



Ming Hsieh Department of Electrical Engineering

## Warped Gating: Gating Aware Scheduling and **Power Gating for GPGPUs**

Mohammad Abdel-Majeed, Daniel Wong and Murali Annavaram

GPGPU Execution units

Introduction

Motivation

S

GATE

Blackout

Results

- · GPU targets application with thousands of threads.
- Large number of execution units in the GPGPU.
- Each unit has an INT and FP pipelines.
- 32/SM in Fermi and 192/SM in Kepler.



Execution units burn massive leakage and dynamic power Why not power gate the execution units?

 Scheduler greedily issues ready instructions (without considering instruction type)



• On average 16 warps are ready to execute any cycle

- Good mix of INT and FP instructions are available each cycle
- INT/FP units turn ON/OFF rather rapidly due to greedy scheduling
  - Power gating needs many consecutive cycles of idleness
  - So no opportunity to power gate
- Power Gating regions
  - A: Detect Idle periods(no Gating)
  - **B**:Gating overhead is higher than saving(Power gated)
  - **C**:Cycles spent in this region will Translate into savings



 Give priority to same instruction type during scheduling

Change the scheduling order based on the instruction mix of the benchmark.





• GATES is able to increase the length of idle period but still not long enough to take advantage



- Idle periods are unable to go past break even time
  - Force idleness until break-even period once a unit goes idle and even if an instruction needs that unit
- Performance Loss? Scheduler Supp • No because one can take advantage of other available Scheduler resources and instruction mix **Issue Logic Execution Units** \_Buffer Busy Issued Arbiter V Dec\_INST INT R Ready, Typ Cycles>idle\_detect PG\_Signals Instruction Ready, Typ INT\_RDY FP\_RDY Dec\_INST FP 30% Ready, Typ Idle\_detect Dec\_INST SFU R SFU\_RDY LDST\_RDY Architectura 54.3% 0.0% 45.7% PG Logic Ready, Typ V Dec\_INST LD R Cycles>wakeup\_delay Frequency 10% BET\_counter Uncompensated Highest priority PG\_status Critical wakeup Priority Logic Cycle 1 Scoreboard Adaptive idle dle\_detect INT\_ACTV FP\_ACTV Count\_info Wakeup value Idle detect detect Critical Wakeup logic Counter 0% Cycle 1+BET 20 15 25 Ω Ready\_instruction\_scheduled Idle period length Compensated Cycles>BET time
- 0 Conv\_PG Gating Aware 📃 Coordinated Blackout 📃 Adaptive Idle\_detect

#### Simulation Setup

- GPGPU-sim cycle accurate simulator.
- Fermi architecture
- 14 cycles BET, 3 cycles wakeup latency, 5 cycles idle detect
- Benchmarks **GPGPU-sim** simulator

# **1.5x**

Leakage power Reduction

## ~0%

Area overhead

Performance overhead

1%





### Ming Hsieh Institute Ming Hsieh Department of Electrical Engineering