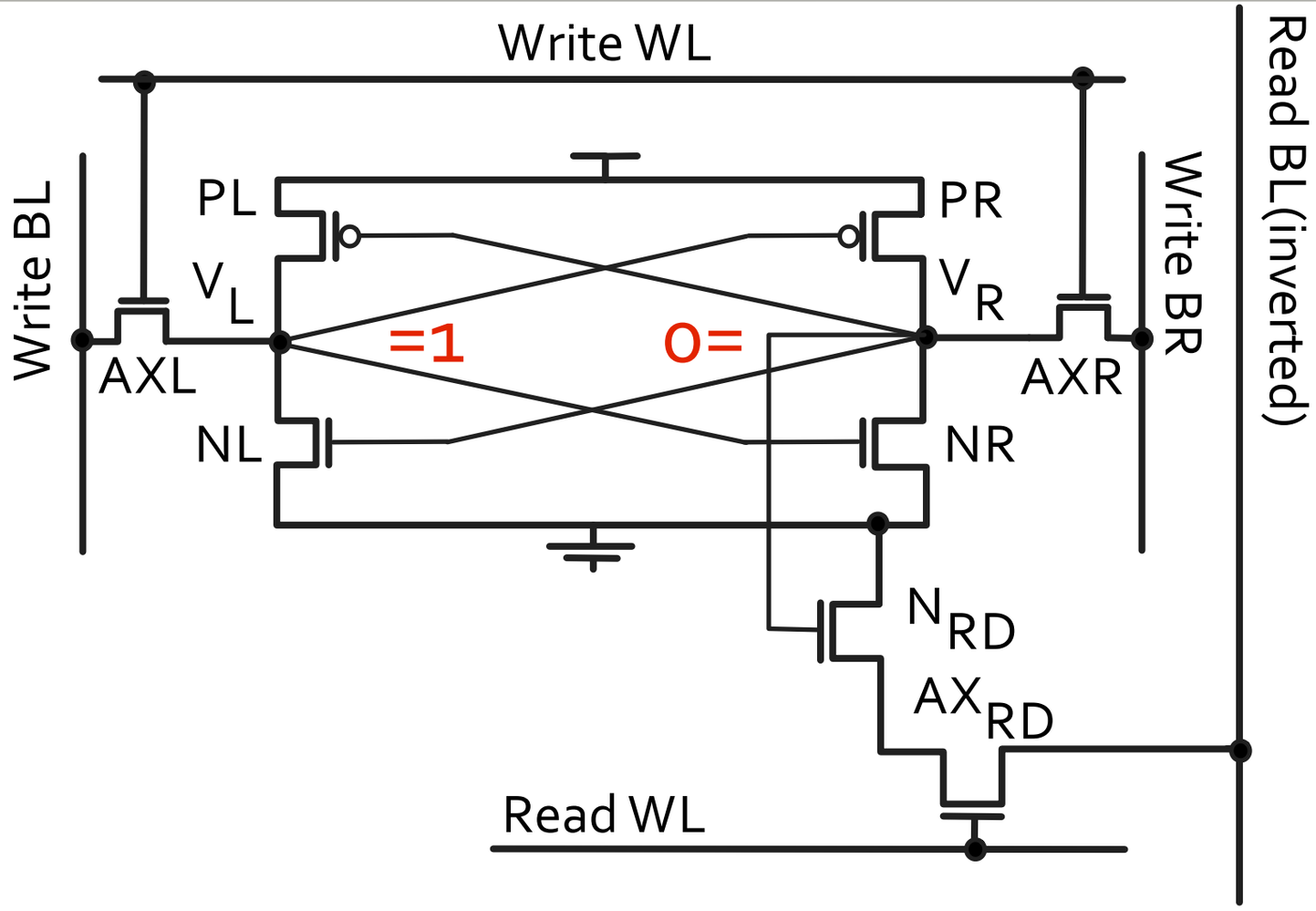
- The cell is not accessed
- Node L looses its state
  - Excessive leakage

VARIUS-NTV

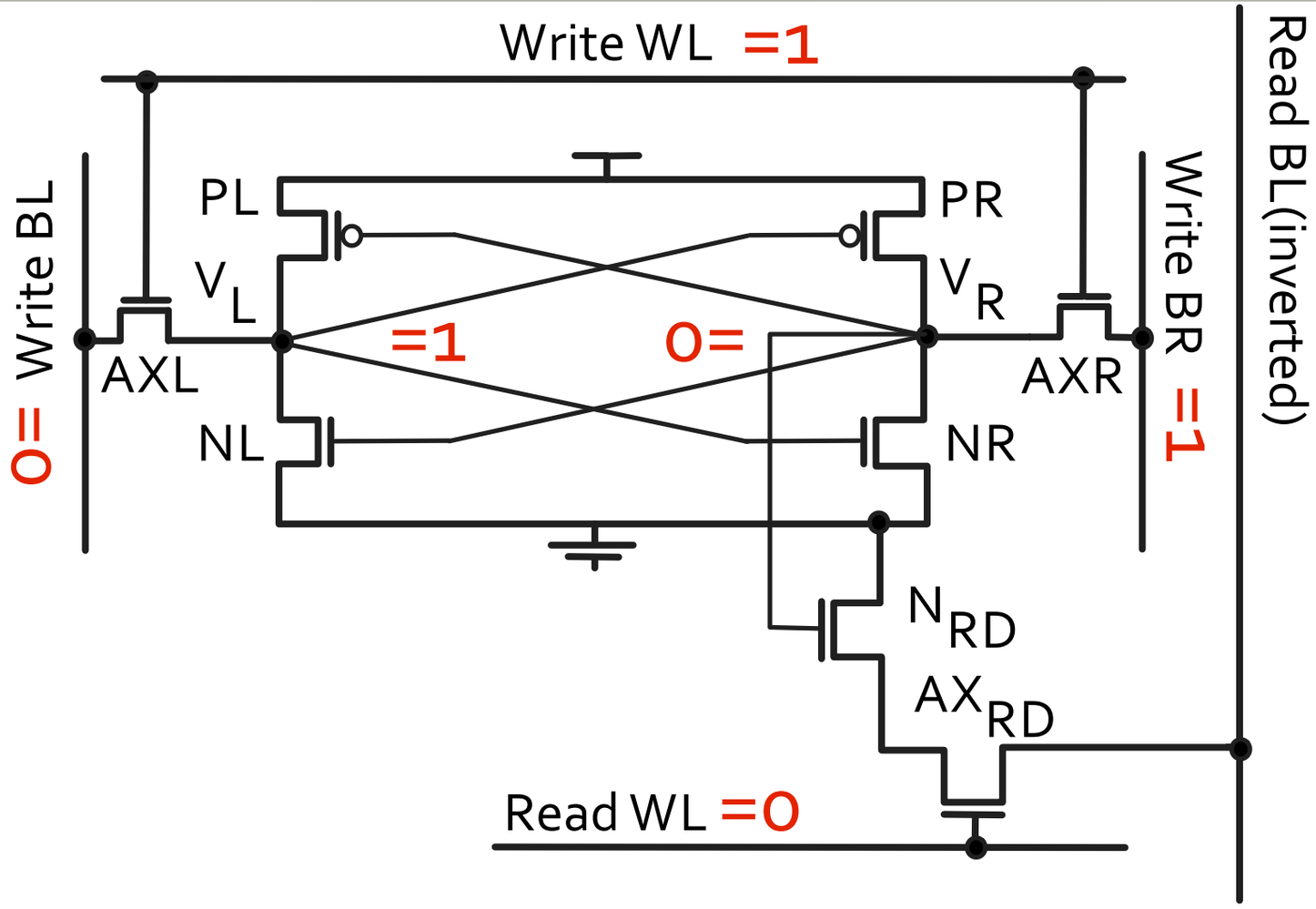
- Minimum V_{dd} (V_{ddMIN}) to exclude state loss
- Hold failure rate at any given V_{dd}



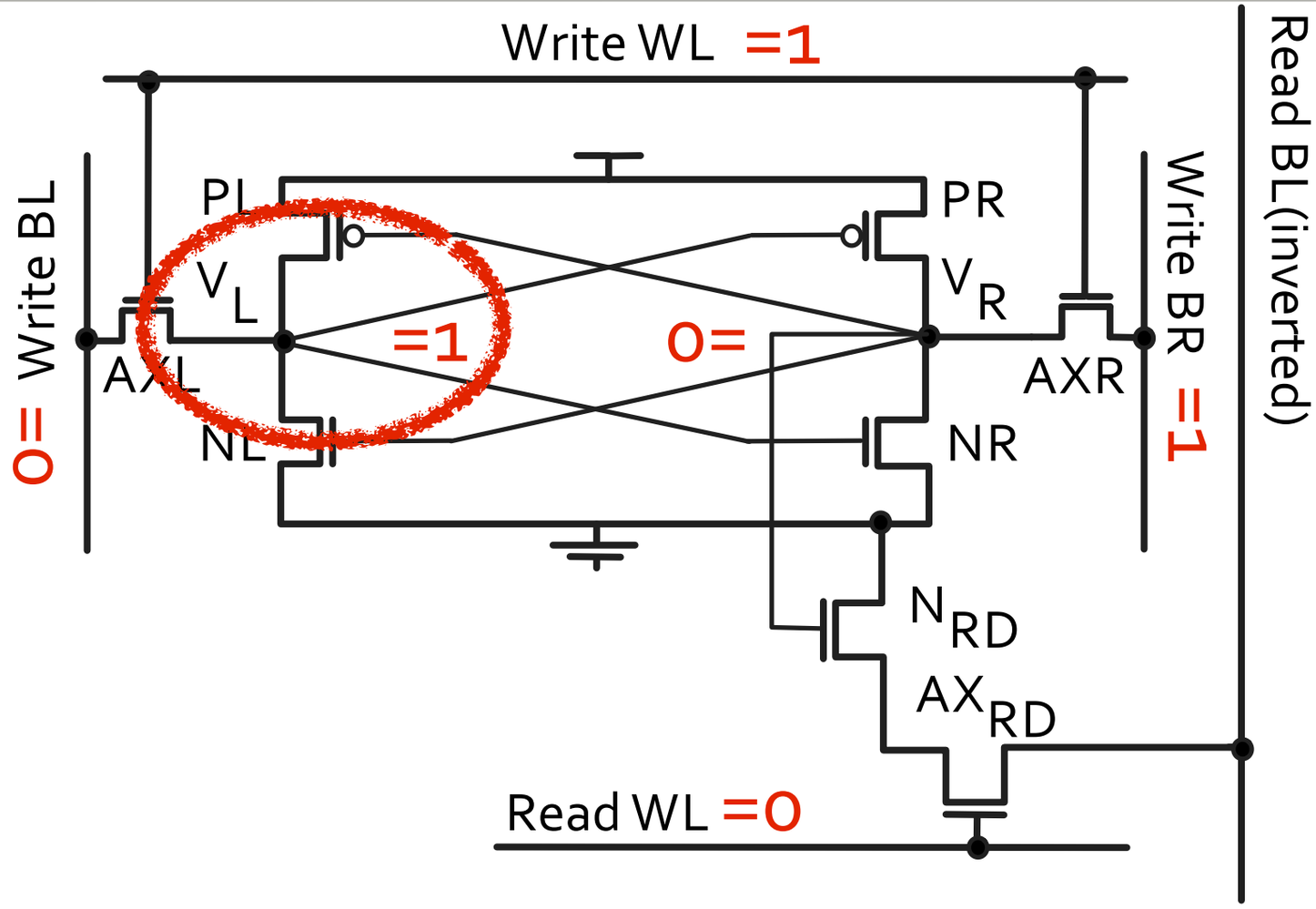
SRAM Timing Analysis



SRAM Timing Analysis

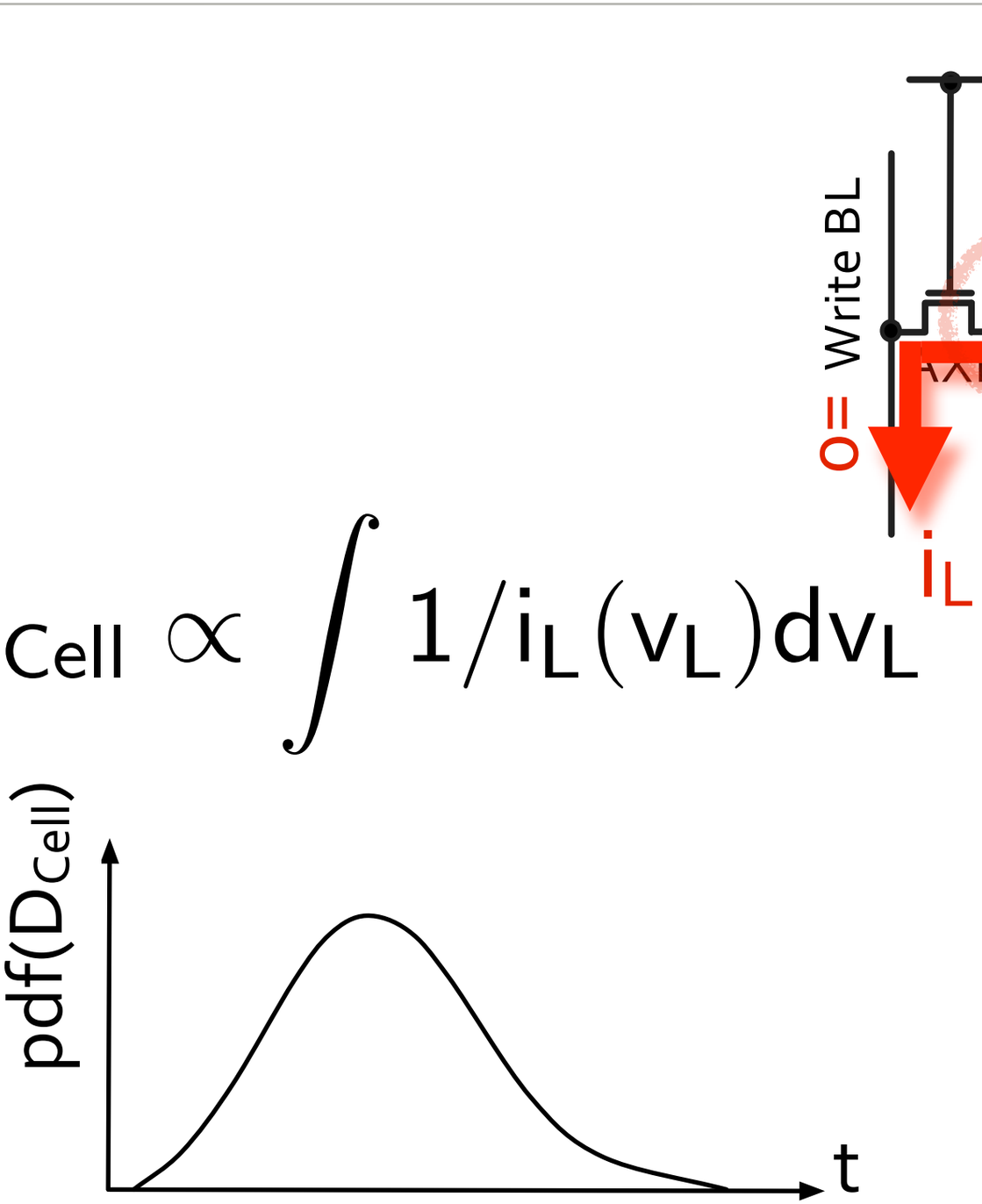


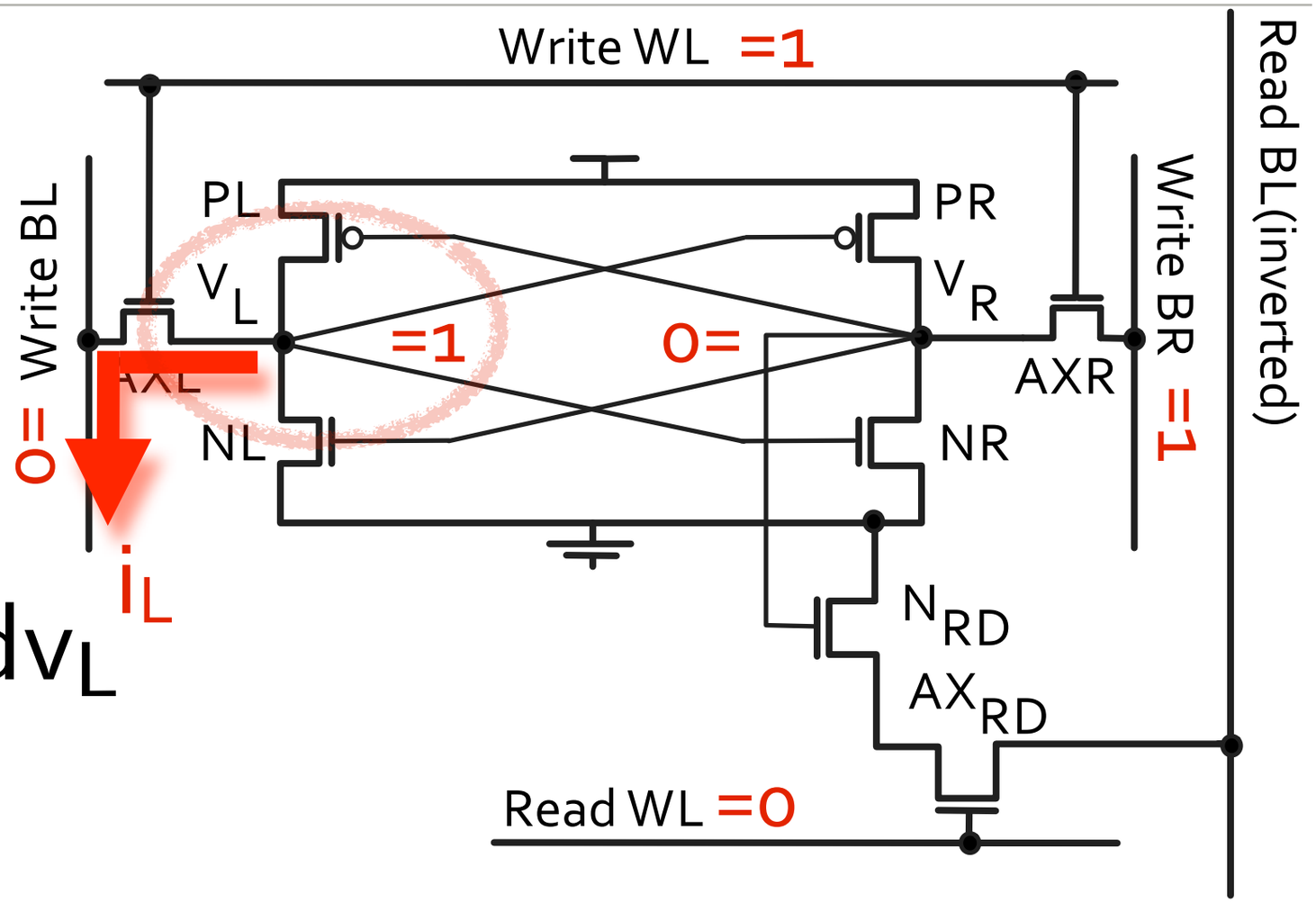
SRAM Timing Analysis



SRAM Timing Analysis

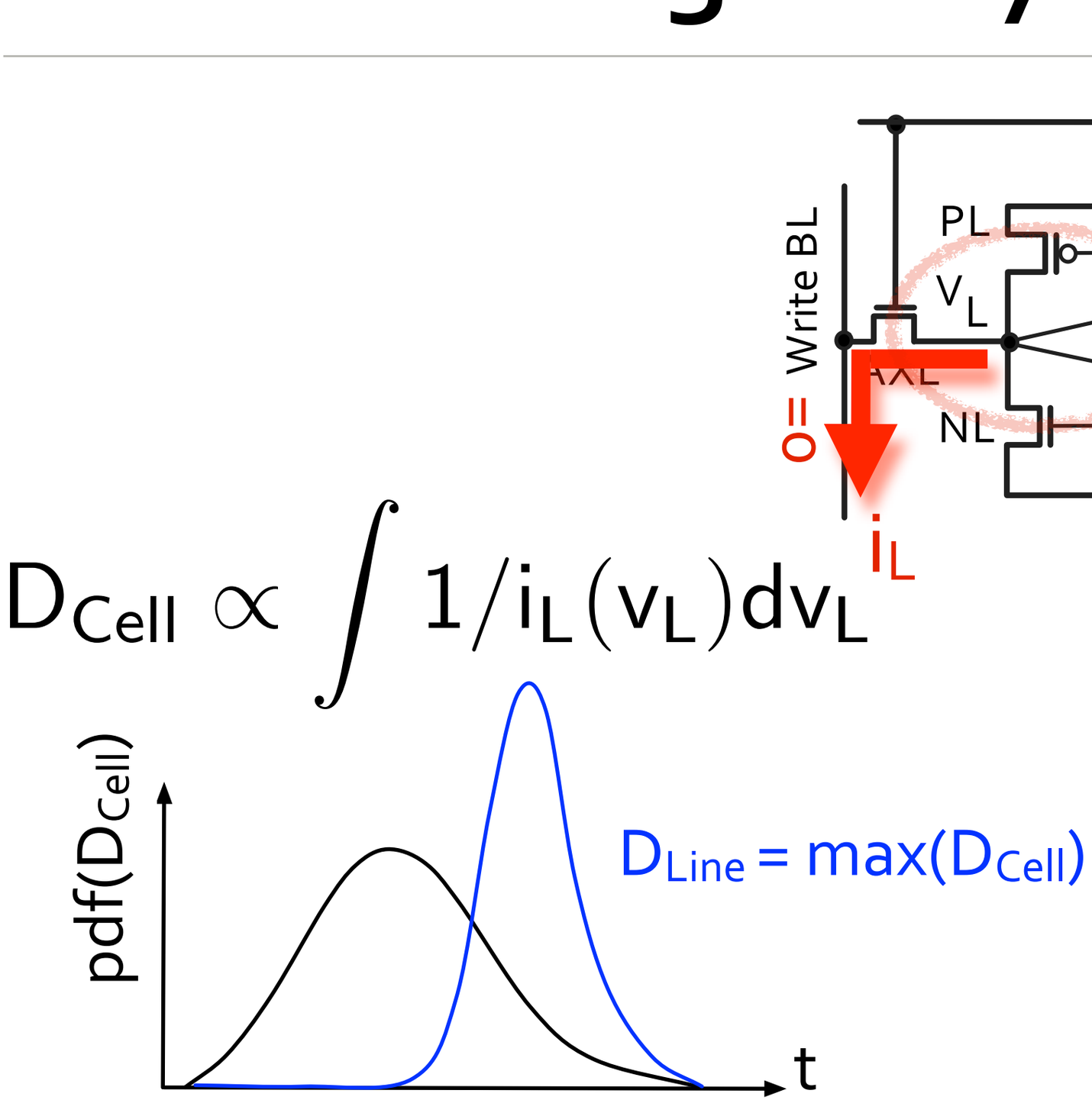
$$D_{Cell} \propto \int 1/i_L(v_L)dv_L$$

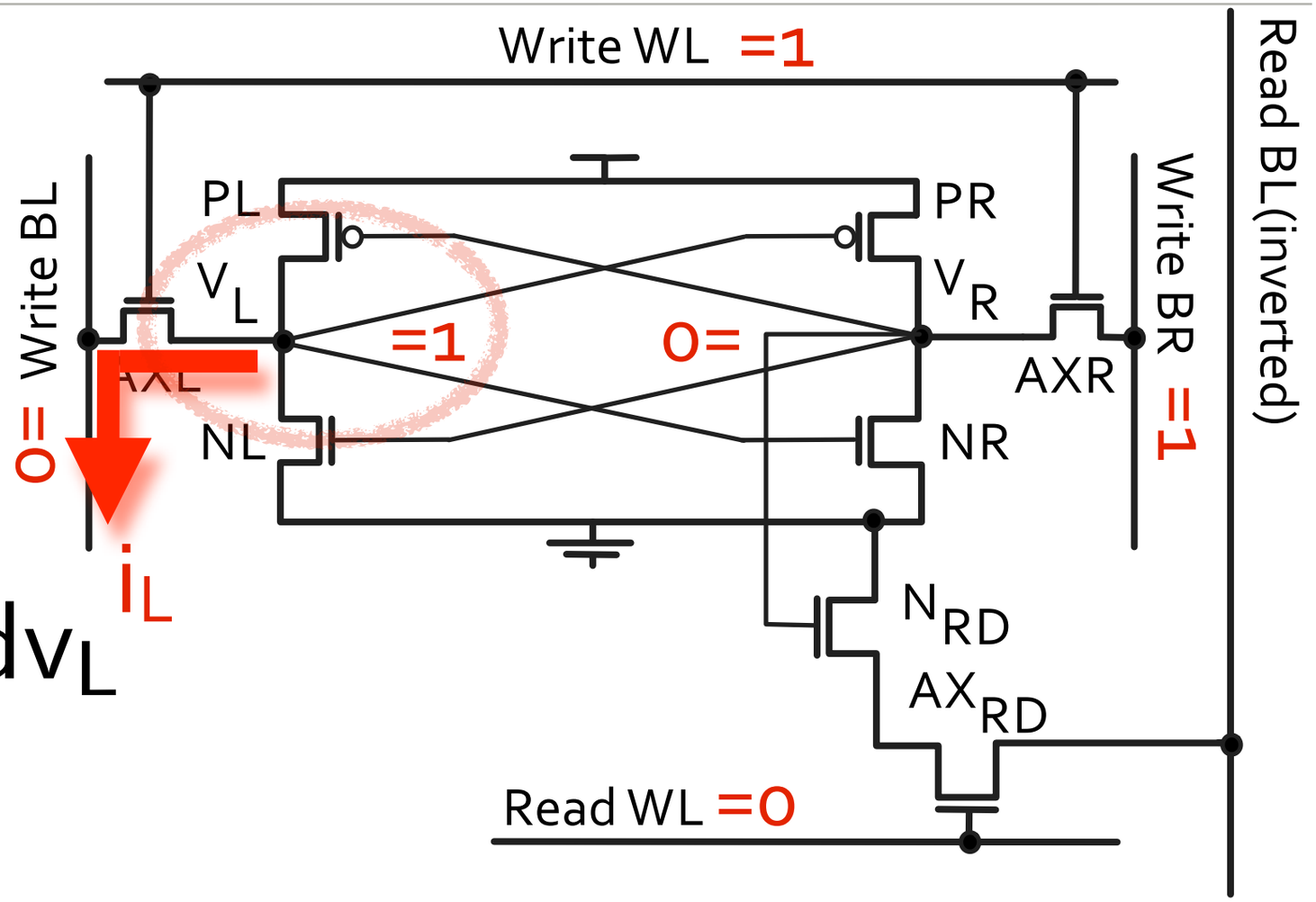




SRAM Timing Analysis

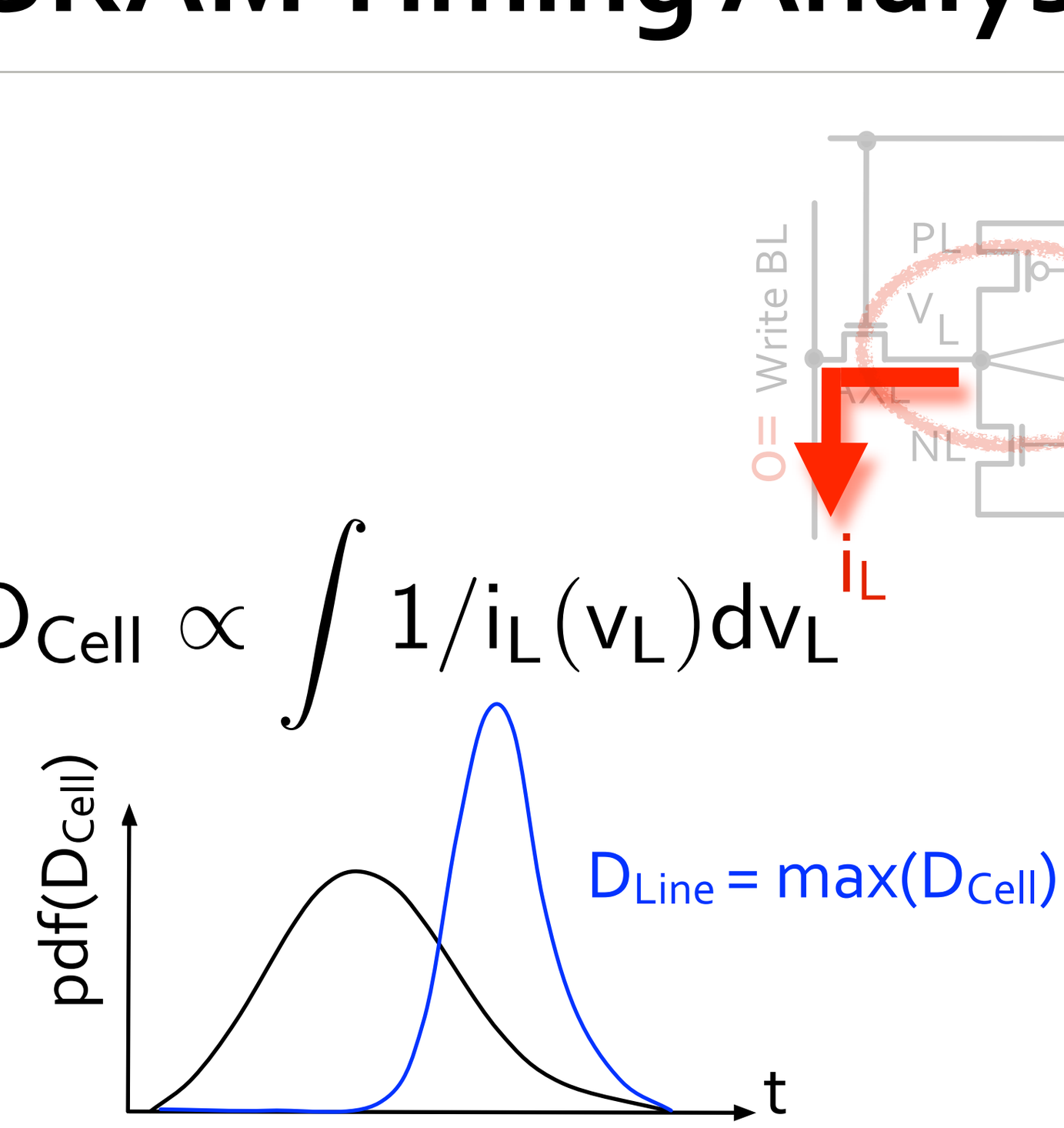
$$D_{Cell} \propto \int 1/i_L(v_L)dv_L$$

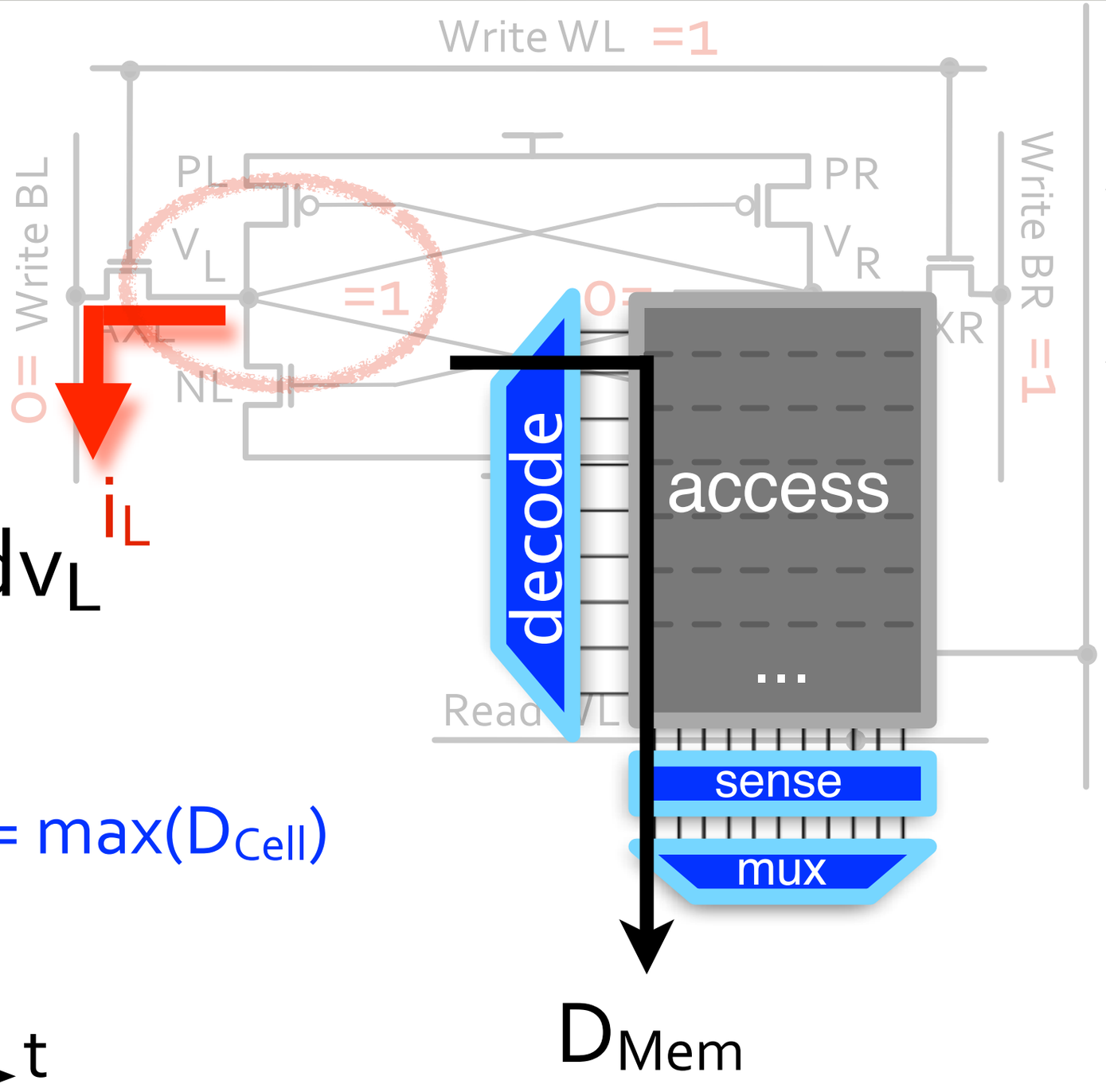




SRAM Timing Analysis

$$D_{Cell} \propto \int 1/i_L(v_L)\,dv_L$$





VARIUS-NTV: Summary

VARIUS-NTV: Summary

Gate Delay Model	EKV Based

VARIUS-NTV: Summary

Gate Delay Model	EKV Based
SRAM Cell Architecture	8T

VARIUS-NTV: Summary

Gate Delay Model	EKV Based
SRAM Cell Architecture	8T
SRAM Failure Modes	Hold Write Stability Write Timing Read Timing

VARIUS-NTV: Summary

Gate Delay Model	EKV Based
SRAM Cell Architecture	8T
SRAM Failure Modes	Hold Write Stability Write Timing Read Timing
Impact of Leakage	✓

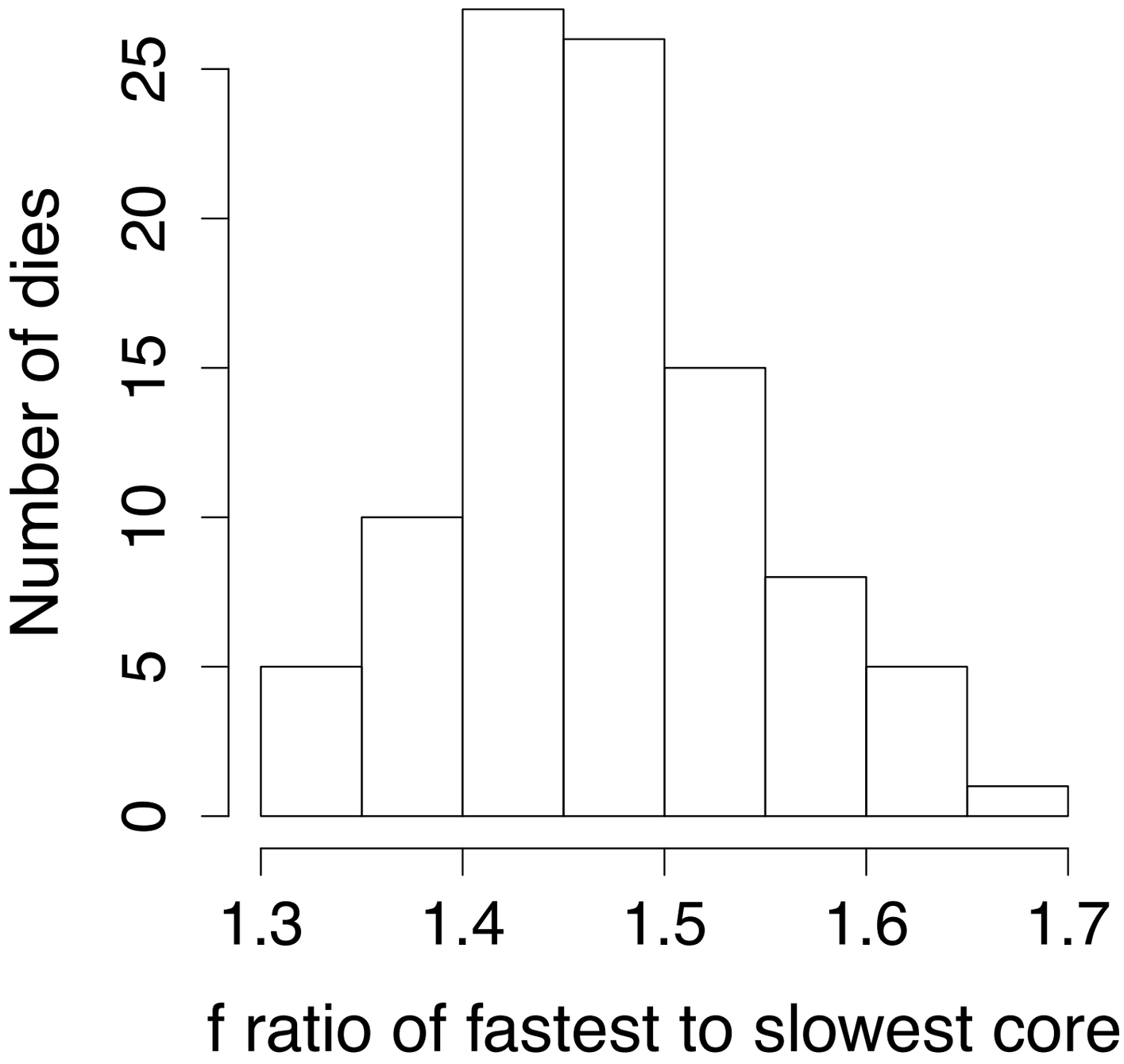
VARIUS-NTV: Validation

VARIUS-NTV: Validation

Validated against Intel 80-core TeraFLOPS

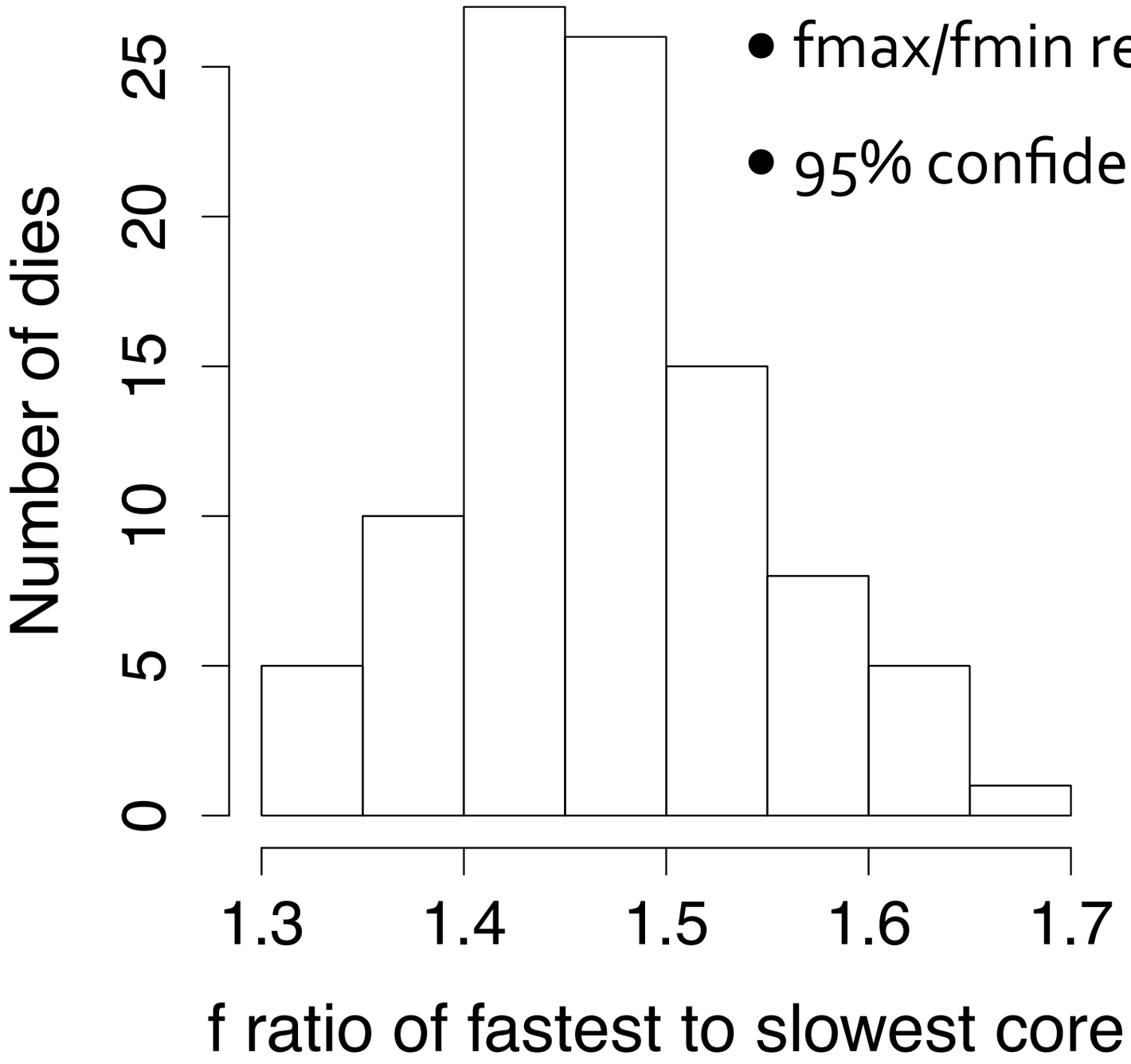
VARIUS-NTV: Validation

Validated against Intel 80-core TeraFLOPS



VARIUS-NTV: Validation

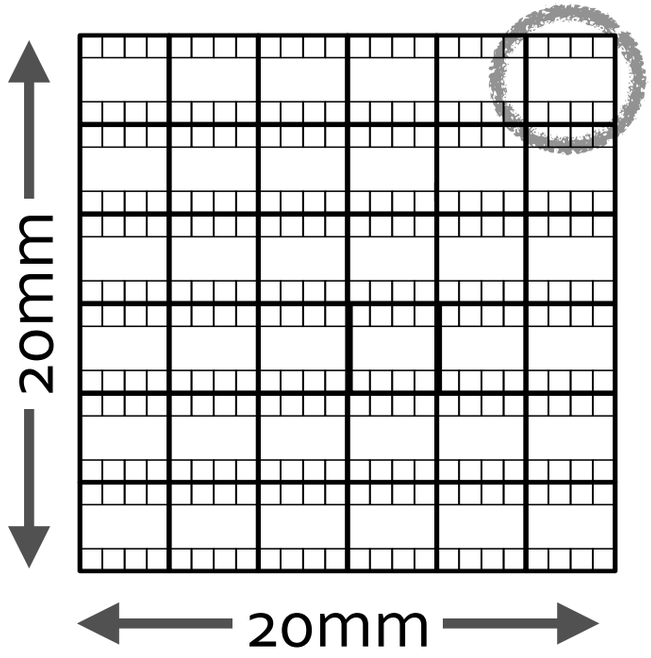
Validated against Intel 80-core TeraFLOPS

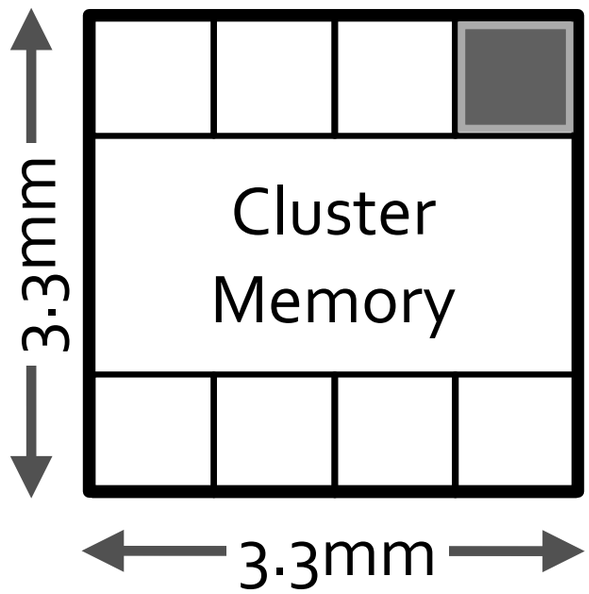


- fmax/fmin reported = 1.6
- 95% confidence interval from VARIUS-NTV = [1.45,1.48]

Evaluation Setup

- 288 core chip:
  - 36 clusters, 8 cores per cluster
  - Core: Single issue in-order
- 11nm process

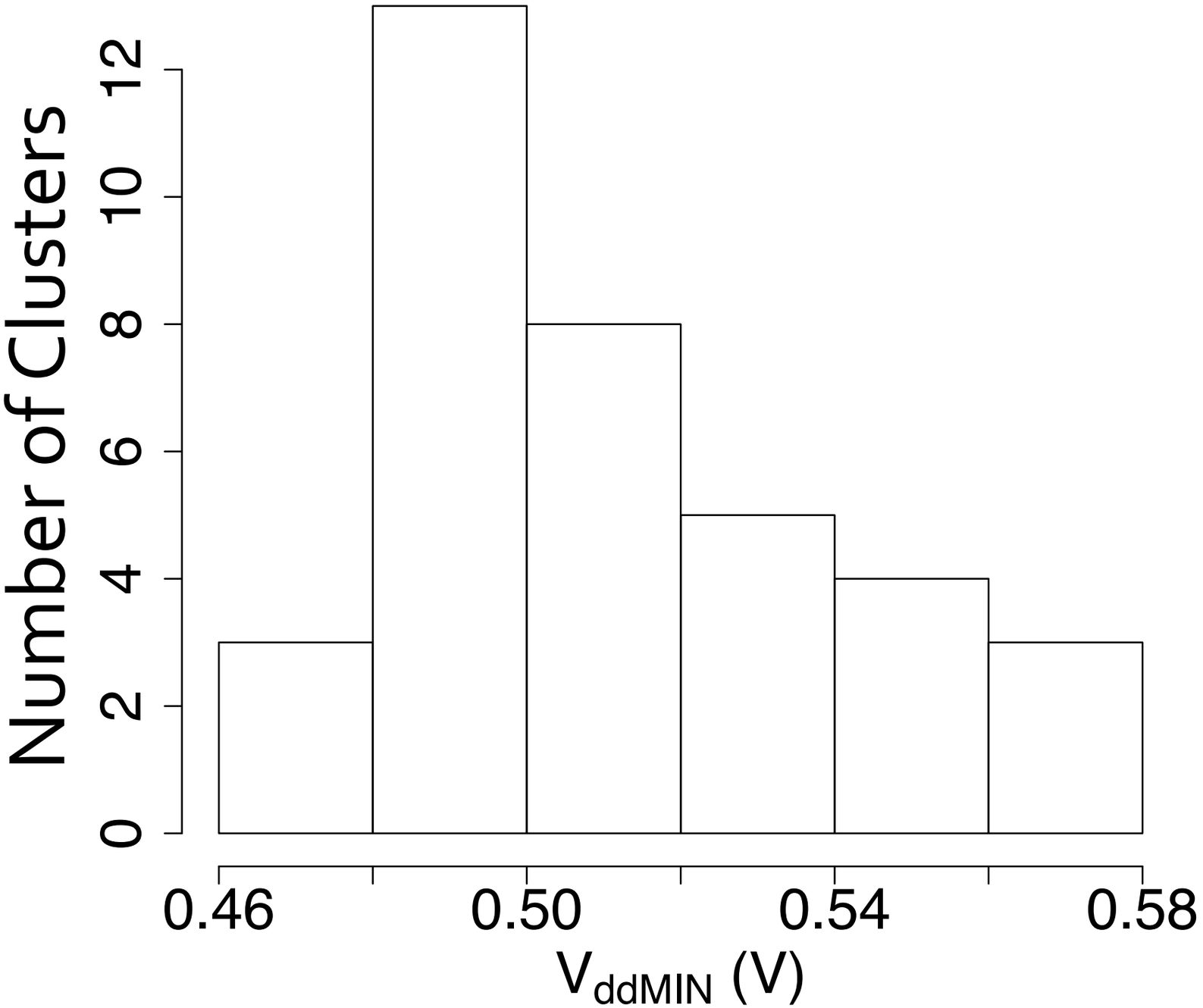




Core + Local Memory

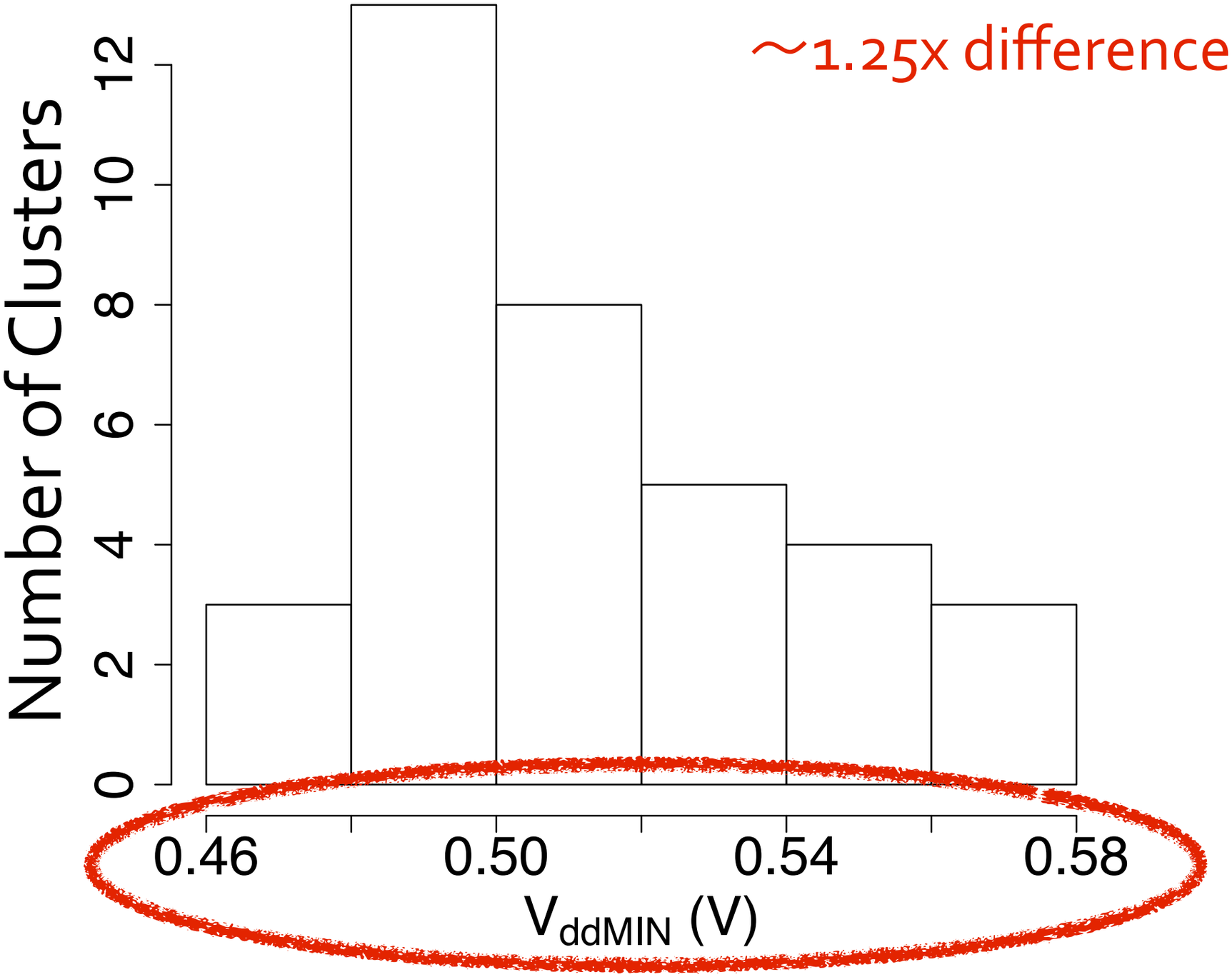
Variation in V_{ddMIN}

Variation in V_{ddMIN}

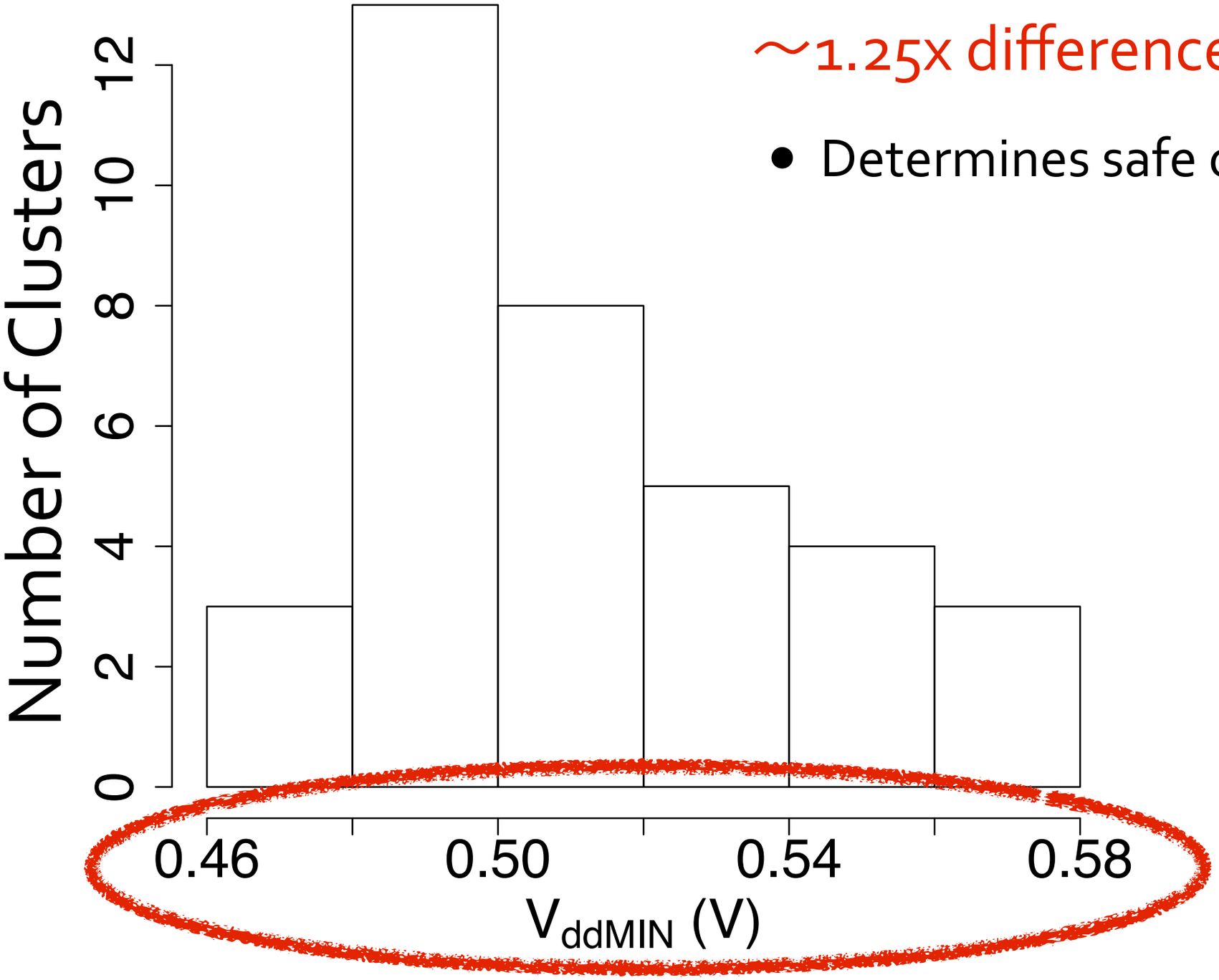


Variation in V_{ddMIN}

~1.25x difference across clusters



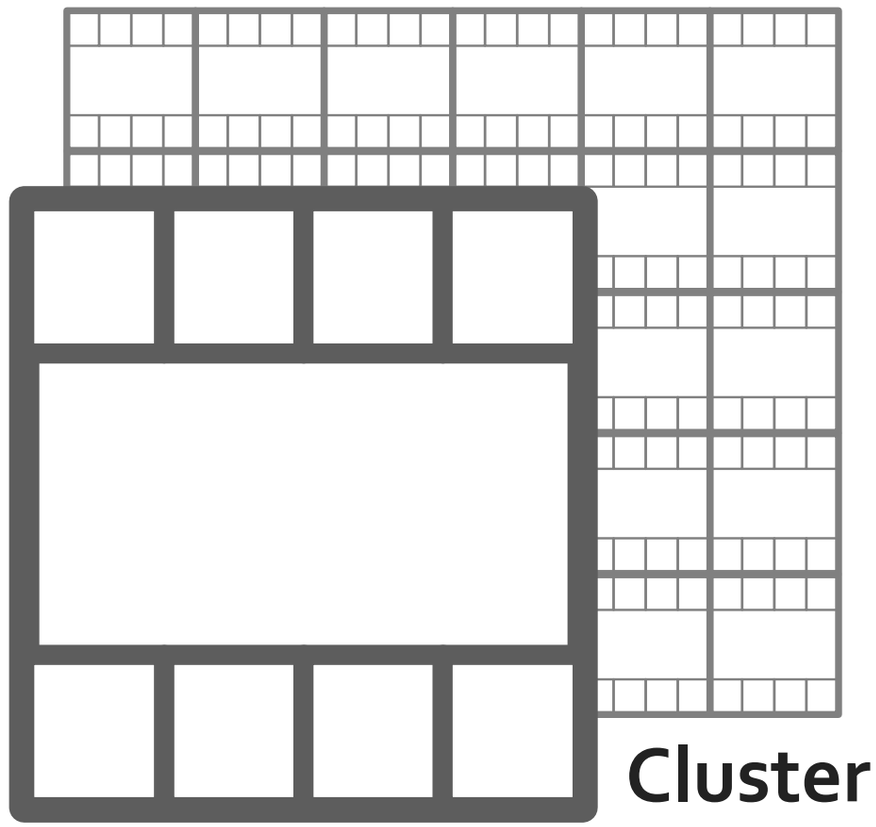
Variation in V_{ddMIN}



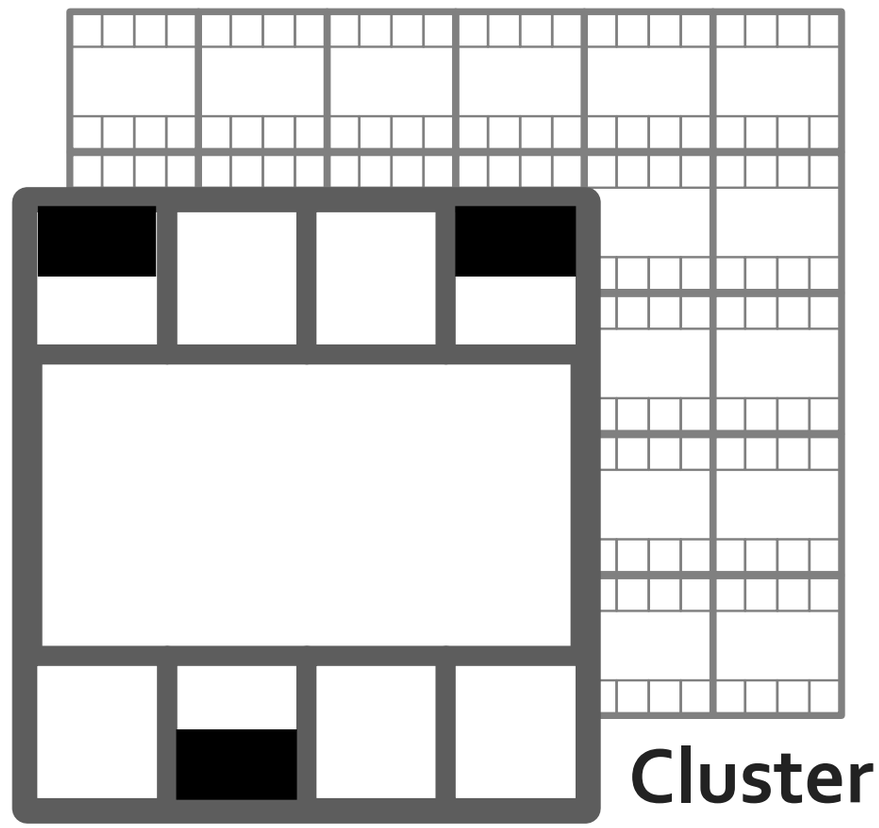
~1.25x difference across clusters

- Determines safe operating Vdd

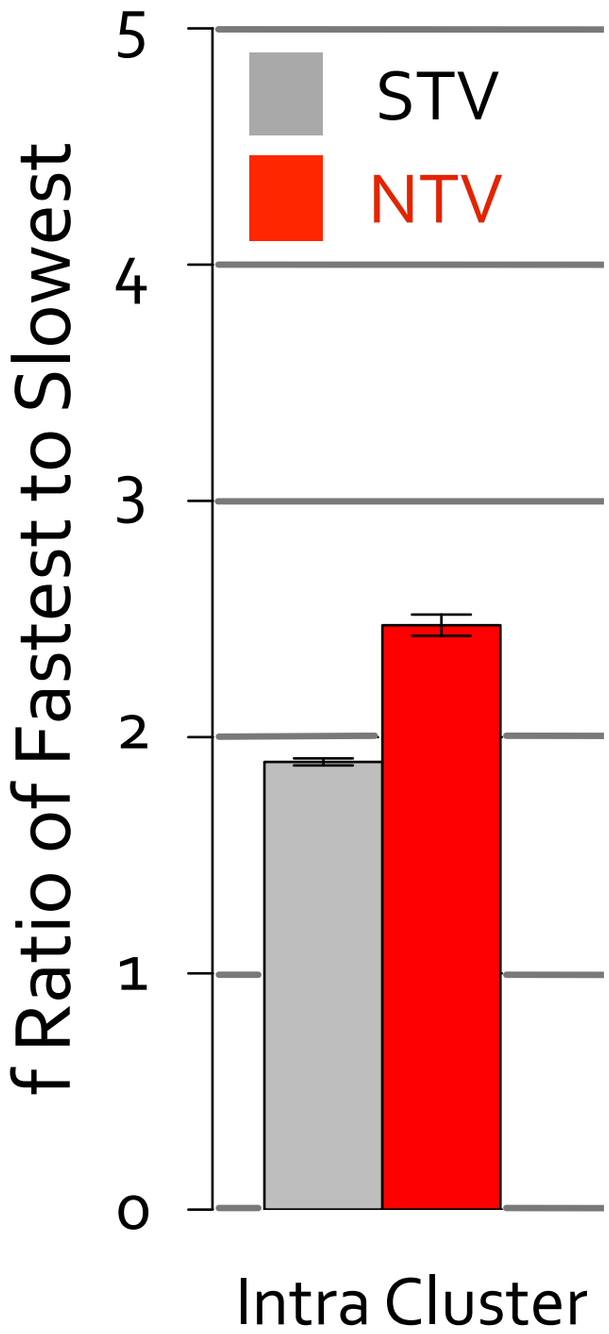
Variation in Frequency

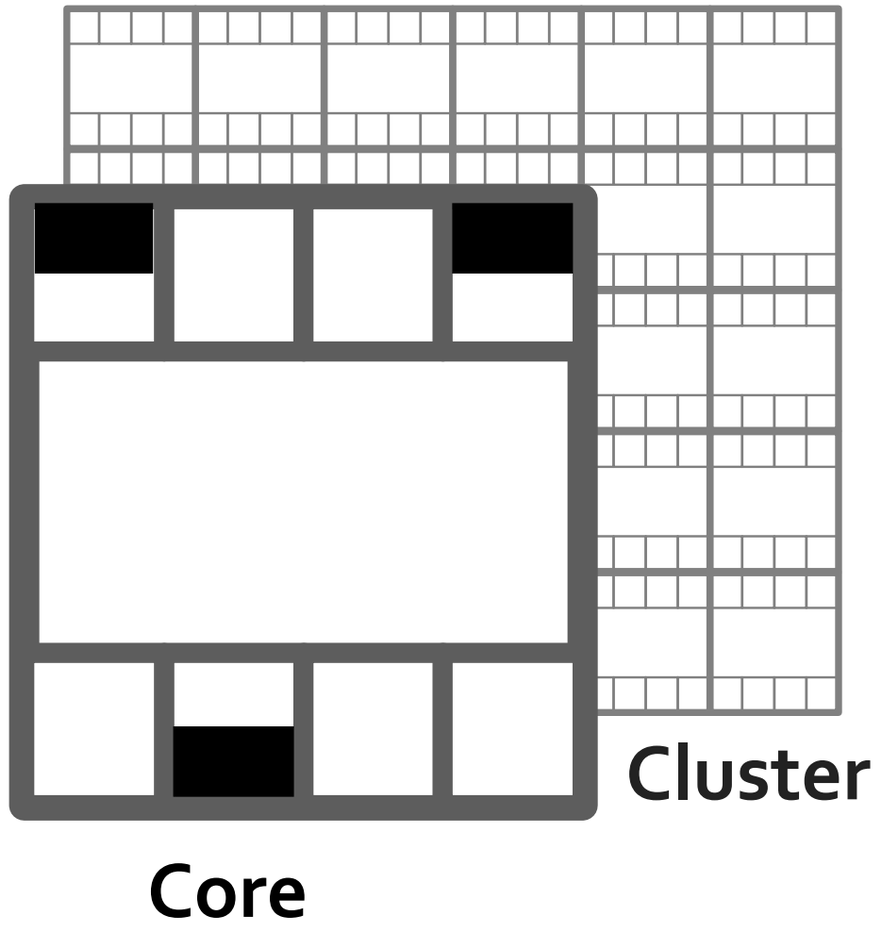


Variation in Frequency

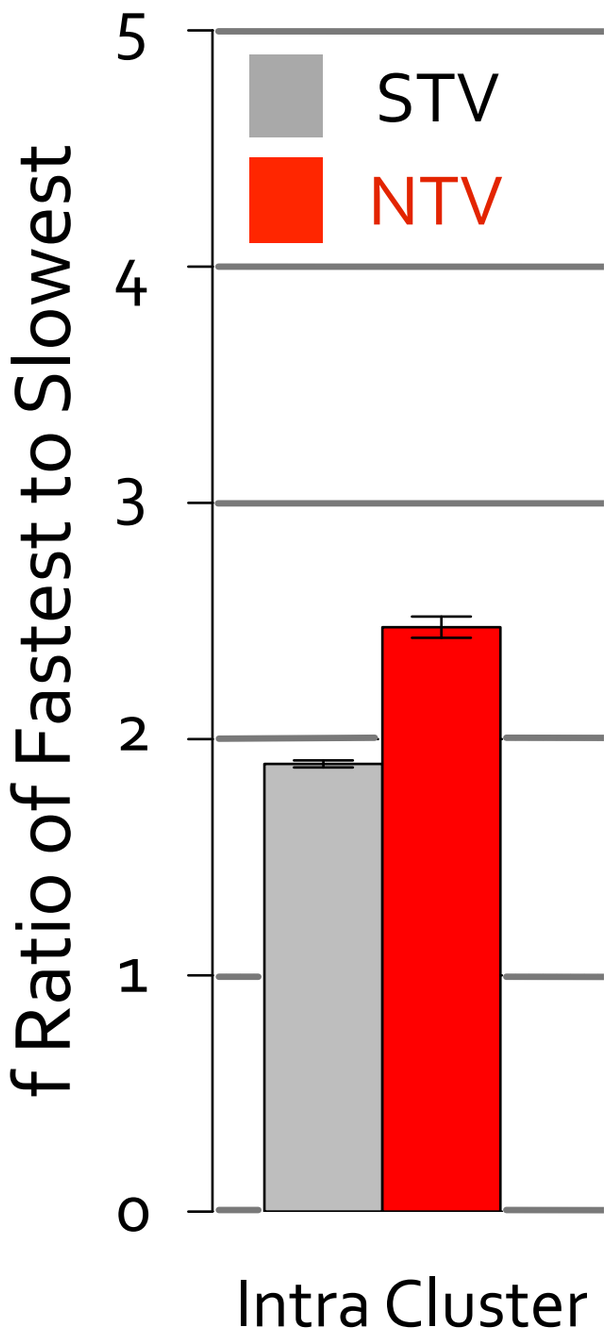


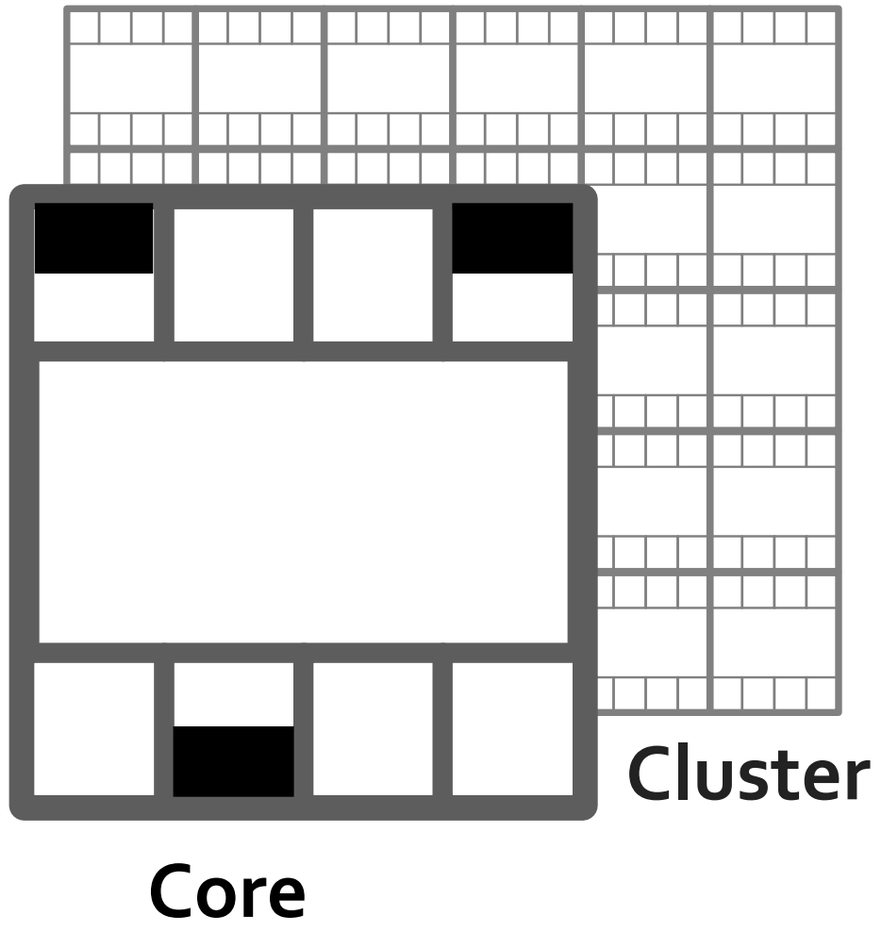
Variation in Frequency





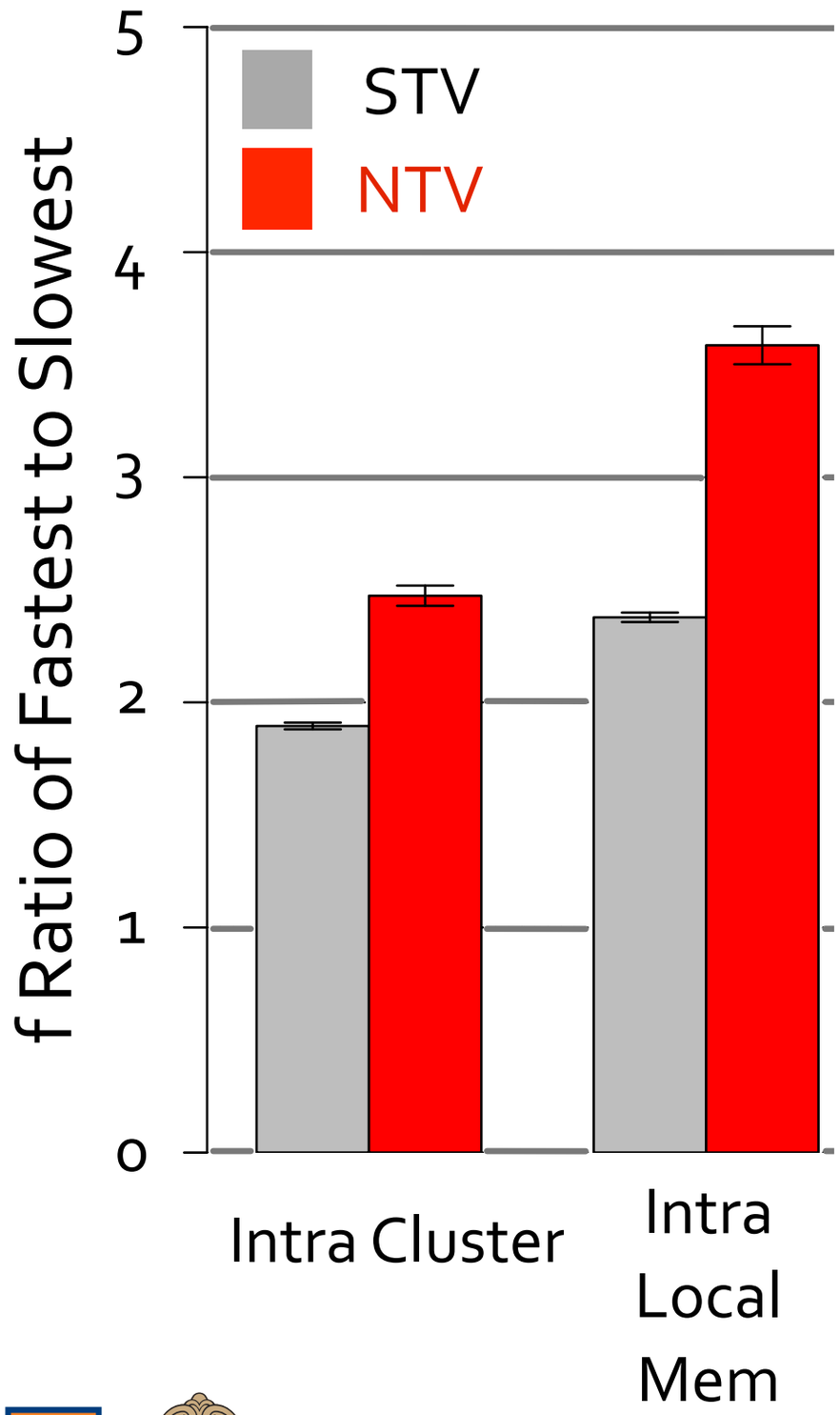
Variation in Frequency

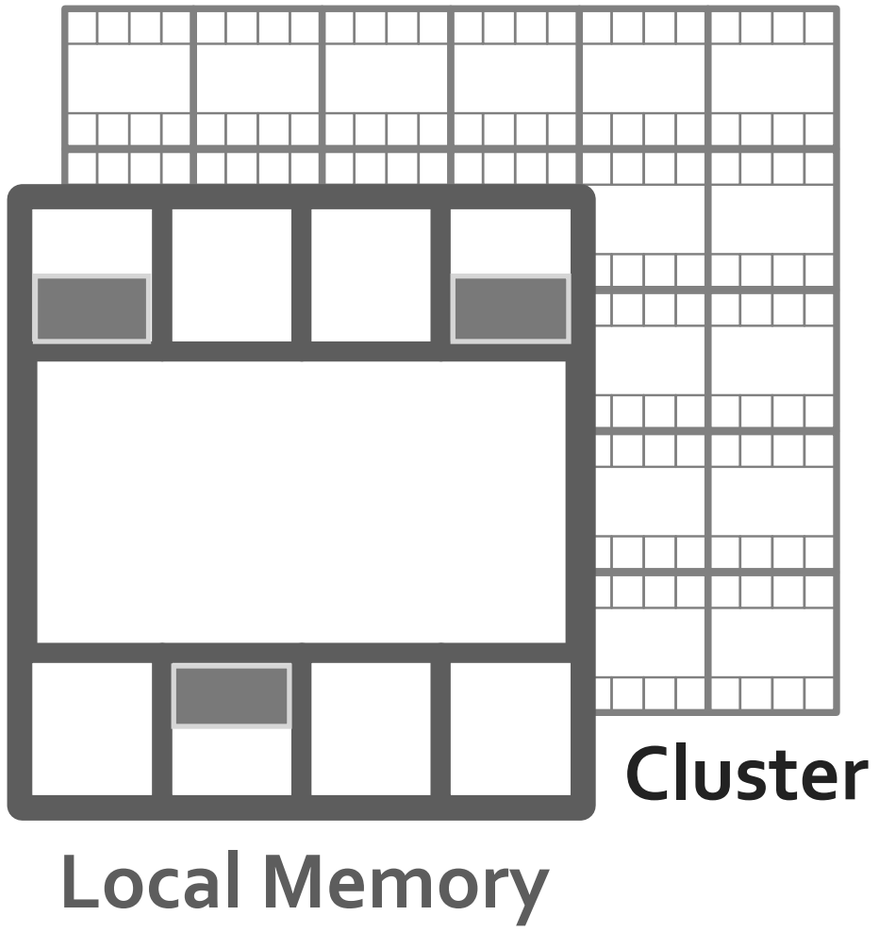




- Larger f variation at NTV
  - 2.5x (NTV) vs. 1.9x (STV)

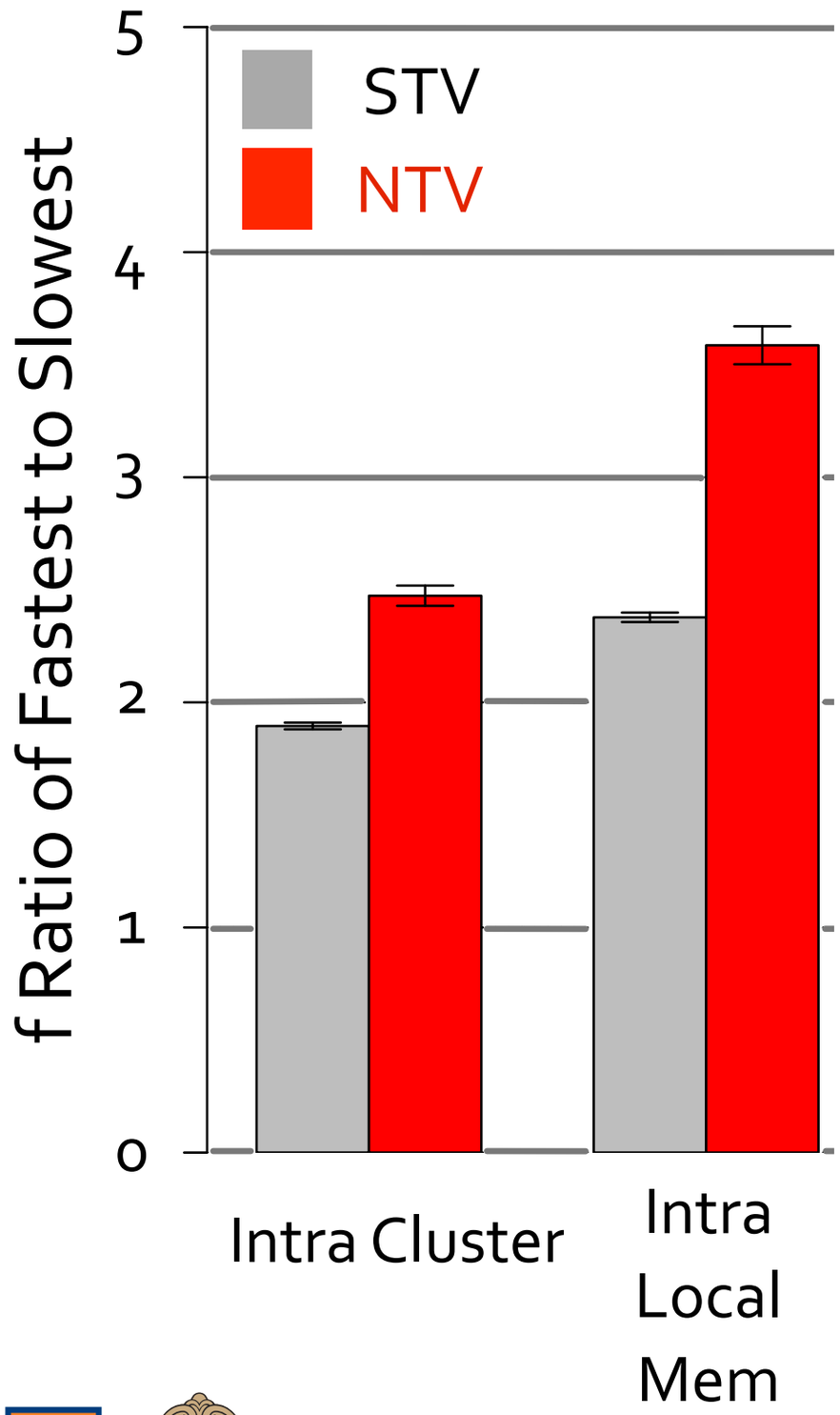
Variation in Frequency

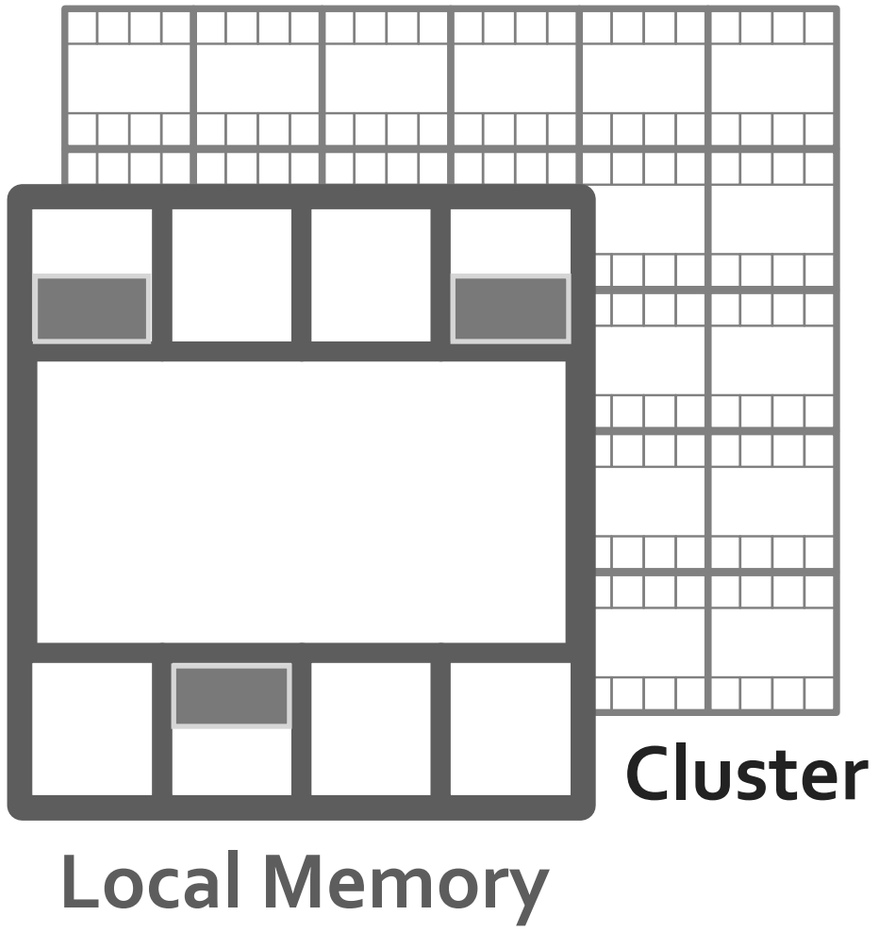




- Larger f variation at NTV
  - 2.5x (NTV) vs. 1.9x (STV)

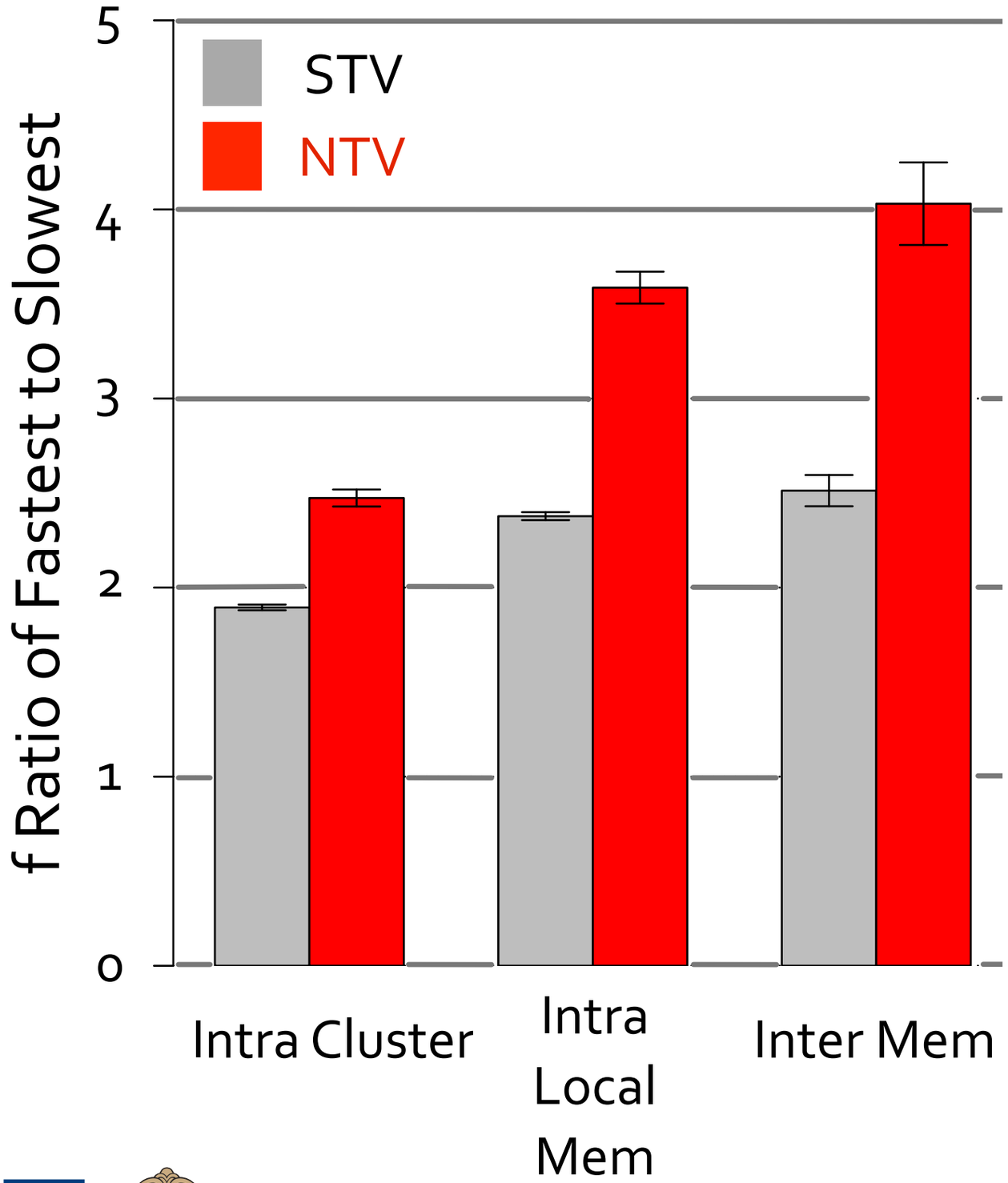
Variation in Frequency

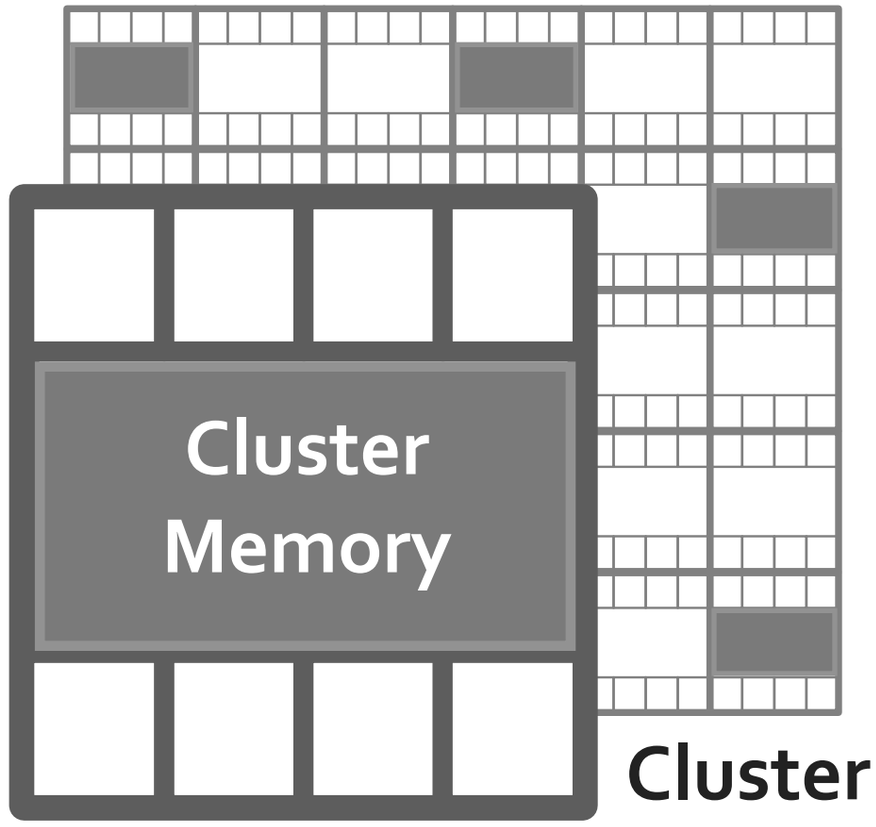




- Larger f variation at NTV
  - 2.5x (NTV) vs. 1.9x (STV)
- Memories more vulnerable

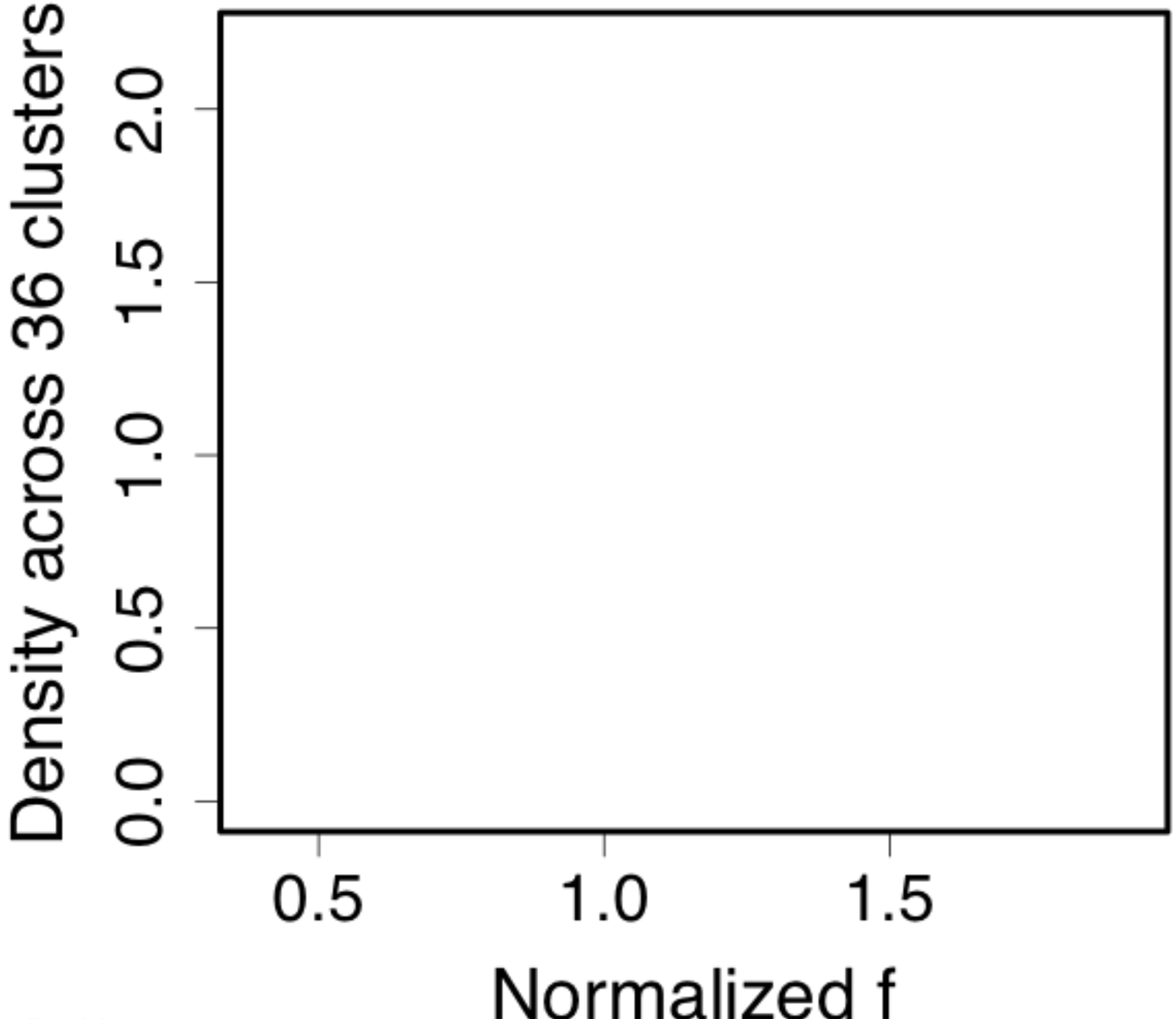
Variation in Frequency



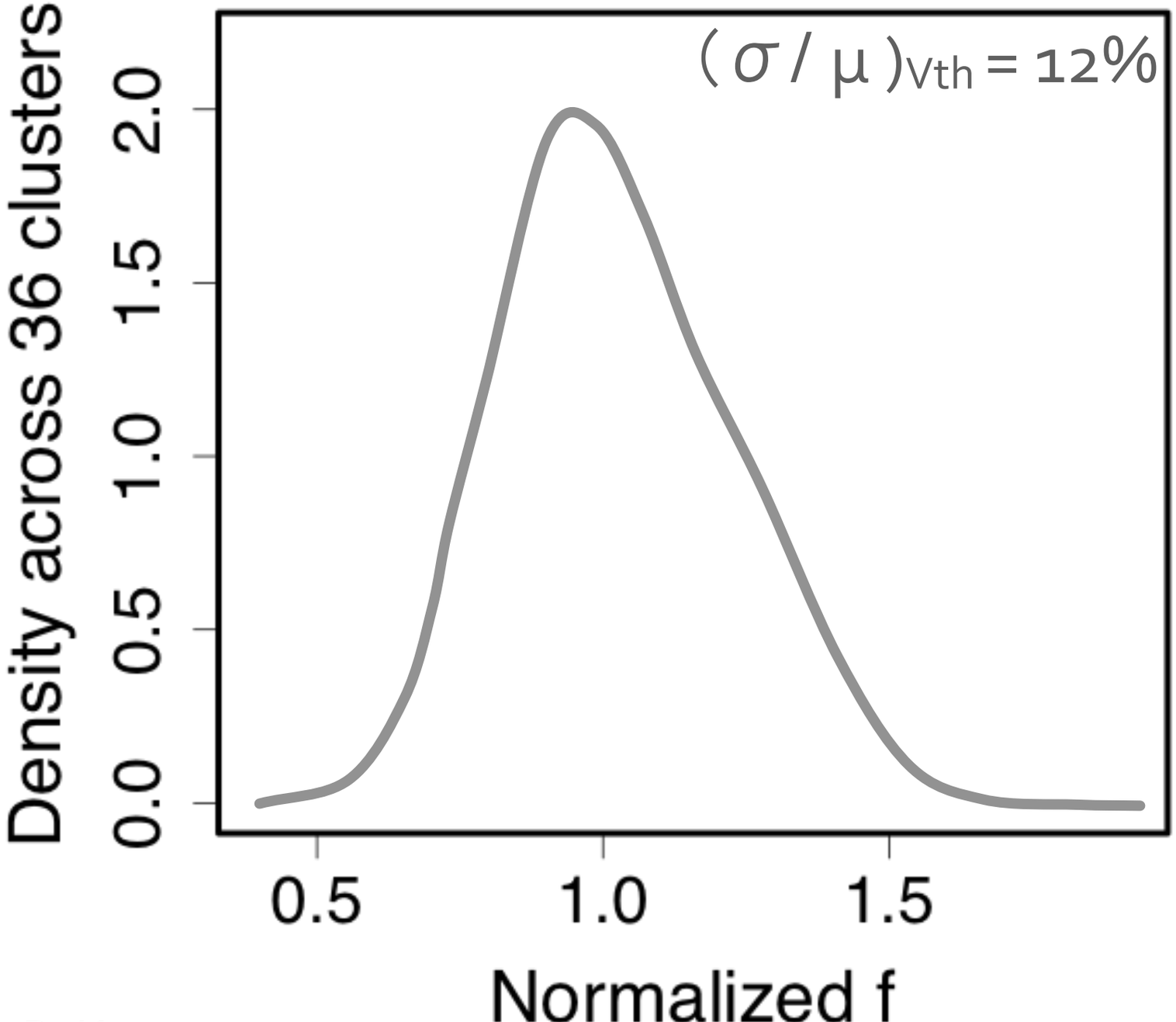


- Larger f variation at NTV
  - 2.5x (NTV) vs. 1.9x (STV)
- Memories more vulnerable

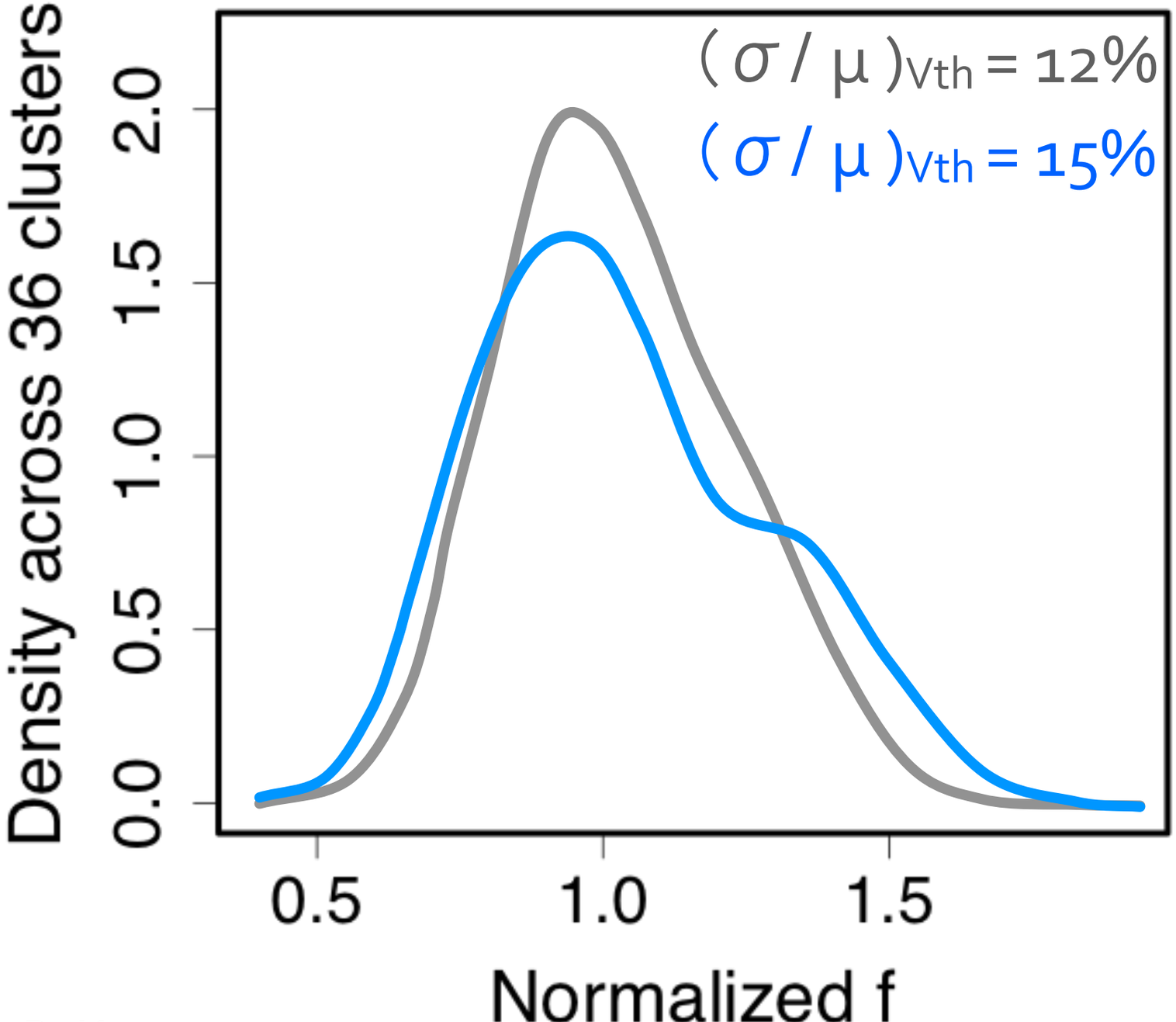
Variation in Frequency



Variation in Frequency

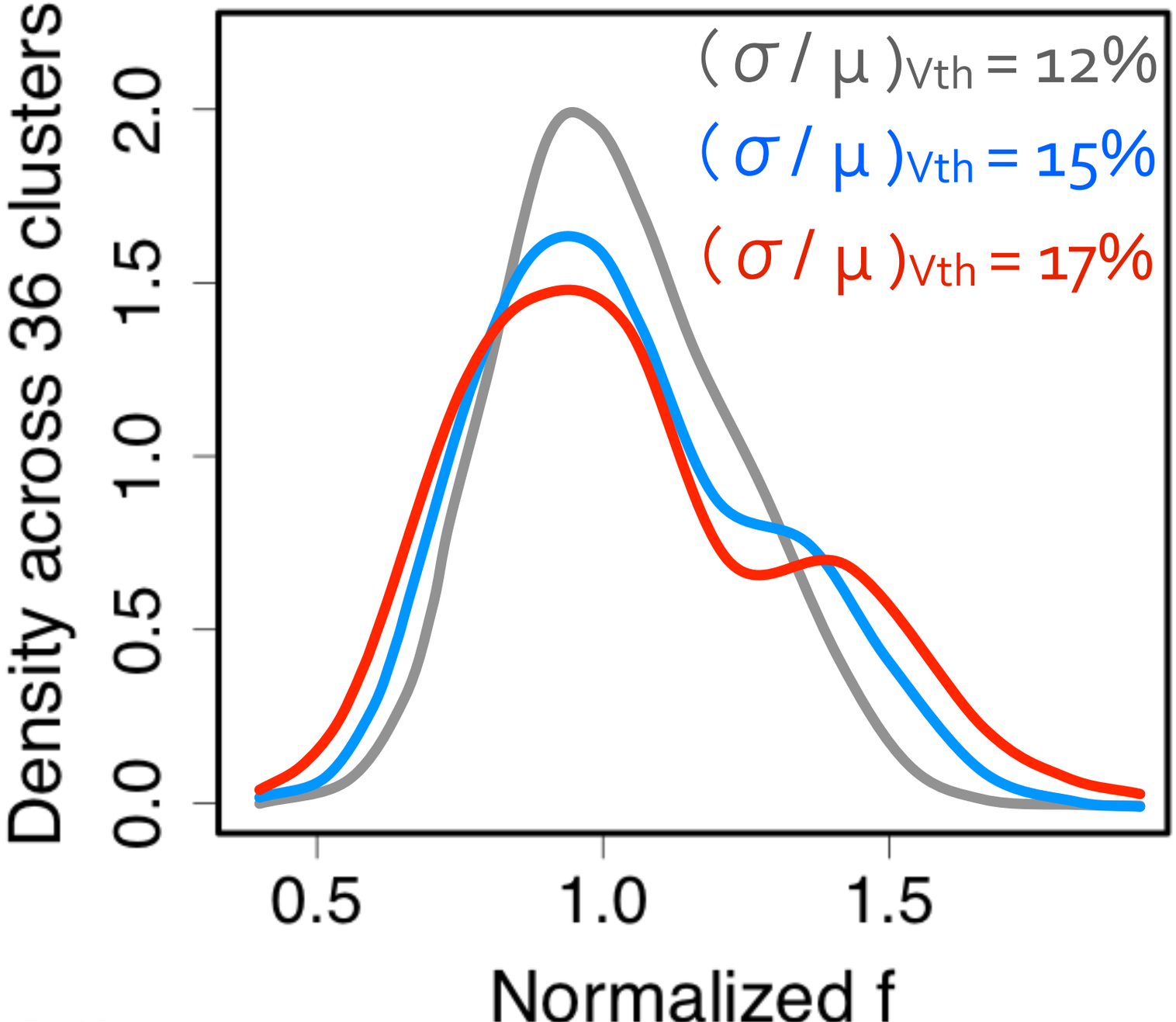


Variation in Frequency



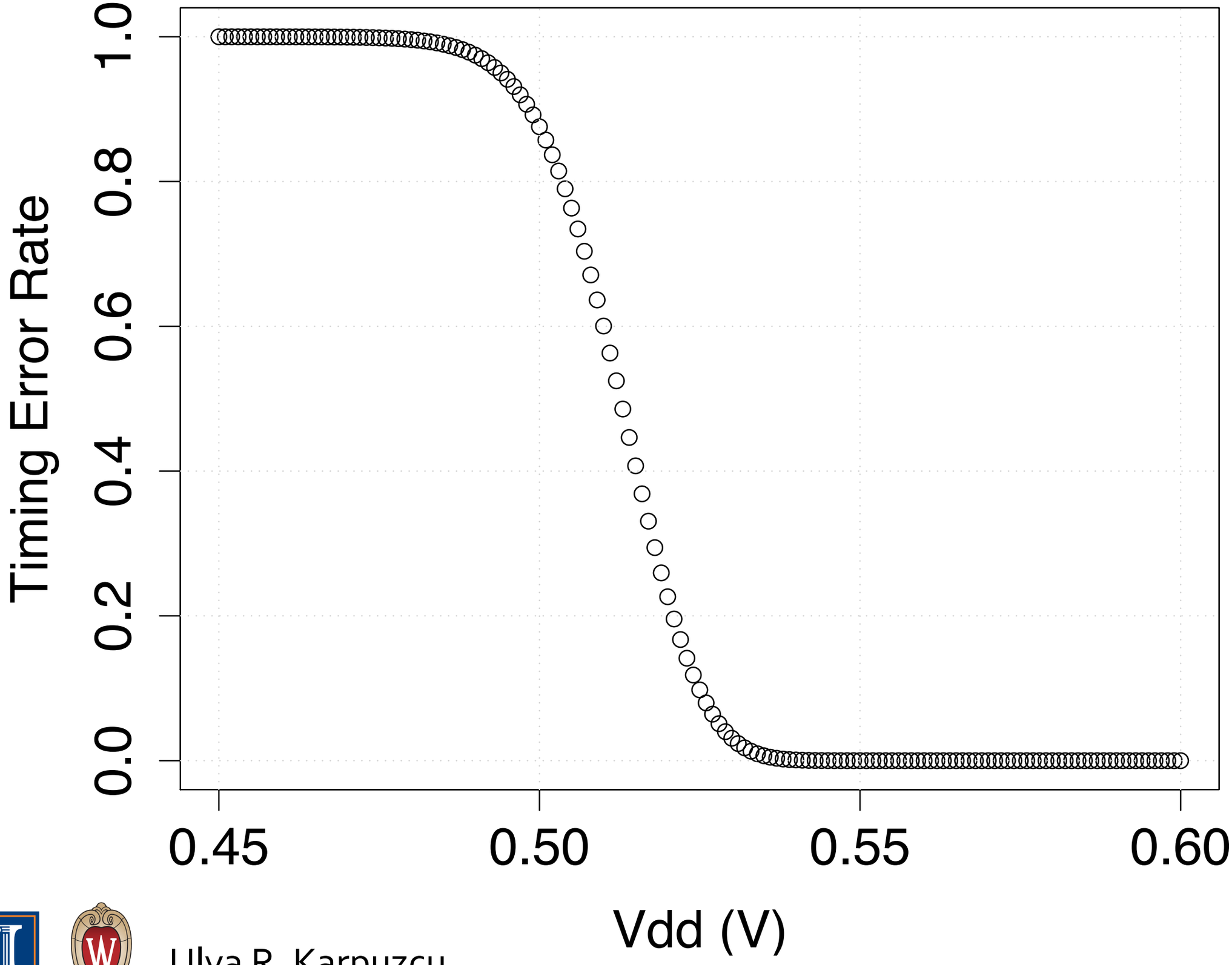
Variation in Frequency

Variation in Vth ↑ ⇒ The spread of distribution ↑

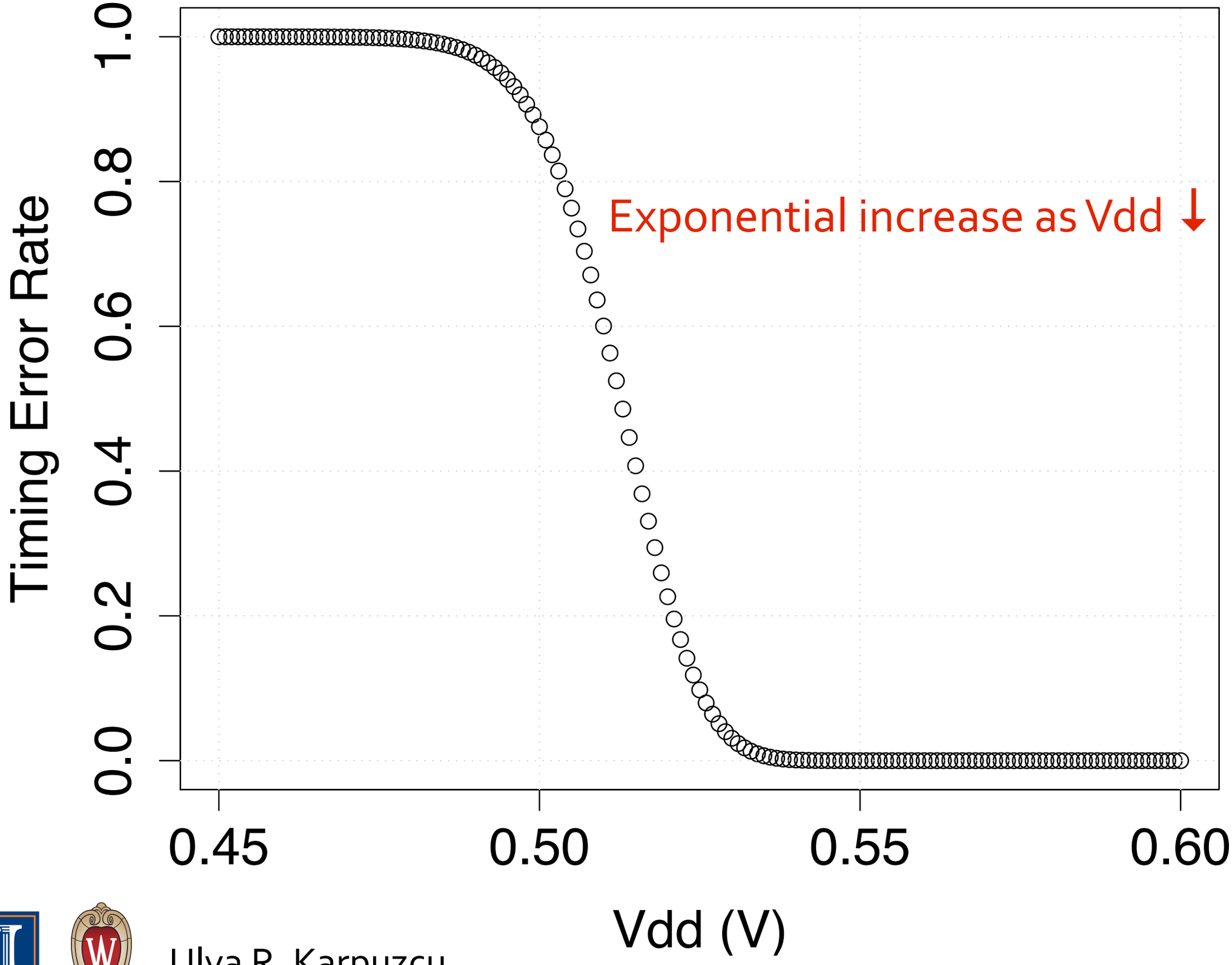


Timing Error Rate (Logic)

Timing Error Rate (Logic)



Timing Error Rate (Logic)



Conclusion

Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV

Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV

Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV
- Various SRAM failure modes

Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV
- Various SRAM failure modes
- Consideration of leakage throughout analysis

Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV
- Various SRAM failure modes
- Consideration of leakage throughout analysis
- Useful for a variety of (μ)architectural studies


Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV
- Various SRAM failure modes
- Consideration of leakage throughout analysis

- Useful for a variety of (μ)architectural studies
  - Design space exploration:
    - Safe Vdd, f for logic and on-chip memory along with induced power


Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV
- Various SRAM failure modes
- Consideration of leakage throughout analysis

- Useful for a variety of (μ)architectural studies
  - Design space exploration:
    - Safe Vdd, f for logic and on-chip memory along with induced power
  - Reliability analysis:
    - Timing/stability error rate for logic and on-chip memory


Conclusion

VARIUS-NTV: A μ-architectural model of variations at NTV

- Gate delay model tailored for NTV
- SRAM cell architecture facilitating robust operation at NTV
- Various SRAM failure modes
- Consideration of leakage throughout analysis

- Useful for a variety of (μ)architectural studies
  - Design space exploration:
    - Safe Vdd, f for logic and on-chip memory along with induced power
  - Reliability analysis:
    - Timing/stability error rate for logic and on-chip memory
- Download: http://iacoma.cs.uiuc.edu/varius/ntv


VARIUS-NTV: A Model of Process Variations at Near-Threshold Voltages

Ulya R. Karpuzcu❖, Krishna Kolluru*, Nam Sung Kim*, Josep Torrellas❖

❖ University of Illinois

* University of Wisconsin

http://iacoma.cs.uiuc.edu/varius/ntv

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON