Energy Efficient Partitioning of Dynamic Reconfigurable MRAM-FPGAs

Ali Ahari, Mojtaba Ebrahimi, Mehdi B. Tahoori
Motivation

- High leakage power in SRAM-FPGAs
  - Non-volatile technologies for FPGA configuration bits
    - Flash-based FPGAs are already in mass production

- MRAM could be potentially replaced with SRAM in configuration bits
  - Fast read
  - Low leakage power
  - Smaller footprint
  - Non-volatile
  - Immunity to soft errors

- Dynamic partial reconfiguration (DPR) is a popular area reduction method
  - Impractical for Flash- and Phase Change Memory-based FPGAs → Low endurance
  - MRAM Technology → Unlimited endurance
Problem Statement

- Write operation in MRAM has high latency and energy
  - ~5X higher latency and ~7X higher energy compared to SRAM [Oboril15TCAD]
  - Higher reconfiguration energy in MRAM-based FPGAs

- Reconfiguration energy is a major challenge in DPR-based MRAM-FPGAs

Proposal: Partitioning technique for reconfigurable MRAM-FPGAs

- Takes into account reconfiguration frequency of tasks
- Exploits entire FPGA area for implementing DPR
- Minimizes reconfigured bits → reconfiguration energy
Outline

- MRAM-FPGA
- Proposed Partitioning Technique
- Experimental Results
- Conclusions
Data stored in Magnetic Tunnel Junction (MTJ)

- Composed of two independent ferromagnetic layers: free and reference layer
- Parallel and anti-parallel orientation of two layers
  - low and high resistance → represent either logic 1 or 0
- Write operation: a high current to change orientation
- Read operation: a low current to sense orientation
6-input MRAM-based Lookup Table [Zhao11MR]
Experimental setup
- TSMC 65 nm for SRAM-LUT
- STT-MRAM model [Bishnoi14ISQED]
- Without power gating

In SRAM-FPGA
- Static power of configuration bit is 38%
- No power-gating is applied
  - Due to volatile contents of SRAM
  - Reconfiguration required at wakeup time

MRAM-FPGA
- Significantly smaller leakage power
  - Only in CMOS-based parts
- Non-volatile → could be power gated
  - Reconfiguration is not required

Energy Efficient Partitioning of Dynamic Reconfigurable MRAM-FPGAs
Reconfiguration energy of MRAM-FPGA is 8X higher than SRAM-FPGA
- Higher write power
- Higher write latency

Similar results by [Zhao11GLSVLSI]

Reconfiguration energy is new concern in dynamic reconfigurable MRAM-FPGAs
- New partitioning technique needed

Energy Efficient Partitioning of Dynamic Reconfigurable MRAM-FPGAs
Outline

- MRAM-FPGA
- Proposed Partitioning Technique
- Experimental Results
- Conclusions
Main Idea

- Order of execution matters

- DPR with power-gating
  - SRAM-FPGA → Both 3 reconfiguration
  - MRAM-FPGA → a) 3  b) 2
    - Non-volatile

- Proposed partitioning technique
  - Input: FPGA size and hardware tasks
  - Objective: Minimize reconfiguration energy
    → Minimizing overall energy
Classification of Timing Conflicts

- Active-active conflict → cannot be mapped to the same partition

- Active-idle conflict → content update in each power-off cycle

- No conflict → only one reconfiguration
Energy Reduction Objective

- Energy of FPGA with power gating capability

\[ \sum_{\text{for each HWT}} \text{Active energy}_{HWTi} + \text{Power gating energy}_{HWTi} + \# \text{ of reconf.}_{HWTi} \times \text{reconfig. energy}_{HWTi} \]

- Active energy: dynamic and leakage energy
- Power gating energy: wakeup and power-off overheads
Proposed Partitioning Technique

- Graph-based representation adapted from [He12ICCAD]
- Constructing exclusion graph according to timing conflicts
  - Vertex → hardware task
  - Edge → active-active conflict between vertices
    - Cannot be mapped to the same partition
Proposed Partitioning Technique

Constructing intersection graph → independent set of exclusion graph

- Independent set: subset of vertices with no adjacent vertices
- Vertex → partitioning candidate
- Edge → two partitions have common tasks
  - Cannot be simultaneously included in final result

**Exclusion graph**

- **Vertex**
  - **Partitioning candidate**
- **Edges**
  - Two partitions have common tasks
  - **Cannot be simultaneously included in final result**

**Intersection graph**

- **Nodes**
  - **P1, P2, P3, P4, P5, P6, P7**

**Execution schedule**

- **Time**:
  - i
  - i+1
  - i+2
  - i+3
  - i+4

- **Tasks**
  - T1, T2, T3, T4, T5

**Energy Efficient Partitioning of Dynamic Reconfigurable MRAM-FPGAs**
Proposed Partitioning Technique

- Weighting intersection graph
  - Based on optimization objective → i.e. reconfiguration energy reduction
  - Weight of partition $P = \{T_n, T_{n+1}, ..., T_m\}$ computed as $W_P = R \times \max_{n \leq i \leq m} a_i$
    - $R$ is # of reconfigurations → dependant on execution schedule
    - $a_i$ is area of $T_i$ → dependant on partitioning strategy
Proposed Partitioning Technique

Obtaining final solution
- By min weight maximal Independent set
  - Over all possible solutions which fit inside FPGA

Execution schedule

Exclusion graph

Intersection graph

Energy Efficient Partitioning of Dynamic Reconfigurable MRAM-FPGAs
Outline

- MRAM-FPGA
- Proposed Partitioning Technique
- Experimental Results
- Conclusions
Experimental Setup

- Experimental setup for evaluating 32-bit configuration frame
  - TSMC 65 nm for SRAM-LUT
  - STT-MRAM model [Bishnoi14ISQED]

- Set of commonly used multimedia, cryptography, and scientific modules
  - adpcm, AES, blowfish, dfsin, dfvin, dfadd, dfmul, Mpeg4

- Test cases: 7 synthetic (TC1 … TC7) and one real world (Mpeg4)
Reconfiguration Energy Improvement

- For five points
  - Min: minimum area partitioning [He12ICCAD]
  - Max: fully static implementation
  - and three points in between
Opportunities for Power Gating

- Time
- Active
- Idle
- Reconfig.
- Active

- Without avoiding active-idle conflicts
- With avoiding active-idle conflicts

![Graph showing improvements in power gating time vs. average task frequency.](image)

- Average Task Frequency (#Execution per 10000 seconds)
- Improvements in Power Gating Time (%)
Outline

- MRAM-FPGA
- Proposed Partitioning Technique
- Experimental Results
- Conclusions
Conclusions

- Reconfiguration energy is a major challenge of DPR in MRAM-FPGA
  - Due to high write energy of MRAM

- An energy-efficient partitioning technique for MRAM-FPGA presented
  - Reducing number of reconfigured bits $\rightarrow$ less reconfiguration energy

- Experimental results
  - 25% additional area $\rightarrow$ 68.1% less reconfiguration energy
  - ~12% improvement in power gating time
Why new partitioning technique?

- Dynamic reconfigurable MRAM-FPGAs
  - Possibility of power gating without data loss
  - High reconfiguration energy
- Different concerns → New partitioning technique needed