

# Architectural Aspects in Design and Analysis of SOT-based Memories

Rajendra Bishnoi, Mojtaba Ebrahimi, Fabian Oboril & Mehdi Tahoori

INSTITUTE OF COMPUTER ENGINEERING (ITEC) – CHAIR FOR DEPENDABLE NANO COMPUTING (CDNC)



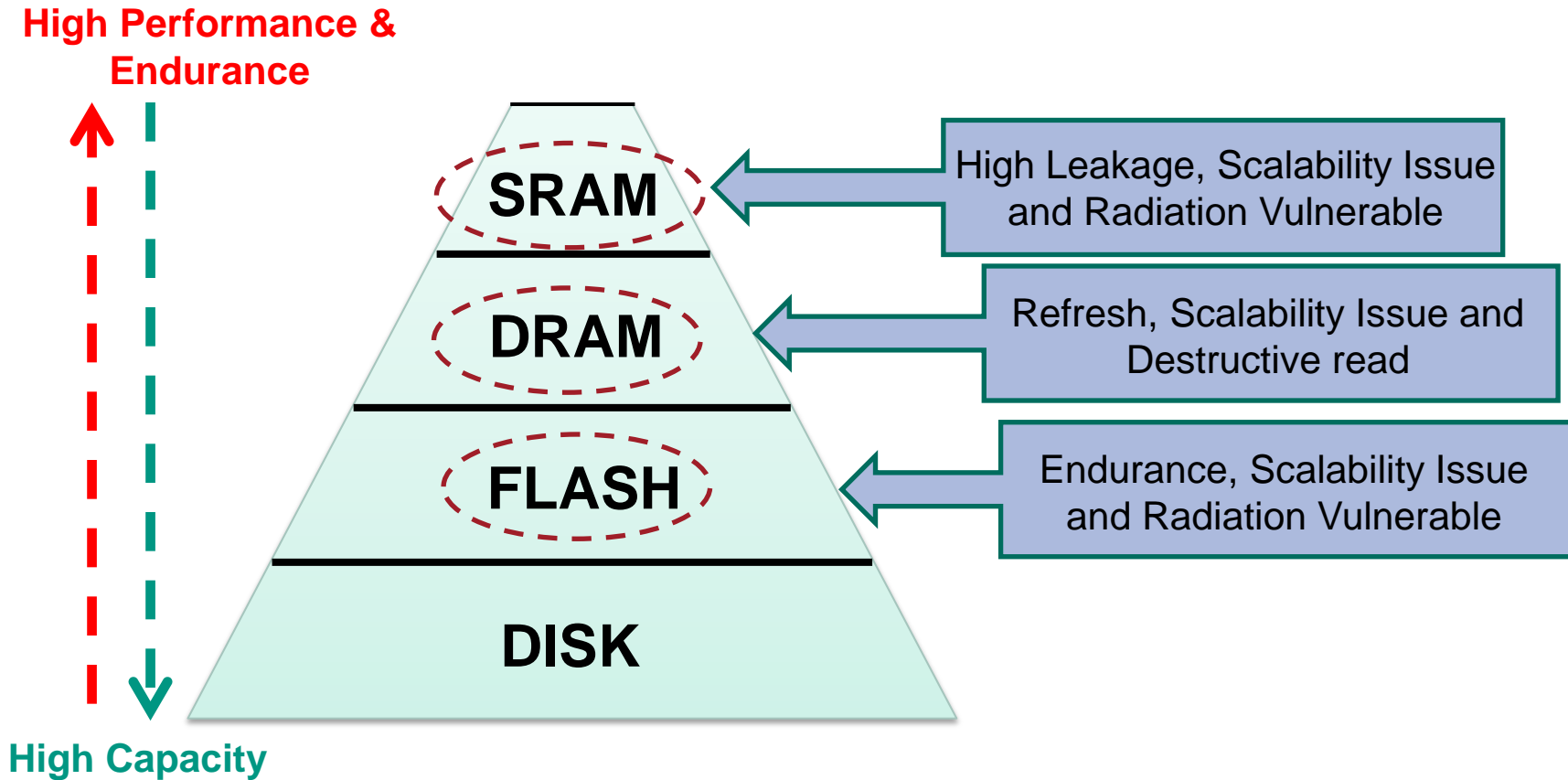
# Outline

- **Motivation**
  
- **SOT based MRAM**
  - STT-MRAM & Limitations
  - Basics of SOT-MRAM
  - Simulation tool flow
  
- **Results**
  - For various memory technologies
  - System-level
  
- **Summary & Conclusion**

# Outline

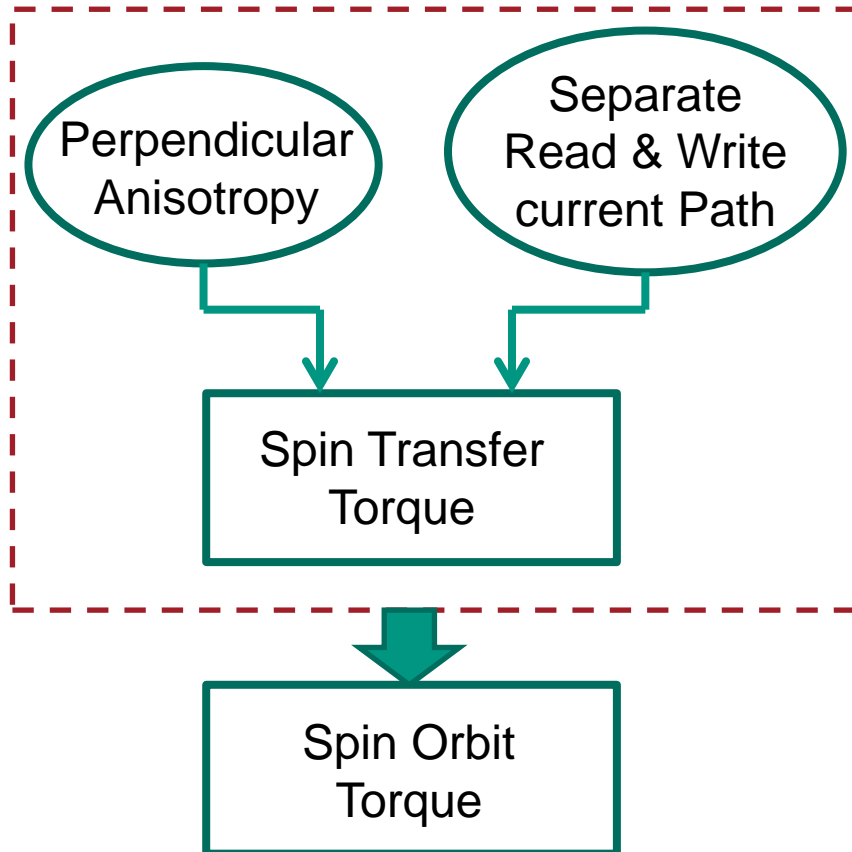
- **Motivation**
  
- **SOT based MRAM**
  - STT-MRAM & Limitations
  - Basics of SOT-MRAM
  - Simulation tool flow
  
- **Results**
  - For various memory technologies
  - System-level
  
- **Summary & Conclusion**

# Memory Hierarchy



- A Universal Memory Required to overcome these limitations
- Non Volatile Magnetic RAM is promising candidate

# Motivation



## Results in

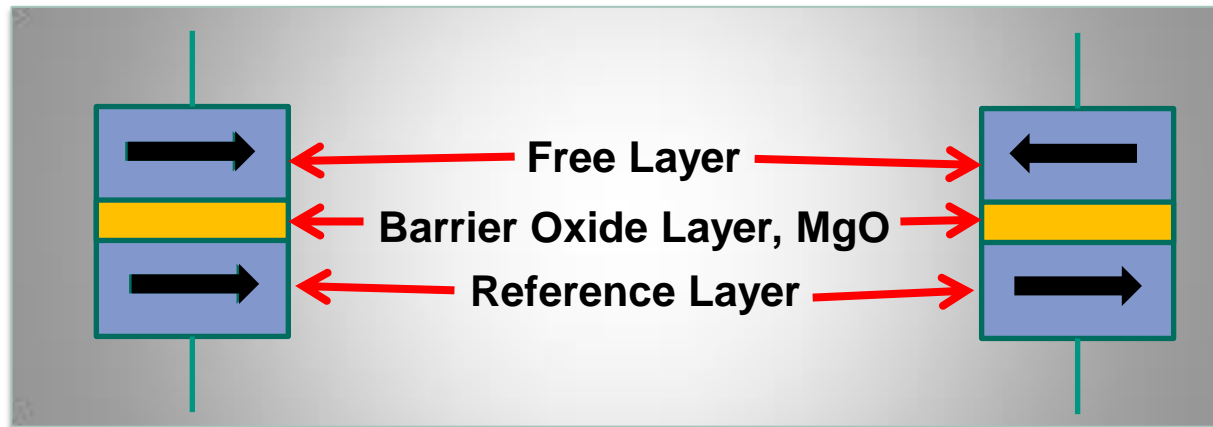
- **Low Write Current**
- **Low Switching Time**
- **Avoid Read Disturb**

- STT-MRAM has potential to become universal memory technology
- However, obstacles are
  - High write current & time
  - “Read Disturb”
- Addressed using Spin Orbit Torque (SOT)

# Outline

- Motivation
  
- **SOT based MRAM**
  - STT-MRAM & Limitations
  - Basics of SOT-MRAM
  - Simulation tool flow
  
- Results
  - For various memory technologies
  - System-level
  
- Summary & Conclusion

# Basics of Spin Transfer Torque (STT)

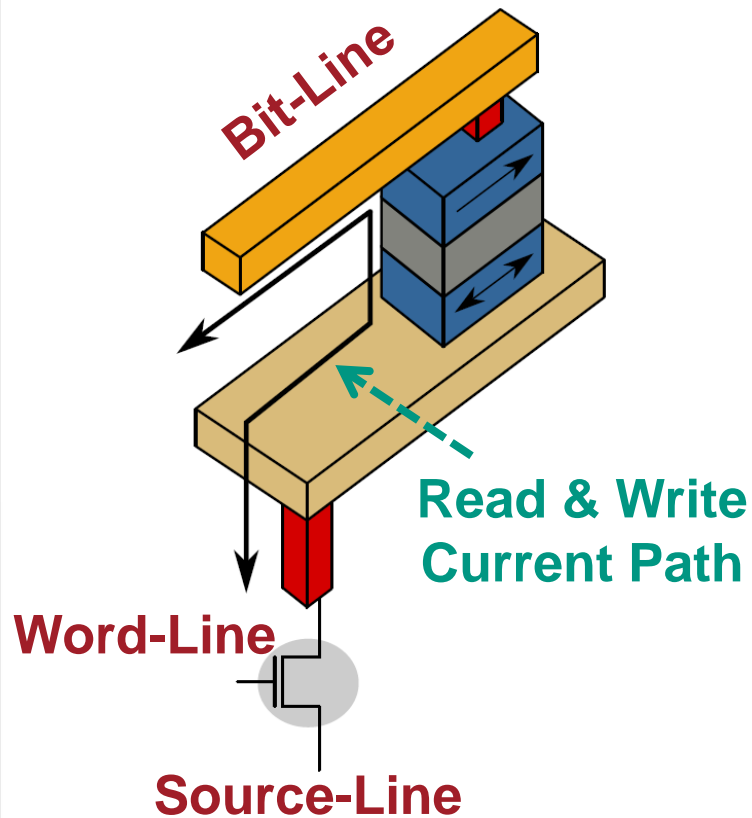


**Parallel Magnetisation (P)**  
**Low Resistance**

**Anti- Parallel Magnetisation (AP)**  
**High Resistance**

- Two ferromagnetic layers separated by a oxide barrier layer
- Magnetic Tunneling Junction (MTJ) Cell is a storing device
- Value stored as a resistance state

# Bit-cell using STT based MTJ cell



- Bit-cell has three terminals:
  - Bit-Line
  - Word-Line
  - Source-Line
- Read current is unidirectional
- Write current is bidirectional
- Possible “Read Disturb”
  - Same path for read and write



# Merits & Demerits of STT

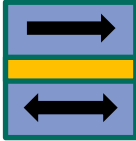
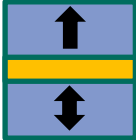
## ■ Merits:

- High Density
- Non-Volatility
- Scalability
- CMOS Compability
- Low Read Latency
- High Endurance
- High Retention
- Radiation Immune

## ■ Demerits:

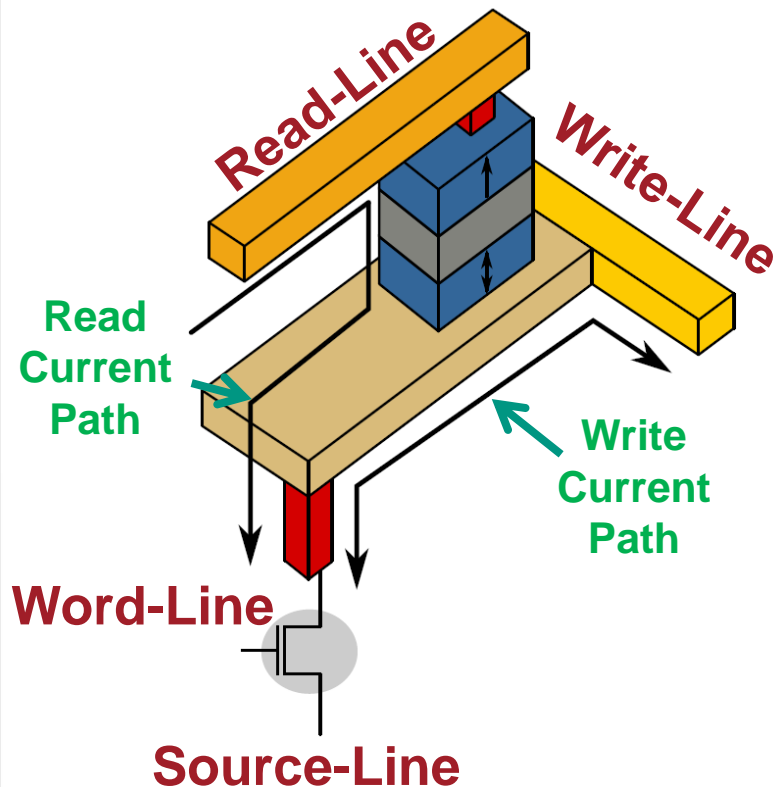
- High Write Power
- High Write Latency
- Aditonal Layer requires
- Read Disturb

# In-Plane Vs Perpendicular Anisotropy

| Parameters   | In-Plane Magnetic Anisotropy  | Perpendicular Magnetic Anisotropy   |
|--|---|---|
| Diagram  |    |  |
| Ratio of critical switching current to thermal stability, $\frac{I_c}{\Delta}$ | $\frac{\alpha}{\eta} \times \left( 1 + \frac{H_d}{2H_k} \right)$ <p>where, <math>\alpha</math>= damping constant,<br/> <math>\eta</math>=STT efficiency, <math>H_d</math> = demagnetization field, <math>H_k</math>=in-plane anisotropy field</p> | $\frac{\alpha}{\eta}$   |
| switching current  | High  | Low   |
| switching time   | More  | Less  |

- Perpendicular magnetic anisotropy
  - Low switching current
  - Less switching time
- “Read Disturb“ still remains challenge

# Spin Orbit Torque

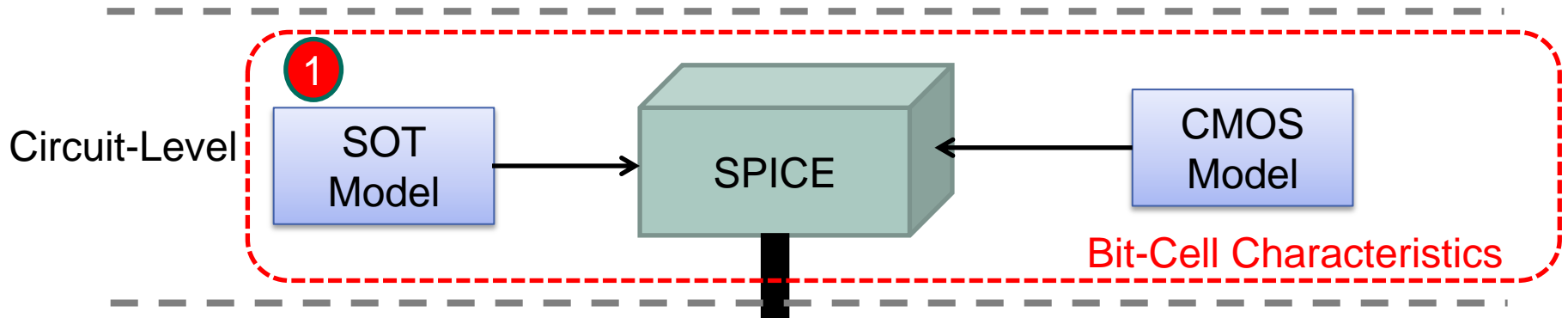


- Separate read and write current paths
  - One additional terminal
  - No read disturb
  - Need not to maintain ratio of  $I_{\text{Read}}/I_{\text{Write}}$
  
- Less current required to flip due to parallel magnetization
  
- Fast switching

# STT-MRAM VS SOT-MRAM

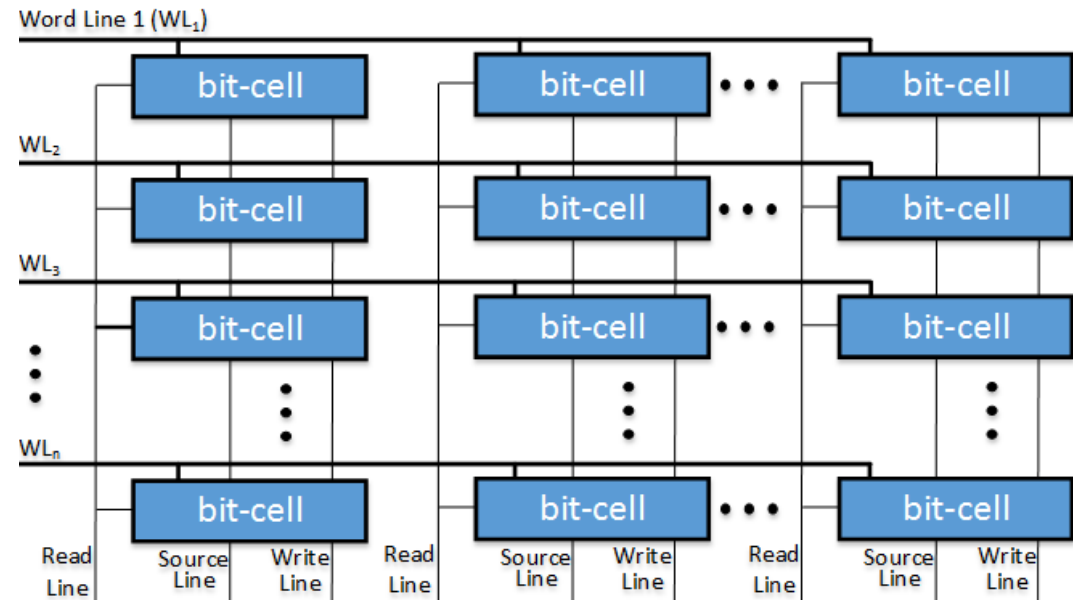
| Parameter                           | STT-MRAM             | SOT-MRAM           |
|-------------------------------------|----------------------|--------------------|
| Bit-cell Terminals<br>(1T1MTJ type) | 3                    | 4                  |
| Access Transistor                   | 8F                   | 2F                 |
| Current (uA)                        | 750                  | 100                |
| Write Current Period (ns)           | 11                   | 0.3                |
| Read Disturb                        | High Probability     | Almost Nil         |
| Read Energy (pJ)                    | 1.8                  | 1.8                |
| Write Energy (pJ)                   | 3.9 (reset)/3.4(set) | 0.1                |
| Switching Behavior                  | Asymmetrical         | Almost Symmetrical |
| Magnetic Anisotropy                 | In-Plane             | Perpendicular      |

# Tool Simulation Flow



# Basic Memory Architecture

- Word-line drive the access transistor of bit-cell
- Write Enable =1, for write operation
- Write Enable =0, for read operation



# Simulation models

## ■ STT-MTJ

- SPICE modelling framework presented in [ W. Guo, JAP-2010]

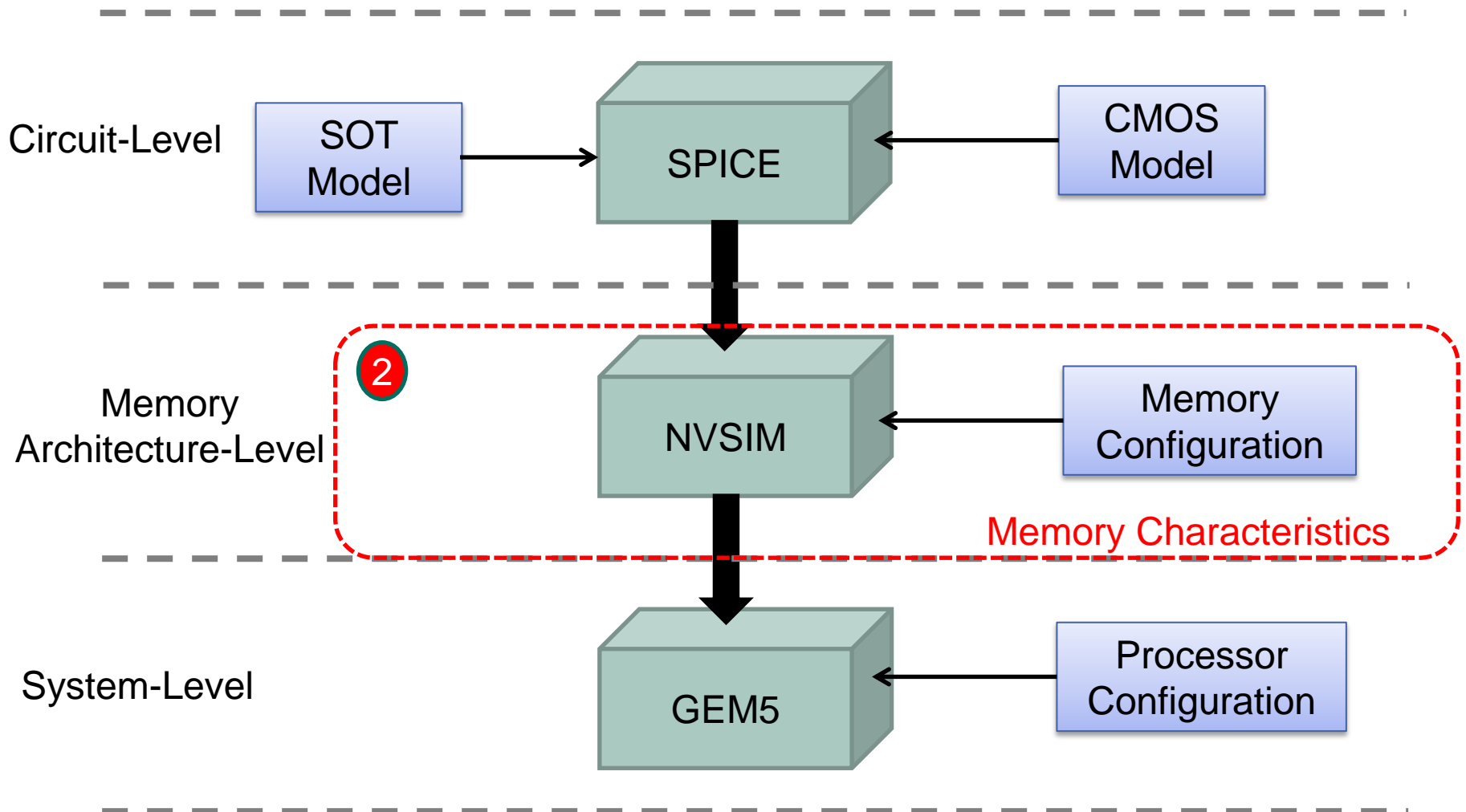
## ■ SOT-MTJ

- Compact Verilog-A framework presented in [ K. Jabeur, IJESSE-2013]

## ■ CMOS

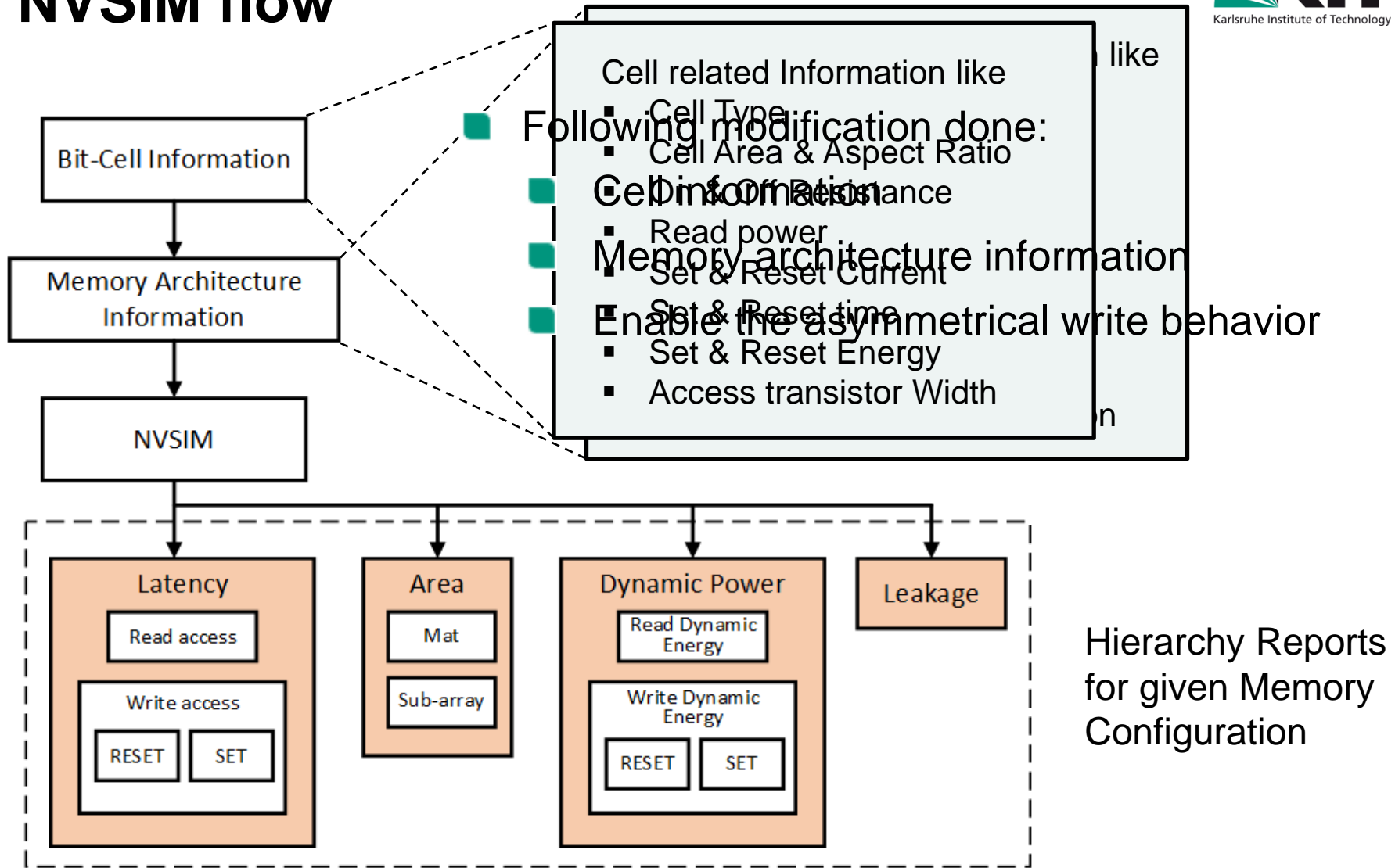
- General purpose TSMC 65nm models.

# Tool Simulation Flow



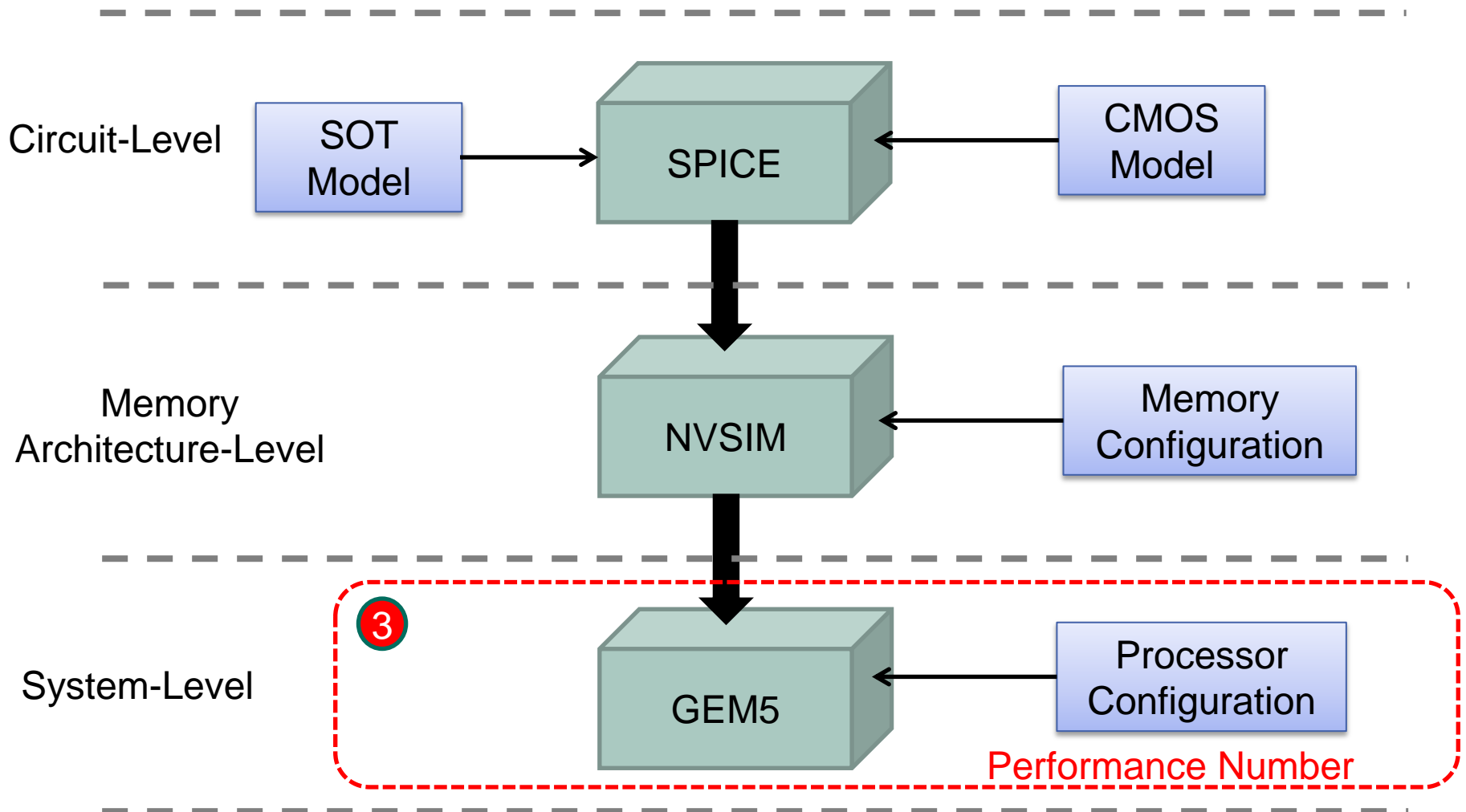


# NVSIM flow



Hierarchy Reports  
for given Memory  
Configuration

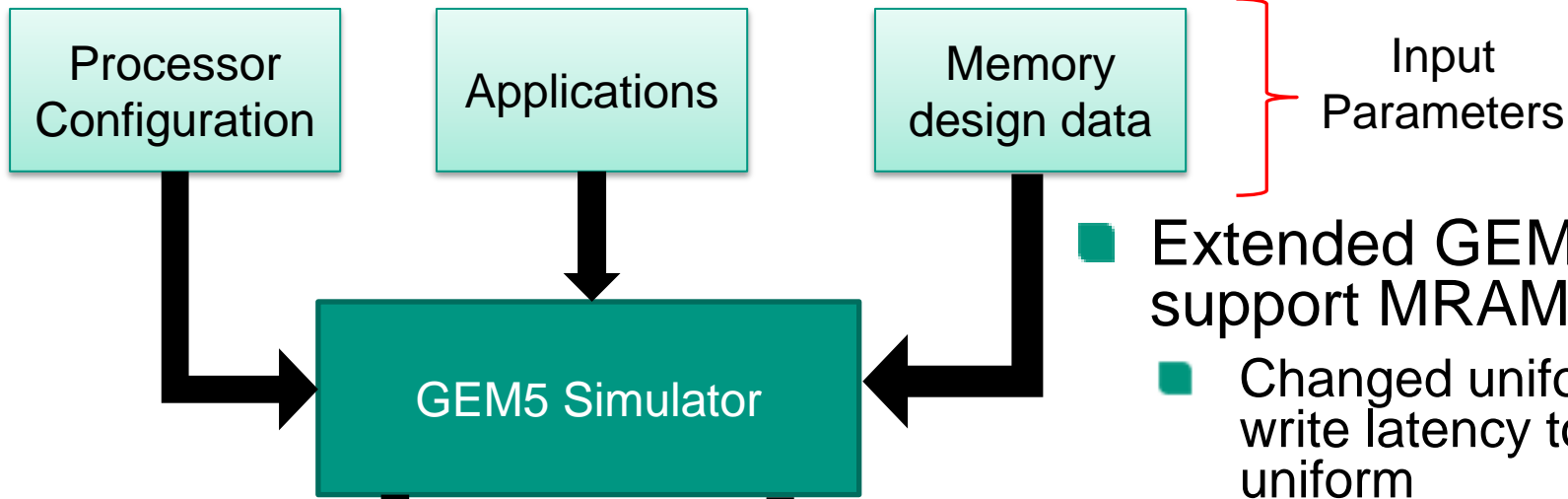
# Tool Simulation Flow



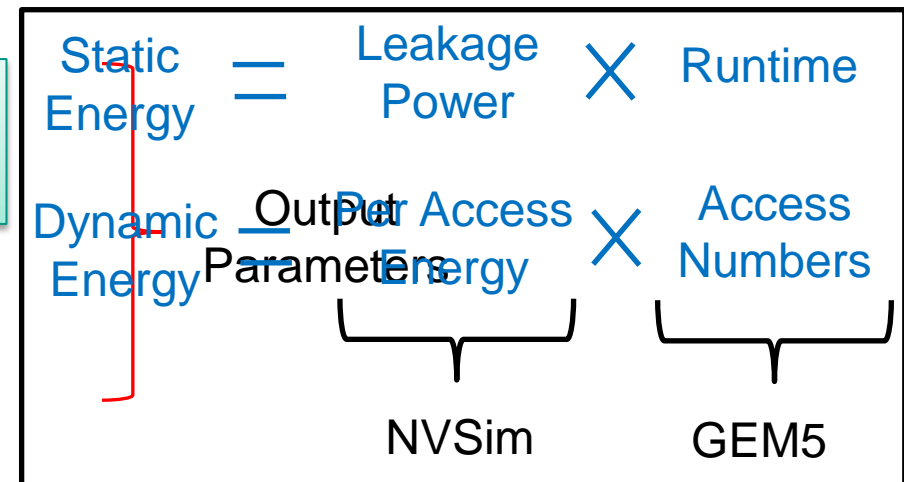
3

Performance Number

# Input & Output Parameters



- Extended GEM5 to support MRAM
- Changed uniform read-write latency to non-uniform



# Outline

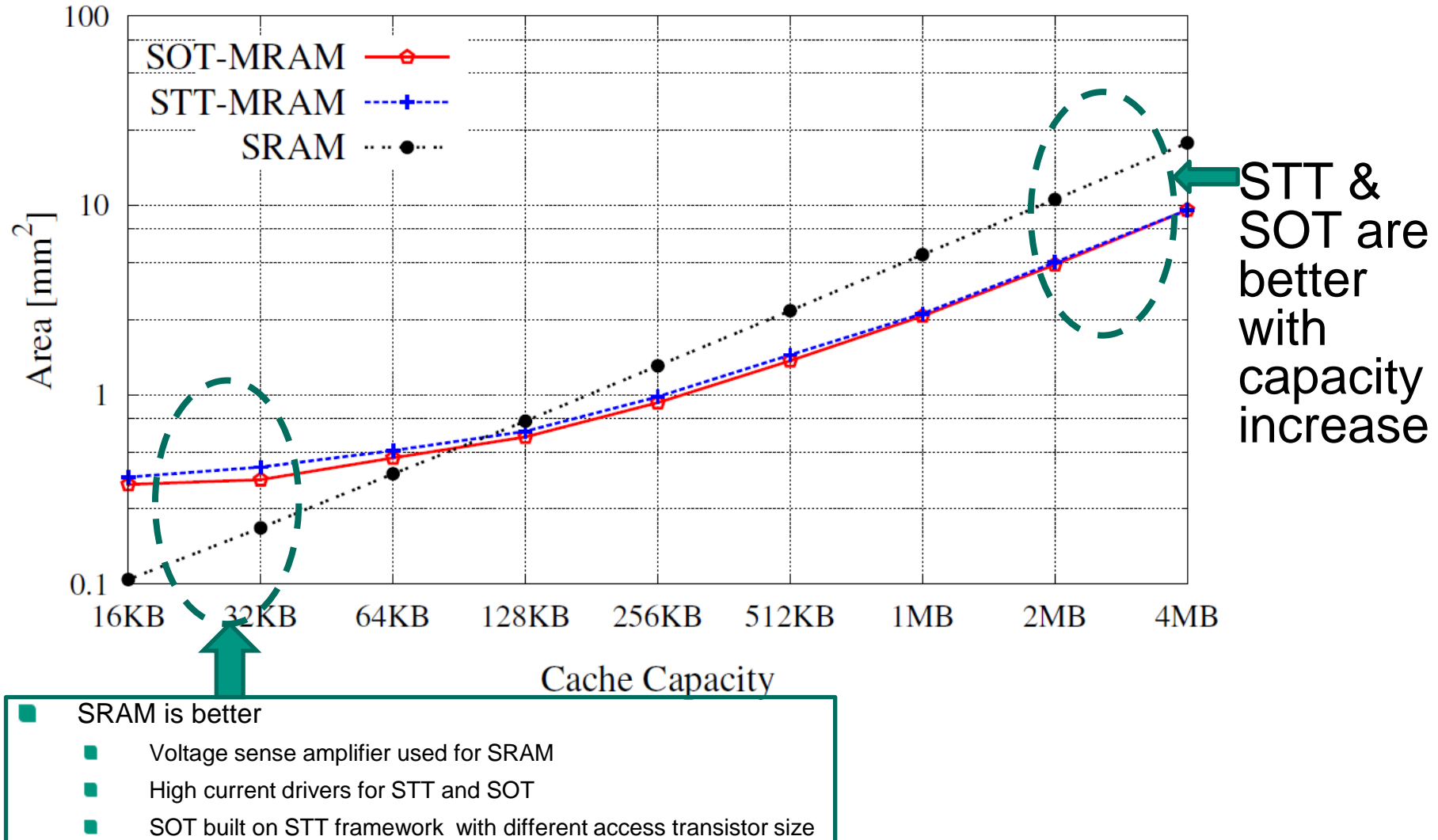
- Motivation
- SOT based MRAM
  - STT-MRAM & Limitations
  - Basics of SOT-MRAM
  - Simulation tool flow
- **Results**
  - For various memory technologies
  - System-level
- Summary & Conclusion

# Comparison of various Memory Technologies

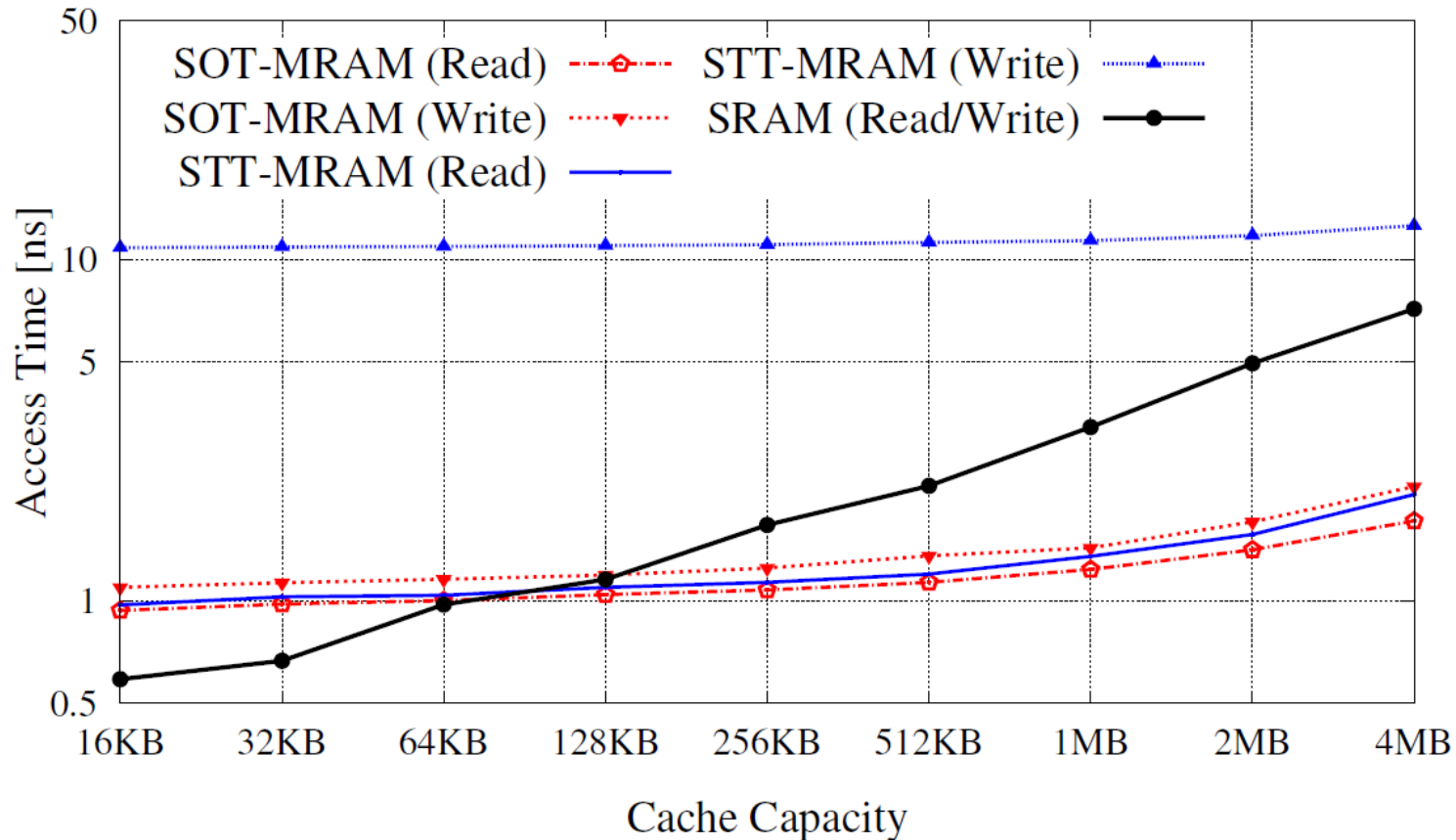
| Parameters              | SRAM | NAND<br>FLASH      | STT-<br>MRAM | SOT-<br>MRAM | PC-<br>RAM | R-<br>RAM |
|-------------------------|------|--------------------|--------------|--------------|------------|-----------|
| Area [mm <sup>2</sup> ] | 2.8  | 0.2                | 1.6          | <b>1.5</b>   | 0.3        | 0.7       |
| Read Latency [ns]       | 2.2  | 565                | 1.2          | <b>1.13</b>  | 0.6        | 1.2       |
| Write Latency [ns]      | 2.0  | 2× 10 <sup>5</sup> | 11.2         | <b>1.4</b>   | 150        | 21        |
| Read Energy [pJ]        | 587  | 3921               | 260          | <b>247</b>   | 363        | 193       |
| Write Energy [pJ]       | 355  | 6902               | 2337         | <b>334</b>   | 63670      | 592       |
| Leakage [mW]            | 932  | 77                 | 387          | <b>254</b>   | 153        | 115       |

- Values are extracted using NVSim for
  - 512 Kbyte capacity
  - Latency optimization

# Area Comparison for various memory sizes

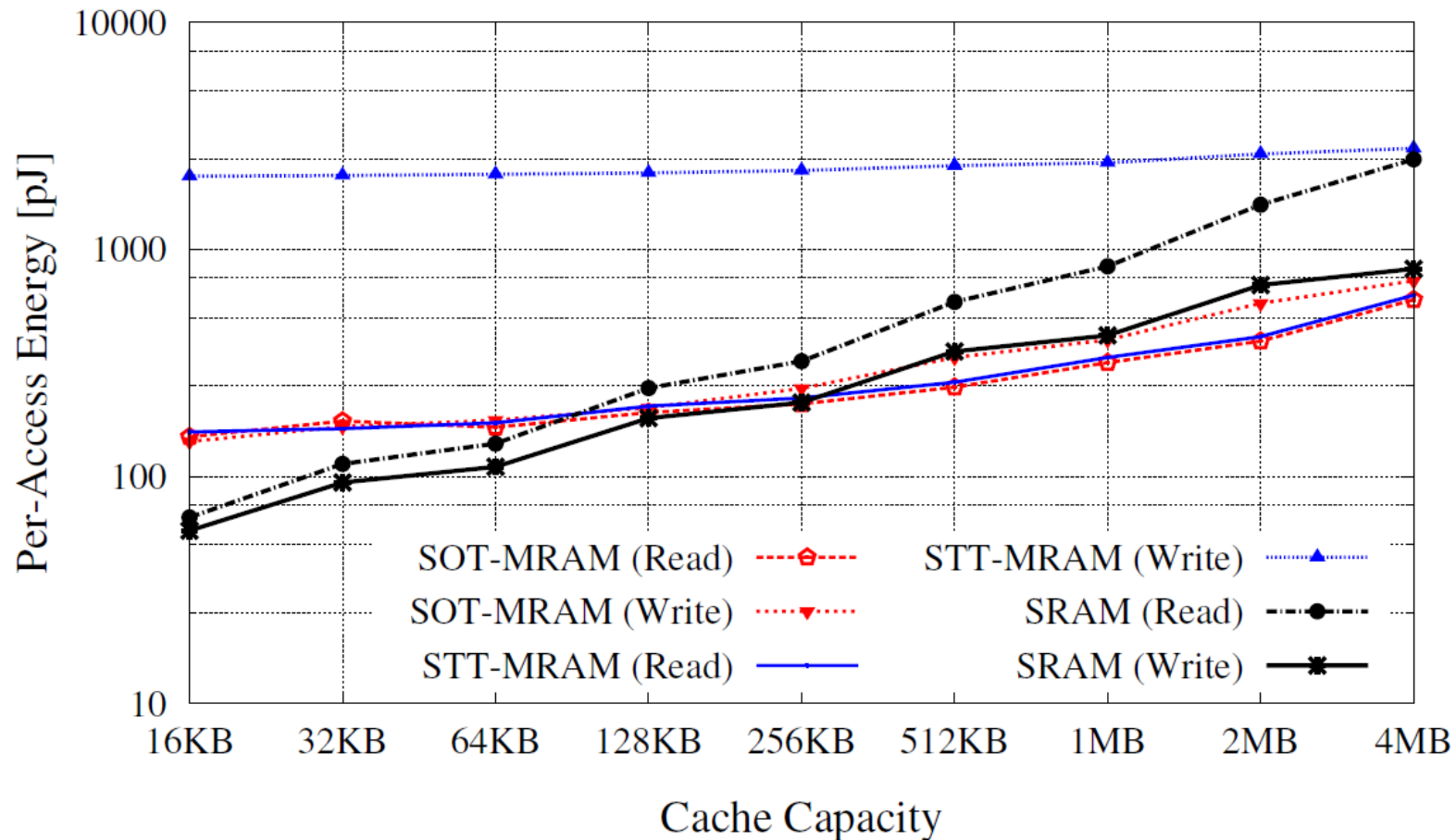


# Read & Write Latency Comparison



- SRAM varies linearly with capacity increase
- STT & SOT, remain almost flat with capacity increase

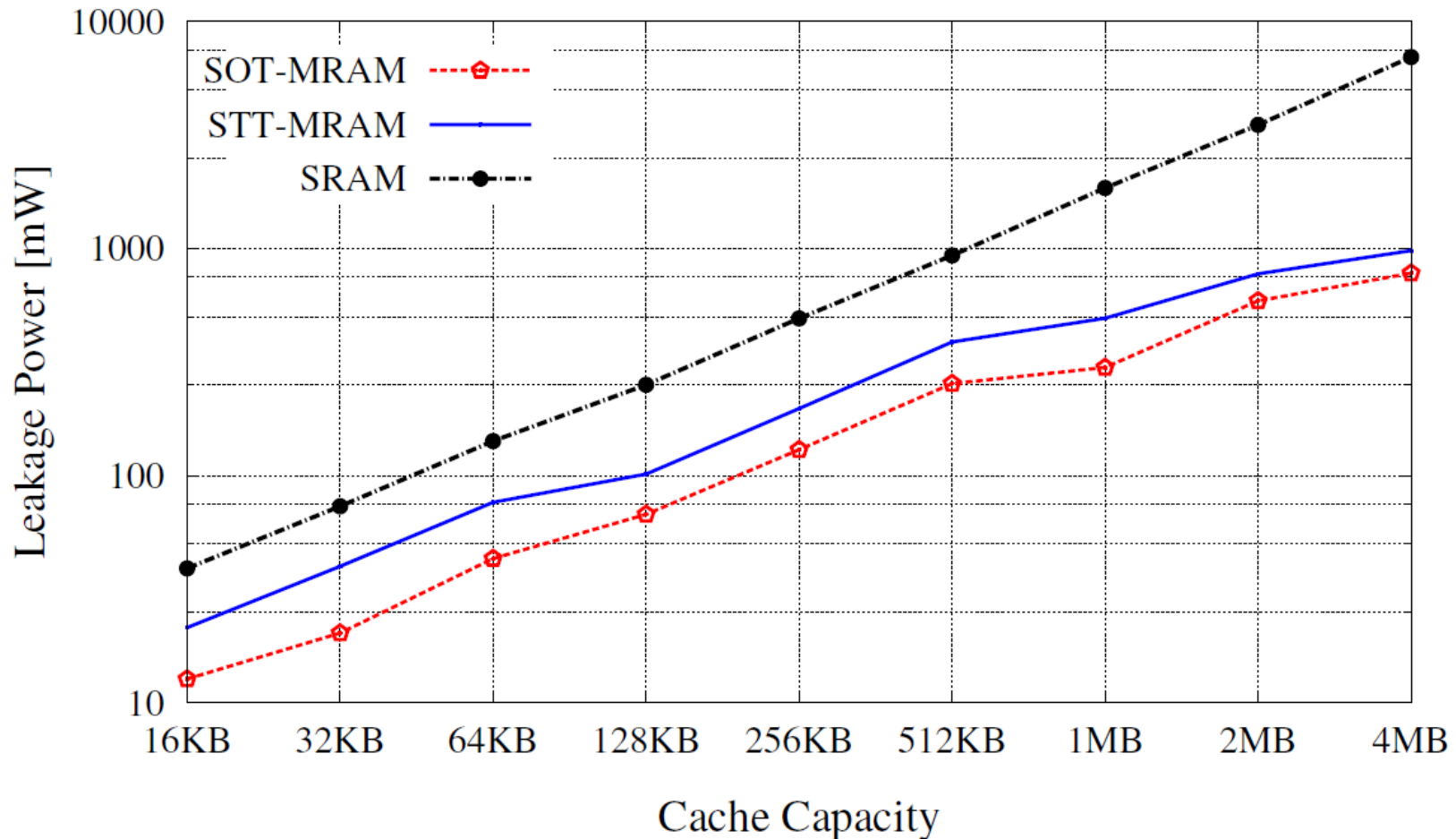
# Energy Comparison



■ SOT has almost same read & write access energy



# Leakage Comparisons

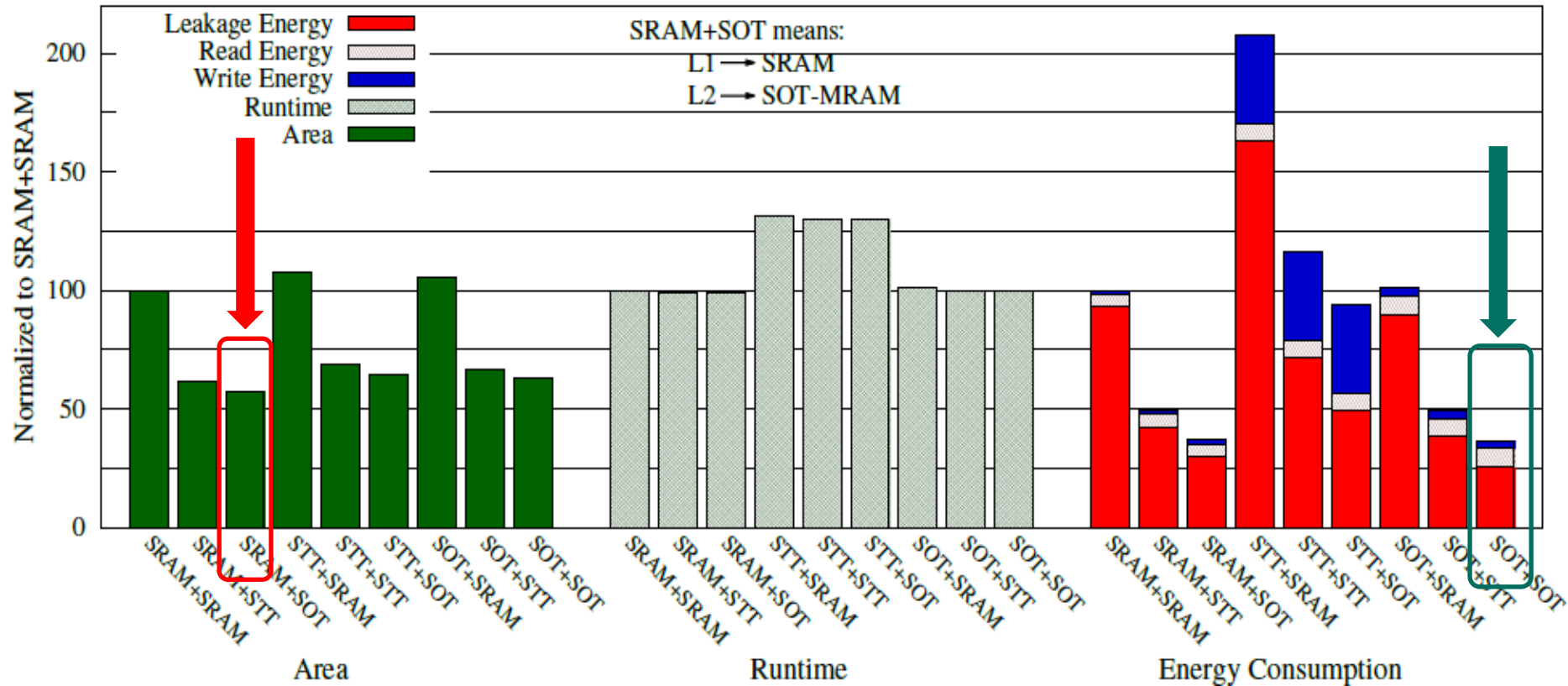


- SRAM varies linearly with capacity increase
- STT & SOT, leakage is due to periphery circuits

# System-Level Evaluation

- Configuration details for Experiments:
  - Processor : single core, 3 GHz
  - L1-Cache : 32 Kbyte with 64B Data Width
  - L2-Cache : 512 Kbyte with 64B Data Width
- Application (MiBench):
  - BasicMath, BitCnt, Qsort, Dijkstra, Patricia, StrSearch, SHA, CRC, FFT

# Comparisons with Various Cache conf.



- SRAM+SOT is best area combination.
- SOT+SOT is best energy configuration

# Benchmark Analysis

|              | Runtime [ms]  |              |              |               | Energy [mJ]   |             |             |             |
|--------------|---------------|--------------|--------------|---------------|---------------|-------------|-------------|-------------|
|              | SRAM+SRAM     | SRAM+STT     | SRAM+SOT     | SOT+SOT       | SRAM+SRAM     | SRAM+STT    | SRAM+SOT    | SOT+SOT     |
| BasicMath    | 61.4          | 59.8         | 59.8         | 60.5          | 66.4          | 31.6        | 23.6        | 22.8        |
| BitCount     | 130.1         | 130.1        | 130.1        | 130.1         | 133.8         | 63.0        | 45.6        | 40.4        |
| CRC          | 998.8         | 998.8        | 998.8        | 1025.5        | 1075          | 531.5       | 398.1       | 395.7       |
| Dijkstra     | 62.7          | 62.4         | 62.4         | 62.6          | 75.5          | 41.2        | 32.9        | 36.8        |
| FFT          | 176.1         | 175.4        | 175.3        | 176.1         | 191.9         | 95.5        | 72          | 71.6        |
| Patricia     | 49.1          | 46.7         | 46.7         | 47.6          | 54.6          | 25.8        | 19.5        | 19.4        |
| QSort        | 35.2          | 34.9         | 34.9         | 34.9          | 36.7          | 17.6        | 12.7        | 11.6        |
| SHA          | 23.3          | 23.3         | 23.3         | 23.3          | 26.1          | 13.4        | 10.3        | 10.7        |
| StringSearch | 1.5           | 1.5          | 1.5          | 1.5           | 1.7           | 0.9         | 0.7         | 0.7         |
| Average      | 170.9 (100 %) | 170.0 (99 %) | 170.0 (99 %) | 173.3 (101 %) | 184.6 (100 %) | 91.2 (49 %) | 68.4 (37 %) | 67.7 (36 %) |

- SOT only solution is best for low power.
- For runtime, the best combination is SRAM+SOT.

# Outline

- Motivation
  
- SOT based MRAM
  - STT-MRAM & Limitations
  - Basics of SOT-MRAM
  - Simulation tool flow
  
- Results
  - For various memory technologies
  - System-level
  
- **Summary & Conclusion**

# Summary & Conclusion

- Developed hybrid memory architecture based on SOT-MRAM
  - A cell-level information is extracted using SPICE simulations
  - NVSim tool is explored to estimate the design data
  - Many applications run using GEM5 simulator
- SOT is the best solution for low power
- Overall best is hybrid memory architecture SRAM+SOT