SAVS: A Self-Adaptive Variable Supply-Voltage Technique for Process-Tolerant and Power-Efficient Multi-issue Superscalar Processor Design

Hai Li, Intel Co.
Yiran Chen, Synopsys Inc.
Kaushik Roy, Purdue University
Cheng-Kok Koh, Purdue University

Presenter: Kunhyuk Kang

Outline

- Introduction
- Mechanism of SAVS
- Implementation of SAVS
- Experimental Results
- Conclusion

Outline

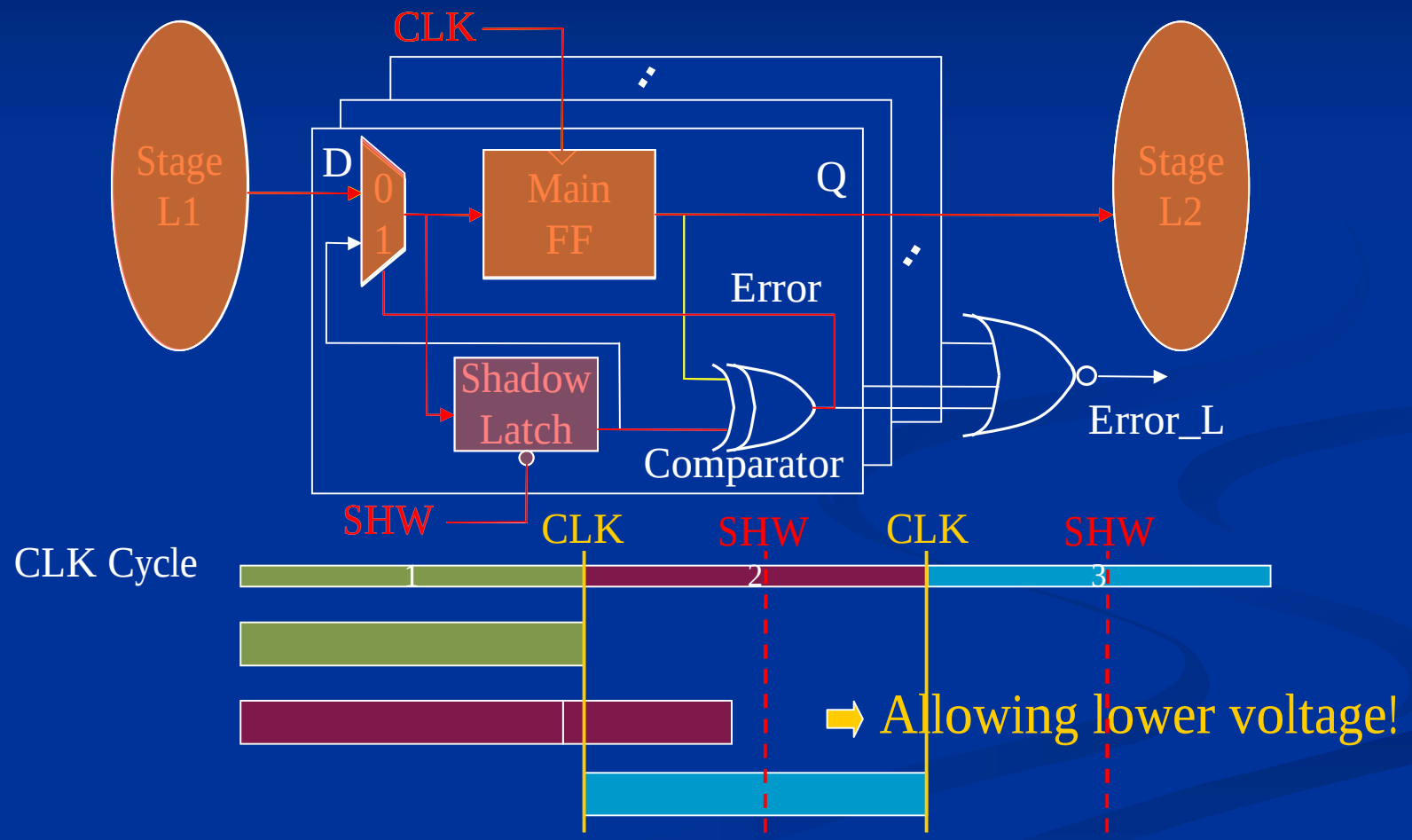
- Introduction
- Mechanism of SAVS
- Implementation of SAVS
- Experimental Results
- Conclusion

"Peak Power" and "Average Power" of Logic VLSI Circuit

Delay distribution of 32-bit CLA (Carry Look-Ahead adder) shows:

Very few operations really go through the longest data path of CLA

Shadow flip-flop mechanism for timing failure correction [1]



[1] D. Ernst, et. al., Razor: A Low-Power Pipeline Based on Circuit-Level Timing Speculation, 36th MICRO., pp. 7-18, Dec. 2003.

Motivation

Timing-failure correction (TFC) has been successfully applied to the logic circuit (i.e., ALU) of embedded system for low power but:

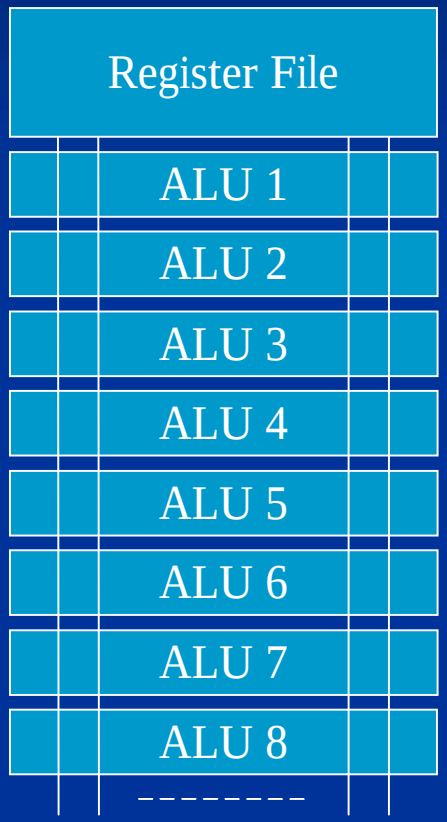
- Can TFC be used for interconnect dominated circuit, i.e., bus? (Wasted power of interconnect?)
- Can TFC be applied to high-performance system, not only for power efficiency, but also for process tolerance (chip yield enhancement)?

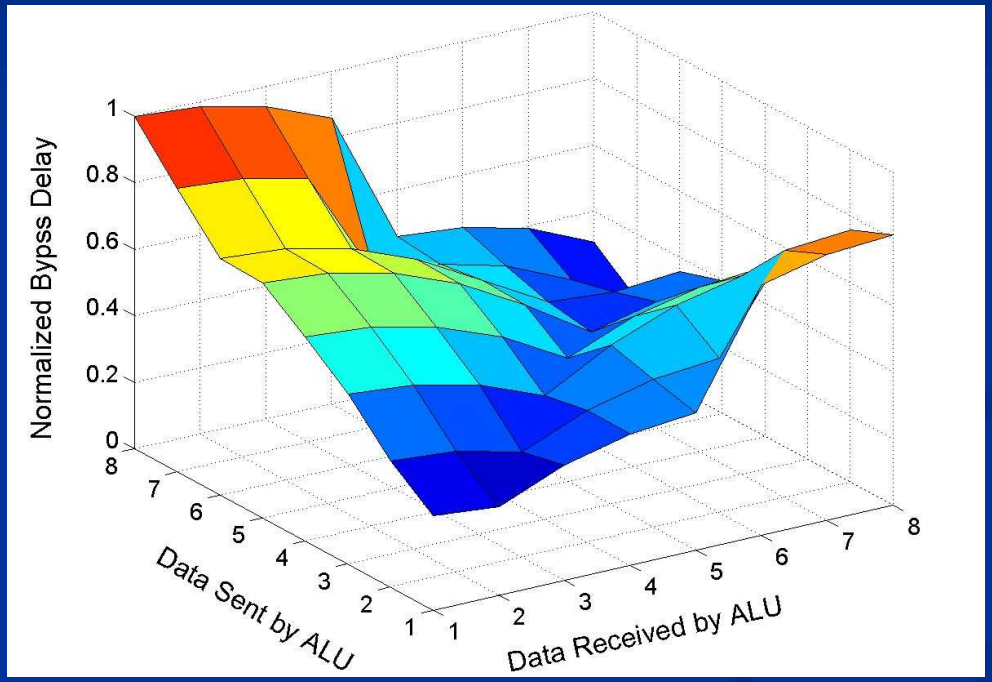
SAVS: Self-Adaptive Variable Supply-Voltage Technique

Outline

- Introduction
- Mechanism of SAVS
- Implementation of SAVS
- Experimental Results
- Conclusion

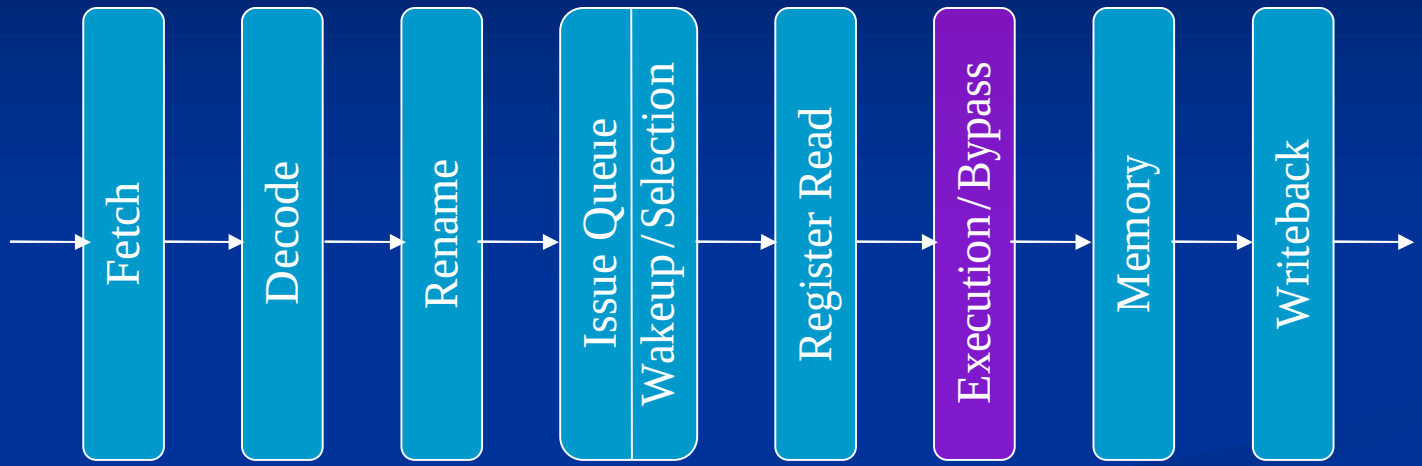
Delay Distribution of Bypass Logic





Most of the time, bypassing is constrained among first several ALU's that are in physical proximity (Instruction level parallelism limitation)

SAVS in High Performance Microprocessor



Stage Delays of 8-issue Superscalar Pipeline

	Delay (FO4)		Delay (FO4)
Fetch	30	Decode	18
Rename	18	Register Read	24
Issue Queue	12	Execution	18
Wakeup / Selection	12	Bypass	18
Load / Store Queue	12	Writeback	24
I-Cache	30	D-Cache	30

Applying SVAS to Pipeline Stages

- The delays of Fetch, Register Read, D-cache and Writeback stages are insensitive to the inputs (data address). – non-SAVS
- Rename logic is a RAM structure but has very short delay. – SAVS w/o TFC
- The delay of Decode greatly relies on ISA (Instruction set architecture) – non-SAVS for conservation
- Recovery from the timing error at Issue Queue and Wakeup/Selection is difficult and expensive– non SAVS
- The load/store queue in memory stage is a CAM structure with short delay – SAVS w/o TFC
- Applying SAVS to the critical stage – Execution/Bypass Stage – helps chip yield enhancement.

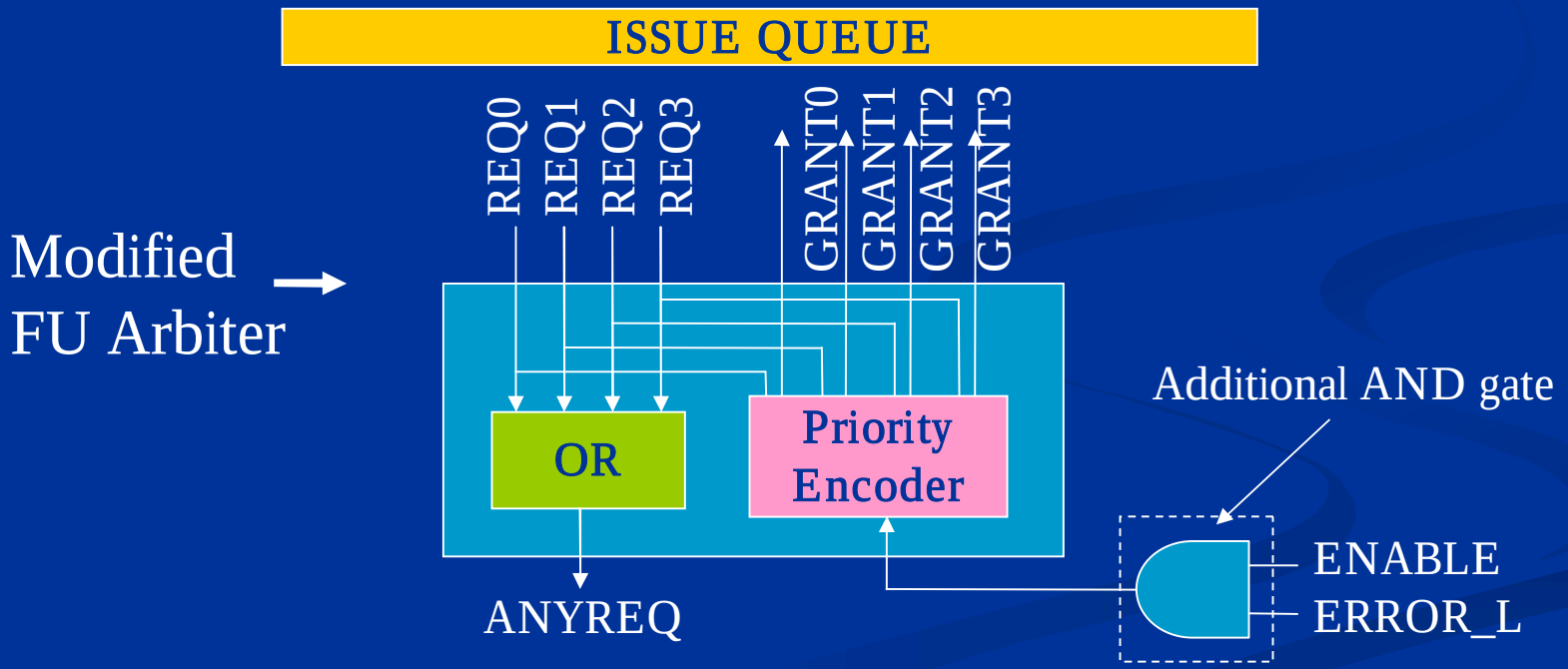
Outline

- Introduction
- Mechanism of SAVS
- Implementation of SAVS
- Experimental Results
- Conclusion

Pipeline Recovery Mechanism (1)

Pipeline recovery mechanism ensures:

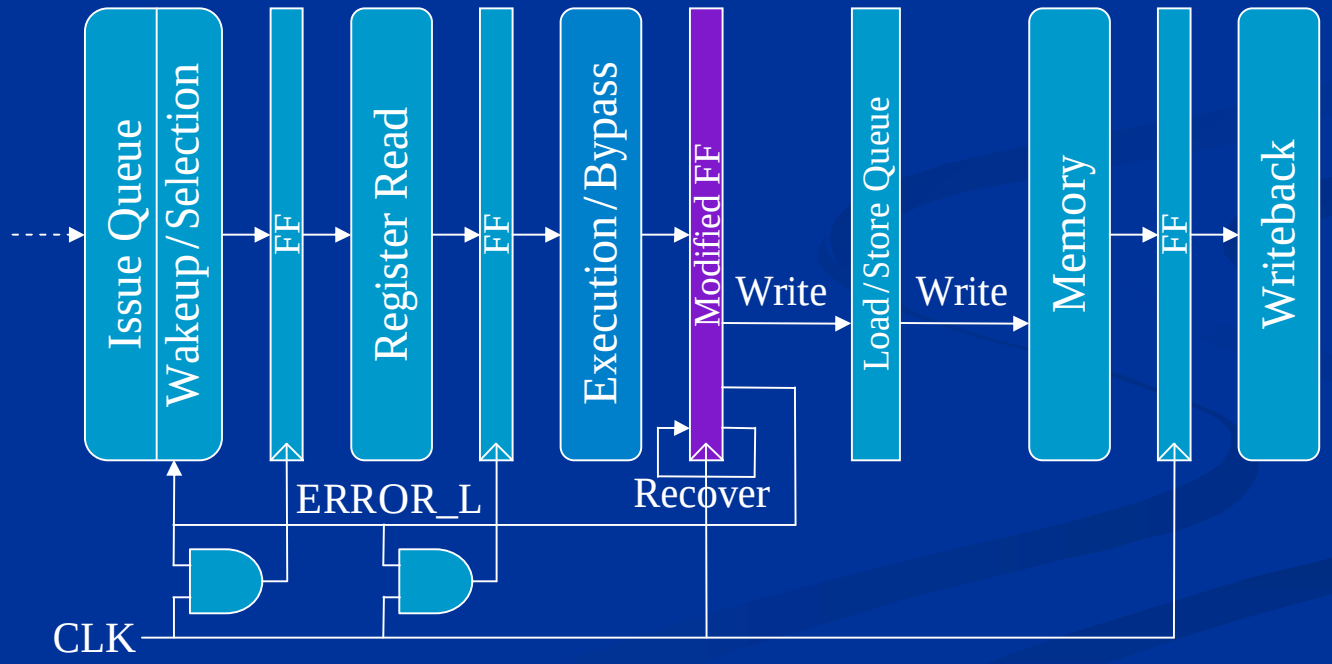
- No new instructions are issued out from the issue queue until the re-executions complete;



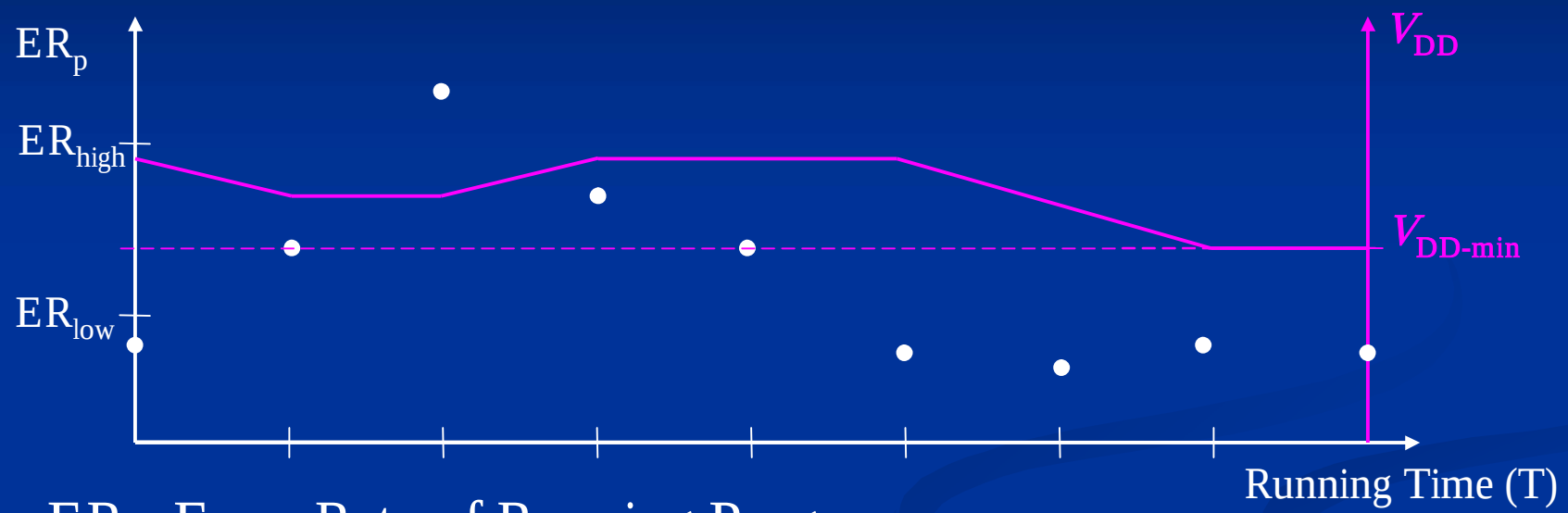
Pipeline Recovery Mechanism (2)

Pipeline recovery mechanism also ensures:

- The incorrect execution results in the previous cycle should be flushed out from the pipeline
- No errant register or cache writing is committed.



Supply Voltage Scaling Control



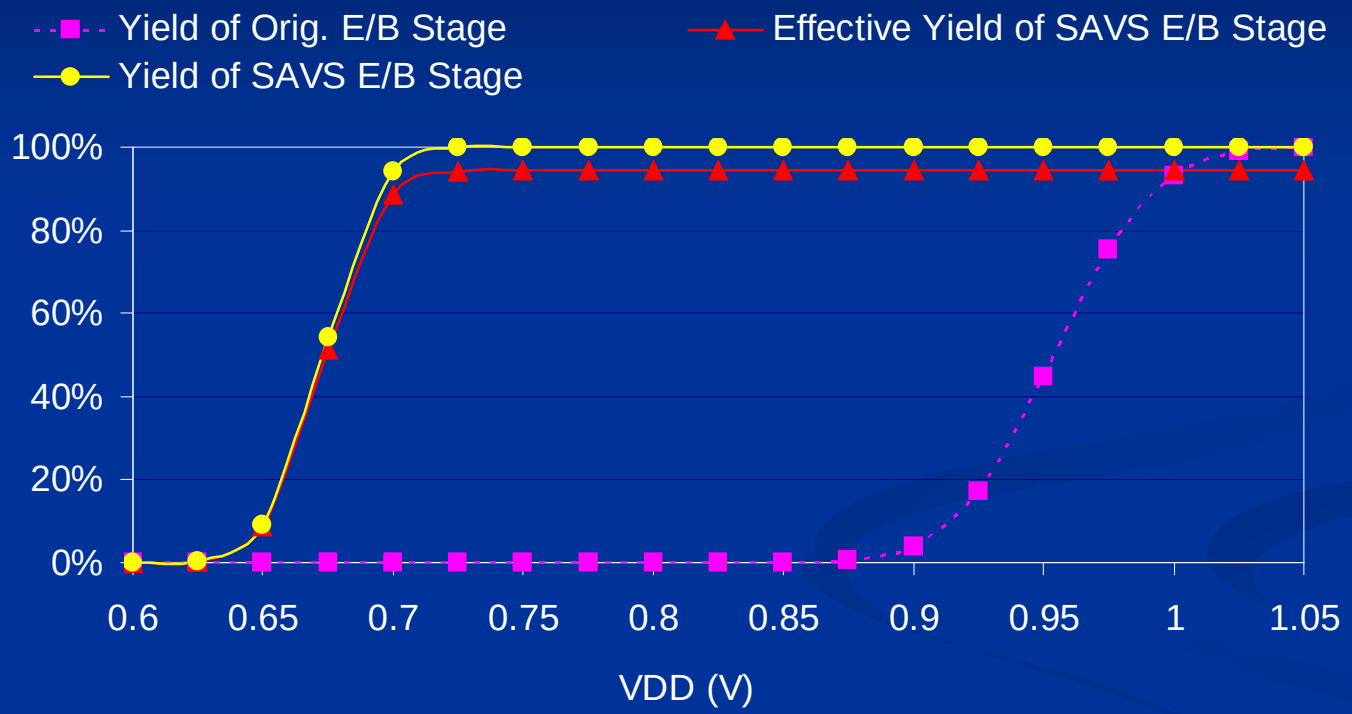
ER_p: Error Rate of Running Program

- Two error rate thresholds are used to indicate the direction of V_{DD} adjustment.
- V_{DD} can never be scaled under $V_{DD\text{-}min}$ for a desired chip yield.
- V_{DD} ramping is limited by the response time of voltage regulator and power supply noise tolerance.

Outline

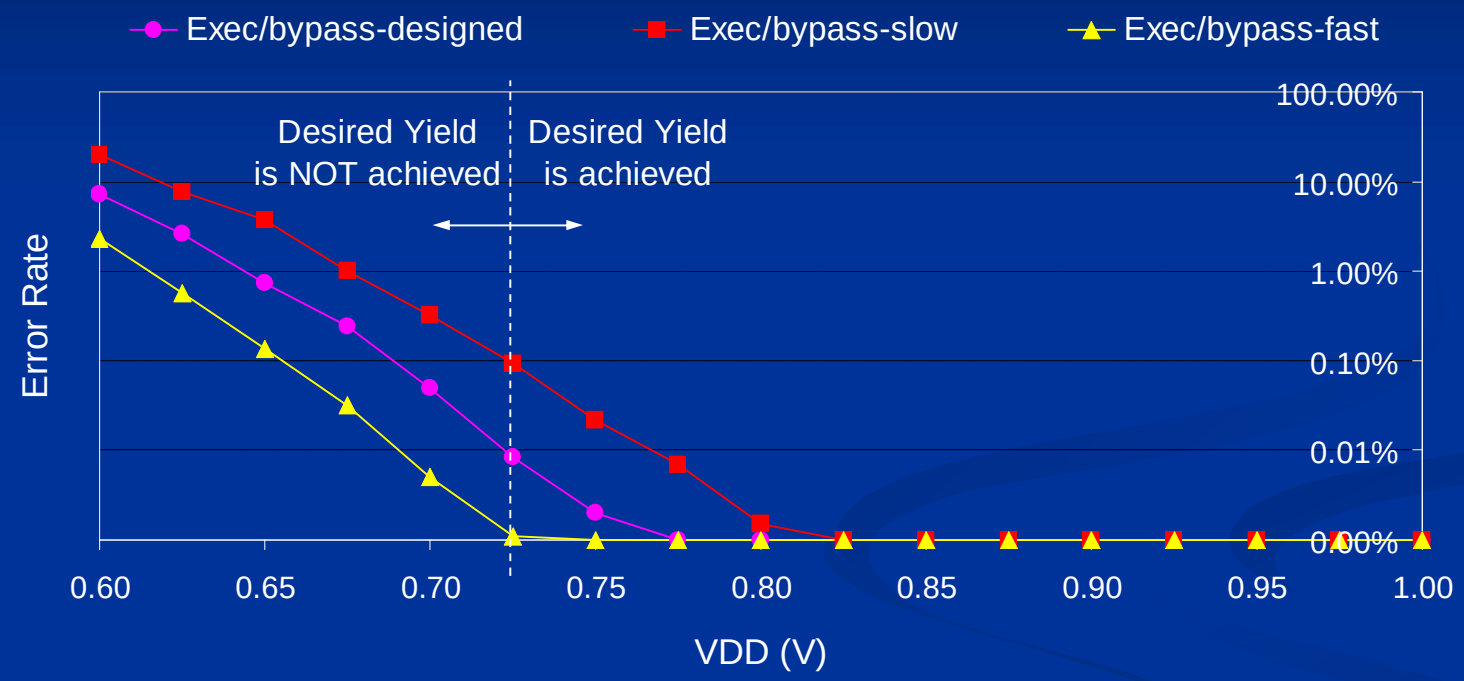
- Introduction
- Mechanism of SAVS
- Implementation of SAVS
- Experimental Results
- Conclusion

Yield of Execution/Bypass Stage under Different Supply Voltages



- 8-issue execution/bypass (E/B) stage with 32-bit CLA, at BPTM 70nm Tech..
- STDs of both inter-die and intra-die V_T are 30mV.
- STD of interconnect width is 10% of nominal value.
- Spatial correlation coefficient is set to 0.4.
- Clock period ensures 93.3% chip yield at the normal VDD (1.0V).

Error Rate of ALU and Execution/Bypass Stage



- $V_{DD\text{-}min}$ is 0.725V.
- "-slow" ("-fast"):
  - V_T has +42.4mV (-42.4mV) deviation from the designed value toward 0V
  - Interconnect width of bypass logic has –10% (+10%) deviation from the designed value toward 0.

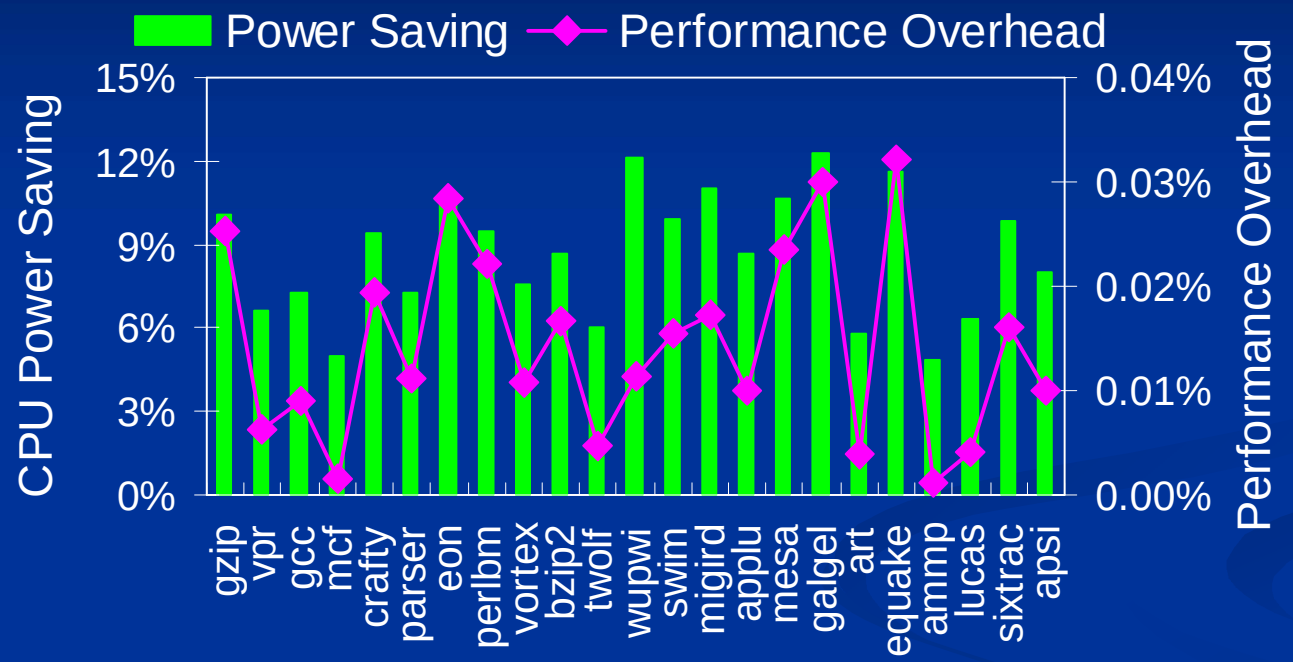
Architectural Level Simulation Setup

Baseline processor configuration

Processor	8-way issue, 128 RUU, 64 LSQ, 8 integer ALUs, 2 integer mul/div units, 4 FP ALUs, 4 FP mul/div units, uses clock gating (DCG) and s/w prefetching
Brach Prediction	8K/8K/8K hybrid predictor; 32-entry RAS, 8192-entry 4-way BTB, 8 cycle misprediction penalty
Caches	64KB 2-way 2-cycle I/D L1, 2MB 8-way 12-cycle L2, both LRU
MSHR	IL1 - 32, DL1 – 32, L2 – 64
Memory	Infinite capacity, 400 cycle latency
Memory Bus	32-byte wide, pipelined, split transaction, 4-cycle occupancy

- Deterministic Clock Gating (DCG) is applied.
- ER_{high} = 0.1%, ER_{low} = 0.05%.
- V_{DD} ramping rate = 5mV/μs.
- Clock frequency = 2GHz.
- Original V_{DD} = 1.0V (baseline), $V_{DD\text{-}min}$ = 0.725V.

Power Saving and Performance Overhead of SAVS



- On average, 8.66% power reduction with 0.014% IPC- (Instruction Per Cycle) -based performance penalty.
- The more IPC is, the more power can be saved.
- V_{DD} applied to other non-SAVS stages can be further reduced to 0.85V w/o introducing performance, power and yield penalties.

Outline

- Introduction
- Mechanism of SAVS
- Implementation of SAVS
- Experimental Results
- Conclusion

Conclusion

- Timing failure correction (TFC) can be applied to interconnect-dominated circuit, when delay varies from case to case, i.e., bypassing bus.
- Our TFC technique, SAVS (Self-Adaptive Variable Supply-Voltage Technique), can be successfully applied to high-performance multi-issue system for power efficiency and process-tolerance.
- SAVS is good for scaled technology with significant process parameter fluctuations.
- SAVS achieves high throughput and low power in pipelined systems.

Thank you!
Q & A

Contact author: Yiran Chen (Yiran.Chen@synopsys.com)