

Swimming Lane: a Composite Design to Mitigate Voltage Droop Effects in 3D Chips

Xing Hu^{1,2,3}, Yi Xu^{3,4}, Yu Hu¹ and Yuan Xie^{3,5}

¹Institute of Computing Technology, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

³AMD Research China Lab

⁴Macau University of Science and Technology

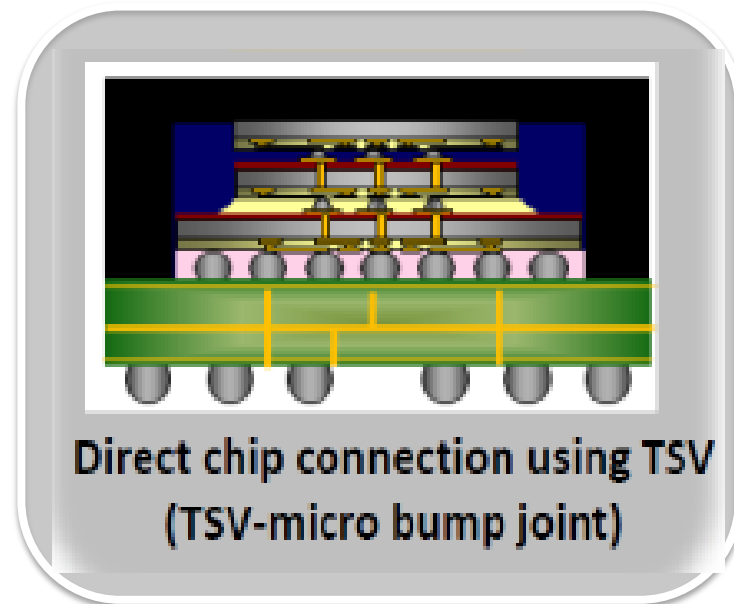
⁵Pennsylvania State University

Outline

- Background and Observation
- Layer independent failsafe design
- Violent-first Scheduling
- Experimental Results
- Conclusions

Background

- Chance of 3d integration:



35% Smaller
Package Size

50% Less
Power
Consumption

8X
Bandwidth
improvement

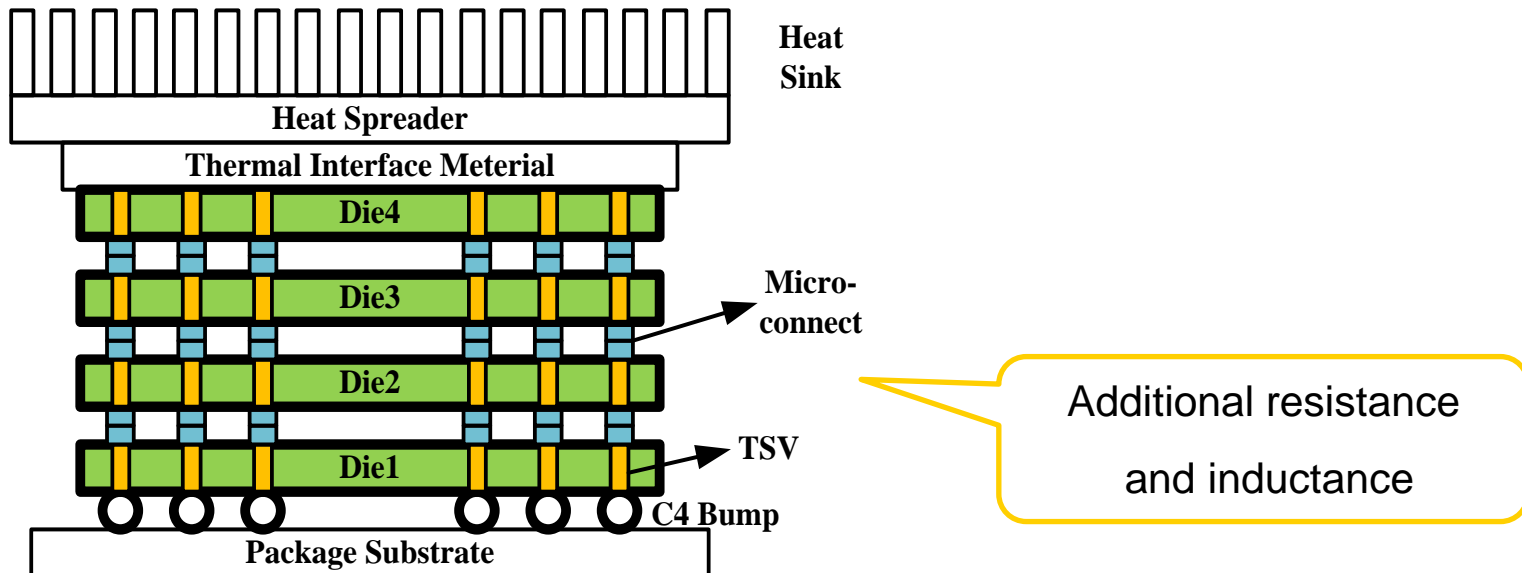
Multiple
process
integration

Power Integrity Challenges

■ Power integrity challenges

- Large load current
 - High device density
 - Limited pins
- Long power delivery path

Large voltage noise



[6] J Gu, et. al. "Multi-story power delivery for supply noise reduction and low voltage operation", *ISLPED 05*.

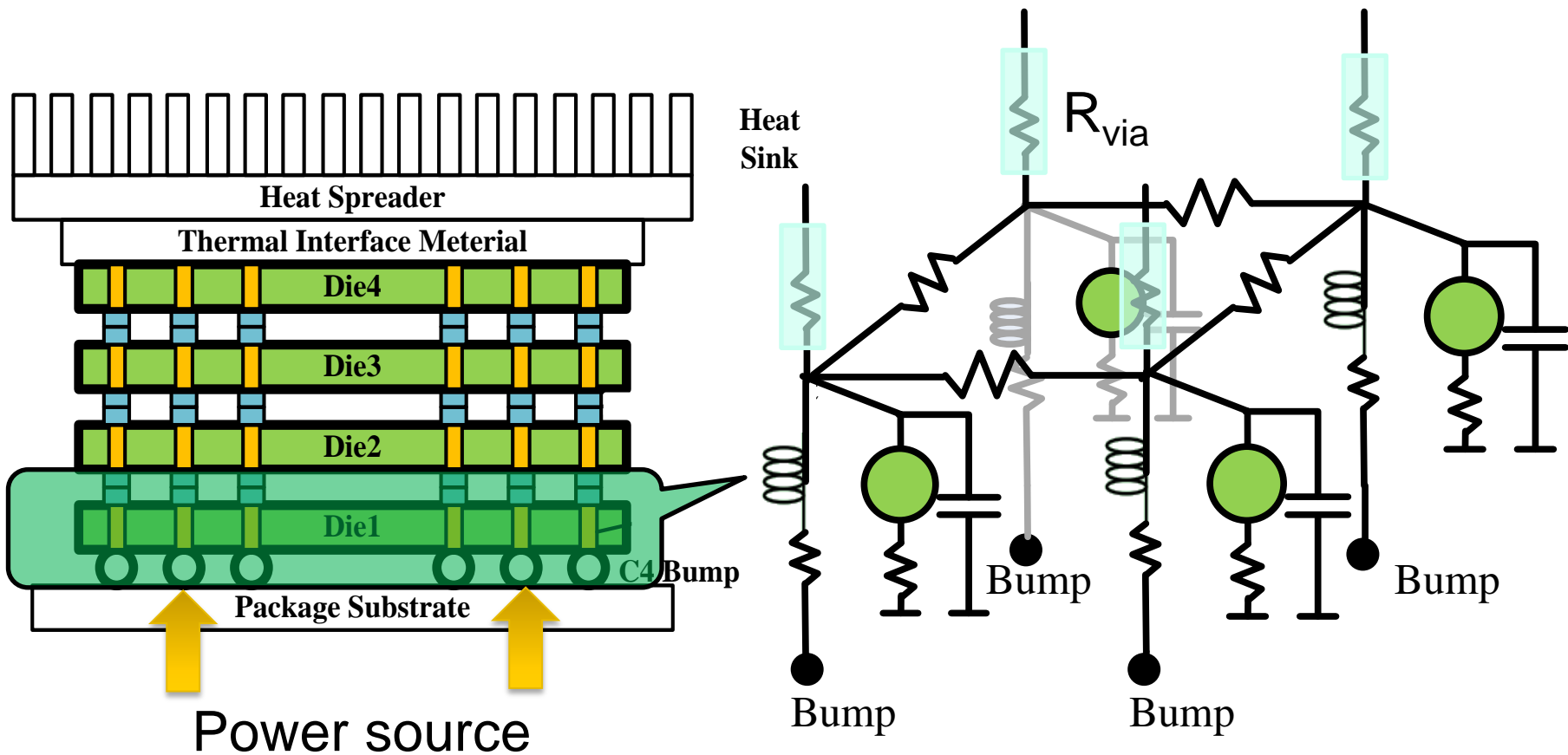
Power Integrity in 3D stacked chips

- Prior mitigation technologies
 - Static physical design
 - Increase power pads or decoupling capacity [1]
 - P/G TSV planning [2]
- Our work
 - Dynamical run-time mitigation
 - Flexible and low-cost

[1]. Taigon Song, et. al., “A Fine-Grained Co-Simulation Methodology for IR-drop Noise in Silicon Interposer and TSV-based 3D IC”, EPEPS 2011

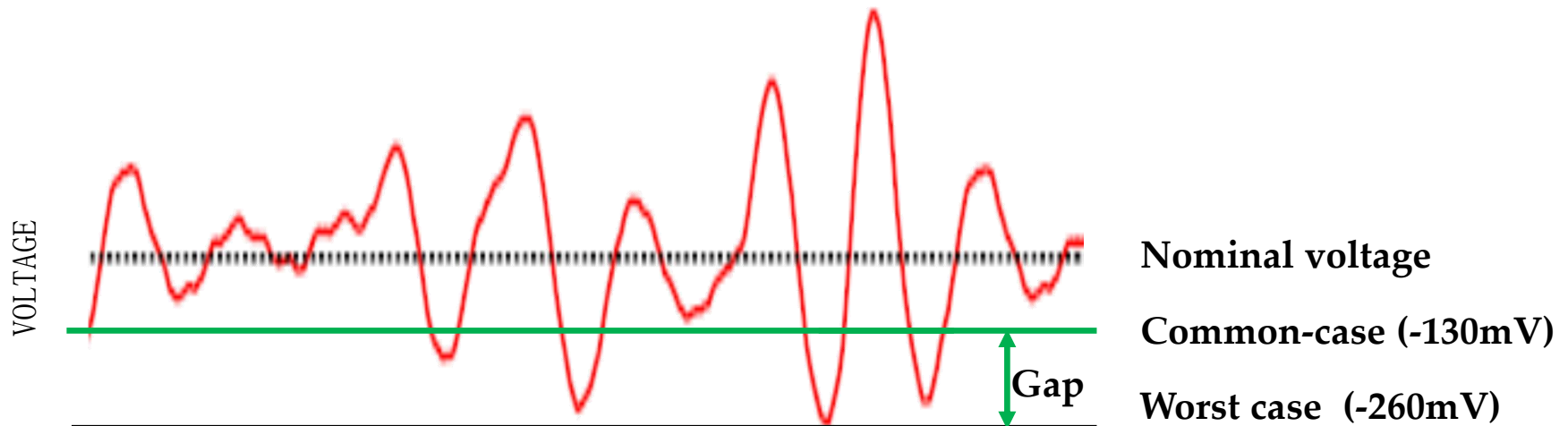
[2]. Zuowei Li, et. al., “Thermal-aware Power Network Design for IR Drop Reduction in 3D ICs,” ASPDAC 2012

3D power delivery network



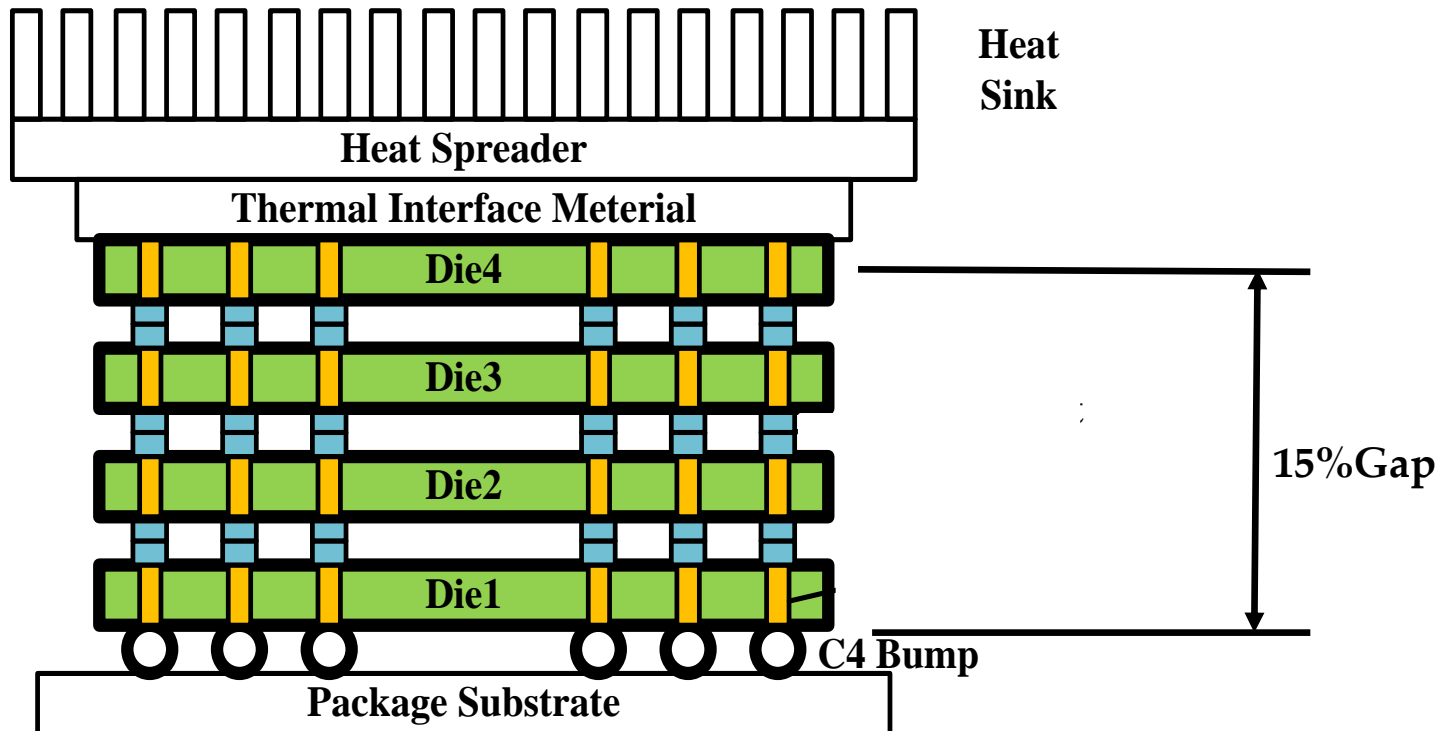
Three key observations of V droop

- 1. Temporal variation
 - Worst-case >> Common-case



Three key observations of V droop

- 2. Spatial variation
 - Top layer > bottom layer



Three key observations of V drop

- 3. Application variation
 - Voltage-violent > Voltage-mild

Voltage-violent



Voltage-mild



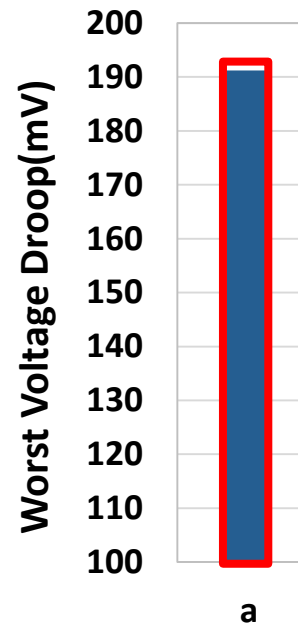
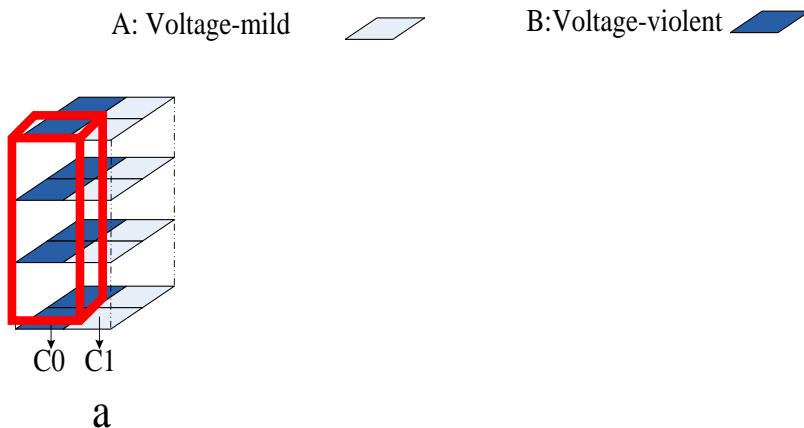
Three key observations of V drop

■ 3. Application variation

- Voltage-violent > Voltage-mild



- Vertical resonance

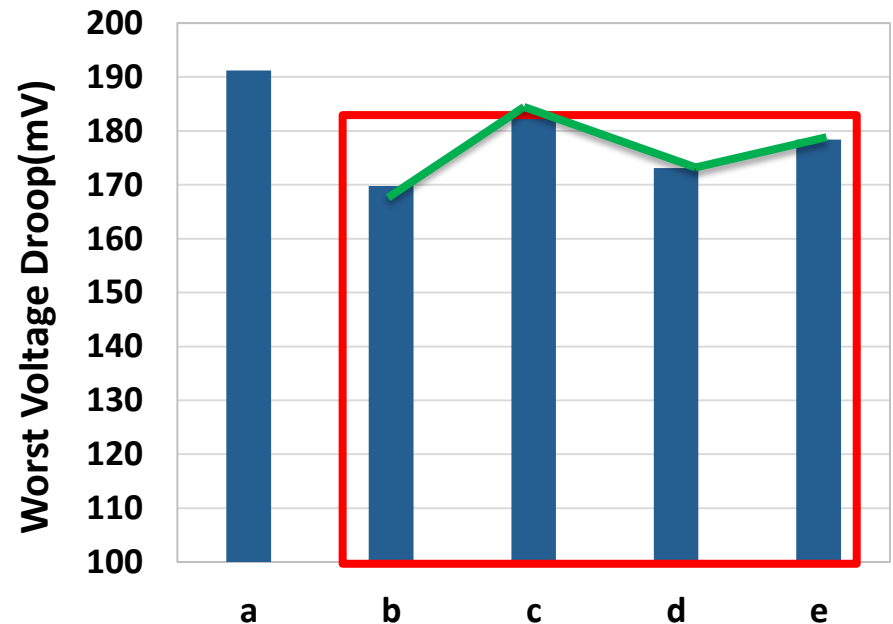
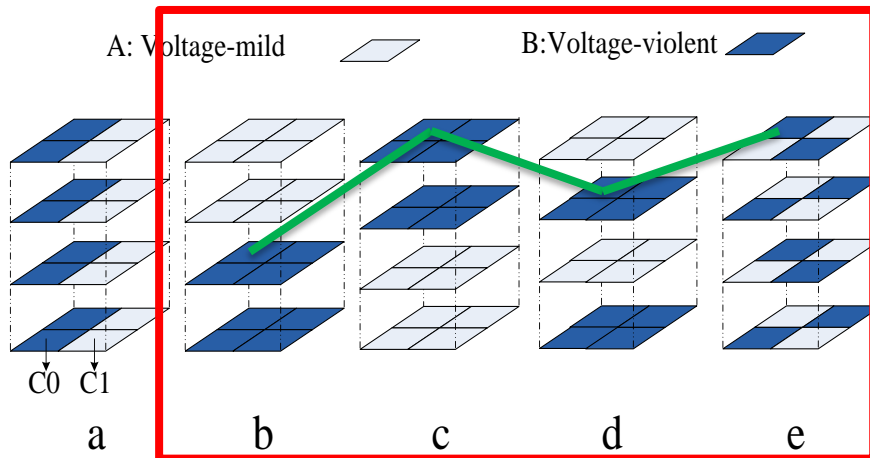


Three key observations of V drop

- 3. Application variation
 - Voltage-violent > Voltage-mild



- Vertical resonance

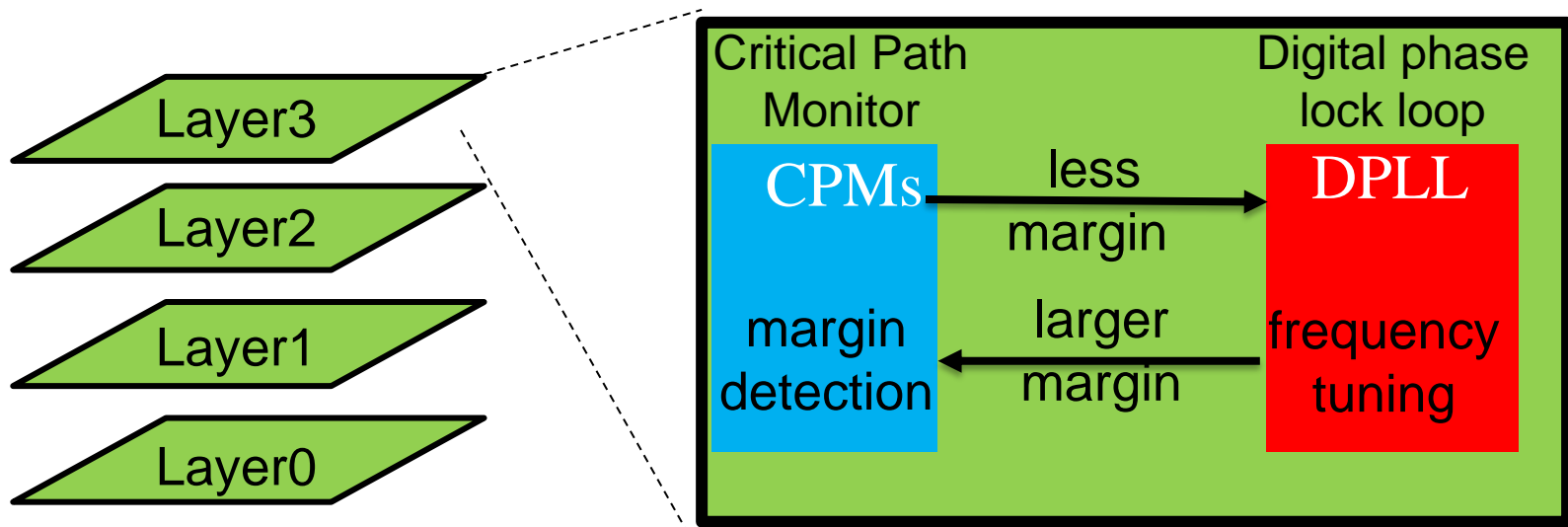


Swimming Lane – Overview

- Temporal variation
 - Common margin (C_Margin) vs. Worst-case margin (W_Margin)
 - Reduce supply voltage
- Spatial variation
 - Layer-independent failsafe design
 - Constrain the voltage droop effect within the layer
- Application variation
 - Voltage droop mitigation based on thread scheduling
 - Reduce intra-layer gap of voltage droop
 - Reduce worst voltage droop of the whole chip

Swimming Lane – Hardware Design

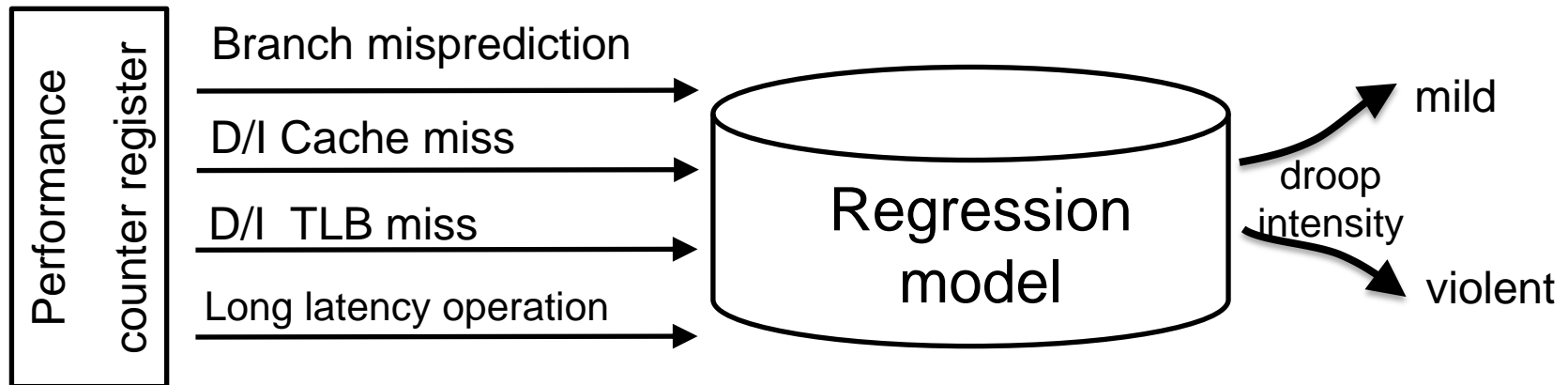
- Failsafe design
 - Rapidly tune the frequency in case of large voltage droop.
- Layer-independent failsafe design
 - Constrain the voltage droop effect within the layer



Swimming Lane – Software Design

Violent-first thread scheduling

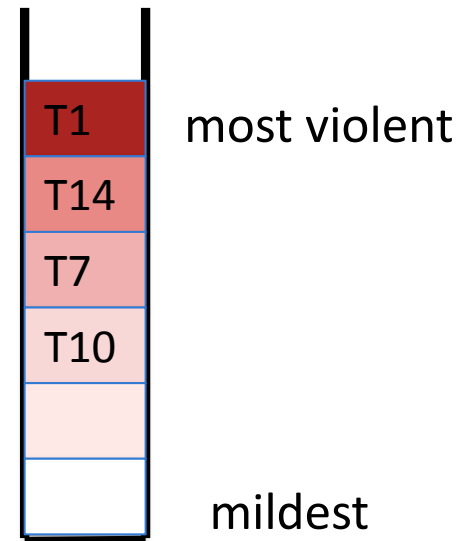
- Determine the priority for thread scheduling
 - Thread emergency Level prediction
 - Use program activities as input
 - Branch mis-prediction
 - Cache miss
 - TLB miss
 - Long latency operation



Swimming Lane Software Design

Violent-first thread scheduling

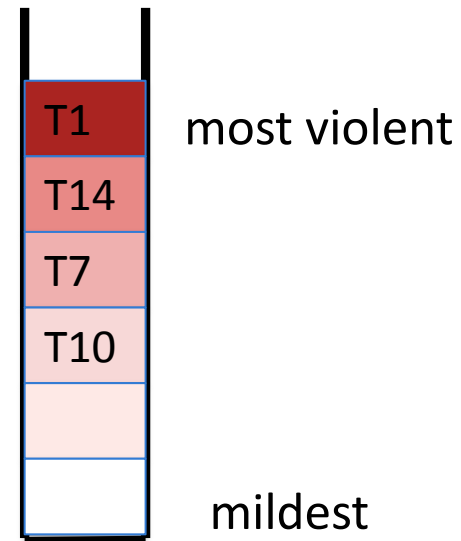
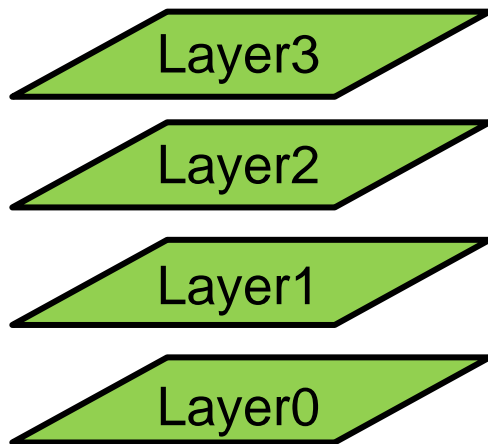
- Determine the priority for thread scheduling
 - Thread emergency Level prediction
 - Sort thread according to their droop intensity
 - If multiple threads have the same IDI, a round-robin algorithm is employed to choose threads from different applications.



SwimmingLane Software Design

Violent-first thread scheduling

- Determine the priority for thread scheduling
 - Thread emergency Level prediction
 - Sort thread according to their droop intensity
 - If multiple threads have the same IDI, a round-robin algorithm is employed to choose threads from different applications.
- Violent-first scheduling



Experimental setup

■ Simulated Layer Configuration

■ GEMS

Parameters	Configuration
Number of Cores	4
Clock Frequency	2.0 GHz
Fetch/Decode Width	4 instructions/cycle
Branch-Predictor Type	64 KB bimodal gshare/chooser, 1K entries
Reorder Buffer Size	128
Unified Load/Store Queue Size	64
Physical Register File	32-entry INT, 32-entry FP
INT ALU, INT Mul/Div, FP ALU, FP Mul/Div	4/2/4/2
L1 Data Cache	16KB, 2-way, 32B line-size, 1-cycle latency
L1 Instruction Cache	16KB, 2-way, 32B line-size, 1-cycle latency
L2 Unified Cache	1MB, 4-way, 64B line-size, 16-cycle latency
I-TLB/D-TLB	64-entry, fully-associative

■ Workloads

■ SPLASH2

■ Power Delivery Network

■ $V_{\text{nominal}} = 1.4 \text{ V}$

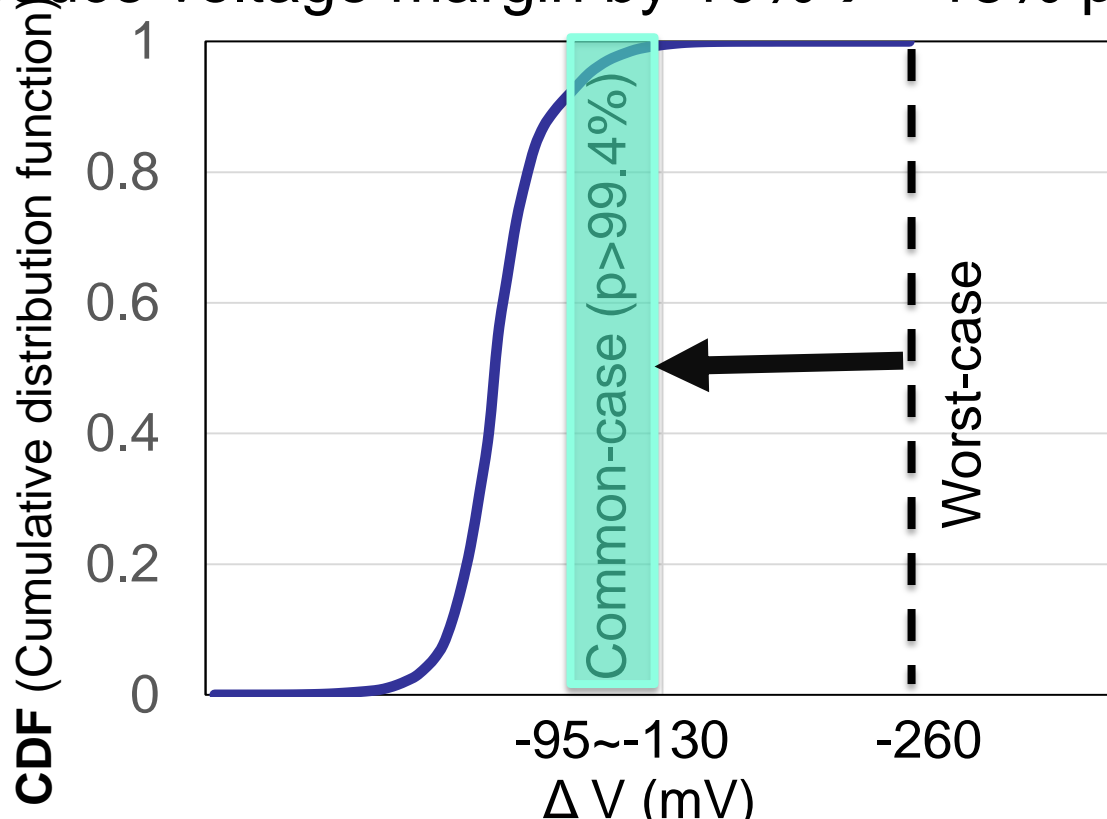
[5] Zheng Xu, et al., “Decoupling Capacitor Modeling and Characterization for Power Supply Noise in 3D Systems,” ASMC 2012.

Experimental Results

- Voltage droop reduction
 - Reduce the worst voltage droop by 26 mV.
 - Reduce the intra-layer voltage gap by 14 mV on average.

Experimental Results

- Voltage droop reduction
 - Reduce the worst voltage droop by 26 mV.
 - Reduce the intra-layer voltage gap by 14 mV on average.
- Voltage margin reduction
 - Reduce voltage margin by 10% → ~18% power saving



Conclusions

- **We observe non-evenly voltage droop distributed across 3D-stacked chips.**
- **Propose a hardware infrastructure and violent-first thread scheduling policy.**
 - Isolate timing error effect within single layer.
 - Characterize thread voltage feature and conduct optimal scheduling.
- **Reduce voltage droop effect**
 - Mitigate 40% of voltage violations
 - Reduce the voltage margin by 10%
 - Save power by 18%

Thank You for Your Attention

Swimming Lane: a Composite Design to Mitigate Voltage Droop Effects in 3D Chips

Xing Hu^{1,2,3}, Yi Xu^{3,4}, Yu Hu¹ and Yuan Xie^{3,5}

¹Institute of Computing Technology, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

³AMD Research China Lab

⁴Macau University of Science and Technology

⁵Pennsylvania State University