A Programmable Resistive Power Grid for Post-Fabrication Flexibility and Energy Tradeoffs

^{1 2}Kyle Craig, ¹Yousef Shakhsheer, ¹Sudhanshu Khanna, ¹Saad Arrabi, ¹John Lach, ¹Benton H. Calhoun, and ²Stephen Kosonocky

¹Dept. Of Electrical and Computer Engineering, University of Virginia, Charlottesville VA ²Advanced Micro Devices, Fort Collins, CO

International Symposium on Low Power Electronics and Design

Motivation

- Many applications impose energy consumption constraints:
 - − High End → Thermal constraints
 - Low End → Battery lifetime constraints



Performance

• Still demand high bursts of performance

<u>Technology scaling alone is not enough</u>. <u>Lowering energy is always ongoing</u>

Background

• Traditional Approaches:

Power gating/Dynamic Voltage & Frequency Scaling (DVFS)

- Power gating:
 - Leakage reduction
 - Register/Memory data is lost
- DVFS:
 - Dynamic energy reduction
 - Multiple *large/slow* DC-DC converters
 - Voltage scaling limited by fastest block

Background

• Traditional Approaches:

Power gating/Dynamic Voltage & Frequency Scaling (DVFS)

- Power gating:
 - Leakage reduction
 - Register/Memory data is lost
- DVFS:
 - Dynamic energy reduction
 - Multiple *large/slow* DC-DC converters
 - Voltage scaling limited by fastest block

<u>Propose: Programmable Resistive Power Grid for state</u> <u>retention leakage reduction & local voltage scaling</u> <u>without requiring extra DC-DC converters</u>.

Outline

- I. Motivation
- II. Background
- III. Programmable Resistive Power Grid
 - I. Implementation
 - II. Local Voltage Scaling
- **IV. Energy Savings**
- V. Practicality in a Commercial Processor
- VI. Large System Modeling
- **VII.** Conclusion

Reconfigurable Implementation

- Monolithic header broken into partitions (w^k_n)
 - Independent gate control
 - Non-uniform sizing
 - Sized for application requirements
 - W total width



Local Dynamic Voltage Scaling

- Allow header resistance (R_{Header}) to increase
 - Number paritions enabled
- V_{rail} droop



<u>No DC-DC regulation required.</u> <u>No extra DC-DC converters needed</u>.

Outline

- I. Motivation
- II. Background
- III. Programmable Resistive Power Grid
- IV. Energy Savings
 - I. Theoretical
 - II. Simulated
 - III. Measured
 - IV. Leakage Reduction
 - V. Activity Factor
- V. Practicality in a Commercial Processor
- **VI. Large System Modeling**
- VII. Conclusion

Theoretical Energy Savings

• Traditional Energy Equation:

$$E_{op}(V_{DD}) = C_{eff}(V_{DD}) * V_{DD}^{2} + V_{DD} * I_{L} (V_{DD}) * t_{op}$$

Dynamic Leakage

• Our Energy Equation:

$$\begin{split} E_{op}(V_{DD},V_{rail}) = \\ C_{eff}(V_{rail}) * V_{DD} * V_{rail} + V_{DD} * I_L \ (V_{rail}) * t_{op} \\ \\ \end{split} \\ \end{split} \\ \end{split} \\ \end{split} \\ \end{split} \\ \vspace{-2mm} \\ \vspace{-2m$$

Greater than linear energy savings expected

Simulated Energy Savings

- 32b Kogge Stone Adder
 - Successive Operations
- 90nm bulk CMOS



 V_{rail} settle \rightarrow Up to 37% energy savings, 2.8X slow down

Measured Energy Savings

- Four 97-Stage Ring Oscillators (RO)
- 90nm bulk CMOS



Measured up to 30% energy savings, 2X slow down

Leakage Reduction

- Reduce leakage while maintaining state
- 32nm SOI four-core x86 SOC
- Enable smallest partition allowing retention



Activity Factor

Simulation of 64 parallel ROs - Activty \rightarrow number enabled **Important for sizing** Informs header partitions 0.02 0.03 0.06 0.13 0.25 0.50 1.00 ■ 0.02 ■ 0.03 ■ 0.06 ■ 0.13 ■ 0.25 ■ 0.50 ■ 1.00 1.20 1.00 0.90 1.00 0.80 ധ 0.70 0.80 0.60 0.50 0.40 0.30 0.20 0.20 0.10 0.00 0.00 16 32 8 4 2 16 2 64 32 8 4 1 64 **Enabled parallel ROs Enabled parallel ROs**

1

Activity Factor



Activity Factor



Outline

- I. Motivation
- II. Background
- III. Programmable Resistive Power Grid
- **IV. Energy Savings**
- V. Opportunity in a Commercial Processor
- **VI. Large System Modeling**
- VII. Conclusion

Opportunity for Regulation

- Commercial four-core x86 SOC
 - Typical p-state occupancy
- Estimate power w/ Programmable Resistive Power Grid



Opportunity for Regulation

- w/o Programmable grid
 - Low P-state cores limited to frequency scaling
- w/ Programmable grid
 - Low P-state cores can reduce voltage and frequency



Outline

- I. Motivation
- II. Background
- **III. Programmable Resistive Power Grid**
- **IV. Energy Savings**
- V. Practicality in a Commercial Processor
- VI. Large System Modeling
 - I. Motivation
 - II. Setup
 - III. Results
- VII. Conclusion

Modeling Programmable Resistive Power Grid

- Modeling a full core?
 - Spice Simulation prohibitively long
 - Not practical
- Commercial power integrity tools exits
 - Apache Redhawk
 - Cadence tool suite
 - etc.

<u>Can we use these commercial tools during design to</u> <u>model our Programmable Resistive Power Grid?</u>

Small Scale Test

- Use Apache Redhawk
- Model Route Level Macro (RLM)
 - Power gated, simulated header partitions
 - − V_{DD} modeled in M11 M9, V_{rail} → M8 M2, V_{SS} → M11 M1
 - 32nm commercial processor



Small Scale Test

- Use Apache Redhawk
- Model Route Level Macro (RLM)
 - Power gated
 - − V_{DD} M11 to M9, V_{rail} → M8 to M2, V_{SS} → M11 to M1
 - 32nm commercial processor



Model Setup

- AMD Bulldozer core
- RLM's modeled as a time dependent current source and capacitance
- Caches excluded
- $V_{DD}/V_{SS} \rightarrow C4$ to M10, Virtual- $V_{SS} \rightarrow M11$ to M10



 Double-precision General Matric Multiply (DGEMM) benchmark



V_{DD} response shows our model is properly working

- Double-precision General Matric Multiply (DGEMM) benchmark
 - All RLMs superimposed



 Double-precision General Matric Multiply (DGEMM) benchmark



- Double-precision General Matric Multiply (DGEMM)
 benchmark
 Small change in footer partition.
 - All RLMs superimposed

Small change in footer partition. Significant change in ΔV



- **Double-precision General Matric Multiply (DGEMM)** benchmark ~50 RLMS: little variance across core.
 - All RLMs superimposed



- Double-precision General Matric Multiply (DGEMM) benchmark
 - All RLMs superimposed



Outline

- I. Motivation
- II. Background
- **III. Programmable Resistive Power Grid**
- **IV. Energy Savings**
- V. Practicality in a commercial processor
- **VI. Large System Modeling**
- VII. Conclusion

Conclusions

- Programmable Resistive Power Grid
 - Measured 30% energy reduction w/ 2X slowdown
 - Measured 90% leakage reduction w/ data retention
- Estimated 15% savings available in commercial fourcore x86 processor
- Demonstrated how to model commercial processor

Thank you

Questions?