



HAL
open science

Applications des technologies mémoires MRAM appliquées aux processeurs embarqués

Luís Vitório Cargnini

► **To cite this version:**

Luís Vitório Cargnini. Applications des technologies mémoires MRAM appliquées aux processeurs embarqués. Micro et nanotechnologies/Microélectronique. Université Montpellier II - Sciences et Techniques du Languedoc, 2013. Français. NNT : 2013MON20091 . tel-01015187

HAL Id: tel-01015187

<https://theses.hal.science/tel-01015187>

Submitted on 26 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

Pour obtenir le grade de
Docteur

Délivré par **UNIVERSITÉ MONTPELLIER 2**

Préparée au sein de l'école doctorale
Information, Structures et Systèmes
Et de l'unité de recherche
LIRMM

Spécialité : **Systèmes Automatiques et Microélectronique**

Présentée par **Luís Vitório Cargnini**

**Applications des technologies
mémoires MRAM
Appliquées aux processeurs embarqués**

Soutenue le 12/11/2013 devant le jury composé de

Dr. Lionel TORRES	Professeur, Université Montpellier 2	Directeur de Thèse
Dr. Gilles SASSATELLI	Directeur de Recherche CNRS, LIRMM UMR CNRS 5506	Co-Encadrant
Dr. Guy Gogniat	Professeur, Université Bretagne-Sud, Lorient	Rapporteur
Dr. Bertrand GRANADO	Professeur, Université Pierre et Marie Curie, Paris	Rapporteur
Dr. Jacques-Olivier Klein	Professeur, Université Paris Sud 11, Paris	Président
Dr. Daniel Étienne	Professeur, Université Paris Sud 11, Paris	Examineur
Dr. Guillaume Prenat	Ingénieur-Docteur, CEA, Laboratoire SPINTEC, Grenoble	Examineur

THÈSE

présentée au Laboratoire d'Informatique de Robotique
et de Microélectronique de Montpellier pour
obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ MONTPELLIER II

Discipline : Génie Informatique, Automatique et Traitement du Signal
Formation Doctorale : Systèmes Automatiques et Microélectronique
École Doctorale : Information, Structures et Systèmes

**MRAM APPLIED TO EMBEDDED PROCESSORS
ARCHITECTURE AND MEMORY HIERARCHY**

NON-VOLATILE MEMORY MRAM INTO THE MEMORY HIERARCHY

par

Luís Vitório Cargnini

Jury composé de :

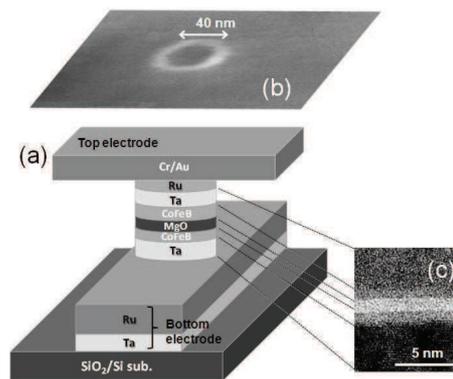
Lionel Torres	Professeur, Université Montpellier II/CNRS, Montpellier, France	<i>Directeur de thèse</i>
Gilles Sassatelli	Directeur de recherche/CNRS, Montpellier, France	<i>Co-encadrant</i>
Guy Gogniat	Professeur, Université de Bretagne Sud, Lorient, France	<i>Rapporteur</i>
Bertrand Granado	Professeur, Université Pierre et Marie Curie, Paris, France	<i>Rapporteur</i>
Jacques-Olivier Klein	Professeur, Institut d'Électronique Fondamentale, Université Paris Sud 11, Paris, France	<i>Président</i>
Daniel Étienne	Professeur, Lab. de Recherche en Informatique, Université Paris Sud 11, Paris, France	<i>Examineur</i>
Guillaume Prenat	Ingenieur, CEA Spintec, Grenoble, France	<i>Examineur</i>

LUÍS VITÓRIO CARGNINI

MRAM APPLIED TO EMBEDDED PROCESSORS ARCHITECTURE
AND MEMORY HIERARCHY

MRAM APPLIED TO EMBEDDED PROCESSORS ARCHITECTURE AND MEMORY HIERARCHY

LUÍS VITÓRIO CARGNINI



Non-volatile Memory MRAM into the memory hierarchy

Ph.D.

I2S

LIRMM – Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
Université de Montpellier 2

November 12, 2013 – version 7.3

Luís Vitório Cargnini: *MRAM applied to Embedded Processors architecture and Memory Hierarchy*, Non-volatile Memory MRAM into the memory hierarchy, Ph.D., ©, November 12, 2013

SUPERVISORS:

Lionel Torres

Gilles Sassatelli

LOCATION & TIME FRAME:

Montpellier, France, November 12, 2013

To my wife, that endured, my years of Ph.D., my absences.
I was lucky to have found you at our *Alma Mater* !
Thanks for all the trust, help and support you give me everyday,
we finally can go home !
To my thesis baby: now, dad have time to play !

New ideas pass through three periods:
1) *It can't be done.*
2) *It probably can be done, but it's not worth doing.*
3) *I knew it was a good idea all along!*
Arthur C. Clarke

Ohana means family.
Family means nobody gets left behind, or forgotten.
— Lilo & Stitch

Ad Verum Ducit
— Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) motto.

ABSTRACT

The Semiconductors Industry with the advent of submicronic manufacturing flows below 45nm began to face new challenges to keep evolving according with the Moore's Law. Regarding the widespread adoption of embedded systems one major constraint became power consumption of Integrated Circuit (IC). Also, memory technologies like the current standard of integrated memory technology for memory hierarchy, the **SRAM**, or the **FLASH** for non-volatile storage have extreme intricate constraints to be able to yield memory arrays at technological nodes below 45nm. One important is up until now Non-Volatile Memory weren't adopted into the memory hierarchy, due to its density and like flash the necessity of multi-voltage operation.

This thesis has the objective to work into these constraints and provide some answers. Into the thesis will be presented methods and results extracted from this methods to corroborate our goal of delineate a roadmap to adopt a new memory technology, non-volatile, low-power, low-leakage, SEU/MEU-resistant, scalable and with similar performance as the current **SRAM**, physically equivalent to **SRAM**, or even better with a area density between 4 to 8 times the area of a **SRAM** cell, without the necessity of multi-voltage domain like FLASH. This memory is the **MRAM** (Magnetic Memory), according with the ITRS one candidate to replace **SRAM** in the near future. **MRAM** instead of storing charge, they store the magnetic orientation provided by the spin-torque orientation of the free-layer alloy in the Magnetic Tunnel Junction (**MTJ**). Spin is a quantical state of matter, that in some metallic materials can have the orientation or its torque switched applying a polarized current in the sense of the desired field orientation.

Once the magnetic field orientation is set, using a sense amplifier, and a current flow through the **MTJ**, the memory cell element of **MRAM**, it is possible to measure the orientation given the resistance variation, higher the resistance lower the passing current, the sense will identify a logic zero, lower the resistance the SA will sense a one logic. So the information is not a charge stored, instead it is a magnetic field orientation, reason why it is not affected by **SEU** or **MEU** caused due to high energy particles. Also it is not due to voltages variations to change the memory cell content, trapping charges in a floating gate.

Regarding the **MRAM**, this thesis has by objective address the following aspects: **MRAM** applied to memory Hierarchy:

- By describing the current state of the art in **MRAM** design and use into memory hierarchy;
- by providing an overview of a mechanism to mitigate the latency of writing into **MRAM** at the cache level (Principle to composite memory bank);
- By analyzing power characteristics of a system based on **MRAM** on CACHE L1 and L2, using a dedicated evaluation flow
- by proposing a methodology to infer a system power consumption, and performance.
- and for last based into the memory banks analyzing a Composite Memory Bank, a simple description on how to generate a memory bank, with some compromise in power, but equivalent latency to the **SRAM**, that keeps similar performance.

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

BOOK CHAPTERS

Weisheng Zhao, Lionel Torres, Luís Vitório Cargnini, Raphael Martins Brum, Yue Zhang, Yoann Guillemenet, Gilles Sassatelli, Yahya Lakys, Jacques-Olivier Klein, Daniel Etiemble, Dafiné Ravelosona, Claude Chappert, "High Performance SoC Design Using Magnetic Logic and Memory", VLSI-SoC: Advanced Research for Systems on Chip, IFIP Advances in Information and Communication Technology, Volume 379, 2012, pp 10-33, DOI: 10.1007/978-3-642-32770-4_2, Print ISBN 978-3-642-32769-8, Online ISBN 978-3-642-32770-4, ISSN 1868-4238, Springer Berlin Heidelberg, http://link.springer.com/chapter/10.1007%2F978-3-642-32770-4_2

INTERNATIONAL CONFERENCES

Torres, L.; Brum, R.M.; Cargnini, L.V.; Sassatelli, G., "Trends on the application of emerging nonvolatile memory to processors and programmable devices," Circuits and Systems (ISCAS), 2013 IEEE International Symposium on , vol., no., pp.101,104, 19-23 May 2013, doi: 10.1109/ISCAS.2013.6571792, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6571792&isnumber=6571764>

Torres, L.; Brum, R.M.; Guillemenet, Y.; Sassatelli, G.; Cargnini, L.V., "Evaluation of hybrid MRAM/CMOS cells for reconfigurable computing," New Circuits and Systems Conference (NEWCAS), 2013 IEEE 11th International , vol., no., pp.1,6, 16-19 June 2013, doi: 10.1109/NEWCAS.2013.6573676, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6573676&isnumber=6573561>

Cargnini, Luis Vitorio; Torres, Lionel; Brum, Raphael Martins; Senni, Sophiane; Sassatelli, Gilles, "Embedded memory hierarchy exploration based on magnetic RAM," Faible Tension Faible Consommation (FTFC), 2013 IEEE , vol., no., pp.1,4, 20-21 June 2013, doi: 10.1109/FTFC.2013.6577780, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6577780&isnumber=6577746>

Weisheng Zhao, Lionel Torres, Yoann Guillemenet, Luís Vitório Cargnini, Yahya Lakys, Jacques-Olivier Klein, Dafine Ravelosona, Gilles Sassatelli, and Claude Chappert. 2011. Design of MRAM based logic circuits and its applications. In Proceedings of the 21st edition of the great lakes symposium on Great lakes symposium on VLSI (GLSVLSI'11). ACM, New York, NY, USA, 431-436., DOI: 10.1145/1973009.1973104, <http://doi.acm.org/10.1145/1973009.1973104>

Zhao, W.S.; Zhang, Y.; Lakys, Y.; Klein, J-O; Etiemble, D.; Revelosona, D.; Chappert, C.; Torres, L.; Cargnini, L.V.; Brum, R. M.; Guillemenet, Y.; Sassatelli, G., "Embedded MRAM for high-speed computing," VLSI and System-on-Chip (VLSI-SoC), 2011 IEEE/I-FIP 19th International Conference on , vol., no., pp.37,42, 3-5 Oct. 2011, doi: 10.1109/VLSISoC.2011.6081627, <http://dx.doi.org/10.1109/VLSISoC.2011.6081627>

Barthe, L.; Cargnini, L.V.; Benoit, P.; Torres, L., "The SecretBlaze: A Configurable and Cost-Effective Open-Source Soft-Core Processor," Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on , vol., no., pp.310,313, 16-20 May 2011, doi: 10.1109/IPDPS.2011.154, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6008911&isnumber=6008799>

Barthe, L.; Cargnini, L.V.; Benoit, P.; Torres, L., "Optimizing an Open-Source Processor for FPGAs: A Case Study," Field Programmable Logic and Applications (FPL), 2011 International Conference on , vol., no., pp.551,556, 5-7 Sept. 2011, doi: 10.1109/FPL.2011.107, <http://dx.doi.org/10.1109/FPL.2011.107> – Nominated the 6th best papers of FPL 2011 –

Cargnini, L.V.; Guillemenet, Y.; Torres, L.; Sassatelli, G., "Improving the Reliability of a FPGA Using Fault-Tolerance Mechanism Based on Magnetic Memory (MRAM)," Reconfigurable Computing and FPGAs (ReConFig), 2010 International Conference on , vol., no., pp.150,155, 13-15 Dec. 2010, doi: 10.1109/ReConFig.2010.10, <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5695297&isnumber=5695271>

NATIONAL CONFERENCES

Luis Vitorio Cargnini, Raphael Martins Brum, Gilles Sassatelli, Lionel Torres, "Exploration of MRAM cache memory for embedded processors", GDR SoC-SiP, 2012, June 13-15, Paris, France

Luís Vitório Cargnini, Raphael Brum, Yoann Guillemenet, Lionel Torres, Gilles Sassatelli, "Improving the Reliability of a FPGA using Fault-Tolerance Mechanism Based on Magnetic Memory (MRAM)", GDR SoC-SiP, 2011, June 15-17, Lyon, France

ACKNOWLEDGMENTS

*The limits of the possible can only be defined
by going beyond them into the Impossible.*
— Arthur C. Clarke

To my advisor Professor Dr. Lionel Torres, who have gave me the chance and opportunity to work in a fascinating field as MRAM. Thank you for your time and patience, to advise me all this years.

To my co-advisor Dr. Gilles Sassatelli for his insights during my research about the best approaches to evaluate an architecture.

To the reviewers of my thesis Professor Dr. Guy Gogniat and Professor Dr. Bertrand Granado, for having the time to read my manuscript and for gently accept being part of my thesis committee.

To Professor Dr. Jacques-Olivier Klein, Professor Dr. Daniel Étiemble, and Dr. Guillaume Prenat to also accept being part of my thesis committee.

To my research network, which I worked and co-authored with them my many thanks for their insights to Jacques-Olivier Klein, Daniel Étiemble, Weisheng Zhao and Claude Chappert.

To professor Dr. Fernando Gehm Moraes (PUCRS), by his help providing me access to the Cadence Physical Synthesis and Characterization tools, so I could fulfill and enrich my research. Also, for all the help trough this years.

To Raphael Garibotti, for daily opportunity on debating, discussing, exchanging ideas regarding both researches. Also, for the help on how to conduct the Integrated Circuit experiments, aggregating ideas to my research.

To Raphael Brum, for all the help during this last years at LIRMM writing a paper with tight dead-lines or discussing about the MTJ technology base in our researches.

To Marcelo Mandelli, to have saved me in the last minute when I needed a Mac to finish one the images of my thesis.

To Agence Nationale de Recherche (ANR) through SPIN,MARS and European Union (MODERN) for funding my research.

To Lorival Regis, who promptly and efficiently provided me all the technical resources that allowed me to conduct my research.

To Cécile Lukasik , for always so kindly help and solve the administrative related issues.

To God, that gave me strength, wisdom and perseverance all this years to conclude my Ph.D..

CONTENTS

i	INTRODUCTION	1
1	INTRODUCTION	3
ii	STATE OF THE ART, DRAFTED METHODOLOGIES AND MATERIALS	9
2	CURRENT STATE OF MEMORY TECHNOLOGY	11
2.1	Memory Taxonomy	17
2.2	Memory Devices	19
2.2.1	Ferroelectric Memory	19
2.2.1.1	Ferroelectric FET	19
2.2.1.2	Ferroelectric Polarization ReRAM	19
2.2.2	Nanoelectromechanical memory (NEMM)	20
2.2.3	Redox Memory	20
2.2.4	Mott Memory	21
2.2.5	Macromolecular Memory	22
2.2.6	Molecular Memory	22
2.2.6.1	Vertical Transistors	23
2.2.6.2	Two-terminal select devices (resistance-based memories)	24
2.2.6.3	Diode-type select devices	24
2.2.7	Resistive-Switch-type select devices	25
2.2.7.1	MIT switch	25
2.2.7.2	Threshold switch	25
2.2.7.3	MIEC switch	25
2.2.7.4	2 Terminal Switches	26
2.3	Magnetic Tunneling Junction (MTJ)	27
2.3.1	Tunnel Magnetoresistance (TMR)	29
2.3.2	Field Induced Magnetic Switching (FIMS) MRAMs	29
2.3.3	Thermally Assisted Switching (TAS) MRAMs	32
2.3.4	Spin Transfer Torque (STT) RAMs	32
2.3.5	STT based Magnetic Tunnel Junction (STT-MTJ)	33
2.3.6	Planar and Perpendicular STT	33
2.4	Current state compared to the memory state of the art	36
2.5	Computer Architecture and Memory Hierarchy	37
2.5.1	Working Principles of the Memory Hierarchy	38
2.6	CACHE Memories	40
2.6.1	Larger block size to reduce MISS rate	41
2.6.2	Bigger caches to reduce MISS rate	41
2.6.3	Higher associativity to reduce MISS rate	41
2.6.4	Multilevel caches to reduce MISS penalty	41
2.6.5	Giving priority to read misses over writes to reduce MISS penalty	41
2.6.6	Avoiding address translation during indexing of the cache to reduce HIT time	42
2.7	MRAM applied into Memory Hierarchy	42

iii	ANALYTICAL METHODOLOGY FLOW	47
3	ANALYTICAL METHODOLOGY FLOW	49
3.1	CACTI: Integrated Memory Simulator	49
3.1.1	CACTI background	50
3.1.2	CACTI Thesaurus	51
3.1.3	NUCA Modeling	52
3.2	NVSim	52
3.2.1	NVSim Non-Volatile Memory (NVM) Physical Mechanisms and Write Operations	53
3.2.2	NAND Flash	53
3.2.3	STT-RAM	53
3.2.4	PCRAM	53
3.2.5	ReRAM	54
3.2.6	Charge Pump	54
3.2.7	Write Endurance Issue	54
3.2.8	Retention Time Issue	55
3.2.9	MOS-Accessed Structure Versus Cross-Point Structure	55
3.3	Comparison of NVSim to CACTI	55
3.4	Reliability of the methodology base-line models	56
3.5	Methodology to Evaluate the Intrinsic aspects of MRAM compared to SRAM memories	57
3.5.1	Memory models used in NVSim for analyses	58
3.5.2	Intrinsic Analyses	58
3.5.3	Intrinsic Analyses - Results	62
3.5.3.1	Memory banks of 45nm	62
3.6	Methodology to Evaluate the aspects of MRAM into Memory Hierarchy	87
3.7	SimpleScalar	89
3.8	Gem5	90
3.8.1	Simulation Capabilities	90
3.8.2	ISAs	90
3.8.3	Execution Modes	90
3.8.4	CPU Models	90
3.8.5	Interconnection Networks	91
3.8.6	Devices	91
iv	ANALYSES OF EMBEDDED MEMORY HIERARCHY	93
4	ANALYSES OF EMBEDDED MEMORY HIERARCHY	95
4.1	Embedded MRAM for Processor Applications	96
4.1.1	MRAM applications for the processor memory hierarchy	96
4.1.2	In-Depth Analysis: Case Study CJPEG	98
4.2	Using Gem5 to evaluate the implications into memory hierarchy of L2 CACHE banks of significant sizes from 2GB down to 256KB.	102
4.2.1	Recalibrating the Gem5 using the NVSim bank details.	106
4.2.2	Power Analysis based into the memory banks generated after the NVSim.	114
4.3	Benchmarks used for evaluations	115

4.4	Comparing a Microprocessor memory hierarchy, synthesized with MRAM and SRAM at 28nm	116
4.4.1	OpenRISC Architecture	119
4.4.1.1	Memory Model	120
4.4.1.2	Memory Synchronize Instruction	121
4.4.1.3	Memory Management	121
4.4.2	MRAM Characterization using Liberty	122
4.4.3	Synthesis Results	123
4.4.4	MRAM and Static Random Access Memory (SRAM) synthesis results	124
4.4.5	Assessments comparing the synthesis results of the two technologies	125
4.5	The Composite Bank	127
4.5.1	Composite Memory banks of 45nm	127
V	CONCLUSIONS & FUTURE INSIGHTS	149
5	CONCLUSIONS & FUTURE INSIGHTS	151
vi	APPENDIX	155
A	MRAM TERMINOLOGY	157
A.1	Anisotropy	157
A.2	Magnetic anisotropy	157
A.3	Superparamagnetism	158
A.4	Magnetic Permeability	158
A.5	Isotropy	158
A.6	Magnetic moment	159
A.7	Magnetic moment and angular momentum	159
A.8	Exchange bias	160
A.9	Antiferromagnetism	160
A.9.1	Measurement	160
A.9.2	Antiferromagnetic materials	160
A.9.3	Geometric frustration	161
B	IMPROVING THE RELIABILITY OF A FPGA USING FAULT-TOLERANCE MECHANISM BASED ON MRAM	163
B.1	Reliability	164
B.2	Mechanism to improve reliability regarding SEU events	165
B.3	Scrubbing Mechanism - Refresh the configuration bits	167
B.4	General approach	167
B.5	Error Detection Code	169
B.6	Experimental Results	170
B.7	Hamming results	170
B.8	BCH Results	171
B.9	Assessments on MRAM applied to FPGA to improve reliability	171
C	INTRINSIC ANALYSES - ADDITIONAL RESULTS	173
D	CACTI ANALYTICAL MODELS - RELEVANT DETAILS	181
D.1	CACTI Analytical Models	181
D.2	Wire Parasitics	181

D.3	Global Wires	181
D.4	Low-swing Wires	183
D.5	Transmitter	183
D.6	Sense Amplifier	184
E	INTRINSIC ANALYSES - MEMORY BANKS OF 28nm	185
F	THE COMPOSITE BANK - ADDITIONAL RESULTS	205
F.1	LOP 45nm	205
F.2	Composite Bank 28nm	225
F.3	LOP 28nm	245
	BIBLIOGRAPHY	265

LIST OF FIGURES

Figure 1.1	Relationship among More Moore, More-than-Moore, and Beyond CMOS [ITRS, 2012a].	4
Figure 1.2	A Taxonomy for Emerging Research Information Processing Devices found in [ITRS, 2012a].	7
Figure 2.1	Physical difference between a vertical and a planar transistor, extracted from [Ou et al., 2010].	23
Figure 2.2	(a) MTJ structure and interconnection with the CMOS circuit and the relation of the MTJ according with the technology of anisotropy, (b) Planar MTJ, (c) Perpendicular MTJ.	28
Figure 2.3	Schematic of a 1-transistor, 1-MTJ memory cell showing the write lines above and below the bit and the read current path [Everspin Technologies Inc., 2010].	30
Figure 2.4	The magnetic tunnel junction (MTJ) material stack used for Toggle MRAM. The Free SAF magnetic moments switch between two states when the proper magnetic field sequence is applied. Electrons tunnel across the alumina (AlOx) tunnel barrier, resulting in a magnetoresistance that is sensitive to the magnetic moment direction of the sense layer [Everspin Technologies Inc., 2010].	31
Figure 2.5	Schematic of a toggle Magnetic Random Access Memory (MRAM) bit with the field sequence used to switch the free layer from one state to the other. The fields, H_1 , H_1+H_2 and H_2 are produced by passing currents, i_1 and i_2 , through the write lines [Everspin Technologies Inc., 2010].	31
Figure 2.6	MTJ process to switch the electromagnetic thin-layer field and store the logic value into the MTJ passing from the electrical layer.	32
Figure 2.7	High performance perpendicular MTJ with ϕ 40nm	35
Figure 2.8	Microprocessor Datapath [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012].	38
Figure 2.9	The levels in a typical memory hierarchy in embedded and server computers. As we move further away from the processor, the memory in the lower level increases in latency and density. Note that the time units change by factors of 10, from picoseconds to milliseconds, and that the size units change by factors of 1000, from bytes to terabytes.	38
Figure 2.10	An example of a possible Set-Associative Cache Organization [Patterson and Hennessy, 2012].	39
Figure 2.11	A general overview of a hypothetical memory hierarchy going from virtual address to L2 cache access. [Hennessy and Patterson, 2006]	42
Figure 3.1	CACHE bank memory physical organization [Tarjan et al., 2006]	50
Figure 3.2	Evaluation Methodology depicted.	57
Figure 3.3	Total Area.	67

Figure 3.4	CACHE Total Write Latency.	68
Figure 3.5	CACHE DATA Array Leakage Power.	69
Figure 3.6	CACHE TAG Leakage Power.	70
Figure 3.7	Leakage Power.	71
Figure 3.8	Low-Power (LOP) Total Leakage, observe only the Low Power Performance (LOP) banks observed in Figure 3.7.	72
Figure 3.9	Hit Dynamic Energy.	73
Figure 3.10	Hit Latency.	74
Figure 3.11	Miss Dynamic Energy.	75
Figure 3.12	Miss Latency.	76
Figure 3.13	Data Read Latency.	77
Figure 3.14	TAG Read Latency.	78
Figure 3.15	DATA Write Latency.	79
Figure 3.16	TAG Write Latency.	80
Figure 3.17	CACHE Write Dynamic Energy	81
Figure 3.18	TAG Dynamic Energy	82
Figure 3.19	TAG Write Latency	83
Figure 3.20	Data Write Latency	84
Figure 3.21	Write Bandwidth.	85
Figure 3.22	Write Bandwidth.	86
Figure 3.23	Evaluation Methodology depicted.	87
Figure 4.1	Overview of the processor performance using low-capacity L1 caches.	98
Figure 4.2	Overview of the processor performance using high-capacity L1 caches.	98
Figure 4.3	Overview of Cycles Per Instruction (CPI) Penalty: best-case, worst-case and average of the Mediabench benchmarks' performance.	99
Figure 4.4	Simulation results for 1000 cycles and 1 cycle of delay for each subsequent word in burst mode.	100
Figure 4.5	Simulation results for 1000 cycles and 10 cycles of delay for each subsequent word in burst mode.	101
Figure 4.6	% Miss Latency Encoder.	103
Figure 4.7	% Miss Latency Decoder.	103
Figure 4.8	Overall Miss Rate Encoder.	104
Figure 4.9	encoder L2 replacements.	104
Figure 4.10	decoder L2 replacements.	105
Figure 4.11	X.264 Execution time.	108
Figure 4.12	X.264 Clock cycles.	110
Figure 4.13	L2 Replacements.	111
Figure 4.14	L2 Overall Access Total.	111
Figure 4.15	L2 Overall Miss Latency.	112
Figure 4.16	L2 ReadReq Access total.	112
Figure 4.17	L2 Write Access Total.	113
Figure 4.18	The OR1200 is a 32-bit scalar RISC with Harvard microarchitecture, 5 stage integer pipeline, virtual memory support (Memory Management Unit (MMU)) and basic Digital Signal Processing (DSP) capabilities.	119

Figure 4.19	Translation of Effective to Physical Address. Simplified block diagram for 32-bit processor implementations	122
Figure 4.20	Synthesized OpenRISC architecture organization.	123
Figure 4.21	Composite CACHE set bank architecture, built with two heterogeneous memory technologies. In the Figure the TAG(green matrix, left) is built using SRAM, while the DATA (blue matrix, right) is built using Spin Transfer Torque MRAM (STT-MRAM).	128
Figure 4.22	Total Area.	128
Figure 4.23	Total Write Latency.	130
Figure 4.24	Composite Leakage Power.	131
Figure 4.25	Low-Power (LOP) Total Leakage, magnification on the LOP banks observed in Figure 4.24.	132
Figure 4.26	Hit Dynamic Energy.	133
Figure 4.27	Hit Latency.	134
Figure 4.28	Miss Dynamic Energy.	135
Figure 4.29	Miss Latency.	136
Figure 4.30	Data Read Latency.	137
Figure 4.31	TAG Read Latency.	138
Figure 4.32	Data array matrix Write Latency.	139
Figure 4.33	TAG Write Latency.	141
Figure 4.34	TAG Dynamic Energy	142
Figure 4.35	CACHE Write Dynamic Energy	143
Figure 4.36	Data Dynamic Energy	145
Figure 4.37	Write Bandwidth.	146
Figure 4.38	Write Bandwidth.	147
Figure B.1	MTJ process to switch the electromagnetic thin-layer field and store the logic value into the MTJ passing from the electrical layer.	166
Figure B.2	TAS-MRAM non-volatile memory cell, with two complimentary MTJs.	166
Figure B.3	The joint reliability of corrosion and Single Event Upset (SEU) in green (A) and in blue (B) the reliability applying the scrubbing technique, improving the reliability regarding SEU in 2x.	168
Figure B.4	The joint reliability of corrosion and SEU in green and in blue the reliability applying the refresh technique, improving the reliability regarding SEU over the joint reliability distribution in 2x.	169
Figure B.5	An overview of our general approach.	170
Figure D.1	Low-swing transmitter (actual transmitter has two such circuits to feed the differential wires).	182
Figure D.2	Sense Amplifier model adopted by CACTI.	184
Figure E.1	Total Area.	185
Figure E.2	Total Write Latency.	186
Figure E.3	Total Write Dynamic Energy.	187
Figure E.4	CACHE Data Array Leakage Power.	188
Figure E.5	TAG Leakage Power.	189
Figure E.6	Leakage Power.	190
Figure E.7	Low-Power (LOP) Total Leakage, zoom into it to observe only the LOP banks observed in Figure E.6.	191

Figure E.8	Miss Dynamic Energy.	192
Figure E.9	Hit Dynamic Energy.	193
Figure E.10	Hit Latency.	194
Figure E.11	Hit Latency.	195
Figure E.12	Miss Dynamic Energy.	196
Figure E.13	Miss Latency	197
Figure E.14	Data Read Latency.	198
Figure E.15	TAG Read Latency.	199
Figure E.16	Write Dynamic Energy	201
Figure E.17	Write Bandwidth.	202
Figure E.18	Write Bandwidth.	203
Figure F.1	Total Area.	205
Figure F.2	Total Write Latency.	206
Figure F.3	CACHE Data Array Leakage Power.	207
Figure F.4	TAG Leakage Power.	208
Figure F.5	Leakage Power.	209
Figure F.6	Low-Power (LOP) Total Leakage, zoom into it to observe only the LOP banks observed in Figure F.5	210
Figure F.7	Hit Dynamic Energy.	211
Figure F.8	Hit Latency.	212
Figure F.9	Miss Latency.	213
Figure F.10	DATA Read Latency.	214
Figure F.11	TAG Read Latency.	215
Figure F.12	DATA Write Latency.	216
Figure F.13	TAG Write Latency.	217
Figure F.14	Write Dynamic Energy	218
Figure F.15	DATA Dynamic Energy	219
Figure F.16	TAG Dynamic Energy	220
Figure F.17	TAG Dynamic Energy	221
Figure F.18	DATA Write Latency	222
Figure F.19	Write Bandwidth.	223
Figure F.20	Write Bandwidth.	224
Figure F.21	Total Area.	225
Figure F.22	Total Write Latency.	226
Figure F.23	Leakage Power.	227
Figure F.24	Low-Power (LOP) Total Leakage, close-up into only the LOP banks observed in Figure F.23	228
Figure F.25	Hit Dynamic Energy.	229
Figure F.26	Hit Latency.	230
Figure F.27	Miss Dynamic Energy.	231
Figure F.28	Miss Latency.	232
Figure F.29	DATA Read Latency.	233
Figure F.30	TAG Read Latency.	234
Figure F.31	DATA array matrix Write Latency.	235
Figure F.32	TAG Write Latency.	236
Figure F.33	CACHE Write Dynamic Energy	237
Figure F.34	DATA Dynamic Energy	238

Figure F.35	TAG Dynamic Energy	239
Figure F.36	TAG Dynamic Energy	240
Figure F.37	DATA Write Latency	241
Figure F.38	Write Bandwidth.	242
Figure F.39	Write Bandwidth.	243
Figure F.40	Total Area.	245
Figure F.41	Total Write Latency.	246
Figure F.42	CACHE Data Array Leakage Power.	247
Figure F.43	TAG Leakage Power.	248
Figure F.44	Leakage Power.	249
Figure F.45	Low-Power (LOP) Total Leakage, zoom into it to observe only the LOP banks observed in Figure F.44	250
Figure F.46	CACHE Hit Dynamic Energy.	251
Figure F.47	CACHE LOP Hit Latency.	252
Figure F.48	CACHE LOP Miss Latency.	253
Figure F.49	DATA Read Latency.	254
Figure F.50	TAG Read Latency.	255
Figure F.51	DATA Write Latency.	256
Figure F.52	TAG Write Latency.	257
Figure F.53	CACHE Write Dynamic Energy	258
Figure F.54	DATA Write Dynamic Energy	259
Figure F.55	TAG Write Dynamic Energy	260
Figure F.56	TAG Write Latency	261
Figure F.57	DATA Write Latency	262
Figure F.58	DATA Write Bandwidth.	263
Figure F.59	TAG Write Bandwidth.	264

LIST OF TABLES

Table 1.1	Semiconductors Constraints and challenges as defined by ITRS in [ITRS, 2012a]	5
Table 2.1	Memory Taxonomy, according last assessment of ITRS [ITRS, 2012a]	12
Table 2.2	Current Baseline and Prototypical Memory Technologies according [ITRS, 2012a]	13
Table 2.3	Emerging Research Memory Devices—Demonstrated and Projected Parameters for the next generation of memory technologies beyond 2020 [ITRS, 2012a]	15
Table 2.4	Potential of the Current Prototypical and Emerging Research Memory Candidates for Storage Class Memory (SCM) Applications [ITRS, 2012a]	18
Table 2.5	Labels detail for Table 2.4	18
Table 2.6	Experimental demonstrations according with [ITRS, 2012a] of vertical transistors in memory arrays.	23

Table 2.7	Benchmark Select Device Parameters, according ITRS report [ITRS, 2012a]	24
Table 2.8	2-Terminals Select Devices demonstrated experimentally [ITRS, 2012a]	26
Table 2.9	Comparison of NVM technologies [itr, 2011; Kim et al., 2011a; Yoda et al., 2012c]	36
Table 2.10	Timeline of the MTJ current power, according with technology. . .	37
Table 3.1	Memory Cell characterization models, employed to simulate the electrical and physical resulting bank.	58
Table 3.2	Details about SRAM and MRAM memory banks, generated using NVSim, this table comprises the 45nm results of the memory banks for LOP	60
Table 3.3	Details about SRAM and MRAM memory banks, generated using NVSim, this table comprises the 28nm results of the memory banks for LOP	61
Table 3.4	Details about SRAM and MRAM DATA memory arrays, this table comprises the 45nm LOP	63
Table 3.5	Details about SRAM and MRAM DATA memory arrays, this table comprises the 28nm LOP	64
Table 3.6	Details about SRAM and MRAM TAG memory arrays, this table comprises the 45nm LOP	65
Table 3.7	Details about SRAM and MRAM TAG memory arrays, this table comprises the 28nm LOP	66
Table 3.8	Total Area (mm ²).	67
Table 3.9	Cache Total Write Latency (ns).	68
Table 3.10	Cache Data Array Leakage Power (nW).	69
Table 3.11	Cache TAG Array Leakage Power (nW).	70
Table 3.12	Cache Total Leakage Power (mW).	71
Table 3.13	Cache Hit Dynamic Energy (nJ).	73
Table 3.14	Cache Hit Latency (ns).	74
Table 3.15	Cache Miss Dynamic Energy (nJ).	75
Table 3.16	Cache Miss Latency (ns).	76
Table 3.17	Data Read Latency (ns).	77
Table 3.18	TAG Read Latency (ns).	78
Table 3.19	DATA Write Latency (ns).	79
Table 3.20	Read Latency (ns).	80
Table 3.21	CACHE Write Dynamic Energy (nJ).	81
Table 3.22	TAG Write Dynamic Energy (nJ).	82
Table 3.23	TAG Write Latency (ns).	83
Table 3.24	Data Write Latency (ns).	84
Table 3.25	Write Bandwidth (GB/s)	85
Table 3.26	Write Bandwidth (GB/s)	86
Table 4.1	SimpleScalar baseline configuration used in all experiments	97
Table 4.2	SimpleScalar configuration used in Figures 4.1-4.3	97
Table 4.3	The Simulation Results of x264 encoder, for same system, only with different L2 CACHE memory bank, and Gem5 re-calibrated with NVSim latency details of the memory banks.	107

Table 4.4	Memory banks characteristics	109
Table 4.5	Power Consumption estimation regarding the Leakage current.	114
Table 4.6	Power Consumption estimation number of WriteBack Total, and the data about the memory bank technology Table 4.4	114
Table 4.7	Summary of instances and area for SRAM banks	124
Table 4.8	Summary of instances and area for MRAM banks	124
Table 4.9	MRAM area summary, regarding logic and memory.	125
Table 4.10	MRAM synthesis summary	125
Table 4.11	SRAM area summary, regarding logic and memory.	125
Table 4.12	SRAM synthesis summary	125
Table 4.13	Total Area (μm^2).	129
Table 4.14	Cache Write Latency (ns).	130
Table 4.15	Cache Total Leakage Power (mW).	132
Table 4.16	Cache Hit Dynamic Energy (nJ).	133
Table 4.17	Cache Hit Latency (ns).	134
Table 4.18	Cache Miss Dynamic Energy (nJ).	135
Table 4.19	Cache Miss Latency (ns).	137
Table 4.20	Data Read Latency (ns).	138
Table 4.21	TAG Read Latency (ns).	139
Table 4.22	DATA array matrix Write Latency (ns).	140
Table 4.23	TAG Write Latency (ns).	141
Table 4.24	TAG Write Dynamic Energy (nJ).	142
Table 4.25	Cache Write Dynamic Energy (nJ).	144
Table 4.26	Data Write Dynamic Energy (nJ).	145
Table 4.27	Write Bandwidth (GB/s)	146
Table 4.28	Write Bandwidth (GB/s)	147
Table B.1	Summary of the ECC synthesis results using CMOS ST 65nm standard cell library.	171
Table C.1	Details about SRAM and MRAM memory banks, generated using the NVSim, this table comprises the 45nm results of the memory banks for High Performance (HP)	174
Table C.2	Details about SRAM and MRAM memory banks, generated using the NVSim, this table comprises the 28nm results of the memory banks for HP	175
Table C.3	Details about SRAM and MRAM DATA memory arrays,this table comprises the 45nm HP	176
Table C.4	Details about SRAM and MRAM DATA memory arrays,this table comprises the 28nm HP	177
Table C.5	Details about SRAM and MRAM TAG memory arrays,this table comprises the 45nm HP	178
Table C.6	Details about SRAM and MRAM TAG memory arrays,this table comprises the 28nm HP	179
Table E.1	Total Area (μm^2).	186
Table E.2	Cache Write Latency (ns).	187
Table E.3	Cache Write Dynamic Energy (nJ).	188
Table E.4	Cache Data Array Leakage Power (nW).	189
Table E.5	Cache Tag Array Leakage Power (nW).	190

Table E.6	Cache Total Leakage Power (mW).	191
Table E.7	LOP Cache Total Leakage Power (mW).	192
Table E.8	Cache Miss Dynamic Energy (nJ).	193
Table E.9	Cache Hit Dynamic Energy (nJ).	194
Table E.10	Cache Hit Latency (ns).	195
Table E.11	Cache Hit Latency (ns).	196
Table E.12	Cache Miss Dynamic Energy (nJ).	197
Table E.13	Cache Miss Latency (ns).	198
Table E.14	Read Latency (ns).	199
Table E.15	Read Latency (ns).	200
Table E.16	Cache Write Dynamic Energy (nJ).	201
Table E.17	Write Bandwidth (B/s)	202
Table E.18	Write Bandwidth (B/s)	204
Table F.1	Total Area (μm^2).	206
Table F.2	Cache Write Latency (ns).	207
Table F.3	Cache Data Array Leakage Power (nW).	208
Table F.4	Cache Tag Array Leakage Power (nW).	209
Table F.5	Cache Total Leakage Power (mW).	210
Table F.6	Cache Hit Dynamic Energy (nJ).	211
Table F.7	Cache Hit Latency (ns).	212
Table F.8	Cache Miss Latency (ns).	213
Table F.9	DATA Read Latency (ns).	214
Table F.10	TAG Read Latency (ns).	215
Table F.11	Write Latency (ns).	216
Table F.12	Read Latency (ns).	217
Table F.13	Cache Write Dynamic Energy (nJ).	218
Table F.14	DATA Write Dynamic Energy (nJ).	219
Table F.15	TAG Write Dynamic Energy (nJ).	220
Table F.16	TAG Write Latency (ns).	221
Table F.17	DATA Write Latency (ns).	222
Table F.18	Write Bandwidth (B/s)	223
Table F.19	Write Bandwidth (GB/s)	224
Table F.20	Total Area (μm^2).	225
Table F.21	Cache Write Latency (ns).	226
Table F.22	Cache Total Leakage Power (mW).	227
Table F.23	Cache Hit Dynamic Energy (nJ).	229
Table F.24	Cache Hit Latency (ns).	230
Table F.25	Cache Miss Dynamic Energy (nJ).	231
Table F.26	Cache Miss Latency (ns).	232
Table F.27	DATA Read Latency (ns).	233
Table F.28	TAG Read Latency (ns).	234
Table F.29	DATA array matrix Write Latency (ns).	235
Table F.30	TAG Write Latency (ns).	236
Table F.31	Cache Write Dynamic Energy (nJ).	237
Table F.32	DATA Write Dynamic Energy (nJ).	238
Table F.33	TAG Write Dynamic Energy (nJ).	239
Table F.34	TAG Write Latency (ns).	240

Table F.35	DATA Write Latency (ns).	241
Table F.36	Write Bandwidth (GB/s)	242
Table F.37	Write Bandwidth (GB/s)	244
Table F.38	Total Area (μm^2).	245
Table F.39	Cache Write Latency (ns).	246
Table F.40	Cache Data Array Leakage Power (nW).	247
Table F.41	Cache Tag Array Leakage Power (nW).	248
Table F.42	Cache Total Leakage Power (mW).	249
Table F.43	Cache Hit Dynamic Energy (nJ).	250
Table F.44	Cache Hit Latency - LOP (ns).	252
Table F.45	Cache Miss Latency LOP (ns).	253
Table F.46	DATA Read Latency (ns).	254
Table F.47	TAG Read Latency (ns).	255
Table F.48	DATA Write Latency (ns).	256
Table F.49	TAG Read Latency (ns).	257
Table F.50	Cache Write Dynamic Energy - LOP (nJ).	258
Table F.51	DATA Write Dynamic Energy (nJ).	259
Table F.52	TAG Write Dynamic Energy (nJ).	260
Table F.53	TAG Write Latency (ns).	261
Table F.54	DATA Write Latency (ns).	262
Table F.55	DATA Write Bandwidth (B/s)	263
Table F.56	TAG Write Bandwidth (GB/s)	264

LIST OF ALGORITHMS

1	Stretch of the automated script.	88
2	Fragment of a TCL file to demonstrate how to characterize a memory library using Liberate MX.	118

ACRONYMS

ASIC	Application Specific Integrated Circuit	
BCH	Bose-Chaudhuri-Hocquenghem.	168
CMOS	Complementary Metal Oxide Semiconductor	3
CCS	Composite Current Source	117

CMB	Composite Memory Bank	79
CPU	Central Processing Unit	90
CNN	Cellular Nonlinear Network	6
CNT	Carbon Nano Tubes	20
CeRAM	Correlated Electron Random Access Memory	21
CIMS	Current induced magnetic switching	33
CPI	Cycles Per Instruction	xx
DFM	Design For Manufacturing	116
DRAM	Dynamic Random Access Memory	3
DNUCA	Dynamic Non-Uniform CACHE Access	44
DSP	Digital Signal Processing	xx
EDA	Electronic Design Automation	118
ECC	Error Correcting Code	163
FPGA	Field Programmable Gate Arrays	36
FeM	Ferromagnetic	33
FLASH	Floating-gate transistor NAND or NOR	4
FIMS	Field Induced Magnetization Switching	29
FeFET	Ferroelectric Field Effect Transistor	19
FL	Free-Layer	27
GMR	Giant Magnetoresistance	29
HDD	Hard Disk Drive	17
HP	High Performance	xxv
HPC	High Performance Computing	117
HRS	High Resistance State	24
IC	Integrated Circuit	ix
ITRS	International Technology Roadmap for Semiconductors	11
iMTJ	in-Plane or Parallel Magnetic Tunnel Junction	33
ISA	Instruction Set Architecture	89
LOP	Low Power Performance	xx
MTJ	Magnetic Tunnel Junction	ix
MRAM	Magnetic Random Access Memory	xix
MMU	Memory Management Unit	xx
MtM	More than Moore	4
MLC	Multi-Level Cell	17
MIEC	Mixed Ionic and Electronic Conduction Materials	25
MIM	Metal-Insulator-Metal	20
MIT	Metal Insulator Transition	25

MEU	Multiple Event Upset	
NVM	Non-Volatile Memory	xvi
NUCA	Non-Uniform CACHE Access	51
NEMM	Nanoelectromechanical Memory	20
NoC	Network-on-Chip	154
PCRAM	Phase Change Random Access Memory	4
PMA	Perpendicular Magnetic Anisotropy	34
PLL	Phase Locked Loop	116
PL	Pinned Layer	33
PCM	Phase-Change Memory	36
pMTJ	Perpendicular Magnetic Tunnel Junction	34
RRAM	Resistive Random Access Memory	4
ReRAM	Redox Resistive Random Access Memory	4
RTL	Register Transfer Level	
RL	Reference-Layer	27
RISC	Reduced Instruction Set Computing	119
STT	Spin Transfer Torque	7
STT-MRAM	Spin Transfer Torque MRAM	xxi
SEU	Single Event Upset	xxi
SRAM	Static Random Access Memory	xvii
SCM	Storage Class Memory	xxiii
SSD	Solid State Drive	17
SoC	System-on-Chip	40
SAF	Synthetic Antiferromagnet	29
SMT	Simultaneous Multi-Threading	91
SMP	Symmetric Multi-Processing	119
TAS-MRAM	Thermally Assisted Switching MRAM	32
TMR	Tunneling Magnetoresistance	29
TLB	Translation Lookaside Buffer	42
UCA	Uniform CACHE Access	50
VLSI	Very Large Scale Integration	7
VCM	valence change mechanism	20
WORM	Write Once Read Many	114

Part I

INTRODUCTION

From approximately 1.8 million years ago (*Homo Ergaster* appearance) to the current state, the mankind evolved and, since the advent of the CMOS circuits, the human knowledge is shifting exponentially at the speed of the Semiconductors industry. Such is the influence of Semiconductors in the society daily life that the advent of new memory technologies will, again, create a huge impact in the society. This thesis is just a fragment of this evolutionary cycle and will be another forgotten milestone in the years to come.

INTRODUCTION

Any sufficiently advanced technology is indistinguishable from magic.
— Arthur C. Clarke

This research work began sixty years ago with the advent of the Semiconductors [Museum, a,b]. There are two great milestones that made this thesis possible: the first one is the Williams tube, that won the race for a practical random-access memory [Williams and Kilburn, 1949]. The second one was the Core-Memory [Forrester, 1951] by Jay W. Forrester, who to put it simply, invented the magnetic memory applied to computing machines, long before the advent of pipelines, memory hierarchy or cache memories [Evans].

The magnetic memories were already experimented and debunked by integrated MOS memory [Moore, 1970; Vadasz et al., 1969]. So why try it again ? Well, the answer is quite simple: when Forrester first experimented with it, the materials and technology available were completely different. Currently we have sub-micronic nodes below 32nm. The planar manufacturing process did not have been invented yet or, at least, Fairchild Semiconductors did not make it public at the time. Also, the current state-of-the-art for the next generation is around the 10 ~ 14nm, therefore in the existing memory technologies like SRAM, FLASH and Dynamic Random Access Memory (DRAM), the memory cells cannot scale down below 28nm without major problems.

Due to the Moore's law [Moore, 2006], the downscaling of Semiconductors is driving information processing technology into a broadening spectrum of new applications. Many of these applications are enabled by performance gains and increased complexity. Since dimensional scaling of Complementary Metal Oxide Semiconductor (CMOS) eventually will reach fundamental limits, several new alternatives in micro/nano electronics and microarchitectures are being explored to sustain the historical integrated circuit evolution into the years to come. This is generating interest in new devices for information processing and memory, new technologies for heterogeneous integration of multiple functions (a.k.a. **More than Moore**) [Bergeron, 2008; ITRS, 2012a; Jammy, 2010; Kahng, 2010; Loke and Lai, 2008; Roy et al., 2013; van Roosmalen, 2004; Vigna, 2005; Wang et al., 2007; Zhang, 2007; Zhang et al., 2006a,b], and new paradigms for systems architecture.

To address this world beyond the Moore's law [Moore, 2006] it is necessary to extend to the functionality of CMOS, via heterogeneous integration of new technologies, and new information processing paradigms. The relationship between these domains is schematically illustrated in Figure 1.1. The expansion of the CMOS platform by conventional dimensional and functional scaling has been denominated **More Moore**. The CMOS platform can be further extended by the **More-than-Moore** approach, which is a new sub-

ject. On the other hand, new information processing devices and architectures have been classified as **Beyond CMOS** technologies. Owing to the fact that the Memory technologies like the **NVM** as the **MRAM**, Phase Change Random Access Memory (**PCRAM**), Resistive Random Access Memory (**RRAM**) can be categorized into this context, we can classify this thesis fundamentally as a Beyond CMOS thesis having its main approach on how to integrate such technology into existing platform. The heterogeneous integration of **Beyond CMOS**, as well as **More-than-Moore**, into **More Moore** will extend the **CMOS** platform functionality to form ultimately the **Extended CMOS** [ITRS, 2012a].

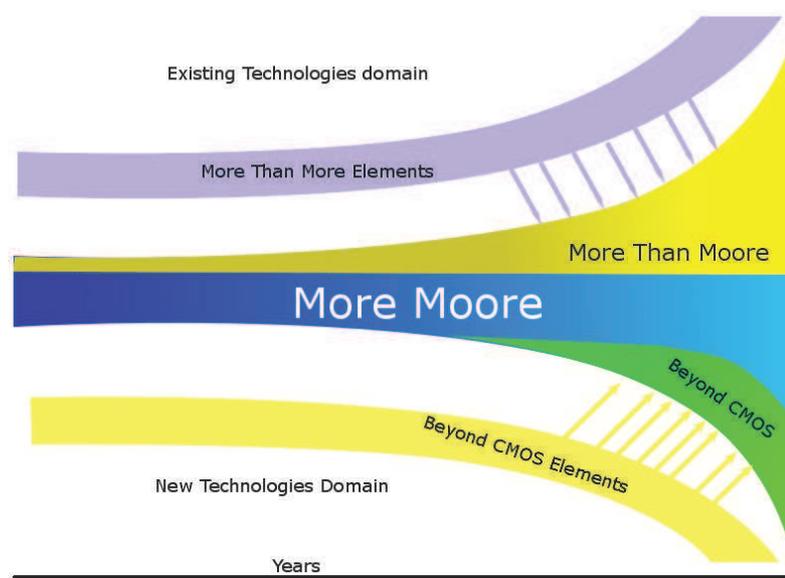


Figure 1.1: Relationship among More Moore, More-than-Moore, and Beyond CMOS [ITRS, 2012a].

According to [ITRS, 2012a], two emerging memory technologies: **STT-MRAM** and Redox Resistive Random Access Memory (**ReRAM**), exhibit potential to likely be ready for production within a decade at most. Also, the ITRS suggests that these technologies are attractive candidates for accelerated development and replacement of **SRAM** and Floating-gate transistor NAND or NOR (**FLASH**) [Masuoka et al.].

The semiconductor industry currently faces three major issues into extends the integrated circuit technology for new applications and beyond **CMOS** dimensional scaling. First is to propel **CMOS** to its ultimate density limits integrating a new memory technology that possesses performance, density and low-power on **CMOS** platform. Second, find alternative solutions to technologies that address existing More than Moore (**MtM**). For last, how to extend information processing beyond **CMOS**, combining on new devices, interconnection and architectural approaches to extend **CMOS**, developing a new information processing platform technology. These drawbacks, addressing the long term period of 2018 – 2026, are presented in Table 1.1 and more detailed in [ITRS, 2012a].

Table 1.1: Semiconductors Constraints and challenges as defined by ITRS in [ITRS, 2012a]

Challenges 2018 – 2026	Summary of Issues and opportunities
Scale performance, density, integrability, volatility, and non-volatile memory technologies to replace SRAM and FLASH for manufacture by 2018.	SRAM and FLASH scaling in 2D will reach definite limits in coming years [ITRS, 2012a]. These limits are driving the need for new memory technologies to replace SRAM and FLASH memories by 2018.
Scale CMOS to and beyond 2018 – 2026	Research and develop the next generation of materials to replace silicon, as an alternate channel and source/drain to increase the velocity and to further reduce V _{dd} and power dissipation in MOSFETs, while minimizing leakage currents.
Extend scaled CMOS as a platform technology into new domains of application.	Discover and improve, new device technologies and low-level architecture, providing special purpose optimized Application Specific Integrated Circuit (ASIC) cores (e. g., accelerator functions, microfluidics analyzes chips) heterogeneously integrable with CMOS. Research a technology capable of replacing CMOS. Ensure that a new information processing technology is compatible with any new memory technology. Such logic technology must also provide the access function in a new memory technology.
Research and reduce long term alternative solutions to technologies that address existing MtM, according with ITRS in wireless/analog and eventually in power devices, MEMS and image sensors	The industry is now faced with the increasing importance of a new trend, MtM, where added value to devices is provided by incorporating functionalities that do not necessarily scale according to Moore’s Law. Heterogeneous integration of digital and non-digital functionalities into compact systems that will be the key driver for a variety of application fields, such as communication, automotive, environmental control, health care, security and entertainment.

The drawbacks and issues driving development of emerging research devices are divided into those related to memory technologies, those related to information processing or logic devices, and those related to heterogeneous integration of multi-functional components (a.k.a. MtM) or Functional Diversification like denoted in Table 1.1 [ITRS, 2012a]. The current concern of Semiconductors industry is the need for a new memory technology that combines the best features of current memories in a fabrication technology compatible with CMOS process flow scaled beyond the present limits of SRAM and FLASH. This would provide a memory device technology required for both, off-the-chip and embedded memory applications.

Information processing to accomplish a specific system function, in general, requires several different interactive layers of technology. An objective in this section is to carefully delineate a taxonomy of these layers to further distinguish the scope of this thesis from that of other research thesis, delineating the current Semiconductors scenario and contextualizing this thesis into this taxonomy.

As shown in [Figure 1.2](#), a bottom-up representation of this taxonomy begins with the lowest physical layer represented by a computational state variable and ends with the highest layer represented by the architecture. In this more schematic representation, focused on generic information processing at the circuit level, a fundamental unit of information (e. g., a bit) is represented by a computational state variable, for example, the charge or voltage state of a node capacitance in CMOS logic. A device provides the physical means of representing and manipulating a computational state variable among its discrete states.

The data representation is how the computational state variable is encoded by the assembling of devices to process the bits or data. Two of the most common examples of data representation are binary digital and continuous or analog signaling. This layer is within the scope of the present thesis. The architecture plane encompasses three subclasses of this Taxonomy:

1. nano-architecture or the physical arrangement or assembling of devices to form higher level functional primitives to represent and execute a computational model;
2. the computational model that describes the algorithm by which information is processed using the primitives, e. g., logic, arithmetic, memory, Cellular Nonlinear Network (CNN);
3. the system-level architecture that describes the conceptual structure and functional behavior of the system or computational model.

The elements shown in the red-lined yellow boxes represent the current CMOS platform technology that is based on electronic charge so far. This state variable serves as the foundation for the Von Neumann computational system architecture [[Hennessy and patterson, 2006](#); [Patterson and Hennessy, 2012](#)]. Analog data representation also is included in the current CMOS platform technology. The other entries grouped in these five categories summarize individual approaches that, combined, may provide a new highly scalable information processing paradigm.

Based on the overview of the current state of Semiconductors as described above and in [[ITRS, 2012a](#)]. We can depict the taxonomy on [Figure 1.2](#). This taxonomy briefly describes the state of Semiconductors from 2016 and beyond. STT-MRAM can be classified into Devices → Spintronics and State Variable → Spin Orientation. That, as said, is the opening for the new paradigm of information process technologies.

Taking this factors as premisses, this thesis has as main objective to pave the way towards this new paradigm of processing information, discussing how to combine a emerging technology like MRAM into an existing CMOS based logic architecture of a microprocessor and its memory hierarchy. Also, exploring and evaluating scenarios and possibilities, given the available resources. To better understand the shift towards this new paradigm and during the course of the research for this thesis, we achieved some milestone contributions:

- An overall state of the art on the MRAM technology and its use into embedded systems;
- A characterization process to evaluate performance for a system based on MRAM. A methodology flow is proposed to compare SRAM and MRAM performance, especially in term of power consumption. We demonstrate that MRAM can bring

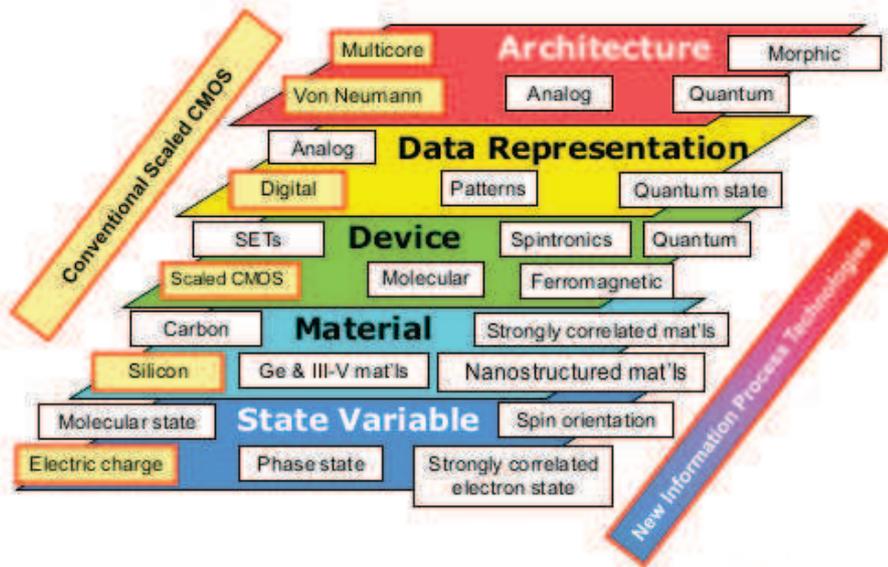


Figure 1.2: A Taxonomy for Emerging Research Information Processing Devices found in [ITRS, 2012a].

significant added value in the memory hierarchy of embedded systems, as well for cache memories L1 and L2;

- The third contribution is focused on the use of **MRAM**. For such, we performed an in-depth study on how the memory hierarchy affects the overall performance based on pure **MRAM** or combining it with **SRAM**. Several scenarios are depicted depending of the applications target and the chosen architecture. The main result is that we can retain similar performances with **MRAM**, while substantially reducing the power consumption, especially that induced by leakage.

The dynamics of the **MRAM** field evolves really fast, specially with the renewed interest in perpendicular Spin Transfer Torque (**STT**). Also, Logic integrated circuit, where intelligent systems are integrated on a single chip die, is the key technology in modern systems. Ground-breaking innovation that forces paradigm shift now rarely occurs in the design and manufacturing technologies of Very Large Scale Integration (**VLSI**), which have already seen tremendous growth. However, based on the achievements of past projects, it has become evident that the fusion of nonvolatile spintronic devices, which retain information without energy usage with the semiconductor integrated circuits, will bring about a revolutionary change. This fusion will realize new integrated computing systems with high performance and ultra low power consumption that dramatically exceed conventional levels.

According with professor Ohno from Tohoku University [IKEDA et al., 2012; Ohno, 2012], the next breakthrough resides in Spintronics VLSI, logic built in Spintronics. So, **MRAM** is intrinsically tied to this new paradigm shift in information storage and VLSI manufacturing devices.

Part II

STATE OF THE ART, DRAFTED METHODOLOGIES AND MATERIALS

The present chapter establishes the state of the art in memory technology, provides a taxonomy on how to classify the existing memories and the emerging technologies according with ITRS. It will also introduce the Magnetic Tunnel Junction ([MTJ](#)) and the Magnetic Random Access Memory ([MRAM](#)) technology and the Spin Transfer Torque ([STT](#)) principle, the cornerstones of the thesis background. Also, it will explain how to explore and where into the memory hierarchy we envisage the [MRAM](#) usage.

 CURRENT STATE OF MEMORY TECHNOLOGY

*We have to abandon the idea that schooling
is something restricted to youth. How can it be,
in a world where half the things a man knows at 20
are no longer true at 40 - and half the things he knows at 40
hadn't been discovered when he was 20?
— Arthur C. Clarke*

The present chapter has as objective to provide an overview of the current state of Memory technologies. To such, we are heavily based into the reports of International Technology Roadmap for Semiconductors (ITRS), specifically the reports [for Semiconductors, 2011; ITRS, 2012b].

www.itrs.net

Specifically for this thesis, the most difficult aspect for any emerging memory technology, regarding new materials, is to find compounds with properties that will enable operation of emerging research devices in high density at the nanometer scale. It is known that perpendicular STT-MRAM challenges the SRAM technology performance and scalability. But the major drawbacks are the MTJ compounds and manufacturing process, key elements for all emerging memories.

Table 2.1 is an organization or taxonomy for the existing and emerging memory technologies into four categories according with [ITRS, 2012a]. A strong boundary exists on how to integrate these memory technologies into existing CMOS technology platform in a seamless manner. A goal is to provide the end user with a device that behaves similar to the existing CMOS silicon memory chips.

Details for Table 2.1

- A 1T1R – 1 transistor-1 resistor
 - 1D1R – 1 diode-1 resistor
 - 1T1C – 1 transistor-1 capacitor
 - 1T – 1 transistor
 - FB DRAM – floating body DRAM
 - FeFET – ferroelectric FET
 - Multiple T – multiple transistor
- B FeRAM –ferroelectric RAM with one transistor and one ferroelectric capacitor
- C Floating gate or charge-trapping

Table 2.1: Memory Taxonomy, according last assessment of ITRS [ITRS, 2012a].

Cell Element	Type	Non-volatility	Retention Time
1T1R or 1D1R [A]	STT-MRAM	Nonvolatile	> 10 years
	Phase change memory	Nonvolatile	> 10 years
	Nano-electro-mechanical memory	Nonvolatile	> years
	RedOx Memory	Nonvolatile	> years
	Mott Memory	Nonvolatile	> years
	Macromolecular memory	Nonvolatile	> years
	Molecular memory	Nonvolatile	> years
1T1C [A]	DRAM	Volatile	~ seconds
	FeRAM [B]	Nonvolatile	> 10 years
1T [A]	FB DRAM [A]	Volatile	< seconds
	FeFET memory [A]	Nonvolatile	> years
	Flash [C]	Nonvolatile	> 10 years
Multiple T [A]	SRAM	Volatile	large

Seeing that each of these new technologies attempts to mimic and improve on the capabilities of a present day memory technology, performance parameters are provided in Table 2.2 for existing baseline and prototypical memory technologies. These parameters provide relevant details, with the current and expected or projected performance.

Detailed notes for Table 2.2:

- A 2011 ITRS PIDS chapter, Tables PIDS7&8
- B Estimated as $E \sim 0.5 * CV^2$ for $C = 25\text{fF}$, $V_c = 0.55$ Volts (in 2011) and $V_c = 0.43$ Volts in 2024 (energy to refresh is not included).
- C Embedded applications (see the Embedded Memory Requirements table in the System Drivers chapter of [ITRS, 2012a])
- D SRAM memory state is preserved so long as voltage is applied.
- E Estimated for hot-electron injection as: $E_w \sim h * N * eV_d$ for the hot-electron injection efficiency $h \sim 10 - 5$, the drain voltage $V_d = 5$ Volt, and the number of stored electrons $N \sim 1000$ in 2011 ($F = 90\text{nm}$) and $N \sim 100$ in 2024 ($F = 25\text{nm}$)
- F Lower bound for Fowler Nordheim write/erase.
- G [Takashima et al., 2011]
- H [Waser, 2012]
- I Estimated as $E \sim 0.5 * q * A * V$ for $q = 13.5\text{mC/cm}^2$, $A = 0.33\text{mm}^2$, $V_c = 1.5$ Volts (in 2011) and $q = 30\text{mC/cm}^2$, $A = 0.069\text{mm}^2$, $V_c = 0.7$ Volts (in 2024).
- J [Kawahara, 2011]
- K [De Sandre et al., 2011]

Table 2.2: Current Baseline and Prototypical Memory Technologies according [ITRS, 2012a]

		Baseline Technologies					Prototypical technologies [A]		
		DRAM		SRAM [C]	Flash		FeRAM	STT-MRAM	PCM
Cell Elements		Off [A]	Embedded [C]		NOR [C]	NAND [A]			
		1T1C		6T		1T	1T1C	1(2)T1R	1T(D)1R
Feature size	2011	36	65	45	90	22	180	65	45
F(nm)	2024	9	20	10	25	8	65	16	8
Cell Area	2011	6F ²	(12 – 30)F ²	140F ²	10F ²	4F ²	22F ²	20F ²	4F ²
	2024	4F ²	(12 – 50)F ²	140F ²	10F ²	4F ²	12F ²	8F ²	4F ²
Read Time	2011	< 10ns	2ns	0.2ns	15ns	0.1ms	40ns[G]	35ns[J]	12ns[K]
	2024	< 10ns	1ns	70ps	8ns	0.1ms	< 20ns[H]	< 10ns	< 10ns
W/E Time	2011	< 10ns	2ns	0.2ns	1ms/10ms	1/0.1ms	65ns[G]	35ns[J]	100ns[K]
	2024	< 10ns	1ns	70ps	1ms/10ms	1/0.1ms	< 10ns[H]	< 1ns	< 50ns
Retention Time	2011	64ms	4ms	[D]	10y	10y	10y	> 10y	> 10y
	2024	64ms	1ms	[D]	10y	10y	10y	> 10y	> 10y
Write Cycles	2011	> 1E16	> 1E16	> 1E16	1E5	1E4	1E14	> 1E12	1E9
	2024	> 1E16	> 1E16	> 1E16	1E5	5E3	> 1E15	> 1E15	1E9
Write Operating Voltage (V)	2011	2.5	2.5	1	10	15	1.3 – 3.3	1.8	3[K]
	2024	1.5	1.5	0.7	9	15	0.7 – 1.5	< 1	< 3
Read Operating Voltage (V)	2011	1.8	1.7	1	1.8	1.8	1.3 – 3.3	1.8	1.2
	2024	1.5	1.5	0.7	1	1	0.7 – 1.5	< 1	< 1
Write Energy (J/bit)	2011	4E – 15[B]	5.00E – 15	5.00E – 16	1E – 10[E]	> 2E – 16[F]	3E – 14[I]	2.5E – 12[A]	6E – 12[L]
	2024	2E – 15[B]	2.00E – 15	3.00E – 17	1E – 11[E]	> 2E – 17[F]	7E – 15[I]	1.5E – 13[A]	~ 1E – 15[M]

L Estimated as $E \sim 0.5 * I^2 R * t_w$ for $I = 224\text{mA}$, $R = 2.2\text{k}\Omega$, $t_w = 100\text{ns}$

M [Xiong et al., 2011]

The emerging technologies for beyond 2026 are organized around a set of six technologies presented in Table 2.3. These Technologies, according with the ITRS, will be the next milestone on memory technologies, following the MtM trend.

This technologies are in the initial states and still cannot be considered as replacement for existing technologies.

Detailed legend for Table 2.3:

A1 [Fitsilis et al., 2005]

A2 [Van Hai et al., 2010]

A3 [Ishiwara, 2009a]

A4 [Ishiwara, 2009b]

A5 [Tang et al., 2011]

A6 [Kaneko et al., 2011]

A7 Calculated based on the parameters of scaled FE capacitor projected in [Fitsilis et al., 2005]

B1 The projections for WRITE voltage and WRITE energy depend on the feature size (length) of nanoelectromechanical element. For very small length, the operating voltage might be too high for practical use, as follows from theoretical analysis in:[Choi et al., 2008]

B2 Corresponds to low-voltage operation, e.g. $\leq 1\text{V}$ [Choi et al., 2008]

B3 [Han et al., 2009]

B4 [Han et al., 2010]

B5 Estimated using 'nearly minimal' NEMS dimensions given in Table I of [B1]: the NEMS beam of length $L_{\text{beam}} = 75\text{nm}$, width $W_{\text{beam}} = 18\text{nm}$, thickness $T_{\text{beam}} = 10\text{nm}$, air gap $\chi = 3.5\text{nm}$, operating at voltage $V = 3\text{Volts}$. The switching energy can be estimated as the energy of charging the NEMS capacitance: $E = CV^2/2$. Alternatively, the 'mechanical' energy of the spring can be estimated as $E = kx^2/2$, where k is the cantilever spring constant. For the parameters, given in Table I of [Choi et al., 2008] both 'electrostatic' and 'mechanical' estimates yield a very close result

C1 [Zhirnov et al., 2011]

C2 [Lee et al., 2011a]

C3 [Dietrich et al., 2007]

C4 [Zhirnov et al., 2010]

C5 [Lee et al., 2010b]

Table 2.3: Emerging Research Memory Devices—Demonstrated and Projected Parameters for the next generation of memory technologies beyond 2020 [ITRS, 2012a].

		A. Emerging Ferroelectric memory	B. Nanomechanical Memory	C. Redox Memory	D. Mott Memory	E. Macromolecular Memory	F. Molecular Memories
<i>Storage Mechanism</i>		Remnant polarization on a ferroelectric dielectric	Electrostatically-controlled mechanical switch	Ion transport and redox reactions	Multiple mechanisms	Multiple mechanisms	Multiple mechanisms
<i>Cell Elements</i>		1T or 1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R	1T1R or 1D1R
<i>Device Types</i>		1) FET with FE gate insulator 2) FE barrier effects	NEMS	1) cation migration 2) anion migration	Mott transition	M-I-M (nc)-I-M	Bi-stable switch
<i>Feature size F</i>	Min. required	<65 nm	<65 nm	<65 nm	<65 nm	<65 nm	<65 nm
	Best projected	22 nm [A1]	>50 nm [B1, B2]	5 nm [C1]	5-10 nm	5-10 nm	5 nm [F1]
	Demonstrated	0.6 μm [A2]	500 nm [B3, B4]	30 nm [C2], 9nm [C7]	10 μm [D1]	130 nm [E1]	30 nm [F2]
<i>Cell Area</i>	Min. required	8F ²	8F ²	8F ²	8F ²	8F ²	8F ²
	Best projected	4F ²	4F ²	4F ²	4F ²	4F ²	4F ²
	Demonstrated	Data not available	Data not available	4F ² [C2], 8F ² [C3]	Data not available	4F ² [E1]	Data not available
<i>Read Time</i>	Min. required	<15 ns	<15 ns	<15 ns	< 15 ns	<15 ns	<15 ns
	Best projected	2.5 ns	<10 ns	<10 ns	< 10 ns	<10 ns	<10 ns [F1]
	Demonstrated	20 ns [A3]	Data not available	<50 ns [C3]	Data not available	10 ns [E1]	Data not available
<i>W/E time</i>	Min. required	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent
	Best projected	2.5 ns [A1]	<1 ns [B1, B2]	<1 ns [C4]	<1 ns [D2]	<10 ns	<40 ns [F1]
	Demonstrated	20 ns [A4]	~5 ns [B3, B4]	0.3ns [C5]	< 20 ns [D3]	15 ns [E2]	10s [F6], 0.2 s [F3]
<i>Retention Time</i>	Min. required	>10 y	>10 y	>10 y	>10 y	>10 y	>10 y
	Best projected	>10 y [A4]	>10 y	>10 y	Not known	Not known	Not known
	Demonstrated	~3.5 month [A6]	~days	>10 y [C2]	Not known	~year [E3]	1 hour [F6], 2 months [F4]
<i>Write Cycles</i>	Min. required	>1E5	>1E5	>1E5	>1E5	>1E5	>1E5
	Best projected	>1E16	>1E16	>1E16	>1E16	>1E16	>1E16
	Demonstrated	2E11 [A5]	~1E3 [B4]	1E12 [C2]	~1E2 [D4]	~1E5 [E4]	~2E3 [F2]
<i>Write operating voltage (V)</i>	Min. required	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent
	Best projected	<0.9 V [A1]	>1 V [B1, B2]	<0.5 V [C6]	Not known	<1 V [E5]	80 mV [F5]
	Demonstrated	±4[A4]	5 V [B3, B4]	0.6/-0.2 [C3]	1.25/0.75 V [D1]	~±2 V [E3]	4V [F6], ~±1.5 V [F2]
<i>Read operating voltage (V)</i>	Min. required	2.5	2.5	2.5	2.5	2.5	2.5
	Best projected	0.7	0.7	<0.2 V [C6]	Not known	0.7	0.3 [F1]
	Demonstrated	2.5 [A3]	1 [B3]	0.15 [C3]	0.2 [D1]	0.5 V [E3]	0.5V [F6], 0.5 V [F2]
<i>Write energy (J/bit)</i>	Min. required	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent	Application dependent
	Best projected	2E-15 [A7]	1E-17 [B5]	1E-17 [C4]	Not known	Not known	2E-19 [F6]
	Demonstrated	Data not available	Data not available	1E-13 [C7]	5E-13 [D5]	5E-11 [E6]	Data not available
<i>Comments</i>		Potential for non-destructive readout	Inverse voltage scaling presents a problem; Limited endurance	Potential for multi-bit storage; Low read voltage presents a problem	Retention requires additional mechanisms to maintain Mott transition conditions		160 Kbit prototype chip demonstrated [F3]

- C6 Electrochemical cell potentials control the write voltage. In appropriate combinations, 0.5V will leave some safety margin. Read voltages will be significantly smaller.
- C7 [Ho et al.]
- C8 Estimated based on experimental data reported in [Lee et al., 2011a]: $E \sim 0.5 * V * I * t_w$, for $V = 2V$, $I = 10\mu A$, $t_w = 10ns$.
- D1 [McWilliams et al., 2011]
- D2 The minimum value of the characteristic time for Mott transition was estimated to be $\sim 1ps$ based on the single element RC considerations [Stefanovich et al., 2000]
- D3 [Chae et al., 2005]
- D4 The number represent a recent experimental report on the Mott memory device [D1]. In an earlier paper on Mott switch (without memory functionality) up to $1E8$ switchnong cycles have been reported in [Guzman et al., 1996]
- D5 Estimated based on experimental data reported in [McWilliams et al., 2011] and [Chae et al., 2005]: $E \sim 0.5 * V * I * t_w$, for $V = 1V$, $I = 50\mu A$, $t_w = 20ns$
- E1 [Sang-Sun et al., 2010]
- E2 [Kuang et al., 2010]
- E3 [Park et al., 2011]
- E4 [Son et al., 2011]
- E5 [Müller et al., 2006]
- E6 Estimated based on experimental data reported in Ref. [G2] for the cell size $5x5\mu m^2$:
 $E \sim 0.5 * V * I * t_w$, for $V = 3Volts$, $I = 1mA$, $t_w = 15ns$.
- F1 [DeHon et al., 2005]
- F2 [Wu et al., 2005]
- F3 [Green et al., 2007]
- F4 [Chen et al., 2003]
- F5 [Meunier et al., 2007]
- F6 [Mukherjee et al., 2010]

2.1 MEMORY TAXONOMY

The purpose of a memory system is to store information and, therefore, memory capacity (or memory density) which is a critical parameter. In a typical memory system, the memory cells are connected to form a two-dimensional array, and it is essential to consider performance of memory cells in the context of their array architecture [Haraszti, 2000; Iniewski, 2010]. A memory cell in array is composed of two components: the **Storage node** and the **Selector**, which allows a given memory cell in an array to be addressed for read or write. Both components impact scaling limits for memory. For several concepts of resistance-based memories, the storage node can, in principle, be scaled down below 10nm [ITRS, 2012a]. The memory density will be limited by the Selector device.

The Table 2.1 provides a way to categorize memory technologies. In this sense, equivalent functional elements that compose a cell are identified. For example, the familiar DRAM cell, that consists of an access transistor and a capacitor storage node, is labeled as a 1T1C technology. STT-MRAM, where data is stored as the spin state in a magnetic oriented material, is represented as a 1T1R technology. Here, the resistance **R** indicates that the cell readout is accomplished by sensing the current through the cell. The usage of this approach of classification is to simplify cells (reduce cell area) by reducing the number of equivalent elements to a minimum. Therefore, early in the development of a given technology, it is common to see multi-transistor multi-x (x equals capacitor or resistor) cells. As knowledge evolves, the structures are scaled down to a 1T1x form [Haraszti, 2000; Iniewski, 2010; ITRS, 2012a]. In ultra-dense nanoelectronic memory arrays, instead of the transistor **T**, a two terminal non-linear diode-like element may be used with a resistive memory element. Such structure is represented as 1D1R technology.

An important property that differentiates emerging research memory technologies is whether data can be retained when power is not present. Nonvolatile memory offers essential use advantages, and the degree to which non-volatility exists is measured in terms of the period that data can be expected to be retained. Volatile memories also have a characteristic retention time that can vary from few milliseconds to the entire period of operation that power remains active.

The Table 2.4 list the Storage Class Memory (SCM) potential candidates and how they are classified, based on technologies listed in Table 2.2. Such device technologies combine the benefits of solid-state memory, as high performance and robustness, with the archival capabilities and low cost per bit of conventional magnetic Hard Disk Drive (HDD) or Solid State Drive (SSD). This devices requires a nonvolatile memory technology that can be manufactured at a low cost per bit. The potential of prototypical and emerging research memory devices for SCM applications, with or without Multi-Level Cell (MLC), is assessed in the context of existing commercialized storage technologies, namely the magnetic HDD and nonvolatile semiconductor FLASH memory [ITRS, 2012a].

The table Table 2.4, in this thesis is updated according current state-of-the-art

Table 2.4: Potential of the Current Prototypical and Emerging Research Memory Candidates for SCM Applications [ITRS, 2012a]

Parameter	Prototypical (Table 2.2)			Emerging (Table 2.3)					
	FeRAM	STT-MRAM	PCRAM	Emerging ferro-electric memory	Nano mechanical memory	Redox memory	Mott Memory	Macro molecular memory	Molecular Memory
Scalability									
MLC									
3D integration									
Fabrication cost									
Endurance									

Table 2.5: Labels detail for Table 2.4

Definitions	
	Scalability $F_{min} > 45nm$
	MLC difficult
	3D integration difficult
	Fabrication cost high
	Endurance $\leq 10^5$ write cycles demonstrated
	Scalability $F_{min} = 10 \sim 45nm$
	MLC feasible
	3D integration feasible
	Fabrication cost medium
	Endurance $\leq 10^{10}$ write cycles demonstrated
	Scalability $F_{min} < 10nm$
	MLC solutions anticipated
	3D integration feasible
	Fabrication cost potentially low
	Endurance $> 10^{10}$ write cycles demonstrated

2.2 MEMORY DEVICES

This section presents the next generation of memory technologies, that are currently under intensive research. Many of these ideas are not even in prototype stage yet. They represent the cutting-edge technology, **ReRAM** by instance seems to be the contender that will challenge **MRAM**.

2.2.1 *Ferroelectric Memory*

Emerging Ferroelectric Memory consists of two classes: 1) Ferroelectric FET and 2) Ferroelectric polarization **ReRAM**. This should not be misinterpreted with the conventional ferroelectric capacitor-based memory (**FeRAM** or **FRAM**), which is addressed in [ITRS, 2012a] and Table 2.2.

2.2.1.1 *Ferroelectric FET*

The Ferroelectric Field Effect Transistor (**FeFET**) memory is a **1T** memory device where a ferroelectric capacitor is integrated into the gate stack of a FET. The ferroelectric polarization directly affects charges in the channel and leads to a defined shift of the output characteristics of the FET. A typical **FeFET** memory element uses inorganic complex oxides or fluorides, such as $\text{PbZr}_x\text{Ti}_{1-x}\text{O}_3$, $\text{SrBa}_2\text{Ta}_2\text{O}_9$, BiMgF_4 , in the gate stack of a silicon FET. The main difficulty with these materials is interdiffusion and chemical reaction between the stack interfaces at the high deposition temperatures and high oxygen concentrations needed for deposition of the ferroelectric films on a Si substrate [Gerber et al., 2010; Ishiwara, 2009b].

Short retention of the **FeFET** memory raises question of its potential for application as nonvolatile memory, e.g. for the **SCM** technologies. Besides, **DRAM**-like applications are envisioned and the **FeFET** memory may have a potential for **SCM**, if scalability below 50nm can be demonstrated. Currently, new materials for the **FeFET** stacks are being actively investigated, such as organic ferroelectrics [Gerber et al., 2010; Heremans et al., 2011], nanotubes [Fu et al., 2009], nanowires [Sohn et al., 2010], and graphene [Zheng et al., 2011]. The **FeFET** memory scaling is projected to end approximately with the 22nm generation, due to the fact that the insulation layer becomes too thin and the properties of the ferroelectric, with respect to thickness dependency of the coercive field, will not allow further reduction [Fitsilis, 2005; Fitsilis et al., 2005].

2.2.1.2 *Ferroelectric Polarization ReRAM*

The ferroelectric polarization **ReRAM** is based on a structure where changing ferroelectric polarization can modify the charge transport properties of Ferroelectric films. The correlations between the resistance change and the ferroelectric switching are explained in terms of different mechanisms, such as modulation of the Schottky barrier [Kohlstedt et al., 2005], Ferroelectric tunnel junctions [Blom et al., 1994] and polarization-induced lattice strains [Bourim et al., 2011]. A serious challenge for practical ferroelectric **ReRAM** is typically low ferroresistive current (most ferroelectrics are insulating wide bandgap materials) [Jiang et al., 2011]. In order to obtain sufficiently high currents needed for the stable detection of the memory state, thin ferroelectric layers are required [Jiang et al., 2011], which constitutes a significant practical issue.

2.2.2 Nanoelectromechanical memory (NEMM)

The Nanoelectromechanical Memory (NEMM) is based on a bi-stable Nanoelectromechanical switch (NEMS). In this concept, mechanical digital signals are represented by displacements of solid nanoelements (e.g. nanowires, nanorods, or nanoparticles), which result in closing or opening of an electrical circuit. Several different modifications of suspended-beam/cantilever NEMMs are currently being explored with different materials, including Si [Ng et al., 2011], Ge [Andzane et al., 2009], TiN [Lee and Choi, 2011] and Carbon Nano Tubes (CNT) [Loh et al., 2011]. NEMM scaling analysis [Choi et al., 2008] suggests that it might be difficult to achieve low-voltage (~ 1 V) operation for the beam length of less than 50nm. Vertically oriented cantilever switches could reduce the NEMM area footprint [Ng et al., 2011]. In addition, nanoelectromechanical torsion switches has recently been demonstrated [Rubin et al., 2011; Xiang and Lee, 2010], which are claimed to have better scaling properties [Rubin et al., 2011].

There are also proposals for hybrid NEMS-floating gate memory devices aiming at the improvement of write/erase characteristics. In these devices, either floating gate [Tasuku Nagami et al., 2010] or control gate [Garcia-Ramirez et al., 2010] are built as a suspended bridge or cantilever [Lee et al., 2011b] separated by an air gap. The suspended bridge electrode can move within the gap under applied voltage, thus changing the separation between the control and floating gates, smaller for the fast write/erase and larger for longer retention time for the storage mode. Limited endurance is a serious issue of experimentally demonstrated NEMM devices, as they often fail after ~ 100 switching cycles [Andzane et al., 2009; Choi et al., 2009; Lee and Choi, 2011; Loh et al., 2011].

2.2.3 Redox Memory

The redox-based nanoionic memory operation is based on a change in resistance of a Metal-Insulator-Metal (MIM) structure caused by ion (cation or anion) migration combined with redox processes involving the electrode material or the insulator material, or both [Akinaga and Shima, 2010; Waser et al., 2009]. So far, were identified three classes of electrically induced phenomena that involves chemical effects, which are related to redox processes in the MIM cell. In these three ReRAM classes, there is a competition between thermal and electrochemical driving forces involved in the switching mechanism. The bipolar electrochemical metallization mechanism or memory effect (ECM), relies on an electrochemically active electrode metal, such as Ag. The drift of the highly mobile Ag^+ cations in the ion conducting I layer and their electromigration towards the (inert) counter electrode, leads to a growth of Ag dendrites. These dendrites form a highly conductive **filament** connecting the metal electrodes that results in the ON state of the cell [Valov et al., 2011]. Upon reversal of polarity of the applied voltage, an electrochemical dissolution of these filaments takes place, resetting the system into the high-resistance OFF state. The valence change mechanism (VCM) or memory effect occurs in specific transition metal oxides and is triggered by a migration of anions, such as oxygen anions. A subsequent change of the stoichiometry leads to a redox reaction expressed by a valence change of the cation and a change in the conductivity. This bipolar memory switching is induced by voltage pulses, where the polarity of the pulse determines the direction of the change, i.e. reduction or oxidation. A third class relies on a unipolar

thermo-chemical mechanism or memory effect (TCM, often called fuse-antifuse memory) which leads to a change of the stoichiometry due to a current induced increase of the temperature [Ielmini et al., 2011].

The material class for redox memory is comprised of oxides, higher chalcogenides (including glasses), semiconductors, as well as organic compounds including polymers. In some cases, a formation process required before the bi-stable switching can be started [Akinaga and Shima, 2010]. Often, the conduction is of filamentary nature. If this effect can be controlled, memories based on this bi-stable switching process can be scaled to very small feature sizes. The switching speed is limited by the ion transport. If the active distance, which is relevant for the redox controlled bi-stable switching, is small (around $< 10\text{nm}$) the switching time can be as low as a few nanoseconds. Many details of the mechanism of the phenomena are still unknown. Developing an understanding of the physical mechanisms governing switching of the redox memory is a key aspect for this technology. Recent experimental demonstrations of scalability, retention and endurance seems encouraging, and there is still room for improvements [Lee et al., 2011a; Yang et al., 2010]

stoichiometry: the relationship between the relative quantities of substances taking part in a reaction or forming a compound, typically a ratio of whole integers

2.2.4 Mott Memory

In the Mott memory, charge injection induces a transition from strongly correlated to weakly correlated electrons, resulting in an insulator-metal transition (IMT) or Mott transition. Electronic switches and memory elements based on the Mott transition (sometimes referred as Correlated Electron Random Access Memory (CeRAM)) has been explored using several materials systems, such as VO_2 [Ruzmetov et al., 2009], SmNiO_3 [Ha et al., 2011], NiO [Celinska et al., 2011; McWilliams et al., 2011; Xue et al., 2011b] and others. It is argued that the switching mechanism is activated by a critical electron population described by the Mott-Hubbard model [Haule and Kotliar, 2007; HELD et al., 1996; Hubbard, 1963; Lieb, 1993; Takahashi, 1994]. Recently, a reversible and non-volatile resistive switching has been reported for a class of Mott insulators AM_4X_8 ($A = \text{Ga, Ge}; M = \text{V, Nb, Ta}; X = \text{S, Se}$), and their potential for memory devices has been discussed [Cario et al., 2010].

A critical issue for this type of device is the sensitivity of the behavior of correlated electrons to small changes in parameters. Thereby, precise control of the physical and chemical structure of the material compounds are critical. For IMT in NiO , it was found that fine tuning of electronic phase transition is possible by $\text{Ni}(\text{CO})_4$ doping [Celinska et al., 2011; McWilliams et al., 2011]. Such doping stabilizes the oxygen vacancies resulting in a pure Mott transition system [McWilliams et al., 2011]. More recently, a new metal-insulator transition effect has been explored, which is based on formation of a quasi two-dimensional electron gas (2DEG) at the interface between two complex oxides [Niranjan et al., 2009; Ohtomo and Hwang, 2004; Park et al., 2009; Thiel et al., 2006]. For example, room-temperature switching of 2DEG nanowires $\text{LaAlO}_3/\text{SrTiO}_3$ grown on Si substrate has been demonstrated and the opportunities for nanoscale memory devices discussed [Park et al., 2009]

2.2.5 *Macromolecular Memory*

Macromolecular memory, also referred as polymer or organic resistive memory, consists of a memory element, with a film of organic material between two metal electrodes [Heremans et al., 2011; Scott and Bozano, 2007]. The organic film is typically relatively thick (~many monolayers, or few nanometers thick). Low fabrication cost is generally presented as the primary motivation for this type of memory, while extreme scaling is de-emphasized [Heremans et al., 2011]. The memory operation mechanisms are still unclear. Some research suggests that the changes in resistance could be due to intrinsic molecular mechanisms [Heremans et al., 2011], charge trapping [Lee et al., 2010a; Prime et al., 2009], or redox/ionic mechanisms [Heremans et al., 2011]. Material systems for macromolecular memory devices include different polymers and small-molecule organic compounds, e.g. polyimides [Park et al., 2011], polyfluorenes [Liu et al., 2011], PMMA [Son et al., 2011] (PMMA=poly(methylmethacrylate)), TCNQ (TCNQ = 7,7,8,8- tetracyano-p-quinodimethane) [Müller et al., 2009]. The active organic insulator layer in the macromolecular memory often contains embedded conductive components, such as metal nanoparticles [Park et al., 2011] and ultrathin graphite layers [Son et al., 2011]. The interaction of these conductive compounds remains unclear. Small macromolecular resistive memory arrays have been demonstrated [Kuang et al., 2010; Lee et al., 2010a; Xing-Hua et al., 2010], including a 3D-stack of three active layers [Cho et al., 2010; Song et al., 2010b].

2.2.6 *Molecular Memory*

Molecular memory is an extensive term, that englobes different proposals for using individual molecules, or small clusters of molecules, as building blocks of memory cells. In the molecular memory, data are stored by applying an external voltage that causes a transition of the molecule into a possible conduction states. The data is read by measuring resistance changes in the molecular cell. The concept emphasizes scaling: essentially, an element of information can be stored in the space of a single molecule [Pasupathy, 2003; Song et al., 2011]. Computing with molecules as circuit blocks is a concept with advantages over conventional circuit elements. Given their small size, highly dense circuits could be built, and bottom-up self-assembly of molecules in complex structures could be applied to improve top-down lithography fabrication techniques. As all molecules of one type are identical, molecular switches should have identical characteristics, therefore, reducing the problem of variability of components. However, the success of molecular electronics depends into the understanding of the phenomena accompanying molecular switching. Currently is unclear how it works. Early experiments on the reversible change in electrical conductance generated some interest [Reed et al., 2001; Tour et al., 2003].

Further studies revealed several drawbacks for single/few molecule devices due to extreme sensitivity of the device characteristics to the exterior parameters such as contacts, reproducible nanogap, environment and the list keeps growing. Also, there are multiple mechanisms contributing to the electrical characteristics of the molecular devices. The conductivity switching, as an intrinsic behavior of molecular switches, may often be masked by other effects, such as, in some cases, formation of metal filaments along the molecule attached between two metal electrodes [Lau et al., 2004]. In other

cases, intrinsic molecular switching has been reported and a 160-kbit molecular memory has been fabricated [Green et al., 2007]. Molecular memory is viewed as a long term research goal. The knowledge base for molecular electronics needs further fundamental work, which is currently under way [Cummings et al., 2011; Pro et al., 2009].

2.2.6.1 Vertical Transistors

Examples of experimental demonstrations of vertical select transistors used in memory arrays are presented in Table 2.6. While a vertical select transistor allows for the highest planar array density ($4F^2$), this technology is more difficult to integrate into stacked 3D memory than the conventional $8F^2$ technology using planar FETs. To avoid thermal stress on the memory elements of the existing layers, the processing temperature of the vertical transistor, as selection devices in 3D stacks, must be low. Also, making contact to the third terminal (gate) of vertical FET constitutes an additional integration challenge, which usually results in cells size larger than $4F$ [Zainuddin et al., 2011], although $4F^2$ arrays can, in principle, be implemented with 3-terminal select devices [Liu, 2009].

Table 2.6: Experimental demonstrations according with [ITRS, 2012a] of vertical transistors in memory arrays.

Reference	Technology	Memory Type	Array size	Cell size	Transistor	I_{on}	V_{on}	I_{on}/I_{off}
[Schloesser et al., 2004]	170nm	DRAM	1Mb	$8F^2$	DG FET	$50\mu A$	1.8V	10^{10}
[Song et al., 2010a]	80nm	DRAM	50Mb	$4F^2$	GAA FET	$30\mu A$	1.2V	10^{11}
[Kim et al., 2010a]	54nm	Z-RAM	ϕ	ϕ	DG FET	ϕ	0.5V	ϕ
[Servalli, 2009]	45nm	PCM	1Gb	$5.5F^2$	BJT	$300\mu A$	2V	ϕ
[Wang et al., 2010a]	180nm	ReRAM	ϕ	$4F^2$	BJT	$100\mu A$	1.2V	ϕ
[Yang et al., 2008a]	25nm (NW dia)	ϕ	ϕ	ϕ	NWGAA FET	$25\mu A$	1.2V	10^7

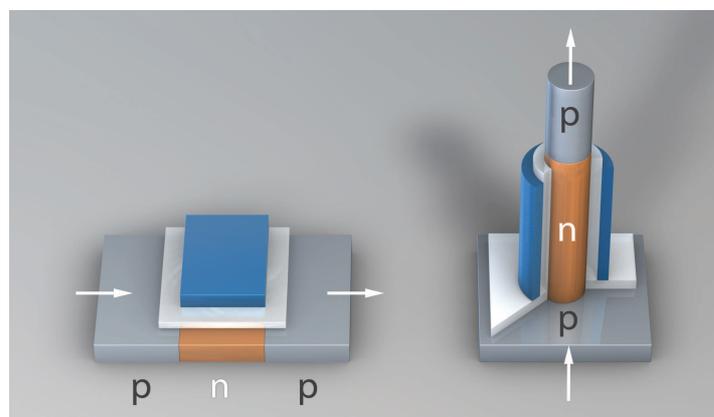


Figure 2.1: Physical difference between a vertical and a planar transistor, extracted from [Ou et al., 2010].

2.2.6.2 Two-terminal select devices (resistance-based memories)

In order to achieve the highest planar array density of $4F^2$ without considerable constraints associated with vertical select FETs, passive memory arrays with two-terminal select device are currently being investigated [Kügeler et al., 2011; Wong et al., 2010]. Two-terminal devices with nonlinear behavior (e.g. diodes) can be integrated with resistive storage nodes in a cross-bar array. General requirements for such two-terminal switches are sufficient ON currents at proper bias to support read and write operations and sufficient ON/OFF ratio to enable selection. The minimum ON current required for fast read operation is $\sim 1\mu\text{A}$ (Table 2.7). The required ON/OFF ratio depends on the size of the memory block, $m \times m$, like using a standard scheme of array biasing the required ON/OFF ratio should be in the range of $10^7 - 10^8$ for $m = 10^3 - 10^4$, in order to minimize the **sneak currents**[GH et al., 2010]. These specifications are quite challenging, and the experimental scaled select devices have yet to meet them. Thus, select devices are becoming a critical part of emerging memory and there is a need for detailed analysis on the performance requirements. Currently, two approaches to integrate the select device with storage node are being pursued. The first approach is to use an external select device in series with the storage node, integrated in a multilayer stack. The second approach is to use a storage element with inherent nonlinear properties.

Table 2.7: Benchmark Select Device Parameters, according ITRS report [ITRS, 2012a]

Parameter	Value	Driver
ON Voltage (V_T)	$\sim 1\text{V}$	Compatibility with logic; low-power operation
ON current (I_T)	$\sim 10^{-6}\text{A}$	Sensing of memory state (fast read)
Operanting Temperature	85°C	top end spec for servers
	50°C	NAND spec (the very embodiment of non-volatile memory for the current state-of-the-art)

2.2.6.3 Diode-type select devices

The simplest realizations of two-terminal memory select devices uses semiconductor diode structures, such as a pn-junction diode, Schottky diode, or heterojunction diode. Such devices are suitable for a unipolar memory cell. For bipolar memory cells, selectors with two-way switching are needed. Proposed examples include Zener diodes [Toda, 2012], BARITT diodes [Woerlee, 2005], reverse breakdown Schottky diode [Puthenthadam et al., 2011b], and complementary resistive switches [Linn et al., 2010a; Rosezin et al., 2011]. In the latter approach, the memory cell is composed of two identical non-volatile **ReRAM** switches connected back-to-back, like in the Pt/GeSe/Cu/GeSe/Pt structure [Linn et al., 2010a], or vertically integrated in a Pt/SiO₂/Cu/SiO₂/Pt structure [Rosezin et al., 2011]. In such configuration, one of the switches will always be at the high-resistance state, so the sneak current can be suppressed at low bias. The read operation, however, is destructive, and the cell state needs to be recovered by reprogramming the disturbed switch back to High Resistance State (HRS) afterwards. It should be noted that several read modes have been suggested which can be adapted to specific applications [Li et al., 2013].

One important detail is found in [Tulapurkar et al., 2005], where they demonstrate Spin-torque diode effect using magnetic tunnel junctions.

2.2.7 Resistive-Switch-type select devices

The category of **switch-based selector** refers to recent innovative device concepts that exhibit resistive switching behavior. In fact, in some of these concepts, the device structure physics of operation is similar to the structure of the storage node. In other words, a modified memory element could act as a select device. The main difference between the two is that a **nonvolatile** switch is required for the storage node, while for the select device, depending on the approaches, non-volatility may not be necessary. A brief description of several proposed select devices is given below.

2.2.7.1 MIT switch

This device is based on the Metal Insulator Transition (MIT), such as Mott transition and exhibits a low resistance above a critical electric field (threshold voltage). The select device will exhibit a high-resistance state if the voltage is below a hold voltage. To achieve reliable read, the select device needs to be volatile to ensure rapid transition from the low-resistance state to high-resistance state at low bias. If the electronic conditions that triggered Mott transition can relax within the memory device operation time scale, the Mott transition device is essentially a volatile resistive switch, which can be used as a select device. A VO₂-based device has been demonstrated as a select device for NiOx RRAM element [Lee et al., 2007a]. However, the switching mechanism is not clear and the feasibility of the Mott-transition switch, as a select device, still needs further research. It should be also noted that VO₂ undergoes a phase transition to the metal state at temperature around 68°C, thus its operation as MIT switch is restricted to temperatures below 68°C. This limits practical applications of VO₂ in memory devices, since current specifications require operational temperature of 85°C or more. Suitable Mott materials with higher transition temperatures need to be investigated. Recently, metal insulator transitions at ~ 130°C and electrically driven MIT switching were observed in thin films of SmNiO₃ [Ha et al., 2011].

2.2.7.2 Threshold switch

This type of device is based on the threshold-switching effect observed in thin-film based MIM structures. The threshold switching is caused by electronic charge injection. Operation of the threshold switch is then governed by an electronic switching process. Significant resistance reduction can occur at a threshold voltage and, when the applied voltage falls below a holding voltage, this low-resistance state quickly recovers to the original high-resistance state. One example is the threshold switching, which occurs before the structural change in phase change materials [Kau et al., 2009].

2.2.7.3 MIEC switch

This device is based on the exponential I-V characteristics observed in materials that conduct both ions and electronic charges, called Mixed Ionic and Electronic Conduction Materials (MIEC). The resistive switching mechanism of MIEC switch is similar to the

ionic memories [Chen, 2008]. With appropriate control, the resistive switching in MIEC devices is volatile and provides device selections functions [Gopalakrishnan et al., 2010].

2.2.7.4 2 Terminal Switches

As follows from Table 2.8 of [ITRS, 2012a], the required device characteristics have not yet been demonstrated, remaining a significant scientific and technical challenge. For scaled two-terminal select devices, two fundamental challenges are contact resistance [Sasago et al., 2009] and lateral depletion effects [Simpkins et al., 2008; Zhirnov et al., 2011]. A high concentration of dopants is needed to minimize both effects. Furthermore, high dopant concentrations creates an increased reverse bias currents in classical diode structures and, therefore, in reduced I_{on}/I_{off} ratio. For switch-type select devices, the main challenge is identify the right material and the switching mechanism to achieve the required drive current density, I_{on}/I_{off} ratio, and reliability.

Table 2.8: 2-Terminals Select Devices demonstrated experimentally [ITRS, 2012a]

Select Device	Material System	V_{on1}	I_{on1} (J_{on1})	V_{on2}	I_{on2}	ON/OFF	F	Ref
pn-diode	sc-Si (E)	1.8V	1.8mA ($3.107A/cm^2$) ϕ		ϕ	108	90nm	[C]
	poly-Si (E)	2V	400mA ($8.106A/cm^2$) ϕ		ϕ	105	80nm	[D]
	n-ZnO (E)	1V	45mA ($500A/cm^2$) ϕ		ϕ	105	3mm	[E]
Schottky diode	Ge NW (E)	1V	1mA ($500A/cm^2$) ϕ		ϕ	102	0.5mm	[F]
	a-Si (I)	1V	100nA ($1000A/cm^2$) ϕ		ϕ	106	100nm	[G]
	p-Si (E)	1V	10mA ($1000A/cm^2$) ϕ		ϕ	103	1mm	[H]
	Pt/TiO ₂	1V	6mA($10A/cm^2$) ϕ		ϕ	109	245mm	[I]
Reverse break-down Schottky diode	Cu/n-Si	1V	10mA	-3	10mA	103	2mm	[L]
Complementary resistive switch	Pt/GeSe/Cu/GeSe/Pt		600mA	-1	600mA	103	5mm	[M]
MIT switch [A]	Pt/VO ₂ /Pt	0.4/0.6V	($400A/cm^2$) ϕ		ϕ	103	ϕ	[N]
Threshold switch	Chalcogenide alloy (undisclosed)	ϕ	ϕ	ϕ	ϕ	106	ϕ	[O]

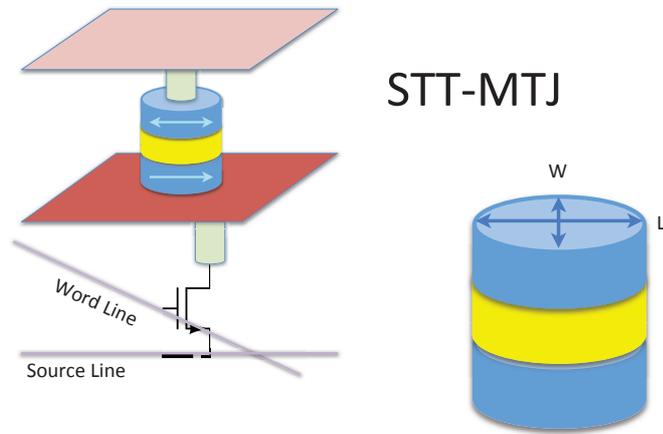
References and acronyms for Table 2.8:

- A Metal Insulator Transition (MIT)
- B Mixed Ionic and Electronic Conduction Materials (MIEC)
- C [Oh et al., 2006]
- D [Sasago et al., 2009]
- E [Huby et al., 2008]
- F [Kim et al., 2008]
- G [Kim et al., 2010b]
- H [Cho et al., 2010]
- I [Park et al., 2010]
- J [Choi et al., 2010]
- K [Ahn et al., 2009]
- L [Puthentheradam et al., 2011a]
- M [Linn et al., 2010b]
- N [Lee et al., 2007b]
- O [Kau et al., 2009]
- P [Gopalakrishnan et al., 2010]

2.3 MAGNETIC TUNNELING JUNCTION (MTJ)

In the MTJ, the information is stored as the anisotropy (magnetic orientation) in one of the two ferromagnetic layers separated by a thin tunnel barrier, which is called the energy barrier. The switchable layer manipulated to contain the information is denominated Free-Layer (FL). The last layer is called Reference-Layer (RL) and is designed in such a way that it is nearly impossible to reverse its magnetization. There are two types of MTJ in relation to the direction of magnetic free axes of the two magnetic electrodes. Figure 2.2a shows a schematic diagram of a typical 1T – MTJ cell, as a memory cell architecture in MRAM [Huai, 2008]. One is MTJ with in-plane axis (i-MTJ, see Figure 2.2b), and the other is that with perpendicular axis (p-MTJ, see Figure 2.2c). One important detail about the MTJ, is that the magnetic state changes from parallel (P) to antiparallel (AP). Since when using a MTJ based memory in fact we use a sense amplifier to identify the magnetic state, so the resistance currently set on the MTJ, we have the resistance in parallel state R_P and in antiparallel state R_{AP} , as depicted in Figure 2.2c, providing us the ability to identify the logic state [Huai, 2008]. .

Magnetic anisotropy is the directional dependence of a material's magnetic properties. In the absence of an applied magnetic field, a magnetically isotropic material has no preferential direction for its magnetic moment, while a magnetically anisotropic material will align its moment with one of the easy axes. An easy axis is an energetically favorable direction of spontaneous magnetization

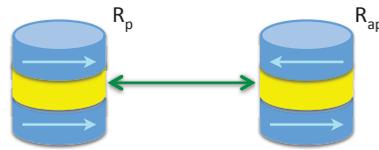


$$\text{Aspect Ratio} = K = \frac{L}{W}$$

(a) MTJ layout, 1T-MTJ, for STT

Plane (Planar) Anisotropy

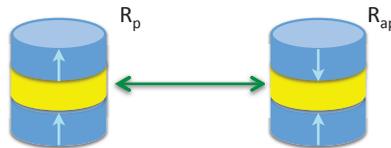
MTJ(i-MTJ) $K > 1$



(b) i-MTJ

Perpendicular Anisotropy

MTJ(p-MTJ) $K = 1$



(c) p-MTJ

Figure 2.2: (a) MTJ structure and interconnection with the CMOS circuit and the relation of the MTJ according with the technology of anisotropy, (b) Planar MTJ, (c) Perpendicular MTJ.

2.3.1 Tunnel Magnetoresistance (TMR)

Tunneling Magnetoresistance (TMR) is a magnetoresistive effect that occurs in a MTJ, if the insulating layer is thin enough (typically a few nanometers), electrons can tunnel from one ferromagnet into the other. Since this process is forbidden in classical physics, the tunnel magnetoresistance is a strictly quantum mechanical phenomenon.

MTJ are manufactured in thin film technology. On industrial scale, the film deposition has been done by magnetron sputter deposition [Sun et al., 2003]. On laboratory scale, molecular beam epitaxy, pulsed laser deposition and electron beam physical vapor deposition are also employed. The junctions are etched using standard photolithography.

The effect of magnetization field in the MTJ was originally discovered in 1975 by M. Jullière (University of Rennes, France) in Fe/Ge/Co, Fe/Ge/Pb junctions at $T \leq 4.2\text{K}$, while Albert Fert and Peter Grünberg discovered the Giant Magnetoresistance (GMR) effect in thin-film structures in 1988 [Fert and Grünberg, 1997]. The relative change of resistance was around 14%, and did not attract much attention [Julliere, 1975]. In 1991 T. Miyazaki (University Tohoku, Japan) found an effect of 2.7% at room temperature [Miyazaki et al., 1991]. Later, in 1995, Miyazaki found 18% in junctions of iron separated by an amorphous aluminum oxide insulator [Miyazaki and Tezuka, 1995] and J. Moodera found 11.8% in junctions with electrodes of CoFe and Co [Moodera et al., 1995]. The highest effects observed with aluminum oxide insulators are around 70% at room temperature. Since the year 2000, tunnel barriers of crystalline magnesium oxide (MgO) have been under development. TMR effects up to 600% at room temperature and currently TMR with more than 1100% were observed, also at room temperature, in junctions of CoFeB/MgO/CoFeB [Amara et al., 2012; Arakawa et al., 2011; Fukushima et al., 2012; Ikeda et al., 2010; Ishibashi et al., 2011; Jiang et al., 2009; Min et al., 2010; Nakayama et al., 2008; Tanaka and Moodera, 1996; Zhu et al., 2012].

2.3.2 Field Induced Magnetic Switching (FIMS) MRAMs

The Toggle or Field Induced Magnetization Switching (FIMS) MRAM technology eliminates the single-line disturb phenomenon that exists in other approaches to MRAM switching. Through the use of a proprietary layer structure, bit orientation and current pulse sequence, the MRAM bit state can be programmed via a **Savtchenko switching toggle** mode [Durlam et al., 2004; Savtchenko et al., 2001]. Toggle switching employs the same pulse sequence to either write from the '0' state to the '1' state or write from the '1' state to the '0' state. Each time the sequence is executed, the device changes from its current magnetic state to the opposite state. The switching is different from the previous switching technologies, where the free layer magnetic moment simply followed the applied field [Everspin Technologies Inc., 2010].

Savtchenko switching relies on the unique behavior of a Synthetic Antiferromagnet (SAF) free layer, that is formed from two ferromagnetic layers separated by a non-magnetic coupling spacer layer. This is shown schematically in Figure 2.4, the moment-balanced SAF free layer responds to an applied magnetic field differently than the single ferromagnetic layer of conventional MRAM. Rather than following an applied magnetic field, the two anti-parallel layer magnetizations will rotate to be approximately orthogonally to the applied field. A current pulse sequence is used to generate a rotating magnetic field that moves the free-layer moments through the 180° switch, from one

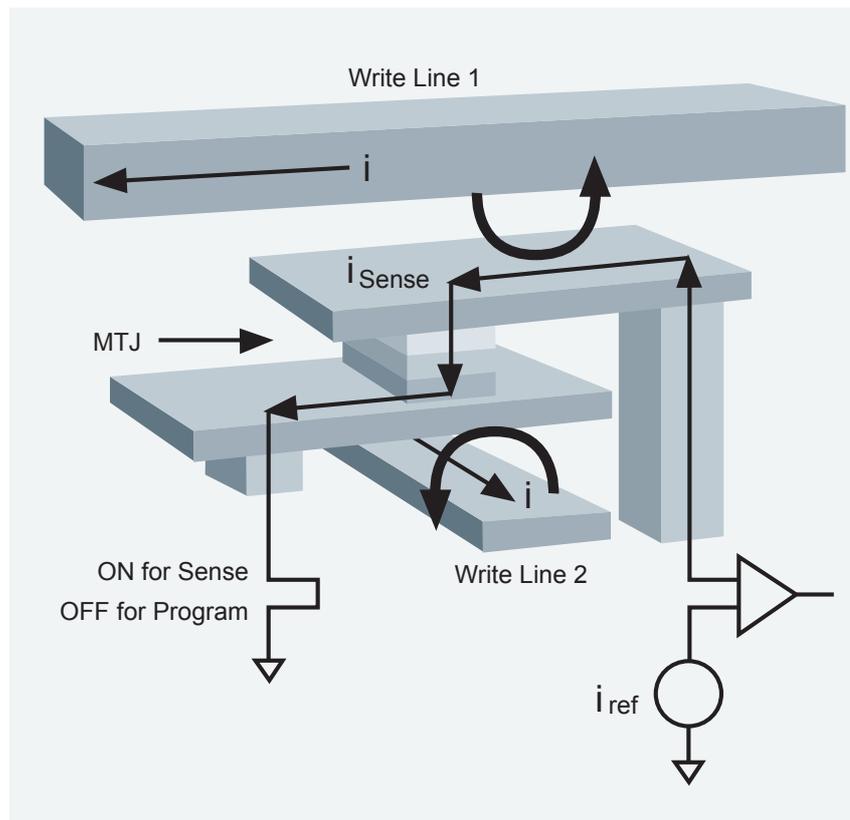


Figure 2.3: Schematic of a 1-transistor, 1-MTJ memory cell showing the write lines above and below the bit and the read current path [Everspin Technologies Inc., 2010].

state to the other, as shown in Figure 2.5. To exploit the unique field response of this free layer, a two-phase programming pulse sequence, shown in Figure 2.3, is applied to rotate the magnetic moments of the SAF by 180 degrees. Due to its inherent symmetry, this sequence toggles the bit to the opposite state, regardless of existing state. A pre-read is therefore used to determine if a write is required. Given the way a SAF responds to applied fields, a single line alone cannot switch the bit, providing enhanced selectivity over the previous approaches to MRAM switching.

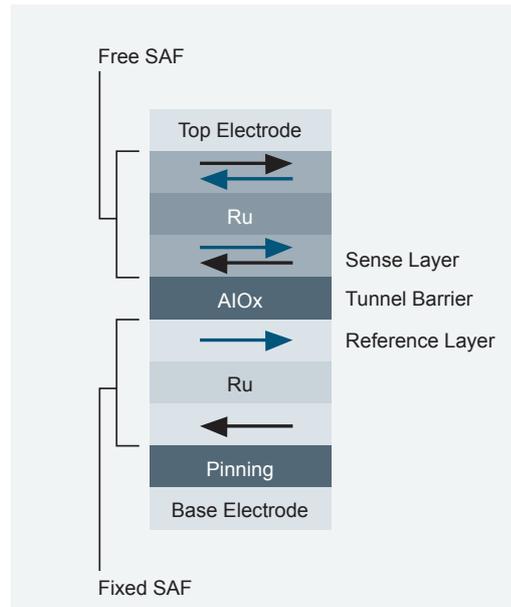


Figure 2.4: The magnetic tunnel junction (MTJ) material stack used for Toggle MRAM. The Free SAF magnetic moments switch between two states when the proper magnetic field sequence is applied. Electrons tunnel across the alumina (AlOx) tunnel barrier, resulting in a magnetoresistance that is sensitive to the magnetic moment direction of the sense layer [Everspin Technologies Inc., 2010].

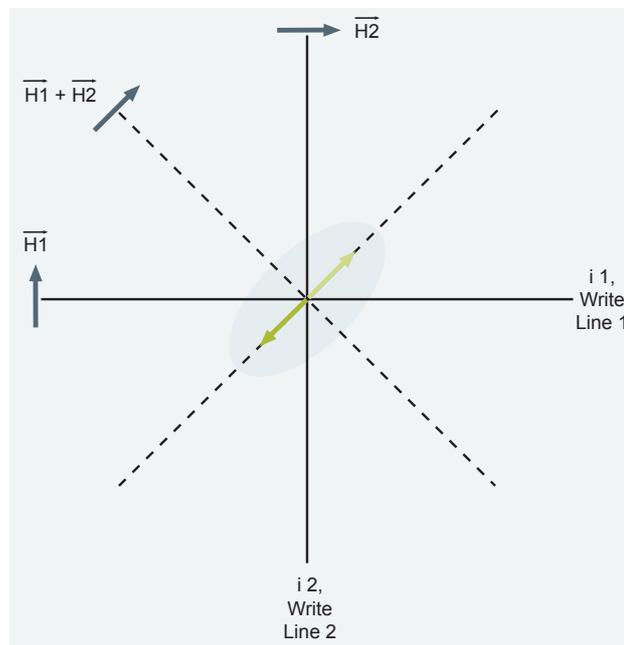


Figure 2.5: Schematic of a toggle MRAM bit with the field sequence used to switch the free layer from one state to the other. The fields, H_1 , H_1+H_2 and H_2 are produced by passing currents, i_1 and i_2 , through the write lines [Everspin Technologies Inc., 2010].

2.3.3 Thermally Assisted Switching (TAS) MRAMs

The Thermally Assisted Switching MRAM (TAS-MRAM), already employed in previous works like [Guillemenet et al., 2010], [Guillemenet et al., 2009] and [Guillemenet et al., 2008b], has shown improvements in terms of required writing current and power consumption during write operation. TAS-MRAM is an improvement, or a parallel development, of the Toggle-MRAM and is classed as a second generation MRAM, together with Toggle [BRUCHON, 2007a].

Thermally Assisted Switching approach combines a local heating of the junction and a single low amplitude magnetic field. This writing method also requires several steps that are depicted in Figure 2.6. When the junction is heated above the blocking temperature, $\sim 150^\circ\text{C}$ by a current ($I_{\text{heat}} \sim 340\mu\text{A}$) flowing through the junction, the magnetization of the ferromagnetic layer is freed and can be reversed under the application of a single low amplitude magnetic field. This entire operation is performed in $\sim 35\text{ns}$ [Guillemenet et al., 2010].

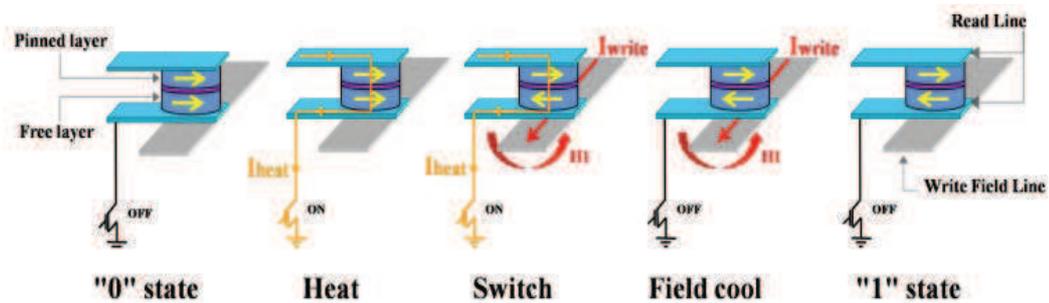


Figure 2.6: MTJ process to switch the electromagnetic thin-layer field and store the logic value into the MTJ passing from the electrical layer.

2.3.4 Spin Transfer Torque (STT) RAMs

The Spin Transfer Torque (STT) was not widely known until 1996. It is common sense that the beginning of Spin Transfer Torque research, applied to semiconductor devices, start with the patent [International Business Machines Corporation, 1996] it also referenced [Edw; Hornreich and M]. Given the potential applications of the technology, academy and industry have been proactively researching it, and lots of progresses have been made in recent years: from the first experimental verification of spin-torque transfer in GMR film, to magnetic tunneling junction devices. A large portion of this effort was made towards the development of practical MRAM chips based on STT. A key milestone in STT research has been reached, in early 2004, by first demonstration of STT switching in Al_2O_3 based MTJs by Huai et al. [Huai et al., 2005a; Rippard et al., 2005]. In late 2004, the revolutionary high TMR above 200% has been reported in MTJs with crystalline MgO barrier [Parkin et al., 2004; Yuasa et al., 2004], following the early theoretical predictions [Butler et al., 2001; Mathon and Umerski, 2001].

The following sections covers the basic in physics of spin-torque transfer, the switching properties of the magnetic tunnel junction and, also, the technological aspects of STT-MRAM technology.

2.3.5 STT based Magnetic Tunnel Junction (STT-MTJ)

MTJ is the basic cell of MRAM [Chappert et al., 2007b]. It is a nano-pillar composed of two Ferromagnetic (FeM) layers and one oxide thin barrier (see Figure 2.2). As the magnetization direction of the two FeM layers is either in parallel or anti-parallel, a MTJ shows two different resistance values R_P and R_{AP} . For practical applications, the magnetization direction of one FM layer is pinned as reference and that of the other ferromagnetic layer is free to be switched to store binary state [Hoya et al., 2006; Wolf et al., 2001].

STT-MRAM technology has significant advantages over magnetic-field-switched (FIMS, Toggle) MRAM or TAS-MRAM. The main challenge for implementing STT writing mode, in high-density and high-speed memory, is the substantial reduction of the intrinsic current density J_{c0} (defined in Equation 2.1), required to switch the magnetization of the FL, while maintaining high thermal stability needed for long-term data retention. Minimal switching (write) current is necessary, mainly, for reducing the size of the access transistor in series, with MTJ in one transistor and one MTJ design (1T1J), in order to achieve the highest possible memory density, since the channel width (in unit of F) of the transistor is proportional to the write current for a given transistor current drivability ($\mu\text{A}/\mu\text{m}$). Minimal channel width of 1F, or at least the width of MTJ element, is expected for achieving ultimate smallest STT-MRAM cell size. Also, smaller voltage across MTJ decreases the probability of tunneling barrier degradation and breakdown, ensuring endurance for the junction. This is particularly important for STT-MRAM, for the reason that both, sense and write currents are driven through MTJ cells [Huai, 2008].

2.3.6 Planar and Perpendicular STT

As described in [Theodonis et al., 2006], when a spin-polarized current passes through a magnetic multilayer structure, the Magnetic Tunnel Junction (MTJ), it can transfer spin angular momentum from one ferromagnetic electrode to another, exerting this way a torque on the magnetic moments of the electrodes. Using sufficiently high current densities, this spin transfer can stimulate spin-wave excitations and even reverse the magnetization of one of the alloy layers, denominating the process as Current induced magnetic switching (CIMS) a.k.a. Spin Transfer Torque (STT).

This angular spin transfer is important to understand that it will be tied with the MTJ manufacturing process. According with the mathematical model, in the in-Plane or Parallel Magnetic Tunnel Junction (iMTJ), the anisotropy vectors are contained inside the thin-film physical structure. The field orientation is given by the equation Equation 2.1, that is explained in details in [Huai, 2008], that defines the MTJ magnetization in the film plane be expressed as Equation 2.1.

$$J_{c0} = \frac{2e\alpha M_s t_F (H + H_k + 2\pi M_s)}{\hbar\eta} \quad (2.1)$$

In Equation 2.1 the H is the field applied along the easy axis, M_s and t_F are the magnetization and thickness of the free layer respectively, α is the damping constant, and H_k is the effective anisotropy field including magnetocrystalline anisotropy and shape anisotropy. The spin transfer efficiency η , is a function of the current polarity, polarization, and the relative angle between the FL and Pinned Layer (PL).

According with [Huai, 2008], observing the Equation 2.1, the intrinsic current density is primarily governed by the thin film easy-plane anisotropy $4\pi M_s$ ($> 10\text{kOe}$). The anisotropy field H_k , mainly dominated by shape anisotropy, is about several hundred Oe. Therefore, most efficient means of reducing J_{c0} would be using a FL with perpendicular anisotropy with $H_{k\perp} > 4\pi M_s$ as stated in [Huai, 2008]. In this case, the magnetization is out of the film plane, which would eliminate the $2\pi M_s$ term in Equation 2.1, where the H_k becomes the effective perpendicular anisotropy $H_{k\perp} = H_{k\perp} - 4\pi M_s$ [Huai, 2008]. Additional sources into this subject are [Chappert et al., 2007a; Chshiev et al., 2008; Heinonen et al., 2010; Huai et al., 2005b; Jung et al., 2010; Mejdoubi et al., 2013; Oh et al., 2009; Stiles and Zangwill, 2002; Wilczyński et al., 2008; Xiao et al., 2008].

So, it needs less energy to switch the anisotropy in Perpendicular Magnetic Tunnel Junction (pMTJ) than in a iMTJ, according with Equation 2.1 and [Huai, 2008; Theodonis et al., 2006], this is true. Also, it is easy to control the angular momentum of a MTJ in perpendicular orientation. Additionally, regarding the architectural aspects, since the MTJ can be built in a circular shape instead of a oval shape due to the angular field orientation because of the alloy etching to create the MTJ nanopillar, it can scale below 45nm, use less energy for field switching, it is also possible to create denser memory banks than using iMTJ.

According with [Chun et al., 2013], they explored the scalability of in-plane (iMTJ) and perpendicular (pMTJ) MTJ based STT – MRAMs from 65nm to 8nm, focusing on the read and write performances of a STT-MRAM based cache, rather than the obvious advantages such as the denser bit-cell and zero static power. An accurate MTJ macromodel capturing key MTJ properties was adopted for efficient Monte Carlo simulations. For the simulation of access devices and peripheral circuitries, ITRS projected transistor parameters were utilized and calibrated, using the MASTAR tool that has been widely employed in industry. The 6T SRAM and STT-MRAM arrays were implemented to mimic industrial memory designs. The thermal stability factor ensuring a 10 year retention time was obtained by adjusting the FL thickness, as well as assuming improvement in the crystalline anisotropy. The paper demonstrated that in-plane STT-MRAM can outperform SRAM from 15nm node, while its perpendicular counterpart requires further innovations in MTJ material in order to overcome the poor write performance scaling from 22nm node onwards.

In [Kim et al., 2011b] they report the first experimental demonstration of sub-20nm MTJ cells investigating the downscaling feasibility of Spin Transfer Torque MRAM (STT-MRAM). They demonstrated a STT switching of 17nm node pMTJ cell, allegedly the smallest feature size reported until now, using perpendicular materials possessing with high interface anisotropy of $2.5\text{erg}/\text{cm}^2$ and an integration processes, to achieve reproducible switching, with critical current (I_c) of $44\mu\text{A}$, TMR ratio of 70% and thermal stability factor ($E/k_B T$) of 34.

It is widely recognized that Perpendicular Magnetic Anisotropy (PMA) should be implemented into MTJs at such a small node to obtain high enough thermal stability. Although many PMA materials have been studied, for instance $L1_0$ ordered alloys (FePt and CoPt), Co based multilayers (Co/Pd and Co/Pt), rare-earth and transition metal alloys [Kim et al., 2011b]. Adopting the PMA materials as a free layer in the MTJs is a highly complex process [Kim et al., 2011b]. They demonstrate the feasibility of sub-20nm MRAM device adopting the free layer of interface driven PMA (i-PMA) [Kim et al., 2011b] with enhanced anisotropy energy.

These two studies are among the most elucidating and innovative regarding the MTJ technologies. It is strongly advised to consult the references [Chun et al., 2013; Huai et al., 2011] to have an overview and comparative studies regarding the in-plane and perpendicular MTJ structures. Specifically for perpendicular MTJ [Chun et al., 2013; Driskill-Smith et al., 2011; Huai et al., 2011; Kim et al., 2011b; Krivorotov et al., 2012; You, 2012] this are the latest available publications about the subject.

The research group lead by professor Hideo Ohno of Tohoku University, have published works on MRAM applied for building logic circuits as well the recent advances in Perpendicular STT [Brataas et al., 2012; IKEDA et al., 2012; Matsunaga et al., 2009, 2012].

The proposed MTJ of Tohoku university is depicted in Figure 2.7, with a cross-section photo in Figure 2.7a. At Figure 2.7b we have depicted the material schematic organization of the MTJ, with its respective thicknesses. This MTJ architecture is the pMTJ, currently adopted and is the leading edge on the field.

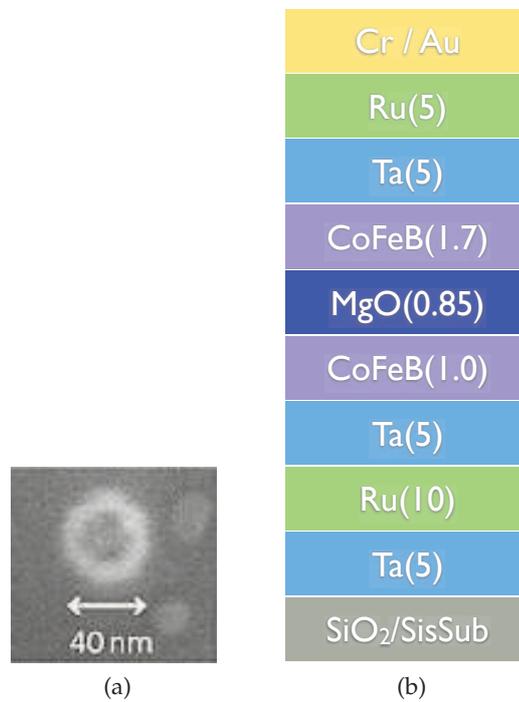


Figure 2.7: High performance perpendicular MTJ with ϕ 40nm

2.4 CURRENT STATE COMPARED TO THE MEMORY STATE OF THE ART

SRAM has been the dominant technology for building cache memories in processors and programmable logic in Field Programmable Gate Arrays (FPGAs). It is a fast, yet nowadays very power-hungry, variety of memory. DRAM comes next in the hierarchy, serving as a larger, but not so fast, volatile memory. Finally, in embedded systems and FPGAs, secondary storage is usually made with solid-state devices based on Flash memory.

Many obstacles threaten continued scaling of these three technologies: from increasing leakage power to lithography issues. It has been estimated that, by 2018, SRAM, DRAM and Flash technologies will likely be replaced if Moore's law is to stand [itr, 2011].

This scenario motivated the appearance of a number of NVMs technologies in the past years. STT-MRAM, Phase-Change Memory (PCM) and Resistive Random Access Memory (RRAM), among others, are considered by ITRS as the most promising candidates to take over the mainstream market. In Table 2.9, a quick comparison of those technologies is provided.

Since 2006, a certain number of proposals on the NVM field have emerged, from architecture-level enhancements to actual circuit implementations [Augustine et al., 2012; Chen et al., 2012; Chen and Li, 2011; Chun et al., 2013; Dason et al., 2011; Dingler et al., 2012; Dorrance et al., 2013; Fong et al., 2013, 2011; Gupta et al., 2012; Huda and Sheikholeslami, 2013; Iba et al., 2011; Jadidi et al., 2011; Jiang et al., 2012; Kim et al., 2011b; Kitagawa and Fujita, 2013; Kitagawa et al., 2012; Lakys et al., 2012; Lee et al., 2012b; Li et al., 2011a; Mandhdapu and Samson Immanuel, 2012; Mao et al., 2013; Mishra et al., 2011; Morris et al., 2012; Panagopoulos et al., 2012; Park et al., 2012a,b; Sun et al., 2012a,b, 2011a, 2012d; Torres and Zhao, 2011; Tsunoda et al., 2012; Venkatesan et al., 2012; Wang et al., 2013; Wang and Amiri, 2012; Wang et al., 2012; Wen et al., 2012; Yoda et al., 2012a,c; Yoshida et al., 2012; Yu et al., 2013].

Table 2.9: Comparison of NVM technologies [itr, 2011; Kim et al., 2011a; Yoda et al., 2012c]

Technology	Min. cell size (F ²)	Endurance (cycles)	Read time(ns)	Write time (ns)
SRAM	150	ϕ	2	2
STT-MRAM	20	10^{16}	5	15
Perpend. STT	ϕ	ϕ	3	3
TAS-MRAM *	ϕ	10^{12}	30	30
NAND-Flash	4	10^4	10^4	10^6
NOR-Flash	10	10^5	15	10^3
FeRAM	22	10^{12}	40	65
RRAM	30	10^5	100	100
PCM	4	10^{12}	12	100

* Data provided by Crocus Technology

Table 2.10: Timeline of the MTJ current power, according with technology.

	FIMS (130nm) Everspin	TAS (120nm) Crocus Spintec	STT (45nm) / SPINTEC ¹	pSTT (28nm) SPINTEC ¹	pSTT (20nm) Toshiba
write (μA)	16000	14320	29	17	50
Read (μA)	10	10	20	7	ϕ
Period (ns)	40	35	17	6	3

In Table 2.10 is denoted the evolution of the currents used for read and write into each specific MTJ technology. Also, details regarding the Toshiba's pMTJ can be found in [Inc., 2011; Kitagawa and Fujita, 2013; Yoda et al., 2012b].

2.5 COMPUTER ARCHITECTURE AND MEMORY HIERARCHY

The founders of Computer Science predicted that programmers would desire unlimited amounts of fast memory. Since memory is not infinity, and its performance is inversely proportional to its density, the trade-off created to supply this eagerness for memory is a memory hierarchy, which takes advantage of locality and cost-performance of memory technologies.

The principle of locality denotes that most programs do not access all code or data uniformly. Locality occurs in time (temporal locality) and in space (spatial locality)[Hennessy and Patterson, 2007b]. This principle evolved to hierarchies based on memories of different speeds and sizes. Figure 2.9 shows a multilevel memory hierarchy, including typical sizes and speeds of access [Hennessy and Patterson, 2007b; Patterson, 1980; Patterson and Hennessy, 2012].

Since fast memory is expensive, a memory hierarchy is organized into several levels. The levels nearest to the microprocessor, beginning by the registers, are extremely expensive and work at the core frequency. Further we advance in the memory hierarchy the density increases, to the expense of performance [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]. Note that each level maps addresses from a slower, larger memory to a smaller but faster memory higher in the hierarchy. As part of address mapping, the memory hierarchy is given the task of address checking and translation (virtual to physical and vice-versa). Hence, protection schemes for addresses segments are also part of the memory hierarchy.

The importance of the memory hierarchy has increased with advances in performance of processors [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]. So, is feasible to assert that: still, there is room in computer architecture to close the processor-memory gap.

¹ MTJ Model, established during discussions with Guillaume Prenat and Gregory Di Pedina of CEA-Spintec

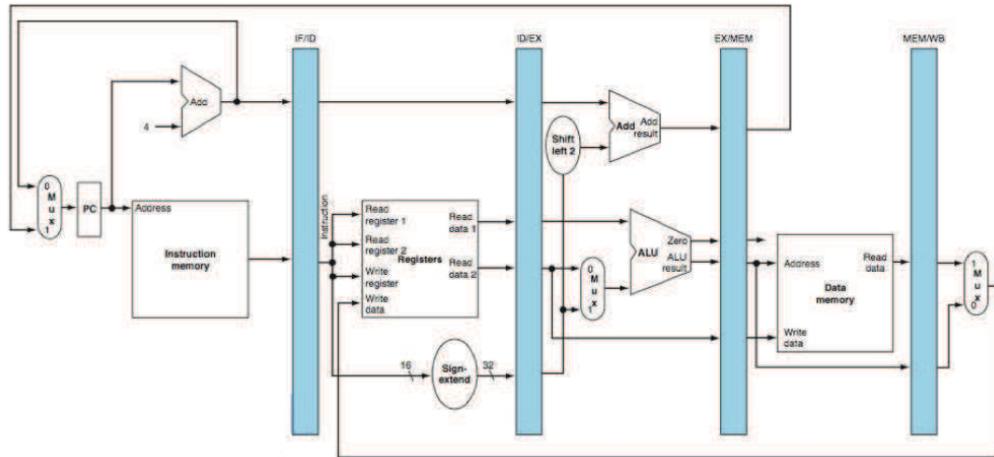


Figure 2.8: Microprocessor Datapath [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012].

2.5.1 Working Principles of the Memory Hierarchy

When a data/instruction word is not found in the cache, the word must be fetched from the memory and placed in the cache before continuing [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]. Multiple words, called a block (or line), are moved for efficiency reasons: instead one word each fetch, you move a block (Spatial Locality). Each cache block includes a tag, to see which memory address it corresponds to.

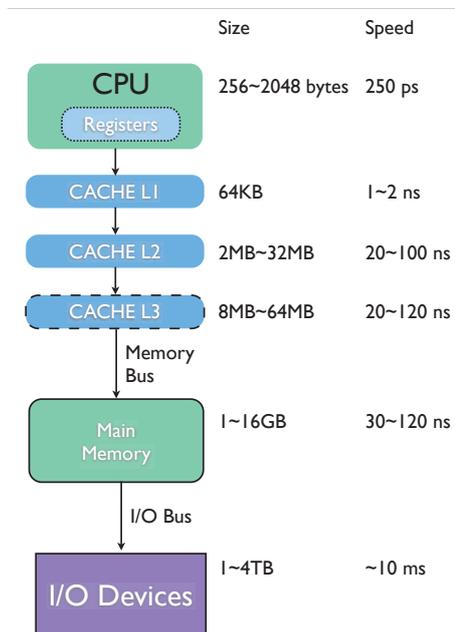


Figure 2.9: The levels in a typical memory hierarchy in embedded and server computers. As we move further away from the processor, the memory in the lower level increases in latency and density. Note that the time units change by factors of 10, from picoseconds to milliseconds, and that the size units change by factors of 1000, from bytes to terabytes.

A cornerstone design decision, is the blocks (or lines) placement in a cache. The most popular scheme is set associative, where a set is a group of blocks in the cache. A block is first mapped onto a set, and then the block can be placed anywhere within that set. Finding a block consists in mapping the block address to the set, and then searching the set, usually in parallel, to find the block. The set is chosen by the address of the data [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]:

$$(\text{Block address}) \text{MOD} (\text{Number of sets in cache}) \quad (2.2)$$

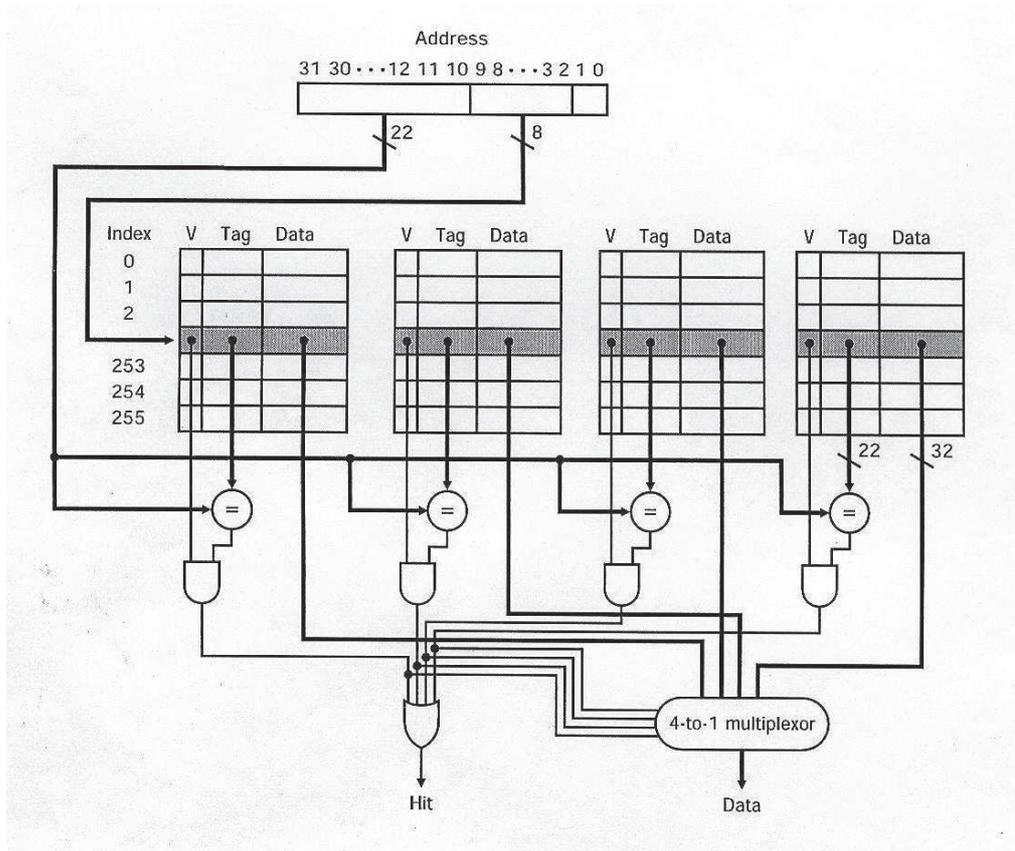


Figure 2.10: An example of a possible Set-Associative Cache Organization [Patterson and Hennessy, 2012].

If there are n blocks in a set, the cache placement is called n -way set associative. A direct-mapped cache has just one block per set (so a block is always placed in the same location), and a fully associative cache has just one set (so a block can be placed anywhere). Read access to the CACHE, that is only read, is easy, since the copy in the cache and memory will be identical. Writing access is more difficult: the circuit must, kept the CACHE copies consistent in all levels of the hierarchy. There are two main strategies for that: the **write-through** and **write-back**. A **write-through** cache updates the item in the cache and writes through to update main memory. A **write-back** cache only updates the copy in the cache. When the block is about to be replaced, it is copied back to memory. Both write strategies can use a write buffer to allow the cache to proceed as soon as the data is placed in the buffer, rather than wait the full latency to

write the data into memory [Hennessy and Patterson, 2006; Patterson and Hennessy, 2012].

A measurement unit for performance evaluation of different cache organizations is MISS rate. Miss rate is simply the fraction of cache accesses that result in a miss, that is, the number of accesses that MISS divided by the number of accesses. To gain insights into the causes of high MISS rates, which can inspire better cache designs, the three C_s model sorts all misses into three categories [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]:

- Compulsory – The very first access to a block cannot be in the cache, so the block must be brought into the cache. Compulsory misses are those that occur even if you had an infinite cache.
- Capacity – If the cache cannot contain all the blocks needed during execution of a program, capacity misses (in addition to compulsory misses) will occur because of blocks being discarded and later retrieved.
- Conflict – If the block placement strategy is not fully associative, conflict misses (in addition to compulsory and capacity misses) will occur because a block may be discarded and later retrieved if conflicting blocks map to its set.

Miss rate can be a misleading measure for several reasons. Hence, some designers prefer measuring misses per instruction rather than misses per memory reference (miss rate) [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]. These two are related:

$$\frac{\text{Misses}}{\text{Instructions}} = \frac{\text{Miss rate} * \text{Memory Accesses}}{\text{Instruction Count}} = \text{Miss rate} * \frac{\text{Memory Accesses}}{\text{Instruction}} \quad (2.3)$$

. For speculative processors, we only count instructions that are committed [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]. The problem with both measures is that they do not factor in the cost of a miss. A better measure is the average memory access time:

$$\text{Average Memory Access Time} = \text{Hit Time} + \text{Miss rate} * \text{Miss penalty} \quad (2.4)$$

where Hit time is the time to HIT in the cache and Miss penalty is the time to replace the block from memory (that is, the cost of a miss). Average memory access time is still an indirect measure of performance; although it is a better measure than MISS rate, it is not a substitute for execution time [Hennessy and Patterson, 2006; Patterson and Hennessy, 2012].

2.6 CACHE MEMORIES

CACHE memories are probably the most important advent for microprocessors, because it marks the appearance of the memory hierarchy concept [Hennessy and Patterson, 2007b; Patterson and Hennessy, 2012]. It is in place for the last 30 years and have granted the fast evolution of microprocessors and the advent of System-on-Chip (SoC) for embedded systems.

There are six basic cache optimizations approaches, which we quickly review here, as described in [Hennessy and patterson, 2006], that will be listed in the following subsections.

2.6.1 Larger block size to reduce MISS rate

The simplest way to reduce the MISS rate is to take advantage of spatial locality and increase the block size. Note that larger blocks also reduce compulsory misses, but they also increase the MISS penalty.

2.6.2 Bigger caches to reduce MISS rate

The obvious way to reduce capacity misses is to increase cache capacity. Drawbacks include potentially longer HIT time of the larger cache memory and higher cost and power (Indeed bigger the caches, higher the latencies as will be seen in Section 3.5.3).

2.6.3 Higher associativity to reduce MISS rate

Obviously, increasing associativity reduces conflict misses. Greater associativity can come at the cost of increased HIT time too.

2.6.4 Multilevel caches to reduce MISS penalty

A difficult decision is whether to make the cache HIT time fast, to keep pace with the increasing the frequency and bandwidth of processors, or to make the cache large, to overcome the widening gap between the processor and main memory. Adding another level of cache, between the original cache and memory, simplifies the decision (observe Figure 2.9 and Figure 2.11). The first-level cache can be small enough to match a fast clock cycle time, yet the second-level cache can be large enough to gather many accesses that would go to the main memory. The focus on misses in second-level caches leads to larger blocks, bigger capacity, and higher associativity. Given that L1 and L2 refer to first and second level caches, we can define, like in [Hennessy and patterson, 2006; Patterson and Hennessy, 2012], the average memory access time as:

$$\text{Hit time L1} + \text{Miss rate L1} * (\text{Hit time L2} + \text{Miss rate L2} * \text{Miss penalty L2}) \quad (2.5)$$

2.6.5 Giving priority to read misses over writes to reduce MISS penalty

A write buffer is a neat place to implement this choice of optimization. Write buffers create hazards, since they hold the updated value of a location needed on a read miss—that is, a read-after-write hazard through memory [Hennessy and patterson, 2006; Patterson and Hennessy, 2012]. A solution is to check the contents of the write buffer on a read miss. If there are no conflicts, and if the memory system is available, sending the read before the writes reduces the MISS penalty. Most processors give reads priority over writes.

2.6.6 Avoiding address translation during indexing of the cache to reduce HIT time

Memory Management Unit (MMU), within CACHES must cope with the translation of a virtual address from the processor to a physical address to access memory. Figure 2.11 shows a general flow between caches, Translation Lookaside Buffers (TLBs), and virtual memory. A common optimization is to use the page offset, the part that is identical in both virtual and physical addresses, to index the cache. The virtual address is translated while the cache is read using that index, so the physical addressing can be employed for TAG matching. The drawback of this virtually indexed, physically tagged optimization is that the size of the page limits the size of the cache. Higher associativity can keep the cache index in the physical part of the address.

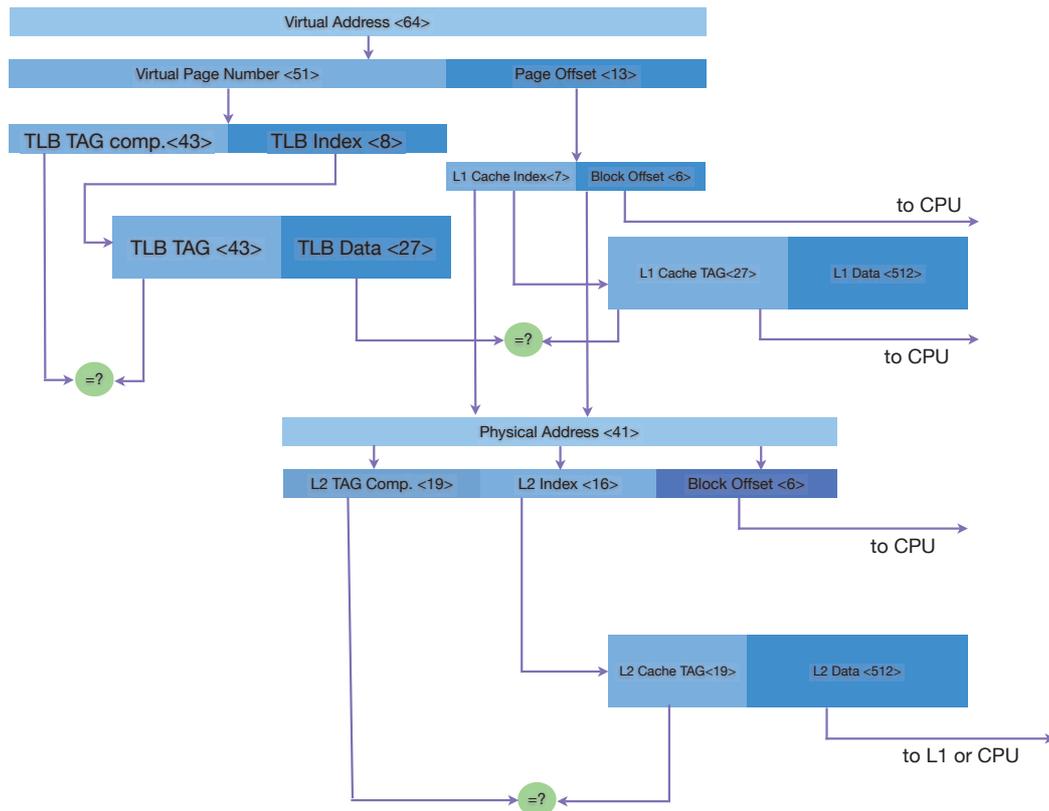


Figure 2.11: A general overview of a hypothetical memory hierarchy going from virtual address to L2 cache access. [Hennessy and patterson, 2006]

2.7 MRAM APPLIED INTO MEMORY HIERARCHY

To close this chapter, the idea is take into account the concepts of memory hierarchy and search how the research in the field of computer architecture is exploring this elements of memory hierarchy. This way, better positioning our research goals, targets and exploratory experiments.

In [Kim et al., 2011b] article, they report the first experimental demonstration of sub-20nm MTJ cells for investigating the downscaling feasibility of STT-MRAM. They demonstrate the STT switching of 17nm node P-MTJ cells, the smallest feature size re-

Extended scalability of perpendicular STT-MRAM towards sub-20nm MTJ node.

ported, utilizing perpendicular materials possessing high interface anisotropy of $2.5\text{erg}/\text{cm}^2$ and improved integration processes to achieve reproducible switching, with critical current (I_c) of $44\mu\text{A}$, TMR ratio of 70% and thermal stability factor (E/kBT) of 34.

In the paper [Mishra et al., 2011], they study the integration of STT-MRAM in a 3D multicore environment and propose solutions at the on-chip network level, to mitigate the write overhead problem in the CACHE architecture with STT-MRAM technology. Their scheme is based on the observation that, instead of staggering requests to a write-busy STT-MRAM bank, the network should schedule requests to other idle CACHE banks for effectively hiding the latency. Thus, they prioritize CACHE accesses to the idle banks by delaying accesses to the STT-MRAM CACHE banks, that are currently serving long latency write requests. Through a detailed characterization of the CACHE access patterns of 42 applications, they propose a mechanism to facilitate delayed writes to CACHE banks by estimating the busy time of each CACHE bank through logical partitioning of the CACHE layer and prioritizing packets in a router requesting accesses to idle banks. Evaluations on a 3D architecture, consisting of 64 cores and 64 STT-MRAM CACHE banks, according with the authors the proposed approach achieves 14% average IPC improvement for multi-threaded benchmarks, 19% instruction throughput improvement for multi-programmed workloads, and 6% latency reduction compared to a proposed write buffering mechanism.

In recent years, NVM technologies have emerged as candidates for future universal memory. NVMs generally have advantages, such as low leakage power, high density, and fast read speed. At the same time, NVMs also have disadvantages. For example, NVMs often have asymmetric read and write speed and energy cost, which poses new challenges when adopting NVMs. In [Xue et al., 2011a] they present four contributions, comparing three emerging NVM technologies, their characteristics, potential challenges, and new opportunities that they may bring forward in memory systems.

In [Sun et al., 2011b], they propose both, L1 and lower level CACHE designs, that use STT-MRAM. In particular, their designs use STT-MRAM cells with multiple data retention time and write performances, made possible by different MTJ designs. For the fast STT-MRAM bits with reduced data retention time, a counter controlled dynamic refresh scheme is proposed to maintain the data validity. According with them, their dynamic scheme saves more than 80% refresh energy compared to the simple refresh scheme proposed in previous works. A L1 CACHE built with ultra low retention STT-MRAM coupled with the proposed dynamic refresh scheme can achieve 9.2% in performance improvement, and saves up to 30% of the total energy, when compared to one that uses traditional SRAM. For lower level CACHES, with relative large CACHE capacity, they propose a data migration scheme, that moves data between portions of the CACHE with different retention characteristics to maximize the performance and power. Their experiments show that on the average, the proposed multi retention level STT-MRAM CACHE reduces 30 ~ 70% of the total energy compared to previous works, while improving IPC performance for both 2-level and 3-level CACHE hierarchy.

The recently proposed retention-relax design could improve STT-MRAM write access performance. Nevertheless, the process variations could affect the writability of STT-MRAM cells. The situation for retention-relax design is even more severe. In [Sun et al., 2012c], they study the impact of process variations, including those from both, CMOS and magnetic technologies, on STT-MRAM design. Also, they propose process variation aware non-uniform CACHE access (PVA-NUCA) technique for large

STT-MRAM CACHE design. Besides the varying interconnect latencies determined by memory locations, PVA-NUCA compensates write time variations of **STT-MRAM** cells resulted by process variations. Two algorithms, conservative promotion and aggressive prediction, have been introduced and evaluated. A conflict-reduction mechanism is utilized to degrade the data access miss rate, caused by conflicts of access-intensive data blocks. Compared to the traditional **STT-MRAM** Dynamic Non-Uniform CACHE Access (**DNUCA**), their proposed dynamic PVA-NUCA can improve 25.29% of IPC performance and reduce 26.4% of **STT-MRAM** CACHE energy consumption, with $< 1\%$ of area overhead, according with the authors.

The simulation results in [Sun et al., 2012b], on the **STT-MRAM** based last-level CACHE, show that their technique can improve the system performance by 4% while obtaining 35% power reduction on average. According with them, emerging **NVM** memory technology, suffer from a common issue in their write operations: the switching processes at different write operations (i.e., $0 \rightarrow 1$ and $1 \rightarrow 0$) are asymmetric. Using a pessimistic design corner to cover the worst case of a write operation incurs large power and performance cost in the existing emerging memory technology designs. In this work, they demonstrated a universal log style write methodology to mitigate this asymmetry issue, by operating two switching processes in separate stages. The dedicated design optimizations are allowed on either switching process.

While GPUs are designed to hide memory latency with massive multi-threading, the tremendous demands for memory bandwidth and power consumption constrain the system performance scaling. In [Zhao and Xie, 2012], they propose a hybrid graphics memory architecture with different memory technologies (**DRAM**, **STT-MRAM**, and **RRAM**), to improve the memory bandwidth and reduce the power consumption. In addition, they present an adaptive data migration mechanism that exploits various memory access patterns of GPGPU applications, for further memory power reduction. They evaluate their design with a set of multi-threaded GPU workloads. Compared to traditional **GDDR5** memory, their design leads to 16% of GPU system power reduction, and improves the system throughput and energy efficiency by 12% and 33%.

Emerging memory technologies are explored as potential alternatives to traditional **SRAM/DRAM**-based memory architecture in future microprocessor designs. Among various emerging memory technologies, **STT-MRAM** has the benefits of fast read latency, low leakage power, and high density, and therefore has been investigated as a promising candidate for last-level CACHE (**LLC**). One of the major disadvantages for **STT-MRAM** is the latency and energy overhead associated with the write operations. In particular, a long-latency write operation to **STT-MRAM** CACHE may obstruct other CACHE accesses, resulting in performance degradation. Consequently, mitigation techniques to minimize the write overhead are required, in order to successfully adopt this new technology for CACHE design. In [Wang et al., 2013], they propose an obstruction-aware CACHE management policy called OAP. OAP monitors the CACHE to periodically detect **LLC**-obstruction processes and manage the CACHE accesses from different processes. The experimental results, on a 4-core architecture with an 8MB **STT-MRAM** **L3** CACHE, shows that the performance can be improved by 14% on average and up to 42%, with a reduction of energy consumption by 64%.

Data prefetching is a common mechanism to mitigate the bottleneck of off-chip memory bandwidth in modern computing systems. A major side effect is increase off-chip communication number of write operations. With the adoption of **STT-MRAM** at last-

level CACHES (LLCs) for their high density and low power consumption, the increase of write pressure to the CACHE from prefetching, associated with the long write access in MRAM, exacerbates the performance cost of prefetching schemes. In [Mao et al., 2013], they propose two techniques to reduce the negative performance impact induced by aggressive prefetching on multicore systems, employing STT-MRAM based LLC. First, basic priority assignment focus on the different types of access requests of LLC by their criticality, responding to them based on priority. Second, priority boosting differentiates requests by application, prioritizing the relatively few requests from applications with non-intensive accesses to the LLC, which usually creates the most severe performance degradation in multi-core systems. Combining these two prioritization policies they suggest that it can alleviate the negative effect induced by aggressive prefetching. The results show that these techniques could achieve an 8.3% average application speedup compared to a baseline, design with prefetch without prioritization.

Compared to the reviewed literature, our approaches to solve the architectural problems are strictly VLSI oriented, our results are based in physical factors, compared with real integrated circuits, a exception of the architectural simulation, where is quite difficult to present a complete time frame of an ARM v7 ISA in details, give the intellectual property protection. In at least two referenced researchs of the literature, they are proposing solutions, seems based in software solutions, which is not our intention. Besides that, a common factor between this work and the state of the art is the drawback of MRAM, due to the high current to switch the anisotropy. Furthermore, we do not have by objective circumvent or mitigate the latency of a MRAM, instead we try to present the advantages and drawbacks of adopting MRAM into the memory hierarchy and how this would affect an existing SoC performance and power consumption.

In the following sections and chapters, we will demonstrate our methodology to analyze individual memory banks for a specific technology and how we incorporate the details of this memory banks into an architectural level simulator. Finally, we will present the Composite Bank, a set of CACHE set associative that contains an original and simple to implement solution in order to obtain a better trade-off to incorporate MRAM into the memory hierarchy.

Part III

ANALYTICAL METHODOLOGY FLOW

This chapter will present our analytical methodology to evaluate memory banks and architectural systems given a memory technology. The set of tools employed to compose the methodology and how to combine them to extract useful data. Also, it will present in the end the results for intrinsic analyses.

ANALYTICAL METHODOLOGY FLOW

*If an elderly but distinguished scientist says that something is possible,
he is almost certainly right; but if he says that it is impossible,
he is very probably wrong.*
— Arthur C. Clarke

The main problematic addressed by this thesis is how and when to incorporate MRAM into memory hierarchy. The evaluations will denote that there are constraints about the adoption of MRAM into memory hierarchy, with the current technological state. Of course, if the technology achieve the expectation of ITRS, it will be possible to use it in the coming years. One major achievement to make possible the adoption of MRAM is indicated in [Section 2.3.6](#), the perpendicular MTJ. There are a series of problems to be addressed before the adoption of MRAM into mainstream designs. One major, if not the biggest drawback of MRAMs, is the MTJ write current (I_w). Once this is solved, MRAM becomes a viable technology using the current state of semiconductors manufacturing.

CACTI is used for the purpose to simulate [SRAM](#) L2 CACHE banks. It simulates bank-sets with the organization of TAG and DATA contained into the same element. This was due to the usage of associativity (≥ 4) as one of the parameters. Since NVSim is based on CACTI, the same pattern is observed on it.

NVSim is used on the same fashion as CACTI. Its main purpose is to simulate our [NVM](#) banks. Particularly for this thesis, the focus is [MRAM](#). Experiments were conducted for other technologies, but comparisons and results are not presented due to the fact that they were neither relevant nor substantial, to provide any valuable conclusion.

Further results were generated to simply compare [SRAM](#) vs [MRAM](#) banks. At this point, to obtain convergence and reliable results, we used only the NVSim simulator, since it produces nearly the same results as CACTI, with improvements and higher accuracy.

3.1 CACTI: INTEGRATED MEMORY SIMULATOR

CACTI is an integrated CACHE and memory simulator, developed by Hewlett-Packard, to evaluate a CACHE memory bank. By combining a set of models, we can have confidence that tradeoffs between time, power and area are all based on the same assumptions and, hence, are mutually consistent. CACTI is designed to be used to explore computer architecture at memory hierarchy level, to evaluate and understand the performance tradeoffs inherent in memory system organizations [[Labs](#)].

With the technological node shrink, the disparity between transistor and wire delay increases. The properties of future CACHE memories will depend on the characteristics of the interconnection network that connects various sub-modules of a CACHE memory. CACTI 6.5 is a significantly enhanced version that focuses on memory interconnect design [Tarjan et al., 2006].

3.1.1 CACTI background

This section presents some basics on the CACTI CACHE access model. Figure 3.1 shows the basic logical structure of a Uniform CACHE Access (UCA) organization. The address request to the CACHE is first provided as input to the decoder, which then activates a **wordline** in the **DATA** array and **TAG** array. The contents of an entire row are routed on the sense amplifiers through a network of wires. The routed outputs of the TAG arrays are compared against the input address, to detect if one of the ways of the set does contain the requested DATA. This comparator logic drives the multiplexor that finally forwards at least one of the routed output of the DATA array, from one of the sets, back to the requesting processor.

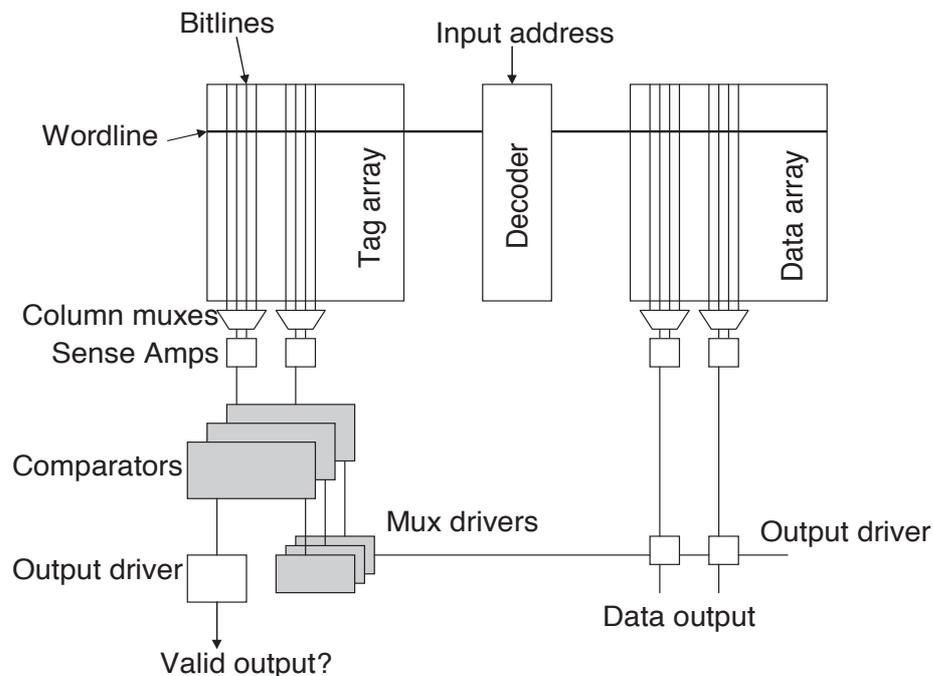


Figure 3.1: CACHE bank memory physical organization [Tarjan et al., 2006]

The CACTI CACHE access model [Tarjan et al., 2006] takes in the following major parameters as input: CACHE capacity, CACHE block size (also known as CACHE line size), CACHE associativity, technology node, number of ports and number of independent banks (not sharing address and DATA lines).

As result, it generates the CACHE configuration that minimizes delay, along with its power and area characteristics. CACTI models the delay, power and area of eight major CACHE components: decoder, wordline, bitline, sense amplifier, comparator, multiplexor, output driver and interbank wires. The wordline and bitline delays are two of

the most significant components of the access time. They are both quadratic functions of the width and height of each array, respectively [Tarjan et al., 2006].

In practice, the TAG and DATA arrays are complex and physically large structures, that makes it inefficient to implement them as single individual structures. Hence, CACTI partitions each storage array (in the horizontal and vertical dimensions) to produce smaller sub-arrays and reduce wordline and bitline delays. The bitline is partitioned into N_{dbl} different segments, the wordline is partitioned into N_{dwl} segments, and so on. Each sub-array has its own decoder and some central pre-decoding is now required to route the request to the correct sub-array. CACTI carries out an exhaustive search across different sub-array counts (different values of N_{dbl} , N_{dwl} , etc.) and sub-array aspect ratios to compute the CACHE organization with optimal total delay. A CACHE may be organized into a handful of banks. An example of a CACHE's physical structure is shown in Figure 3.1

3.1.2 CACTI Thesaurus

To clearly understand CACTI and its functioning, you should comprehend the terminology used by CACTI.

The following is a list of keywords introduced in various releases of CACTI [Tarjan et al., 2006].

- Bank: A memory structure that consists of a DATA and a TAG array. A CACHE may be partitioned into banks and CACTI assumes enough bandwidth to access these banks simultaneously. The network topology that interconnects these banks can change depending on the CACHE model (UCA or Non-Uniform CACHE Access (NUCA)).
- Sub-arrays: A DATA or TAG array is divided into a number of sub-arrays to reduce the delay due to wordline and bitline. The total number of sub-arrays in a CACHE is equal to the product of N_{dwl} and N_{dbl} .
- Mat: A group of four sub-arrays (2x2). They share a common central predecoder. CACTI's deep search starts from a minimum of one mat.
- Sub-bank: In a typical CACHE, a block is distributed across multiple sub-arrays to improve the reliability of a CACHE. Regardless of the CACHE organization, CACTI assumes that every CACHE block in a CACHE is distributed across a row of mats and the row number corresponding to a particular block is determined based on the block address. Each row (of mats) in an array is referred as a sub-bank.
- N_{twl}/N_{dwl} : Number of horizontal partitions in a TAG or DATA array i.e., the number of segments that a single wordline is partitioned into.
- N_{tbl}/N_{dbl} : Number of vertical partitions in a TAG or DATA array i.e., the number of segments that a single bitline is partitioned into.
- N_{tspd}/N_{spd} : Number of sets stored in each row of a sub-array. For a given N_{dwl} and N_{dbl} values, N_{spd} decides the aspect ratio of the sub-array.

- Ntcm/Ndcm: Degree of bitline multiplexing.
- Ntsam/Ndsam: Degree of sense-amplifier multiplexing.

3.1.3 *NUCA Modeling*

Earlier versions of CACTI assumed a **UCA** model, in which the access time of a CACHE is determined by the delay to access the farthest sub-array. To enable pipelining, an H-tree network is employed to connect all the sub-arrays of a CACHE. For large CACHES, this uniform model can suffer from a very high HIT latency. A more scalable approach for future large CACHES is to replace the H-tree bus with a packet-switched on-chip grid network. The latency for a bank, is determined by the delay between to route the request and the response from the bank that contains the DATA and the CACHE controller. Such **NUCA** model was first proposed in [Kim et al., 2002] and has been the subject of many architectural evaluations. CACTI builds upon this model and adopts the following algorithm to identify the optimal **NUCA** organization.

3.2 NVSIM

A series of new **NVM** technologies have emerged in the last years. Among all these emerging candidates for the so called **Universal Memory**, **STT-MRAM** or simply **MRAM**, **PCRAM** and **ReRAM** are regarded as the most promising candidates in replacing **DRAM**, **FLASH** and **SRAM**, while competing with others technologies for the spot into Registers, L1, L2, L3, **SSD** and **HDD** in the memory hierarchy [Dong et al., 2012b; Natarajan et al., 2009].

Since one of the purposes of this research is to evaluate the aspects of deploying **NVM** into the levels of the memory hierarchy, is necessary to explore the design space and find the most adequate implementations at different memory hierarchy levels. While some tools are available to analyze **SRAM/DRAM** design, like CACTI, similar tools for **NVM** designs were currently unavailable. So far to obtain base-line systems, only physical parameters and circuit designs were used, to analyze **NVM** memory banks, which means you had to apes through a design cycle to obtain some physical and electrical details. Currently, the main focus on this thesis is the usage of **MRAM**, but it can be extended to others **NVM** technologies, such as **RRAM**.

According with [Dong et al., 2012a], the NVSim is a circuit-level model simulator for **NVM** performance, energy, and area estimation, which supports **NVM** technologies, differently from CACTI, its base line, that only support **SRAM**. The supported **NVM** includes **STT-MRAM**, **PCRAM**, **ReRAM** and legacy NAND Flash. NVSim was validated comparing to industrial **NVM** prototypes and helps to boost architecture-level **NVM**-related exploration.

Seeing that we have to devise the impact of **MRAM** memory across multiple layers of the memory hierarchy, we need to evaluate the **MRAM** in a wide design space that covers a spectrum from high performance microprocessor caches to highly dense secondary storage.

However, since few of these **NVM** technologies are mature so far, only a limited number of prototype chips have been demonstrated, covering just a small portion of the emerging technologies. Therefore, additional data from physical implementations

would help to improve the precision of NVSim. Such data would help on architectural research.

3.2.1 NVSim *NVM Physical Mechanisms and Write Operations*

NVSim was made to support different *NVM* technologies with their particular storage mechanisms and write methods. These models as foreseeing by NVSim are described below.

3.2.2 *NAND Flash*

The physical mechanism of the Flash memory consists in store floating electrons through the floating gate and control the gate threshold voltage. The series bit-cell string of NAND Flash eliminates contacts between the cells and approaches the minimum cell size of $4F^2$ [Iizuka and Masuoka; Masuoka et al.]. The small cell size, low cost and strong application demands, make the NAND Flash dominates the traditional non-volatile memory market. A Flash memory cell consists of a floating gate and a control gate aligned vertically [Chen, 2012; Grupp et al., 2009]. The Flash memory cell modifies its threshold voltage V_T by adding electrons to, or subtracting electrons, from the isolated floating gate. NAND Flash usually charges or discharges the floating gate by using Fowler–Nordheim tunneling or hot electron injection [Masuoka et al.]. A program operation adds tunneling charges to the floating gate and the threshold voltage becomes negative, while an erase operation subtracts charges and the threshold voltage returns positive.

3.2.3 *STT-RAM*

The details of *STT-MRAM* are already well detailed presented in Section 2.3.4. The important details regarding NVSim are, When writing 0 state into *STT-MRAM* cells (RESET operation in NVSim modelisation), positive voltage difference is established between Source Line (SL) and Bit Line (BL). When writing 1 state (SET operation), the opposite voltage difference is established between BL and SL [Dong et al., 2012b]. The current amplitude and duration required to reverse the direction of the free ferromagnetic layer is determined by the size and aspect ratio of the *MTJ*, according with physical parameters off the *MTJ*.

3.2.4 *PCRAM*

PCRAM uses chalcogenide material to store information. The chalcogenide materials can be switched from a crystalline phase (SET state) to an amorphous phase (RESET state) with the application of heat (current induced). The crystalline phase shows low resistance, while the amorphous phase is characterized by high resistance. The SET operation crystallizes by heating it above its crystallization temperature, and the RESET operation melts it to make the material amorphous. The temperature is controlled by passing a specific electrical current profile and generating the required heat. High-power pulses are required for the RESET operation to heat the memory cell above the melting

temperature. In contrast, moderate power, but longer duration pulses, are required for the SET operation to heat the cell above the crystallization temperature, but below the melting temperature [141, 2004; 454, 2007; Dong et al., 2009; Raoux et al., 2008].

3.2.5 *ReRAM*

Although many nonvolatile memory technologies are based on electrically induced resistive switching effects, they define **ReRAM** as the one that involves electro and thermochemical effects in the resistance change of a metal/oxide/metal system. In addition, the authors of NVSim in [Dong et al., 2012a] confine their definition to bipolar **ReRAM**. An **ReRAM** cell consists of a metal oxide layer (e.g., Ti, Ta, and Hf [Chen et al., a; Yang et al., 2008b, 2010]) sandwiched by two metal (e.g., Pt [Yang et al., 2008b]) electrodes. The electronic behavior of metal/oxide interfaces depends on the oxygen vacancy concentration of the metal oxide layer. Typically, the metal/oxide interface shows Ohmic behavior in the case of very high doping and rectifying behavior in the case of low doping [Yang et al., 2008b]. The oxygen vacancy in metal oxide is n-type dopant, whose draft under the electric field can cause the change of doping profiles. Thus, applying electronic current can modulate the I-V curve of the **ReRAM** cell and, further, switch the cell from one state to the other state [479, 2008; Wong et al., 2012]. Usually, for bipolar **ReRAM**, the cell can be switched ON (SET operation), only by applying a negative bias, and OFF (RESET operation), only by applying the opposite bias [Yang et al., 2008b].

3.2.6 *Charge Pump*

The write operations in NAND Flash and **PCRAM** circuits require voltage higher than the chip supply voltage. Therefore, a charge pump that uses capacitors as energy storage elements to create a higher voltage is implemented in its designs.

In NVSim, the silicon area occupied by charge pump is neglected, since the charge pump area can vary, depending on its underlying circuit design techniques, and is relatively small compared to the cell array area in a large-capacity NAND circuit [Grupp et al., 2009].

However, the simulator models the energy dissipated by charge pumps during the operations on the memory cells, due to their contribution to the total energy consumption. The energy consumed by charge pumps is referred from an actual NAND Flash chip design [Ishida et al., 2009], which specifies that a conventional charge pump consumes 0.25μJ at 1.8V supply voltage. NVSim use this value as its default.

3.2.7 *Write Endurance Issue*

Write endurance is the number of times that an **NVM** cell can be overwritten. Among all the **NVM** technologies modeled in NVSim, only **STT-MRAM** doesn't suffer from the write endurance issue. NAND Flash, **PCRAM** and **ReRAM** all have limited write endurance. NAND Flash has write endurance of only $10^5 \sim 10^6$. The **PCRAM** endurance is now in the range between 10^5 and 10^9 [Ahn et al.; Lee et al., 2008]. **ReRAM** research currently shows endurance numbers in the range between 10^5 and 10^{10} [Kim et al., 2010b; Seevinck et al., 1991]. A projected plan by ITRS for 2024 for emerging NVM, i.e.,

PCRAM and ReRAM, highlight endurance in the order of 10^{15} , or more, write cycles [ITRS, 2012a,b]. In NVSim, the write endurance limit is not modeled.

3.2.8 Retention Time Issue

Retention time is the time that data can be retained in NVM cells. Typically, NVM technologies require retention time of higher than 10 years. However, in some cases, such a high retention time is not necessary. For example, in [Smullen et al.] they relaxed the retention time requirement to improve the timing and energy profile of STT-MRAMs. Since the tradeoff among NVM retention time and other NVM parameters (e.g., the duration and amplitude of write pulses) is on the device level, as a circuit-level tool, NVSim does not model this tradeoff directly, but instead, takes different sets of parameters with various retention periods as the device-level input.

3.2.9 MOS-Accessed Structure Versus Cross-Point Structure

Some NVM technologies (e.g., PCRAM in [Kau et al., 2009; Xu et al., 2011] and ReRAM in [479, 2008; Akinaga and Shima, 2010; Chen et al., b; Kau et al., 2009; Kim et al., 2010b; Lee and Wong, 2011; Lee et al., 2012a; Marinella, 2013; Park et al., 2012c; Terai et al., 2009; Valov et al., 2011; Wang et al., 2010a, 2012; Wong et al., 2012, 2010; Yoon et al., 2009; Zhirnov et al., 2010]) have the capability of building cross-point memory arrays without access devices. Conventionally, in the MOS-accessed structure, memory cell arrays are isolated by MOS access devices. The cell size is dominated by the access gate, that is necessarily dimensioned to drive enough current, even though the NVM cell itself is much smaller. However, taking advantage of the cell nonlinearity, a NVM array can be accessed without any extra access devices. The removal of MOS access devices leads to a memory cell size of only $4F^2$, where F is the process feature size. Unfortunately, the cross-point structure also brings extra peripheral circuitry design challenges and a tradeoff among performance, energy and area is always necessary as discussed in [Xu et al., 2011]. NVSim models both the MOS-accessed and the cross-point structures.

3.3 COMPARISON OF NVSIM TO CACTI

NVSim was also tested against CACTI, by simulating identical SRAM caches and DRAM chips. The results denoted that NVSim models SRAM and DRAM more accurately than CACTI does, since some false assumptions in CACTI are fixed in NVSim [Dong et al., 2012a]. Also, according with a detailed description of the original CACTI that is given in the CACTI technical report [Wilton and Jouppi, 1993], it is known that the CACTI has a 4% difference between CACTI and SPICE simulation for the same technological node and this performance have been kept in subsequent releases of CACTI [Muralimanohar et al., 2009; Reinman and Jouppi, 2000; Shivakumar and Jouppi, 2001; Tarjan et al., 2006; Thoziyoor et al., 2008b]. NVSim is validated by comparing against industrial prototype chips within the error range of 30% [Dong et al., 2012b].

The main reasons for adopting the NVSim is that, since it has improvements over CACTI, like better models for the SRAM cells than compared to CACTI, it is easier to concentrate in one tool to simplify the analytical flow. Also, NVSim allowed us estimate

the latencies, energy and silicon imprint of **NVM** memory banks before the effort of manufacturing a physical prototype. Besides, CACTI does not support **NVM** memory technologies.

NVSim also has as target, to always search for the optimal **NVM** memory bank organization, optimized for one design metric while keeping other metrics under constraints.

Still comparing to CACTI, NVSim also allow us to move sense amplifiers from inner memory subarrays to the outer bank level of the memory bank. Also, provides flexibility in the array organizations and data activation modes, by considering any combinations of memory data allocation and address distribution. It enables to try different models of sense amplifying, not only Voltage sensing. Another aspect of differentiation is that allows trying memory banks organized in a bus manner and not only as H-trees. NVSim also, permits provide the choices of buffers to use, not only for latency optimization, that is achieved on CACTI through logical effort.

Regarding modern memory banks organizations, NVSim is capable of modelize a cross-point memory cell [Johnson, 2008; Li et al., 2011b], not only MOS accessed memory cells. Also, considers the subarray size limit by analyzing the current sneak path. The last main advantage over CACTI, NVSim allows advanced target for users to redefine memory cell properties by providing a customized interface.

That is the reasoning, together with the fact that NVSim is based on CACTI, to choose the NVSim as the main tool used to perform our intrinsic evaluations.

3.4 RELIABILITY OF THE METHODOLOGY BASE-LINE MODELS

On the models mentioned at Table 3.1, the **SRAM** is a generically **SRAM** model assumption, based into the physics for the CMOS at 45nm, so the **SRAM** cell was dimensionally configured that way. If taken into account that the TSMC had at 45nm a $296\mu\text{m}^2$ and at 40nm a $242\mu\text{m}^2$ [TSMC, 2007, 2008], while at 28nm a $127\mu\text{m}^2$ [Wilson, 2009] **SRAM** cells, for 45nm we are almost with the same dimensioning as TSMC, slightly bigger. Also, based in [Iwai, 2008] as base-line, we are having pessimistic assumptions regarding the 28nm model, because ours is bigger compared to the TSMC model. Which is good, in my opinion, given the fact that the **MRAM** did not surpass the **SRAM** neither with a **SRAM** cell supersized. This denotes that **MRAM** still having lots of room to improve its operation. Since the physical parameters of the **MRAM** are coming from the SPINTEC, that means they are near exact to the real physical junction, providing more than 90% of confidence into them. For my **SRAM** model at 28nm, since it was dimensionate with twice the TSMC dimension, I would say that the accuracy of the results is, at most, of 70% ~ 80%. Which means the results are not the best, nor the most accurate, but the evaluation flow sustain its validity. This is proved by the fact that, for the 45nm node technology, we almost match the TSMC 45nm **SRAM** cell area dimensioning, which provides a great confidence into the results, based on the assumed models.

We found similar analyses in [Abu-Rahma et al., 2010; Argyrides et al., 2008; Cao et al., 2002; Chellappa et al., 2011; Chen et al., 2009; Do et al., 2006, 2007; Dong et al., 2008; Donkoh et al., 2012; Fang et al., 2005; Hu et al., 2002; Kanda et al., 2004; Kim et al., 2007, 2003; Kim and Guthaus, 2011; Kolar et al., 2011; Okabe and Abe, 2010; Park et al., 2002; Plass and Chan, 2007; Qureshi et al., 2009; Singhee, 2011; Sun et al., 2008; Swonger et al., 2012; Takeda et al., 2006; Tan et al., 2009; Wang et al., 2009b; Wu et al., 2010; Yueh et al., 2013]. All these publications evaluate **SRAM** cells. There are works

from dual-rail compilers of SRAM in 45nm to 32nm, high-k gate first process to SRAM cells, even circuit simulation and how to calibrate electrical simulators regarding the leakage current, in order to obtain a more fine-grain simulation model.

In the publications, the SRAM cells dimensioning and power detailing are similar to our models in one way or another. Ours seems to be one of the most accurate so far. There is not exactly a consensus about the right values, for reasons that goes from intellectual property protection, to foundry process protection. Others like us, that knows the details, cannot disclose them, so they assume a synthetical model that can be combined with synthetic libraries, having a high likelihood with real existing physical cells in same technological node, but that fits a synthetic library technology, so the results can be disclosed and published.

3.5 METHODOLOGY TO EVALUATE THE INTRINSIC ASPECTS OF MRAM COMPARED TO SRAM MEMORIES

As mentioned before, to evaluate the memory technology, specifically the MRAM, we have two main approaches: one is the intrinsic analyses, that means compare the memory banks in different technologies of memory with similar characteristics like size and associativity. This approach is depicted in [Figure 3.2](#), and we name it intrinsic analyses. In the second approach, we combine the results of the intrinsic analyses, applying them into the memory hierarchy analyses. This approach is depicted in [Figure 3.23](#).

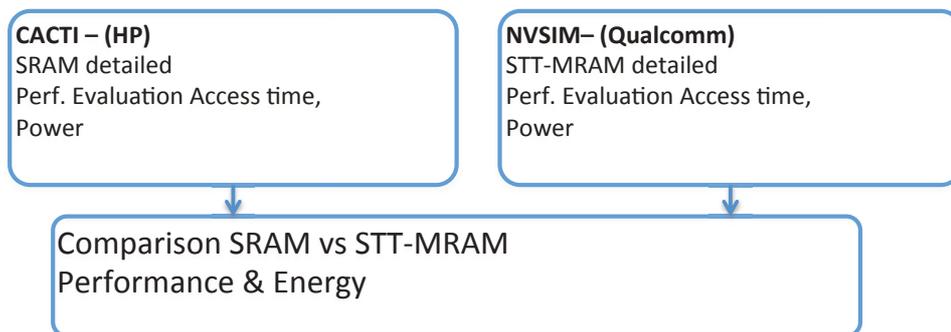


Figure 3.2: Evaluation Methodology depicted.

Depending on the objectives with the proposed model of [Figure 3.2](#), the researcher have the possibility to use NVSim or CACTI, we do not encourage the usage of both for comparisons purposes. If possible try to concentrate all your experiments with the same simulator, unless you are comparing the results from same technology in both tools. Our flow supports both CACTI or NVSim, remembering we discourage the usage of both in parallel. Is possible that you already have all you [SRAM](#) cells models for CACTI and validated against your own model of [STT-MRAM](#), that you built using CACTI, there is no reason to mix NVSim in this case, you can choose one of the tools, but your flow will have to use only one of them.

3.5.1 Memory models used in NVSim for analyses

For the purpose of comparing our memory models, we configured the NVSim memory models for two specific technologies: MRAM and SRAM.

The MRAM MTJ and the SRAM cells specifications employed for analyses on bank comparisons MRAM vs SRAM, are defined on [Table 3.1](#) specifications.

Table 3.1: Memory Cell characterization models, employed to simulate the electrical and physical resulting bank.

	pMTJ	SRAM	pMTJ
Technology	45nm	45/28nm	28nm
Temperature Operation	300K	300K	300K
Cell Area	10F ²	146F ²	20F ²
Aspect Ratio	1.0	1.46	1.0
R _p	7100	ϕ	18000
R _{ap}	15600	ϕ	36000
Read Current	20μA	ϕ	7μA
Write Current	29μA	ϕ	17μA
Write Pulse	10ns	ϕ	6ns
Access CMOS Width	6F	1.31F	2F
NMOS Width	ϕ	2.08F	ϕ
PMOS Width	ϕ	1.23F	ϕ

The [STT-MRAM](#) is evaluated using a MTJ model, for a specific technological node. A SRAM cell model was also described. These physical models of the memory cells are described in [Table 3.1](#). The NVSim based into the cell physical model began to search and calculate, using the equational models of CACTI and NVSim, in a spread search mode. Then, pruning the search trees looking for the best solution for a given memory bank given the set of constraints, like size, technological node, word size and associativity.

Besides, in the intrinsic analyses ([Section 3.5.3](#)) for the technology nodes of 45nm and 28nm, we also evaluate the memory banks regarding two possible implementations: High Performance ([HP](#)) and Low Power Performance ([LOP](#)). Specifically for the SRAM, these two models assume different kinds of SRAM cells of 6T and 8T cells respectively. In addition to that, the [LOP](#) has two additional transistors to control access to the memory cells, plus, it assumes that the memory bank has tension scaling for low power operation.

3.5.2 Intrinsic Analyses

The first milestone using the simulators was evaluate how the memory banks compare one-to-one. In that sense we used the NVSim and generated a series of memory banks:

- 16KB/4-way/32-bytes;

- 32KB/4-way/32-bytes;
- 64KB/4-way/32-bytes;
- 128KB/4-way/32-bytes;
- 256KB/4-way/32-bytes;
- 512KB/8-way/64-bytes;
- 1MB/8-way/64-bytes;
- 2MB/8-way/64-bytes;
- 4MB/8-way/64-bytes;
- 8MB/8-way/64-bytes.

The [Table 3.2](#), [Table 3.3](#), [Table 3.4](#), [Table 3.5](#), [Table 3.6](#) and [Table 3.7](#) are the sources for the depicting at [Section 3.5.3](#). The results presented in this tables are plotted there for a better understanding. The tables that do not appear into this chapter are included in [Appendix C](#), specifically the [HP](#) tables. This section it will be focused in the [LOP](#) models, however, the additional results for [HP](#) are found in [Appendix C](#). Analyzing the available data, result tables and the depicting results, one major conclusion we arrive is that the TAG array is the responsible for the low performance of the MRAM memory bank set. So, based on that, we devised and denominated the **Composite Bank**, wich is well detailed and discussed in [Section 4.5](#). In [Figure 3.3](#), [Figure 3.4](#), [Figure 3.5](#), [Figure 3.9](#), [Figure 3.10](#), [Figure 3.7](#), [Figure 3.8](#), [Figure 3.11](#), [Figure 3.12](#), [Figure 3.6](#), [Figure 3.17](#) are depicted the relations of power, latency, dynamic power and leakage. This are by far the most critical important details about the memory banks alone. Also, these are the factors that will affect the systems performance in the end.

Based on these results, and comparing with the SRAM, we could observe one critical factor: the TAG array. The TAG array is the critical component for [NVM](#) memory bank based on MRAM. This conclusion, as well as the graphics, is based analyzing the simulation results obtained with NVSim. Also, our results were obtained using NVSim [[Dong et al., 2012a](#)]. The results presented here match similar results presented in [[Chen et al., 2008](#); [De and Borkar, 1999](#); [Gangwal et al., 2006](#); [Hentrich et al., 2009](#); [Jan et al., 2009](#); [Jarollahi and Hobson, 2010](#); [Kuang et al., 2005](#); [Luo et al., 2004](#); [Mohammad et al., 2012](#); [Panda et al., 2009](#); [Pilo et al., 2013](#); [Rajendra Prasad et al., 2012](#); [Tuan and Lai, 2003](#); [Valaee and Al-Khalili, 2011, 2012](#); [von Arnim et al., 2007](#); [Wang et al., 2009a, 2010b](#)], given the technological node difference.

Table 3.2: Details about SRAM and MRAM memory banks, generated using NVSim, this table comprises the 45nm results of the memory banks for LOP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface Latency(ns)	0.048	0.053	0.064	0.086	0.109	0.278	0.392	0.618	0.996	1.830
Write (ns)	109.442	118.732	118.204	117.723	117.199	135.875	134.731	133.609	170.296	167.692
Hit (ns)	10.490	19.792	19.265	18.788	18.350	37.213	36.086	35.038	72.565	70.240
Miss (ns)	10.322	19.624	19.097	18.620	18.113	36.771	35.645	34.593	71.303	68.973
Power (operation)										
Write (nJ)	0.146	0.145	0.145	0.147	0.148	0.151	0.157	0.171	0.176	0.206
Read (nJ)	0.034	0.034	0.034	0.033	0.035	0.150	0.149	0.148	0.162	0.161
leakage (mW)	0.022	0.036	0.062	0.115	0.120	0.130	0.234	0.443	0.459	0.875
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface Latency(ns)	0.057	0.108	0.212	0.414	0.819	1.445	2.870	5.664	11.267	22.341
Write (ns)	1.099	1.977	1.960	2.553	3.768	3.740	4.884	11.245	13.168	38.388
Hit (ns)	1.293	2.287	2.271	3.134	4.352	5.323	6.474	16.702	18.632	59.228
Miss (ns)	1.158	2.036	2.019	2.610	3.825	3.797	4.942	7.658	9.890	16.632
Power (operation)										
Write (nJ)	0.002	0.002	0.003	0.004	0.006	0.009	0.009	0.016	0.017	0.030
Read (nJ)	0.007	0.013	0.022	0.038	0.077	0.139	0.249	0.523	0.964	2.025
leakage (mW)	0.141	0.263	0.517	0.998	1.969	3.426	6.807	13.354	26.577	52.550
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table 3.3: Details about SRAM and MRAM memory banks, generated using NVSim, this table comprises the 28nm results of the memory banks for LOP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface Latency(ns)	0.017	0.020	0.027	0.040	0.062	0.141	0.223	0.385	0.680	1.298
Write (ns)	9.132	9.048	12.134	11.983	11.801	11.802	18.150	17.737	32.194	31.125
Hit (ns)	4.784	4.700	7.796	7.646	7.566	7.892	14.252	13.856	29.663	28.654
Miss (ns)	4.644	4.560	7.656	7.505	7.326	7.317	13.675	13.276	27.752	26.738
Power (operation)										
Write (nJ)	0.005	0.005	0.006	0.006	0.006	0.007	0.008	0.009	0.011	0.013
Read (nJ)	0.044	0.043	0.043	0.043	0.043	0.200	0.200	0.199	0.203	0.202
leakage (mW)	0.112	0.152	0.235	0.391	0.518	0.677	1.098	1.920	1.982	3.624
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface Latency(ns)	0.022	0.042	0.083	0.160	0.317	0.560	1.111	2.193	4.362	8.655
Write (ns)	0.889	1.620	1.707	2.175	3.380	3.724	4.523	12.024	14.123	41.959
Hit (ns)	1.039	1.879	1.968	2.732	3.940	5.365	6.170	18.104	20.209	65.532
Miss (ns)	0.932	1.663	1.749	2.216	3.421	3.399	4.311	7.156	8.960	16.512
Power (operation)										
Write (nJ)	0.001	0.001	0.002	0.002	0.003	0.004	0.004	0.007	0.008	0.014
Read (nJ)	0.003	0.006	0.011	0.019	0.038	0.069	0.124	0.261	0.483	1.016
leakage (mW)	0.565	1.063	2.093	3.940	7.775	13.528	26.877	52.727	104.938	207.909
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

3.5.3 *Intrinsic Analyses - Results*

In the [Figure 3.3](#), [Figure 3.4](#), [Figure 3.5](#), [Figure 3.9](#), [Figure 3.10](#), [Figure 3.7](#), [Figure 3.8](#), [Figure 3.11](#), [Figure 3.12](#), [Figure 3.6](#), [Figure 3.17](#) are depicted the relations of power, latency, dynamic power, leakage and area. These are by far one of the most critical and important details about the memory banks alone. Also, these are factors that will affect the systems performance in the end.

The basic idea of the intrinsic analyses is to compare one-to-one memory banks of equivalent sizes, and understand how a system would perceive the impact of a memory technology switch.

As is observed in [Section 4.4](#), a synthesis was performed using a memory library characterized for that purpose. In this sense, similar results were generated for 45nm ([Section 3.5.3.1](#)) and 28nm ([Appendix E](#)). These results in two different technology nodes, also, specifically for 28nm, serves to illustrated the pattern behavior when adopting advanced submicronic nodes to build MTJs, that composes memory arrays, the impacts on power and latency for similar banks in 45nm.

3.5.3.1 *Memory banks of 45nm*

In [Figure 3.3](#) we can observe that, increasing the memory size, the area increases, which is expected. However, the area of [SRAM](#) memory banks starting in 1MB increases exponentially, while [MRAM](#) keeps a steady, almost linear, increase. One reason is the memory density: is known that the [MRAM](#) density is 4 ~ 8 times the [SRAM](#) density, e. g., [MRAM](#) at 45nm has a dimension of $10F^2$ while the [SRAM](#) has $146F^2$. So, for every increase in the memory array organization, the impact regarding the density in [SRAM](#) are much more critic than in the [MRAM](#), to keep the time coherence, power distribution and so on.

Table 3.4: Details about SRAM and MRAM DATA memory arrays, this table comprises the 45nm LOP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	29564.739	34460.487	44154.149	63878.447	80814.420	235478.564	339539.867	544975.240	881857.849	1638009.333
Latency(ns)										
Write (ns)	100.373	100.414	100.538	100.868	100.942	101.954	103.541	107.194	108.031	117.639
Read (ns)	1.208	1.218	1.238	1.359	1.586	2.207	2.536	3.603	6.089	10.046
Power (operation)										
Write (nJ)	67.254	68.085	69.746	73.069	74.175	77.057	83.707	96.986	101.329	127.887
Read (nJ)	16.051	16.054	16.061	16.075	18.498	116.539	116.571	116.612	126.226	126.308
leakage (mW)	0.019	0.032	0.058	0.110	0.115	0.122	0.226	0.433	0.445	0.860
Bandwith										
Write (GB/s)	0.149	0.149	0.148	0.148	0.148	0.293	0.288	0.279	0.277	0.254
Read (GB/s)	13.829	11.434	8.413	5.365	5.305	10.669	5.864	2.776	2.768	1.139
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	44105.891	84110.422	166207.457	326283.895	649941.472	1276301.711	2543600.249	5035430.745	10052727.329	19996890.560
Latency(ns)										
Write (ns)	0.534	0.727	0.869	1.464	1.817	3.617	4.401	11.245	13.168	38.388
Read (ns)	0.534	0.727	0.869	1.464	1.817	3.617	4.401	11.245	13.168	38.388
Power (operation)										
Write (nJ)	0.959	1.559	1.690	2.887	3.150	5.803	6.328	10.732	11.781	20.584
Read (nJ)	5.186	10.000	16.941	33.441	61.162	122.689	233.479	465.140	908.113	1814.240
leakage (mW)	0.112	0.208	0.409	0.793	1.573	3.030	6.046	11.883	23.740	47.063
Bandwith										
Write (GB/s)	37.599	25.833	20.939	11.608	9.442	8.796	7.269	2.741	2.339	0.788
Read (GB/s)	41.773	34.016	21.093	16.000	9.701	14.686	8.432	5.741	3.190	2.090
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table 3.5: Details about SRAM and MRAM DATA memory arrays, this table comprises the 28nm LOP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	9730.754	12459.250	17891.657	28841.432	47022.296	120235.151	193534.812	339864.533	603552.626	1161659.749
Latency(ns)										
Write (ns)	6.245	6.259	6.285	6.375	6.480	6.957	7.282	7.981	9.336	11.192
Read (ns)	1.722	1.727	1.734	1.786	1.928	2.306	2.401	2.626	4.226	5.041
Power (operation)										
Write (nJ)	2.572	2.621	2.719	2.914	3.305	4.251	4.639	5.416	6.839	8.393
Read (nJ)	20.325	20.327	20.332	20.340	21.287	158.001	158.016	158.045	161.692	161.750
leakage (mW)	0.080	0.120	0.198	0.354	0.479	0.616	1.027	1.847	1.892	3.531
Bandwith										
Write (GB/s)	2.423	2.419	2.410	2.389	2.352	4.367	4.191	3.823	3.266	2.723
Read (GB/s)	9.806	9.627	9.279	8.590	8.470	17.084	14.491	10.305	10.232	5.526
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	17132.668	33009.969	65227.487	126324.259	251631.183	494132.809	984786.763	1949511.361	3892008.360	8.655
Latency(ns)										
Write (ns)	0.439	0.631	0.766	1.396	1.735	3.724	4.523	12.024	14.123	41.959
Read (ns)	0.439	0.631	0.766	1.396	1.735	3.724	4.523	12.024	14.123	41.959
Power (operation)										
Write (nJ)	0.444	0.720	0.786	1.325	1.457	2.674	2.939	4.974	5.502	9.579
Read (nJ)	2.500	4.834	8.325	16.444	30.385	60.936	116.655	232.514	455.296	909.720
leakage (mW)	0.449	0.844	1.666	3.130	6.211	11.963	23.873	46.917	93.736	186.245
Bandwith										
Write (GB/s)	44.045	28.607	22.905	11.813	9.572	8.399	6.941	2.539	2.161	0.718
Read (GB/s)	44.235	35.824	21.657	16.063	9.543	14.053	7.930	5.336	2.898	1.891
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table 3.6: Details about SRAM and MRAM TAG memory arrays, this table comprises the 45nm LOP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	18041.728	18479.358	19611.676	22265.910	27753.110	42397.835	52607.607	73367.961	113903.168	192248.146
Latency(ns)										
Write (ns)	109.442	118.732	118.204	117.723	117.199	135.875	134.731	133.609	170.296	167.692
Read (ns)	10.322	19.624	19.097	18.620	18.113	36.771	35.645	34.593	71.303	68.973
Power (operation)										
Write (nJ)	78.624	77.202	75.587	74.397	74.041	73.715	73.270	74.380	74.215	78.053
Read (nJ)	17.454	18.403	17.913	17.421	16.935	33.811	32.834	31.860	35.584	34.463
leakage (mW)	0.003	0.004	0.004	0.005	0.005	0.008	0.009	0.009	0.014	0.015
Bandwith										
Write (GB/s)	0.040	0.036	0.035	0.035	0.034	0.029	0.029	0.028	0.021	0.021
Read (GB/s)	4.601	4.385	3.827	3.083	2.212	2.963	2.129	1.325	1.264	0.679
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	12649.157	24219.092	45747.695	87898.091	168985.746	169165.249	326355.013	628850.958	1214583.959	2343805.568
Latency(ns)										
Write (ns)	1.099	1.977	1.960	2.553	3.768	3.740	4.884	7.601	9.834	16.576
Read (ns)	1.158	2.036	2.019	2.610	3.825	3.797	4.942	7.658	9.890	16.632
Power (operation)										
Write (nJ)	0.609	0.858	1.481	1.451	2.653	2.785	2.734	5.014	4.900	9.014
Read (nJ)	1.597	2.566	4.606	4.486	15.825	16.033	15.585	57.468	55.706	210.971
leakage (mW)	0.029	0.055	0.108	0.205	0.396	0.396	0.761	1.471	2.837	5.487
Bandwith										
Write (GB/s)	4.816	2.421	2.443	1.766	1.142	1.151	0.839	0.518	0.383	0.220
Read (GB/s)	5.808	2.775	1.562	1.399	0.621	0.624	0.565	0.222	0.204	0.070
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table 3.7: Details about SRAM and MRAM TAG memory arrays, this table comprises the 28nm LOP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	7214.550	7690.061	8757.068	10886.561	15173.828	20819.341	29007.691	45183.339	76329.344	136770.376
Latency(ns)										
Write (ns)	9.132	9.048	12.134	11.983	11.801	11.802	18.150	17.737	32.194	31.125
Read (ns)	4.644	4.560	7.656	7.505	7.326	7.317	13.675	13.276	27.752	26.738
Power (operation)										
Write (nJ)	2.782	2.739	2.877	2.851	2.874	2.983	3.286	3.384	3.998	4.234
Read (nJ)	23.530	22.915	22.820	22.189	21.560	42.109	41.835	40.578	41.109	39.797
leakage (mW)	0.032	0.032	0.037	0.038	0.038	0.060	0.071	0.073	0.090	0.092
Bandwith										
Write (GB/s)	0.490	0.481	0.348	0.344	0.339	0.339	0.213	0.212	0.113	0.113
Read (GB/s)	2.953	2.846	2.743	2.616	2.448	2.464	2.364	2.123	2.019	1.663
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	4904.042	9392.464	17711.643	34030.664	65424.502	65494.212	126351.239	243464.797	470237.689	8.655
Latency(ns)										
Write (ns)	0.889	1.620	1.707	2.175	3.380	3.358	4.269	7.115	8.920	16.473
Read (ns)	0.932	1.663	1.749	2.216	3.421	3.399	4.311	7.156	8.960	16.512
Power (operation)										
Write (nJ)	0.287	0.411	0.722	0.708	1.306	1.361	1.336	2.471	2.415	4.475
Read (nJ)	0.767	1.254	2.278	2.219	7.907	7.994	7.770	28.808	27.925	105.974
leakage (mW)	0.115	0.219	0.427	0.809	1.564	1.565	3.004	5.809	11.202	21.664
Bandwith										
Write (GB/s)	5.837	2.916	2.746	2.044	1.255	1.264	0.951	0.548	0.420	0.219
Read (GB/s)	6.139	2.878	1.533	1.389	0.586	0.588	0.539	0.200	0.186	0.060
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

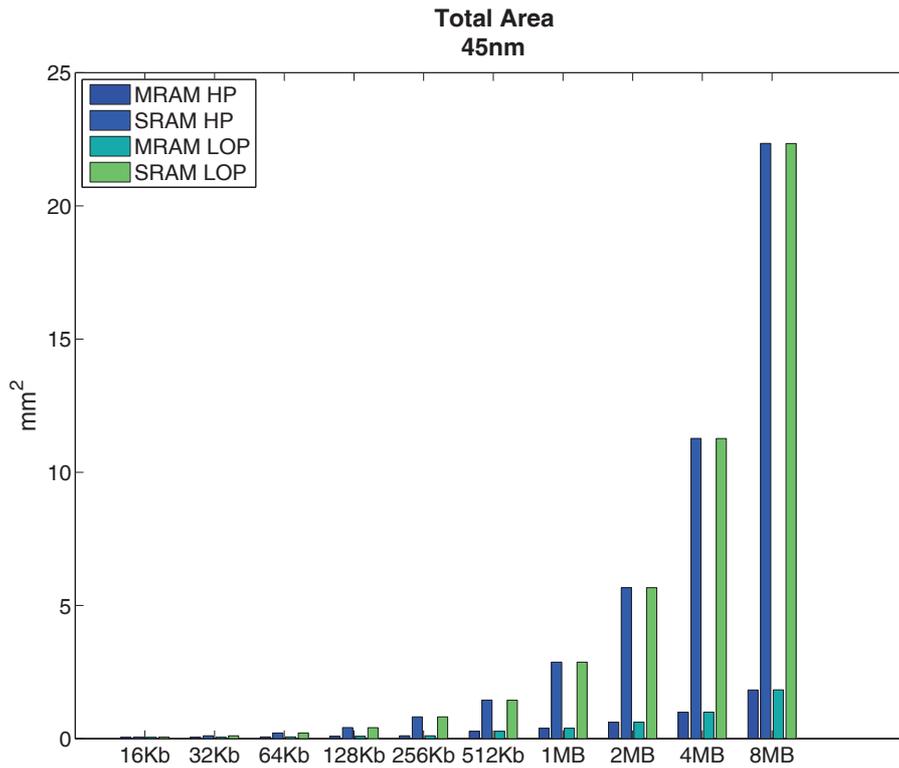


Figure 3.3: Total Area.

Table 3.8: Total Area (mm²).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.0474	0.0569	0.0476	0.0568
32KB	0.0528	0.1086	0.0529	0.1083
64KB	0.0637	0.2126	0.0638	0.2120
128KB	0.0863	0.4145	0.0861	0.4142
256KB	0.1085	0.8195	0.1086	0.8189
512KB	0.2757	1.4461	0.2779	1.4455
1MB	0.3904	2.8712	0.3921	2.8700
2MB	0.6178	5.6655	0.6183	5.6643
4MB	0.9919	11.2698	0.9958	11.2673
8MB	1.8285	22.3432	1.8302	22.3407

In [Figure 3.4](#) we can observe the stigma of a [MRAM](#) memory bank: the write latency. Assuming that a system has at least 80% readings [[Hennessy and patterson, 2006](#)], this would configure a great problem at the initialization time of a system, during the called cold start, when a cache has only compulsory misses. During this moment the application is loaded to the main memory at the boot or load moment.

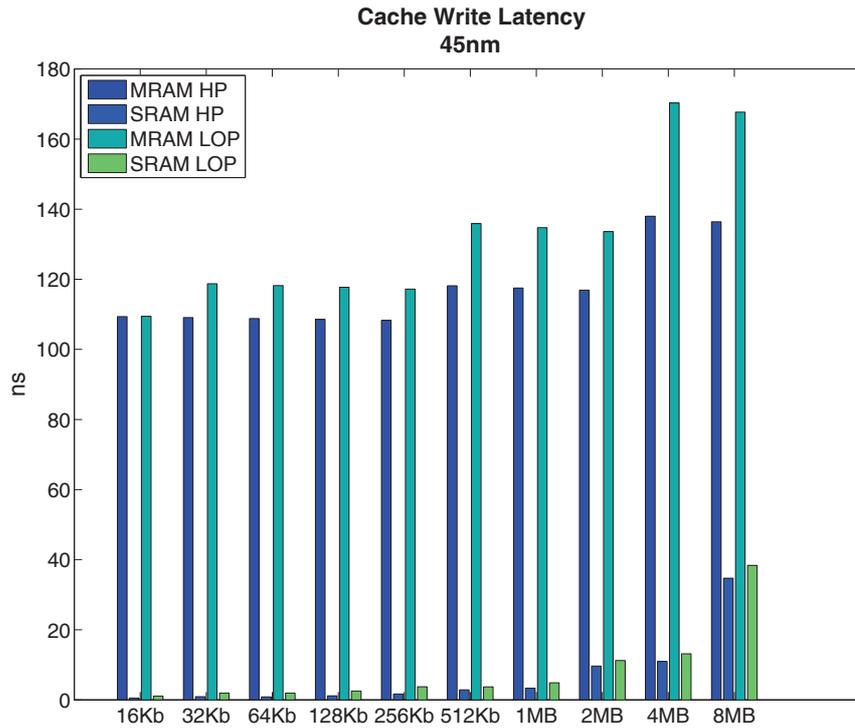


Figure 3.4: CACHE Total Write Latency.

Table 3.9: Cache Total Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	109.3390	0.5315	109.4420	1.0987
32KB	109.0710	0.9391	118.7320	1.9773
64KB	108.8020	0.8787	118.2040	1.9596
128KB	108.5620	1.1819	117.7230	2.5528
256KB	108.2970	1.6796	117.1990	3.7679
512KB	118.1400	2.8795	135.8750	3.7399
1MB	117.5150	3.3515	134.7310	4.8844
2MB	116.9060	9.6606	133.6090	11.2448
4MB	137.9750	10.9909	170.2960	13.1675
8MB	136.4120	34.7428	167.6920	38.3881

In Figure 3.7, Figure 3.5, Figure 3.6 and Figure 3.8 we can observe the major advantages of MRAM until the present moment, the leakage power. Regarding the HP set of memories, is almost unproductive to compare SRAM versus MRAM, since is practically impossible to SRAM achieve so low leakage power as MRAM. Leakage power is one of the biggest strengths of MRAM technology, that makes it competitive, even with the higher latency to write in the current technology state. Observing Table 3.10 and Table 3.11, is possible to conclude that, the TAG banks represent roughly a small-%

(exception in HP for SRAM) of the total leakage in any technology, being irrelevant to the calculus of total leakage.

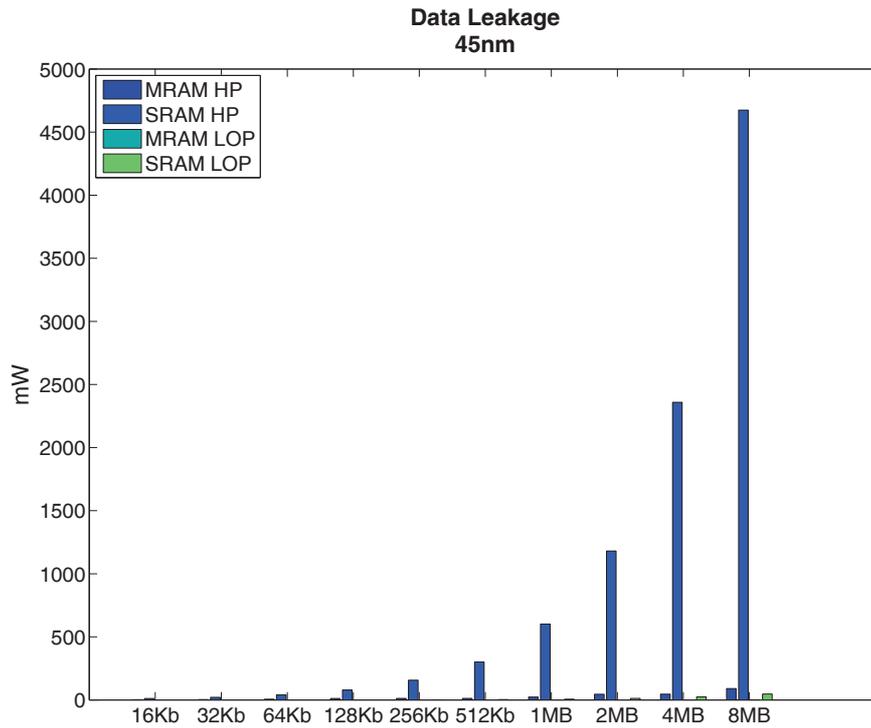


Figure 3.5: CACHE DATA Array Leakage Power.

Table 3.10: Cache Data Array Leakage Power (nW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	1.9093	11.2182	0.0189	0.1123
32KB	3.2807	20.7374	0.0320	0.2075
64KB	6.0130	40.8760	0.0580	0.4093
128KB	11.4824	79.0373	0.1101	0.7928
256KB	12.0386	156.7850	0.1155	1.5731
512KB	12.2694	301.3650	0.1216	3.0298
1MB	23.2024	601.3570	0.2258	6.0462
2MB	44.9838	1180.6800	0.4334	11.8826
4MB	46.1475	2358.7800	0.4448	23.7402
8MB	89.7104	4673.7500	0.8601	47.0629

In Figure 3.6 is depicted the TAG array leakage power. On HP memory banks, the leakage of the tag array increases proportionally to the DATA array. In this depicting, the LOP seems to have a better power relation, almost similar to the MRAM.

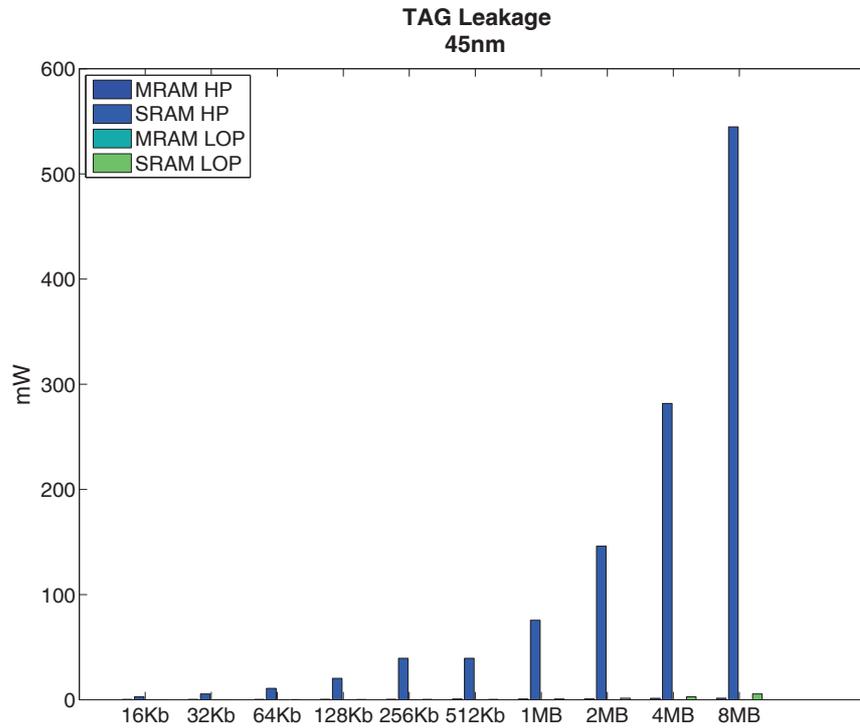


Figure 3.6: CACHE TAG Leakage Power.

Table 3.11: Cache TAG Array Leakage Power (nW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.3762	2.8944	0.0029	0.0291
32KB	0.3815	5.4894	0.0044	0.0552
64KB	0.3881	10.7580	0.0044	0.1081
128KB	0.4106	20.3837	0.0046	0.2050
256KB	0.4403	39.3581	0.0049	0.3960
512KB	0.7370	39.4043	0.0084	0.3965
1MB	0.7651	75.5880	0.0086	0.7610
2MB	0.8375	146.1160	0.0093	1.4713
4MB	1.3674	281.6980	0.0144	2.8372
8MB	1.4678	544.7340	0.0154	5.4868

In Figure 3.7, we observe the total leakage power, so TAG+DATA arrays of a cache memory bank. Adding up the TAG with the DATA does not help the SRAM case regarding leakage. Here in this case, the leakage is a great problem for HP memory banks. The great advantage of MRAM is that, despite its high latency, on power aspects it is always better. SRAM cannot overcome MRAM on leakage and general power consumption, due to the low power threshold for reading.

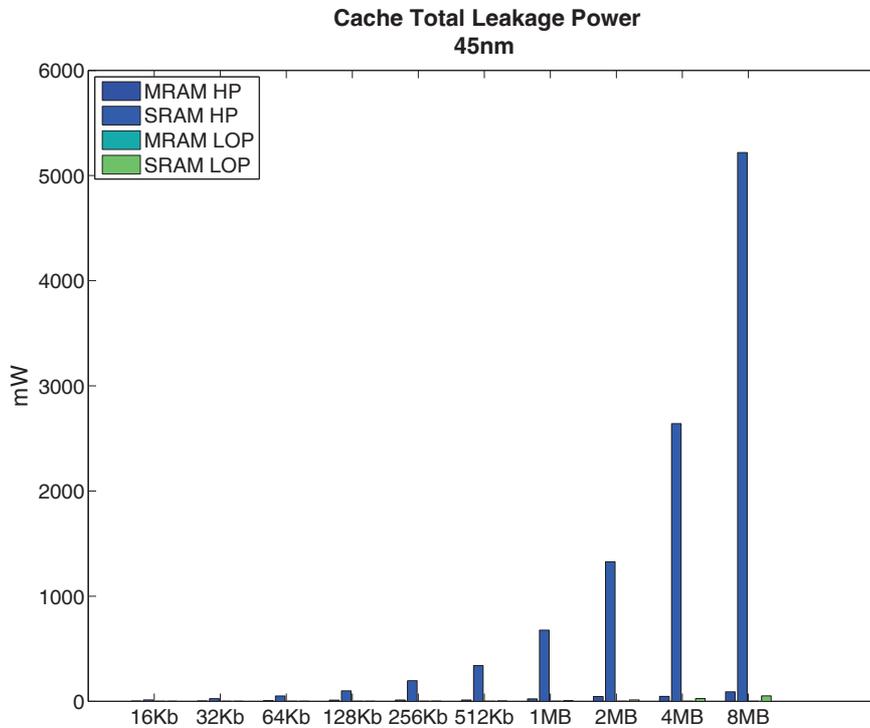


Figure 3.7: Leakage Power.

Table 3.12: Cache Total Leakage Power (mW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	2.2855	14.1126	0.0218	0.1414
32KB	3.6622	26.2267	0.0363	0.2627
64KB	6.4012	51.6340	0.0624	0.5174
128KB	11.8930	99.4210	0.1148	0.9978
256KB	12.4789	196.1430	0.1204	1.9691
512KB	13.0065	340.7700	0.1300	3.4262
1MB	23.9675	676.9460	0.2344	6.8072
2MB	45.8213	1326.8000	0.4427	13.3539
4MB	47.5148	2640.4800	0.4593	26.5774
8MB	91.1782	5218.4800	0.8754	52.5496

As said before in the previous depicting, the **SRAM LOP** seems to have a better power relation, almost similar to the **MRAM**, in fact this is not entirely true. In [Figure 3.8](#), we have a magnification, focusing only into the **LOP** banks. We can observe that the power leakage relation between **MRAM** and **SRAM** is maintained. It also shows two things: first) the **HP** memory banks architecture based in **SRAM**, consumes a great amount of energy and has a high leakage, as expected. What wasn't expected was that they would consume as depicted in [Figure 3.7](#); second) the **LOP** relation is not that advantageous, or more fair, compared against **MRAM**. The reason is that the **LOP SRAM** banks still having a higher leakage than **MRAM** counterparts. The **SRAM** leakage power increases exponentially even in **LOP** memory banks model. Surely, the **LOP** is also tied to mechanisms like clock-gating and low voltage (multi-voltage) to 'deactivate' the circuit during execution time in order to control power consumption. Regarding clock-gating it also greatly increases the latency as it will be seen.

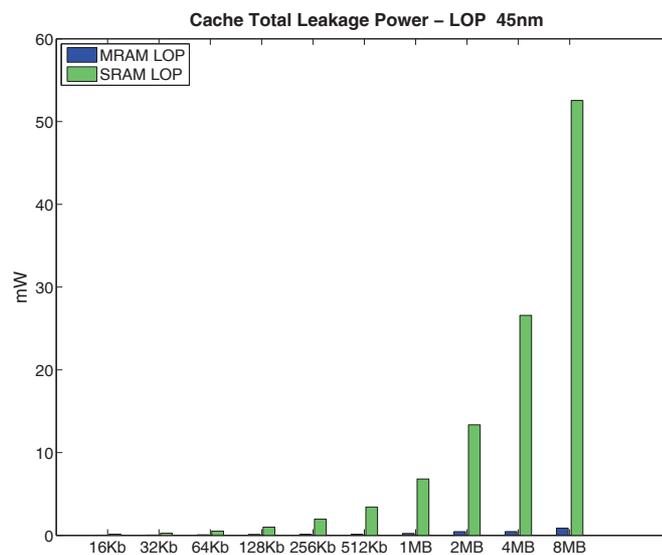


Figure 3.8: Low-Power (LOP) Total Leakage, observe only the **LOP** banks observed in [Figure 3.7](#).

In Figure 3.9 we have the energy depicting regarding a HIT into the memory bank, that means the information was found in the desired memory bank, without the latency to wait the MMU search for the block, in lower memory levels and updating the current memory bank. In our particular case, a HIT means the memory line is present and it will be updated or just read.

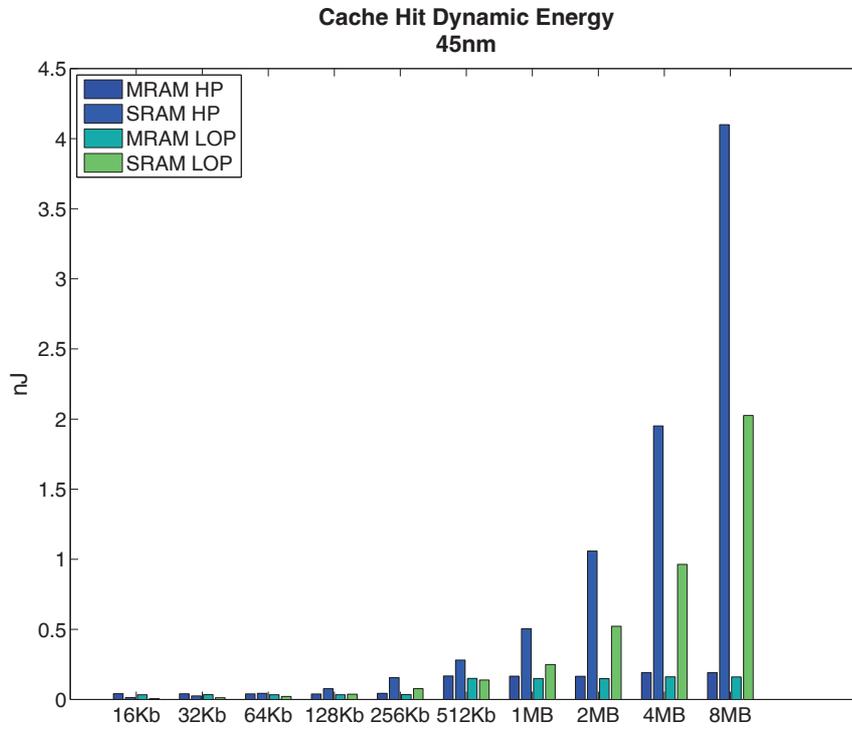


Figure 3.9: Hit Dynamic Energy.

Table 3.13: Cache Hit Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.0411	0.0138	0.0335	0.0068
32KB	0.0405	0.0256	0.0345	0.0126
64KB	0.0400	0.0438	0.0340	0.0215
128KB	0.0394	0.0770	0.0335	0.0379
256KB	0.0436	0.1561	0.0354	0.0770
512KB	0.1675	0.2815	0.1504	0.1387
1MB	0.1665	0.5047	0.1494	0.2491
2MB	0.1654	1.0586	0.1485	0.5226
4MB	0.1924	1.9511	0.1618	0.9638
8MB	0.1912	4.0991	0.1608	2.0252

In [Figure 3.10](#) we observe the HIT latency, due to the higher write latency of MRAM. At 45nm comparing equipotential memory banks, the MRAM always loses at the present state of the memory. Is known, based into the [[Zhao et al., 2012](#)], comparing with the present results, that there is a direct relation between the MTJ technological node to power and Latency of the memory bank. The reason is: lower the techonological node, lower the layers thickness, which needs lower electric currents to switch the anisotropy of the FL.

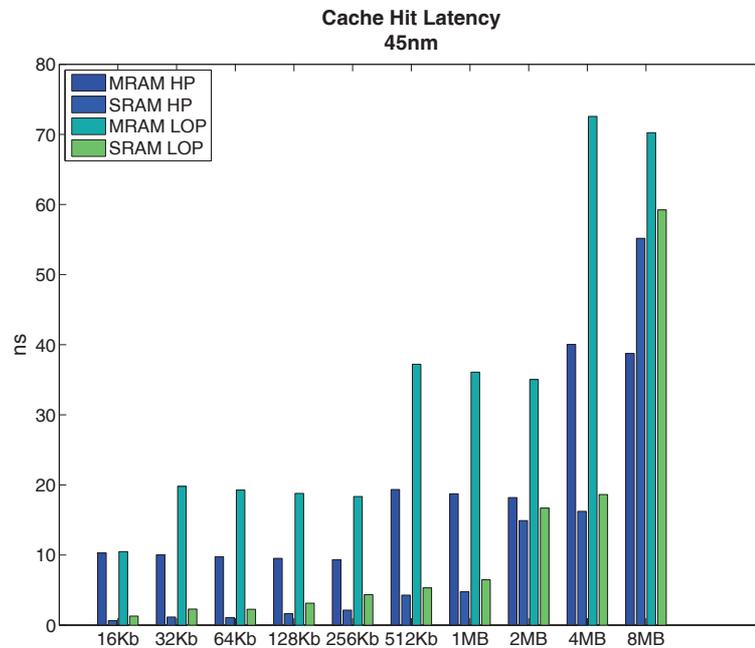


Figure 3.10: Hit Latency.

One factor that is easily noticeable is: regarding [Figure 3.11](#), it is the same as [Figure 3.9](#). So, miss or hit, the energetic cost will be almost the same.

Table 3.14: Cache Hit Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	10.2953	0.6533	10.4902	1.2926
32KB	10.0276	1.1525	19.7923	2.2867
64KB	9.7603	1.0939	19.2655	2.2712
128KB	9.5246	1.6477	18.7884	3.1341
256KB	9.3351	2.1476	18.3503	4.3523
512KB	19.3512	4.2859	37.2127	5.3230
1MB	18.7428	4.7618	36.0863	6.4738
2MB	18.2018	14.9050	35.0382	16.7024
4MB	40.0442	16.2397	72.5652	18.6320
8MB	38.7401	55.1709	70.2395	59.2280

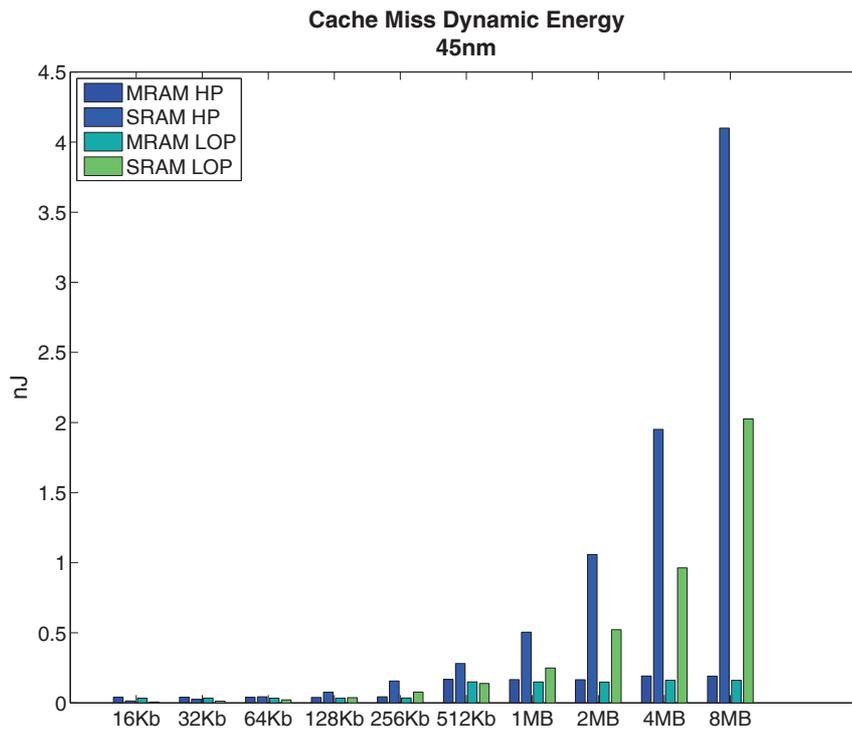


Figure 3.11: Miss Dynamic Energy.

Table 3.15: Cache Miss Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.0411	0.0138	0.0335	0.0068
32KB	0.0405	0.0256	0.0345	0.0126
64KB	0.0400	0.0438	0.0340	0.0215
128KB	0.0394	0.0770	0.0335	0.0379
256KB	0.0436	0.1561	0.0354	0.0770
512KB	0.1675	0.2815	0.1504	0.1387
1MB	0.1665	0.5047	0.1494	0.2491
2MB	0.1654	1.0586	0.1485	0.5226
4MB	0.1924	1.9511	0.1618	0.9638
8MB	0.1912	4.0991	0.1608	2.0252

On the other hand, if we compare the latencies depicted in Figure 3.12 with those depicted in Figure 3.10, we can observe that for SRAM a miss costs more, while for the MRAM case it is the same cost. The reason can be due to its high latency compared to the SRAM.

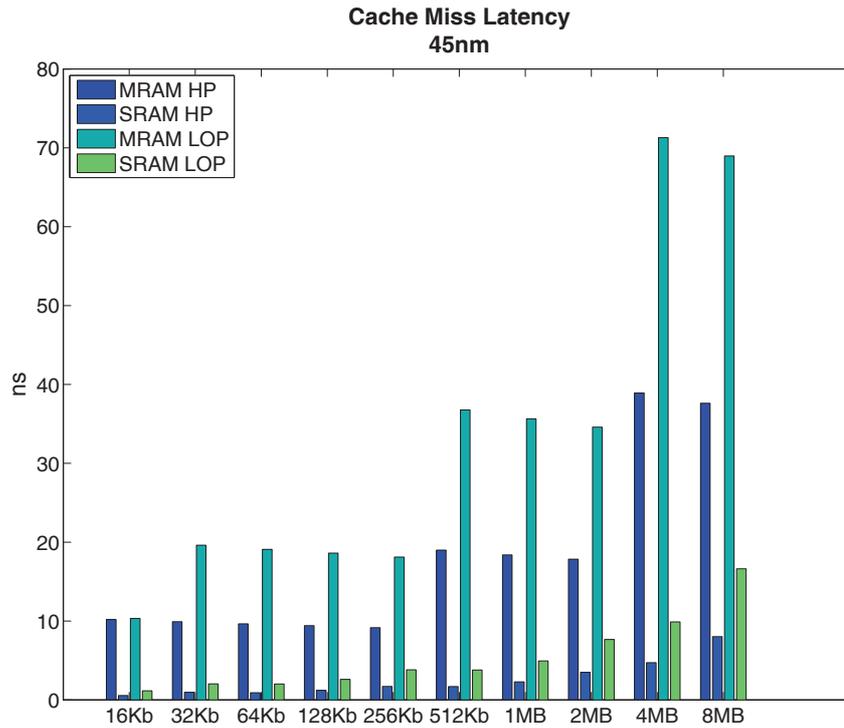


Figure 3.12: Miss Latency.

Table 3.16: Cache Miss Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	10.1874	0.5605	10.3223	1.1577
32KB	9.9195	0.9681	19.6243	2.0363
64KB	9.6520	0.9077	19.0972	2.0186
128KB	9.4158	1.2102	18.6196	2.6101
256KB	9.1674	1.7079	18.1132	3.8252
512KB	18.9930	1.6960	36.7710	3.7973
1MB	18.3854	2.2987	35.6448	4.9418
2MB	17.8419	3.4984	34.5929	7.6584
4MB	38.9224	4.7310	71.3032	9.8897
8MB	37.6153	8.0312	68.9729	16.6321

Here in [Figure 3.13](#), we have, one of the most interesting results of this experiment. Observing the figure, we see that the [MRAM LOP](#) latency is practically the same of the [SRAM HP](#). This has, at least, 3 possible implications: 1) you could replace the [SRAM HP](#) by [MRAM LOP](#), and have a considerable reduction in power consumption; 2) this is mainly valid to the data array matrix; 3) the TAG array represents the bottleneck of a [MRAM](#) memory bank. If the TAG bank had a better efficiency, the [MRAM](#) banks would have had a better outcome.

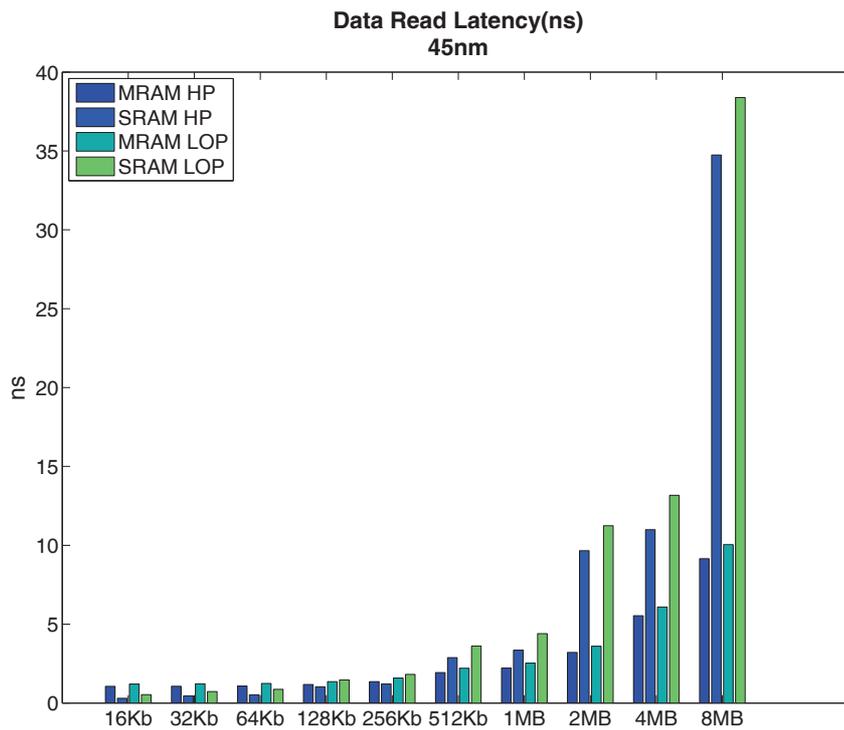


Figure 3.13: Data Read Latency.

Table 3.17: Data Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	1.0553	0.3040	1.2077	0.5339
32KB	1.0621	0.4597	1.2178	0.7269
64KB	1.0803	0.5260	1.2382	0.8690
128KB	1.1762	1.0291	1.3590	1.4635
256KB	1.3564	1.2156	1.5860	1.8168
512KB	1.9328	2.8795	2.2071	3.6172
1MB	2.2221	3.3515	2.5359	4.4008
2MB	3.2074	9.6606	3.6027	11.2448
4MB	5.5340	10.9909	6.0894	13.1675
8MB	9.1545	34.7428	10.0463	38.3881

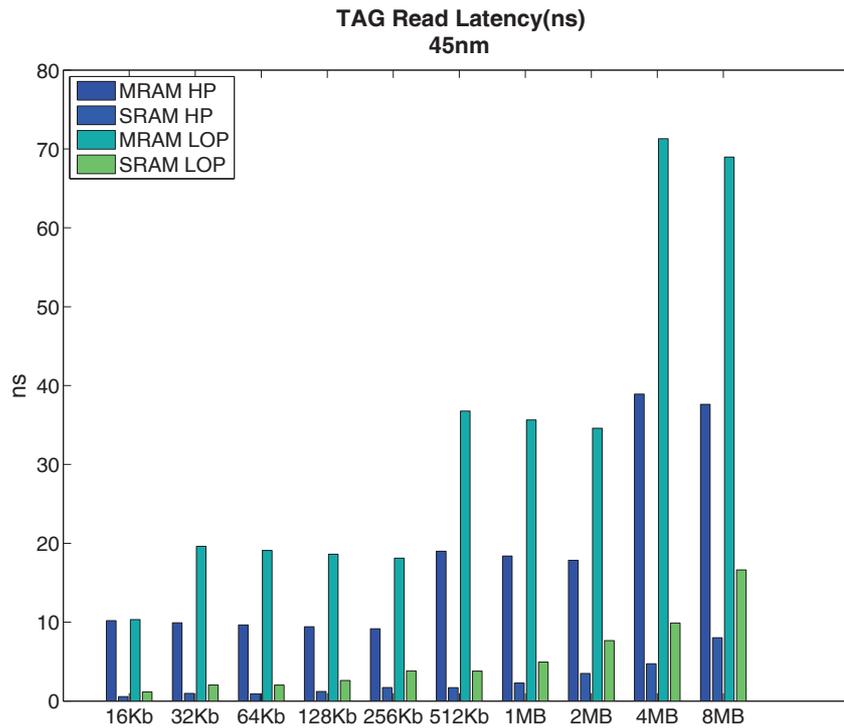


Figure 3.14: TAG Read Latency.

Table 3.18: TAG Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	10.1874	0.5605	10.3223	1.1577
32KB	9.9195	0.9681	19.6243	2.0363
64KB	9.6520	0.9077	19.0972	2.0186
128KB	9.4158	1.2102	18.6196	2.6101
256KB	9.1674	1.7079	18.1132	3.8252
512KB	18.9930	1.6960	36.7710	3.7973
1MB	18.3854	2.2987	35.6448	4.9418
2MB	17.8419	3.4984	34.5929	7.6584
4MB	38.9224	4.7310	71.3032	9.8897
8MB	37.6153	8.0312	68.9729	16.6321

Observing the results in Table 3.24 and Table 3.23, we noticed the latency factor played by the TAG array into the total performance of the CACHE memory bank. This observation, after many essays helped us to devise the Composite Memory Bank (CMB).

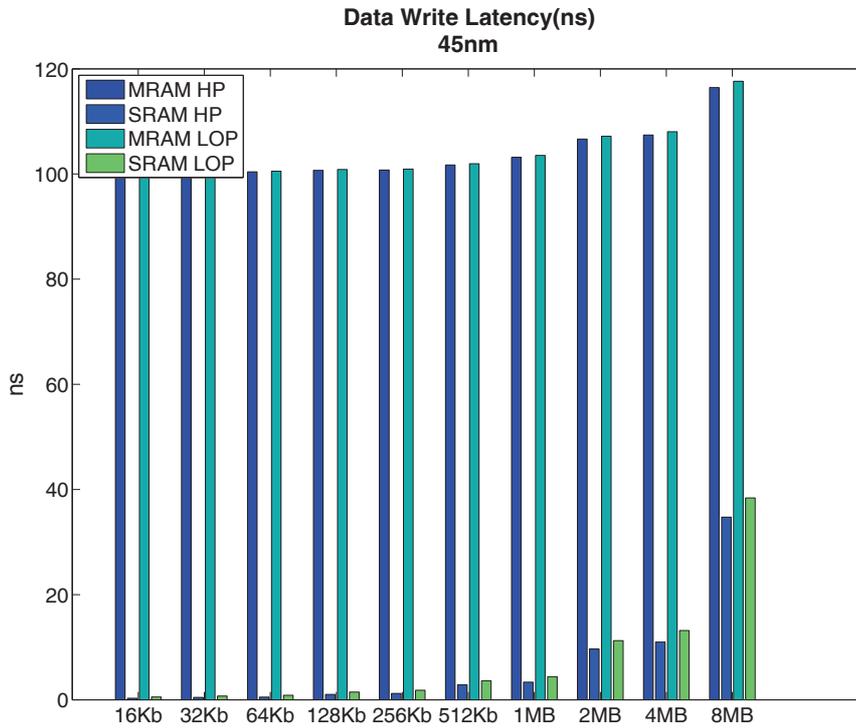


Figure 3.15: DATA Write Latency.

Table 3.19: DATA Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	100.2340	0.3040	100.3730	0.5339
32KB	100.2870	0.4597	100.4140	0.7269
64KB	100.4030	0.5260	100.5380	0.8690
128KB	100.6970	1.0291	100.8680	1.4635
256KB	100.7590	1.2156	100.9420	1.8168
512KB	101.7170	2.8795	101.9540	3.6172
1MB	103.1950	3.3515	103.5410	4.4008
2MB	106.6340	9.6606	107.1940	11.2448
4MB	107.4050	10.9909	108.0310	13.1675
8MB	116.4490	34.7428	117.6390	38.3881

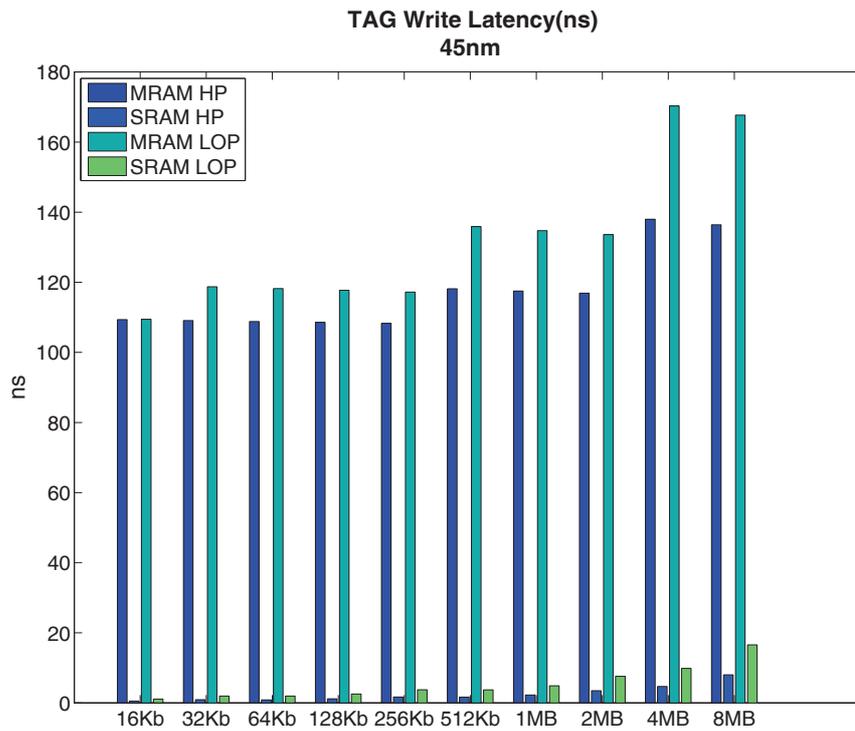


Figure 3.16: TAG Write Latency.

Table 3.20: Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	109.3390	0.5315	109.4420	1.0987
32KB	109.0710	0.9391	118.7320	1.9773
64KB	108.8020	0.8787	118.2040	1.9596
128KB	108.5620	1.1819	117.7230	2.5528
256KB	108.2970	1.6796	117.1990	3.7679
512KB	118.1400	1.6677	135.8750	3.7399
1MB	117.5150	2.2704	134.7310	4.8844
2MB	116.9060	3.4701	133.6090	7.6010
4MB	137.9750	4.7034	170.2960	9.8340
8MB	136.4120	8.0036	167.6920	16.5764

In Table 3.21 we can observe, supported by the depicting Figure 3.17, is possible to observe that due to the high current, necessary to store the information into the MTJ, this affects performance of the MRAM. It departs already with 63.96 times more energy than SRAM in 16KB in HP, and in LOP it began with 93.125 times more dynamic energy to write than SRAM. This helps to demonstrate the biggest, drawback so far of the MRAM technology.

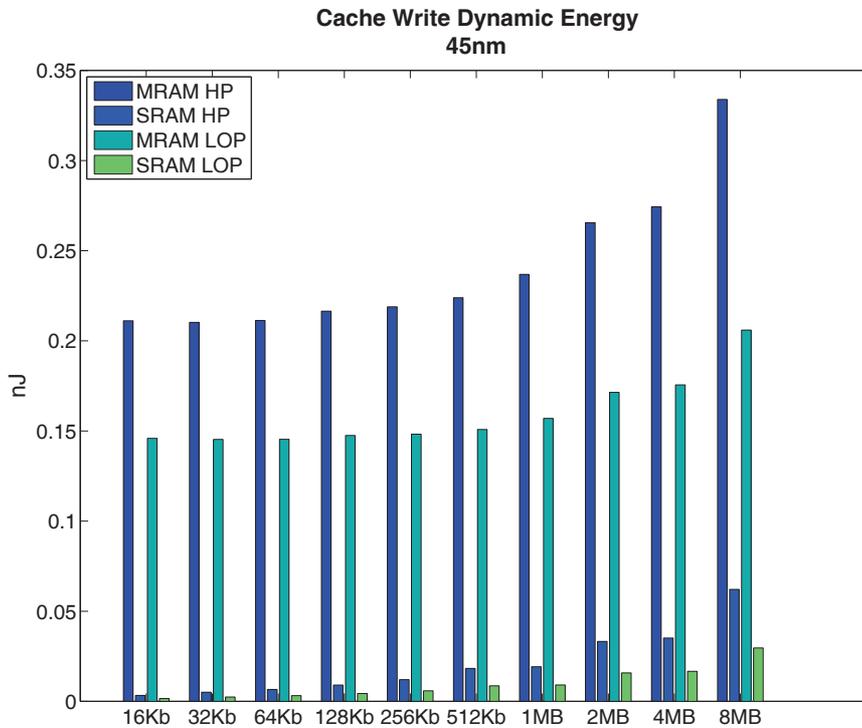


Figure 3.17: CACHE Write Dynamic Energy

Table 3.21: CACHE Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.2111	0.0033	0.1459	0.0016
32KB	0.2102	0.0050	0.1453	0.0024
64KB	0.2113	0.0066	0.1453	0.0032
128KB	0.2164	0.0090	0.1475	0.0043
256KB	0.2188	0.0121	0.1482	0.0058
512KB	0.2239	0.0182	0.1508	0.0086
1MB	0.2369	0.0192	0.1570	0.0091
2MB	0.2656	0.0332	0.1714	0.0157
4MB	0.2743	0.0351	0.1755	0.0167
8MB	0.3339	0.0621	0.2059	0.0296

For the TAG arrays the problem with dynamic energy to write is not as critical as in the DATA array, as is possible to observe in Table 3.22 .

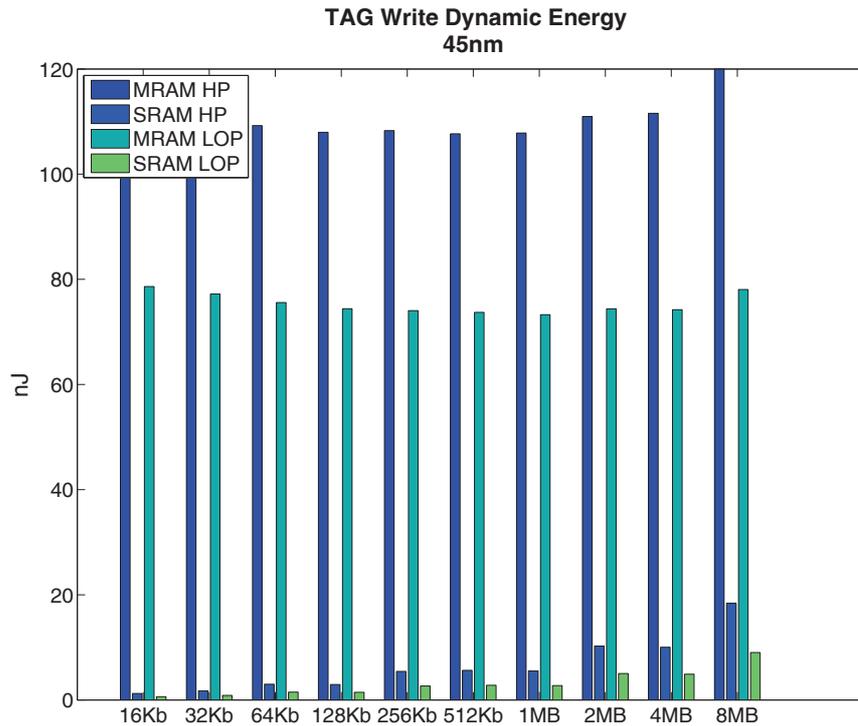


Figure 3.18: TAG Dynamic Energy

Table 3.22: TAG Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	113.8640	1.2262	78.6242	0.6093
32KB	111.3400	1.7284	77.2016	0.8576
64KB	109.2400	2.9948	75.5873	1.4812
128KB	107.9570	2.9350	74.3972	1.4513
256KB	108.2780	5.4298	74.0408	2.6533
512KB	107.6720	5.6278	73.7150	2.7846
1MB	107.8230	5.5258	73.2701	2.7341
2MB	110.9670	10.2586	74.3801	5.0145
4MB	111.5730	10.0250	74.2152	4.9001
8MB	119.9700	18.3804	78.0532	9.0142

Observing the results [Table 3.23](#) and comparing it to [Table 3.24](#), we notice, supported by [Figure 3.19](#) and [Figure 3.20](#), that the burden caused by the TAG array built using [MTJ](#), due to the write operation, given the fact that the total write time is mixture among write latency of TAG+DATA, is tremendous. Based into this facts, we began to evaluate and finally achieved the proposed solution of the Composite Memory Bank (CMB), to help mitigate this performance penalty caused by the TAG array built in MRAM as is explained in [Section 4.5](#).

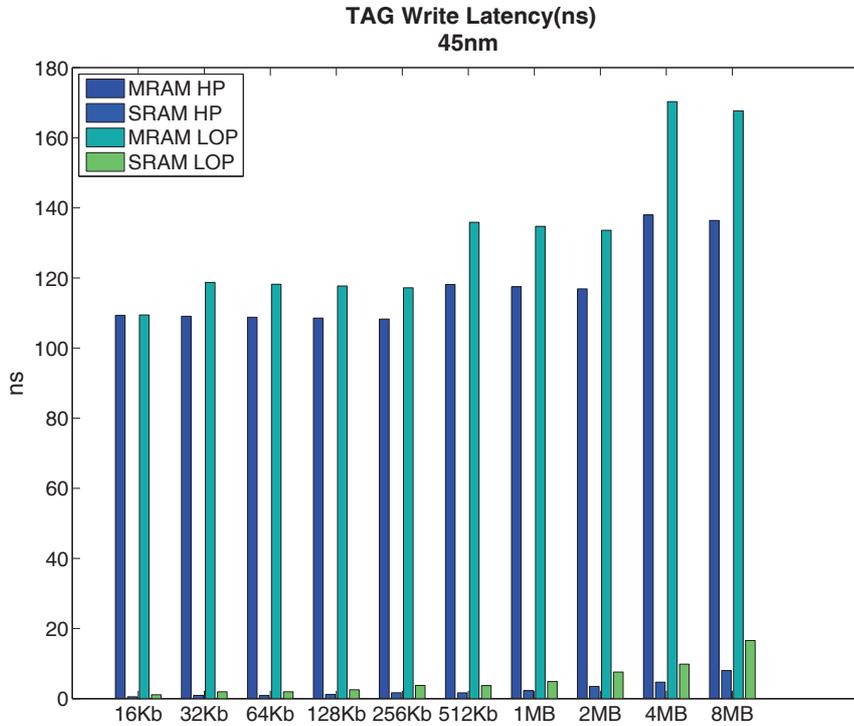


Figure 3.19: TAG Write Latency

Table 3.23: TAG Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	109.3390	0.5315	109.4420	1.0987
32KB	109.0710	0.9391	118.7320	1.9773
64KB	108.8020	0.8787	118.2040	1.9596
128KB	108.5620	1.1819	117.7230	2.5528
256KB	108.2970	1.6796	117.1990	3.7679
512KB	118.1400	1.6677	135.8750	3.7399
1MB	117.5150	2.2704	134.7310	4.8844
2MB	116.9060	3.4701	133.6090	7.6010
4MB	137.9750	4.7034	170.2960	9.8340
8MB	136.4120	8.0036	167.6920	16.5764

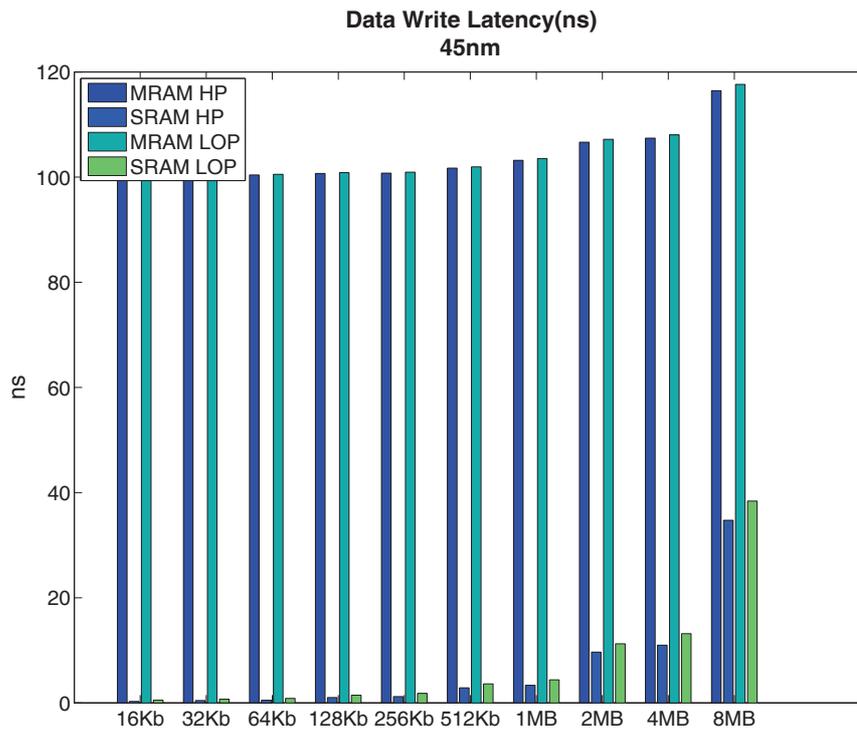


Figure 3.20: Data Write Latency

Table 3.24: Data Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	100.2340	0.3040	100.3730	0.5339
32KB	100.2870	0.4597	100.4140	0.7269
64KB	100.4030	0.5260	100.5380	0.8690
128KB	100.6970	1.0291	100.8680	1.4635
256KB	100.7590	1.2156	100.9420	1.8168
512KB	101.7170	2.8795	101.9540	3.6172
1MB	103.1950	3.3515	103.5410	4.4008
2MB	106.6340	9.6606	107.1940	11.2448
4MB	107.4050	10.9909	108.0310	13.1675
8MB	116.4490	34.7428	117.6390	38.3881

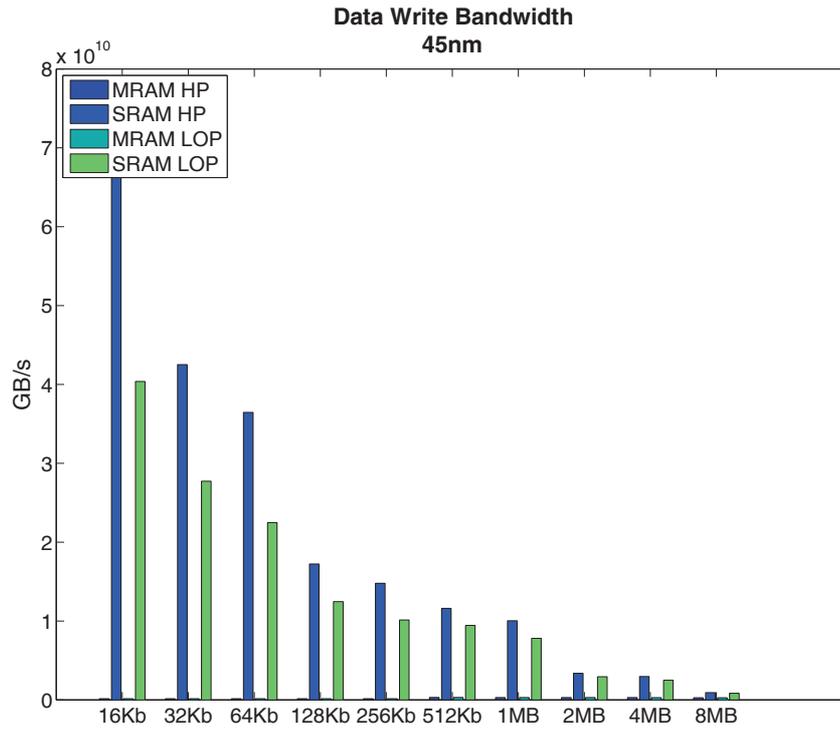


Figure 3.21: Write Bandwidth.

Table 3.25: Write Bandwidth (GB/s)

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.1488	65.38	0.1486	37.60
32KB	0.1487	39.60	0.1486	25.83
64KB	0.1485	33.96	0.1484	20.94
128KB	0.1481	16.05	0.1480	11.61
256KB	0.1480	13.77	0.1479	9.442
512KB	0.2933	10.82	0.2928	8.796
1MB	0.2891	9.347	0.2884	7.269
2MB	0.2798	3.149	0.2786	2.741
4MB	0.2778	2.767	0.2765	2.339
8MB	0.2563	0.8649	0.2540	0.7880

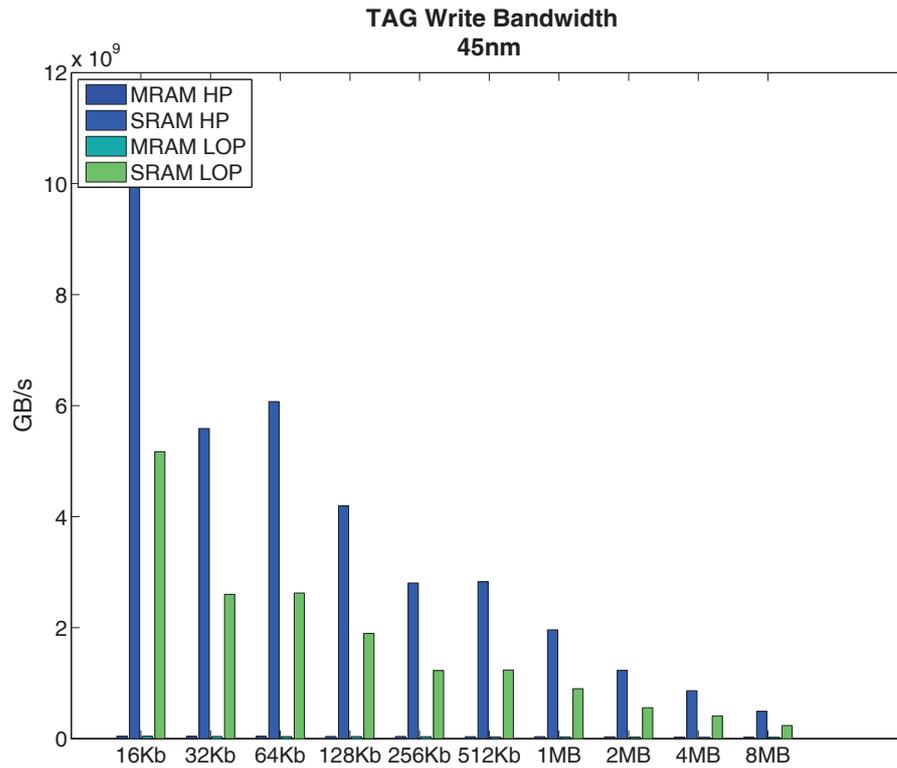


Figure 3.22: Write Bandwidth.

Table 3.26: Write Bandwidth (GB/s)

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP
16KB	0.04049	10.29	0.04047	4.816
32KB	0.03952	5.204	0.03632	2.421
64KB	0.03855	5.656	0.03550	2.443
128KB	0.03757	3.904	0.03466	1.766
256KB	0.03658	2.613	0.03383	1.142
512KB	0.03354	2.634	0.02917	1.151
1MB	0.03272	1.825	0.02856	0.8387
2MB	0.03190	1.144	0.02793	0.5179
4MB	0.02618	0.8042	0.02122	0.3830
8MB	0.02563	0.4558	0.02086	0.2195

Observing this preliminary results we were able to understand the strengths and weaknesses of MRAM compared to SRAM. Also, we could identify the architectural targets to explore and propose improvements, like the write access timing, that can be exploited with perpendicular MTJ. In addition to that, we observed that: the write bandwidth and latency are affected by the TAG arrays in the CACHE set banks. Another important conclusion is: as we increase the bank sizes and due to that the complexity of internal circuitry to access the matrix of memory cells proportionally increases, the differences in bandwidth and write latency between MRAM and SRAM also decreases.

3.6 METHODOLOGY TO EVALUATE THE ASPECTS OF MRAM INTO MEMORY HIERARCHY

After has tested, validated and worked on experiments with the described simulators Section 3.1, Section 3.2 and Section 3.8 a methodology on how to combine them was drafted.

The methodology is quite straightforward, as depicted in Figure 3.23. The memory bank simulators are used to generate memory bank models, obtaining the essential details like latency, silicon imprint and power of the memory banks. After that, these values are used to calibrate the architectural simulator. Then the architectural simulator provides support for the analyses, with the behavior of an existing architecture and the performance of a system equipped with this kind of memory banks into its memory hierarchy. After that, results are gathered to estimate power consumption, using power details from the memory bank simulator with the architectural outcome for this bank.

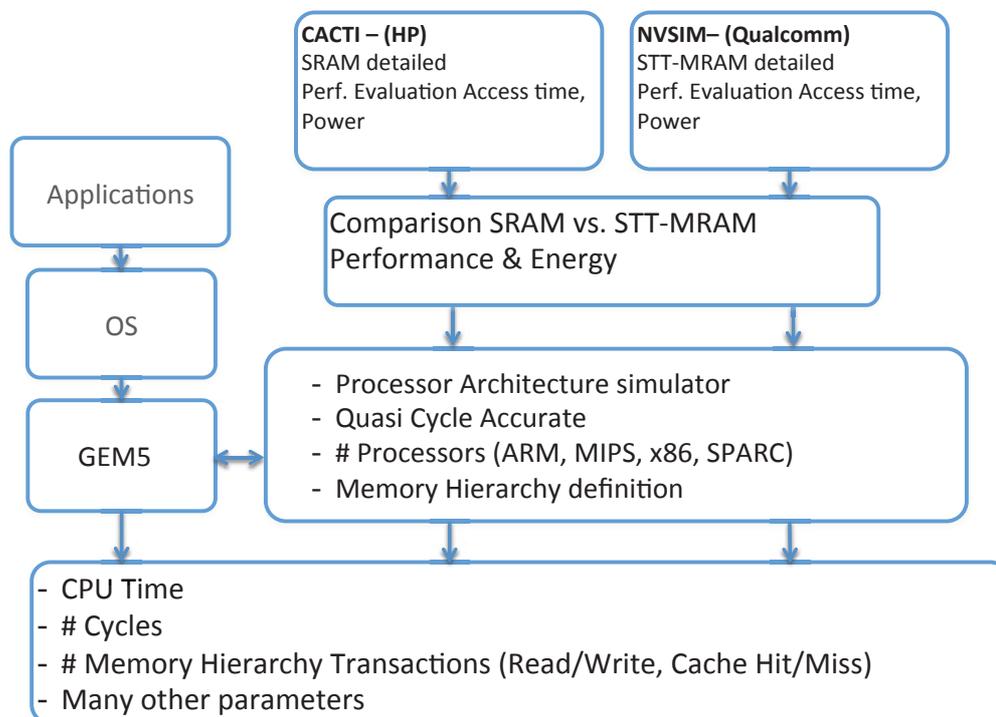


Figure 3.23: Evaluation Methodology depicted.

The Gem5 is used to measure performance and architectural statistics regarding the overall system. It was modified, so is possible to set the architecture parameters on-the-fly, automatizing the workload flow.

Gem5 generates statistics data regarding the access to the memory hierarchy like number of accesses for read and write, misses, hits, replacements, for the entire memory hierarchy. It is also possible to configure it and analyze, according with our own system, to the approximate performance of the target SoC.

The script code below is merely illustrative it is not the integral script used.

```
#!/usr/bin/perl -w
use strict;
use POSIX qw/strftime/;

my @cpu_cores=('1','2');#,3,4,5,6,7,8,9,10,12,20)
#my @l1_size=("1kB","2kB","4kB","8kB","16kB","32kB","64kB");
my @l1_size=("64kB");#,"32kB","16kB","4kB","2kB","1kB");

#my @l1_assoc = ('1','2',
my @l1_assoc = ('4');#,'8');
my @l1_latency = ('1ns');
my @l1_line = ('32');#,'64');
my @l2_size=("2GB","1GB","512MB","256MB","128MB","64MB","32MB","16MB","8MB",
"4MB","2MB","1MB","512kB","256kB","128kB");
##EMERGENCY
#my @l2_assoc = (1,2,4,8);
my @l2_assoc = (8);
#my @l2_latency = ('1ns','2ns','4ns','8ns','12ns','16ns','18ns','20ns','100ns');
my @l2_latency = ('1ns','8ns','12ns','100ns');

#my @freq=("600MHz","800MHz","1GHz","1.66GHz","2GHz","2.53GHz","4GHz");
my @freq=("1.5GHz");
my @mode=("timing");#,"detailed","inorder");

my $GEM5 = "build/ARM/gem5.opt";

....
foreach ...{
system("$GEM5 configs/lirmm/mram.py --cpu-type=$param_mode -n
$param_cpu --num-dirs=$param_cpu --caches --l2cache
--l1i_size=$param_l1size --l1i_assoc=$param_l1assoc
--l1i_latency=$param_l1latency --l1d_size=$param_l1size
--l1d_assoc=$param_l1assoc --l1d_latency=$param_l1latency
--l2_size=$param_l2size --l2_assoc=$param_l2assoc
--l2_latency=$param_l2latency --cacheline_size=$param_l1line
--clock=$param_freq -b $param_bench");
...}
...
```

Algorithm 1: Stretch of the automated script.

3.7 SIMPLESCALAR

SimpleScalar [Austin and Burger, 1997; Burger and Austin, 1997] is an open source computer architecture simulator developed at the University of Wisconsin Madison. It simulates a series of Instruction Set Architecture (ISA), its original purpose was to be used to demonstrate how differences among architectures would affect the performance of a specific application running on top of architecture A would be better than architecture B without necessarily building architecture A or B.

SimpleScalar is a set of tools that model a virtual computer system with CPU and Memory Hierarchy. Using the SimpleScalar tools, users can build modeling applications that simulate real programs running on a range of modern processors architectures. The tool set includes sample simulators ranging from a fast functional simulator to a detailed, dynamically scheduled processor model that supports non-blocking caches, speculative execution, and state-of-the-art branch prediction. In addition to simulators, the SimpleScalar tool set includes performance visualization tools, statistical analysis resources, and debug and verification infrastructure.

The SimpleScalar tools are used widely for research and instruction, according with SimpleScalar site, since 2000 around one third of all papers published in top computer architecture conferences used the SimpleScalar tools to evaluate their designs.

<http://www.simplescalar.com/>

The tool set itself consists of a collection of architectural simulators that emulates microprocessor ISA at different levels of detail. The available tools in SimpleScalar toolset are:

- SIM-FAST: Fast instruction interpreter, optimized for speed. This simulator does not account for the behavior of pipelines, caches, or any other part of the microarchitecture. It performs only functional simulation using in-order execution of the instructions (i.e., they are executed in the order they appear in the program);
- SIM-SAFE: Slightly slower instruction interpreter, as it checks for memory alignment and memory access permission on all memory operations. This simulator can be used if the simulated program causes sim-fast to crash without explanation;
- SIM-PROFILE: Instruction interpreter and profiler. This simulator keeps track of and reports dynamic instruction counts, instruction class counts, usage of address modes, and profiles of the text and data segments;
- SIM-CACHE: Memory system simulator. This simulator can emulate a system with multiple levels of instruction and data caches, each of which can be configured for different sizes and organizations. This simulator is ideal for fast cache simulation if the effect of cache performance on execution time is not needed;
- SIM-BPRED: Branch predictor simulator. This tool can simulate difference branch prediction schemes and reports results such as prediction hit and miss rates. Like sim-cache, this does not simulate accurately the effect of branch prediction on execution time;
- SIM-OUTORDER: Detailed microarchitectural simulator. This tool models in detail and out-of-order microprocessor with all of the bells and whistles, including branch prediction, caches, and external memory. This simulator is highly parameterized and can emulate machines of varying numbers of execution units.

For our purposes we began our experiments, using the SimpleScalar, later switching to the GEM5, after had verified that we obtained similar results to the SimpleScalar with GEM5, this experiments was conducted by my colleague Raphael Brum, to evaluate SimpleScalar and GEM5.

3.8 GEM5

The Gem5 [Binkert et al., 2006] is a simulator, resulting from the joining of simulators M5 [Binkert et al., 2011] and GEMS [Martin et al., 2005]. Currently, Gem5 supports most commercial ISAs (ARM, ALPHA, MIPS, Power, SPARC, and x86), including booting Linux on three of them (ARM, ALPHA, and x86).

3.8.1 *Simulation Capabilities*

The Gem5 simulator has a wide range of simulation capabilities, ranging from the selection of ISA, CPU model and coherence protocol to the instantiation of interconnection networks and devices. This section describes some of the different options available in these categories.

3.8.2 *ISAs*

The Gem5 simulator currently supports ISAs including Alpha, ARM, MIPS, Power, SPARC, and x86. However, not all combinations of ISAs and other components are currently known to work.

3.8.3 *Execution Modes*

The Gem5 simulator can operate in two modes: System-call Emulation (SE) or FullSystem (FS). In SE mode, Gem5 emulates most common system calls. When the program executes a system call, Gem5 emulates it, often by passing it to the host operating system. Currently, there is no thread scheduler in SE mode, so threads must be statically mapped to cores, limiting its use with multi-threaded applications. In FS mode, Gem5 simulates a bare-metal environment capable of running an OS. It includes support for interrupts, exceptions, privilege levels and I/O devices. Compared to SE mode, FS mode improves both the simulation accuracy and variety of workloads that Gem5 can execute.

3.8.4 *CPU Models*

The Gem5 simulator supports four different Central Processing Unit (CPU) models: AtomicSimple, TimingSimple, In-Order, and O3. AtomicSimple and TimingSimple are non-pipelined CPU models that attempt to fetch, decode, execute and commit a single instruction on every cycle. The AtomicSimple CPU is a minimal, single IPC! (IPC!) CPU.

The InOrder model is an **execute-in-execute** CPU model emphasizing instruction timing and simulation accuracy with an in-order pipeline. InOrder can be configured to model different numbers of pipeline stages, issue width, and numbers of hardware threads. Finally, the O3 CPU is a pipelined, out-of-order model that simulates depen-

dencies between instructions, functional units, memory accesses, and pipeline stages. Parameterizable pipeline resources, such as the load/store queue and reorder buffer, allow O3 to simulate superscalar architectures and **CPUs!** (**CPUs!**) with multiple hardware threads (Simultaneous Multi-Threading (**SMT**)).

3.8.5 *Interconnection Networks*

The Ruby memory model supports arrays of interconnections topologies and includes two different network models. In essence, Ruby can create any arbitrary topology as long as it is composed of point-to-point links. After Ruby creates the links and routing tables, it can implement the resulting network in one of two ways.

3.8.6 *Devices*

The Gem5 simulator supports I/O devices, ranging from timers to complex network interface controllers. Base classes are available that encapsulates common device interfaces such as PCI. Currently implemented models includes NICs, an IDE controller, a frame buffer, DMA engines, UARTs, and interrupt controllers.

In this chapter, we presented the problematic dealt by this thesis, delineating in which stages of the memory hierarchy we have worked. Also, we provide the analytical flow we establish to evaluate different memory banks and their impact into the memory hierarchy. Furthermore, we provided details about the tools we employed to make such evaluations, and how do we combine these different tools, their strengths, to obtain the conclusions and results presented here.

Part IV

ANALYSES OF EMBEDDED MEMORY HIERARCHY

This chapter will provide results to the assertions we made as viable solutions to employ MRAM into memory hierarchy. In addition, will demonstrate how to use a simple technique to mitigate the MRAM delay, only taking advantage of density. In addition, will be presented a more invasive approach, proposing a change into the basic memory bank to address such high latency of MRAM. Furthermore presents how the intrinsic analyses are employed at the architectural analyses to calibrate it. Describes how, using standards EDA tools, we demonstrate a more precise method to extract electrical results from a SoC based into the synthesis results combining a characterized MRAM library in Liberty. For last, it will present the Composite banks as an alternative that has the best trade-of among all constraints compared to memory banks to integrate MRAM into the memory hierarchy.

4

ANALYSES OF EMBEDDED MEMORY HIERARCHY

*The idea is to try to give all the information to help others
to judge the value of your contribution;
not just the information that leads to judgment
in one particular direction or another.
— Richard Phillips Feynman*

In this chapter will be presented the results regarding the methodology proposed in [Section 3.5](#) and [Section 3.5.3](#). In the [Section 4.2](#) is presented an evaluation at system level, based on the results obtained in [Section 3.5.3](#). Using these values, the latency of the memory hierarchy is calibrated into the Gem5, producing the results discussed in [Section 4.2](#).

Furthermore, three main approaches used to evaluate the memory banks will be described. The approach described in [Section 4.1](#) will demonstrate the initial experiments performed comparing memory banks for cache with a relation of 4-to-1. Assuming that [MRAM](#) has an area four times bigger than the area of the [SRAM](#) based on [[Diao et al., 2006](#); [Huai et al., 2008](#); [Torres et al., 2013](#); [Zhao et al., 2011, 2012](#)], we evaluated, using the SimpleScalar, which would be the impacts on the memory hierarchy performance replacing a [SRAM](#) of size S by a [MRAM](#) of size $4S$, with its particularities of latency being accounted for. This was assumed as a microprocessor without L2 cache, only L1 and the memory banks replaced at L1 level. It is exactly as it is suggested in [Section 2.6](#) of mechanisms for cache optimization: to compensate the [MRAM](#) delays, intrinsic to the [MTJ](#) technology, regarding the necessary current and period of the applied current to modify the anisotropy, we take advantage of the higher density, to mitigate the cache misses. The idea in this scenario is evaluate the Microcontrollers application space, it describes the kind of microprocessor designed with these characteristics.

In the [Section 3.5.3](#), we discuss and depict the results of the so called intrinsic analyses. By intrinsic analyses, we mean the analysis of a 1-to-1 bank. In this case, we compare memory banks of same capacity, for cache L2, using the proposed methodology in [Section 3.5](#) and [Section 3.6](#). The key point here is: latency at the level L2 of the memory hierarchy is less critical than at L1, so in theory we can keep the near same performance as a microprocessor with [SRAM](#), by just employing [MRAM](#) on L2 cache, but obtaining advantage due to power leakage reduction, and paying for the high latency to write into the [MRAM](#). This particular scenario is also subdivided into two memory technology approaches the High Performance ([HP](#)) and the Low Power Performance ([LOP](#)) for the memory banks, so a system with [HP](#) and the same one with [LOP](#) memory banks. The intention here is have a more detailed and feasible analysis, without helping the case for [MRAM](#), trying to place a fair comparison. Also, the data is relative to the data

for 45nm and 28nm provided in [Section 3.6](#). This scenario is more devoted to evaluate embedded SoC, like embedded microprocessors with computing performance in mind.

In the [Section 4.2](#), performed previously to the [Section 3.5.3](#), the idea was explore the possibility of having a gigantic cache L2 like 4GB, with same associativity. With latency from utopic 1ns, 6ns to the more realistic 12ns and the realistic and pessimistic scenario of 100ns of latency into the L2 cache. The main observation in this case was that latency at the L2 cache level doesn't have a major role into the performance. The most important conclusion was that the associativity and the line size play a predominant role than latency at L2.

4.1 EMBEDDED MRAM FOR PROCESSOR APPLICATIONS

In this section is introduced, the space scenario of [MRAM](#) in memory hierarchy applied for microcontrollers. Specifically regarding the L1 cache aspects, and mainly the influence of density versus latency are presented in [Section 4.1.1](#).

4.1.1 *MRAM applications for the processor memory hierarchy*

[MRAM](#) can be used for building the processor memory hierarchy. As mentioned earlier, even though the [MRAM](#) density is between two and four times higher than that of [SRAM](#)'s, its access time is between three and ten times higher. When using the same silicon area for implementing both variations, the former might compensate the latter.

In [\[Sun et al., 2009b\]](#), for example, a 2MB L2 [SRAM](#) Cache was replaced with an 8MB L2 [MRAM](#) Cache, using roughly the same silicon fingerprint. In their particular case, the increase on the cache size was not enough to compensate the penalty due to the cache access delay. By employing write buffers and a novel cache access policy, they managed to achieve similar performance while reducing the power consumption by almost 74%. They also present a hybrid [MRAM/SRAM](#) cache organization, having 31 sets implemented in [MRAM](#) and 1 set implemented in [SRAM](#). The write-intensive data is kept in the [SRAM](#) part, in order to mitigate the higher write delay. A method for determining which data is suitable for being placed in the [SRAM](#) set is also discussed.

We propose here a preliminary study on the impact of cache delays on the processor's performance, based on the SimpleScalar simulator [\[Burger and Austin, 1997\]](#) and the set of benchmarks Mediabench 1 [\[Lee et al., 1997\]](#). This simulator implements a MIPS-like architecture, and the memory hierarchy can be fully configured in terms of capacity, access speed and access policies. Our goal was to determine whether replacing L1 [SRAM](#) caches by L1 [MRAM](#) caches, while keeping the same silicon fingerprint, is worthwhile.

SimpleScalar was configured to mimic a processor designed for embedded applications, such as LEON3 [\[Gaisler, 2010\]](#). The baseline configuration is described in [Table 4.1](#). It consists of a single processor having a single cache level and a large external memory, assumption that can be considered for many systems.

Table 4.1: SimpleScalar baseline configuration used in all experiments

Option	Value	Meaning	Configuration
-cache:memlat 1st	1000	Ext. memory latency - 1 st word	1000 cycles
-cache:memlat burst	10	Ext. memory - burst	10 cycles per word
-cache:dl2	none	L2 Data Cache Parameters	No L2 Cache
-cache:il2	none	L2 Instr. Cache Parameters	No L2 Cache
-res:ialu	1	# of integer ALUs	1 integer ALU
-res:falu	1	# of floating-point ALUs	1 FP ALU
-res:imult	1	# of integer multipliers	1 Multiplier
-fetch:ifqsize	1	Instruction Fetch (IF) Queue Size	1 IF per cycle
-fetch:mplat	1	Branch misprediction latency	1 extra cycle
-cache:dl1 linesize	32	L1 Data Cache Line Size	32 bits per cache line
-cache:dl1 policy	1	L1 Data Cache Replacement Policy	Least Recently Used (LRU)
-cache:il1 linesize	32	L1 Instr. Cache Line Size	32 bits per cache line
-cache:il1 policy	1	L1 Instr. Replacement Policy	Least Recently Used (LRU)

Table 4.2: SimpleScalar configuration used in Figures 4.1-4.3

Option	Value	Meaning	Configuration
-cache:dl1 assoc	2	L1 Data Cache Line Size	2-way associative
-cache:dl1lat	3	L1 Data Cache Access Latency	SRAM: 1 cycle, MRAM: 3 cycles
-cache:il1 assoc	2	L1 Instr. Cache Line Size	2-way associative
-cache:il1lat	3	L1 Instr. Cache Access Latency	SRAM: 1 cycle, MRAM: 3 cycles

Differently from our previous work in [Zhao et al., 2011], we assumed that the MRAM density is four times the SRAM's [Mackay, 2011]. We are then comparing, for instance, a 4 KB SRAM-based cache with a 16KB MRAM-based cache.

For this set of experiments, we assumed a latency of 3 clock cycles during each cache access. It means that the processor will stall upon each cache request, waiting for the data to become available. We also assumed a latency of 1000 cycles for the external memory to make the first word available, and 10 cycles for each subsequent word while doing burst reading [JC-42.3, 2008].

In Figure 4.1, we compare a 1KB SRAM cache with a 4KB MRAM-based memory. We can see that the increase in capacity can easily compensate for the delay in such a case.

In the same manner, as shown in Figure 4.2, where a 128 KB SRAM cache is compared with its 512 KB MRAM counterpart, the latter shows comparable performance to the smaller, yet faster SRAM.

In order to generalize this conclusion, let us then define the CPI penalty as the increase in the CPI caused by replacing an SRAM cache with an MRAM cache using the same silicon area, as follows:

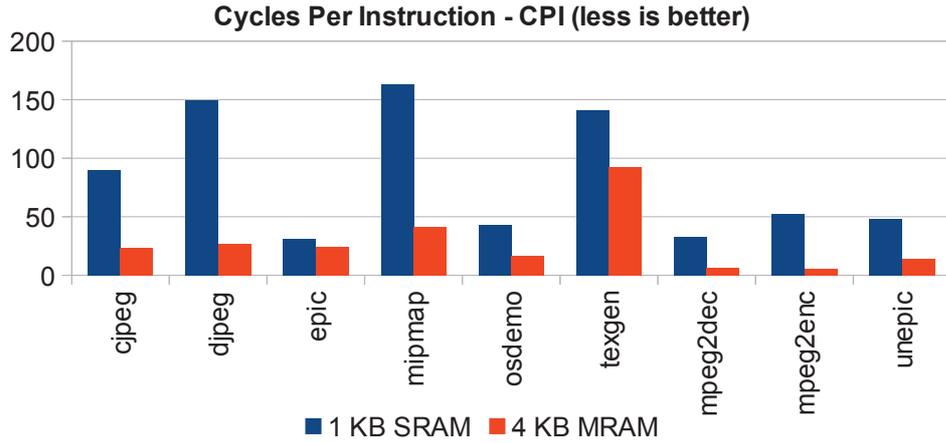


Figure 4.1: Overview of the processor performance using low-capacity L1 caches.

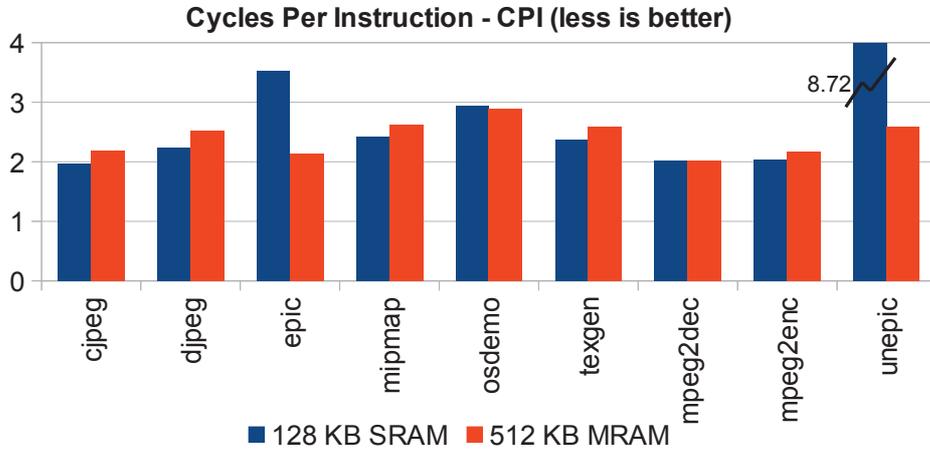


Figure 4.2: Overview of the processor performance using high-capacity L1 caches.

$$CPI_{\text{penalty}} = \frac{CPI_{\text{MRAM}}}{CPI_{\text{SRAM}}} - 1 \tag{4.1}$$

Based on the CPI_{penalty} , in Figure 4.3, the best case, the worst-case and the average performance over the benchmark set are shown as a function of the cache capacity. Given our assumptions are valid, MRAM does present a CPI gain rather than a CPI penalty for most cases. Once the cache capacity is large enough to contain the whole benchmark data, the CPI gain turns into a penalty that can no longer be compensated if no specific technique is employed.

4.1.2 In-Depth Analysis: Case Study CJPEG

The objective for this experiments were to focus in one of the benchmarks and extrapolate some of the possibilities tuning associativity and size with fixed latencies. Comparing the results and depicting it for analyses of memory banks with sizes up to 256 KB.

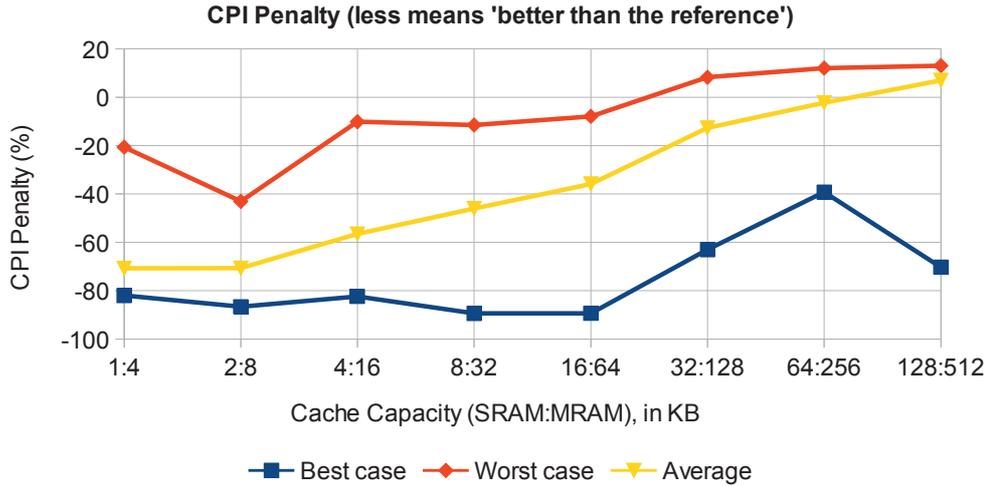


Figure 4.3: Overview of CPI Penalty: best-case, worst-case and average of the Mediabench benchmarks' performance.

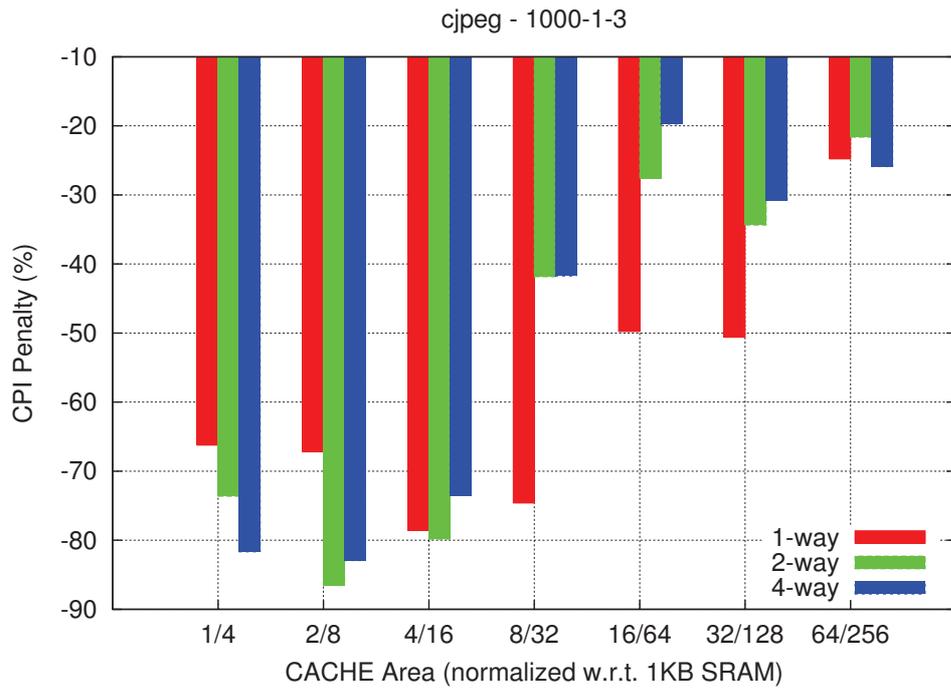
Based into the data analysis for the set of benchmarks, one was selected for an in-depth analysis, the CJPEG. The choice for this particular algorithm is because it is a data-driven benchmark performing a substantial number of accesses to the cache memories.

Also for this particular benchmark, we performed the tests for different associativities 1, 2 and 4-way caches. We also adjusted the delays of the cache latency, focusing on three and ten for the MRAM latency. Furthermore, in this particular case was assumed as our reference that the SRAM has one cycle delay [Alvarez et al., 2004; Boschma et al., 2004; Hennessy and Patterson, 2007a; Nambu et al., 1998; Patterson and Hennessy, 2005; Powell et al., 2001].

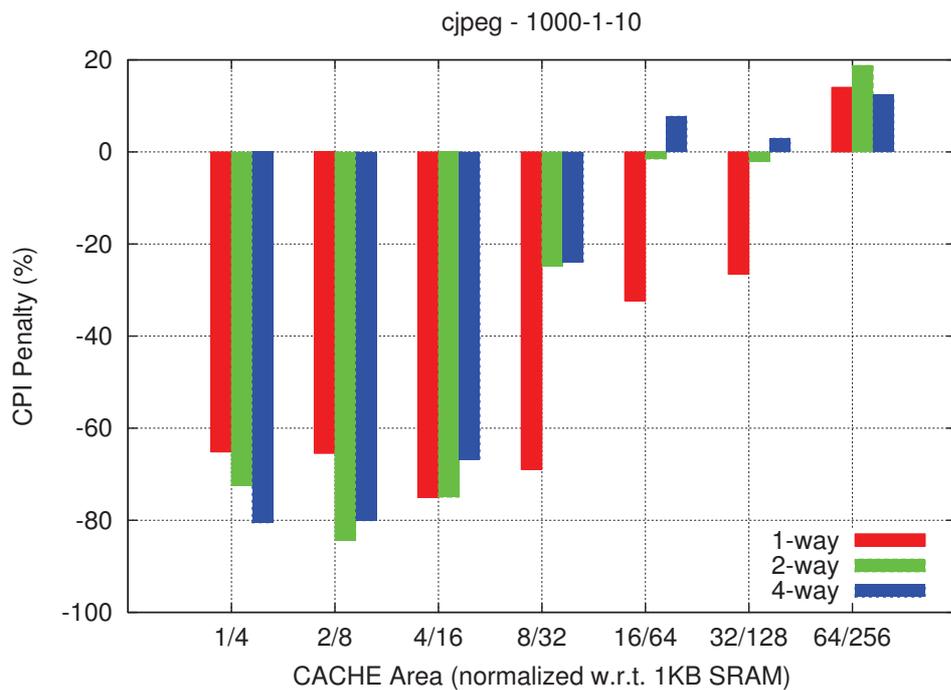
The Figure 4.4 and Figure 4.5 are depicting the two cases regarding delay of 1000 clock cycles for the first word been retrieved from the external memory [Gharachorloo et al., 4; JC-42.3, 2008] and one cycle between subsequent words in burst. A latency of 1000 cycles is assumed to be more conservative, regarding the external memory access and the circuitry in the path to access it [Gutierrez et al., 2011].

This case study assumes that for the same silicon fingerprint the MRAM integration density is $\times 4$ in comparison to SRAM (for instance for 1KB of SRAM, meaning it is possible to integrate in the same silicon area a 4KB MRAM memory), considering also as parameter the associativity available to access to the cache memory (1, 2 and 4 way). All the results are normalized by the baseline SRAM-L1 of 1KB, delay of 1 cycle. This way we compare SRAM [1, 2, 4, 8, 16, 32, 64]KB with MRAM [4, 16, 32, 64, 128, 256]KB. The comparisons are based into the CPI penalty, and CPI penalty is defined as specified in Equation 4.1.

Observing the Figure 4.4a and Figure 4.4b, it is possible to notice that an increase of seven cycles into the cache latency was necessary to affect the MRAM cache performance for sizes larger than 16 KB. We also note that the 1-way cache kept steady for sizes 16 and 32 KB despite the MRAM cache delays, this result is widely described into [Hennessy and Patterson, 2007a]. Similar results are observed in Figure 4.5a and Figure 4.5b, for a different MRAM cache parameters, also in this case a delay of ten cycles between subsequent words are assumed for read burst from the main memory.



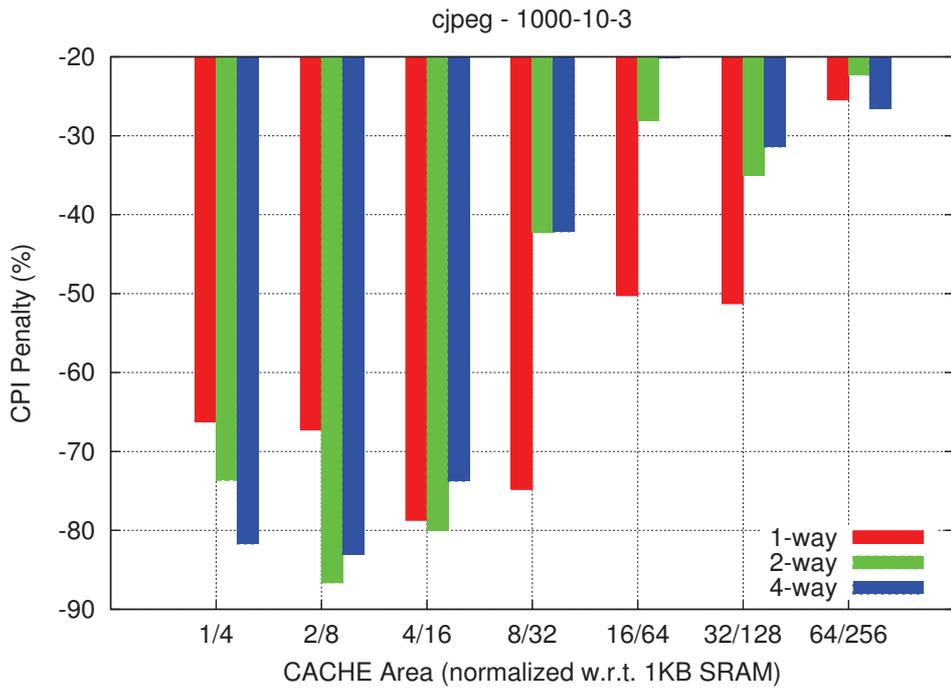
(a) 1000 cycles for the first word, than 1 for the following, 3 cycles cache delay



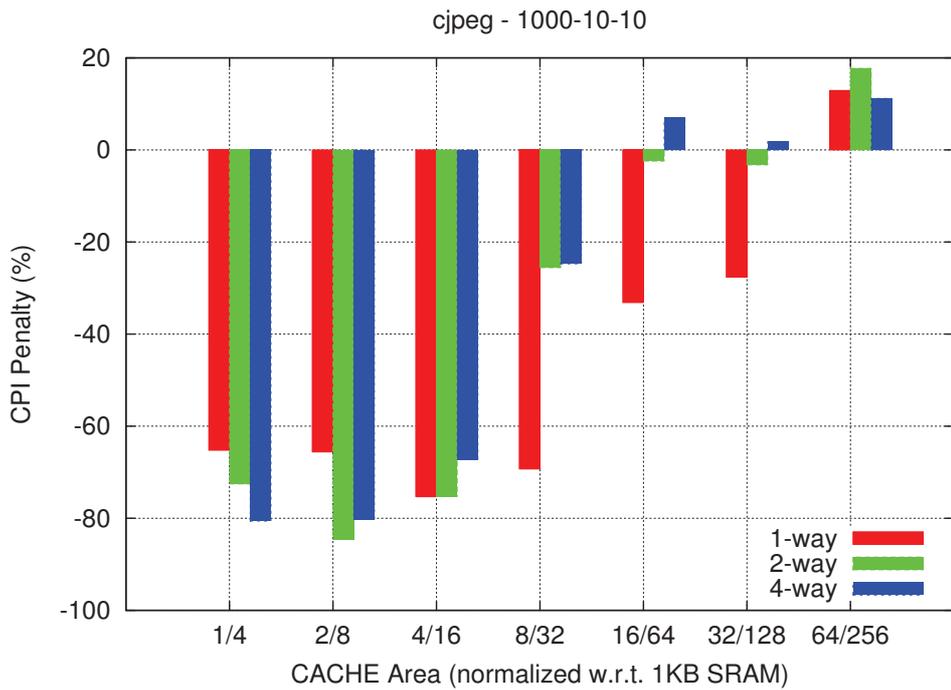
(b) 1000 cycles for the first word, than 1 for the following, 10 cycles cache delay

Figure 4.4: Simulation results for 1000 cycles and 1 cycle of delay for each subsequent word in burst mode.

So the strategy of a n-way cache is interesting for caches larger than 16 KB. The **SRAM** despite the density of 4 times smaller than the **MRAM** (for the same silicon area) outperforms the **MRAM** for sizes larger than 32 KB.



(a) 1000 cycles for the first word, than 10 for the following, 3 cycles cache delay



(b) 1000 cycles for the first word, than 10 for the following, 10 cycles cache delay

Figure 4.5: Simulation results for 1000 cycles and 10 cycles of delay for each subsequent word in burst mode.

The cache sizes combined with the associativity have a relevant impact on the cache miss ratio. For this reason, MRAM obtained a better result, despite the higher access de-

lay. But for caches larger than 16KB we should use an additional mechanism to mitigate the MRAM delay, assuming the worst case scenario of ten cycles of cache latency.

So, when using an MRAM L1 cache in a microprocessor, the higher density mitigates the lower delay up to 64 KB in this particular experiment. For L1 caches beyond this capacity, the MRAM density itself is not enough to mitigate the delay.

For these cases one possible solution could be the usage of write-buffers, or a MRAM cache working in a higher frequency and phase-aligned. Another approach would be a hybrid MSRAME cell, for L1 cache, for L2 caches the access delay of the current state of the technology are not an issue, only adopting a L2 MRAM it already cuts the leakage current in 70% for a SoC as described in [Sun et al., 2009b].

4.2 USING GEM5 TO EVALUATE THE IMPLICATIONS INTO MEMORY HIERARCHY OF L2 CACHE BANKS OF SIGNIFICANT SIZES FROM 2GB DOWN TO 256KB.

In this specific experiment, the objective was evaluating the outcome of a system with a high amount of embedded memory, in particular at CACHE L2 level. Given the limitations with SimpleScalar and processing time to simulate, also the difficulty to simulate a complete Operating System on top of the system using SimpleScalar we conducted the experiments using the Gem5, since it is easier and faster compared to SimpleScalar. A system was evaluated assuming only the L2 varying, the remaining of the system was kept unchanged. The parameters evaluated were size and the range was: 2GB, 1GB, 512MB, 256MB, 128MB, 64MB, 32MB, 16MB, 8MB, 4MB, 2MB, 1MB, 512KB, 256KB and 128KB. For all the results, the cache L2 was configured assuming 8-way associativity for all memory sizes in this experiment. We also played with the L2 latency, as 1ns, 8ns, 12ns (SRAM average) and worst case scenario 100ns. When this experiment was performed, wasn't knew that an 8MB STT-MRAM 8 – way has an average of 70ns in latency, as later verified and discussed in Section 3.5.3.

The system had a CACHE L1-I/D of 64KB, DDR of 256MB, running a Linux OS compiled for ARM v7 ISA, on top of the Gem5.

Also, in this case, we used the MPEG2 encoder and decoder as Benchmark for the measurements and evaluations depicted in Figure 4.6, Figure 4.7, Figure 4.8, Figure 4.9 and Figure 4.10.

Observing the Figure 4.6 and Figure 4.7 we can conclude that the influence of latency in the Miss latency for both encoder and decoder. For this specific algorithm size of CACHE is more critic than the latency, e.g., at 1MB from 1ns to 12ns the % Miss Rate is nearly the same, increasing at 100ns. In both cases the percentage is lower than 1%. It is visible that the size of CACHE L2 has much more influence on performance them latency

Observing the Figure 4.8, the overall Miss rate is visible that the memory size has the major role in performance regarding L2 CACHE, in fact the latency has not a major role into the performance.

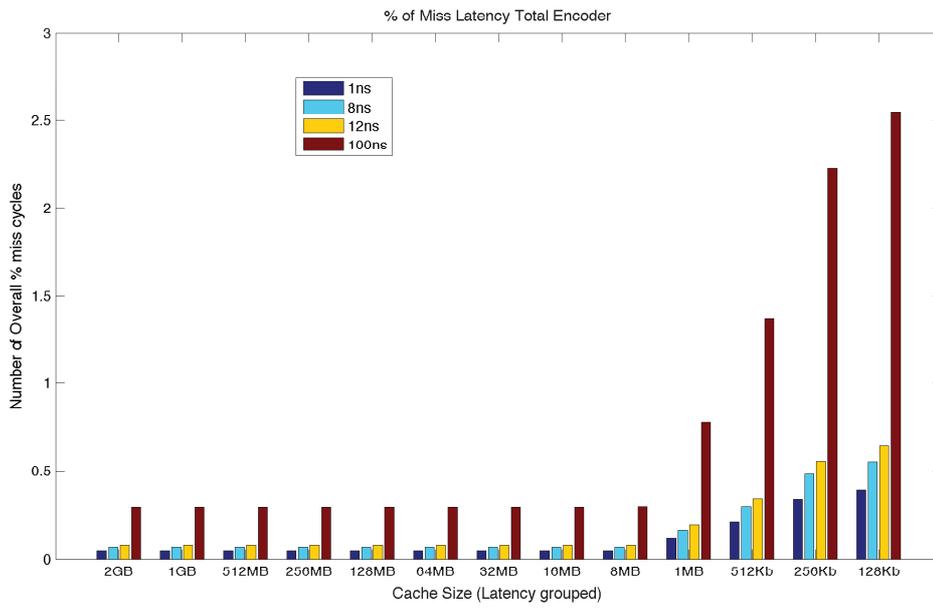


Figure 4.6: % Miss Latency Encoder.

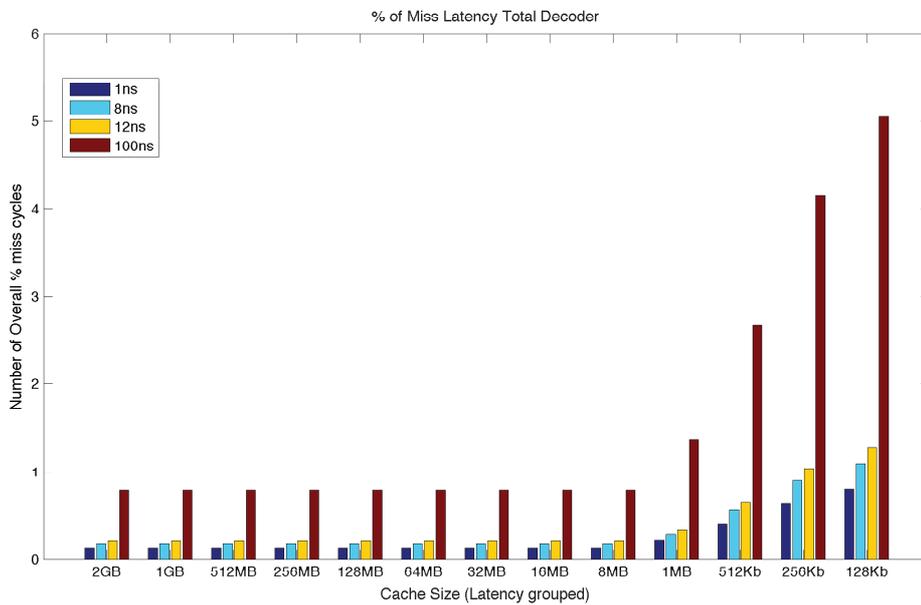


Figure 4.7: % Miss Latency Decoder.

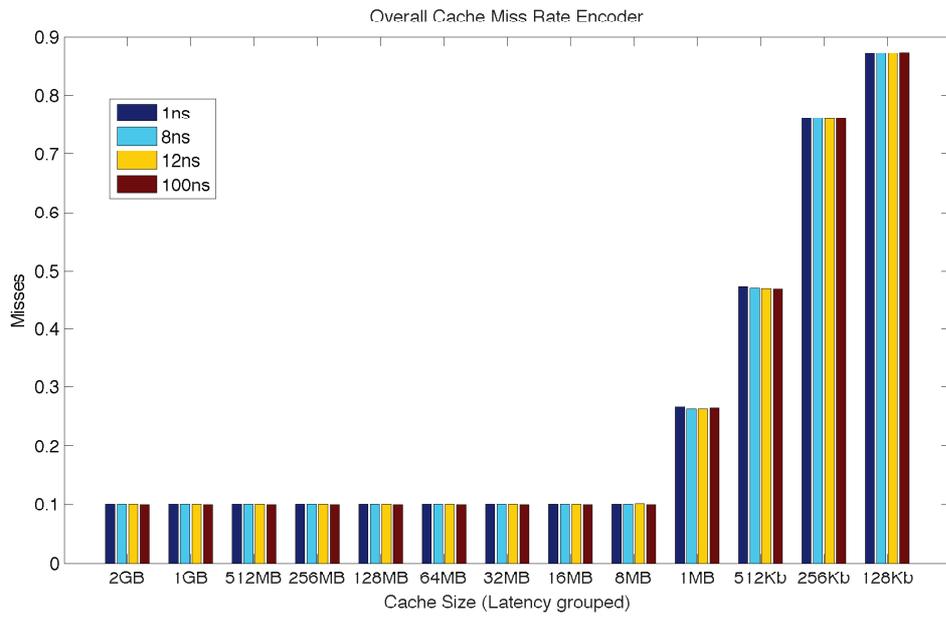


Figure 4.8: Overall Miss Rate Encoder.

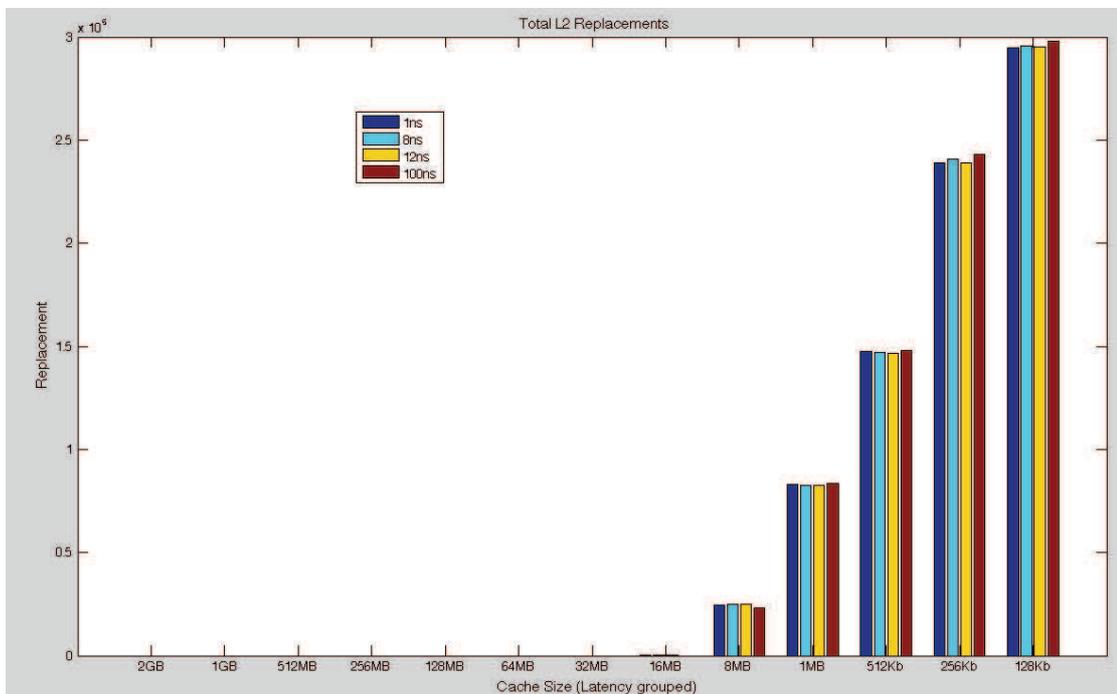


Figure 4.9: encoder L2 replacements.

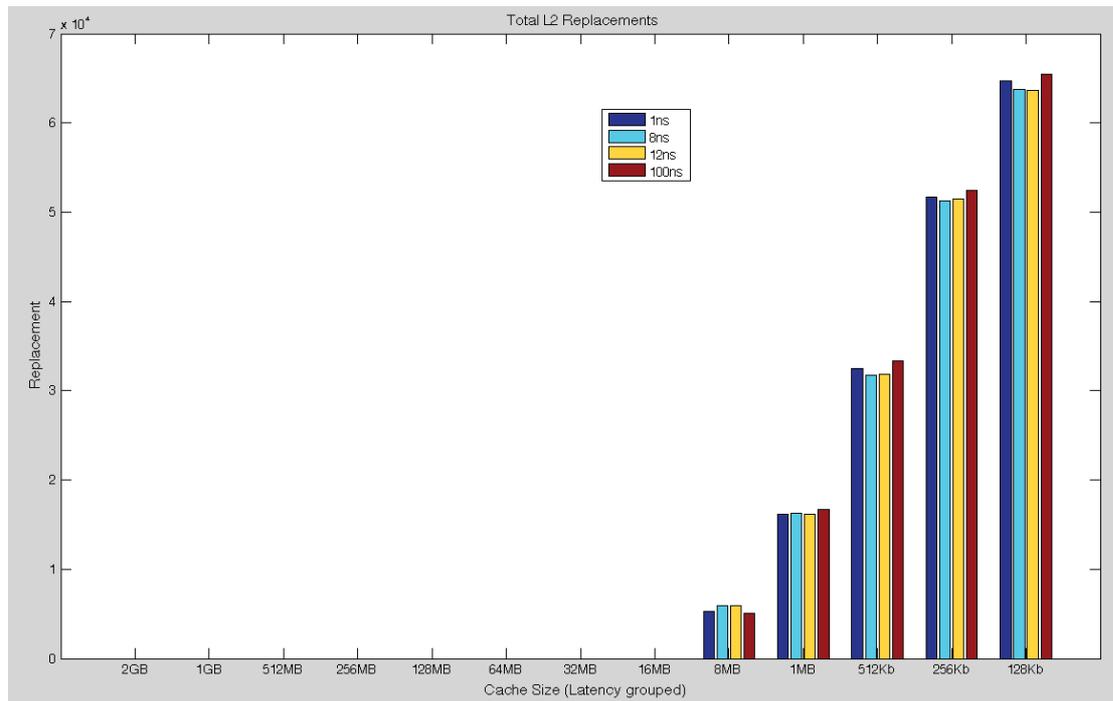


Figure 4.10: decoder L2 replacements.

The expected with this experiment, was to observe, mainly, how the cache L2 sizing affects the overall performance of the system and how the latency combined with the size, how the interaction of both affects the system performance.

At this point we have a rough idea that independently of the bank size in the L2-CACHE, the latency does not have a major effect to the overall performance. But the size plays an important role. It is known that more cache you have, better the performance [Patterson and L Hennessy, 2012]. A new fact eluded was related to the latency. It is supposed, in general, that lower the latency, higher the performance. But in the experiments demonstrated in Section 4.2 and Section 4.2.1, the obtained result was that: the latency of the L2 CACHE memory banks has a minimal, not to say almost null, effect into the overall performance of a system.

In this sense we evolved to the next stage of evaluations described in Section 4.2.1.

4.2.1 *Recalibrating the Gem5 using the NVSim bank details.*

Initially, Gem5 was used based on empirical assumptions. By empirical, it means we fully believed that the memory banks had a well-known pattern into the timeframe and power. Instead, they had a well-unknown pattern. Minimal changes into the memory bank like associativity and size of memory matrix-cell would affect other elements, like the sense amplifiers, by consequence affecting the TAG array size. In the same chain, it would affect the metal tracks for routing, affecting capacitance and resistance. Finally, all of this would affect the overall performance, latency, area and power characteristics of the memory bank.

So, based on observation, it was noticed that the best approach would be not make the power analysis of the system based on a possible memory bank time frame. Instead we generate the memory bank for a NVM specific technology, and re-calibrate the time frame of the Gem5 simulator, constrained by the memory banks generation. This way we obtained a precise unbiased model of a system performance, according with the memory technology used to simulate the system, instead of empirical presumptions.

The results are contained in Table 4.3. The last column is the relation, in percentage, of how much the MRAM technology impacts of each aspect of the system according with the ISA simulation. These results were obtained with the Gem5 calibrated using the details of the memory banks described in Table 4.4.

Table 4.3: The Simulation Results of x264 encoder, for same system, only with different L2 CACHE memory bank, and Gem5 re-calibrated with NVSim latency details of the memory banks.

Field	SRAM	MRAM	Impact(ratio)
sim_seconds	16.222047	17.133641	0.0561
sim_ticks	16222047298831	17133640621041	0.0561
final_tick	18853541587097	19984674238114	0.0599
sim_insts	8267542176	8267755985	2.5861E-005
sim_ops	8692429862	8692742955	0.00003
DDR.bytes_read::total	812829600	856487704	0.0537
L2.replacements	1021451	1025877	0.0043
L2.tagsinuse	65123.176249	65108.818324	-0.0002
L2.total_refs	16436632	17312300	0.0532
L2.sampled_refs	1021451	1025877	0.0043
L2.ReadReq_hits::total	18992618	20173225	0.0621
L2.Writeback_hits::writebacks	5879046	5944740	0.0111
L2.Writeback_hits::total	5879046	5944740	0.0111
L2.ReadExReq_hits::total	1098766	1109432	0.0097
L2.demand_hits::total	20091384	21282657	0.0592
L2.overall_hits::total	20091384	21282657	0.0592
L2.ReadReq_misses::total	263136	267399	0.0162
L2.UpgradeReq_misses::total	3622	3651	0.0080
L2.ReadExReq_misses::total	759467	759565	0.0001
L2.overall_misses::total	1022603	1026964	0.0042
L2.ReadReq_miss_latency::total	13603298609	34466314518	1.533
L2.UpgradeReq_miss_latency::total	273591	990683	2.6210
L2.ReadExReq_miss_latency::total	40300754564	98074660663	1.433
L2.demand_miss_latency::total	53904053173	132540975181	1.458
L2.overall_miss_latency::total	53904053173	132540975181	1.458
L2.ReadReq_accesses::total	19255754	20440624	0.0615
L2.Writeback_accesses::writebacks	5879046	5944740	0.0111
L2.Writeback_accesses::total	5879046	5944740	0.0111
L2.UpgradeReq_accesses::total	3633	3662	0.0079
L2.ReadExReq_accesses::total	1858233	1868997	0.0057
L2.demand_accesses::total	21113987	22309621	0.0566
L2.overall_accesses::total	21113987	22309621	0.05662
L2.ReadReq_miss_rate::total	0.013665	0.013082	-0.0426
L2.UpgradeReq_miss_rate::total	0.996972	0.996996	2.4072E-005
L2.ReadExReq_miss_rate::total	0.408704	0.406402	-0.0056
L2.demand_miss_rate::total	0.048432	0.046032	-0.0495
L2.overall_miss_rate::total	0.048432	0.046032	-0.0495
L2.writebacks::writebacks	813623	812790	-0.0010
L2.writebacks::total	813623	812790	-0.0010

Based on the results presented in [Table 4.3](#), we can depict the following elements to explain in fine grain, the important aspects of the MRAM impact into the memory hierarchy:

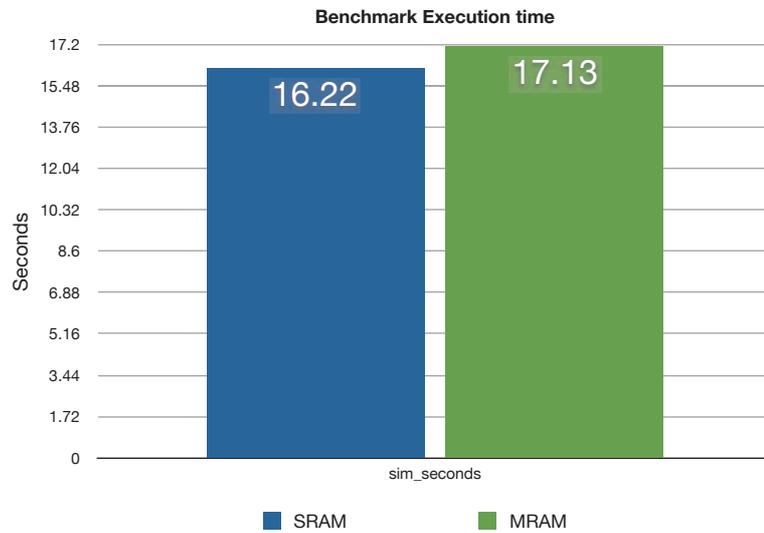


Figure 4.11: X.264 Execution time.

The simulations condition in this case, are similar to the conditions in [Section 4.2](#), a CACHE L1-I/D of 64KB, DDR of 256MB, running a Linux OS compiled for ARM v7 ISA, on top of the Gem5. In addition to that the timing for the L1 is assumed as 2ns, the details for the L2 can be found in [Table 4.4](#).

In [Figure 4.11](#) we can observe, as foremost importance, the exchange of the L2 memory bank based on SRAM by the MRAM counterpart, with higher latency, 276% higher for hit and 271% higher for response: the outcome was a increase of 0.05% on the total execution time. We passed from total execution time of 16.22 to 17.13, so 0.911 seconds of increase in total execution time. This is a significant impact in total performance for a systems engineer, meanwhile, for a user of a embedded system like a smartphone, the person will not even notice the increase in latency.

Table 4.4: Memory banks characteristics

field	SRAM	MRAM
Technology	45nm	45nm
Size	2MB	2MB
Associativity	8	8
Area		
Total Area	5.665mm ²	2.262mm ²
DATA Array Area	(1352.235μm x 3724.638μm)	(1470.019μm x 1274.689μm)
	5.037mm ²	1.874mm ²
TAG Array Area	(5426.310μm x 115.898μm)	(165.444μm x 2349.076μm)
	0.629mm ²	0.389mm ²
Timing		
Cache Hit Latency	18.795ns	70.806ns
Cache Miss Latency	2.968ns	66.038ns
Cache Write Latency	10.115ns	75.153ns
Power		
Hit Dynamic Energy ¹	1.076nJ	0.213nJ
Miss Dynamic Energy ²	1.076nJ	0.213nJ
Write Dynamic Energy ³	0.033nJ	0.223nJ
Total Leakage Power	1326.713mW	26.514mW
DATA Array Leakage Power	1180.597mW	24.283mW
TAG Array Leakage Power	146.116mW	2.231mW
hit(ns)	18.795	70.806
response (ns)	10.115	75.153

In Figure 4.12 is depicted the clock cycles added to the total amount of the processing for this benchmark.

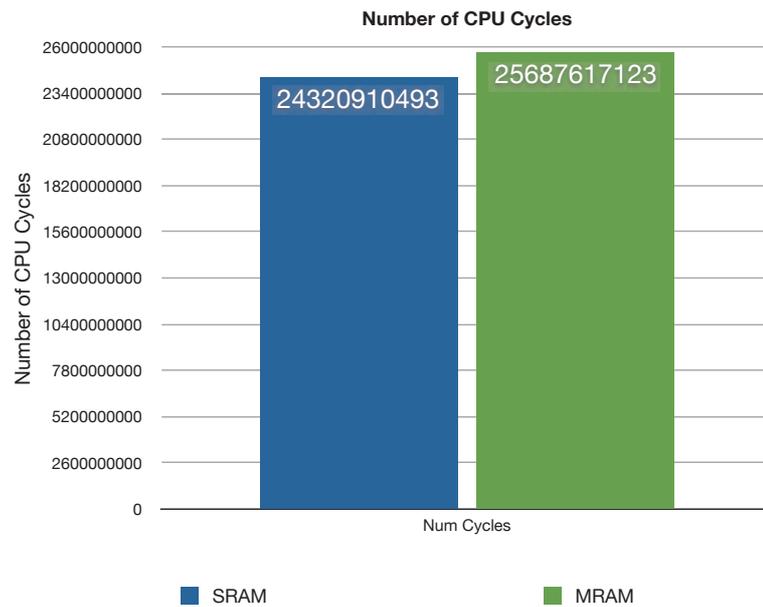


Figure 4.12: X.264 Clock cycles.

The latency is reflected as shown in Table 4.3, increasing the overall accesses to the L2 and the main memory, due to increase in latency of the memory banks. The reason for the increase in main memory access is the ARM v7 architecture MMU, which contains a policy of deadlines to avoid long periods of processor stalls. A deadline means that the system can wait to fetch the data from the memory. If the memory fails to supply the data in the specified time, it will fetch from main memory. This would explain the increase on the main memory access, using the same application *Ipsis litteris*, changing only the L2 CACHE memory bank.

The conclusion we obtain accessing all this data in this experiment is: even the increase of 270% in latency at CACHE L2, does not create a significant bottleneck in performance, so is safe to assert that is possible to replace a SRAM bank by an MRAM memory bank of same memory size. The resulting system has a smaller silicon footprint, due to the higher density of the MRAM technology of 4 to 8 times higher compared to the SRAM. The system has a smaller leakage using MRAM. The MRAM-LOP also provides the same performance of the SRAM-HP as depicted in Figure 3.13. This is important, since allow us built a memory bank adopting additional technology, for Low-Power operation, decreasing the total power consumption and yet keeping the performance of SRAM High-Performance found on server class systems.

In Figure 4.13, Figure 4.14, Figure 4.15, Figure 4.16, and Figure 4.17 is depicted the impacting factors of performance. These factors are also related with power. It's important to notice that the behavior depicted previously at Figure 4.11 will repeat across all over the parameters evaluations.

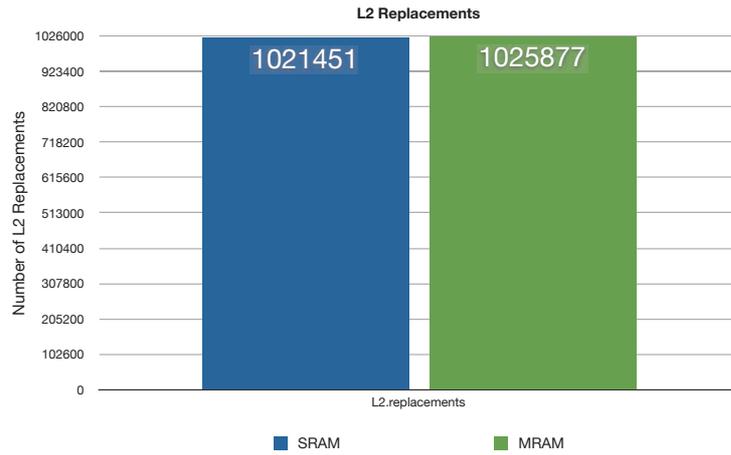


Figure 4.13: L2 Replacements.

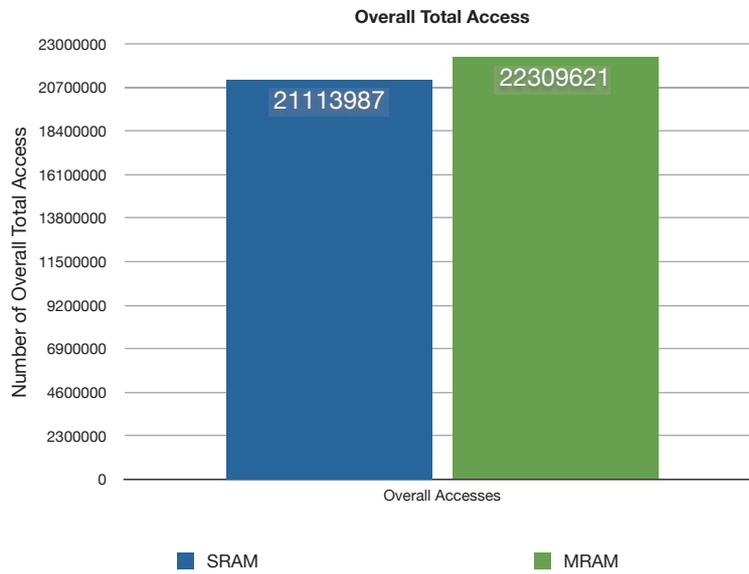


Figure 4.14: L2 Overall Access Total.

As is possible to observe in Figure 4.15, the increases in miss latency more than doubled, due to the difference in latency of the two memory banks. In the end the total processing time, observe Figure 4.11, have increased in less than a second, despite the average miss latency drastic increase.

So the conclusion is: the performance is not heavily affected by the increase in latency in more than 270% comparing SRAM vs MRAM. But the important factor to evaluate is the power consumption. The way the technology shift will affect the power consumption is next topic to evaluate.

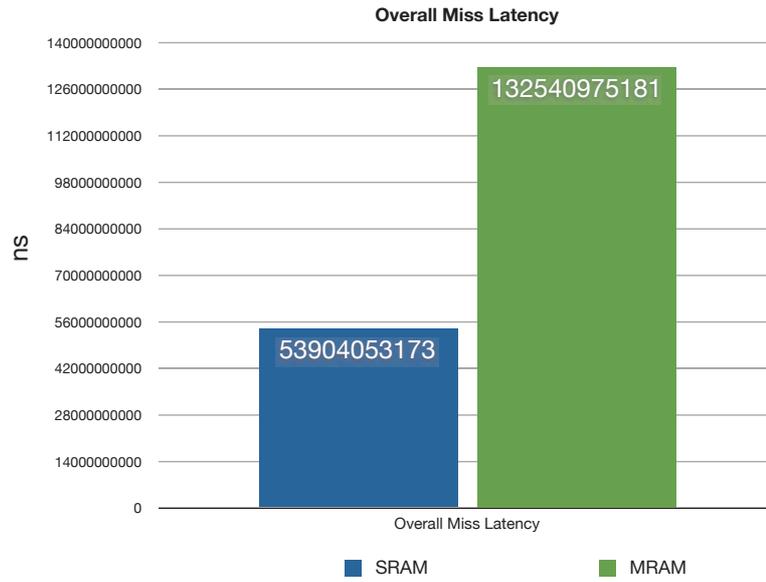


Figure 4.15: L2 Overall Miss Latency.

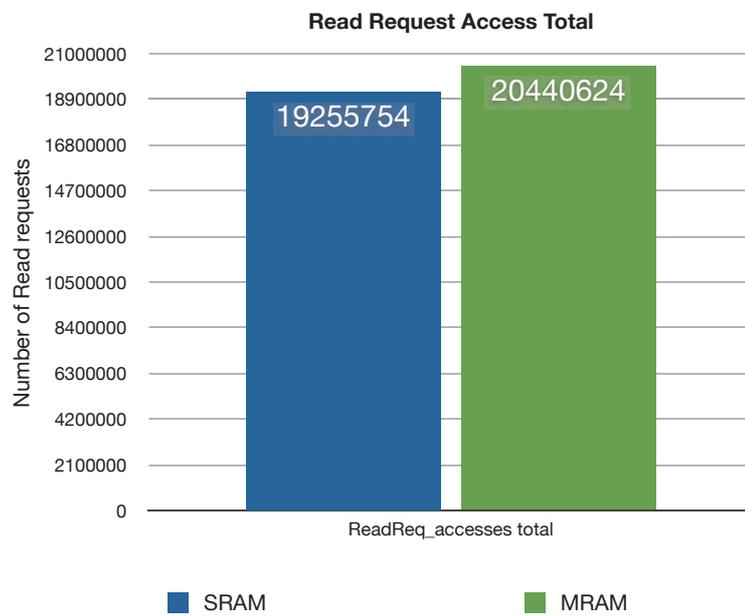


Figure 4.16: L2 ReadReq Access total.

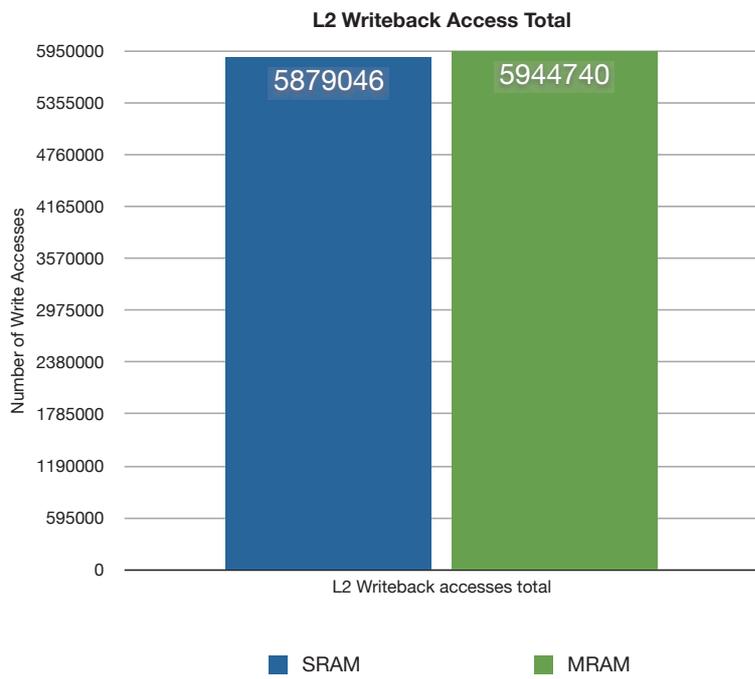


Figure 4.17: L2 Write Access Total.

4.2.2 Power Analysis based into the memory banks generated after the NVSim.

Based on the assessments detailed on [Table 4.4](#) and [Table 4.3](#), we can now cross the data from these two tables. This way, we can generate a profile of the power usage, having the number of accesses and knowing the power dissipated for each access. The simple manner is just multiplying the power for the number of accesses to obtain the total power for each characteristic.

Assuming the demonstrated in [Figure 4.11](#) is the total time of the execution of the X.264 encoder. Now lets take the leakage current denoted for each technology in [Table 4.4](#).

Joining the values we obtain a simple table like [Table 4.5](#). The Total energy presented there is the worst possible case scenario, assuming the memory banks are leaking during the entire process lifetime. This way we can obtain a rough idea of how much power the same system will consume at L2 of the memory hierarchy.

Table 4.5: Power Consumption estimation regarding the Leakage current.

	SRAM	MRAM
Execution Time:	16.222047	17.133641
Leakage Current:	3.386W	38.36mW
Total Static Energy (J):	54.927	0.657

Observing [Table 4.5](#) is already quite clear the first advantage of using MRAM to replace the SRAM at the L2 CACHE in the memory hierarchy. We can apply the same principle for all access from the microprocessor to the L2, this way we can obtain the total energetic cost for each operation into the memory. Lets repeat the process this time for the total Write Access as described in [Table 4.6](#).

Table 4.6: Power Consumption estimation number of WriteBack Total, and the data about the memory bank technology [Table 4.4](#).

	SRAM	MRAM
Write Back Total (Write):	5879046	5944740
Overall Access (Read):	21113987	22309621
Write Energy:	22.865nJ	170.657nJ
Read Energy:	957.780pJ	150.454pJ
Total Write Energy (nJ):	0.1344	1.0145
Total Read Energy (nJ):	0.0202	0.0033

Comparing both memory technologies for Write we can detail that the MRAM takes 7.54 times more energy than the SRAM for write operations, while on the read operations the SRAM takes 6.024 times more energy. The MRAM took a 1.25 or 25% more energy than the SRAM for overall operation for this specific application, the X.264 encoder. Analyzing the MRAM, only taking into account the dynamic energy is a mistake, especially because this is the biggest flaw of MRAM. Taking into account the leakage, the MRAM has the advantage. Nevertheless, the clear usage where MRAM would total excel is when executing algorithms Write Once Read Many (WORM) style, the kind of

expected scenario even from storage devices [Sion and Chen, 2012; Wang and Zheng, 2004].

For the current state of the technology, MRAM will consume more energy than SRAM for dynamic operation, as presented in Table 4.6 (at least for our particular case a X.264 encoder, available in all embedded devices of-the-shelf). The particular application that could take real advantage of the leakage reduction and reduced read power of a MRAM memory is the class WORM of applications and systems. Also, an architecture similar to the Kepler from NVIDIA that has a read-only cache would take advantage of MRAM memory cells [NVIDIA Corporation, a,b].

Embedded systems can benefit from the reduced leakage current, but the performance impact over power consumption is unclear and only real test in silicon will provide this answer. It is known that SRAM, at least in L3, is around 30% of the power consumption. So, assuming that L2 is at least ten percent less, and so on, we could extrapolate that the memory hierarchy roughly represents 40 ~ 50% of total power consumption in a SoC [Monchiero et al., 2008; Thoziyoor et al., 2008a]. Nevertheless, as will be seen in Section 4.4, the trend with MRAM is the power consumption shifts from the memory hierarchy (the CACHES) to the logic and interconnections of the circuit.

For the present state of the technology, unless a L3 cache is adopted, replacing the L2 by MRAM will not provide any real benefit, except of the leakage current. Regarding this is based on the X.264 Encoder only, for this kind of application in particular, the L2 technology replacement makes no difference at all, considering that the only evident benefits are the leakage current and reduced read power. Again, this reinforces the idea for WORM systems.

4.3 BENCHMARKS USED FOR EVALUATIONS

The conjectures, results and analyses presented in Section 4.2.1, were based on one specific benchmark. The chosen benchmark was the X.264. The X.264 is a free software for encoding video streams in the H.264/MPEG-4 AVC format.

Also, the idea was to stress the memory hierarchy, since it is a stream codec for high resolution videos. The particular encoder used has the following features:

- 8x8 and 4x4 adaptive spatial transform
- Adaptive B-frame placement
- B-frames as references / arbitrary frame order
- CAVLC/CABAC entropy coding
- Custom quantization matrices
- Intra: all macroblock types (16x16, 8x8, 4x4, and PCM with all predictions)
- Inter P: all partitions (from 16x16 down to 4x4)
- Inter B: partitions from 16x16 down to 8x8 (including skip/direct)
- Interlacing (MBAFF)
- Multiple reference frames

- Ratecontrol: constant quantizer, constant quality, single or multipass ABR, optional VBV
- Scenecut detection
- Spatial and temporal direct mode in B-frames, adaptive mode selection
- Parallel encoding on multiple CPUs
- Predictive lossless mode
- Psy optimizations for detail retention (adaptive quantization, psy-RD, psy-trellis)
- Zones for arbitrarily adjusting bitrate distribution

For our particular case we, used it in a single core at 1.5 GHz, 64 KB of L1 I/D, 2MB of L2, 256 MB of DDR2. It was launched on top of a Linux on a ARM ISA v7, with 8 parallel threads and 16 frames. It encodes a set of images that generates a 1 second video of 720p.

This benchmark was chosen with the intention of totally stress the CACHE L2 usage, since it is a streaming algorithm. Also, it has SIMD optimizations for ARM v7 architecture, the one implemented in Gem5.

For the experiments depicted in [Figure 4.6](#), [Figure 4.6](#), [Figure 4.8](#), [Figure 4.9](#) and [Figure 4.10](#), the used benchmark was the MPEG-2 encoder and decoder. The same experiment was performed using ADPCM, G721, G711, aLaw, μ Law, PCM, MPEG-2. Considering that the results suffer variations comparing among the benchmarks, was opted-out to keep the focus into the MPEG – 2, since it was stressing the L2 more than the others. Also, the spent time performing the simulations seems useless, regarding the other benchmarks.

The purpose of the experiments with the MPEG-2 was identify how the associativity and cache-L2 size would be affected, given certain combinations of delays and sizes.

4.4 COMPARING A MICROPROCESSOR MEMORY HIERARCHY, SYNTHESIZED WITH MRAM AND SRAM AT 28nm

The objective of this section is set in place all the conjectures already made, in a silicon level: with current available EDA tools, how to put in place a library of **MRAM** memory banks, and how to use it for physical analyses into synthesis level, bridging this way the gap between physical models, simulated models and a physical implementation. Therefore, it will be possible obtain more reliable information, instead of simply speculative results. Simulation results are a quick way to provide a general overview of possible scenarios and possible results. However, in the end, they are only simulative results. Consequently, the objective here is close this gap between the simulative speculation and the expected observed in circuit level, without the necessity of undergoing a circuit implementation for every design. Of course no model will ever be accurate without a integrated circuit implementation to close the final gap of an entire model.

For this particular experiment we adopted the technological library, a synthetic library provided by Synopsys, for 28nm **CMOS** technology. It contains the Phase Locked Loop (**PLL**), standard cells, **SRAM** memory banks, High-performance and Low-Power sets. So, you can even perform the low-power stage of a Design For Manufacturing (**DFM**)

flow. Based on this premisses, I modeled the [MTJ](#) and generated the memory models with NVSim. With the memory banks details provided by the NVSim, I generated a Liberty file containing the data extracted from the NVSim, to synthesize the design with my [MRAM](#) memory banks instead of the [SRAM](#). The base for my liberty file was the library for the [SRAM](#), where I replaced the Composite Current Source ([CCS](#)) matrices of the [SRAM](#), when necessary. Of course this is not a simple task, and the availability of better tools would be welcome. The result exposed here is not perfect yet, there is plenty of room for improvements. Remember that this discussion is based in a near twelve thousand lines file, that describes electrically and physically how this set of memory banks will compose the time framing computation of the synthesis.

The best manner to characterize this library to be error prone, would be using a SPICE model of the [MTJ](#), which I did not had available for 28nm, neither through my research network it was available in the moment I performed the experiment. The SPICE model is loaded into a characterization system such as ALTOS (Cadence package) Liberate MX from Cadence. With the SPICE model and the characterization tool, we provide a series of configurations like the operation conditions:

```
set_operating_condition -voltage $vdd -temp $temp
```

, than we provide the [CCS](#) template matrices for power, timing, setup and hold times. The ALTOS uses the SPICE model of a memory cell to characterize the electrical behavior of a memory bank and its matrices of memory cells.

With all this information in place, the characterization tool plus a few hours, days or weeks, according with the desired complexity and number of possible banks, into an High Performance Computing ([HPC](#)) machine, will generate the liberty files with all the [CCS](#) matrices embedded into it. The [CCS](#) is a troublesome feature to generate, but

once you have it, your synthesis reports results will provide a near exact to the final circuit results, at least for each IP block, regarding power and timing.

```

set pwd [exec pwd]

# Define operating conditions ##
set vdd 1.2
set gnd 0.0
set temp 25
set_operating_condition -voltage $vdd -temp $temp
...
# define templates
...
SRAM_delay_template
define_template -type constraint \
-index_1 { 0.010 0.050 0.200 0.400 1.000 } \
-index_2 { 0.010 0.050 0.200 0.400 1.000 } \
...
# define cell, pins, truth tables
set cell { sram }
set mxtables {}
lappend mxtables $pwd/tables/sram.tbl
...
# read spice
set spicein {}
lappend spicein $pwd/data/cl_models.sp
lappend spicein $pwd/data/sram.sp
read_spice $spicein

# clock prop
mx_set_clockprop {{ME enable CLK} {TME enable CLK}}
...
set characterization_simulator "hspice"
...
# specify models the memory should be characterized for
# partition (with fast spice) and characterize (with real spice)
...
# write models
write_library -overwrite sram.lib
foreach model $models {
    write_library -overwrite $model sram.${model}.lib
}

```

Algorithm 2: Fragment of a TCL file to demonstrate how to characterize a memory library using Liberate MX.

The target microprocessor used was the OpenRISC architecture and the implementation is the OR1200. The RTL is freely available. Replacing the memory banks is quite simple and add new memory technology is also a straight-forward process. This way we provide a unbiased manner to demonstrate how using a of-the-shelf Electronic De-

sign Automation (EDA) tool will a SoC built in 28nm with SRAM differentiates from the same SoC with MRAM, regarding area, power and timing.

4.4.1 OpenRISC Architecture

The OpenRISC architecture is an open architecture. It defines the architecture of an open source, Reduced Instruction Set Computing (RISC) microprocessor core. The OpenRISC architecture allows a wide spectrum of SoC designs implementations. It is a 32/64-bit load and store RISC architecture.

Performance features include a full 32/64-bit architecture, vector DSP and floating-point instructions, virtual memory support, cache coherency, optional Symmetric Multi-Processing (SMP) and SMT support and support for fast context switching. The architecture defines several features, such as several instruction extensions, a configurable number of general-purpose registers, configurable cache and TLB sizes, dynamic power management support, and space for user-provided instructions.

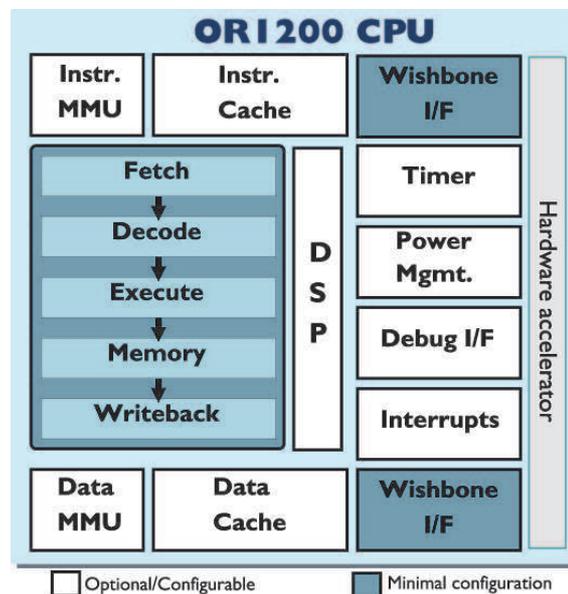


Figure 4.18: The OR1200 is a 32-bit scalar RISC with Harvard microarchitecture, 5 stage integer pipeline, virtual memory support (MMU) and basic DSP capabilities.

The full source for implementations of the OpenRISC architecture is available at www.opencores.org and is supported with GNU software development tools. Most OpenRISC implementations are designed to be modular and vendor-independent. They can be interfaced with other open-source cores available at www.opencores.org.

The OpenRISC architecture includes the following principal features:

- A royalty free and open architecture;
- A linear, 32-bit or 64-bit logical address space with implementation-specific physical address space;
- Simple and uniform-length instruction formats featuring different instruction set extensions;

- OpenRISC Basic Instruction Set (ORBIS_{32/64}) with 32-bit wide instructions aligned on 32-bit boundaries in memory and operating on 32- and 64-bit DATA;
- OpenRISC Vector/DSP eXtension (ORVDX₆₄) with 32-bit wide instructions aligned on 32-bit boundaries in memory and operating on 8-, 16-, 32- and 64-bit DATA;
- OpenRISC Floating-Point eXtension (ORFPX_{32/64}) with 32-bit wide instructions aligned on 32-bit boundaries in memory and operating on 32- and 64-bit DATA;
- Two simple memory addressing modes, where memory address is calculated by immediate or effective address;
- All operations are register based, including operands and results
- Shadowed or single 32-entry or 16-entry general purpose register file
- Optional branch delay slot for keeping the pipeline as full as possible
- Support for separate instruction and DATA caches/MMUs (Harvard architecture) or for unified instruction and DATA caches/MMUs (Stanford architecture)
- A flexible architecture definition that allows certain functions to be performed either in hardware or with the assistance of implementation-specific software
- Number of different, separated exceptions simplifying exception model
- Fast context switch support in register set, caches, and MMUs

4.4.1.1 *Memory Model*

Memory is byte-addressed with halfword accesses aligned on 2-byte boundaries, single-word accesses aligned on 4-byte boundaries, and doubleword accesses aligned on 8-byte boundaries.

The OpenRISC architecture specifies a weakly ordered memory model for uniprocessor and shared memory multiprocessor systems. This model has the advantage of a higher-performance memory system but places the duty for access ordering on the programmer.

The order in which the processor performs memory access, the order in which those accesses complete in memory, and the order in which those accesses are viewed by another processor may all be different. There are two mechanisms of enforcing memory access ordering provided to allow programs in uniprocessor and multiprocessor system to share memory.

An OpenRISC processor implementation can also implement a more restrictive, strongly ordered memory model.

The important detail about memory ordering is, considering that the control is placed into the program side as specified, the architecture implements a critical instruction **l.msync** used to synchronize memory.

4.4.1.2 *Memory Synchronize Instruction*

The `l.msyc` instruction permits the program to control the order in which load and store operations are performed. The memory sync instruction ensures that all memory accesses initiated by a program have been performed before the next instruction is executed. OpenRISC processor implementations, that implement the strongly-ordered memory model instead of the weakly-ordered one, can execute memory synchronization instruction as a no-operation instruction.

4.4.1.3 *Memory Management*

This section describes the virtual memory and access protection mechanisms for memory management within the OpenRISC architecture. Also, it describes the structure of the page tables, the MMU conditions that cause its related exceptions and the MMU registers. The hardware implementation details that are invisible to the OpenRISC programming model, such as MMU organization and TLB size, are not contained in the architectural definition.

The OpenRISC memory management unit includes the following principal features:

- Support for effective address (EA) of 32 bits and 64 bits
- Support for implementation specific size of physical address spaces up to 35 address bits (32 GByte)
- Three different page sizes:
 1. Level 0 pages (32 Gbyte; only with 64-bit EA) translated with D/I Area Translation Buffer (ATB)
 2. Level 1 pages (16 MByte) translated with D/I Area Translation Buffer (ATB)
 3. Level 2 pages (8 KByte) translated with D/I Translation Lookaside Buffer (TLB)
- Address translation using one-, two- or three-level page tables
- Powerful page based access protection with support for demand-paged virtual memory
- Support for SMT

The primary functions of the MMU in an OpenRISC processor are to translate effective addresses to physical addresses for memory accesses. In addition, the MMU provides various levels of access protection on a page-by-page basis. Note that this is just an conceptual overview model of the OpenRISC MMU implementations can differ in the specific hardware used to implement this model. Generally, the address translation mechanism is defined in terms of page tables used by OpenRISC processors to locate the effective to physical address mapping for instruction and DATA accesses.

Translation Lookaside Buffer (TLB) are commonly implemented in OpenRISC processors to keep recently-used page address translations on-chip. Although their exact implementation is not specified, the general concepts regarding the system software are described.

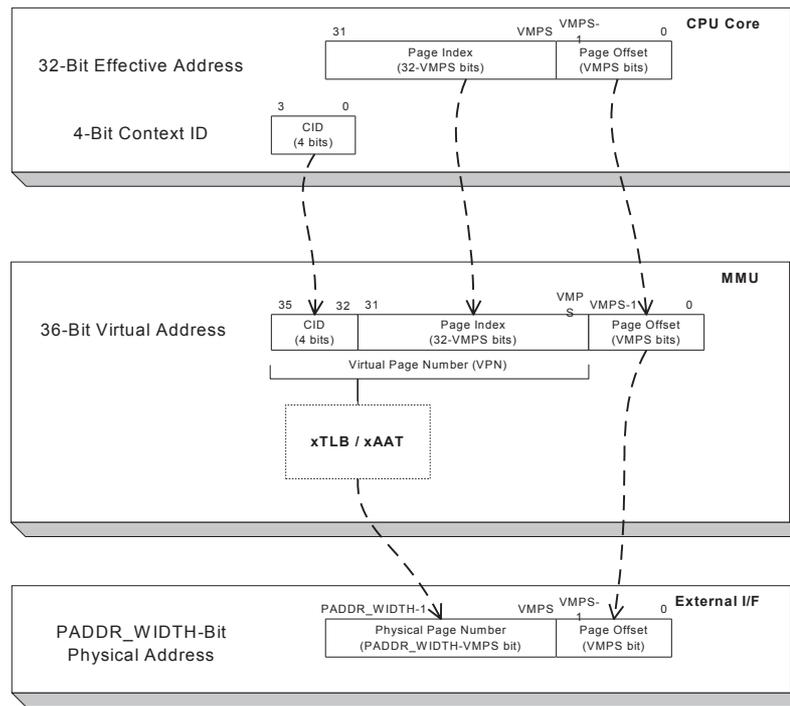


Figure 4.19: Translation of Effective to Physical Address. Simplified block diagram for 32-bit processor implementations

Large areas can be translated with optional facility called Area Translation Buffer (ATB). ATBs translate 16MB and 32GB pages. xTLB is used if xTLB and xATB have a match on the same virtual address.

The MMU with the exception processing mechanism, provides the necessary support for the operating system to implement a paged virtual memory environment and enforcing protection of designated memory areas.

4.4.2 MRAM Characterization using Liberty

The process to characterize a library using Liberty format is not a straight-forward approach. It is necessary to take into account the information of [Linser, 2010; Synopsys Inc., 2006, 2008, 2013a,b]. Sometimes, the information available seems not enough and the process became a try-and-error. Still, this seems to be the best way to obtain results about the final characteristics of a SoC that integrates no matter which technology. For our case, the characterization was for memory banks, in order to perform a complete Integrated Circuit (IC) synthesis flow. This way, we could obtain the SoC reports using an EDA tool of-the-shelf, linking our library that characterizes MRAM with the 28nm synthetic library, provided by Synopsys Inc..

The reason to use CCS coupled with Liberty was based in [TSMC and Synopsys Inc., 2007] and [Bao et al., 2006]. In [Bao et al., 2006] they found that CCS modeling delivered accuracy within 2% of difference compared to SPICE. This means that you can obtain a highly accurate model tight with your circuit synthesis, near the final physical implementation of your Integrated Circuit (IC). This way you will have a better insight about your MRAM memory banks. In our case, the Liberty was enough to obtain the

<http://www.OpensourceLiberty.org/>

areas comparisons and leakage. For a more accurate and detailed series of reports, the library needs to pass to a more detailed characterization process, checking all the CCS matrices.

4.4.3 Synthesis Results

In this section, is presented and compared the results of two physical synthesis of the OpenRISC targeting a 28nm technological node. The tables will expose the summary data about the synthesis result for the 28nm technology library, with the add-on of the characterization library for MRAM. The mram_28nm_ff1p16v25c, for fast-fast case, power supply of 1.6V and 25°C of operating temperature, assumes as base line for the characterization the sram_28nm_ff1p16v125c, that is similar to the MRAM, fast-fast case, power supply of 1.6V and 25°C of operating temperature. So, they are similar regarding the electrical parameters of the technology, due to the fact that both carry the same interconnection circuitry.

The difference between my characterization library to that of SRAM, is that I modified removing the unnecessary cells, setting the area and leakage according with NVSim. The transition curves, I kept similar to the SRAM, due to the fact that the MRAM banks interconnections that I'm adopting are the same as the SRAM banks, regarding line and column decoder as well as the sense amplifiers. Which means the access logic in the CMOS level is the same in SRAM and MRAM, the biggest difference residing in the data array matrix.

For this experiment the OpenRISC was configured with 32KB of data and Instruction CACHE L1 and Instruction and data TLB, each one with 64 entries, with page size of 8KB, addressing up to 4GB of address space. The architecture is similar as presented in Figure 4.20.

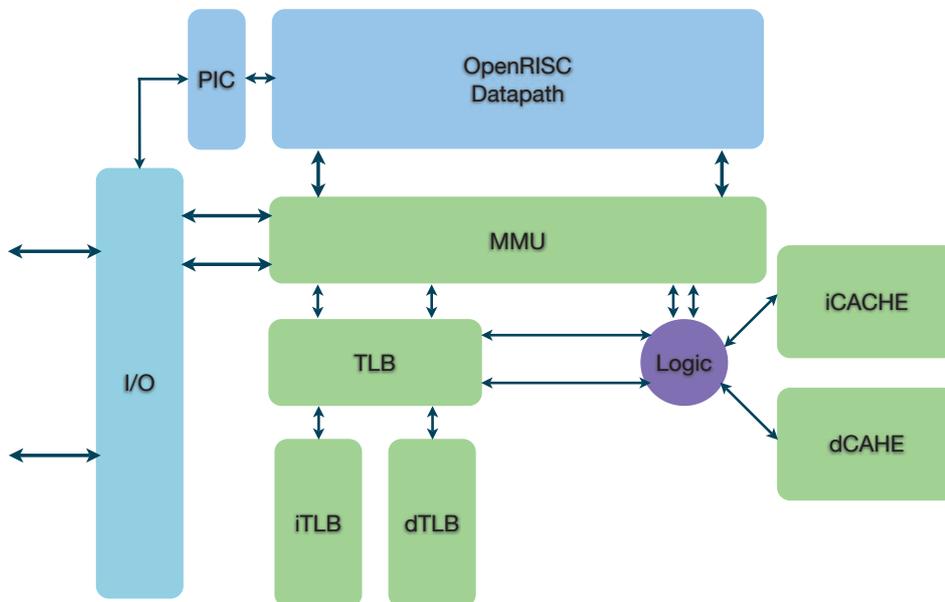


Figure 4.20: Synthesized OpenRISC architecture organization.

Specifically for the CACHES, the RTL design is based on a composite of memory banks chosen during synthesis, based on the address and word sizes to match the available **SRAM** banks into the tech library 28nm. For this reason, the **MRAM** banks were built equally to the **SRAM** banks listed in [Table 4.7](#) and [Table 4.8](#). This way, was simple to place conditional synthesis flags into the Verilog RTL of the microprocessor architecture.

So, the purpose here is to use a standard state-of-the-art Integrated Circuit (**IC**) flow, at the synthesis level, linking a **STT-MRAM** library, composed of memory banks of different word sizes, and different lengths, integrating it into this flow. The purpose was, also, demonstrate how to use this set of existing tools, characterizing a memory bank, based into a **MTJ** physical parameters, and how to evaluate a given circuit to analyze the required performance, latencies and power dissipation in a way that would be less speculative, or more precise and analytic, based onto the memory physical factors.

In [Table 4.7](#) and [Table 4.8](#), is presented the number of banks of intrinsic memories, instantiate by the synthesizer, used for TLB, instruction and data **CACHE**, composing the memory hierarchy of this particular OpenRISC implementation.

Table 4.7: Summary of instances and area for **SRAM** banks

Gate	Instances	Area (μm^2)
SRAM1RW512x32	32	1621365.825
SRAM1RW512x8	8	123818.056
SRAM1RW64x32	4	34854.011
SRAM2RW128x16	16	190224.400
SRAM2RW32x32	2	18617.339
		1988879.632

Table 4.8: Summary of instances and area for **MRAM** banks

Gate	Instances	Area (μm^2)
MRAM1RW512x32	32	160000.000
MRAM1RW512x8	8	24000.000
MRAM1RW64x32	4	16000.000
MRAM2RW128x16	16	48000.000
MRAM2RW32x32	2	8000.000
		256000.000

4.4.4 *MRAM and SRAM synthesis results*

Herein are the synthesis results regarding the **MRAM** and **SRAM**. In [Table 4.9](#) and [Table 4.11](#) are presented the results about number of cells and sum of the area for this set of cells. So, with this two tables you will be able to understand the area relation between **SRAM** and **MRAM**. To explain one minor detail, some tables shows **Instances**

others Cells, they are the same thing, it is just how the tools calls a mapped element like a NAND2.

Table 4.10 Table 4.12

Table 4.9: MRAM area summary, regarding logic and memory.

Library	Cells	Area (μm^2)	Instances %
CMOS cells 28nm	23941	169 282.408	99.7
MRAM cells	62	256 000.000	0.3
total	24 003	425 282.408	100.0

Table 4.10: MRAM synthesis summary

Type	Instances	Area (μm^2)	Area %
memory hierarchy	62	256 000.000	60.0
logic core	24 572	170 322.555	40
total	24 634	426 322.555	100.0

Table 4.11: SRAM area summary, regarding logic and memory.

Library	Cells	Area (μm^2)	Instances %
CMOS cells 28nm	28 976	96 911.715	99.8
SRAM cells	62	1 988 879.632	0.2
total	29 038	2 085 791.347	100.0

Table 4.12: SRAM synthesis summary

Type	Instances	Area (μm^2)	Area %
memory hierarchy	62	1 988 879.632	96.4
logic core	23 739	74 760.015	3.6
total	23 801	2 063 639.647	100.0

4.4.5 Assessments comparing the synthesis results of the two technologies

The relevant report in the gates report is shown in Table 4.12. This present the total area regarding the final circuit as shown in Table 4.11. So, the first inference we can assert is that the SRAM accounts for $1\,988\,879.632\mu\text{m}^2$ of $2\,063\,639.647$, which represents 91.34%, taking into account the total SoC area presented in Table 4.12, of the total area of the final circuit, targeting SRAM based memory banks. While in same manner the MRAM accounts for $256\,000.000\mu\text{m}^2$ of $425\,282.408$, which represents $\sim 60\%$ of the surface.

So the first great aspect of MRAM in a memory hierarchy of a SoC is to reduce in nearly $\sim 30\%$ the amount of area in the circuit. Also, in SRAM the power driver is

the **SRAM** cells, adopting **MRAM** the power driver becomes the logic of the circuit, and this means two things: 1) the power consumption is greatly reduced, since **CMOS** consumes when it is switching. Basically, in the case of **SRAM** is always consuming, 2) improvements into the low power technics on the core will have a much higher impact in the total power consumption. Based in each memory technology, as well as each final **SoC** power characteristics. So, essentially the power will be driven by the logic circuit, and not the memory banks, in circuits that implements **MRAM** memories. This leads to low power technics that can, further, reduce the power consumption without affecting performance like spintronics circuits.

The result that was unexpected is the leakage Power for **MRAM**, that in fact was increased. But the total **SoC** power remained below 50% of the power on same **SoC** built in **SRAM**. A possibility could be the gate dimensioning to support the current variation, due to the high write current for the **MTJ** into the memory banks. Another possibility, could be the fact that the 28nm library for **SRAM** memory banks do not have the leakage of each memory bank cell, so the **RTL** compiler cannot calculate accurately the leakage for them.

Another important aspect, more than just the areas of the banks, is that the **SoC** based in **MRAM** has a decrease of 4.84 times in area. Which means it is possible to place on the wafer near four dies based in **MRAM** in the area of one die based in **SRAM**. Consequently, an increase of four times of the yielding per wafer, just switching the memory technology.

So, a few points we observe in this section, once we characterized the **MRAM** memory bank cells, performing a logical, physical synthesis using Cadence **RTL** compiler, also we performed in before hand the same process using **SRAM**. Observing the results and comparing them we can assess that:

1. The **MRAM** needs at least 4 times less substrate surface compared to the **SRAM**;
2. The leakage with **MRAM** increases, probably due to the sense amplifiers or access gates, that are dimensionate to support write current of the **MTJ** into the memory arrays;
3. we can increase in four times the number of silicon dies per wafer using **MRAM**, to replace **SRAM**, with same sizes;
4. the final **SoC**, truly has a leakage in memory increased, but the total power consumption is lower than the **SRAM** counterpart. Also the **SRAM** library does not contain the leakage of the memory bank cells, so the synthesis tool has to extract it as best as it can;
5. adopting **MRAM**, the power constraints shift from the memory back to the circuit logic. The memory is not anymore the main drain of power;
6. low power techniques applied to logic produces better results. Currently, the low power technics like clock-gating, multi voltage due to the physical aspects of the **SRAM** cannot achieve better results, considering that the **SRAM** is always consuming (volatile memory), so the low power technics can't push further improvements;

To close this section, **MRAM** seems to produce smaller and energy-efficient circuits. This is theoretically expected, but so far no one have demonstrated, in practice, how

to physically obtain such results. The most important is the characterization library of your memory banks. This is the crucial step needed, so the synthesizer will be able to compute and make all necessary extrapolations, in order to achieve the desired time frame for a given technological node, obtaining results that will match the technological DFM.

4.5 THE COMPOSITE BANK

This is the closing experiment of my research. It occurs that observing the results in [Section 3.5.3](#), as well as analyzing the source code of the NVSim and CACTI, a few things can be noticed. The first immediate observation is that the analyses of MRAM were affected by the TAG memory array. The TAG memory array is the bottleneck of the cache memory banks. They will define your throughput for reading. Also, since they are replaced in every context-switch, the latency to write can affect the overall system performance. The reason is, using MRAM memory cells that possess higher write latency, will create a memory bank with higher latency, that can perform less transactions in the same amount of time as SRAM, so lower bandwidth than SRAM.

The second factor observed was that, even with a higher leakage on SRAM, the TAG array was a minimal slice of this power consumption compared to the DATA array matrix in SRAM. Less than half of the leakage and the energy to access it. You can confirm this looking into the results at [Section 3.5.3](#). So, based on this results, the lead conclusion was, instead of built-in a hybrid memory cell like in [[BRUCHON, 2007b](#); [Sun et al., 2009a](#)], the best trade-off would be a composite bank, with the TAG memory array built-in in SRAM and the DATA array matrix in MRAM. Additional results can be found in [Appendix F](#). There it can be found the LOP-only for 45nm and the results for 28nm.

Generally the TAG size is calculated, observing the [Figure 2.11](#), generally the dimensioning of address sizing depend on how you build your memory hierarchy architecture. In general the index is defined as $\log_2(\text{cache rows})$, since the effective address is in bytes the block offset is $\log_2(\text{bytes per block})$, leaving for the TAG dimensioning as $\text{address_length} - \text{index_length} - \text{block offset}$ this is well explained in [[Hennessy and Patterson, 2006](#); [Patterson and Hennessy, 2012](#)].

The [Figure 4.21](#) depicts the general idea, where the TAG is in SRAM, while the DATA is in MRAM, offering a better trade-off for a CACHE memory bank as will be shown.

So, the composite bank is, literally, a bank composed by n memory technologies to mitigate the MRAM delay, without losing the advantages of power consumption, trying to obtain the best trade-off.

4.5.1 Composite Memory banks of 45nm

In [Figure 4.22](#) we can observe that, despite the area usage is not as good as that of MRAM only memory bank, but is better if compared to the SRAM-only bank. We adopted a trade-off between the technologies obtaining this way better results in latency and power, compromising some area.

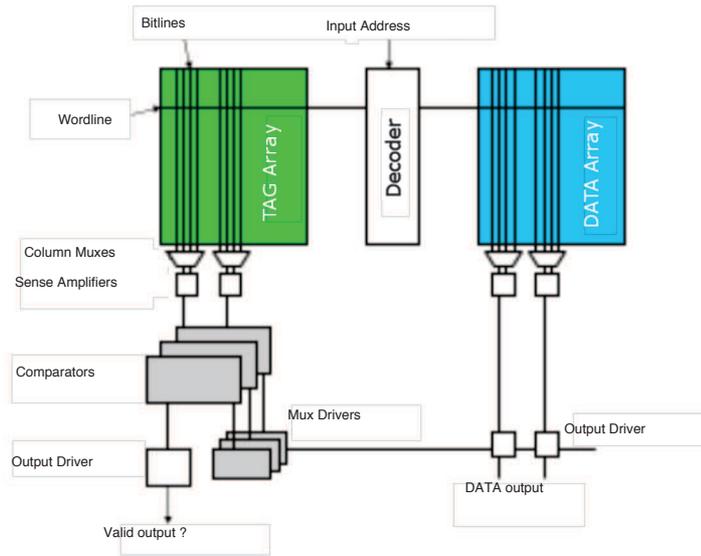


Figure 4.21: Composite CACHE set bank architecture, built with two heterogeneous memory technologies. In the Figure the TAG(green matrix, left) is built using SRAM, while the DATA (blue matrix, right) is built using STT-MRAM.

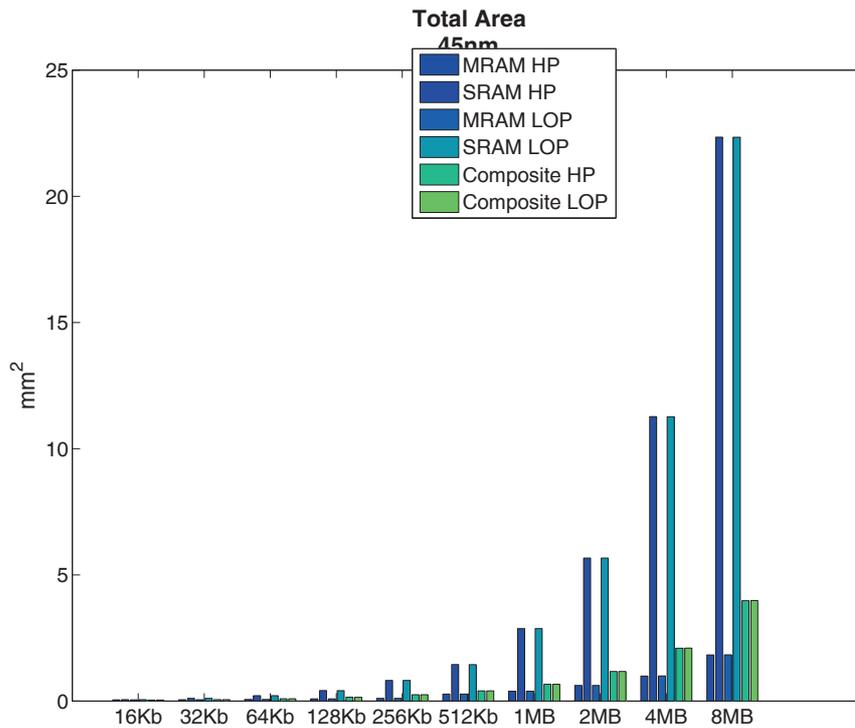


Figure 4.22: Total Area.

Observing the depicting in Figure 4.22, generated from Table 4.13, we can observe that is losen a little of surface, but managing to keep a great margin of advantage compared to the SRAM-only banks in HP or LOP model. This is also a better deal than the proposed MSRAM or RSRAM [BRUCHON, 2007b; Sun et al., 2009a], that uses a

Table 4.13: Total Area (μm^2).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0474	0.0569	0.0476	0.0568	0.0422	0.0422
32KB	0.0528	0.1086	0.0529	0.1083	0.0588	0.0587
64KB	0.0637	0.2126	0.0638	0.2120	0.0901	0.0899
128KB	0.0863	0.4145	0.0861	0.4142	0.1522	0.1518
256KB	0.1085	0.8195	0.1086	0.8189	0.2500	0.2498
512KB	0.2757	1.4461	0.2779	1.4455	0.4029	0.4046
1MB	0.3904	2.8712	0.3921	2.8700	0.6646	0.6659
2MB	0.6178	5.6655	0.6183	5.6643	1.1736	1.1738
4MB	0.9919	11.2698	0.9958	11.2673	2.0935	2.0964
8MB	1.8285	22.3432	1.8302	22.3407	3.9808	3.9818

SRAM latch as sense amplifiers for the MTJs, in a hybrid cell fashion. Due to the fact that, in a hybrid cell you have all the leakage of a SRAM cell, plus all the penalty for high latency to write to the MTJs. Also, any additional set of transistors to access the MTJs in parallel without disturbing the latch content, will increase the amount of transistors per cell, not forgetting the access MOS of the MTJ, that has to be dimensioned to support the write current.

In [Figure 4.23](#) we have depicted the Write Latency of the cache Bank. Since the DATA array is in MRAM, it has a general impact into the total latency to write that cannot be avoided. Either way, the final result is still better. Observing [Table 4.14](#), we have a slightly better result compared to MRAM-only, e.g., if we assume the values of 8MB LOP, MRAM and Composite we have a reduction of 50.053ns of latency. Not impressive? Now, multiply it by a three million accesses. You would have in 5030760000ns of total access time for the MRAM-only and 3529170000ns for the Composite bank, merely 1501590000ns of access time reduction, for only three million accesses. Which nowadays, with demands of streaming processing like X.264, on embedded systems three million or more accesses in burst, are not far from possible.

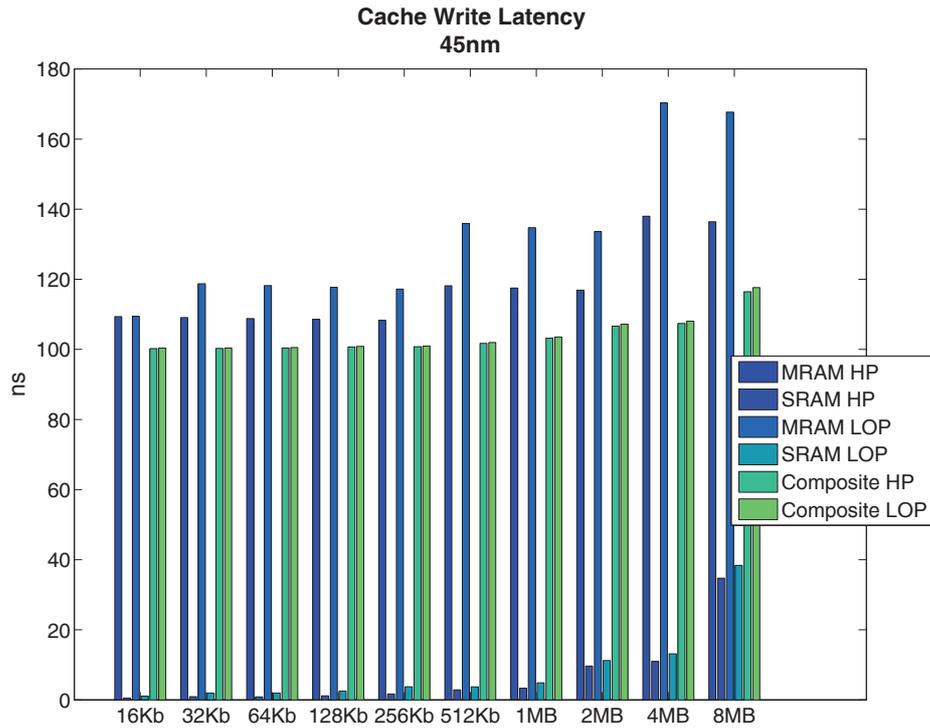


Figure 4.23: Total Write Latency.

Table 4.14: Cache Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	109.3390	0.5315	109.4420	1.0987	100.2340	100.3730
32KB	109.0710	0.9391	118.7320	1.9773	100.2870	100.4140
64KB	108.8020	0.8787	118.2040	1.9596	100.4030	100.5380
128KB	108.5620	1.1819	117.7230	2.5528	100.6970	100.8680
256KB	108.2970	1.6796	117.1990	3.7679	100.7590	100.9420
512KB	118.1400	2.8795	135.8750	3.7399	101.7170	101.9540
1MB	117.5150	3.3515	134.7310	4.8844	103.1950	103.5410
2MB	116.9060	9.6606	133.6090	11.2448	106.6340	107.1940
4MB	137.9750	10.9909	170.2960	13.1675	107.4050	108.0310
8MB	136.4120	34.7428	167.6920	38.3881	116.4490	117.6390

In Figure 4.24, Figure 4.25 we observe the leakage compared to the Composite. This is one of the major breaks of this approach, since is much more simpler to build such memory structure. Also, is easier to integrate into the DFM flow, and we have a really fair trade-of.

In Figure 4.24 and Figure 4.25 is depicted the leakage power for all kinds of memory banks and the LOP only, respectively. The leakage of the tag array increases proportionally to the DATA array on HP memory banks.

In Figure 4.24, we observe the total leakage power, so TAG+DATA arrays of a cache memory bank. Adding up the TAG with the DATA does not help the SRAM case regarding leakage. Here in this case, the leakage is a great problem for HP memory banks. The great observation is, MRAM despite its high latency, on power it is always better regarding leakage. SRAM never overcome MRAM on leakage. Also, look into the numerical results in Table 4.15 and will observe that, despite the Composite presenting a worst leakage than the only MRAM banks, when verifying and comparing to the SRAM, e. g., assume the 8MB memory bank on LOP, the MRAM 0.87, SRAM 52.5 and the Composite 6.34, the leakage relation from Composite to MRAM is 7.25 worst, due to the adoption of SRAM in the TAG array. But, compared to the SRAM, the SRAM has 8.27 worst leakage than the Composite, which means the Composite is really in the mean/average term of trade-off relation.

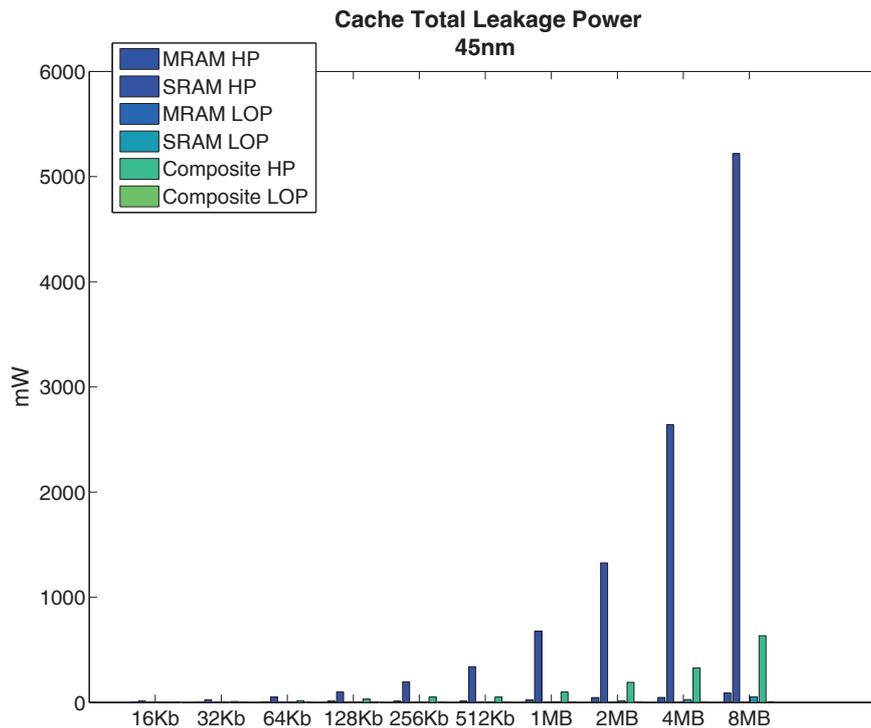


Figure 4.24: Composite Leakage Power.

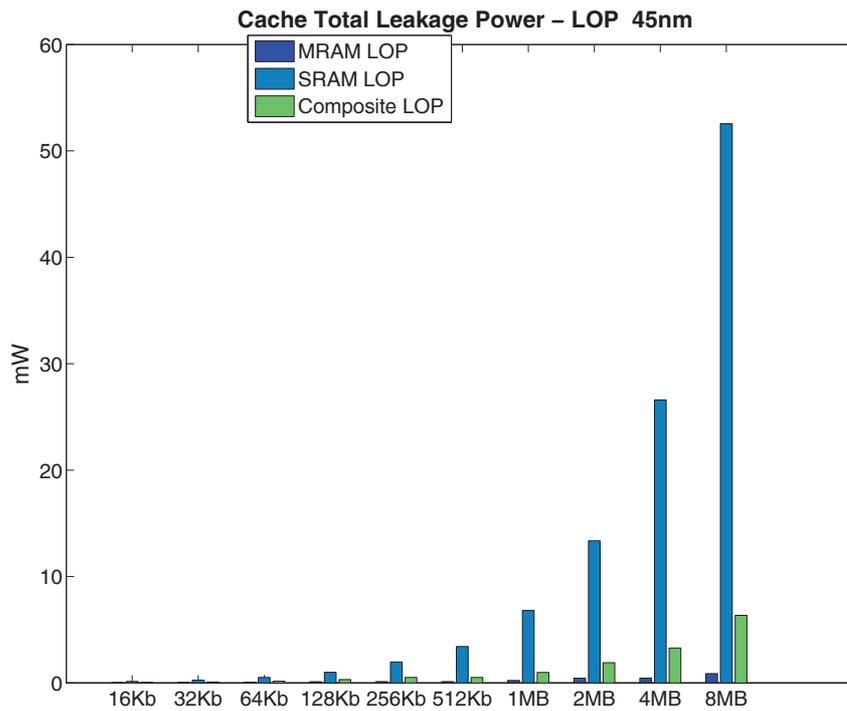


Figure 4.25: Low-Power (LOP) Total Leakage, magnification on the LOP banks observed in Figure 4.24.

Table 4.15: Cache Total Leakage Power (mW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	2.2855	14.1126	0.0218	0.1414	4.8037	0.0480
32KB	3.6622	26.2267	0.0363	0.2627	8.7701	0.0872
64KB	6.4012	51.6340	0.0624	0.5174	16.7710	0.1661
128KB	11.8930	99.4210	0.1148	0.9978	31.8661	0.3152
256KB	12.4789	196.1430	0.1204	1.9691	51.3967	0.5115
512KB	13.0065	340.7700	0.1300	3.4262	51.6737	0.5181
1MB	23.9675	676.9460	0.2344	6.8072	98.7904	0.9868
2MB	45.8213	1326.8000	0.4427	13.3539	191.0998	1.9047
4MB	47.5148	2640.4800	0.4593	26.5774	327.8455	3.2821
8MB	91.1782	5218.4800	0.8754	52.5496	634.4444	6.3468

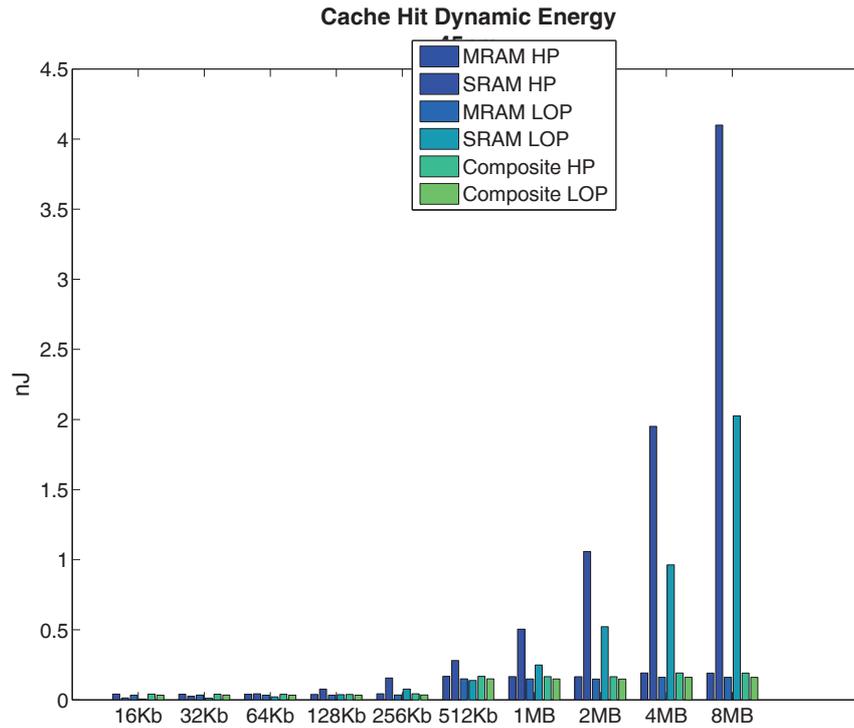


Figure 4.26: Hit Dynamic Energy.

Table 4.16: Cache Hit Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0411	0.0138	0.0335	0.0068	0.0411	0.0335
32KB	0.0405	0.0256	0.0345	0.0126	0.0405	0.0345
64KB	0.0400	0.0438	0.0340	0.0215	0.0400	0.0340
128KB	0.0394	0.0770	0.0335	0.0379	0.0394	0.0335
256KB	0.0436	0.1561	0.0354	0.0770	0.0436	0.0354
512KB	0.1675	0.2815	0.1504	0.1387	0.1675	0.1504
1MB	0.1665	0.5047	0.1494	0.2491	0.1665	0.1494
2MB	0.1654	1.0586	0.1485	0.5226	0.1654	0.1485
4MB	0.1924	1.9511	0.1618	0.9638	0.1924	0.1618
8MB	0.1912	4.0991	0.1608	2.0252	0.1912	0.1608

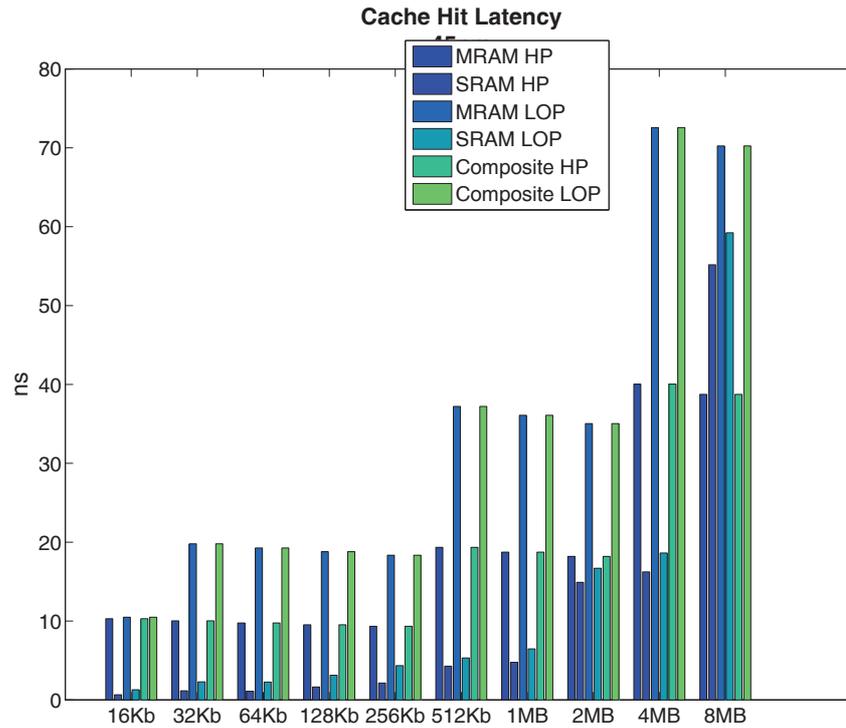


Figure 4.27: Hit Latency.

One factor that is easily noticeable is: regarding [Figure 4.28](#), it is the same as [Figure 4.26](#). So, miss or hit, does not matter, the energetic cost will be the almost the same. However, in the Composite memory bank, again is achieved a better trade-off in energetic cost compared to SRAM, narrowing the gaps between SRAM and MRAM and obtaining the best advantages of each one in a composite approach.

Table 4.17: Cache Hit Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	10.2953	0.6533	10.4902	1.2926	10.2953	10.4902
32KB	10.0276	1.1525	19.7923	2.2867	10.0276	19.7923
64KB	9.7603	1.0939	19.2655	2.2712	9.7603	19.2655
128KB	9.5246	1.6477	18.7884	3.1341	9.5246	18.7884
256KB	9.3351	2.1476	18.3503	4.3523	9.3351	18.3503
512KB	19.3512	4.2859	37.2127	5.3230	19.3512	37.2127
1MB	18.7428	4.7618	36.0863	6.4738	18.7428	36.0863
2MB	18.2018	14.9050	35.0382	16.7024	18.2018	35.0382
4MB	40.0442	16.2397	72.5652	18.6320	40.0442	72.5652
8MB	38.7401	55.1709	70.2395	59.2280	38.7401	70.2395

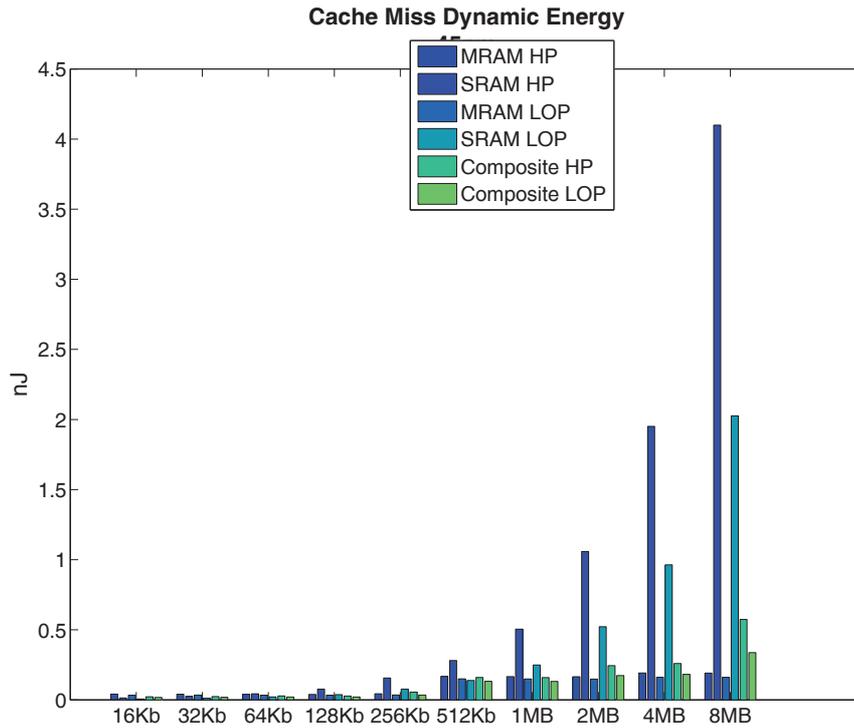


Figure 4.28: Miss Dynamic Energy.

Table 4.18: Cache Miss Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0411	0.0138	0.0335	0.0068	0.0222	0.0176
32KB	0.0405	0.0256	0.0345	0.0126	0.0241	0.0186
64KB	0.0400	0.0438	0.0340	0.0215	0.0283	0.0207
128KB	0.0394	0.0770	0.0335	0.0379	0.0281	0.0206
256KB	0.0436	0.1561	0.0354	0.0770	0.0558	0.0343
512KB	0.1675	0.2815	0.1504	0.1387	0.1603	0.1326
1MB	0.1665	0.5047	0.1494	0.2491	0.1595	0.1322
2MB	0.1654	1.0586	0.1485	0.5226	0.2444	0.1741
4MB	0.1924	1.9511	0.1618	0.9638	0.2598	0.1819
8MB	0.1912	4.0991	0.1608	2.0252	0.5740	0.3373

On the other hand, the latencies depicted in [Figure 4.29](#), if compared to [Figure 4.27](#), we can observe that for SRAM a miss costs more, while for the MRAM, it is the same cost. So, when using the Composite approach, we obtain a better bargain in access latency, due to misses. Since we are using only the TAG as SRAM and the access to the MRAM DATA array to writeback between CACHE layers is faster, assuming it works as modeled in [Figure 2.11](#), once the DATA is found, the CPU does not have to wait it to finishing uploading from L2 to L1. It has the value available immediately.

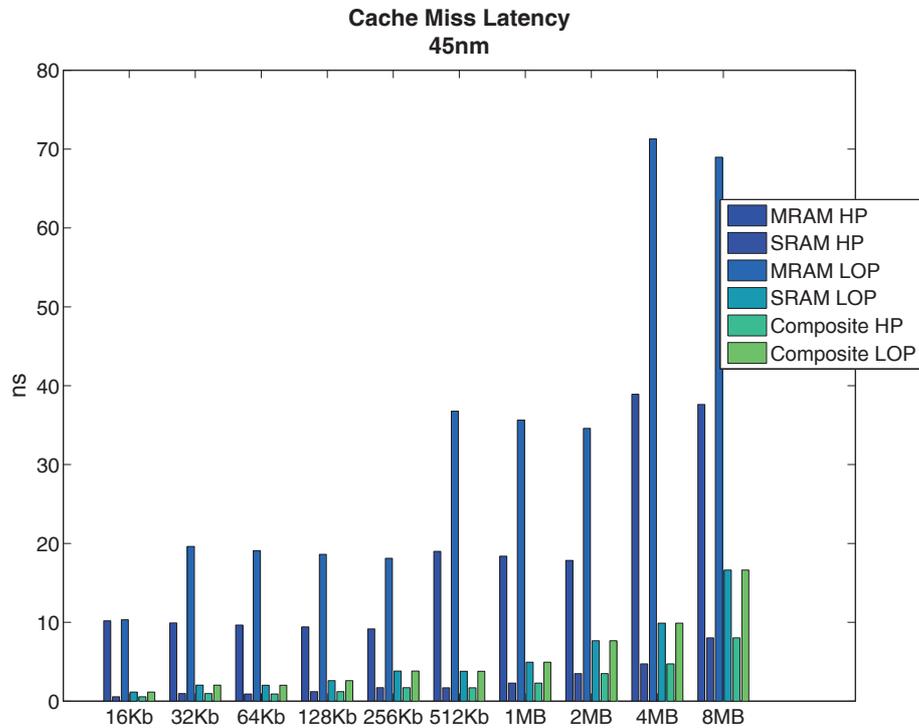


Figure 4.29: Miss Latency.

As observed in [Figure 4.29](#) and [Figure 4.30](#), the Composite is keeping up with the SRAM performance sustaining the MRAM energy footprint. The MTJ will set the bank speed at DATA array matrix and considering the SRAM aspects introduced in the TAG, like leakage and so on. The Composite is a memory bank that obtains the best of both technologies, inheriting the latency to write into the DATA array when is the case, as depicted in [Figure 4.30](#), whose latency is similar to the MRAM for reading. Assuming, this way, a linear pattern like MRAM. The SRAM, starting in 1MB it assumes a exponential tendency as depicted.

Table 4.19: Cache Miss Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	10.1874	0.5605	10.3223	1.1577	0.5605	1.1577
32KB	9.9195	0.9681	19.6243	2.0363	0.9681	2.0363
64KB	9.6520	0.9077	19.0972	2.0186	0.9077	2.0186
128KB	9.4158	1.2102	18.6196	2.6101	1.2102	2.6101
256KB	9.1674	1.7079	18.1132	3.8252	1.7079	3.8252
512KB	18.9930	1.6960	36.7710	3.7973	1.6960	3.7973
1MB	18.3854	2.2987	35.6448	4.9418	2.2987	4.9418
2MB	17.8419	3.4984	34.5929	7.6584	3.4984	7.6584
4MB	38.9224	4.7310	71.3032	9.8897	4.7310	9.8897
8MB	37.6153	8.0312	68.9729	16.6321	8.0312	16.6321

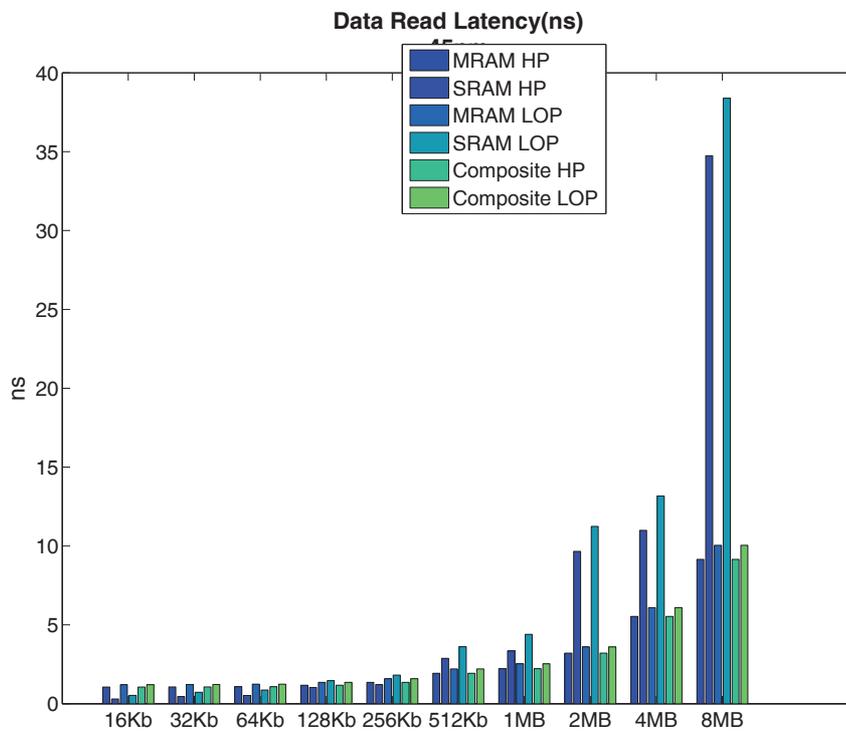


Figure 4.30: Data Read Latency.

In Figure 4.31 and Table 4.21, is just the TAG arrays latencies, that are the same for SRAM, so no changes here.

Table 4.20: Data Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	1.0553	0.3040	1.2077	0.5339	1.0553	1.2077
32KB	1.0621	0.4597	1.2178	0.7269	1.0621	1.2178
64KB	1.0803	0.5260	1.2382	0.8690	1.0803	1.2382
128KB	1.1762	1.0291	1.3590	1.4635	1.1762	1.3590
256KB	1.3564	1.2156	1.5860	1.8168	1.3564	1.5860
512KB	1.9328	2.8795	2.2071	3.6172	1.9328	2.2071
1MB	2.2221	3.3515	2.5359	4.4008	2.2221	2.5359
2MB	3.2074	9.6606	3.6027	11.2448	3.2074	3.6027
4MB	5.5340	10.9909	6.0894	13.1675	5.5340	6.0894
8MB	9.1545	34.7428	10.0463	38.3881	9.1545	10.0463

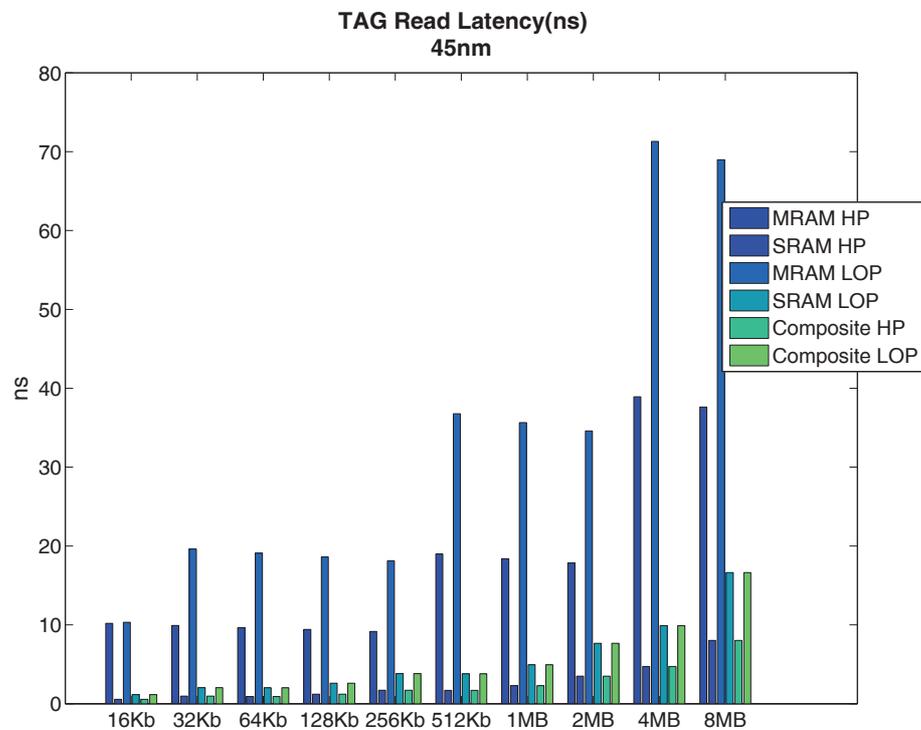


Figure 4.31: TAG Read Latency.

In the [Figure 4.32](#) and [Table 4.22](#), are exposed the details of the DATA array Write latencies, the biggest drawback of MRAM memory banks. Once this issue of anisotropy (see [Appendix A](#)) switching latency is solved, MRAM will present a great advantage over SRAM memory banks.

Table 4.21: TAG Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	10.1874	0.5605	10.3223	1.1577	0.5605	1.1577
32KB	9.9195	0.9681	19.6243	2.0363	0.9681	2.0363
64KB	9.6520	0.9077	19.0972	2.0186	0.9077	2.0186
128KB	9.4158	1.2102	18.6196	2.6101	1.2102	2.6101
256KB	9.1674	1.7079	18.1132	3.8252	1.7079	3.8252
512KB	18.9930	1.6960	36.7710	3.7973	1.6960	3.7973
1MB	18.3854	2.2987	35.6448	4.9418	2.2987	4.9418
2MB	17.8419	3.4984	34.5929	7.6584	3.4984	7.6584
4MB	38.9224	4.7310	71.3032	9.8897	4.7310	9.8897
8MB	37.6153	8.0312	68.9729	16.6321	8.0312	16.6321

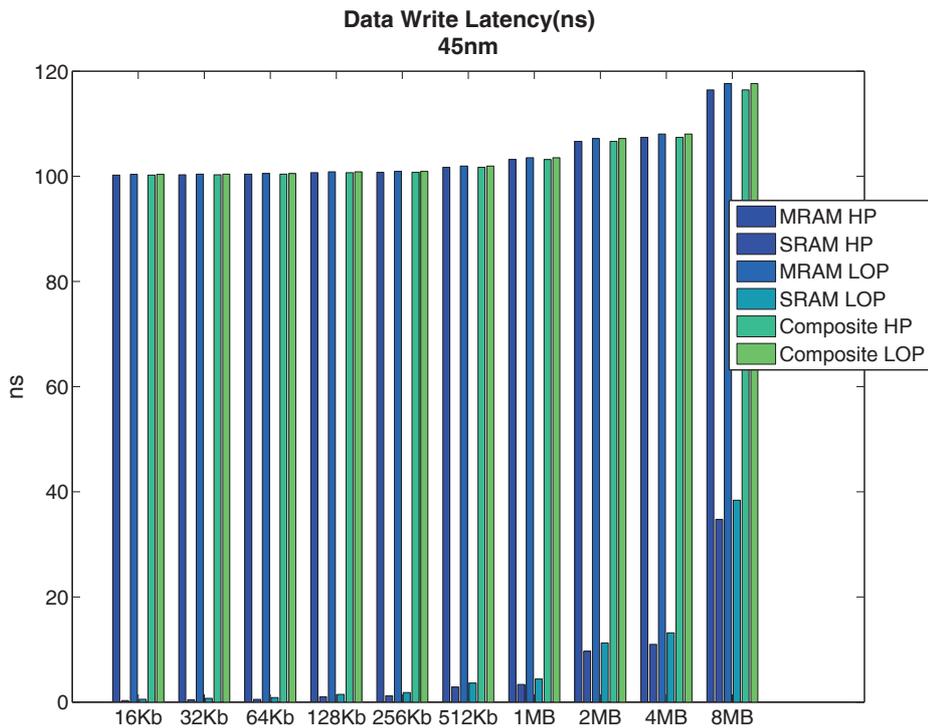


Figure 4.32: Data array matrix Write Latency.

Table 4.22: DATA array matrix Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	100.2340	0.3040	100.3730	0.5339	100.2340	100.3730
32KB	100.2870	0.4597	100.4140	0.7269	100.2870	100.4140
64KB	100.4030	0.5260	100.5380	0.8690	100.4030	100.5380
128KB	100.6970	1.0291	100.8680	1.4635	100.6970	100.8680
256KB	100.7590	1.2156	100.9420	1.8168	100.7590	100.9420
512KB	101.7170	2.8795	101.9540	3.6172	101.7170	101.9540
1MB	103.1950	3.3515	103.5410	4.4008	103.1950	103.5410
2MB	106.6340	9.6606	107.1940	11.2448	106.6340	107.1940
4MB	107.4050	10.9909	108.0310	13.1675	107.4050	108.0310
8MB	116.4490	34.7428	117.6390	38.3881	116.4490	117.6390

We can observe in Figure 4.33, supported by Table 4.23, with CMB we solve the TAG write latency problem. The TAG write latency is the pinnacle to obtain a better performance in a CACHE set-associative. The reason is simple during context switches the applications will dispute CACHE lines, the TAG are fully replaced during the context switch, so is necessary that the TAG performance should be stable to avoid degradations as we see in MRAM-only banks.

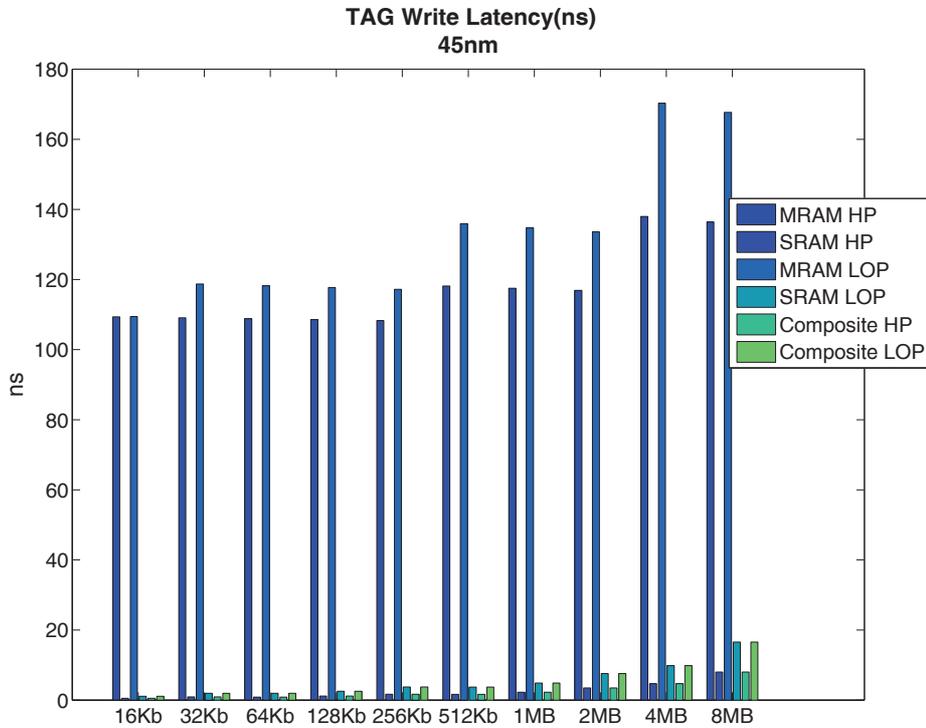


Figure 4.33: TAG Write Latency.

Table 4.23: TAG Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	109.3390	0.5315	109.4420	1.0987	0.5315	1.0987
32KB	109.0710	0.9391	118.7320	1.9773	0.9391	1.9773
64KB	108.8020	0.8787	118.2040	1.9596	0.8787	1.9596
128KB	108.5620	1.1819	117.7230	2.5528	1.1819	2.5528
256KB	108.2970	1.6796	117.1990	3.7679	1.6796	3.7679
512KB	118.1400	1.6677	135.8750	3.7399	1.6677	3.7399
1MB	117.5150	2.2704	134.7310	4.8844	2.2704	4.8844
2MB	116.9060	3.4701	133.6090	7.6010	3.4701	7.6010
4MB	137.9750	4.7034	170.2960	9.8340	4.7034	9.8340
8MB	136.4120	8.0036	167.6920	16.5764	8.0036	16.5764

Another achievement of the **CMB** is the reduction of write dynamic energy for the TAG array. Since the TAG banks are modified so constantly, this energy reduction is important to keep the total memory bank energy consumption low in the long-term.

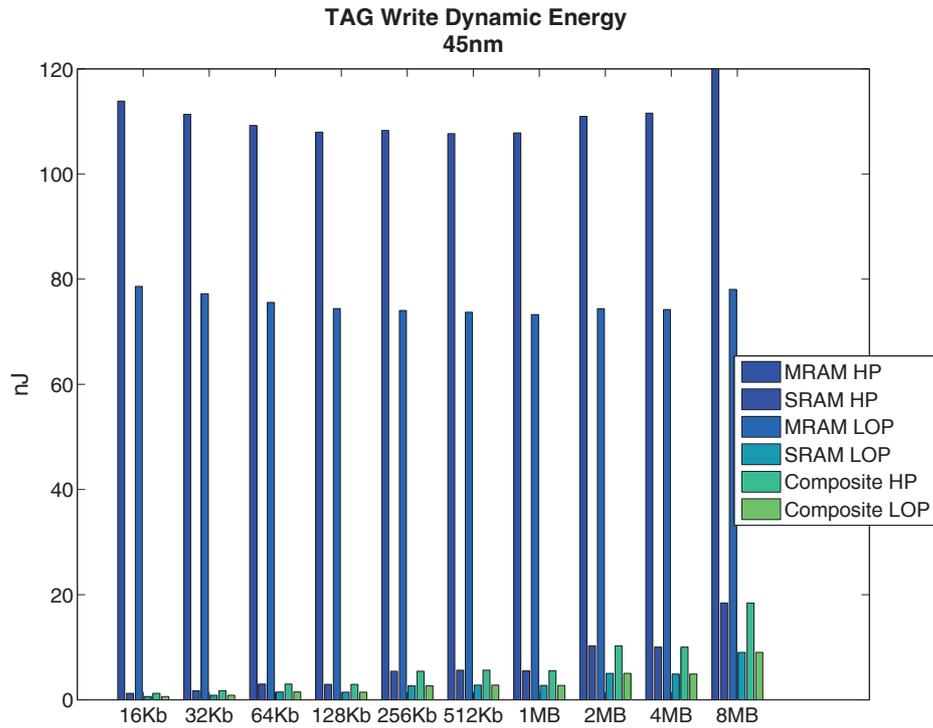


Figure 4.34: TAG Dynamic Energy

Table 4.24: TAG Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	113.8640	1.2262	78.6242	0.6093	1.2262	0.6093
32KB	111.3400	1.7284	77.2016	0.8576	1.7284	0.8576
64KB	109.2400	2.9948	75.5873	1.4812	2.9948	1.4812
128KB	107.9570	2.9350	74.3972	1.4513	2.9350	1.4513
256KB	108.2780	5.4298	74.0408	2.6533	5.4298	2.6533
512KB	107.6720	5.6278	73.7150	2.7846	5.6278	2.7846
1MB	107.8230	5.5258	73.2701	2.7341	5.5258	2.7341
2MB	110.9670	10.2586	74.3801	5.0145	10.2586	5.0145
4MB	111.5730	10.0250	74.2152	4.9001	10.0250	4.9001
8MB	119.9700	18.3804	78.0532	9.0142	18.3804	9.0142

In Figure 4.35 and Table 4.25, is the another great outcome of the Composite bank. Checking the depicting Figure 4.36, you can observe that the MRAM-only has the worst possible Dynamic Energy prospect, while the SRAM has the best, given the physical factors of operation of each memory technology. The Composite reduces the HP in 1.43 times the energy, and LOP in 1.50 times the amount of energy. This is a considerable factor given the fact that you are now 3.74 higher than SRAM instead of 5.37 in HP, and 4.625 instead of 6.95 in LOP. This way, with our Composite approach we are shortening the gap in energy to write, between MRAM and SRAM, which puts MRAM in advantage on long-term.

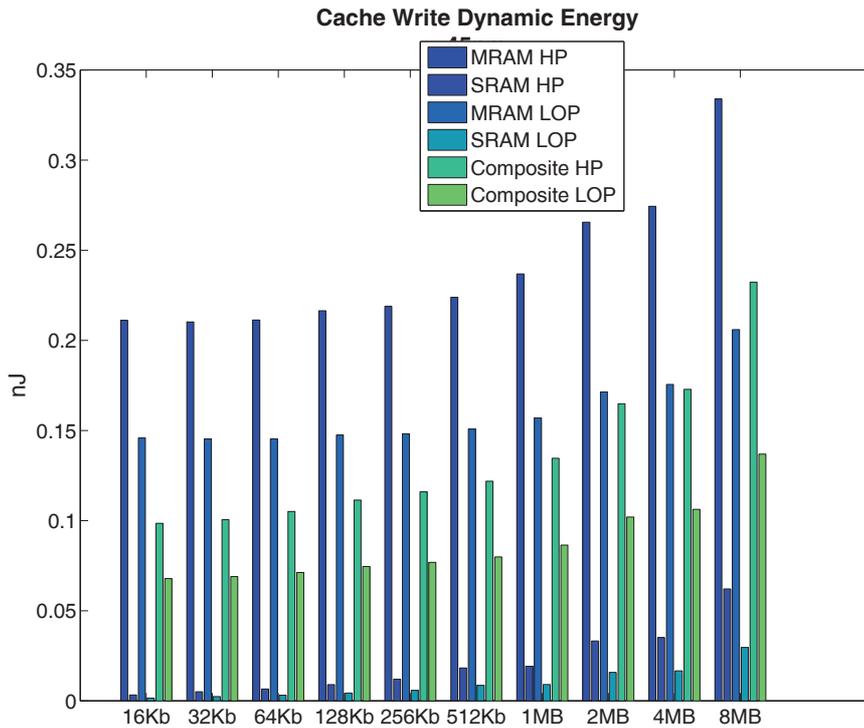


Figure 4.35: CACHE Write Dynamic Energy

Table 4.25: Cache Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.2111	0.0033	0.1459	0.0016	0.0985	0.0679
32KB	0.2102	0.0050	0.1453	0.0024	0.1006	0.0689
64KB	0.2113	0.0066	0.1453	0.0032	0.1050	0.0712
128KB	0.2164	0.0090	0.1475	0.0043	0.1114	0.0745
256KB	0.2188	0.0121	0.1482	0.0058	0.1160	0.0768
512KB	0.2239	0.0182	0.1508	0.0086	0.1219	0.0798
1MB	0.2369	0.0192	0.1570	0.0091	0.1346	0.0864
2MB	0.2656	0.0332	0.1714	0.0157	0.1649	0.1020
4MB	0.2743	0.0351	0.1755	0.0167	0.1728	0.1062
8MB	0.3339	0.0621	0.2059	0.0296	0.2323	0.1369

Regarding the DATA write dynamic energy, there is no solution, unless the MTJ technology is improved, as observed in Figure 4.36 and Table 4.26.

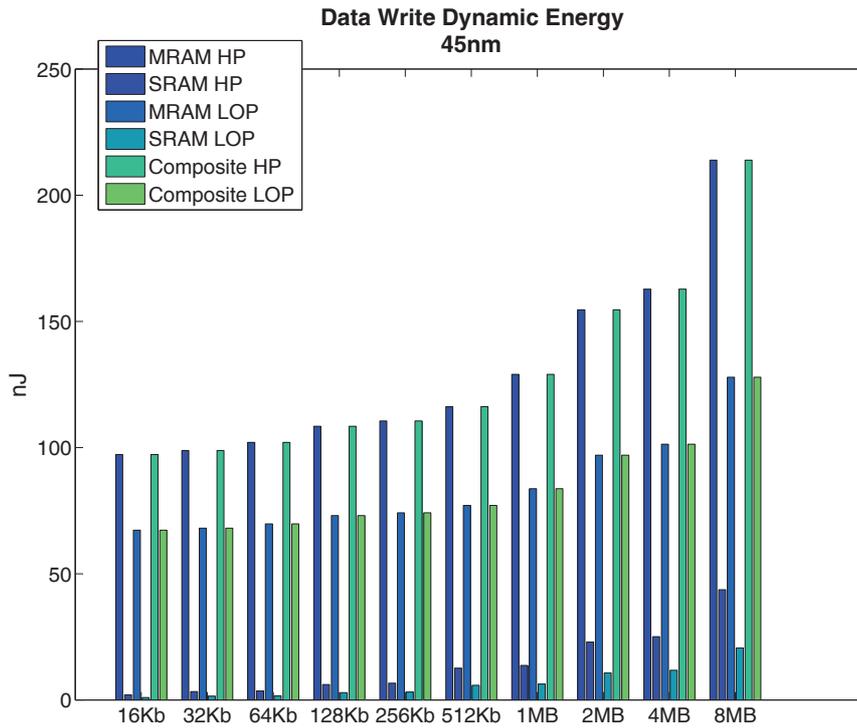


Figure 4.36: Data Dynamic Energy

Table 4.26: Data Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	97.2497	2.0369	67.2541	0.9587	97.2497	67.2541
32KB	98.8485	3.3199	68.0847	1.5590	98.8485	68.0847
64KB	102.0460	3.5854	69.7460	1.6901	102.0460	69.7460
128KB	108.4440	6.1064	73.0693	2.8865	108.4440	73.0693
256KB	110.5420	6.6398	74.1747	3.1498	110.5420	74.1747
512KB	116.2280	12.6088	77.0569	5.8032	116.2280	77.0569
1MB	129.0310	13.6730	83.7073	6.3284	129.0310	83.7073
2MB	154.5920	22.9746	96.9860	10.7319	154.5920	96.9860
4MB	162.7660	25.1002	101.3290	11.7810	162.7660	101.3290
8MB	213.8860	43.6946	127.8870	20.5837	213.8860	127.8870

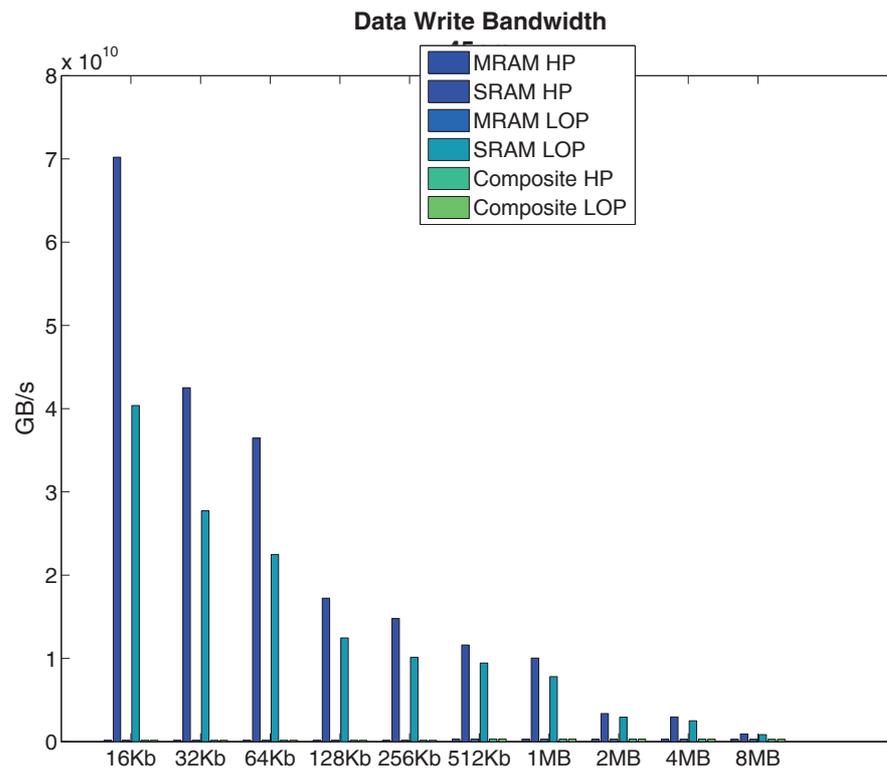


Figure 4.37: Write Bandwidth.

Table 4.27: Write Bandwidth (GB/s)

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	1.488	65.38	0.1486	37.60	0.1488	0.1486
32KB	1.487	39.60	0.1486	25.83	0.1487	0.1486
64KB	1.485	33.96	0.1484	20.94	0.1485	0.1484
128KB	0.1481	16.05	0.1480	11.61	0.1481	0.1480
256KB	0.1480	13.77	0.1479	9.442	0.1480	0.1479
512KB	0.2933	10.82	0.2928	8.796	0.2933	0.2928
1MB	0.2891	9.347	0.2884	7.269	0.2891	0.2884
2MB	0.2798	3.149	0.2786	2.741	0.2798	0.2786
4MB	0.2778	2.767	0.2765	2.339	0.2778	0.2765
8MB	0.2563	0.8649	0.2540	0.7880	0.2563	0.2540

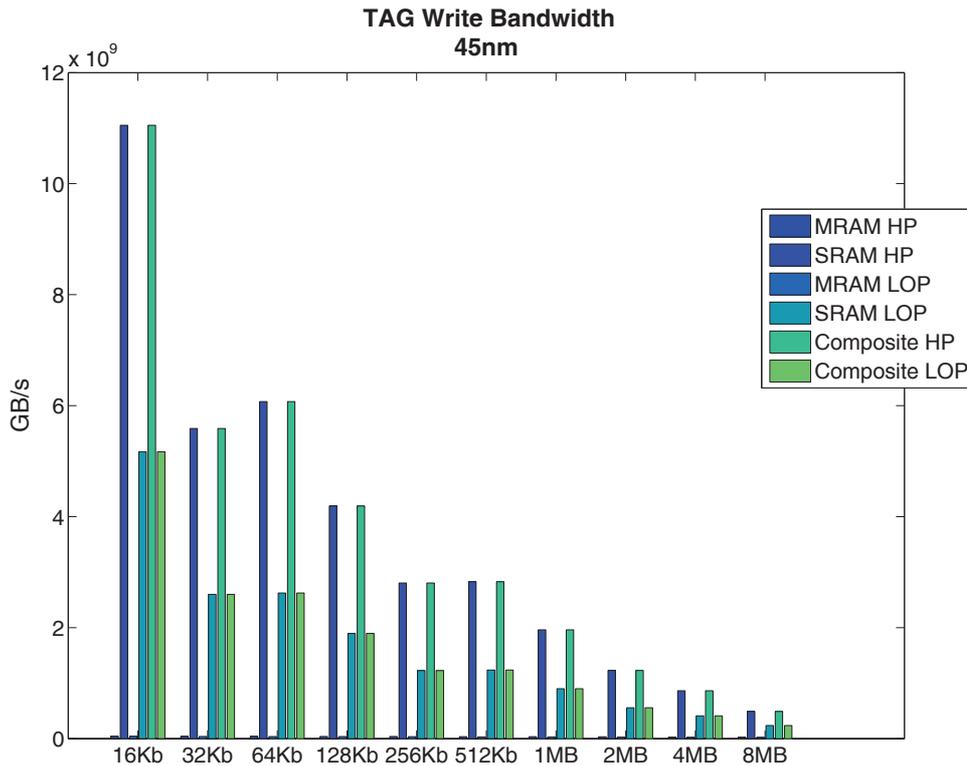


Figure 4.38: Write Bandwidth.

Table 4.28: Write Bandwidth (GB/s)

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.04049	10.29	0.04047	4.816	10.29	4.816
32KB	0.03952	5.204	0.03632	2.421	5.204	2.421
64KB	0.03855	5.656	0.03550	2.443	5.656	2.443
128KB	0.03757	3.904	0.03466	1.766	3.904	1.766
256KB	0.03658	2.613	0.03383	1.142	2.613	1.142
512KB	0.03354	2.634	0.02917	1.151	2.634	1.151
1MB	0.03272	1.825	0.02856	0.8387	1.825	0.8387
2MB	0.03190	1.144	0.02793	0.5179	1.144	0.5179
4MB	0.02618	0.8042	0.02122	0.3830	0.8042	0.3830
8MB	0.02563	0.4558	0.02086	0.2195	0.4558	0.2195

In this section we present our results for exploring the design space of microcontrollers and embedded microprocessors. For the microcontrollers we used a simulator called SimpleSCALAR, that emulates scalar architectures with instruction accuracy. The objective was demonstrate that using the advantage of density, even with write latencies drawbacks, the system can obtain a better performance than a SRAM based system for

some specific algorithms. We also evaluated CACHE L2 with sizes from 2GB decreasing to 128KB, alternating the latency in each size, and we noticed two things: the first latency in L2 is not the biggest impact factor, for each specific size the switching in latencies did not create any great effect. The size in other hand played its role, but to a certain extent. This observation about the latency effect is later confirmed by the experiment calibrating the Gem5 with the Latencies of SRAM and MRAM and performing the same algorithm in the same manner, with memory banks of same size.

We also, based into the intrinsic analyses results calibrate a Gem5 to operate its CACHE L2 at the specific latencies for 2MB, and depicted the results, the conclusion was switching a memory bank of ~ 20 ns by one of ~ 75 ns of latency, the effect in total time execution was less than a second of processing time for an X.264 encoder algorithm with 8 threads in parallel. Therefore, you could replace the CACHE L2 on existing embedded systems without major problems, even with this great difference in latency.

After we presented the initial results using a of-the-shelf EDA system to synthesize a embedded microprocessor using both characterized memory technologies to observe the area results. The power details were left behind, since the SRAM library provided did not have the leakage for each cell, while the MRAM had all details, the resulting synthesis in the reports than had power analyses that could not be well compared. Is possible that new releases or a modified SRAM library solve the problem for future works.

In this section we also presented the Composite Memory Bank (CMB). As stated before the CMB was born of the necessity to achieve a certain trade-off, among performance, latency and power, using MRAM, without losing the advantages of MRAM like area and low leakage. The biggest drawback so far of MRAM is the high current to switch the magnetic orientation of the free-layer. Besides a high current, some times a pulse with high period is necessary to switch the FL orientation. Despite that, the reading operation is faster than SRAM, also consumes much less energy. The leakage is unbeatable, the only way to compete on leakage with MRAM is with another emerging technology like PCRAM or ReRAM.

Part V

CONCLUSIONS & FUTURE INSIGHTS

Here in this chapter we discuss the conclusions of this thesis. Also, to point the directions in the near, middle and long-term or the continuing research related or based into this wok.

CONCLUSIONS & FUTURE INSIGHTS

*We are at the very beginning of time for the human race.
It is not unreasonable that we grapple with problems.
But there are tens of thousands of years in the future.
Our responsibility is to do what we can,
learn what we can, improve the solutions, and pass them on.
— Richard Phillips Feynman*

This thesis, as stated in [Chapter 1](#), has the objective, to evaluate the adoption of [MRAM](#) in memory hierarchy of computer architecture. For such purpose, a series of tools were combined to create an analytical methodology. Such analytical methodology was employed and its results presented in [Section 3.6](#). Also, we performed a series of architectural analyses demonstrated in [Section 3.5.3](#), [Section 4.1](#), [Section 4.2](#) and [Section 4.2.1](#).

The first contribution of this thesis is to be one of the few focused on [MRAM](#) applied to memory hierarchy. For this, we performed an in-depth study on the memory hierarchy, evaluating the effects on the overall performance of a [SoC](#), based on pure [MRAM](#) or combining it with [SRAM](#), in different levels of memory hierarchy as demonstrated in [Section 3.6](#).

Second contribution, is the characterization process of power for a system based on [MRAM](#) on [CACHE L2](#). In this contribution we evaluated all possible aspects that [L2](#) could influence, regarding power on a system adopting a [L2 MRAM](#) based. Also, the performance aspects and how they compare to existing [SRAM](#) were evaluated, as demonstrated in [Section 3.5.3](#).

The third contribution is the result on how to combine the devised methods employed during the research, to create an analytical methodology with well defined algorithm, in order to evaluate a system power consumption and performance. At the present this methodology is based on three tools, but replacing these tools by similar ones and keeping the flow would lead to similar results. It is demonstrated in [Section 3.6](#).

The fourth contribution was the characterization of the memory banks to use in a of-the-shelf [EDA](#), [IC](#) flow. This is probably the right direction to expend the efforts evaluating [MRAM](#) in memory hierarchy. It is demonstrated in [Section 4.4](#)

Finally, after several tests, observing how a cache memory bank timing, power and area is defined, we devised the so called Composite Memory Bank ([CMB](#)). The idea is combining the two technologies, but not in the way it has previously been proposed. We obtained a memory bank, given some tradeoff, with a reasonable area and leakage reduction as demonstrated in [Section 4.5](#), where it was placed besides [MRAM](#) and [SRAM](#) banks to denote its strengths.

The first conclusion achieved is: economically, is better to have the same SoC, using different memory (MRAM), with the same amount of memory, since we can decrease the silicon die area, due to the fact that MRAM has four folding in density. Leading, this way, to an increase in the number of dies per wafer, instead of keeping the same number of dies in the wafer with four times more memory per die. This can lead, not to a cost reduction of the final integrated circuit, but to a profit margin increase of, at least, four folding, given manufacturing cost reduction, since we could add four times more SoC dies in the same wafer. Regarding economical and industrial aspects, this is much more relevant than increasing the amount of memory and keep the silicon die in the same area of a SRAM circuit counterpart. The next generation of 450mm wafers, associated with this possible outcome, can generate an outstanding profit margin in memory and SoC industries. Also, is desirable to have additional logic blocks in every new release of an integrated circuit. For every nano square we can harvest, to integrate additional functionality, like integrated digital baseband radio, is welcome. In this sense, MRAM could provide a great advantage to new circuits.

In the same sense, the increase in density can allow the improvement of embedded system, given the four fold in embedded memory. As was demonstrated in Section 4.2, despite of different latencies on CACHE L2, or higher, in the memory hierarchy, the density plays a major role than latency. Also, highly embedded systems like microcontrollers (e. g., those used in automotive systems) will benefit of this increase in memory, and power consumption reduction.

Also STT-MRAM, can be a great replacement to SRAM in memory hierarchy using pMTJ. However, we can list two drawback regarding STT-MRAM: the pMTJ manufacturing process is still not mature. Also, integrate a magnetic fabrication on top of standard CMOS is not an easy task, given contaminants and α particles generation. Only a few companies have this manufacturing skill and expertise to assimilate this technology. Assuming that this drawbacks are solved, MRAM can successfully replace SRAM. This would lead to two great openings: the first, SoC with reduced leakage current, and the second, memory hierarchy architecture could assume a totally new design to explore, even further, the retention of MTJs.

I truly believe that the expression **Universal Memory** could not be worse used. During the course of the research, what I noticed, in fact, is that SRAM seems to be replaceable by MRAM. The ReRAM have a better chance to succeed in replacing the FLASH technologies. For DRAM, I believe the market will fragment in order to replace it: in extreme performance for MRAM and high density for RRAM.

Although MRAM is an outstanding technology, RRAM is proving to be as good as MRAM or an even better technology. Also, it is easier to integrate into CMOS manufacturing process compared to MRAM. MRAM has better performance, and infinite endurance compared to RRAM. That is the reason I say there is not such thing as Universal Memory. I honestly believe that, after 2018, MRAM and RRAM will be the new dominant and concurrent technologies.

Based on the present work I can have a better perspective in the short, medium and long term of the memory field, regardless the emerging memory technology. In the short-term, I believe that pMTJ and MRAM will dominate the scenery for the next two years. Most of the semiconductors companies involved with MRAM are exploring this solution, specifically for performance and reduced power consumption. Specific sectors in demand for memory are, currently, Big-Data and Cloud computing. Also this trend

of "Cloud" is irreversible: the ubiquitous access to information is becoming every day more prominent, and information will always be stored in some kind of memory. The aspect of Big-Data, demands the ability to store and analyze huge amounts of data at once, or continuously, like running Gaussian Mixture Models composed by mixtures of more than a thousand Gaussians, which is pure HPC over extended amounts of data. This demands lots of memory for processing and it seems that will keep evolving in the coming years.

Also, the current state of SRAM, FLASH and DRAM are reaching its physical limits: the president of Micron, Mark Adams, already said publicly that DRAM has only another three years on the run. No company, neither research facility, is investing time or resources into scaling down DRAM technology, since we reached the full potential of this technology and its limits. For that reason, the industry is looking into a solution for mass production within three years window frame. MRAM so far has the more mature manufacturing process and research available, exception for the pMTJ that is incipient and very few players has the knowledge to build it. I truly believe MRAM will be first solution to reach the mass market. Some companies already have prototypes of systems adopting L3 CACHES based in MRAM. Other companies are already employing MRAM as main memory into its products: Honeywell is commercializing MRAM memory modules for aerospace applications, which is clearly in the same perspective proposals of the ANR-MARS project. The adoption of STT-MRAM using pMTJ, will increase the presence of MRAM, since MRAM is already used in some cases. The only limiting factor for widespread adoption so far is the high write current and specially the high period of the write pulse, that degrades the performance of the memory, issues circumvented by the pMTJ.

I share the opinion diffused among researchers that the future of MRAM resides in STT. Also, I believe that the future of STT applied to computer architecture is intrinsically linked to the success of the Perpendicular MTJ. In my opinion, the state of the art on pMTJ research belongs to the research group lead by professor Hideo Ohno of Tohoku University.

Scientifically and architecturally, I presume that the middle to long term research and development of my thesis theory, not just for memories, will be extrapolated in two tracks. The first is the Spintronics logic, to evaluate the creation of logic using the principles of STT. Instead of having electrons flowing through the CMOS, we have a series of magnetic devices implementing logic through the interactions of its anisotropy. The second track would be to evaluate and build ReRAM banks. Is my understanding that ReRAM will be the biggest contender to face MRAM. The reason for so much speculation on ReRAM is that it is easier to integrate into the manufacturing flow, you do not need magnetic process on top of CMOS, like MRAM. Also, anyone that masters the technology and science behind ReRAM, will be able to build ReRAM cells. Regarding my research work, it can be employed to ReRAM research development as it was for MRAM.

There is also the concept of Memristor, enforced by Hewlett-Packard. This is another trend where both MRAM and ReRAM can fit in. Furthermore, magnetic domain walls associated with race-track architecture could help to create cross-point memories based in MRAM. Another future consideration would be explore the concept of memory banks architecture adopted by NVIDIA to build the L2 of Kepler architecture. Its basically L2 cache where the memory arrays, instead of been interconnect in a H-Tree, are intercon-

nected using Network-on-Chip (NoC). This way you could, using the existing microprocessors architecture state, interconnect more than four cores and keep the bandwidth for each one, using a much less complex structure than the cross-point matrix used in the OpenSPARC. I'm not saying a multiple cores architecture interconnect by a NoC, instead I'm suggesting a sea of memory arrays interconnected using a NoC. In this sense to improve a NoC based memory bank, would be interesting incorporate optoelectronics to build the interconnection channels of the network like optic fibers on-chip. One last idea I would like to pursue is pMTJ built in vertical transistors: I believe this could lead to memory banks with an outstanding density in the near future. Also, I'm really interested in how to use this technology associated with pMTJs.

Seeing that knowledge and technology shifts completely in every twenty four months, this thesis, as I mentioned before, will be a forgotten milestone in some time in the future. In the meantime I hope it can contribute to future research in the field. Using the techniques and data provided in this thesis, researchers can compare the prospects of MRAM with ReRAM in a seamless manner. Furthermore, the methods demonstrated here can be applied for any new emerging memory technology.

Part VI

APPENDIX

MRAM TERMINOLOGY

A.1 ANISOTROPY

Anisotropy is the property of being directionally dependent, as opposed to isotropy, which implies identical properties in all directions. It can be defined as a difference, when measured along different axes, in a material's physical or mechanical properties (absorbance, refractive index, conductivity, tensile strength, etc.) An example of anisotropy is the light coming through a polarizer. An example of an anisotropic material is wood, which is easier to split along its grain than across its grain [Kocks et al., 2000].

A chemical anisotropic filter, as used to filter particles, is a filter with increasingly smaller interstitial spaces in the direction of filtration so that the proximal regions filter out larger particles and distal regions increasingly remove smaller ones, resulting in greater flow-through and more efficient filtration [Kocks et al., 2000].

In NMR spectroscopy, the orientation of nuclei with respect to the applied magnetic field determines their chemical shift. In this context, anisotropic systems refer to the electron distribution of molecules with abnormally high electron density, like the π system of benzene. This abnormal electron density affects the applied magnetic field and causes the observed chemical shift to change [Kocks et al., 2000].

Physicists use the term anisotropy to describe direction-dependent properties of materials. Magnetic anisotropy, for example, may occur in a plasma, so that its magnetic field is oriented in a preferred direction. Plasmas may also show "filamentation" (such as that seen in lightning or a plasma globe) that is directional [Kocks et al., 2000].

A.2 MAGNETIC ANISOTROPY

Magnetic anisotropy is the directional dependence of a material's magnetic properties. In the absence of an applied magnetic field, a magnetically isotropic material has no preferential direction for its magnetic moment, while a magnetically anisotropic material will align its moment with one of the easy axes. An easy axis is an energetically favorable direction of spontaneous magnetization that is determined by the sources of magnetic anisotropy listed below. The two opposite directions along an easy axis are usually equivalent, and the actual direction of magnetization can be along either of them (see spontaneous symmetry breaking).

Magnetic anisotropy is a prerequisite for hysteresis in ferromagnets: without it, a ferromagnet is superparamagnetic [Aharoni, 2000].

There are several sources of magnetic anisotropy:

- Magnetocrystalline anisotropy: the atomic structure of a crystal introduces preferential directions for the magnetisation.
- Shape anisotropy: when a particle is not perfectly spherical, the demagnetizing field will not be equal for all directions, creating one or more easy axes.
- Magnetoelastic anisotropy: tension may alter magnetic behaviour, leading to magnetic anisotropy.
- Exchange anisotropy: a relatively new type that occurs when antiferromagnetic and ferromagnetic materials interact [Meiklejohn and Bean, 1957]

A.3 SUPERPARAMAGNETISM

Superparamagnetism is a form of magnetism, which appears in small ferromagnetic or ferrimagnetic nanoparticles. In sufficiently small nanoparticles, magnetization can randomly flip direction under the influence of temperature. The typical time between two flips is called the Néel relaxation time. In the absence of external magnetic field, when the time used to measure the magnetization of the nanoparticles is much longer than the Néel relaxation time, their magnetization appears to be in average zero: they are said to be in the superparamagnetic state. In this state, an external magnetic field is able to magnetize the nanoparticles, similarly to a paramagnet. However, their magnetic susceptibility is much larger than the one of paramagnets.

A.4 MAGNETIC PERMEABILITY

In electromagnetism, permeability is the measure of the ability of a material to support the formation of a magnetic field within itself. In other words, it is the degree of magnetization that a material obtains in response to an applied magnetic field. Magnetic permeability is typically represented by the Greek letter μ . The term was coined in September, 1885 by Oliver Heaviside. The reciprocal of magnetic permeability is magnetic reluctivity.

In SI units, permeability is measured in henries per meter ($\text{H} \cdot \text{m}^{-1}$), or newtons per ampere squared ($\text{N} \cdot \text{A}^{-2}$). The permeability constant (μ_0), also known as the magnetic constant or the permeability of free space, is a measure of the amount of resistance encountered when forming a magnetic field in a classical vacuum. The magnetic constant has the exact value of $\mu_0 = 4\pi \times 10^{-7} \text{H} \cdot \text{m}^{-1} \approx 1.2566370614 \dots \times 10^{-6} \text{H} \cdot \text{m}^{-1}$ or $\text{N} \cdot \text{A}^{-2}$). A closely related property of materials is magnetic susceptibility, which is a measure of the magnetization of a material in addition to the magnetization of the space occupied by the material.

A.5 ISOTROPY

Isotropy is uniformity in all orientations; it is derived from the Greek isos (equal) and tropos (way). Precise definitions depend on the subject area. Exceptions, or inequalities, are frequently indicated by the prefix an, hence anisotropy. Anisotropy is also used to describe situations where properties vary systematically, dependent on direction. Isotropic

radiation has the same intensity regardless of the direction of measurement, and an isotropic field exerts the same action regardless of how the test particle is oriented.

In Electromagnetic, An isotropic medium is one such that the permittivity, ϵ , and permeability, μ , of the medium are uniform in all directions of the medium, the most simple instance being free space.

A.6 MAGNETIC MOMENT

The magnetic moment of a magnet is a quantity that determines the force that the magnet can exert on electric currents and the torque that a magnetic field will exert on it. A loop of electric current, a bar magnet, an electron, a molecule, and a planet all have magnetic moments. Both the magnetic moment and magnetic field may be considered to be vectors having a magnitude and direction. The direction of the magnetic moment points from the south to north pole of a magnet. The magnetic field produced by a magnet is proportional to its magnetic moment as well. More precisely, the term magnetic moment normally refers to a system's magnetic dipole moment, which produces the first term in the multipole expansion of a general magnetic field. The dipole component of an object's magnetic field is symmetric about the direction of its magnetic dipole moment, and decreases as the inverse cube of the distance from the object.

A.7 MAGNETIC MOMENT AND ANGULAR MOMENTUM

The magnetic moment has a close connection with angular momentum called the gyromagnetic effect. This effect is expressed on a macroscopic scale in the Einstein-de Haas effect, or "rotation by magnetization," and its inverse, the Barnett effect, or "magnetization by rotation" [Cullity and Graham, 2011]. In particular, when a magnetic moment is subject to a torque in a magnetic field that tends to align it with the applied magnetic field, the moment precesses (rotates about the axis of the applied field). This is a consequence of the angular momentum associated with the moment. Viewing a magnetic dipole as a rotating charged sphere brings out the close connection between magnetic moment and angular momentum. Both the magnetic moment and the angular momentum increase with the rate of rotation of the sphere. The ratio of the two is called the gyromagnetic ratio, usually denoted by the symbol γ [Buxton, 2002; Krey and Owen, 2007].

For a spinning charged solid with a uniform charge density to mass density ratio, the gyromagnetic ratio is equal to half the charge-to-mass ratio. This implies that a more massive assembly of charges spinning with the same angular momentum will have a proportionately weaker magnetic moment, compared to its lighter counterpart. Even though atomic particles cannot be accurately described as spinning charge distributions of uniform charge-to-mass ratio, this general trend can be observed in the atomic world, where the intrinsic angular momentum (spin) of each type of particle is a constant: a small half-integer times the reduced Planck constant \hbar . This is the basis for defining the magnetic moment units of Bohr magneton (assuming charge-to-mass ratio of the electron) and nuclear magneton (assuming charge-to-mass ratio of the proton).

A.8 EXCHANGE BIAS

Exchange bias or exchange anisotropy occurs in bilayers (or multilayers) of magnetic materials where the hard magnetization behavior of an antiferromagnetic thin film causes a shift in the soft magnetization curve of a ferromagnetic film. The exchange bias phenomenon is of tremendous utility in magnetic recording, where it is used to pin the state of the readback heads of hard disk drives at exactly their point of maximum sensitivity; hence the term "bias."

A.9 ANTIFERROMAGNETISM

In materials that exhibit antiferromagnetism, the magnetic moments of atoms or molecules, usually related to the spins of electrons, align in a regular pattern with neighboring spins (on different sublattices) pointing in opposite directions. This is, like ferromagnetism and ferrimagnetism, a manifestation of ordered magnetism. Generally, antiferromagnetic order may exist at sufficiently low temperatures, vanishing at and above a certain temperature, the Néel temperature (named after Louis Néel, who had first identified this type of magnetic ordering) [Nee]. Above the Néel temperature, the material is typically paramagnetic.

A.9.1 *Measurement*

When no external field is applied, the antiferromagnetic structure corresponds to a vanishing total magnetization. In an external magnetic field, a kind of ferrimagnetic behavior may be displayed in the antiferromagnetic phase, with the absolute value of one of the sublattice magnetizations differing from that of the other sublattice, resulting in a nonzero net magnetization. Although the net magnetization should be zero at a temperature of absolute zero, the effect of spin canting often causes a small net magnetization to develop. The magnetic susceptibility of an antiferromagnetic material typically shows a maximum at the Néel temperature. In contrast, at the transition between the ferromagnetic to the paramagnetic phases the susceptibility will diverge. In the antiferromagnetic case, a divergence is observed in the staggered susceptibility. Various microscopic (exchange) interactions between the magnetic moments or spins may lead to antiferromagnetic structures. In the simplest case, one may consider an Ising model on a bipartite lattice, e.g. the simple cubic lattice, with couplings between spins at nearest neighbor sites. Depending on the sign of that interaction, ferromagnetic or antiferromagnetic order will result. Geometrical frustration or competing ferromagnetic and antiferromagnetic interactions may lead to different and, perhaps, more complicated magnetic structures.

A.9.2 *Antiferromagnetic materials*

Antiferromagnetic materials occur commonly among transition metal compounds, especially oxides. Examples include hematite, metals such as chromium, alloys such as iron manganese (FeMn), and oxides such as nickel oxide (NiO). There are also numerous examples among high nuclearity metal clusters. Organic molecules can also

exhibit antiferromagnetic coupling under rare circumstances, as seen in radicals such as 5-dehydro-m-xylylene. Antiferromagnets can couple to ferromagnets, for instance, through a mechanism known as exchange bias, in which the ferromagnetic film is either grown upon the antiferromagnet or annealed in an aligning magnetic field, causing the surface atoms of the ferromagnet to align with the surface atoms of the antiferromagnet. This provides the ability to "pin" the orientation of a ferromagnetic film, which provides one of the main uses in so-called spin valves, which are the basis of magnetic sensors including modern hard drive read heads. The temperature at or above which an antiferromagnetic layer loses its ability to "pin" the magnetization direction of an adjacent ferromagnetic layer is called the blocking temperature of that layer and is usually lower than the Néel temperature.

A.9.3 *Geometric frustration*

Unlike ferromagnetism, anti-ferromagnetic interactions can lead to multiple optimal states (ground states—states of minimal energy). In one dimension, the anti-ferromagnetic ground state is an alternating series of spins: up, down, up, down, etc. Yet in two dimensions, multiple ground states can occur. Consider an equilateral triangle with three spins, one on each vertex. If each spin can take on only two values (up or down), there are $2^3 = 8$ possible states of the system, six of which are ground states. The two situations which are not ground states are when all three spins are up or are all down. In any of the other six states, there will be two favorable interactions and one unfavorable one. This illustrates frustration: the inability of the system to find a single ground state. This type of magnetic behavior has been found in minerals that have a crystal stacking structure such as a Kagome lattice or hexagonal lattice.

Synthetic antiferromagnets (often abbreviated by SAF) are artificial antiferromagnets consisting of two or more thin ferromagnetic layers separated by a nonmagnetic layer. Due to dipole coupling of the ferromagnetic layers which results in antiparallel alignment of the magnetization of the ferromagnets. Antiferromagnetism plays a crucial role in giant magnetoresistance, as had been discovered in 1988 by the Nobel prize winners Albert Fert and Peter Grünberg (awarded in 2007) using synthetic antiferromagnets. There are also examples of disordered materials (such as iron phosphate glasses) that become antiferromagnetic below their Néel temperature. These disordered networks 'frustrate' the antiparallelism of adjacent spins; i.e. it is not possible to construct a network where each spin is surrounded by opposite neighbor spins. It can only be determined that the average correlation of neighbor spins is antiferromagnetic. This type of magnetism is sometimes called speromagnetism.

B

IMPROVING THE RELIABILITY OF A FPGA USING FAULT-TOLERANCE MECHANISM BASED ON MRAM

This contribution is not directly linked to the thesis core, but this experiments performed at the beginning of the thesis. It is interesting, the results are also elucidative, so we decide to include it as a contribution into the manuscript.

The current **SRAM** based **FPGA**, are more and more susceptible to Single Event Upset (**SEU**) caused by Neutron particles interference. The problem is exasperated reducing the **CMOS** submicronic scale in the manufacturing process, specially for the next generation of **SRAM**-based **FPGAs**. Nowadays is common practice for **SRAM** manufactures to embed fault tolerant mechanisms like Error Correcting Code (**ECC**) schemes in **SRAM** memory banks for **CMOS** technology below 90nm, to mitigate **SEU**. The present work proposes an approach to improve the reliability of the **FPGAs**, regarding **SEU** events at ground level for the future submicronic scale technologies proposing the adoption of **MRAMs** cells into a simple fault-tolerant system for **FPGAs** manufactured below 65nm submicronic scale.

The main problem addressed into this experiment, was investigate the reliability of **FPGA** regarding **SEU** inducing a 'bit-flip', due to neutron particles crossing the devices. **SRAM** based devices are facing this kind of problem at ground level, especially for **SRAM** based **FPGA**.

Our base-line device is the Virtex-5 110LXT, that according with [Lesea, 2009] whose configuration bits **SRAM** cells are designing using 90nm **CMOS** technology. For example the Virtex-5 suffer of a fault-rate of 140-FIT (one FIT equals 1 failure per 1 billion device hours) due to **SEU** on the configuration bits at ground level [Xilinx, 2010].

Errors due to **SEU** become an increase concern to manufacturers of **FPGA**. For instance Xilinx developed a program called Rosetta [Lesea et al., 2005] in which the effect of **SEU** is evaluated, measuring the neutron crossing particles on devices and observing the **SEU** events at specific locations and altitudes [Xilinx, 2010], [Tallerico, 1995] [Shea and Smart, 2010].

The objective presented in this work is to provide a mechanism permitting to the next generation of **FPGAs** to achieve a reliability level regarding **SEU** events of less than 40 FIT. We expect that our approach can help to achieve this goal, combining new technology like **MRAM** with a mechanism for error detection. Our approach is based in a mechanism using error detection coupled with scrubbing technique into the configuration bits of the **FPGA**.

B.1 RELIABILITY

The reliability of a semiconductor often relies in the following physical factors: oxide defects; silicon defects (bulk); Corrosion (aging); Assembly defects; Electromigration; Mask/Photoresist defects; contamination and charge injection.

Among the physical factors we are interested also in one specific transient fault, the SEU events. SEU events are unavoidable, we only expect that the fault is not induced by an energetic heavy ion striking the CMOS otherwise a transistor could locally transfer enough ionizing dose to affect permanently its physical electrical characteristics [Dufour et al., 1992].

Some companies are adding reliability mechanisms into the produced device to increase its on-field reliability. We can mention [Carmichael, 2006], it presents a classic example of fault-tolerance system, a device replicated based on a voter to decide the result. Also in [Brinkley and Carmichael, 2000], [Carmichael et al., 2000] and [Xilinx, 2005] it presents the use of the read-back mechanism first introduced in [HöFLICH, 1999], but the design had to explicitly verify if the configuration bits are still not affected by an SEU event. In [Tam, 2006] is presented a bit system based on Error Correcting Codes (ECC), the Hamming code for one and two errors detection and correction, but again it remains a design level oriented solution.

Our work proposes a solution at the architectural level of the FPGA, to improve the reliability of the device. We consider that if an SEU event occurs on the configuration bits, this would trigger a catastrophic failure on design level, the proposal is fix this error before the error propagates to the design level.

This work presents a combined technique based on MRAM technology and ECC technique. Our reliability modeling is based on the set of equations for reliability found in [Xilinx, 2010], [Speaks, 2005], [Vigrass, 2010], [DoD-USA, 1995], [Hybrid Memory Products Ltd, 1999] and [SEMATECH, 2000], to construct our model.

Specifically we are adopting the same set of equations found in [DoD-USA, 1995], to describe our reliability model.

The [DoD-USA, 1995] stated that the fault error rate is defined as in Equation B.1, which demands the die complexity failure rate (C_1), the package error rate (C_2), the read/write induced cycling failure rate (λ_{cyc}) only used for EEPROM otherwise is zero, also the environment factor (π_E), quality factor (π_Q), Learning factor (π_L) and the most important the acceleration factor (π_T), that is the Arrhenius equation given on Equation B.2.

$$\lambda = (C_1\pi_T + C_2\pi_E + \lambda_{cyc})\pi_Q\pi_L \quad (B.1)$$

Combining this parameters we calculate our λ , the failure rate for a SRAM based FPGA, considering the failure rate for SRAM cells.

$$AF = \pi_T = \exp \left\{ \frac{E_a}{k} \left(\frac{1}{T_{use}} - \frac{1}{T_{stress}} \right) \right\} \quad (B.2)$$

In Equation B.2 the k value is the Boltzmann constant, E_a is the thermal activation energy. The C_1 factor, is defined as $\chi^2(\alpha, 2f + 2)$, where χ^2 is a distribution, α is the confidence level on the distribution, $2f + 2$ is the free-parameters, that is the same as number of errors, defining the die complexity failure rate, T_{use} is the room temperature

and T_{Stress} is the temperature that will be used to stress the circuit during test, the other parameters are assumed as defined in [DoD-USA, 1995].

Equation B.3, provides the reliability according to time.

$$R(t) = e^{-\lambda * t} \quad (\text{B.3})$$

To calculate the reliability for corrosion we assume as thermal activation energy value found in [Vigrass, 2010] and [Hybrid Memory Products Ltd, 1999].

Our work is based on premises of [DoD-USA, 1995] [DoD-USA, 1999] and [DoD-USA, 1996], with data provided in [Xilinx, 2010] and [Lesea, 2009] [Hybrid Memory Products Ltd, 1999] to construct our model of reliability for a FPGA-SRAM based device. Also, we are using a χ^2 distribution with 30% of confidence level to model the random errors.

To modelize the relations of different distributions, we adopted the joint distribution to create the combined model. The joint distribution is defined by the Equation B.4.

$$R_{jd}(t) = R_{SEU}(t) * R_{aging}(t) \quad (\text{B.4})$$

This way we obtained a joint reliability model distribution, combining the two models with random variables.

Next we present how this reliability modeling will serve as guidance to propose a mechanism to improve the reliability regarding SEU. That is achieved combining two techniques.

B.2 MECHANISM TO IMPROVE RELIABILITY REGARDING SEU EVENTS

This section provides a description of our proposed method to improve the FPGA reliability. Majority of FPGAs at present are SRAM based. Non-volatility is a feature that opens several new opportunities for several devices including FPGAs by providing features like instant on capability and completely shutting down the device in standby, to save static power consumption. Flash is currently the most dominant Non-Volatile Memory (NVM) for both mass storage and embedded NVM.

The use of non-volatile memories such as MRAMs internally helps to overcome the drawbacks of classical SRAM based FPGAs without speed penalty. The MRAM permits new power consuming strategies, since there is no need to load the configuration data from an external non-volatile memory (e.g., instant-on and instant-off). The structure we use to achieve the restore operation is presented and detailed in [Guillemenet et al., 2009].

The Thermally Assisted Switching MRAM (TAS-MRAM) already employed in previous works like [Guillemenet et al., 2010], [Guillemenet et al., 2009] and [Guillemenet et al., 2008b], has shown improvements in term of required writing current and power consumption during write operation. More advanced writing schemes in the MTJ like Spin Transfer Torque (STT) allows further reduction of the required writing current and the die area [BRUCHON, 2007a].

Thermally Assisted Switching approach combines a local heating of the junction and a single low amplitude magnetic field. This writing method also requires several steps that are depicted in Figure B.1. When the junction is heated above the blocking temperature ($\sim 150^\circ\text{C}$) by a current ($I_{\text{heat}} \sim 340\mu\text{A}$) flowing through the junction, the

magnetization of the ferromagnetic layer is freed and can be reversed under the application of a single low amplitude magnetic field, this entire operation is performed in 35ns [Guillemenet et al., 2010].

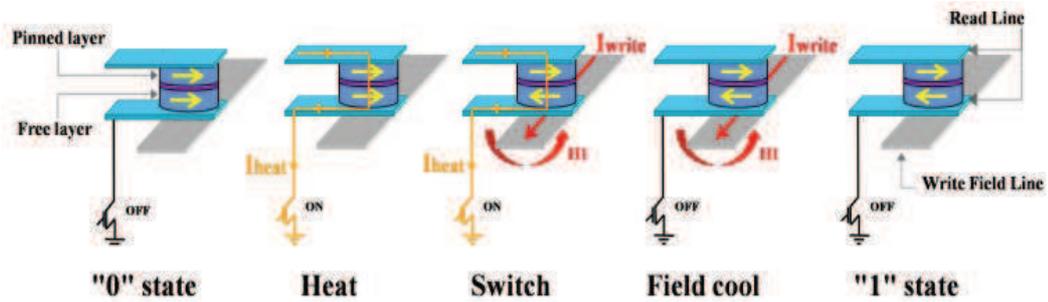


Figure B.1: MTJ process to switch the electromagnetic thin-layer field and store the logic value into the MTJ passing from the electrical layer.

The structure of the non-volatile TAS-MRAM cell (Figure B.2) consists of cross-coupled inverters (MN1 & MP1, MN2 & MP2), two MTJs for a non-volatile storage with complementary values to unbalance the latch during a read step. The writing line is implemented in a U shape such as allowing writing the data and its complement in the two MTJs of each MRAM cell. Two transistors MP3 and MP4 are driven on their gates by a signal *sense*, which act as **isolation** transistors (to preserve the data stored is the latch during the write stage). Selection transistors MN3 and MN4 are used to enable the heat operation.

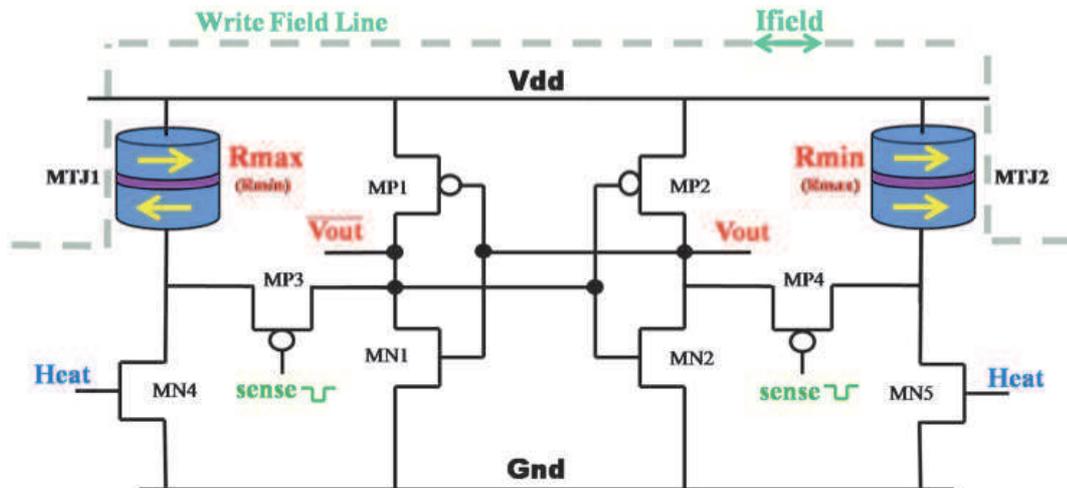


Figure B.2: TAS-MRAM non-volatile memory cell, with two complimentary MTJs.

This structure operates during the restore mode as follows: When the signal *sense* is at low logic level (0), transistors MP3 and MP4 are switched on allowing to pass a read current across MTJs until the latch. The intensity of these current are both different (I_{min} and I_{max}) because these MTJs are in opposite state (R_{max} and R_{min}). In consequence, a differential potential is generated at the boundary of the latch (V_{min} and V_{max}) unbalancing it. The restore operation phase takes less than 1ns to be performed and the necessary power for this operation is inferior to 0.1pJ per bit [Torres et al., 2010].

The main interest of this structure is to employ the [MRAM](#) cell as the storage cell for the configuration bits, directly storing the information in the memory cells, enhancing the configuration reliability of the FPGA. Also, using the [MRAM](#) we can easily update the configuration and implement the scrubbing mechanism [[Guillemenet et al., 2010](#)].

The main idea as described in [[Guillemenet et al., 2010](#)] is to design non-volatile [FPGA](#) architecture based on this cell.

B.3 SCRUBBING MECHANISM - REFRESH THE CONFIGURATION BITS

Our proposal to improve the reliability regarding [SEU](#) events, is to incorporate a correction mechanism based on refresh of the configuration bits, scrubbing the configuration bits from a non-volatile memory like [MRAM](#) [[Guillemenet et al., 2010](#)] to the [SRAM](#) configuration bits. In the [[Guillemenet et al., 2009](#)], [[Guillemenet et al., 2008b](#)] and [[Guillemenet et al., 2008a](#)] it is detailed how the [MRAM](#) can be employed to construct [MRAM](#) based FPGAs.

One simple usage of correction through scrubbing is a temporal refresh in intervals of time t (where t could be minutes, hours or even days). The configuration is refreshed reading the configuration bits for each frame, from a non-volatile [MRAM](#), for instance the configuration bits for each frame.

As mentioned the base-line device for our work is a Virtex – 5110LXT, with 31.2Mb configuration bits, this configuration bits are divided into 23 k frames of 1312 configuration bits [[Dutton and Stroud, 2009](#)].

[Figure B.3](#) depicts the refresh correction scheme in an interval of five days and how reliability is affected. So every five days the [SRAM](#) configuration bits are scrubbed from the [MRAM](#) non-volatile memory.

Applying a zoom in [Figure B.3](#) we can observe the reliability behavior depicted in [Figure B.4](#), it is possible to observe that errors related to [SEU](#), can be reduced scrubbing the configuration bits. It denotes an improving by a factor of 2. We only adopted corrosion as a physical factor, but the [SEU](#) would be limited by the joint reliability of physical factors, so the joint distribution of physical factors will determine the maximum limit of reliability for [SEU](#), we also assumed the Ground Fixed condition as described in [[DoD-USA, 1995](#)].

Based on the constructed reliability model, for the base-line device, we are proposing a more sophisticated method than only scrubbing with fixed intervals of time, aggregating Error-Correcting Codes capabilities for error detection. A refresh of all configurations bits in a FPGA in parallel seems an attractive solution, but very high consuming power.

In the [Section B.6](#) will present the results applying the error detection techniques. In this way, it is not necessary to apply a refresh in fixed intervals of time, but instead in every moment when the presence of an error into the configuration bits is detected.

B.4 GENERAL APPROACH

Our fault-tolerant [FPGA](#) is based on the first stage of a ECC, the Error-detection. The proposal is to adopt the error-detection in the [FPGA](#) at the architectural level. A parity code for each bit stream frame is generated. This parity is stored in a non-volatile

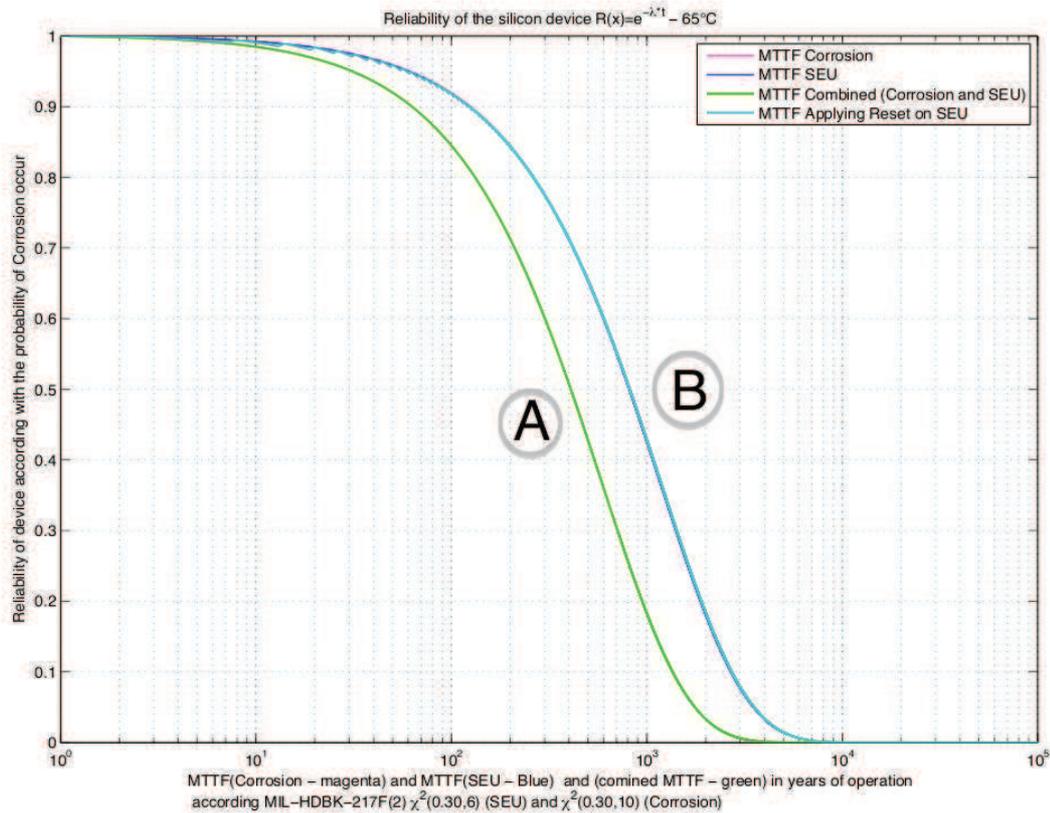


Figure B.3: The joint reliability of corrosion and SEU in green (A) and in blue (B) the reliability applying the scrubbing technique, improving the reliability regarding SEU in 2x.

memory, the MRAM. Remember that the structure of the FPGA is based on a matrix of frames.

As depicted into Figure B.5, each frame will be composed of the configuration bit frame plus the parity generated during synthesis. After the frame is configured with the respective configuration bits, the parity will be stored in a small MRAM memory bank in the frame. Now is possible to have at least two approaches, one will demand the addition of an ECC block in each frame on CMOS level. The other would be add a ECC block for each column or line in the FPGA architecture of frames. In this way it is possible to sequentially sweep each column for the matrix of frames composition, until the entire device is checked against SEU.

The ECC Block is composed by the input of the configuration bits and the parity, also the parity check Matrix if the chosen code is an Bose-Chaudhuri-Hocquenghem (BCH), in this case the parity Matrix, can be the same for all frames, so only one matrix could be stored in a MRAM memory bank.

So, after the ECC block received the input of configuration bits and parity, the system calculates the Syndrome. Internally in the ECC block, a logical OR reduction is applied over the Syndrome bit sequence. If the syndrome is zero no error is detected, otherwise the presence of an SEU error is detected [Sklar, 2000]. The OR reduction is applied to obtain only one signal as output. If the presence of an error is detected, the refresh mechanism is activated to perform a scrubbing of the configuration bits and fixing the SEU fault as depicted in Figure B.5.

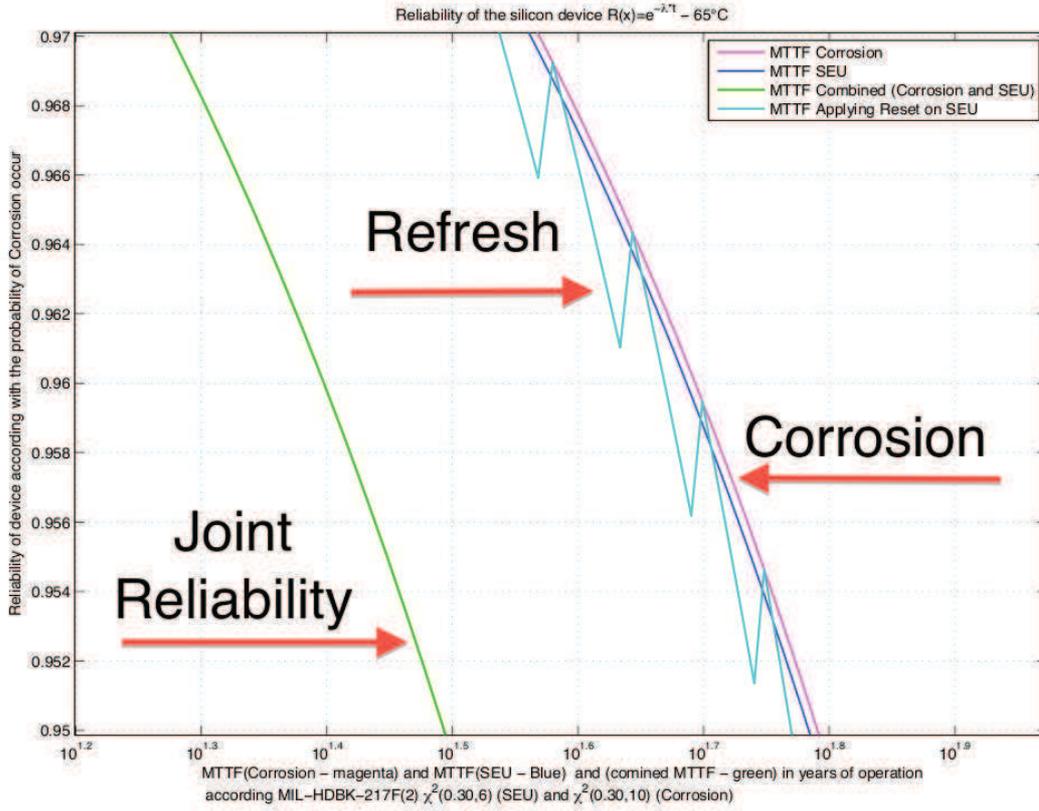


Figure B.4: The joint reliability of corrosion and SEU in green and in blue the reliability applying the refresh technique, improving the reliability regarding SEU over the joint reliability distribution in 2x.

B.5 ERROR DETECTION CODE

First approach is using the entire frame of 1312 bits, based on the C/C++ implementation available at [Zaragoza], we calculated the polynomial generator for the BCH codes. The necessary code using BCH to encode 1312 bits would be BCH(2047, 2024, 5).

Since for the BCH(2047, 2024, 5) would demand too many interconnections and memory elements, we adopted the Hamming code [Hamming, 1950] for the entire bit frame. We also partitioned the configuration bit frame using BCH code, for smaller BCH block codes. That way we could test two different codes for our error detection system, and compare the area and power consumption.

To the partitioning case the 1312 bits were divided into 5 groups of a BCH(255, 247, 1) plus one BCH(127, 78, 1). The polynomial for both codes were obtained from [Sklar, 2000], which consist, in the following polynomials BCH(255, 247, 1), with polynomial generator as defined in Equation B.5.

$$g(x) = x^8 + x^4 + x^3 + x^2 + 1 \tag{B.5}$$

Finally for the code BCH(127, 78, 1) the generator polynomial is denoted in Equation B.6.

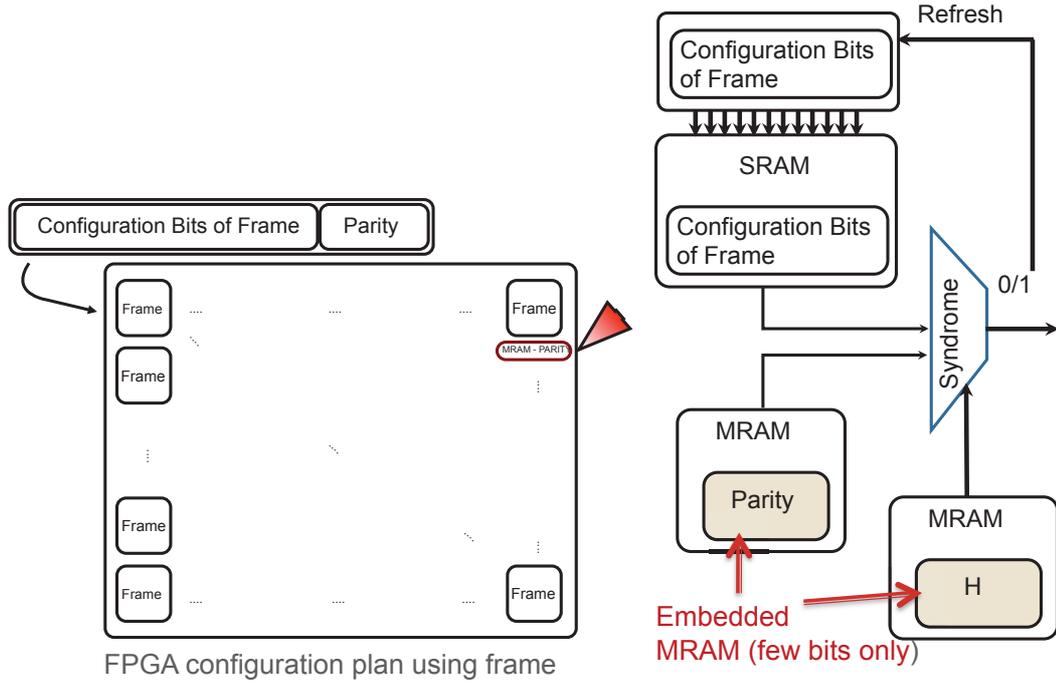


Figure B.5: An overview of our general approach.

$$\begin{aligned}
 g(x) = & x^{49} + x^{48} + x^{45} + x^{44} + x^{43} + x^{40} + x^{36} \\
 & + x^{34} + x^{32} + x^{31} + x^{30} + x^{28} + x^{27} + x^{26} \\
 & + x^{24} + x^{23} + x^{21} + x^{17} + x^{16} + x^{11} + x^{10} \\
 & + x^8 + x^5 + x^2 + 1
 \end{aligned} \tag{B.6}$$

For both Hamming and BCH codes, the results are presented in [Section B.6](#).

B.6 EXPERIMENTAL RESULTS

This section describes the results obtained after synthesizing the Hamming and BCH code using a 65nm CMOS technology. The [Table B.1](#) summarize the synthesis results obtained for area, power of the ECC devices.

The resulting total area and power estimative in [Table B.1](#) are based on standard information provided by the library, without considering placement neither routing.

B.7 HAMMING RESULTS

The Hamming code achieve an area of 0.02mm^2 , which can be considered as a small overhead per frame. One option is to implement only one Hamming code per column, and to perform the verification in slices, since we have 23K frames in the architecture they are pretty much spread in a matrix of 151×151 frames approximately, this will took us to a 151 stages of verification. This solution is more realistic, than placing in all frames one ECC block.

Table B.1: Summary of the ECC synthesis results using CMOS ST 65nm standard cell library.

Code	#Kcells	Area (μm^2)	Power Dynamic (W)	Power Leakage (W)	Worst Delay (ns)
Hamming(1312, 11)	3499	23882	3.6mW	1.7mW	1.2
BCH(255, 247) complete	4108	20542	2.7mW	1.3mW	1.3
BCH(255, 247) simplified	2052	10385	1.3mW	0.714mW	0.9
BCH(127, 78) complete	12494	62931	8.3mW	3.9mW	0.9
BCH(127, 78) simplified	3819	18660	2.7mW	1.3mW	0.9

The Hamming block achieve a better area ratio compared the BCH blocks. The power consumption compared among all ECCs is very similar, because the Hamming(1312, 11) has more input bits than the partitioned BCH codes. Probably smaller codes of Hamming can achieve a better power relation than the BCH codes. The last important characteristic to compare is the maximum delay that determines the maximum frequency of the circuit, if we round the values they are all pretty similar, so the penalty in using the Hamming is not significant.

B.8 BCH RESULTS

For the BCH codes we have four implementations. Two using the BCH(255, 247) and two using BCH(127, 78). The main difference resides that in one implementation (the simplified version) for the H^T matrix we removed all the logical *AND* between input data and the matrix in which positions of $H^T(l, c)$ the value is zero. We reduced the unnecessary connections and unnecessary logic.

As expected removing unnecessary connections reduced considerably the area, dynamic and leakage power, so would be interesting to adopt one specific code. The reduction in area and power are more evident, in the BCH(127, 78) simplified version.

Since the power and area difference from the Hamming to the smallest code the BCH(127, 78) in simplified version is almost minimal. The best approach should be employing a Hamming code to detect an error in the frame configuration bits, regarding area and power usage. In the partitioned code we proposed to use five BCH(255, 247) and one BCH(127, 78) per frame, so the total area and power is much higher than the single Hamming code, only for error detection.

For a circuit that demands a high reliability like space or military grade applications, we can assume a code like BCH should be more interesting, because it is capable to detect more errors. For ground level we are pretty confident that a code like Hamming can help to mitigate the problem of SEU.

B.9 ASSESSMENTS ON MRAM APPLIED TO FPGA TO IMPROVE RELIABILITY

A combined technique of error detection associated with **MRAM** could be used to enhance reliability of FPGA devices. We demonstrated that we can improve the reliability by different methods. Such as scrubbing, using an ECC code or combined techniques

as we implemented, or having a shadowing of the memory states in [MRAM](#), providing a rollback mechanism at the design level, restoring the context of the execution flow to the state before the detection of the error.

C

INTRINSIC ANALYSES - ADDITIONAL RESULTS

Table C.1: Details about SRAM and MRAM memory banks, generated using the NVSim, this table comprises the 45nm results of the memory banks for HP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	0.047	0.053	0.064	0.086	0.108	0.276	0.390	0.618	0.992	1.828
Latency(ns)										
Write (ns)	109.339	109.071	108.802	108.562	108.297	118.140	117.515	116.906	137.975	136.412
Hit (ns)	10.295	10.028	9.760	9.525	9.335	19.351	18.743	18.202	40.044	38.740
Miss (ns)	10.187	9.920	9.652	9.416	9.167	18.993	18.385	17.842	38.922	37.615
Power (operation)										
Write (nJ)	0.211	0.210	0.211	0.216	0.219	0.224	0.237	0.266	0.274	0.334
Read (nJ)	0.041	0.041	0.040	0.039	0.044	0.168	0.166	0.165	0.192	0.191
leakage (mW)	2.285	3.662	6.401	11.893	12.479	13.007	23.968	45.821	47.515	91.178
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	0.057	0.109	0.213	0.414	0.820	1.446	2.871	5.665	11.270	22.343
Latency(ns)										
Write (ns)	0.531	0.939	0.879	1.182	1.680	2.880	3.351	9.661	10.991	34.743
Hit (ns)	0.653	1.152	1.094	1.648	2.148	4.286	4.762	14.905	16.240	55.171
Miss (ns)	0.560	0.968	0.908	1.210	1.708	1.696	2.299	3.498	4.731	8.031
Power (operation)										
Write (nJ)	0.003	0.005	0.007	0.009	0.012	0.018	0.019	0.033	0.035	0.062
Read (nJ)	0.014	0.026	0.044	0.077	0.156	0.281	0.505	1.059	1.951	4.099
leakage (mW)	14.113	26.227	51.634	99.421	196.143	340.770	676.946	1326.800	2640.480	5218.480
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table C.2: Details about SRAM and MRAM memory banks, generated using the NVSim, this table comprises the 28nm results of the memory banks for HP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	0.017	0.020	0.027	0.040	0.062	0.140	0.221	0.384	0.678	1.340
Latency(ns)										
Write (ns)	7.621	7.576	9.310	9.222	9.117	9.118	12.992	12.728	22.534	21.774
Hit (ns)	3.200	3.156	4.895	4.809	4.796	5.114	8.994	8.746	19.850	19.549
Miss (ns)	3.105	3.060	4.799	4.712	4.610	4.606	8.485	8.235	18.051	17.343
Power (operation)										
Write (nJ)	0.008	0.008	0.009	0.009	0.010	0.013	0.014	0.016	0.020	0.040
Read (nJ)	0.046	0.045	0.046	0.045	0.046	0.206	0.207	0.206	0.216	0.234
leakage (mW)	2.605	4.038	7.072	12.750	17.343	17.688	32.940	62.693	65.001	128.174
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	0.022	0.042	0.083	0.160	0.317	0.559	1.111	2.193	4.361	8.655
Latency(ns)										
Write (ns)	0.439	0.791	0.777	1.030	1.549	3.040	3.543	10.481	11.988	38.266
Hit (ns)	0.537	0.980	0.967	1.502	2.023	4.591	5.098	16.401	17.912	61.529
Miss (ns)	0.460	0.812	0.798	1.051	1.571	1.561	2.072	3.407	4.478	8.393
Power (operation)										
Write (nJ)	0.002	0.002	0.003	0.004	0.006	0.009	0.009	0.016	0.017	0.030
Read (nJ)	0.007	0.013	0.022	0.039	0.080	0.144	0.260	0.547	1.010	2.123
leakage (mW)	20.674	38.950	76.655	144.414	285.008	496.123	985.707	1934.100	3849.340	7627.120
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table C.3: Details about SRAM and MRAM DATA memory arrays, this table comprises the 45nm HP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	29564.415	34515.081	44320.923	64268.982	80945.352	233755.073	338221.081	544738.877	878804.297	1636940.364
Latency(ns)										
Write (ns)	100.234	100.287	100.403	100.697	100.759	101.717	103.195	106.634	107.405	116.449
Read (ns)	1.055	1.062	1.080	1.176	1.356	1.933	2.222	3.207	5.534	9.154
Power (operation)								†		
Write (nJ)	97.250	98.849	102.046	108.444	110.542	116.228	129.031	154.592	162.766	213.886
Read (nJ)	18.950	18.957	18.970	18.998	23.776	127.918	127.983	128.066	146.955	147.122
leakage (mW)	1.909	3.281	6.013	11.482	12.039	12.269	23.202	44.984	46.148	89.710
Bandwith										
Write (GB/s)	0.149	0.149	0.149	0.148	0.148	0.293	0.289	0.280	0.278	0.256
Read (GB/s)	15.529	13.711	10.980	7.590	7.525	15.085	8.644	4.029	4.020	1.559
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	44224.273	84400.304	166794.773	326569.230	650526.321	1276872.723	2544776.959	5036607.194	10055128.344	19999291.442
Latency(ns)										
Write (ns)	0.304	0.460	0.526	1.029	1.216	2.880	3.351	9.661	10.991	34.743
Read (ns)	0.304	0.460	0.526	1.029	1.216	2.880	3.351	9.661	10.991	34.743
Power (operation)										
Write (nJ)	2.037	3.320	3.585	6.106	6.640	12.609	13.673	22.975	25.100	43.695
Read (nJ)	10.592	20.404	34.446	67.931	124.010	249.059	473.180	942.233	1838.330	3672.200
leakage (mW)	11.218	20.737	40.876	79.037	156.785	301.365	601.357	1180.680	2358.780	4673.750
Bandwith										
Write (GB/s)	65.381	39.596	33.956	16.051	13.773	10.816	9.347	3.149	2.767	0.865
Read (GB/s)	76.185	58.611	37.020	26.196	15.907	22.191	12.543	8.169	4.357	2.800
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table C.4: Details about SRAM and MRAM DATA memory arrays, this table comprises the 28nm HP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	9681.361	12409.527	17841.633	28793.209	46721.622	119415.600	192518.511	338454.871	601743.182	1203490.690
Latency(ns)										
Write (ns)	6.156	6.166	6.186	6.251	6.343	6.775	7.027	7.603	8.902	9.105
Read (ns)	1.622	1.626	1.631	1.667	1.785	2.157	2.233	2.440	3.978	4.384
Power (operation)										
Write (nJ)	4.092	4.193	4.395	4.797	5.609	7.704	8.506	10.108	13.070	32.188
Read (nJ)	21.548	21.553	21.563	21.579	23.543	162.864	162.895	162.956	170.525	189.643
leakage (mW)	2.240	3.659	6.481	12.119	16.663	16.924	31.767	61.395	63.039	126.079
Bandwith										
Write (GB/s)	2.442	2.438	2.431	2.414	2.380	4.448	4.300	3.974	3.392	3.392
Read (GB/s)	9.984	9.871	9.640	9.140	9.065	18.200	16.029	11.889	11.829	11.829
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	17132.560	32964.472	65137.826	126213.827	251404.322	493912.864	984332.308	1949065.424	3891085.124	7747309.771
Latency(ns)										
Write (ns)	0.254	0.411	0.477	1.021	1.208	3.040	3.543	10.481	11.988	38.266
Read (ns)	0.254	0.411	0.477	1.021	1.208	3.040	3.543	10.481	11.988	38.266
Power (operation)										
Write (nJ)	0.966	1.561	1.699	2.869	3.146	5.947	6.501	10.930	12.035	20.909
Read (nJ)	5.254	10.144	17.441	34.438	63.576	127.642	244.092	486.357	951.959	1901.930
leakage (mW)	16.439	30.918	61.007	114.718	227.652	438.701	875.476	1720.980	3438.340	6832.300
Bandwith										
Write (GB/s)	75.334	42.498	36.089	15.759	13.447	10.123	8.718	2.886	2.523	0.784
Read (GB/s)	79.670	60.816	37.234	25.783	15.213	21.007	11.534	7.528	3.881	2.511
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table C.5: Details about SRAM and MRAM TAG memory arrays, this table comprises the 45nm HP

MRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	17837.030	18242.297	19380.238	22039.256	27553.185	41991.333	52224.984	73025.891	113138.102	191532.853
Latency(ns)										
Write (ns)	109.339	109.071	108.802	108.562	108.297	118.140	117.515	116.906	137.975	136.412
Read (ns)	10.187	9.920	9.652	9.416	9.167	18.993	18.385	17.842	38.922	37.615
Power (operation)										
Write (nJ)	113.864	111.340	109.240	107.957	108.278	107.672	107.823	110.967	111.573	119.970
Read (nJ)	22.143	21.574	21.006	20.433	19.873	39.627	38.497	37.373	45.481	44.070
leakage (mW)	0.376	0.382	0.388	0.411	0.440	0.737	0.765	0.837	1.367	1.468
Bandwith										
Write (GB/s)	0.040	0.040	0.039	0.038	0.037	0.034	0.033	0.032	0.026	0.026
Read (GB/s)	5.056	4.741	4.311	3.704	2.888	3.565	2.781	1.874	1.791	1.002
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM										
	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	12649.157	24239.685	45757.294	87916.563	169007.410	169187.532	326398.032	628898.476	1214677.996	2343892.247
Latency(ns)										
Write (ns)	0.531	0.939	0.879	1.182	1.680	1.668	2.270	3.470	4.703	8.004
Read (ns)	0.560	0.968	0.908	1.210	1.708	1.696	2.299	3.498	4.731	8.031
Power (operation)										
Write (nJ)	1.226	1.728	2.995	2.935	5.430	5.628	5.526	10.259	10.025	18.380
Read (nJ)	3.221	5.181	9.313	9.072	32.072	32.422	31.515	116.364	112.796	426.919
leakage (mW)	2.894	5.489	10.758	20.384	39.358	39.404	75.588	146.116	281.698	544.734
Bandwith										
Write (GB/s)	10.292	5.204	5.656	3.904	2.613	2.634	1.825	1.144	0.804	0.456
Read (GB/s)	11.126	5.167	2.736	2.444	0.982	0.985	0.894	0.311	0.287	0.088
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

Table C.6: Details about SRAM and MRAM TAG memory arrays, this table comprises the 28nm HP

MRAM	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	7190.643	7666.662	8712.797	10843.148	15131.301	20776.765	28919.689	45090.620	76166.852	136598.381
Latency(ns)										
Write (ns)	7.621	7.576	9.310	9.222	9.117	9.118	12.992	12.728	22.534	21.774
Read (ns)	3.105	3.060	4.799	4.712	4.610	4.606	8.485	8.235	18.051	17.343
Power (operation)										
Write (nJ)	4.295	4.244	4.570	4.553	4.639	4.884	5.553	5.792	7.110	7.633
Read (nJ)	24.334	23.701	24.149	23.484	22.824	43.549	44.269	42.950	45.343	43.910
leakage (mW)	0.365	0.380	0.591	0.631	0.679	0.765	1.173	1.298	1.962	2.096
Bandwith										
Write (GB/s)	0.585	0.573	0.453	0.446	0.438	0.438	0.298	0.295	0.161	0.161
Read (GB/s)	2.988	2.891	2.798	2.687	2.548	2.557	2.465	2.265	2.169	1.851
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256
SRAM	16K	32K	64K	128K	256K	512K	1M	2M	4M	8M
Surface	4904.042	9392.464	17707.874	34023.371	65416.029	65485.354	126334.881	243446.047	470201.468	907382.931
Latency(ns)										
Write (ns)	0.439	0.791	0.777	1.030	1.549	1.539	2.051	3.386	4.457	8.372
Read (ns)	0.460	0.812	0.798	1.051	1.571	1.561	2.072	3.407	4.478	8.393
Power (operation)										
Write (nJ)	0.596	0.857	1.507	1.477	2.754	2.840	2.787	5.211	5.092	9.413
Read (nJ)	1.596	2.616	4.755	4.632	16.545	16.697	16.229	60.245	58.399	221.529
leakage (mW)	4.234	8.032	15.648	29.695	57.356	57.422	110.231	213.120	411.006	794.821
Bandwith										
Write (GB/s)	12.134	6.067	6.207	4.390	2.779	2.798	1.996	1.158	0.842	0.431
Read (GB/s)	11.569	5.210	2.572	2.329	0.883	0.884	0.813	0.268	0.250	0.073
Associativity	4	4	4	4	4	8	8	8	8	8
line size bits	128	128	128	128	128	256	256	256	256	256

D

CACTI ANALYTICAL MODELS - RELEVANT DETAILS

D.1 CACTI ANALYTICAL MODELS

The [Appendix D](#) discuss the analytical delay and power models for different wires. All the process specific parameters required for calculating the transistor and wire parasitics are obtained from ITRS [[ITRS, a](#)], we can also take into account the ITRS [[ITRS, b](#)] to evolve the models.

D.2 WIRE PARASITICS

The resistance and capacitance per unit length of a wire is given by the following equations [[Ho et al., 2001](#)]:

$$R_{\text{wire}} = \frac{\rho}{d * (\text{thickness} - \text{barrier}) * (\text{width} - (2 * \text{barrier}))} \quad (\text{D.1})$$

where, $d (< 1)$ is the loss in cross-sectional area due to dishing effect [[ITRS, a](#)] and ρ is the resistivity of the metal.

To calculate the capacitance of the wire we use the following equation:

$$C_{\text{wire}} = \tau_0 \left(2K\tau_{\text{horiz}} \frac{\text{thickness}}{\text{spacing}} + 2\tau_{\text{vert}} \frac{\text{width}}{\text{layerspacing}} \right) + \text{fringe}(\tau_{\text{horiz}}, \tau_{\text{vert}}) \quad (\text{D.2})$$

the first term corresponds to the side wall capacitance, the second term models the capacitance due to wires in adjacent layers, and the last term corresponds to the fringing capacitance between the sidewall and the substrate.

D.3 GLOBAL WIRES

For a long repeated wire, the single pole time constant model for the interconnect fragment is given by,

$$\tau = \left(\frac{1}{l} r_s (c_0 + c_p) + \frac{r_s}{s} C_{\text{wire}} + R_{\text{wire}} s c_0 + (0.5 * R_{\text{wire}} * C_{\text{wire}} * l) \right) \quad (\text{D.3})$$

In [Equation D.3](#) c_0 is the capacitance of the minimum sized repeater, c_p is its output parasitic capacitance, r_s is its output resistance, l is the length of the interconnect segment between repeaters and s is the size of the repeater normalized to the minimum value. The values of c_0 , c_p , and r_s are constant for a given process technology. Wire parasitics R_{wire} and C_{wire} are the resistance and capacitance per unit length. The optimal

repeater sizing and spacing values can be calculated by differentiating Equation D.3 with respect to s and l .

$$L_{\text{optimal}} = \sqrt{\frac{2r_s(c_0 + c_p)}{R_{\text{wire}}C_{\text{wire}}}} \quad (\text{D.4})$$

$$S_{\text{optimal}} = \sqrt{\frac{r_s C_{\text{wire}}}{R_{\text{wire}} c_0}} \quad (\text{D.5})$$

$$(\text{D.6})$$

The delay value calculated using the Equation D.4 and Equation D.5 is guaranteed to have minimum value. The total power dissipated is the sum of three main components Equation D.7

$$P_{\text{total}} = P_{\text{switching}} + P_{\text{short-circuit}} + P_{\text{Leakage}} \quad (\text{D.7})$$

The dynamic and leakage components of the interconnect are computed using Equation D.9 and Equation D.10.

$$P_{\text{dynamic}} = \alpha V_{\text{DD}}^2 f_{\text{clock}} \left(\frac{S_{\text{optimal}}}{L_{\text{optimal}}} (c_p + c_0) + c \right) + (\alpha V_{\text{DD}} W_{\text{min}} I_{\text{SC}} f_{\text{clock}} \log_e 3) S_{\text{optimal}} \frac{\tau}{L_{\text{optimal}}} \quad (\text{D.8})$$

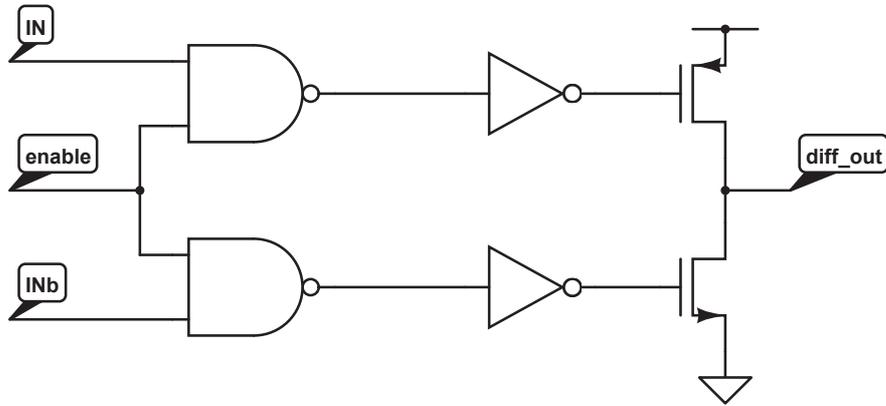


Figure D.1: Low-swing transmitter (actual transmitter has two such circuits to feed the differential wires).

f_{clock} is the operating frequency, W_{min} is the minimum width of the transistor, I_{SC} is the short-circuit current, and the value $(\tau/L)_{\text{optimal}}$ can be calculated from Equation D.9.

$$\left(\frac{\tau}{L} \right)_{\text{optimal}} = 2\sqrt{r_s c_0 r c} \left(1 + \sqrt{0.5 * \left(1 + \frac{c_p}{c_0} \right)} \right) \quad (\text{D.9})$$

$$P_{\text{leakage}} = \frac{3}{2} V_{\text{DD}} I_{\text{leak}} W_n S_{\text{optimal}} \quad (\text{D.10})$$

I_{leak} is the leakage current and W_n is the minimum width of the nMOS transistor.

With Equation D.9 and Equation D.10, CACTI compute the delay and power for global and semi-global wires. Wires faster than global wires can be obtained by increasing the wire width and spacing between the wires. Wires whose repeater spacing and sizing are different from ?? and ?? will incur a delay penalty. The actual calculation involves solving a set of differential equations [Banerjee and Mehrotra, 2002].

D.4 LOW-SWING WIRES

A low-swing interconnect system consists of three main components:

1. a transmitter that generates and drives the low-swing signal;
2. twisted differential wires;
3. a receiver amplifier.

D.5 TRANSMITTER

For an RC tree with a time constant τ , the delay of the circuit for an input with finite rise time is given by Equation D.11

$$\text{delay}_r = t_f \sqrt{\frac{\left[\log \frac{v_{\text{th}}}{V_{\text{dd}}} \right]^2 + 2t_{\text{rise}} b \left(1 - \frac{v_{\text{th}}}{V_{\text{dd}}} \right)}{t_f}} \quad (\text{D.11})$$

where, t_f is the time constant of the tree, v_{th} is the threshold voltage of the transistor, t_{rise} is the rise time of the input signal, and b is the fraction of the input swing in which the output changes (CACTI assumes b to be 0.5).

For falling input, the equation changes to Equation D.12

$$\text{delay}_f = t_f \sqrt{\left[\log \left(1 - \frac{v_{\text{th}}}{V_{\text{dd}}} \right) \right]^2 + \frac{2t_{\text{fall}} b v_{\text{th}}}{t_f V_{\text{dd}}}} \quad (\text{D.12})$$

where, t_{fall} is the fall time of the input. For the falling input, CACTI uses a value of 0.4 for b based on [Wilton and Jouppi, 1993]. To get a reasonable estimate of the initial input signal rise/fall time, we consider two inverters connected in series. Let d be the delay of the second inverter. The t_{fall} and t_{rise} values for the initial input can derivate as:

$$t_{\text{fall}} = \frac{d}{1 - v_{\text{th}}} \quad (\text{D.13})$$

$$t_{\text{rise}} = \frac{d}{v_{\text{th}}} \quad (\text{D.14})$$

For the transmitter circuit depicted in Figure D.1, CACTI employs the model proposed by [Ho et al., 2004]. The total delay of the transmitter is given Equation D.15:

$$t_{\text{delay}} = \text{nand}_{\text{delay}} + \text{inverter}_{\text{delay}} + \text{driver}_{\text{delay}} \quad (\text{D.15})$$

INTRINSIC ANALYSES - MEMORY BANKS OF 28nm

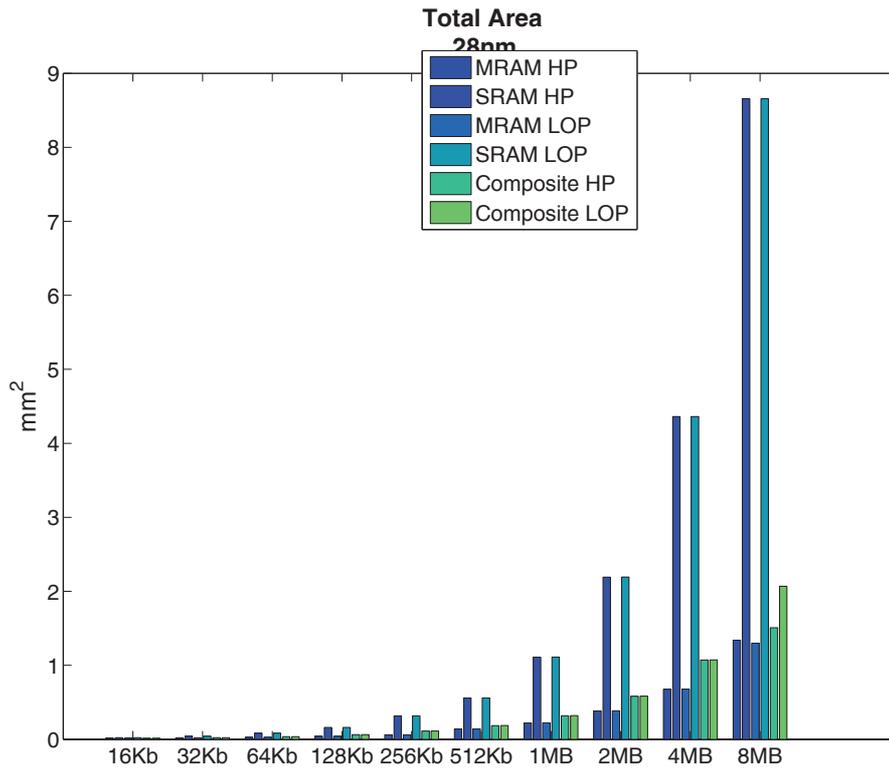


Figure E.1: Total Area.

Table E.1: Total Area (μm^2).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0169	0.0220	0.0169	0.0220	0.0146	0.0146
32KB	0.0201	0.0424	0.0201	0.0424	0.0218	0.0219
64KB	0.0266	0.0828	0.0266	0.0829	0.0355	0.0356
128KB	0.0396	0.1602	0.0397	0.1604	0.0628	0.0629
256KB	0.0619	0.3168	0.0622	0.3171	0.1121	0.1124
512KB	0.1402	0.5594	0.1411	0.5596	0.1849	0.1857
1MB	0.2214	1.1107	0.2225	1.1111	0.3189	0.3199
2MB	0.3835	2.1925	0.3850	2.1930	0.5819	0.5833
4MB	0.6779	4.3613	0.6799	4.3623	1.0719	1.0738
8MB	1.3401	8.6547	1.2984	8.6549	1.5091	2.0691

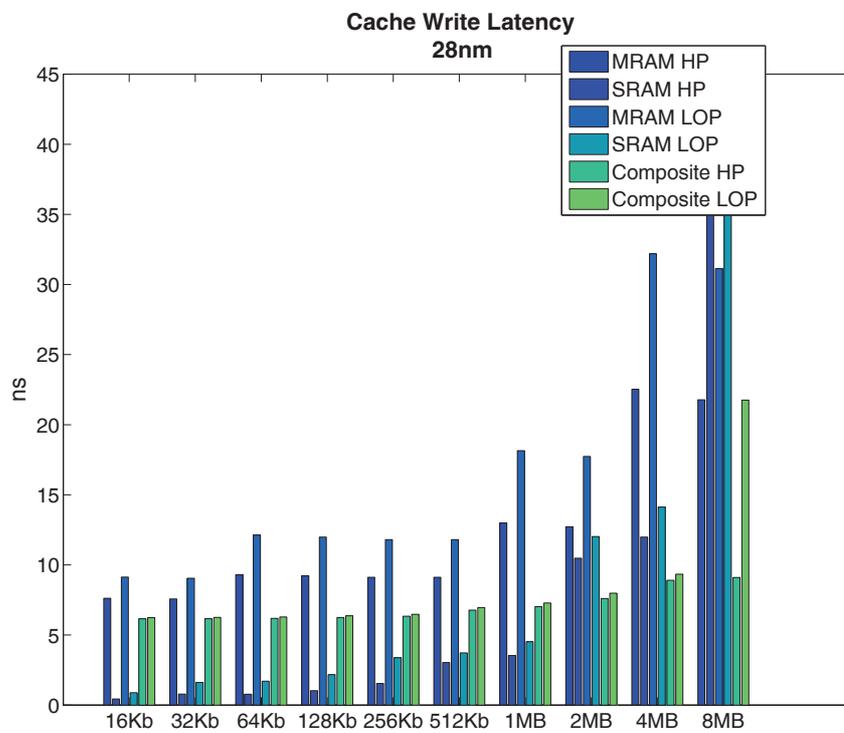


Figure E.2: Total Write Latency.

Table E.2: Cache Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	7.6208	0.4385	9.1320	0.8893	6.1556	6.2446
32KB	7.5755	0.7907	9.0482	1.6204	6.1661	6.2592
64KB	9.3097	0.7766	12.1337	1.7068	6.1856	6.2846
128KB	9.2224	1.0300	11.9828	2.1748	6.2508	6.3747
256KB	9.1170	1.5494	11.8007	3.3801	6.3432	6.4801
512KB	9.1179	3.0395	11.8022	3.7236	6.7750	6.9570
1MB	12.9919	3.5433	18.1500	4.5230	7.0267	7.2825
2MB	12.7278	10.4809	17.7366	12.0243	7.6025	7.9810
4MB	22.5341	11.9878	32.1938	14.1231	8.9019	9.3365
8MB	21.7745	38.2659	31.1254	41.9588	9.1050	21.7534

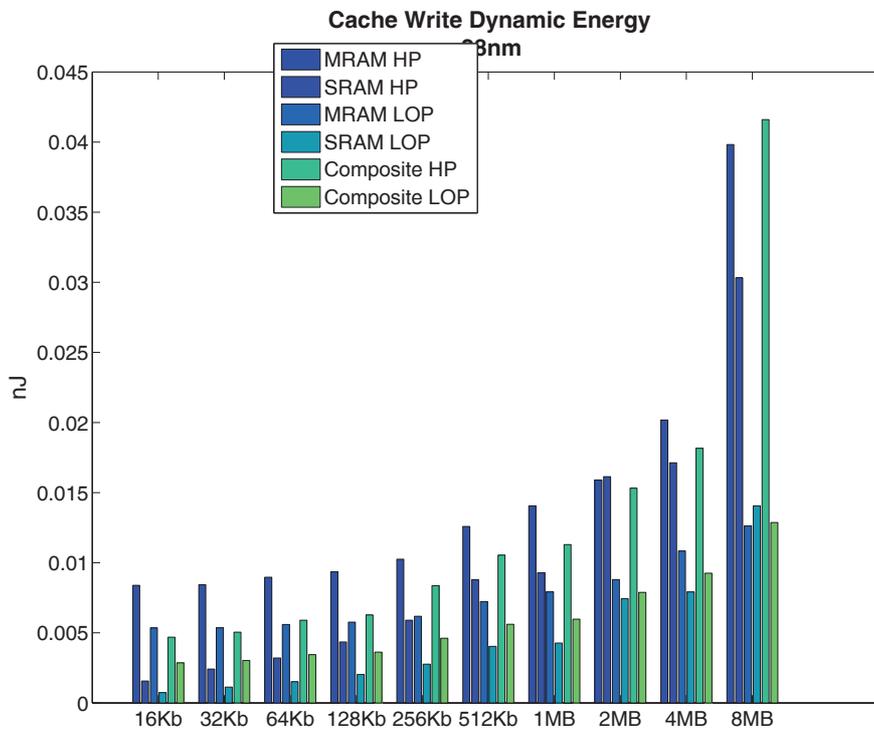


Figure E.3: Total Write Dynamic Energy.

Table E.3: Cache Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0084	0.0016	0.0054	0.0007	0.0047	0.0029
32KB	0.0084	0.0024	0.0054	0.0011	0.0050	0.0030
64KB	0.0090	0.0032	0.0056	0.0015	0.0059	0.0034
128KB	0.0094	0.0043	0.0058	0.0020	0.0063	0.0036
256KB	0.0102	0.0059	0.0062	0.0028	0.0084	0.0046
512KB	0.0126	0.0088	0.0072	0.0040	0.0105	0.0056
1MB	0.0141	0.0093	0.0079	0.0043	0.0113	0.0060
2MB	0.0159	0.0161	0.0088	0.0074	0.0153	0.0079
4MB	0.0202	0.0171	0.0108	0.0079	0.0182	0.0093
8MB	0.0398	0.0303	0.0126	0.0141	0.0416	0.0129

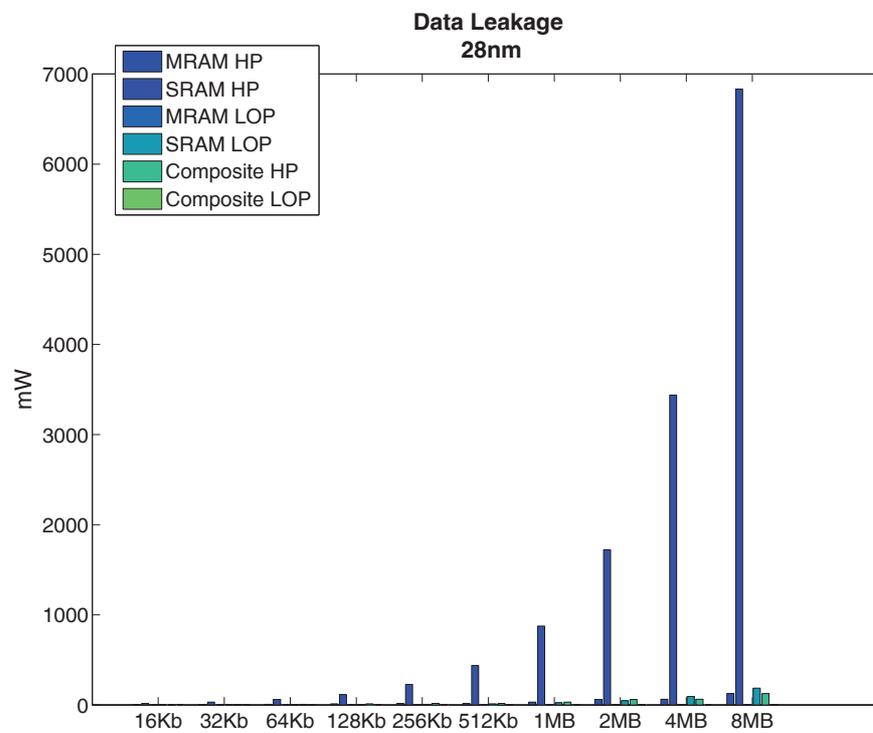


Figure E.4: CACHE Data Array Leakage Power.

Table E.4: Cache Data Array Leakage Power (nW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	2.2398	16.4392	0.0804	0.4492	2.2398	0.0804
32KB	3.6587	30.9181	0.1197	0.8444	3.6587	0.1197
64KB	6.4806	61.0068	0.1978	1.6660	6.4806	0.1978
128KB	12.1186	114.7180	0.3538	3.1302	12.1186	0.3538
256KB	16.6635	227.6520	0.4794	6.2112	16.6635	0.4794
512KB	16.9235	438.7010	0.6165	11.9625	16.9235	0.6165
1MB	31.7669	875.4760	1.0271	23.8729	31.7669	1.0271
2MB	61.3949	1720.9800	1.8469	46.9174	61.3949	1.8469
4MB	63.0393	3438.3400	1.8920	93.7358	63.0393	1.8920
8MB	126.0790	6832.3000	3.5315	186.2450	126.0790	3.5315

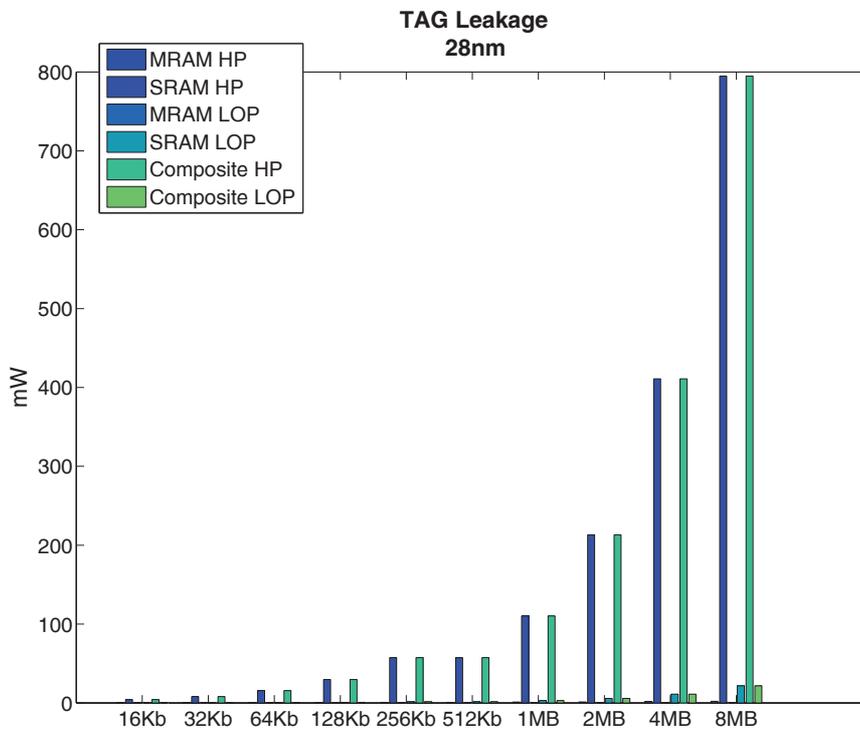


Figure E.5: TAG Leakage Power.

Table E.5: Cache Tag Array Leakage Power (nW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.3647	4.2345	0.0321	0.1153	4.2345	0.1153
32KB	0.3795	8.0315	0.0319	0.2188	8.0315	0.2188
64KB	0.5910	15.6477	0.0371	0.4267	15.6477	0.4267
128KB	0.6313	29.6952	0.0377	0.8094	29.6952	0.8094
256KB	0.6794	57.3559	0.0384	1.5636	57.3559	1.5636
512KB	0.7647	57.4221	0.0605	1.5653	57.4221	1.5653
1MB	1.1735	110.2310	0.0705	3.0045	110.2310	3.0045
2MB	1.2980	213.1200	0.0728	5.8092	213.1200	5.8092
4MB	1.9621	411.0060	0.0899	11.2024	411.0060	11.2024
8MB	2.0956	794.8210	0.0924	21.6640	794.8210	21.6640

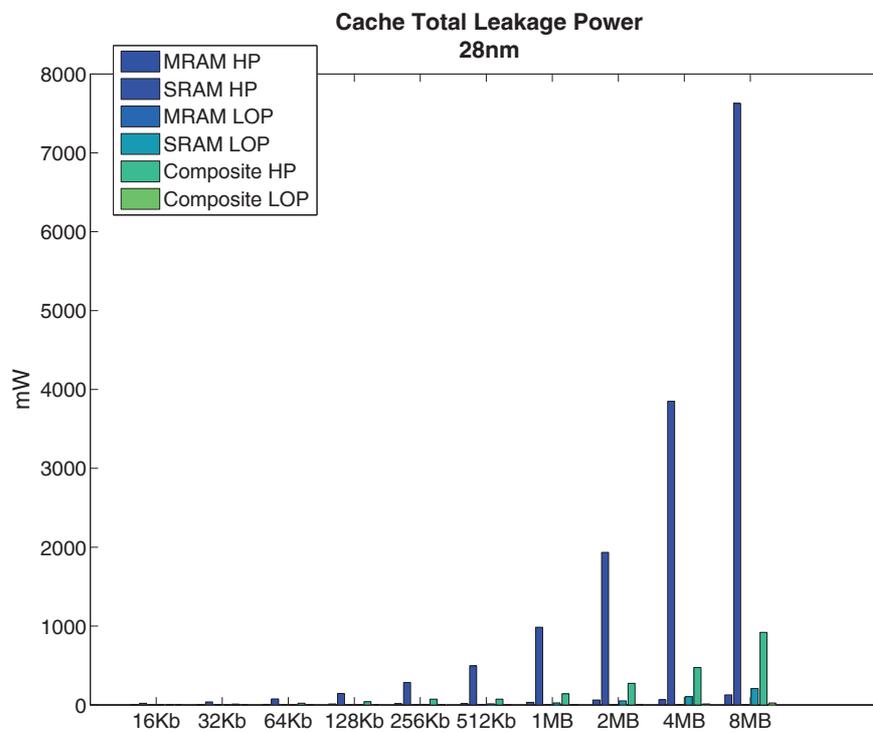


Figure E.6: Leakage Power.

Table E.6: Cache Total Leakage Power (mW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	2.6045	20.6737	0.1125	0.5645	6.4742	0.1958
32KB	4.0383	38.9497	0.1516	1.0632	11.6903	0.3385
64KB	7.0716	76.6546	0.2349	2.0926	22.1283	0.6244
128KB	12.7499	144.4140	0.3914	3.9396	41.8138	1.1632
256KB	17.3429	285.0080	0.5178	7.7749	74.0194	2.0431
512KB	17.6883	496.1230	0.6769	13.5279	74.3456	2.1818
1MB	32.9403	985.7070	1.0977	26.8773	141.9979	4.0316
2MB	62.6929	1934.1000	1.9197	52.7265	274.5149	7.6560
4MB	65.0014	3849.3400	1.9818	104.9380	474.0453	13.0944
8MB	128.1740	7627.1200	3.6239	207.9090	920.9000	25.1955

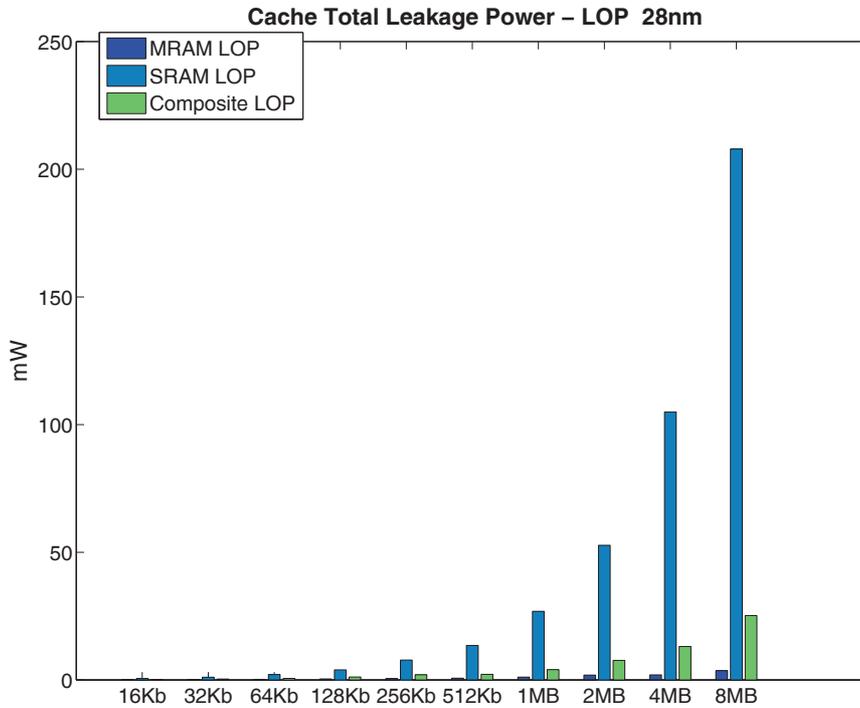


Figure E.7: Low-Power (LOP) Total Leakage, zoom into it to observe only the LOP banks observed in Figure E.6.

Table E.7: LOP Cache Total Leakage Power (mW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.1125	0.5645	0.1958
32KB	0.1516	1.0632	0.3385
64KB	0.2349	2.0926	0.6244
128KB	0.3914	3.9396	1.1632
256KB	0.5178	7.7749	2.0431
512KB	0.6769	13.5279	2.1818
1MB	1.0977	26.8773	4.0316
2MB	1.9197	52.7265	7.6560
4MB	1.9818	104.9380	13.0944
8MB	3.6239	207.9090	25.1955

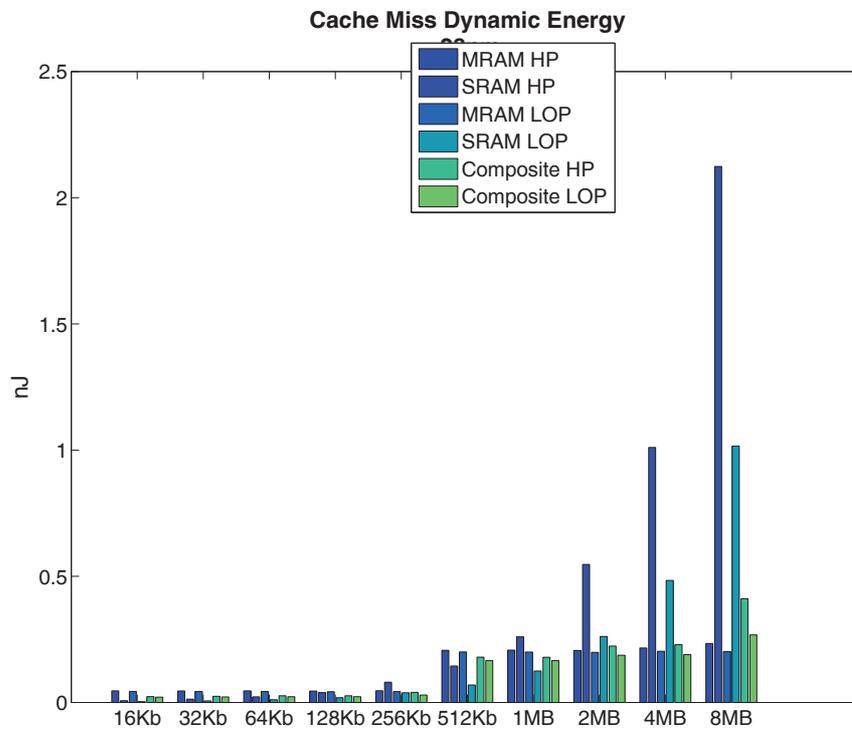


Figure E.8: Miss Dynamic Energy.

Table E.8: Cache Miss Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0459	0.0068	0.0439	0.0033	0.0231	0.0211
32KB	0.0453	0.0128	0.0432	0.0061	0.0242	0.0216
64KB	0.0457	0.0222	0.0432	0.0106	0.0263	0.0226
128KB	0.0451	0.0391	0.0425	0.0187	0.0262	0.0226
256KB	0.0464	0.0801	0.0428	0.0383	0.0401	0.0292
512KB	0.2064	0.1443	0.2001	0.0689	0.1796	0.1660
1MB	0.2072	0.2603	0.1999	0.1244	0.1791	0.1658
2MB	0.2059	0.5466	0.1986	0.2613	0.2232	0.1869
4MB	0.2159	1.0104	0.2028	0.4832	0.2289	0.1896
8MB	0.2336	2.1235	0.2015	1.0157	0.4112	0.2677

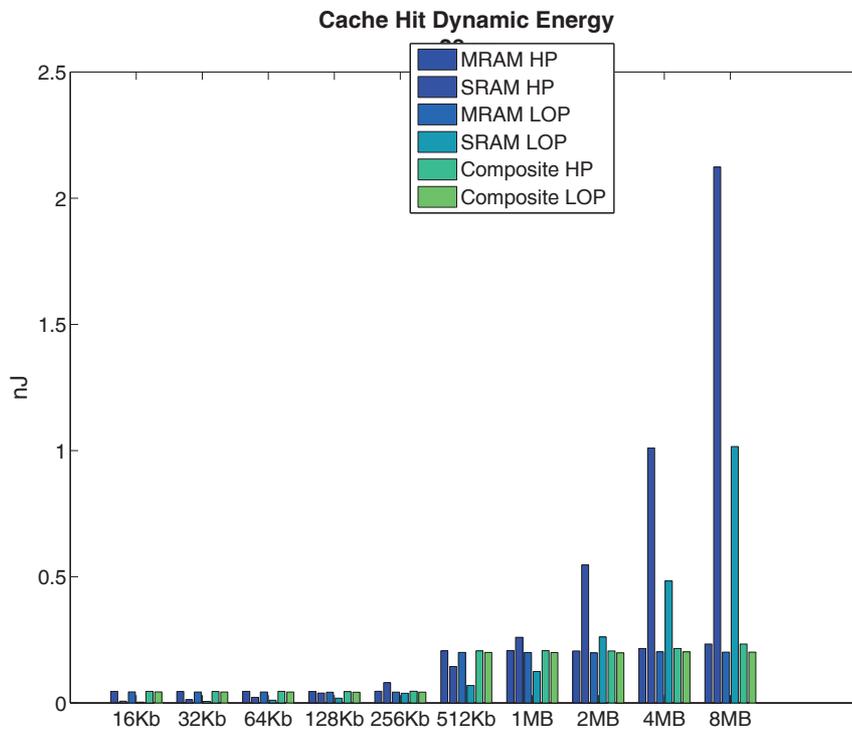


Figure E.9: Hit Dynamic Energy.

Table E.9: Cache Hit Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0459	0.0068	0.0439	0.0033	0.0459	0.0439
32KB	0.0453	0.0128	0.0432	0.0061	0.0453	0.0432
64KB	0.0457	0.0222	0.0432	0.0106	0.0457	0.0432
128KB	0.0451	0.0391	0.0425	0.0187	0.0451	0.0425
256KB	0.0464	0.0801	0.0428	0.0383	0.0464	0.0428
512KB	0.2064	0.1443	0.2001	0.0689	0.2064	0.2001
1MB	0.2072	0.2603	0.1999	0.1244	0.2072	0.1999
2MB	0.2059	0.5466	0.1986	0.2613	0.2059	0.1986
4MB	0.2159	1.0104	0.2028	0.4832	0.2159	0.2028
8MB	0.2336	2.1235	0.2015	1.0157	0.2336	0.2015

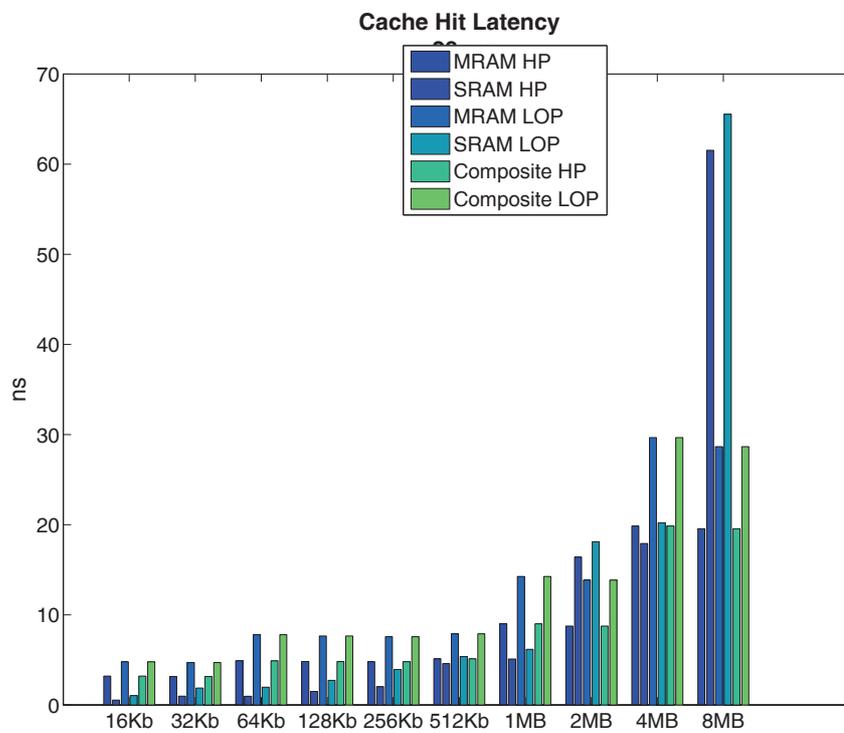


Figure E.10: Hit Latency.

Table E.10: Cache Hit Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.2005	0.5366	4.7837	1.0389	3.2005	4.7837
32KB	3.1556	0.9798	4.7004	1.8794	3.1556	4.7004
64KB	4.8952	0.9671	7.7962	1.9678	4.8952	7.7962
128KB	4.8092	1.5021	7.6462	2.7317	4.8092	7.6462
256KB	4.7959	2.0232	7.5658	3.9396	4.7959	7.5658
512KB	5.1139	4.5914	7.8917	5.3655	5.1139	7.8917
1MB	8.9944	5.0984	14.2518	6.1701	8.9944	14.2518
2MB	8.7464	16.4013	13.8563	18.1043	8.7464	13.8563
4MB	19.8497	17.9117	29.6634	20.2088	19.8497	29.6634
8MB	19.5485	61.5286	28.6538	65.5319	19.5485	28.6538

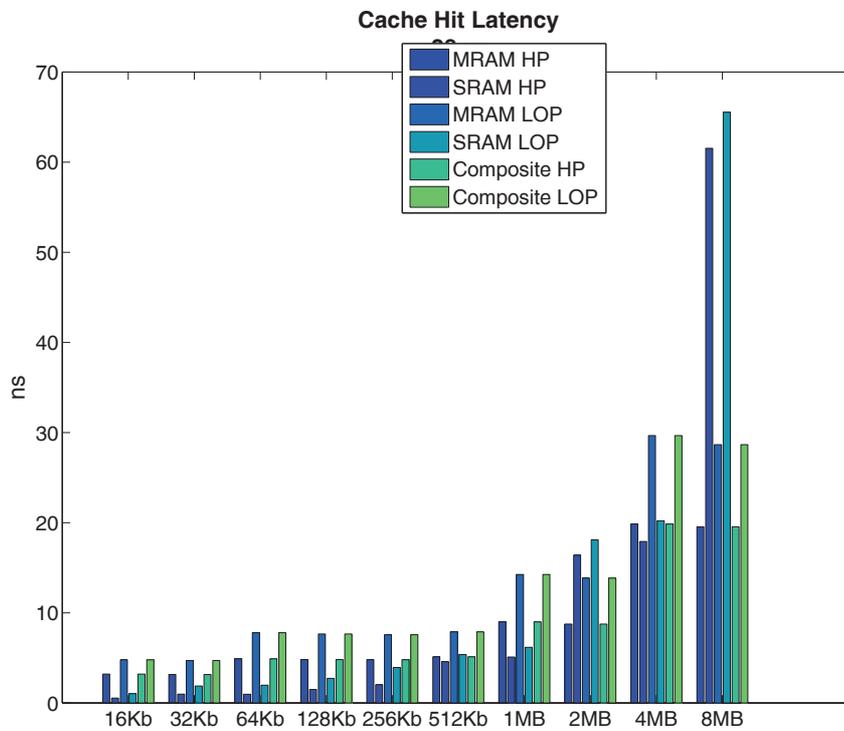


Figure E.11: Hit Latency.

Table E.11: Cache Hit Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.2005	0.5366	4.7837	1.0389	3.2005	4.7837
32KB	3.1556	0.9798	4.7004	1.8794	3.1556	4.7004
64KB	4.8952	0.9671	7.7962	1.9678	4.8952	7.7962
128KB	4.8092	1.5021	7.6462	2.7317	4.8092	7.6462
256KB	4.7959	2.0232	7.5658	3.9396	4.7959	7.5658
512KB	5.1139	4.5914	7.8917	5.3655	5.1139	7.8917
1MB	8.9944	5.0984	14.2518	6.1701	8.9944	14.2518
2MB	8.7464	16.4013	13.8563	18.1043	8.7464	13.8563
4MB	19.8497	17.9117	29.6634	20.2088	19.8497	29.6634
8MB	19.5485	61.5286	28.6538	65.5319	19.5485	28.6538

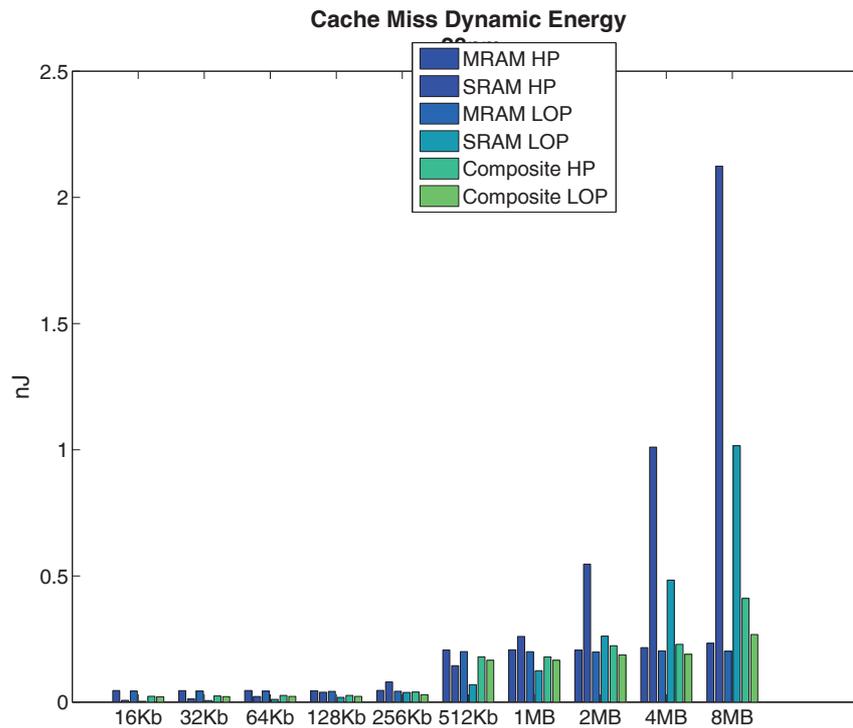


Figure E.12: Miss Dynamic Energy.

Table E.12: Cache Miss Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0459	0.0068	0.0439	0.0033	0.0231	0.0211
32KB	0.0453	0.0128	0.0432	0.0061	0.0242	0.0216
64KB	0.0457	0.0222	0.0432	0.0106	0.0263	0.0226
128KB	0.0451	0.0391	0.0425	0.0187	0.0262	0.0226
256KB	0.0464	0.0801	0.0428	0.0383	0.0401	0.0292
512KB	0.2064	0.1443	0.2001	0.0689	0.1796	0.1660
1MB	0.2072	0.2603	0.1999	0.1244	0.1791	0.1658
2MB	0.2059	0.5466	0.1986	0.2613	0.2232	0.1869
4MB	0.2159	1.0104	0.2028	0.4832	0.2289	0.1896
8MB	0.2336	2.1235	0.2015	1.0157	0.4112	0.2677

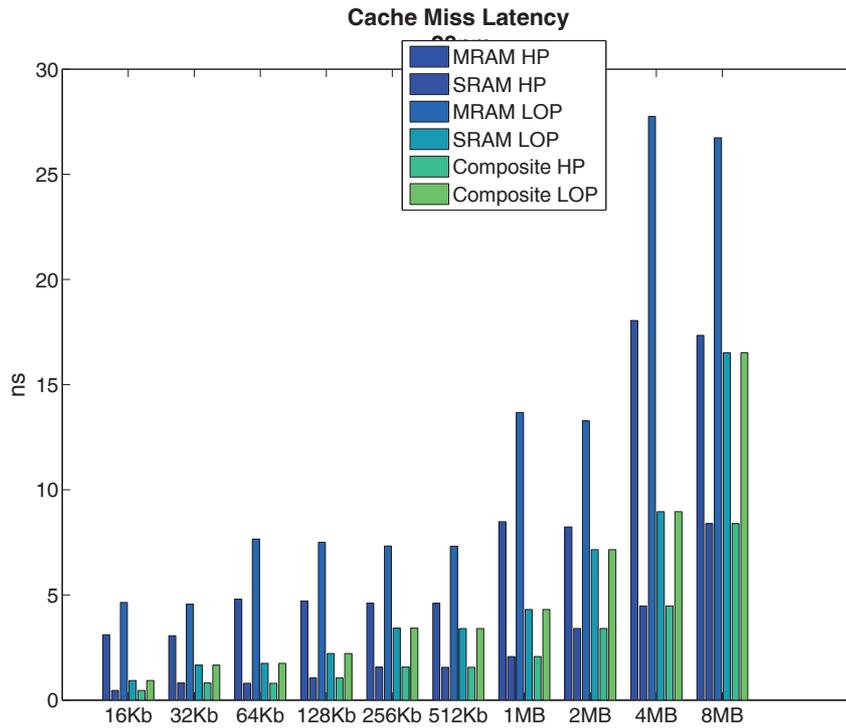


Figure E.13: Miss Latency .

Table E.13: Cache Miss Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.1050	0.4603	4.6437	0.9315	0.4603	0.9315
32KB	3.0600	0.8125	4.5602	1.6627	0.8125	1.6627
64KB	4.7994	0.7984	7.6557	1.7491	0.7984	1.7491
128KB	4.7123	1.0512	7.5046	2.2159	1.0512	2.2159
256KB	4.6104	1.5706	7.3259	3.4212	1.5706	3.4212
512KB	4.6061	1.5607	7.3173	3.3992	1.5607	3.3992
1MB	8.4852	2.0722	13.6752	4.3105	2.0722	4.3105
2MB	8.2346	3.4069	13.2756	7.1559	3.4069	7.1559
4MB	18.0507	4.4777	27.7518	8.9596	4.4777	8.9596
8MB	17.3433	8.3928	26.7375	16.5125	8.3928	16.5125

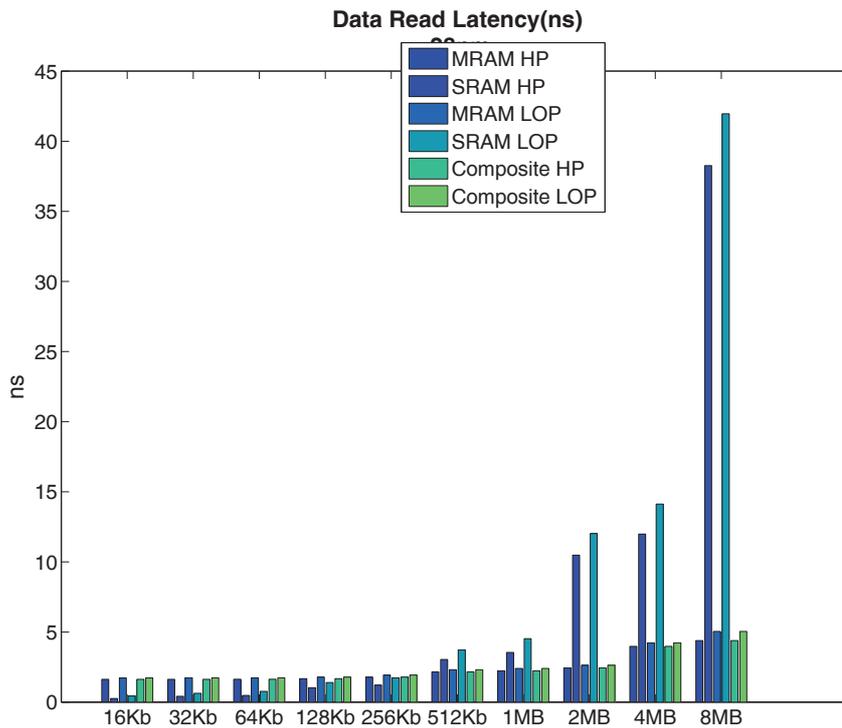


Figure E.14: Data Read Latency.

Table E.14: Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	1.6223	0.2537	1.7222	0.4389	1.6223	1.7222
32KB	1.6256	0.4114	1.7274	0.6305	1.6256	1.7274
64KB	1.6306	0.4769	1.7337	0.7661	1.6306	1.7337
128KB	1.6666	1.0206	1.7856	1.3957	1.6666	1.7856
256KB	1.7851	1.2076	1.9284	1.7345	1.7851	1.9284
512KB	2.1568	3.0395	2.3064	3.7236	2.1568	2.3064
1MB	2.2325	3.5433	2.4014	4.5230	2.2325	2.4014
2MB	2.4403	10.4809	2.6256	12.0243	2.4403	2.6256
4MB	3.9779	11.9878	4.2264	14.1231	3.9779	4.2264
8MB	4.3840	38.2659	5.0413	41.9588	4.3840	5.0413

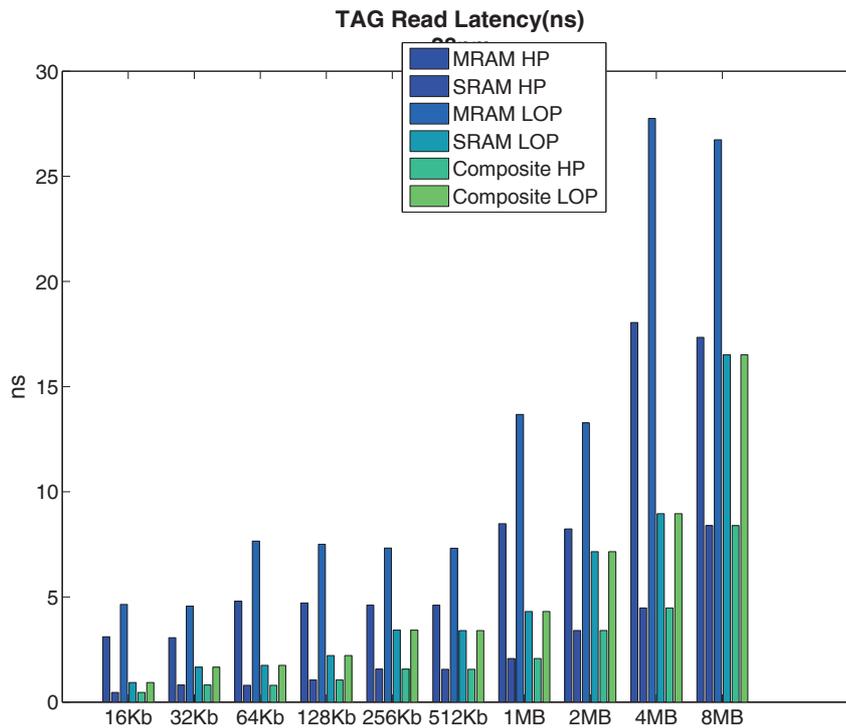


Figure E.15: TAG Read Latency.

Table E.15: Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.1050	0.4603	4.6437	0.9315	0.4603	0.9315
32KB	3.0600	0.8125	4.5602	1.6627	0.8125	1.6627
64KB	4.7994	0.7984	7.6557	1.7491	0.7984	1.7491
128KB	4.7123	1.0512	7.5046	2.2159	1.0512	2.2159
256KB	4.6104	1.5706	7.3259	3.4212	1.5706	3.4212
512KB	4.6061	1.5607	7.3173	3.3992	1.5607	3.3992
1MB	8.4852	2.0722	13.6752	4.3105	2.0722	4.3105
2MB	8.2346	3.4069	13.2756	7.1559	3.4069	7.1559
4MB	18.0507	4.4777	27.7518	8.9596	4.4777	8.9596
8MB	17.3433	8.3928	26.7375	16.5125	8.3928	16.5125

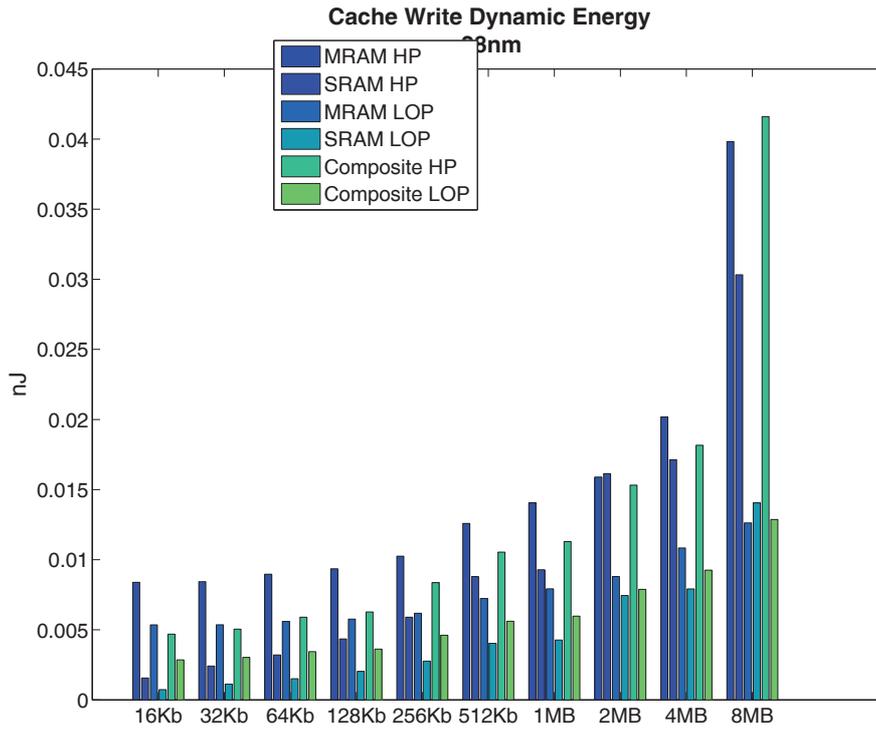


Figure E.16: Write Dynamic Energy

Table E.16: Cache Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0084	0.0016	0.0054	0.0007	0.0047	0.0029
32KB	0.0084	0.0024	0.0054	0.0011	0.0050	0.0030
64KB	0.0090	0.0032	0.0056	0.0015	0.0059	0.0034
128KB	0.0094	0.0043	0.0058	0.0020	0.0063	0.0036
256KB	0.0102	0.0059	0.0062	0.0028	0.0084	0.0046
512KB	0.0126	0.0088	0.0072	0.0040	0.0105	0.0056
1MB	0.0141	0.0093	0.0079	0.0043	0.0113	0.0060
2MB	0.0159	0.0161	0.0088	0.0074	0.0153	0.0079
4MB	0.0202	0.0171	0.0108	0.0079	0.0182	0.0093
8MB	0.0398	0.0303	0.0126	0.0141	0.0416	0.0129

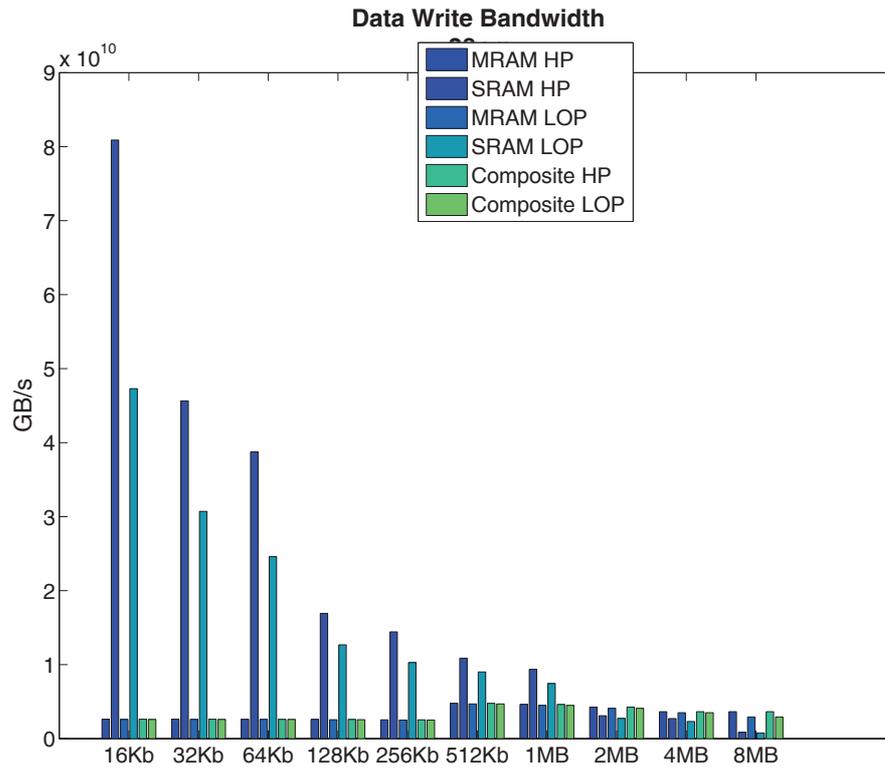


Figure E.17: Write Bandwidth.

Table E.17: Write Bandwidth (B/s)

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	2.442	75.33	2.423	44.04	2.442	2.423
32KB	2.438	42.50	2.419	28.61	2.438	2.419
64KB	2.431	36.09	2.410	22.90	2.431	2.410
128KB	2.414	15.76	2.389	11.81	2.414	2.389
256KB	2.380	13.45	2.352	9.572	2.380	2.352
512KB	4.448	10.12	4.367	8.399	4.448	4.367
1MB	4.300	8.718	4.191	6.941	4.300	4.191
2MB	3.974	2.886	3.823	2.539	3.974	3.823
4MB	3.392	2.523	3.266	2.161	3.392	3.266
8MB	3.392	0.7836	2.723	0.7180	3.392	2.723

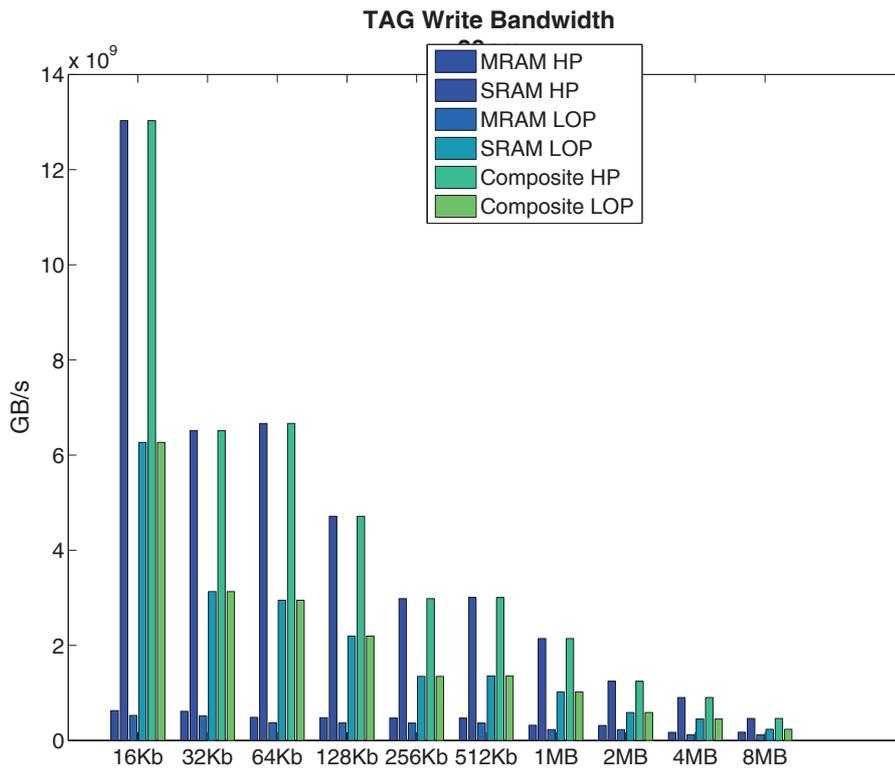


Figure E.18: Write Bandwidth.

Table E.18: Write Bandwidth (B/s)

	MRAM HP		SRAM HP		MRAM LOP		SRAM LOP		Composite HP		Composite LOP	
16KB	5.845e 01	-	1.213e 01	+	4.895e 01	-	5.837e 00	+	1.213e 01	+	5.837e 00	+
32KB	5.726e 01	-	6.067e 00	+	4.811e 01	-	2.916e 00	+	6.067e 00	+	2.916e 00	+
64KB	4.530e 01	-	6.207e 00	+	3.484e 01	-	2.746e 00	+	6.207e 00	+	2.746e 00	+
128KB	4.455e 01	-	4.390e 00	+	3.440e 01	-	2.044e 00	+	4.390e 00	+	2.044e 00	+
256KB	4.379e 01	-	2.779e 00	+	3.394e 01	-	1.255e 00	+	2.779e 00	+	1.255e 00	+
512KB	4.379e 01	-	2.798e 00	+	3.394e 01	-	1.264e 00	+	2.798e 00	+	1.264e 00	+
1MB	2.977e 01	-	1.996e 00	+	2.135e 01	-	9.509e 01	-	1.996e 00	+	9.509e 01	-
2MB	2.949e 01	-	1.158e 00	+	2.120e 01	-	5.479e 01	-	1.158e 00	+	5.479e 01	-
4MB	1.609e 01	-	8.418e 01	-	1.128e 01	-	4.196e 01	-	8.418e 01	-	4.196e 01	-
8MB	1.612e 01	-	4.313e 01	-	1.129e 01	-	2.190e 01	-	4.313e 01	-	2.190e 01	-

THE COMPOSITE BANK - ADDITIONAL RESULTS

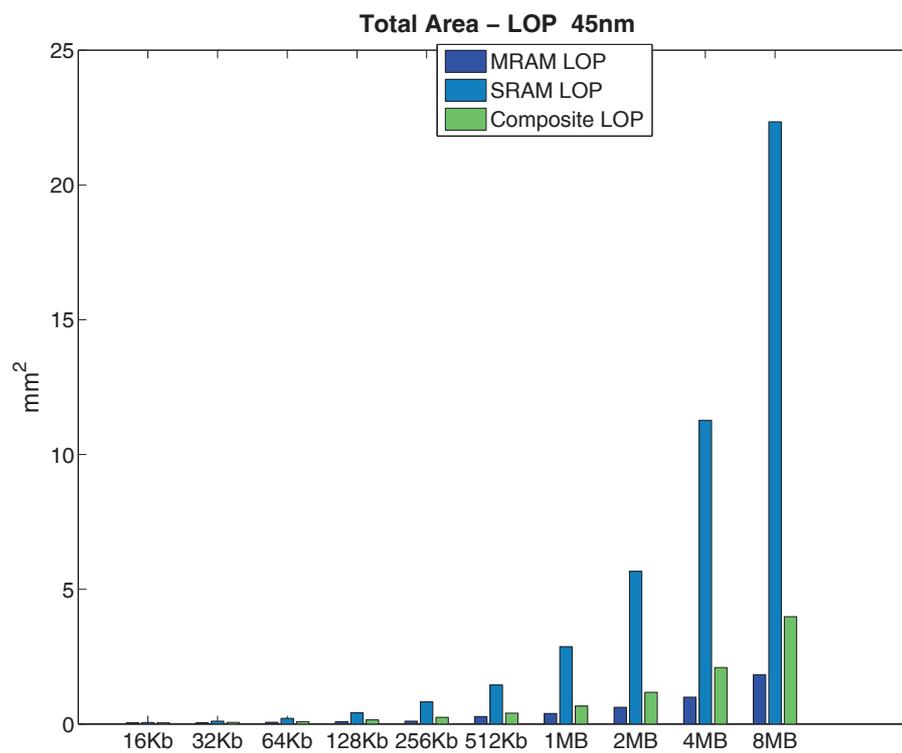
F.1 LOP 45nm

Figure F.1: Total Area.

Table F.1: Total Area (μm^2).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0476	0.0568	0.0422
32KB	0.0529	0.1083	0.0587
64KB	0.0638	0.2120	0.0899
128KB	0.0861	0.4142	0.1518
256KB	0.1086	0.8189	0.2498
512KB	0.2779	1.4455	0.4046
1MB	0.3921	2.8700	0.6659
2MB	0.6183	5.6643	1.1738
4MB	0.9958	11.2673	2.0964
8MB	1.8302	22.3407	3.9818

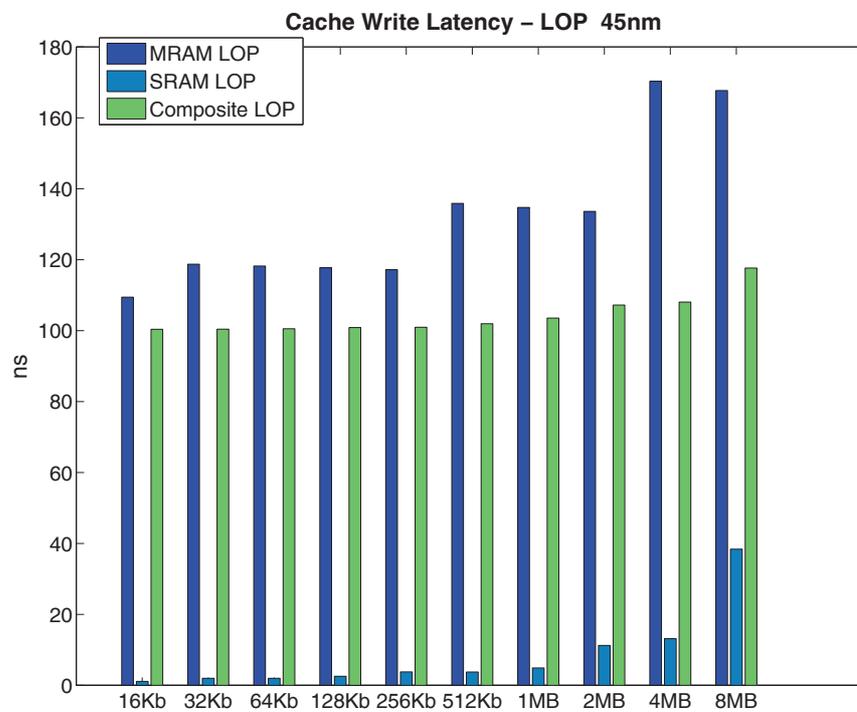


Figure F.2: Total Write Latency.

Table F.2: Cache Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	109.4420	1.0987	100.3730
32KB	118.7320	1.9773	100.4140
64KB	118.2040	1.9596	100.5380
128KB	117.7230	2.5528	100.8680
256KB	117.1990	3.7679	100.9420
512KB	135.8750	3.7399	101.9540
1MB	134.7310	4.8844	103.5410
2MB	133.6090	11.2448	107.1940
4MB	170.2960	13.1675	108.0310
8MB	167.6920	38.3881	117.6390

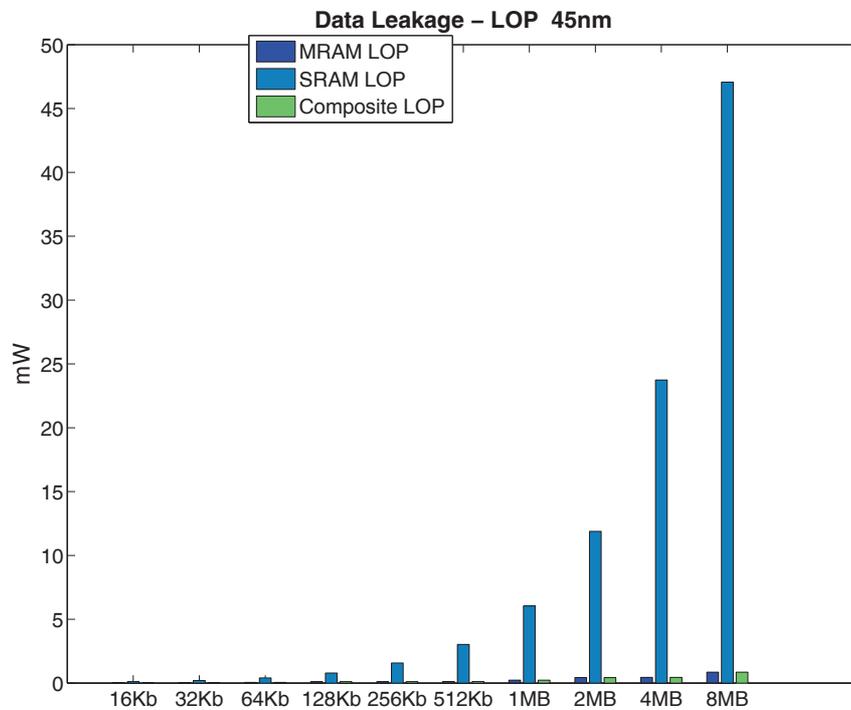


Figure F.3: CACHE Data Array Leakage Power.

Table F.3: Cache Data Array Leakage Power (nW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0189	0.1123	0.0189
32KB	0.0320	0.2075	0.0320
64KB	0.0580	0.4093	0.0580
128KB	0.1101	0.7928	0.1101
256KB	0.1155	1.5731	0.1155
512KB	0.1216	3.0298	0.1216
1MB	0.2258	6.0462	0.2258
2MB	0.4334	11.8826	0.4334
4MB	0.4448	23.7402	0.4448
8MB	0.8601	47.0629	0.8601

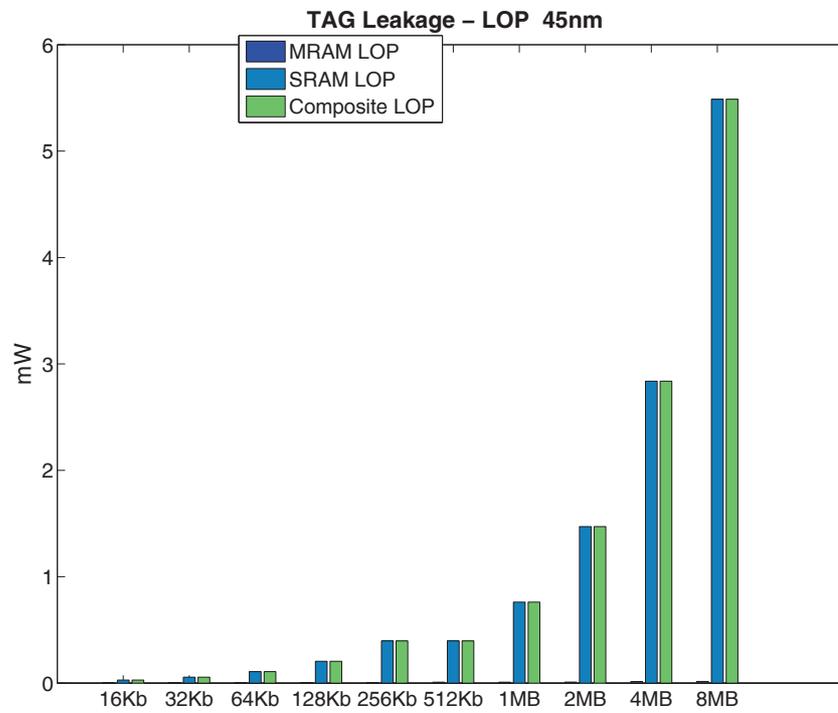


Figure F.4: TAG Leakage Power.

Table F.4: Cache Tag Array Leakage Power (nW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0029	0.0291	0.0291
32KB	0.0044	0.0552	0.0552
64KB	0.0044	0.1081	0.1081
128KB	0.0046	0.2050	0.2050
256KB	0.0049	0.3960	0.3960
512KB	0.0084	0.3965	0.3965
1MB	0.0086	0.7610	0.7610
2MB	0.0093	1.4713	1.4713
4MB	0.0144	2.8372	2.8372
8MB	0.0154	5.4868	5.4868

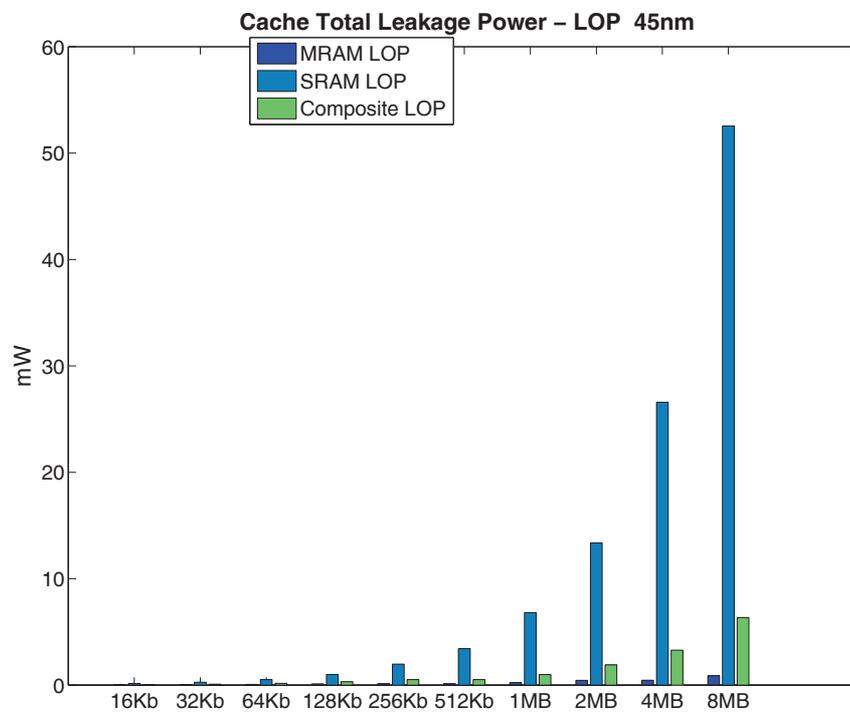


Figure F.5: Leakage Power.

Table F.5: Cache Total Leakage Power (mW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0218	0.1414	0.0480
32KB	0.0363	0.2627	0.0872
64KB	0.0624	0.5174	0.1661
128KB	0.1148	0.9978	0.3152
256KB	0.1204	1.9691	0.5115
512KB	0.1300	3.4262	0.5181
1MB	0.2344	6.8072	0.9868
2MB	0.4427	13.3539	1.9047
4MB	0.4593	26.5774	3.2821
8MB	0.8754	52.5496	6.3468

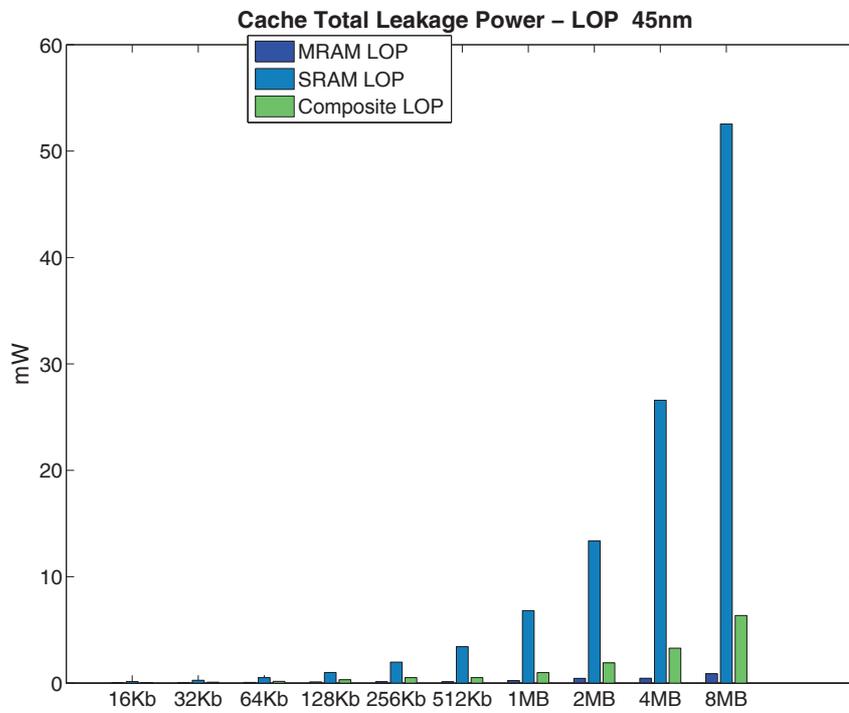


Figure F.6: Low-Power (LOP) Total Leakage, zoom into it to observe only the LOP banks observed in Figure F.5.

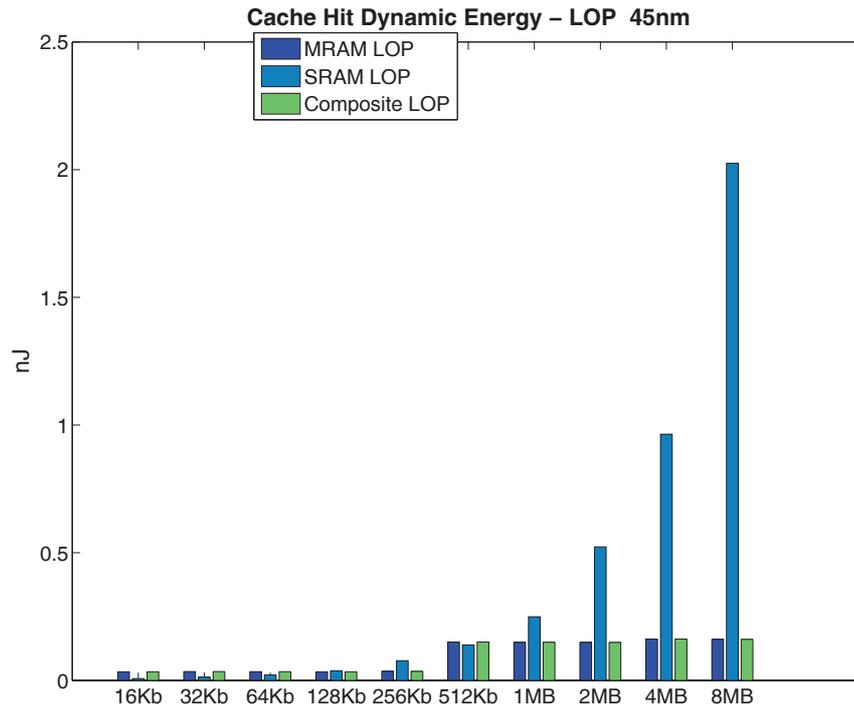


Figure F.7: Hit Dynamic Energy.

Table F.6: Cache Hit Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0335	0.0068	0.0335
32KB	0.0345	0.0126	0.0345
64KB	0.0340	0.0215	0.0340
128KB	0.0335	0.0379	0.0335
256KB	0.0354	0.0770	0.0354
512KB	0.1504	0.1387	0.1504
1MB	0.1494	0.2491	0.1494
2MB	0.1485	0.5226	0.1485
4MB	0.1618	0.9638	0.1618
8MB	0.1608	2.0252	0.1608

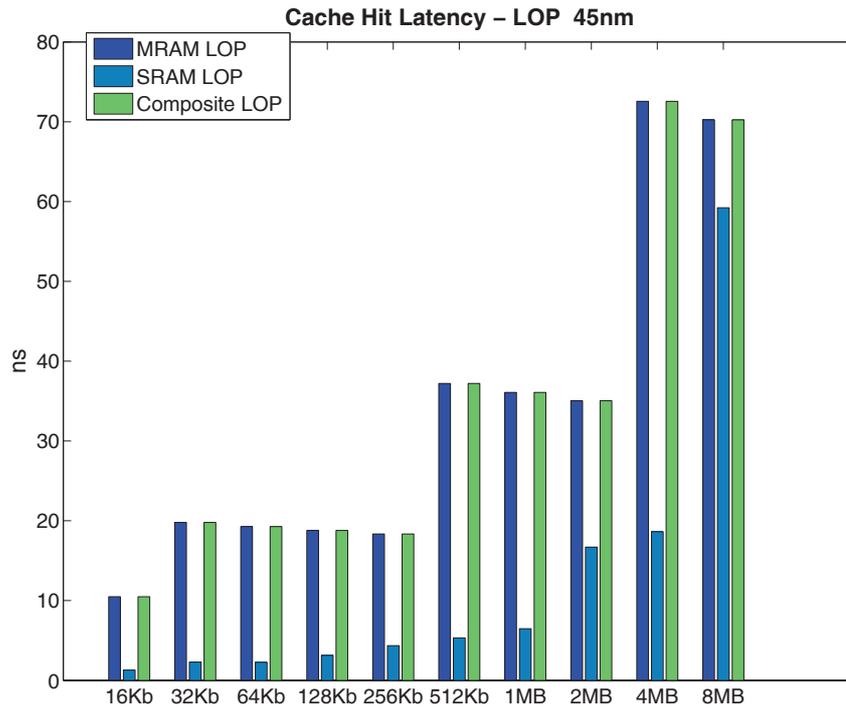


Figure F.8: Hit Latency.

Table F.7: Cache Hit Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	10.4902	1.2926	10.4902
32KB	19.7923	2.2867	19.7923
64KB	19.2655	2.2712	19.2655
128KB	18.7884	3.1341	18.7884
256KB	18.3503	4.3523	18.3503
512KB	37.2127	5.3230	37.2127
1MB	36.0863	6.4738	36.0863
2MB	35.0382	16.7024	35.0382
4MB	72.5652	18.6320	72.5652
8MB	70.2395	59.2280	70.2395

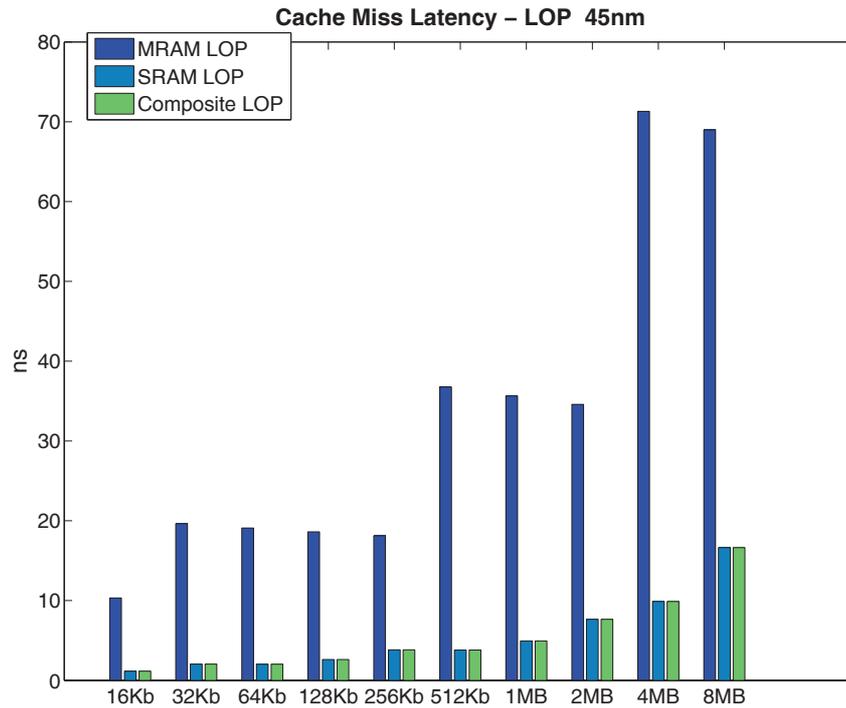


Figure F.9: Miss Latency.

Table F.8: Cache Miss Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	10.3223	1.1577	1.1577
32KB	19.6243	2.0363	2.0363
64KB	19.0972	2.0186	2.0186
128KB	18.6196	2.6101	2.6101
256KB	18.1132	3.8252	3.8252
512KB	36.7710	3.7973	3.7973
1MB	35.6448	4.9418	4.9418
2MB	34.5929	7.6584	7.6584
4MB	71.3032	9.8897	9.8897
8MB	68.9729	16.6321	16.6321

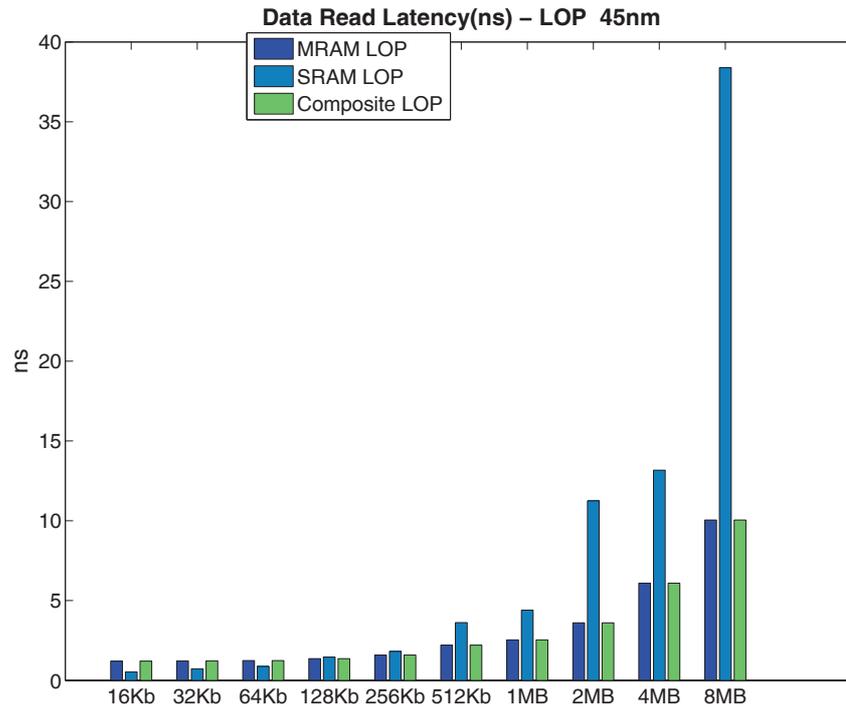


Figure F.10: DATA Read Latency.

Table F.9: DATA Read Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	1.2077	0.5339	1.2077
32KB	1.2178	0.7269	1.2178
64KB	1.2382	0.8690	1.2382
128KB	1.3590	1.4635	1.3590
256KB	1.5860	1.8168	1.5860
512KB	2.2071	3.6172	2.2071
1MB	2.5359	4.4008	2.5359
2MB	3.6027	11.2448	3.6027
4MB	6.0894	13.1675	6.0894
8MB	10.0463	38.3881	10.0463

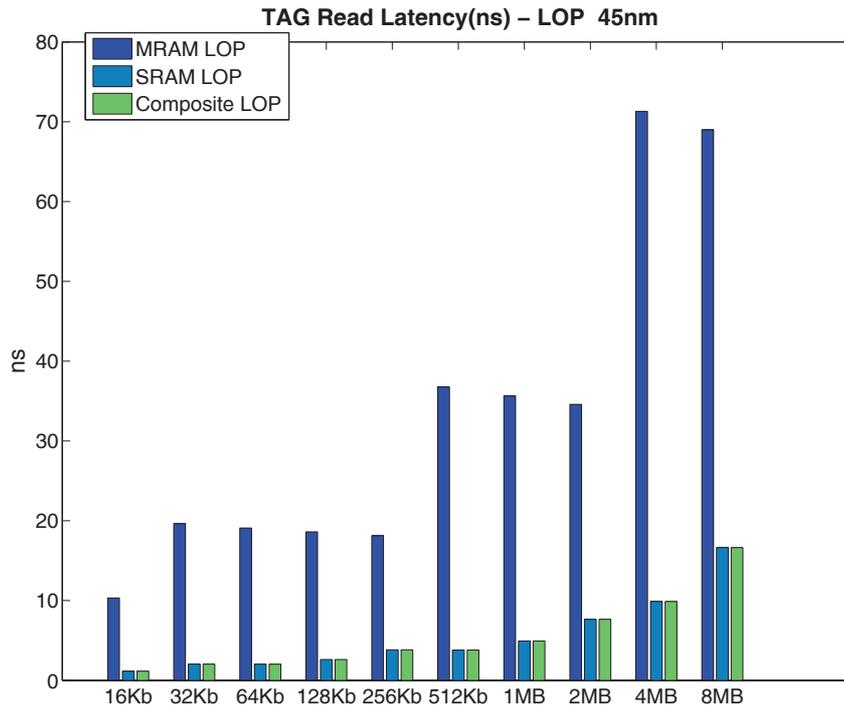


Figure F.11: TAG Read Latency.

Table F.10: TAG Read Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	10.3223	1.1577	1.1577
32KB	19.6243	2.0363	2.0363
64KB	19.0972	2.0186	2.0186
128KB	18.6196	2.6101	2.6101
256KB	18.1132	3.8252	3.8252
512KB	36.7710	3.7973	3.7973
1MB	35.6448	4.9418	4.9418
2MB	34.5929	7.6584	7.6584
4MB	71.3032	9.8897	9.8897
8MB	68.9729	16.6321	16.6321

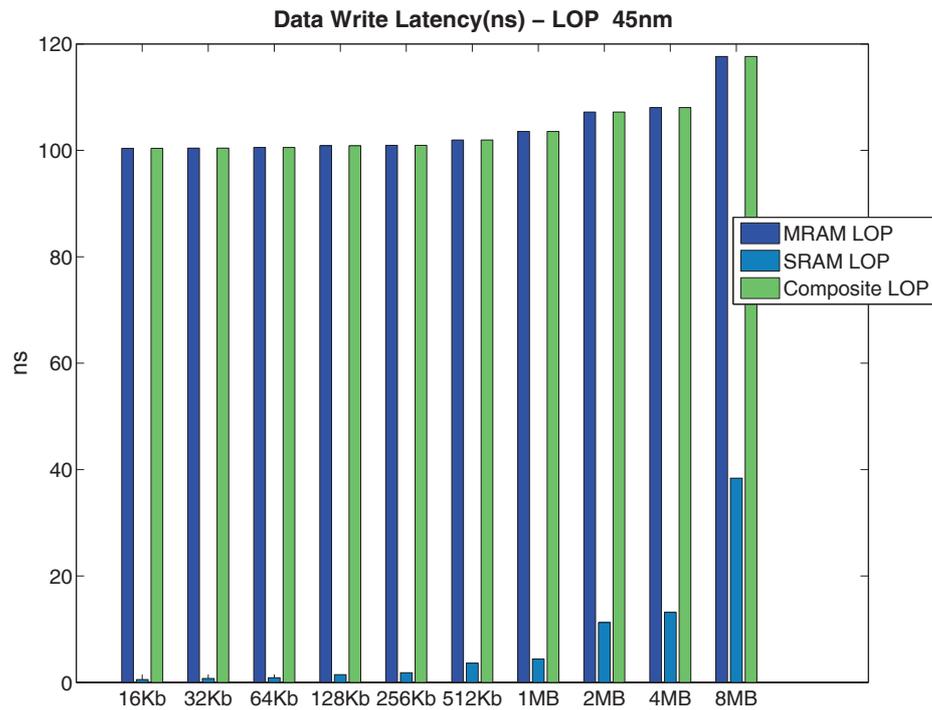


Figure F.12: DATA Write Latency.

Table F.11: Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	100.3730	0.5339	100.3730
32KB	100.4140	0.7269	100.4140
64KB	100.5380	0.8690	100.5380
128KB	100.8680	1.4635	100.8680
256KB	100.9420	1.8168	100.9420
512KB	101.9540	3.6172	101.9540
1MB	103.5410	4.4008	103.5410
2MB	107.1940	11.2448	107.1940
4MB	108.0310	13.1675	108.0310
8MB	117.6390	38.3881	117.6390

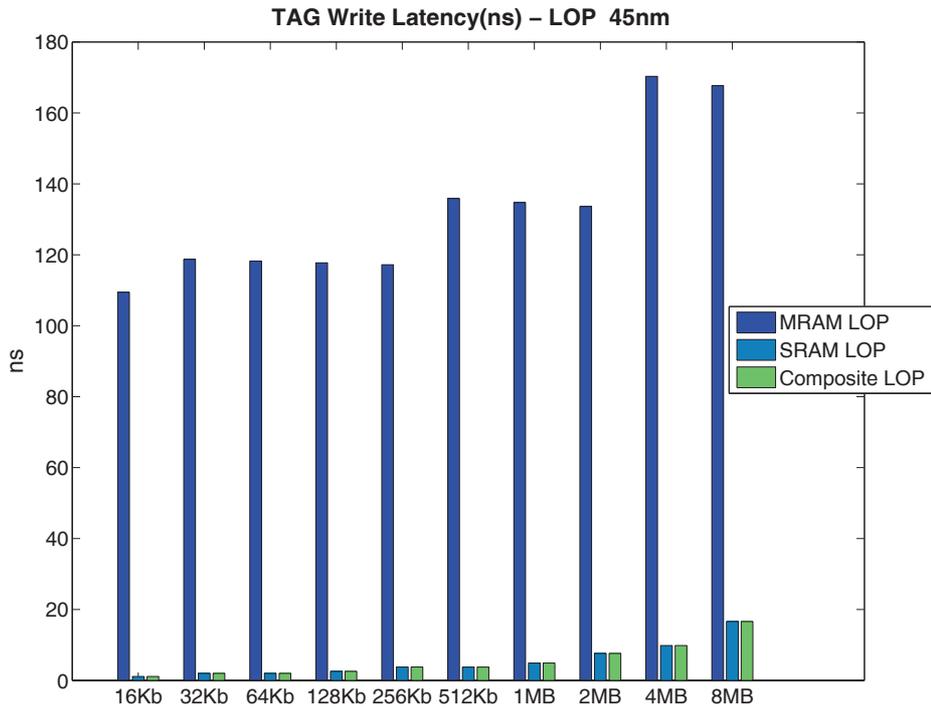


Figure F.13: TAG Write Latency.

Table F.12: Read Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	109.4420	1.0987	1.0987
32KB	118.7320	1.9773	1.9773
64KB	118.2040	1.9596	1.9596
128KB	117.7230	2.5528	2.5528
256KB	117.1990	3.7679	3.7679
512KB	135.8750	3.7399	3.7399
1MB	134.7310	4.8844	4.8844
2MB	133.6090	7.6010	7.6010
4MB	170.2960	9.8340	9.8340
8MB	167.6920	16.5764	16.5764

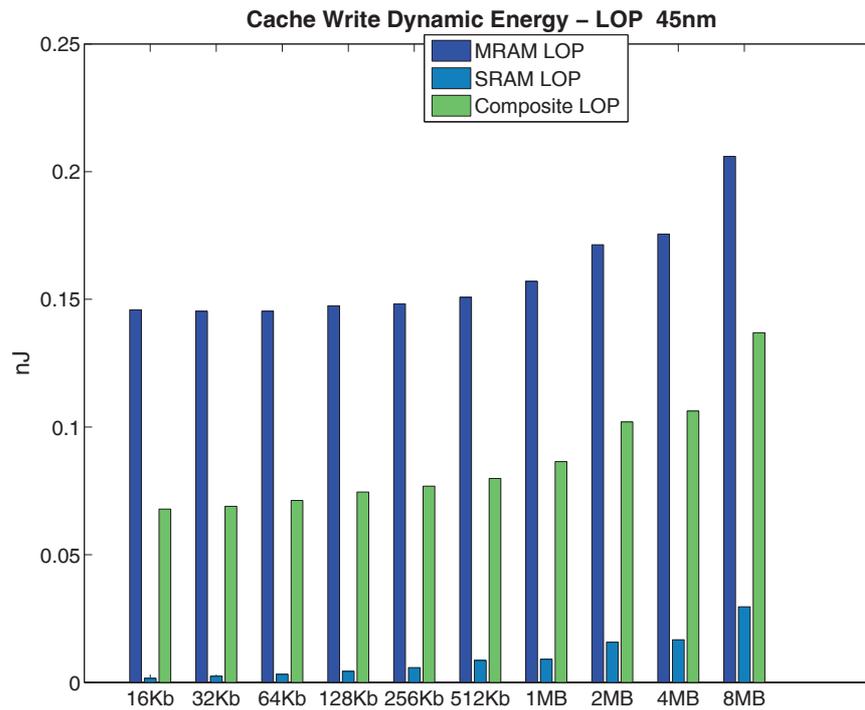


Figure F.14: Write Dynamic Energy

Table F.13: Cache Write Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.1459	0.0016	0.0679
32KB	0.1453	0.0024	0.0689
64KB	0.1453	0.0032	0.0712
128KB	0.1475	0.0043	0.0745
256KB	0.1482	0.0058	0.0768
512KB	0.1508	0.0086	0.0798
1MB	0.1570	0.0091	0.0864
2MB	0.1714	0.0157	0.1020
4MB	0.1755	0.0167	0.1062
8MB	0.2059	0.0296	0.1369

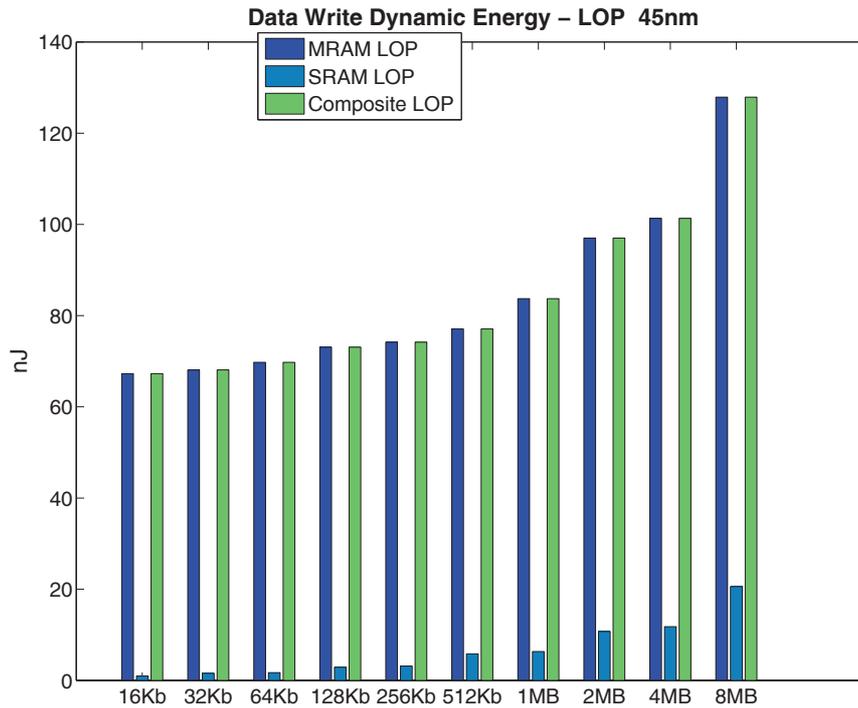


Figure F.15: DATA Dynamic Energy

Table F.14: DATA Write Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	67.2541	0.9587	67.2541
32KB	68.0847	1.5590	68.0847
64KB	69.7460	1.6901	69.7460
128KB	73.0693	2.8865	73.0693
256KB	74.1747	3.1498	74.1747
512KB	77.0569	5.8032	77.0569
1MB	83.7073	6.3284	83.7073
2MB	96.9860	10.7319	96.9860
4MB	101.3290	11.7810	101.3290
8MB	127.8870	20.5837	127.8870

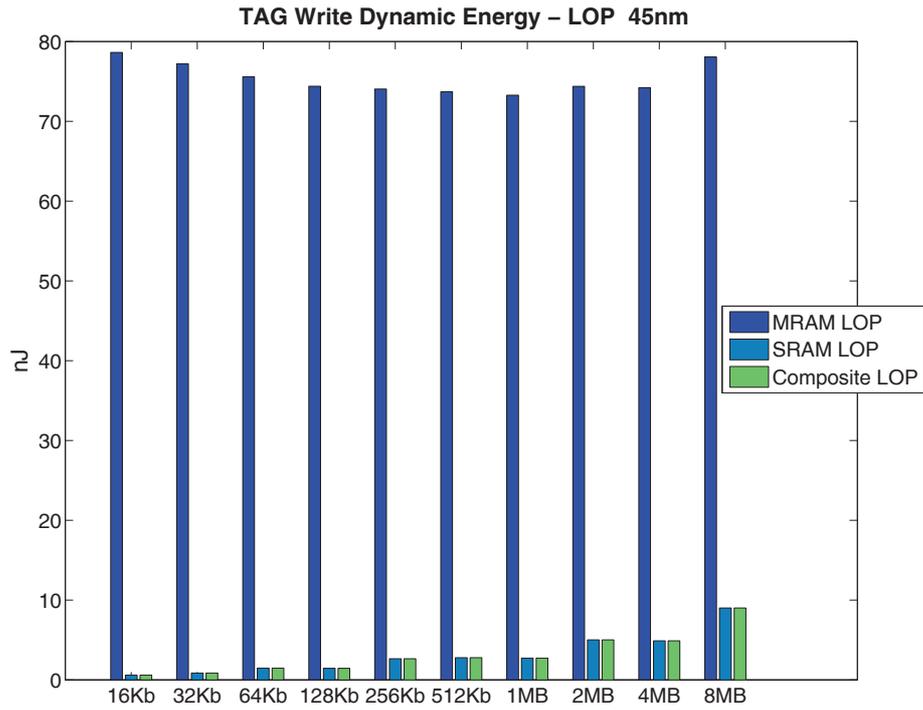


Figure F.16: TAG Dynamic Energy

Table F.15: TAG Write Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	78.6242	0.6093	0.6093
32KB	77.2016	0.8576	0.8576
64KB	75.5873	1.4812	1.4812
128KB	74.3972	1.4513	1.4513
256KB	74.0408	2.6533	2.6533
512KB	73.7150	2.7846	2.7846
1MB	73.2701	2.7341	2.7341
2MB	74.3801	5.0145	5.0145
4MB	74.2152	4.9001	4.9001
8MB	78.0532	9.0142	9.0142

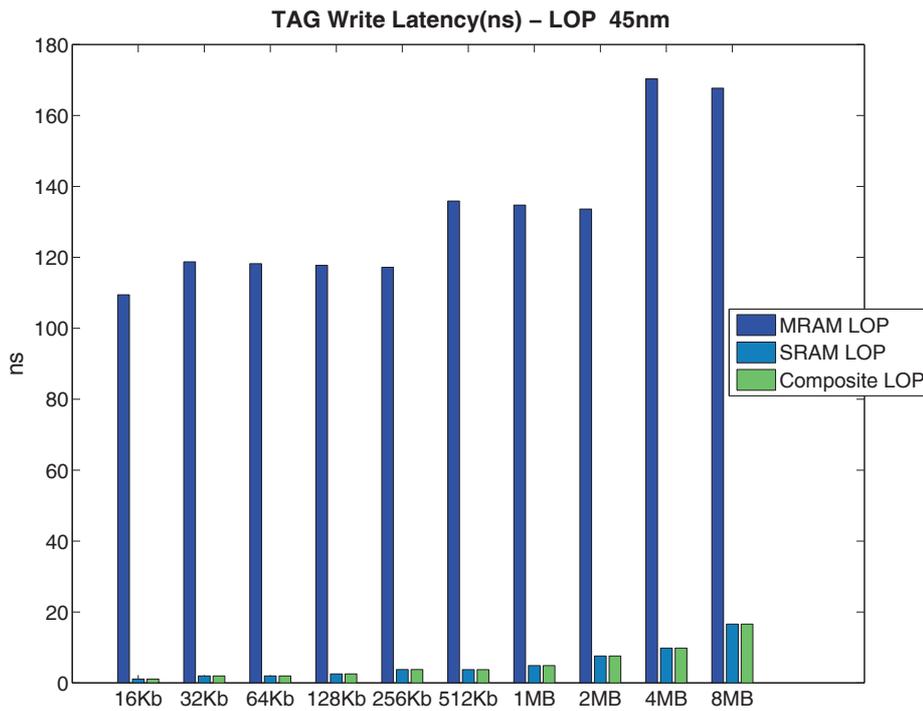


Figure F.17: TAG Dynamic Energy

Table F.16: TAG Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	109.4420	1.0987	1.0987
32KB	118.7320	1.9773	1.9773
64KB	118.2040	1.9596	1.9596
128KB	117.7230	2.5528	2.5528
256KB	117.1990	3.7679	3.7679
512KB	135.8750	3.7399	3.7399
1MB	134.7310	4.8844	4.8844
2MB	133.6090	7.6010	7.6010
4MB	170.2960	9.8340	9.8340
8MB	167.6920	16.5764	16.5764

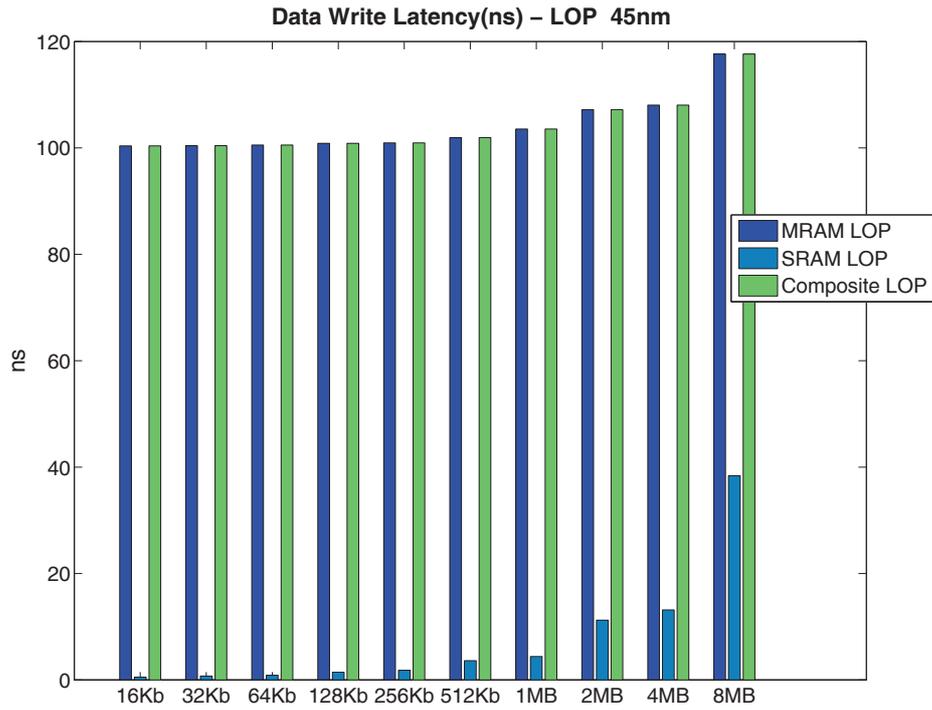


Figure F.18: DATA Write Latency

Table F.17: DATA Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	100.3730	0.5339	100.3730
32KB	100.4140	0.7269	100.4140
64KB	100.5380	0.8690	100.5380
128KB	100.8680	1.4635	100.8680
256KB	100.9420	1.8168	100.9420
512KB	101.9540	3.6172	101.9540
1MB	103.5410	4.4008	103.5410
2MB	107.1940	11.2448	107.1940
4MB	108.0310	13.1675	108.0310
8MB	117.6390	38.3881	117.6390

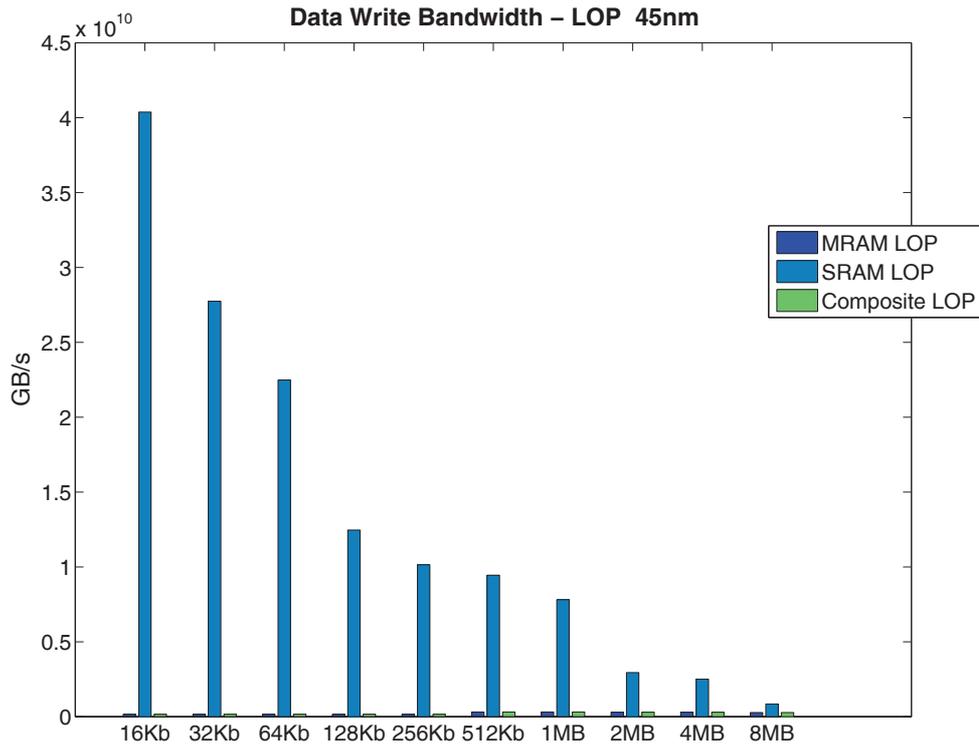


Figure F.19: Write Bandwidth.

Table F.18: Write Bandwidth (B/s)

	MRAM LOP	SRAM LOP	Composite LOP
16KB	1.486e - 01	3.760e + 01	1.486e - 01
32KB	1.486e - 01	2.583e + 01	1.486e - 01
64KB	1.484e - 01	2.094e + 01	1.484e - 01
128KB	1.480e - 01	1.161e + 01	1.480e - 01
256KB	1.479e - 01	9.442e + 00	1.479e - 01
512KB	2.928e - 01	8.796e + 00	2.928e - 01
1MB	2.884e - 01	7.269e + 00	2.884e - 01
2MB	2.786e - 01	2.741e + 00	2.786e - 01
4MB	2.765e - 01	2.339e + 00	2.765e - 01
8MB	2.540e - 01	7.880e - 01	2.540e - 01

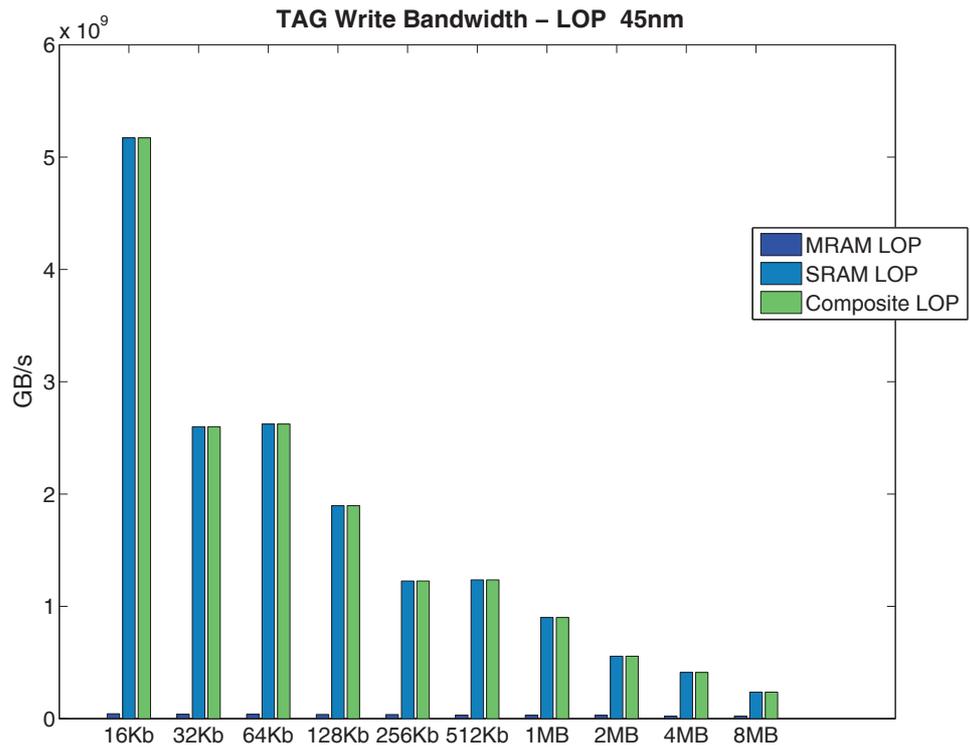


Figure F.20: Write Bandwidth.

Table F.19: Write Bandwidth (GB/s)

	MRAM LOP	SRAM LOP	Composite LOP
16KB	$4.047e - 02$	$4.816e + 00$	$4.816e + 00$
32KB	$3.632e - 02$	$2.421e + 00$	$2.421e + 00$
64KB	$3.550e - 02$	$2.443e + 00$	$2.443e + 00$
128KB	$3.466e - 02$	$1.766e + 00$	$1.766e + 00$
256KB	$3.383e - 02$	$1.142e + 00$	$1.142e + 00$
512KB	$2.917e - 02$	$1.151e + 00$	$1.151e + 00$
1MB	$2.856e - 02$	$8.387e - 01$	$8.387e - 01$
2MB	$2.793e - 02$	$5.179e - 01$	$5.179e - 01$
4MB	$2.122e - 02$	$3.830e - 01$	$3.830e - 01$
8MB	$2.086e - 02$	$2.195e - 01$	$2.195e - 01$

F.2 COMPOSITE BAANK 28nm

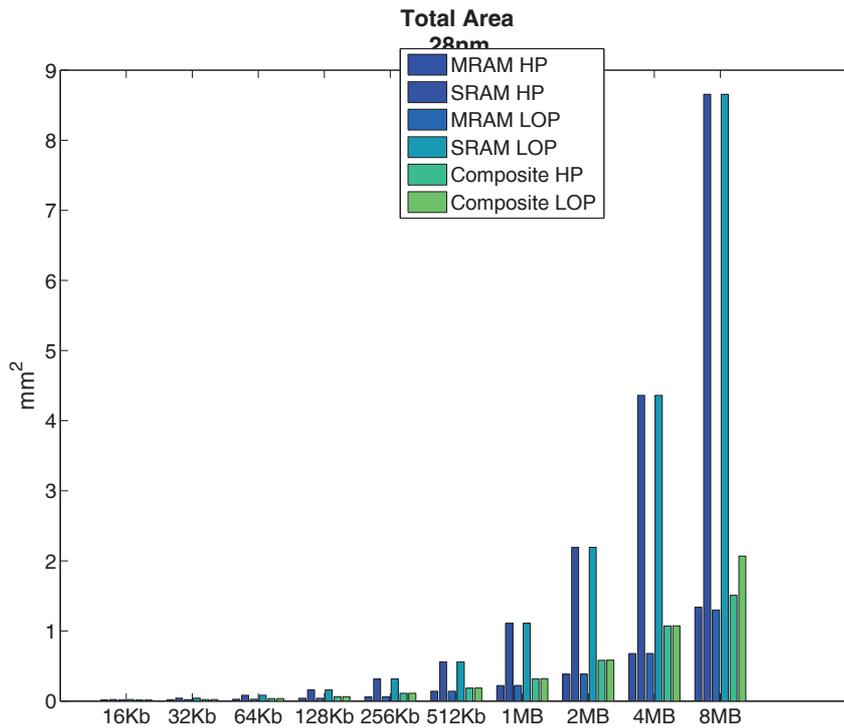


Figure F.21: Total Area.

Table F.20: Total Area (µm²).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0169	0.0220	0.0169	0.0220	0.0146	0.0146
32KB	0.0201	0.0424	0.0201	0.0424	0.0218	0.0219
64KB	0.0266	0.0828	0.0266	0.0829	0.0355	0.0356
128KB	0.0396	0.1602	0.0397	0.1604	0.0628	0.0629
256KB	0.0619	0.3168	0.0622	0.3171	0.1121	0.1124
512KB	0.1402	0.5594	0.1411	0.5596	0.1849	0.1857
1MB	0.2214	1.1107	0.2225	1.1111	0.3189	0.3199
2MB	0.3835	2.1925	0.3850	2.1930	0.5819	0.5833
4MB	0.6779	4.3613	0.6799	4.3623	1.0719	1.0738
8MB	1.3401	8.6547	1.2984	8.6549	1.5091	2.0691

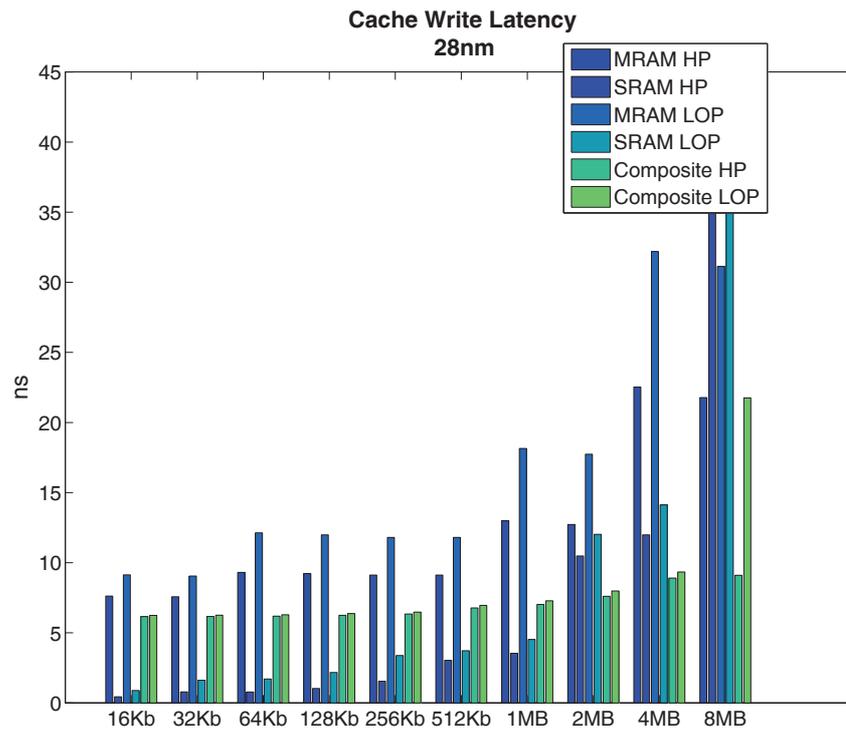


Figure F.22: Total Write Latency.

Table F.21: Cache Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	7.6208	0.4385	9.1320	0.8893	6.1556	6.2446
32KB	7.5755	0.7907	9.0482	1.6204	6.1661	6.2592
64KB	9.3097	0.7766	12.1337	1.7068	6.1856	6.2846
128KB	9.2224	1.0300	11.9828	2.1748	6.2508	6.3747
256KB	9.1170	1.5494	11.8007	3.3801	6.3432	6.4801
512KB	9.1179	3.0395	11.8022	3.7236	6.7750	6.9570
1MB	12.9919	3.5433	18.1500	4.5230	7.0267	7.2825
2MB	12.7278	10.4809	17.7366	12.0243	7.6025	7.9810
4MB	22.5341	11.9878	32.1938	14.1231	8.9019	9.3365
8MB	21.7745	38.2659	31.1254	41.9588	9.1050	21.7534

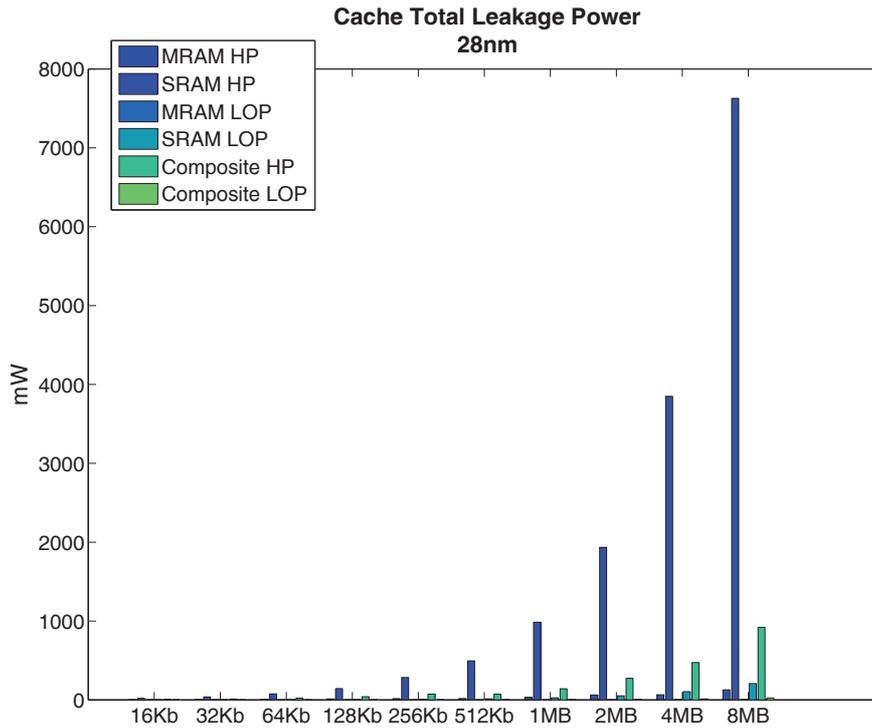


Figure F.23: Leakage Power.

Table F.22: Cache Total Leakage Power (mW).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	2.6045	20.6737	0.1125	0.5645	6.4742	0.1958
32KB	4.0383	38.9497	0.1516	1.0632	11.6903	0.3385
64KB	7.0716	76.6546	0.2349	2.0926	22.1283	0.6244
128KB	12.7499	144.4140	0.3914	3.9396	41.8138	1.1632
256KB	17.3429	285.0080	0.5178	7.7749	74.0194	2.0431
512KB	17.6883	496.1230	0.6769	13.5279	74.3456	2.1818
1MB	32.9403	985.7070	1.0977	26.8773	141.9979	4.0316
2MB	62.6929	1934.1000	1.9197	52.7265	274.5149	7.6560
4MB	65.0014	3849.3400	1.9818	104.9380	474.0453	13.0944
8MB	128.1740	7627.1200	3.6239	207.9090	920.9000	25.1955

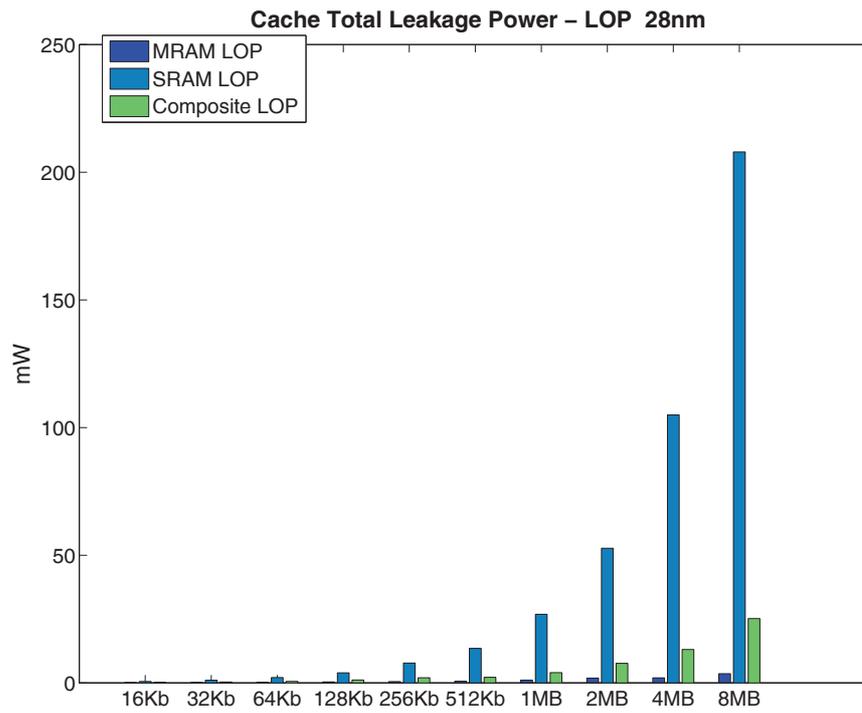


Figure F.24: Low-Power (LOP) Total Leakage, close-up into only the LOP banks observed in Figure F.23.

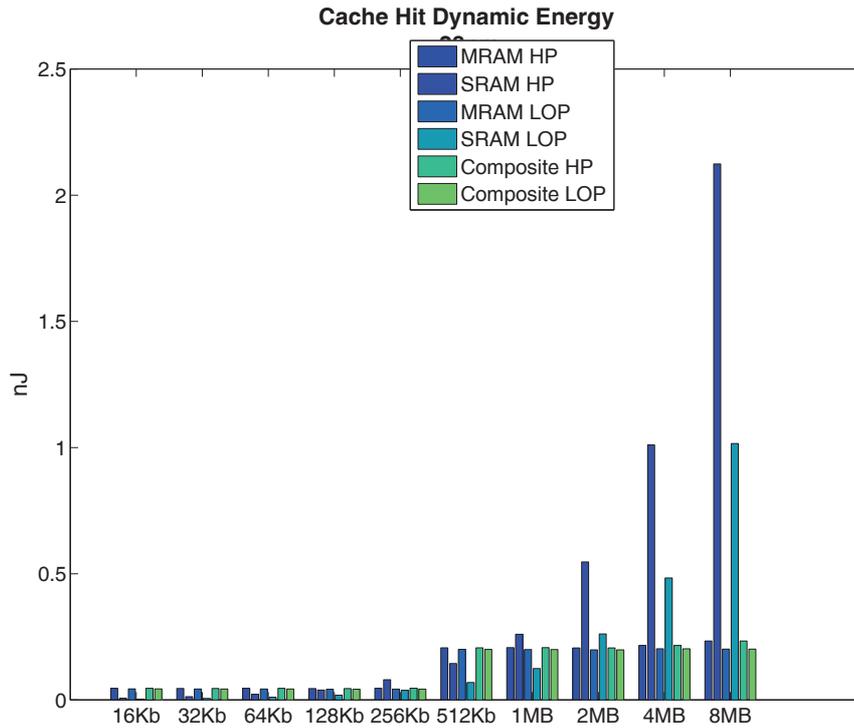


Figure F.25: Hit Dynamic Energy.

Table F.23: Cache Hit Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0459	0.0068	0.0439	0.0033	0.0459	0.0439
32KB	0.0453	0.0128	0.0432	0.0061	0.0453	0.0432
64KB	0.0457	0.0222	0.0432	0.0106	0.0457	0.0432
128KB	0.0451	0.0391	0.0425	0.0187	0.0451	0.0425
256KB	0.0464	0.0801	0.0428	0.0383	0.0464	0.0428
512KB	0.2064	0.1443	0.2001	0.0689	0.2064	0.2001
1MB	0.2072	0.2603	0.1999	0.1244	0.2072	0.1999
2MB	0.2059	0.5466	0.1986	0.2613	0.2059	0.1986
4MB	0.2159	1.0104	0.2028	0.4832	0.2159	0.2028
8MB	0.2336	2.1235	0.2015	1.0157	0.2336	0.2015

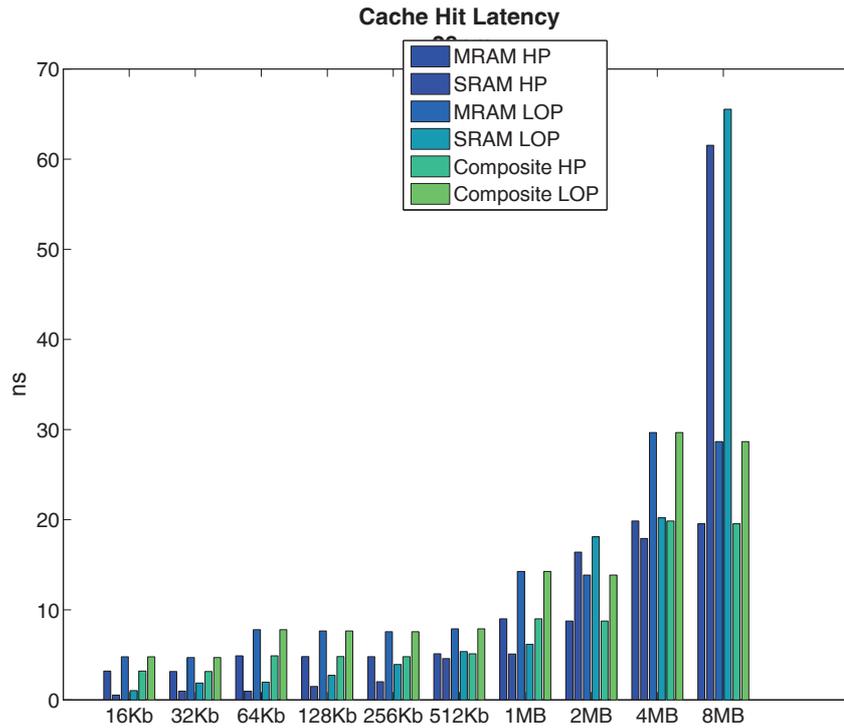


Figure F.26: Hit Latency.

Table F.24: Cache Hit Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.2005	0.5366	4.7837	1.0389	3.2005	4.7837
32KB	3.1556	0.9798	4.7004	1.8794	3.1556	4.7004
64KB	4.8952	0.9671	7.7962	1.9678	4.8952	7.7962
128KB	4.8092	1.5021	7.6462	2.7317	4.8092	7.6462
256KB	4.7959	2.0232	7.5658	3.9396	4.7959	7.5658
512KB	5.1139	4.5914	7.8917	5.3655	5.1139	7.8917
1MB	8.9944	5.0984	14.2518	6.1701	8.9944	14.2518
2MB	8.7464	16.4013	13.8563	18.1043	8.7464	13.8563
4MB	19.8497	17.9117	29.6634	20.2088	19.8497	29.6634
8MB	19.5485	61.5286	28.6538	65.5319	19.5485	28.6538

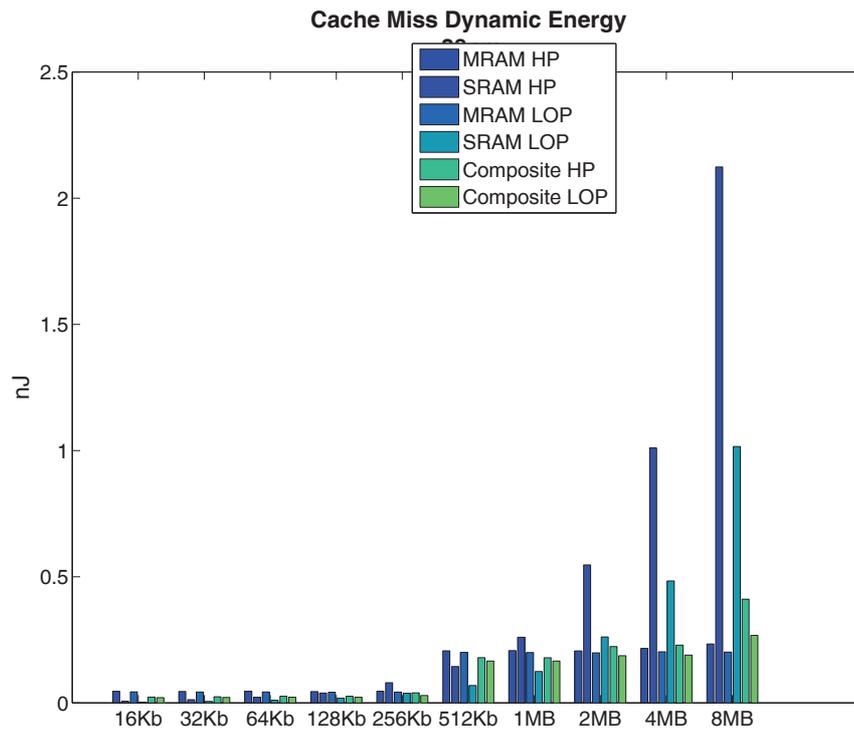


Figure F.27: Miss Dynamic Energy.

Table F.25: Cache Miss Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0459	0.0068	0.0439	0.0033	0.0231	0.0211
32KB	0.0453	0.0128	0.0432	0.0061	0.0242	0.0216
64KB	0.0457	0.0222	0.0432	0.0106	0.0263	0.0226
128KB	0.0451	0.0391	0.0425	0.0187	0.0262	0.0226
256KB	0.0464	0.0801	0.0428	0.0383	0.0401	0.0292
512KB	0.2064	0.1443	0.2001	0.0689	0.1796	0.1660
1MB	0.2072	0.2603	0.1999	0.1244	0.1791	0.1658
2MB	0.2059	0.5466	0.1986	0.2613	0.2232	0.1869
4MB	0.2159	1.0104	0.2028	0.4832	0.2289	0.1896
8MB	0.2336	2.1235	0.2015	1.0157	0.4112	0.2677

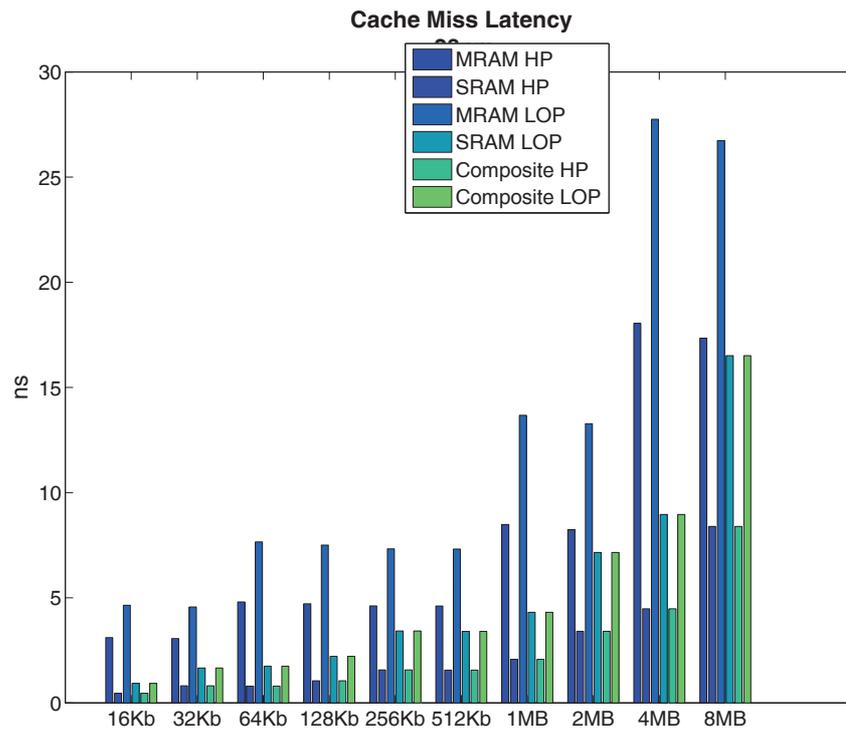


Figure F.28: Miss Latency.

Table F.26: Cache Miss Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.1050	0.4603	4.6437	0.9315	0.4603	0.9315
32KB	3.0600	0.8125	4.5602	1.6627	0.8125	1.6627
64KB	4.7994	0.7984	7.6557	1.7491	0.7984	1.7491
128KB	4.7123	1.0512	7.5046	2.2159	1.0512	2.2159
256KB	4.6104	1.5706	7.3259	3.4212	1.5706	3.4212
512KB	4.6061	1.5607	7.3173	3.3992	1.5607	3.3992
1MB	8.4852	2.0722	13.6752	4.3105	2.0722	4.3105
2MB	8.2346	3.4069	13.2756	7.1559	3.4069	7.1559
4MB	18.0507	4.4777	27.7518	8.9596	4.4777	8.9596
8MB	17.3433	8.3928	26.7375	16.5125	8.3928	16.5125

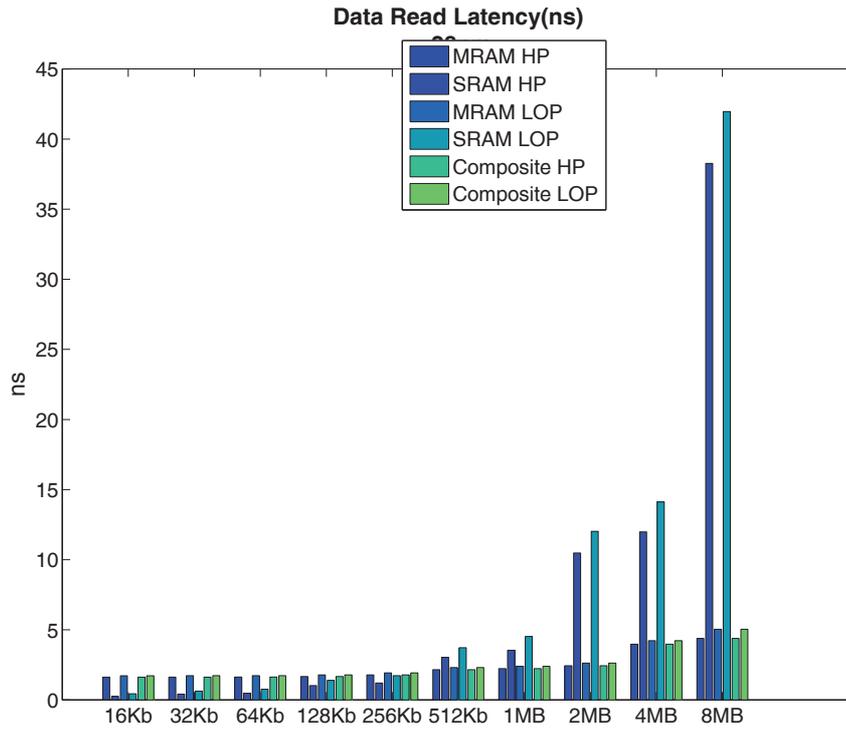


Figure F.29: DATA Read Latency.

Table F.27: DATA Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	1.6223	0.2537	1.7222	0.4389	1.6223	1.7222
32KB	1.6256	0.4114	1.7274	0.6305	1.6256	1.7274
64KB	1.6306	0.4769	1.7337	0.7661	1.6306	1.7337
128KB	1.6666	1.0206	1.7856	1.3957	1.6666	1.7856
256KB	1.7851	1.2076	1.9284	1.7345	1.7851	1.9284
512KB	2.1568	3.0395	2.3064	3.7236	2.1568	2.3064
1MB	2.2325	3.5433	2.4014	4.5230	2.2325	2.4014
2MB	2.4403	10.4809	2.6256	12.0243	2.4403	2.6256
4MB	3.9779	11.9878	4.2264	14.1231	3.9779	4.2264
8MB	4.3840	38.2659	5.0413	41.9588	4.3840	5.0413

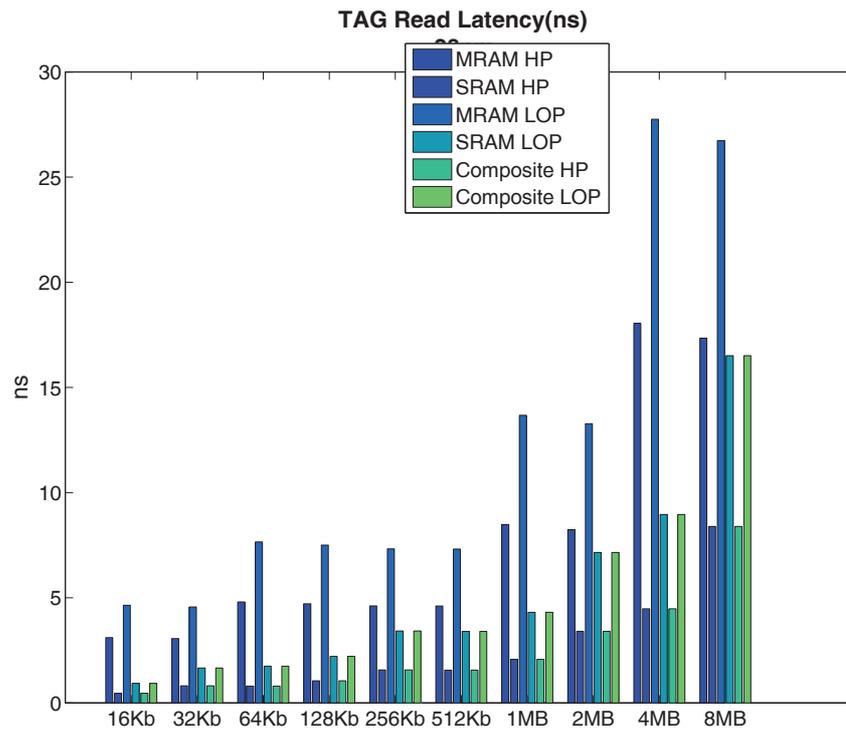


Figure F.30: TAG Read Latency.

Table F.28: TAG Read Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	3.1050	0.4603	4.6437	0.9315	0.4603	0.9315
32KB	3.0600	0.8125	4.5602	1.6627	0.8125	1.6627
64KB	4.7994	0.7984	7.6557	1.7491	0.7984	1.7491
128KB	4.7123	1.0512	7.5046	2.2159	1.0512	2.2159
256KB	4.6104	1.5706	7.3259	3.4212	1.5706	3.4212
512KB	4.6061	1.5607	7.3173	3.3992	1.5607	3.3992
1MB	8.4852	2.0722	13.6752	4.3105	2.0722	4.3105
2MB	8.2346	3.4069	13.2756	7.1559	3.4069	7.1559
4MB	18.0507	4.4777	27.7518	8.9596	4.4777	8.9596
8MB	17.3433	8.3928	26.7375	16.5125	8.3928	16.5125

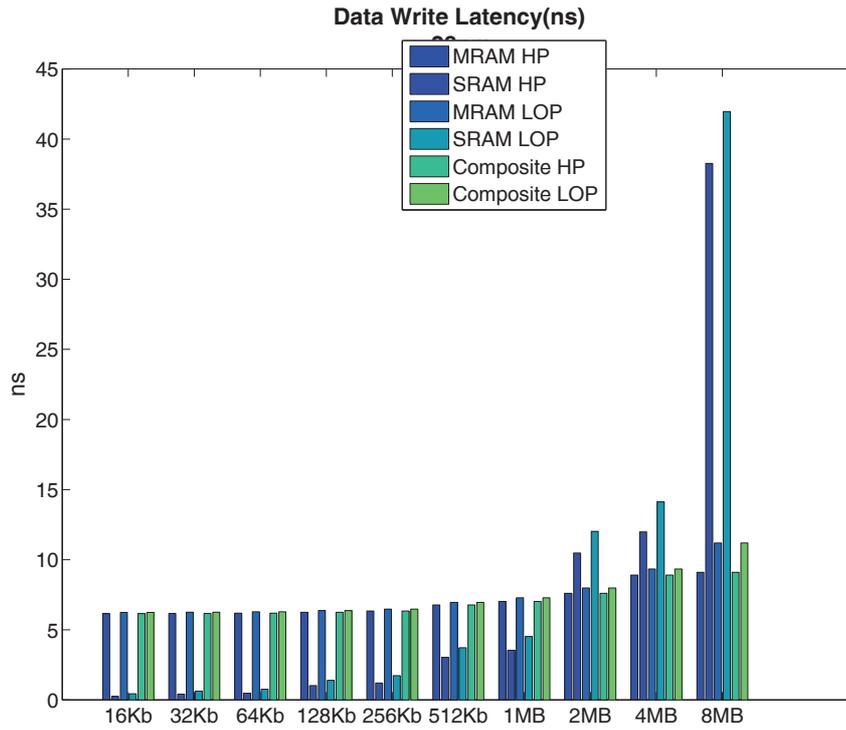


Figure F.31: DATA array matrix Write Latency.

Table F.29: DATA array matrix Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	6.1556	0.2537	6.2446	0.4389	6.1556	6.2446
32KB	6.1661	0.4114	6.2592	0.6305	6.1661	6.2592
64KB	6.1856	0.4769	6.2846	0.7661	6.1856	6.2846
128KB	6.2508	1.0206	6.3747	1.3957	6.2508	6.3747
256KB	6.3432	1.2076	6.4801	1.7345	6.3432	6.4801
512KB	6.7750	3.0395	6.9570	3.7236	6.7750	6.9570
1MB	7.0267	3.5433	7.2825	4.5230	7.0267	7.2825
2MB	7.6025	10.4809	7.9810	12.0243	7.6025	7.9810
4MB	8.9019	11.9878	9.3365	14.1231	8.9019	9.3365
8MB	9.1050	38.2659	11.1918	41.9588	9.1050	11.1918

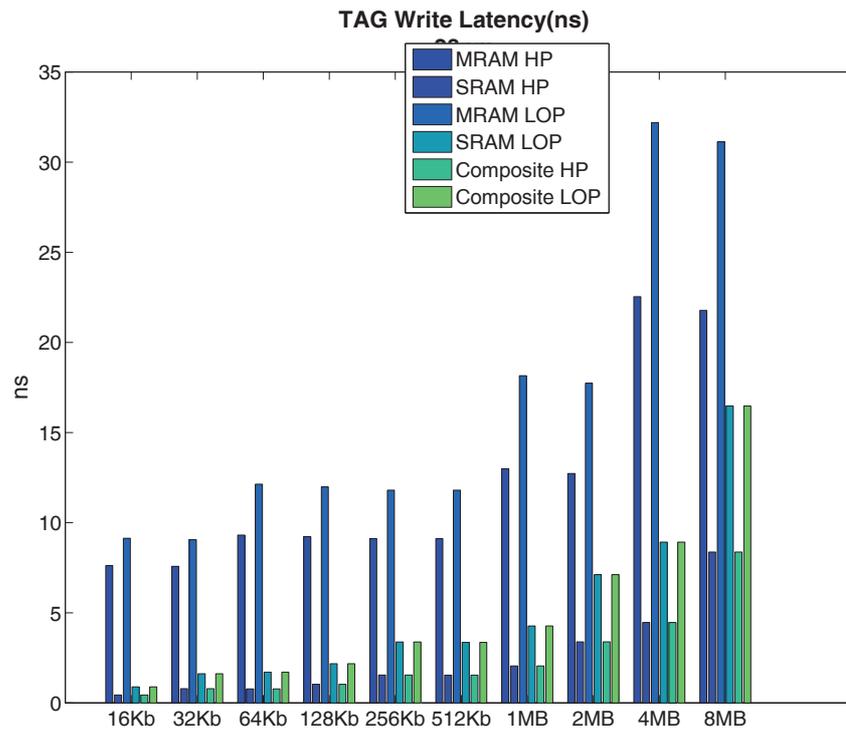


Figure F.32: TAG Write Latency.

Table F.30: TAG Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	7.6208	0.4385	9.1320	0.8893	0.4385	0.8893
32KB	7.5755	0.7907	9.0482	1.6204	0.7907	1.6204
64KB	9.3097	0.7766	12.1337	1.7068	0.7766	1.7068
128KB	9.2224	1.0300	11.9828	2.1748	1.0300	2.1748
256KB	9.1170	1.5494	11.8007	3.3801	1.5494	3.3801
512KB	9.1179	1.5395	11.8022	3.3582	1.5395	3.3582
1MB	12.9919	2.0509	18.1500	4.2694	2.0509	4.2694
2MB	12.7278	3.3857	17.7366	7.1148	3.3857	7.1148
4MB	22.5341	4.4570	32.1938	8.9197	4.4570	8.9197
8MB	21.7745	8.3721	31.1254	16.4726	8.3721	16.4726

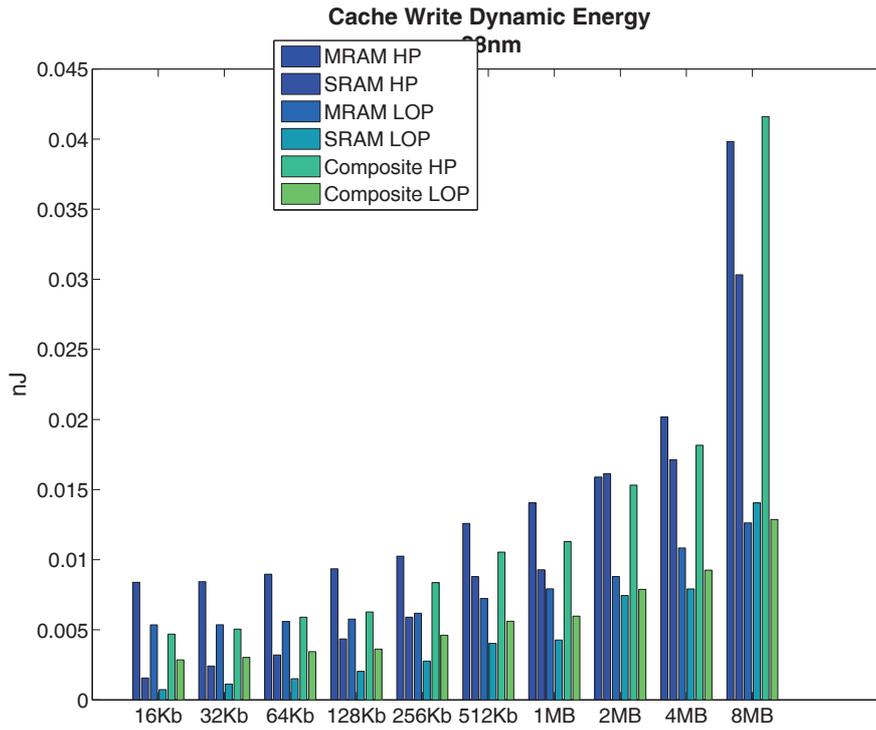


Figure F.33: CACHE Write Dynamic Energy

Table F.31: Cache Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	0.0084	0.0016	0.0054	0.0007	0.0047	0.0029
32KB	0.0084	0.0024	0.0054	0.0011	0.0050	0.0030
64KB	0.0090	0.0032	0.0056	0.0015	0.0059	0.0034
128KB	0.0094	0.0043	0.0058	0.0020	0.0063	0.0036
256KB	0.0102	0.0059	0.0062	0.0028	0.0084	0.0046
512KB	0.0126	0.0088	0.0072	0.0040	0.0105	0.0056
1MB	0.0141	0.0093	0.0079	0.0043	0.0113	0.0060
2MB	0.0159	0.0161	0.0088	0.0074	0.0153	0.0079
4MB	0.0202	0.0171	0.0108	0.0079	0.0182	0.0093
8MB	0.0398	0.0303	0.0126	0.0141	0.0416	0.0129

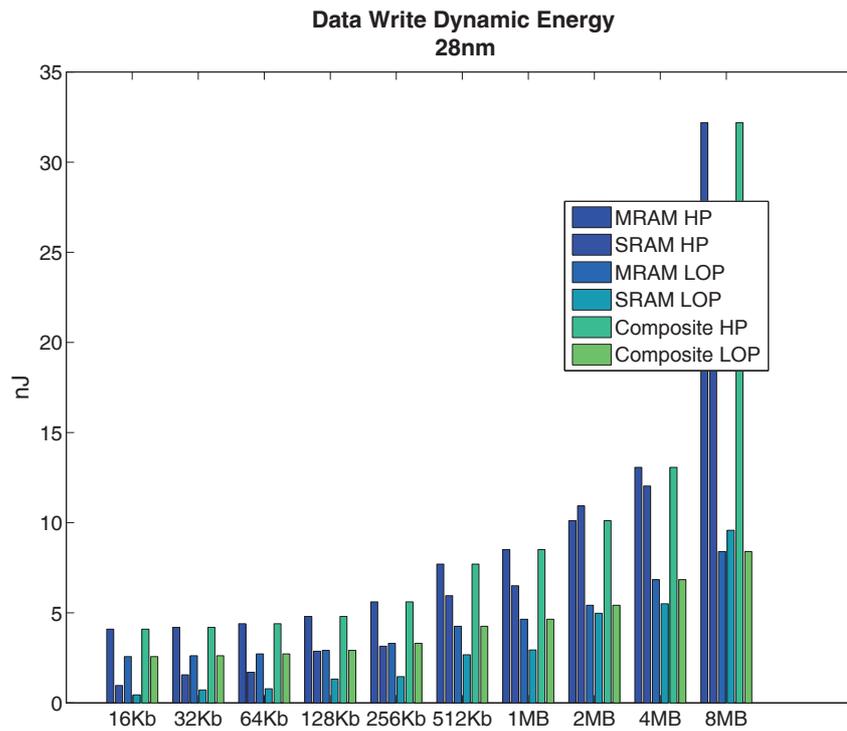


Figure F.34: DATA Dynamic Energy

Table F.32: DATA Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	4.0915	0.9664	2.5719	0.4442	4.0915	2.5719
32KB	4.1928	1.5611	2.6210	0.7195	4.1928	2.6210
64KB	4.3949	1.6992	2.7190	0.7855	4.3949	2.7190
128KB	4.7970	2.8691	2.9141	1.3248	4.7970	2.9141
256KB	5.6094	3.1463	3.3050	1.4572	5.6094	3.3050
512KB	7.7041	5.9473	4.2506	2.6743	7.7041	4.2506
1MB	8.5057	6.5006	4.6393	2.9387	8.5057	4.6393
2MB	10.1079	10.9298	5.4161	4.9736	10.1079	5.4161
4MB	13.0702	12.0353	6.8393	5.5018	13.0702	6.8393
8MB	32.1876	20.9087	8.3928	9.5791	32.1876	8.3928

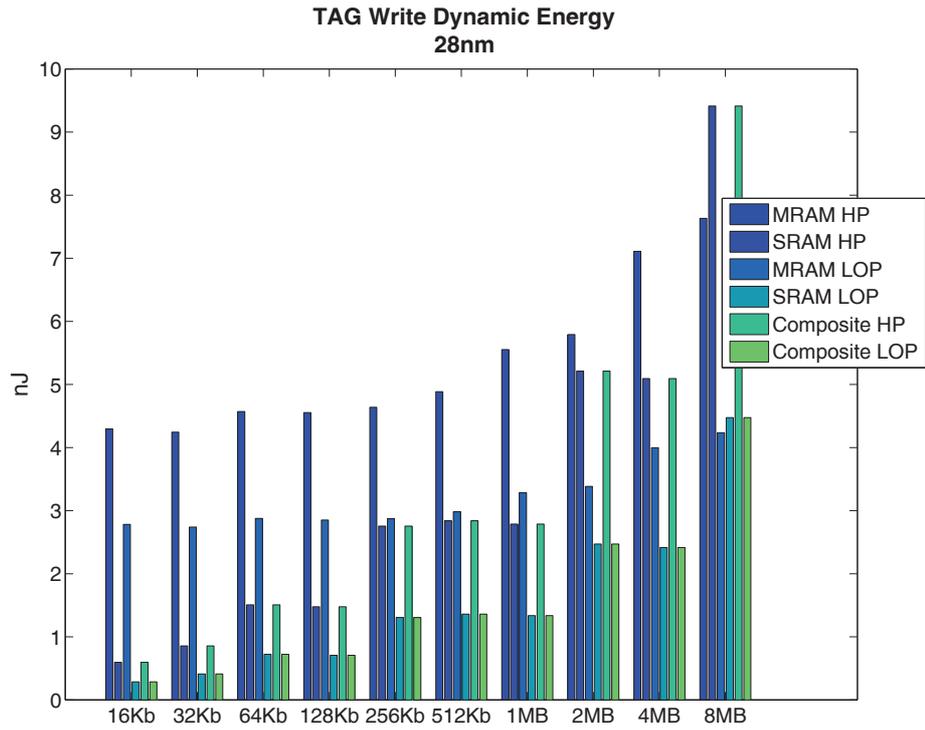


Figure F.35: TAG Dynamic Energy

Table F.33: TAG Write Dynamic Energy (nJ).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	4.2948	0.5962	2.7823	0.2868	0.5962	0.2868
32KB	4.2440	0.8568	2.7392	0.4114	0.8568	0.4114
64KB	4.5703	1.5070	2.8772	0.7224	1.5070	0.7224
128KB	4.5535	1.4767	2.8507	0.7077	1.4767	0.7077
256KB	4.6390	2.7541	2.8738	1.3061	2.7541	1.3061
512KB	4.8844	2.8399	2.9834	1.3608	2.8399	1.3608
1MB	5.5534	2.7874	3.2861	1.3356	2.7874	1.3356
2MB	5.7920	5.2110	3.3839	2.4711	5.2110	2.4711
4MB	7.1098	5.0923	3.9978	2.4147	5.0923	2.4147
8MB	7.6326	9.4130	4.2337	4.4750	9.4130	4.4750

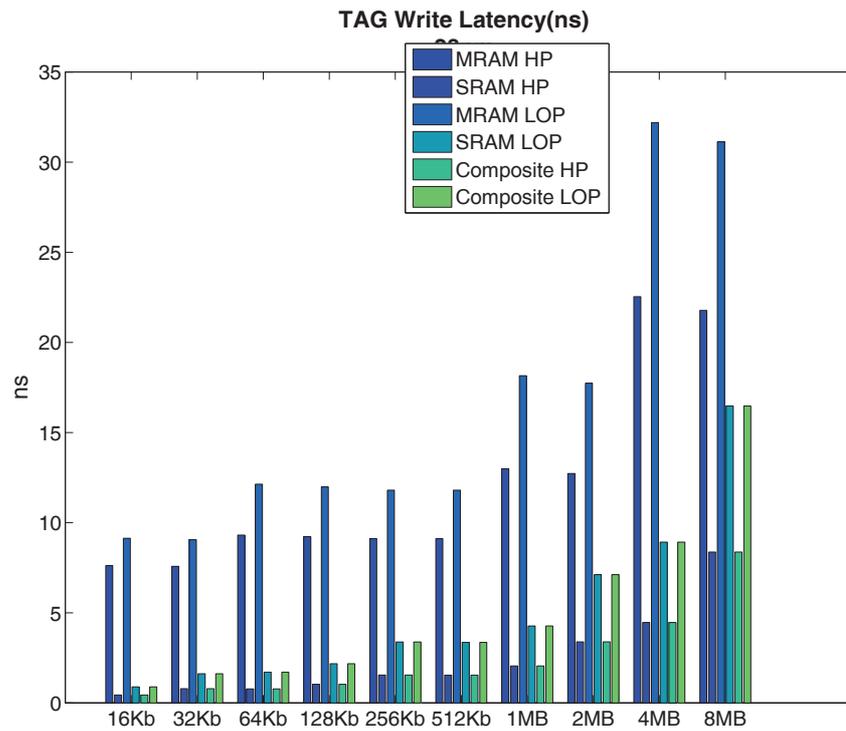


Figure F.36: TAG Dynamic Energy

Table F.34: TAG Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	7.6208	0.4385	9.1320	0.8893	0.4385	0.8893
32KB	7.5755	0.7907	9.0482	1.6204	0.7907	1.6204
64KB	9.3097	0.7766	12.1337	1.7068	0.7766	1.7068
128KB	9.2224	1.0300	11.9828	2.1748	1.0300	2.1748
256KB	9.1170	1.5494	11.8007	3.3801	1.5494	3.3801
512KB	9.1179	1.5395	11.8022	3.3582	1.5395	3.3582
1MB	12.9919	2.0509	18.1500	4.2694	2.0509	4.2694
2MB	12.7278	3.3857	17.7366	7.1148	3.3857	7.1148
4MB	22.5341	4.4570	32.1938	8.9197	4.4570	8.9197
8MB	21.7745	8.3721	31.1254	16.4726	8.3721	16.4726

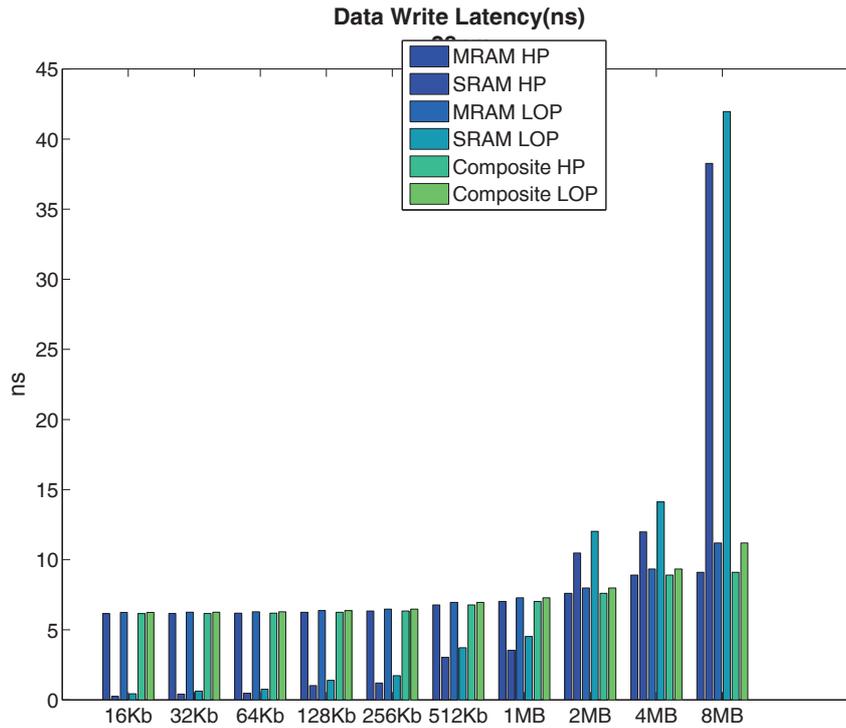


Figure F.37: DATA Write Latency

Table F.35: DATA Write Latency (ns).

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	6.1556	0.2537	6.2446	0.4389	6.1556	6.2446
32KB	6.1661	0.4114	6.2592	0.6305	6.1661	6.2592
64KB	6.1856	0.4769	6.2846	0.7661	6.1856	6.2846
128KB	6.2508	1.0206	6.3747	1.3957	6.2508	6.3747
256KB	6.3432	1.2076	6.4801	1.7345	6.3432	6.4801
512KB	6.7750	3.0395	6.9570	3.7236	6.7750	6.9570
1MB	7.0267	3.5433	7.2825	4.5230	7.0267	7.2825
2MB	7.6025	10.4809	7.9810	12.0243	7.6025	7.9810
4MB	8.9019	11.9878	9.3365	14.1231	8.9019	9.3365
8MB	9.1050	38.2659	11.1918	41.9588	9.1050	11.1918

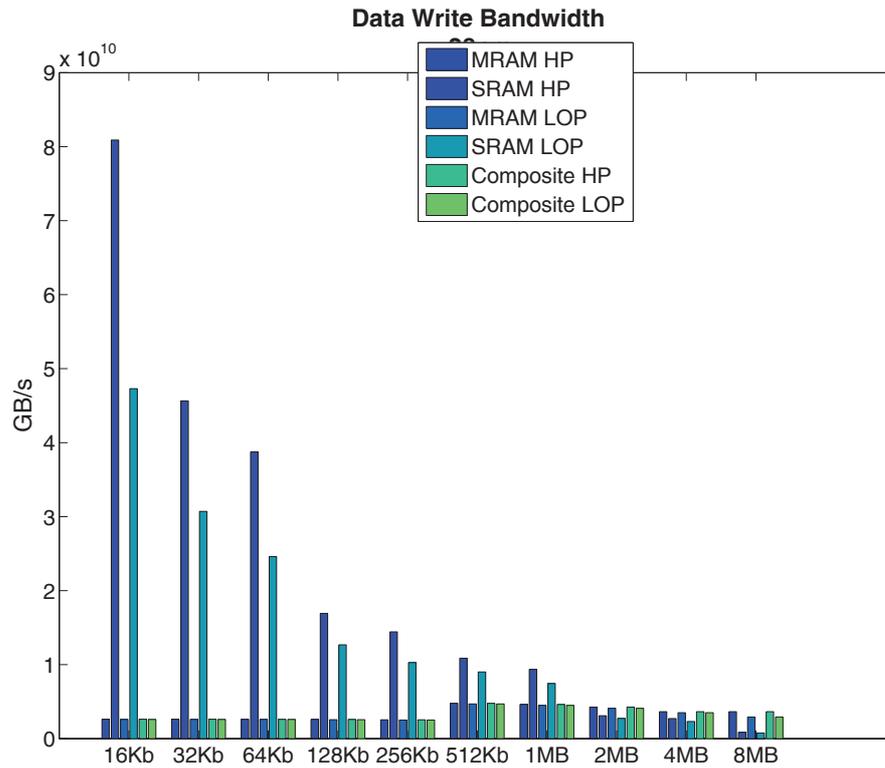


Figure F.38: Write Bandwidth.

Table F.36: Write Bandwidth (GB/s)

	MRAM HP	SRAM HP	MRAM LOP	SRAM LOP	Composite HP	Composite LOP
16KB	2.442	75.33	2.423	44.04	2.442	2.423
32KB	2.438	42.50	2.419	28.61	2.438	2.419
64KB	2.431	36.09	2.410	22.90	2.431	2.410
128KB	2.414	15.76	2.389	11.81	2.414	2.389
256KB	2.380	13.45	2.352	9.572	2.380	2.352
512KB	4.448	10.12	4.367	8.399	4.448	4.367
1MB	4.300	8.718	4.191	6.941	4.300	4.191
2MB	3.974	2.886	3.823	2.539	3.974	3.823
4MB	3.392	2.523	3.266	2.161	3.392	3.266
8MB	3.392	0.7836	2.723	0.7180	3.392	2.723

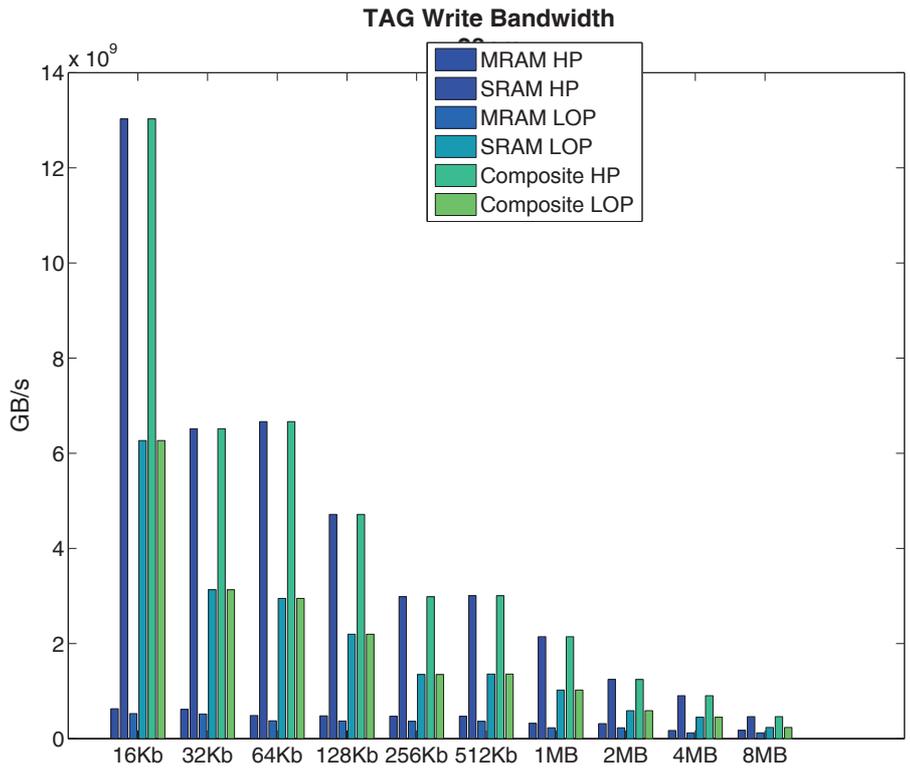


Figure F.39: Write Bandwidth.

Table F.37: Write Bandwidth (GB/s)

	MRAM HP		SRAM HP		MRAM LOP		SRAM LOP		Composite HP		Composite LOP
16KB	5.845e 01	-	1.213e 01	+	4.895e 01	-	5.837e 00	+	1.213e 01	+	5.837e 00
32KB	5.726e 01	-	6.067e 00	+	4.811e 01	-	2.916e 00	+	6.067e 00	+	2.916e 00
64KB	4.530e 01	-	6.207e 00	+	3.484e 01	-	2.746e 00	+	6.207e 00	+	2.746e 00
128KB	4.455e 01	-	4.390e 00	+	3.440e 01	-	2.044e 00	+	4.390e 00	+	2.044e 00
256KB	4.379e 01	-	2.779e 00	+	3.394e 01	-	1.255e 00	+	2.779e 00	+	1.255e 00
512KB	4.379e 01	-	2.798e 00	+	3.394e 01	-	1.264e 00	+	2.798e 00	+	1.264e 00
1MB	2.977e 01	-	1.996e 00	+	2.135e 01	-	9.509e 01	-	1.996e 00	+	9.509e 01
2MB	2.949e 01	-	1.158e 00	+	2.120e 01	-	5.479e 01	-	1.158e 00	+	5.479e 01
4MB	1.609e 01	-	8.418e 01	-	1.128e 01	-	4.196e 01	-	8.418e 01	-	4.196e 01
8MB	1.612e 01	-	4.313e 01	-	1.129e 01	-	2.190e 01	-	4.313e 01	-	2.190e 01

F.3 LOP 28nm

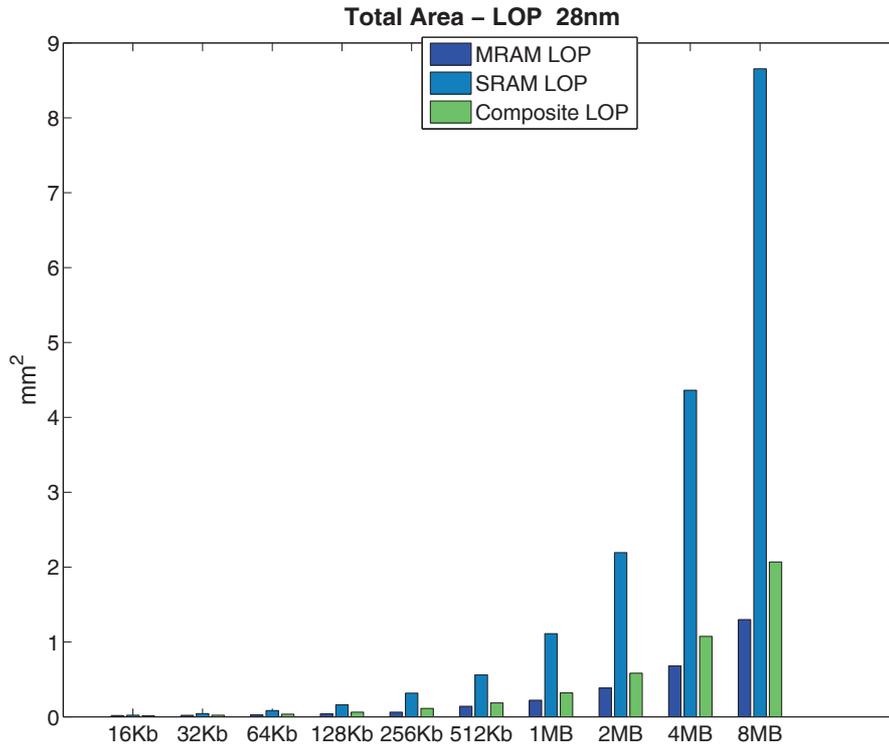


Figure F.40: Total Area.

Table F.38: Total Area (μm²).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0169	0.0220	0.0146
32KB	0.0201	0.0424	0.0219
64KB	0.0266	0.0829	0.0356
128KB	0.0397	0.1604	0.0629
256KB	0.0622	0.3171	0.1124
512KB	0.1411	0.5596	0.1857
1MB	0.2225	1.1111	0.3199
2MB	0.3850	2.1930	0.5833
4MB	0.6799	4.3623	1.0738
8MB	1.2984	8.6549	2.0691

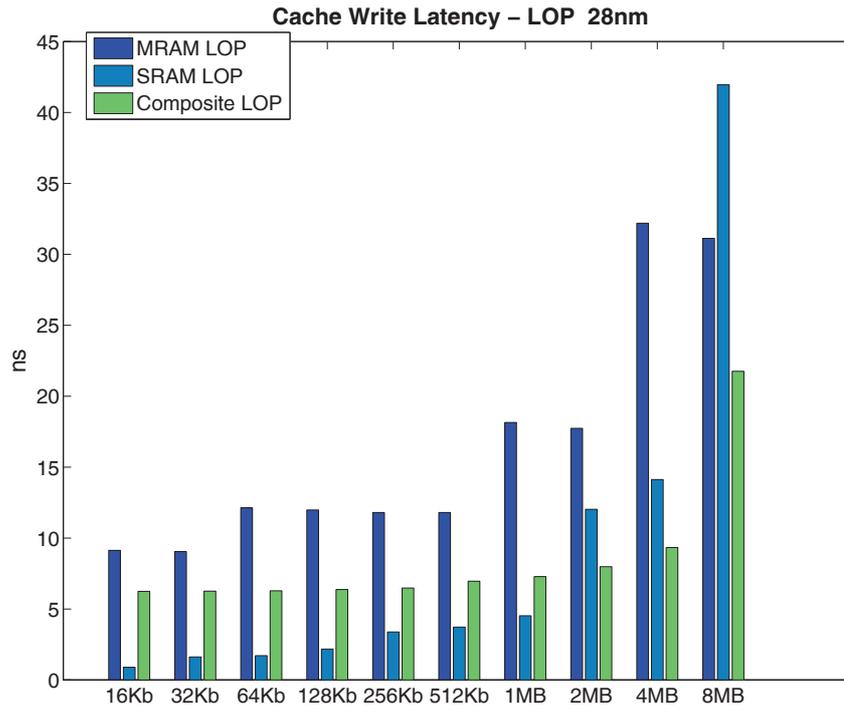


Figure F.41: Total Write Latency.

Table F.39: Cache Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	9.1320	0.8893	6.2446
32KB	9.0482	1.6204	6.2592
64KB	12.1337	1.7068	6.2846
128KB	11.9828	2.1748	6.3747
256KB	11.8007	3.3801	6.4801
512KB	11.8022	3.7236	6.9570
1MB	18.1500	4.5230	7.2825
2MB	17.7366	12.0243	7.9810
4MB	32.1938	14.1231	9.3365
8MB	31.1254	41.9588	21.7534

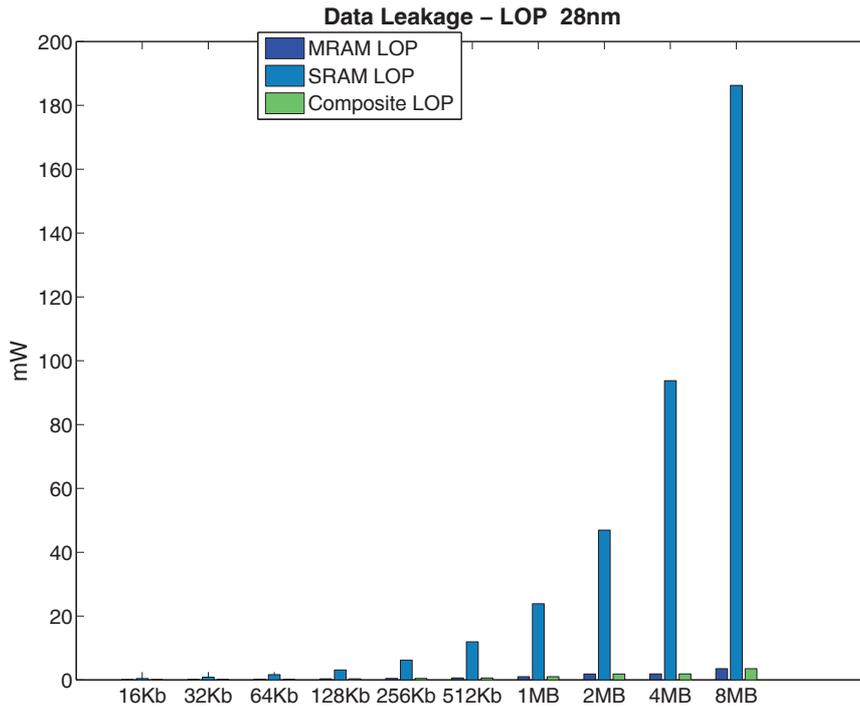


Figure F.42: CACHE Data Array Leakage Power.

Table F.40: Cache Data Array Leakage Power (nW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0804	0.4492	0.0804
32KB	0.1197	0.8444	0.1197
64KB	0.1978	1.6660	0.1978
128KB	0.3538	3.1302	0.3538
256KB	0.4794	6.2112	0.4794
512KB	0.6165	11.9625	0.6165
1MB	1.0271	23.8729	1.0271
2MB	1.8469	46.9174	1.8469
4MB	1.8920	93.7358	1.8920
8MB	3.5315	186.2450	3.5315

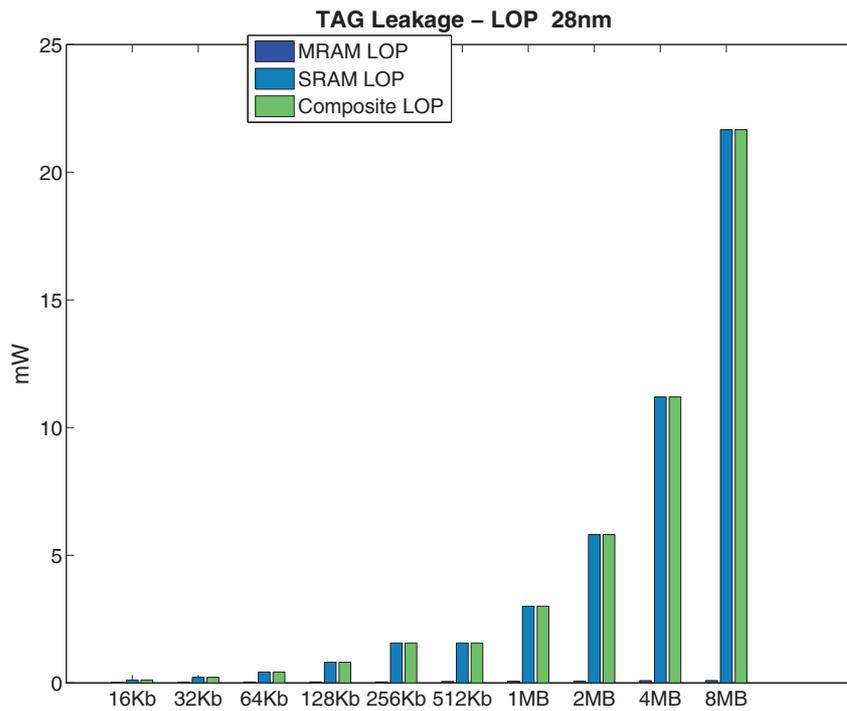


Figure F.43: TAG Leakage Power.

Table F.41: Cache Tag Array Leakage Power (nW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0321	0.1153	0.1153
32KB	0.0319	0.2188	0.2188
64KB	0.0371	0.4267	0.4267
128KB	0.0377	0.8094	0.8094
256KB	0.0384	1.5636	1.5636
512KB	0.0605	1.5653	1.5653
1MB	0.0705	3.0045	3.0045
2MB	0.0728	5.8092	5.8092
4MB	0.0899	11.2024	11.2024
8MB	0.0924	21.6640	21.6640

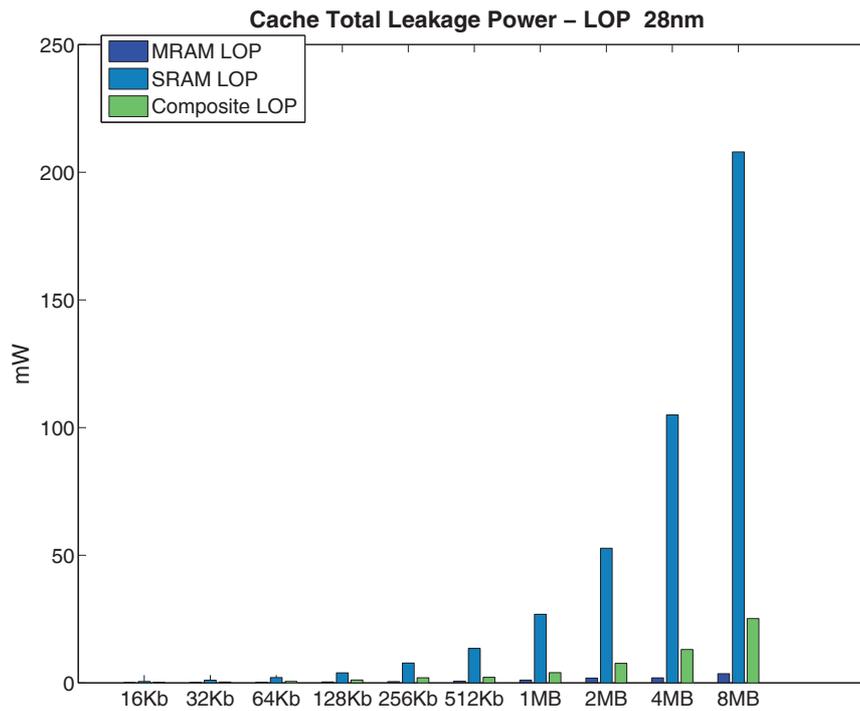


Figure F.44: Leakage Power.

Table F.42: Cache Total Leakage Power (mW).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.1125	0.5645	0.1958
32KB	0.1516	1.0632	0.3385
64KB	0.2349	2.0926	0.6244
128KB	0.3914	3.9396	1.1632
256KB	0.5178	7.7749	2.0431
512KB	0.6769	13.5279	2.1818
1MB	1.0977	26.8773	4.0316
2MB	1.9197	52.7265	7.6560
4MB	1.9818	104.9380	13.0944
8MB	3.6239	207.9090	25.1955

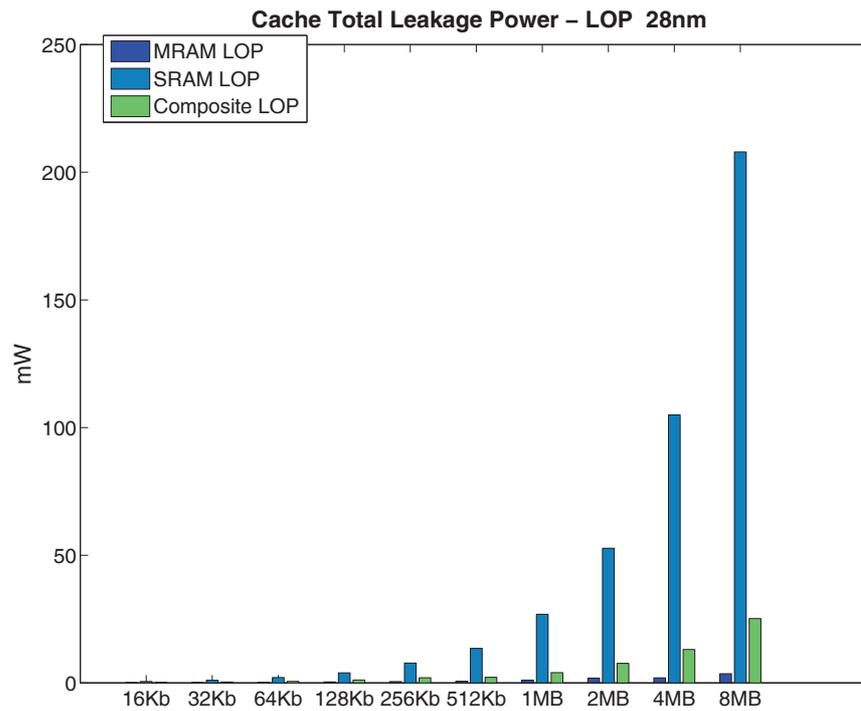


Figure F.45: Low-Power (LOP) Total Leakage, zoom into it to observe only the LOP banks observed in Figure F.44.

Table F.43: Cache Hit Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0439	0.0033	0.0439
32KB	0.0432	0.0061	0.0432
64KB	0.0432	0.0106	0.0432
128KB	0.0425	0.0187	0.0425
256KB	0.0428	0.0383	0.0428
512KB	0.2001	0.0689	0.2001
1MB	0.1999	0.1244	0.1999
2MB	0.1986	0.2613	0.1986
4MB	0.2028	0.4832	0.2028
8MB	0.2015	1.0157	0.2015

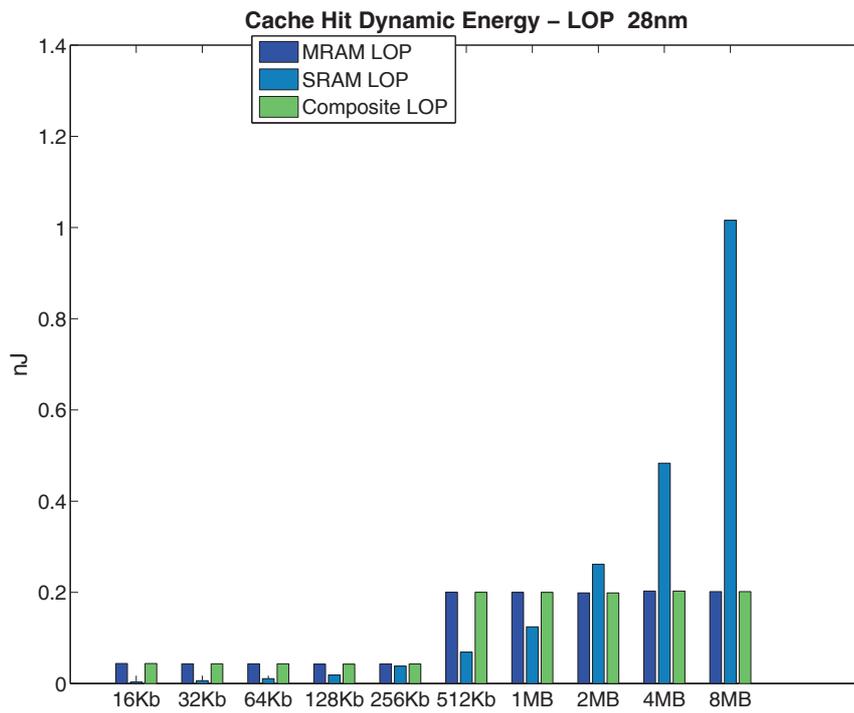


Figure F.46: CACHE Hit Dynamic Energy.

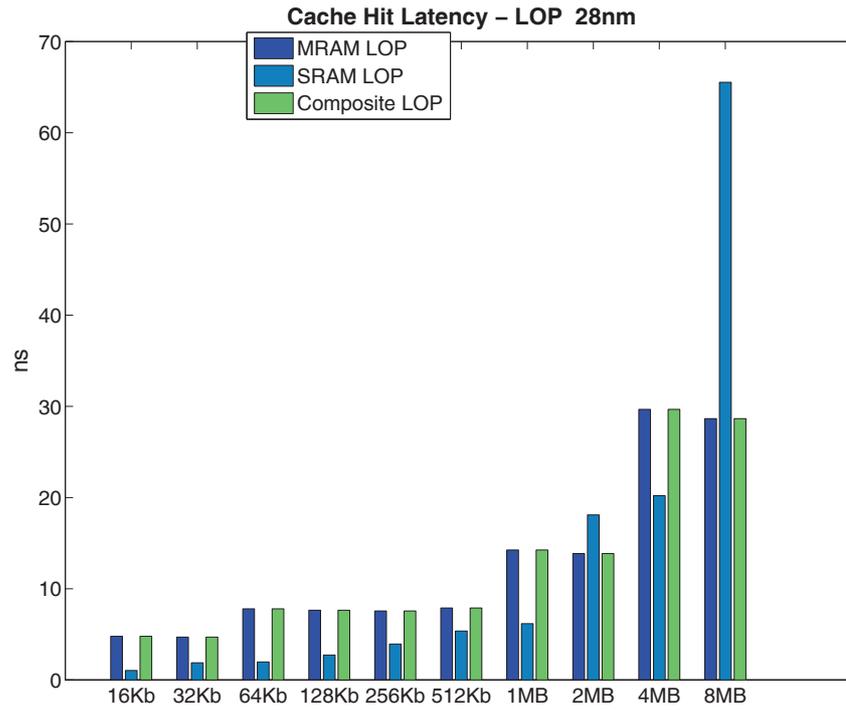


Figure F.47: CACHE LOP Hit Latency.

Table F.44: Cache Hit Latency - LOP (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	4.7837	1.0389	4.7837
32KB	4.7004	1.8794	4.7004
64KB	7.7962	1.9678	7.7962
128KB	7.6462	2.7317	7.6462
256KB	7.5658	3.9396	7.5658
512KB	7.8917	5.3655	7.8917
1MB	14.2518	6.1701	14.2518
2MB	13.8563	18.1043	13.8563
4MB	29.6634	20.2088	29.6634
8MB	28.6538	65.5319	28.6538

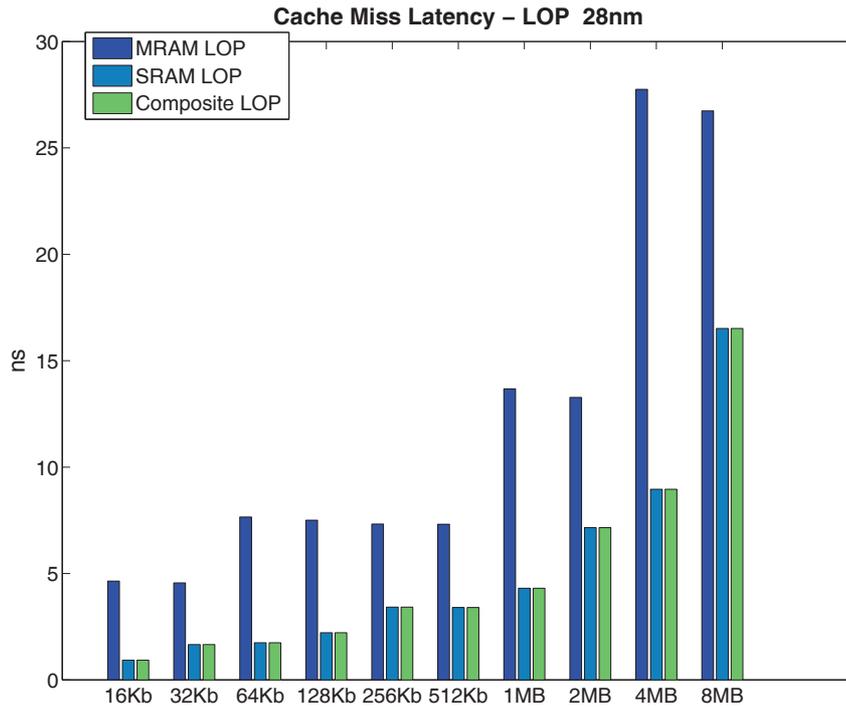


Figure F.48: CACHE LOP Miss Latency.

Table F.45: Cache Miss Latency LOP (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	4.6437	0.9315	0.9315
32KB	4.5602	1.6627	1.6627
64KB	7.6557	1.7491	1.7491
128KB	7.5046	2.2159	2.2159
256KB	7.3259	3.4212	3.4212
512KB	7.3173	3.3992	3.3992
1MB	13.6752	4.3105	4.3105
2MB	13.2756	7.1559	7.1559
4MB	27.7518	8.9596	8.9596
8MB	26.7375	16.5125	16.5125

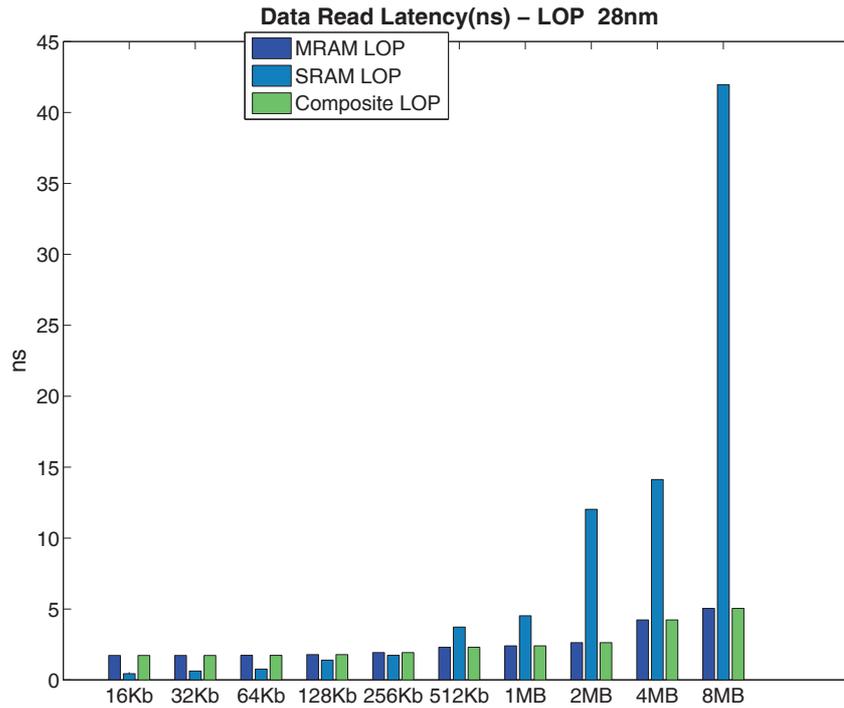


Figure F.49: DATA Read Latency.

Table F.46: DATA Read Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	1.7222	0.4389	1.7222
32KB	1.7274	0.6305	1.7274
64KB	1.7337	0.7661	1.7337
128KB	1.7856	1.3957	1.7856
256KB	1.9284	1.7345	1.9284
512KB	2.3064	3.7236	2.3064
1MB	2.4014	4.5230	2.4014
2MB	2.6256	12.0243	2.6256
4MB	4.2264	14.1231	4.2264
8MB	5.0413	41.9588	5.0413

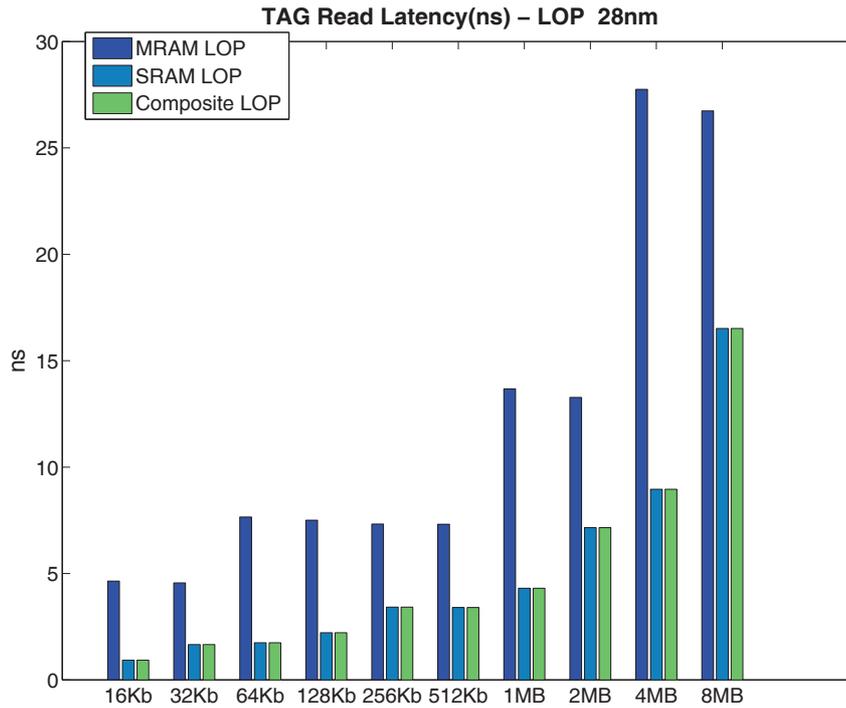


Figure F.50: TAG Read Latency.

Table F.47: TAG Read Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	4.6437	0.9315	0.9315
32KB	4.5602	1.6627	1.6627
64KB	7.6557	1.7491	1.7491
128KB	7.5046	2.2159	2.2159
256KB	7.3259	3.4212	3.4212
512KB	7.3173	3.3992	3.3992
1MB	13.6752	4.3105	4.3105
2MB	13.2756	7.1559	7.1559
4MB	27.7518	8.9596	8.9596
8MB	26.7375	16.5125	16.5125

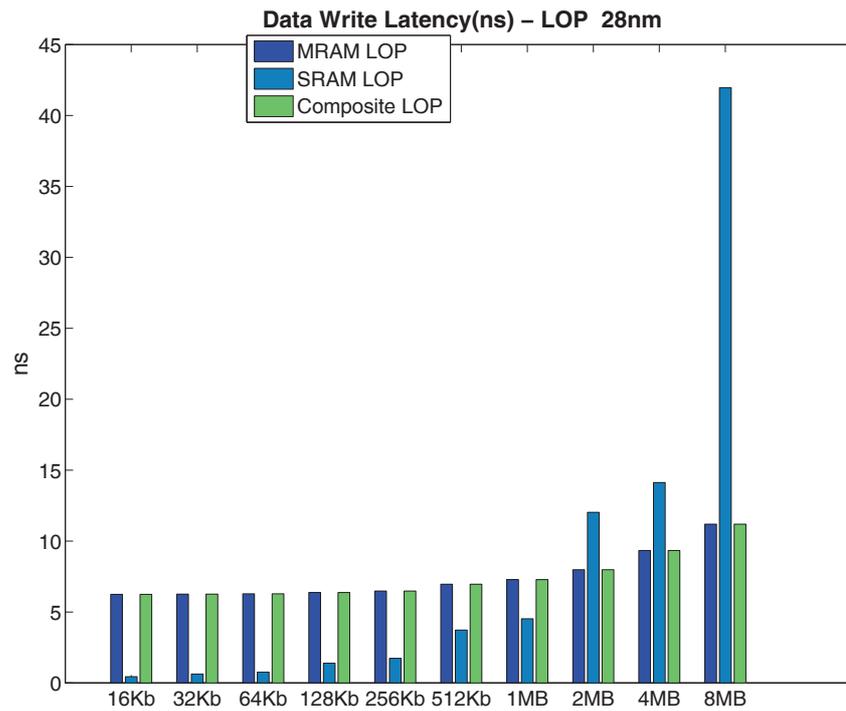


Figure F.51: DATA Write Latency.

Table F.48: DATA Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	6.2446	0.4389	6.2446
32KB	6.2592	0.6305	6.2592
64KB	6.2846	0.7661	6.2846
128KB	6.3747	1.3957	6.3747
256KB	6.4801	1.7345	6.4801
512KB	6.9570	3.7236	6.9570
1MB	7.2825	4.5230	7.2825
2MB	7.9810	12.0243	7.9810
4MB	9.3365	14.1231	9.3365
8MB	11.1918	41.9588	11.1918

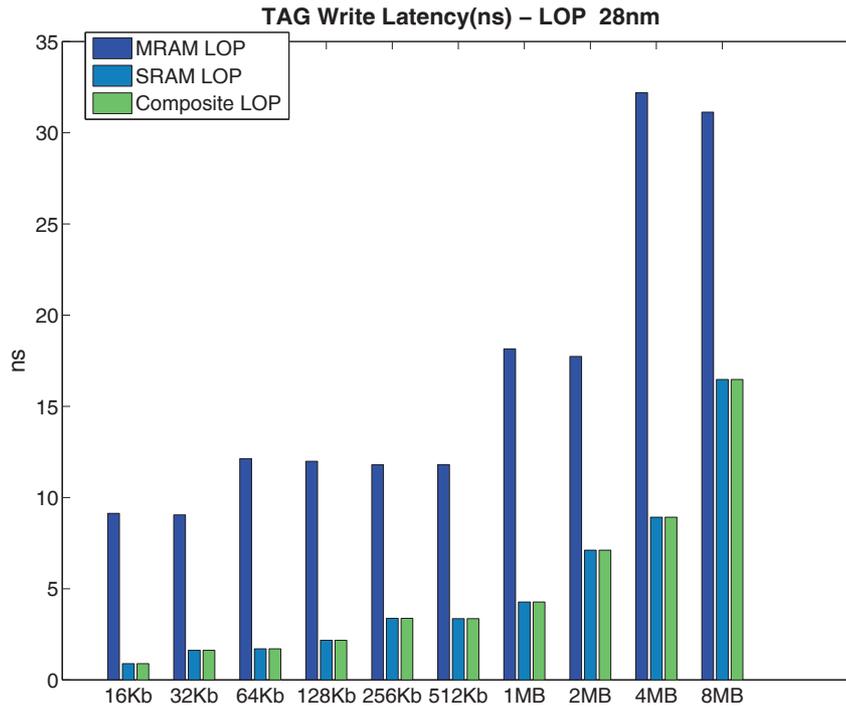


Figure F.52: TAG Write Latency.

Table F.49: TAG Read Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	9.1320	0.8893	0.8893
32KB	9.0482	1.6204	1.6204
64KB	12.1337	1.7068	1.7068
128KB	11.9828	2.1748	2.1748
256KB	11.8007	3.3801	3.3801
512KB	11.8022	3.3582	3.3582
1MB	18.1500	4.2694	4.2694
2MB	17.7366	7.1148	7.1148
4MB	32.1938	8.9197	8.9197
8MB	31.1254	16.4726	16.4726

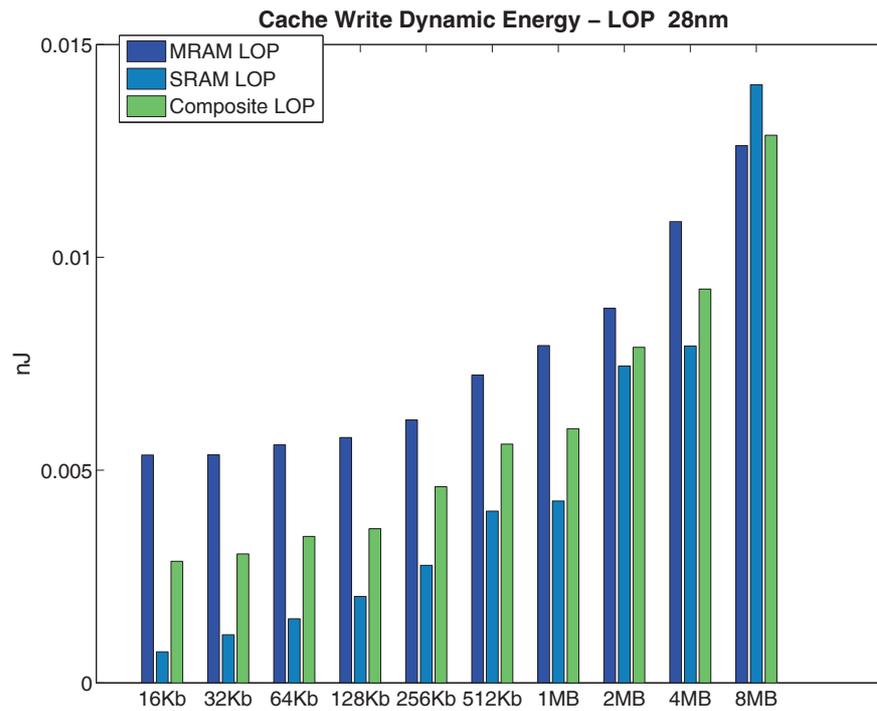


Figure F.53: CACHE Write Dynamic Energy

Table F.50: Cache Write Dynamic Energy - LOP (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	0.0054	0.0007	0.0029
32KB	0.0054	0.0011	0.0030
64KB	0.0056	0.0015	0.0034
128KB	0.0058	0.0020	0.0036
256KB	0.0062	0.0028	0.0046
512KB	0.0072	0.0040	0.0056
1MB	0.0079	0.0043	0.0060
2MB	0.0088	0.0074	0.0079
4MB	0.0108	0.0079	0.0093
8MB	0.0126	0.0141	0.0129

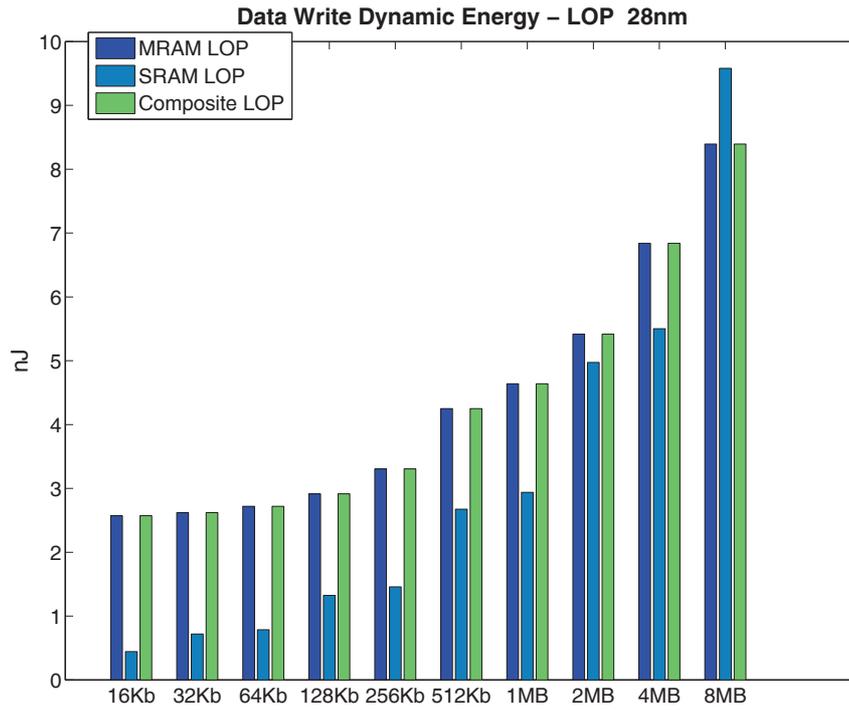


Figure F.54: DATA Write Dynamic Energy

Table F.51: DATA Write Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	2.5719	0.4442	2.5719
32KB	2.6210	0.7195	2.6210
64KB	2.7190	0.7855	2.7190
128KB	2.9141	1.3248	2.9141
256KB	3.3050	1.4572	3.3050
512KB	4.2506	2.6743	4.2506
1MB	4.6393	2.9387	4.6393
2MB	5.4161	4.9736	5.4161
4MB	6.8393	5.5018	6.8393
8MB	8.3928	9.5791	8.3928

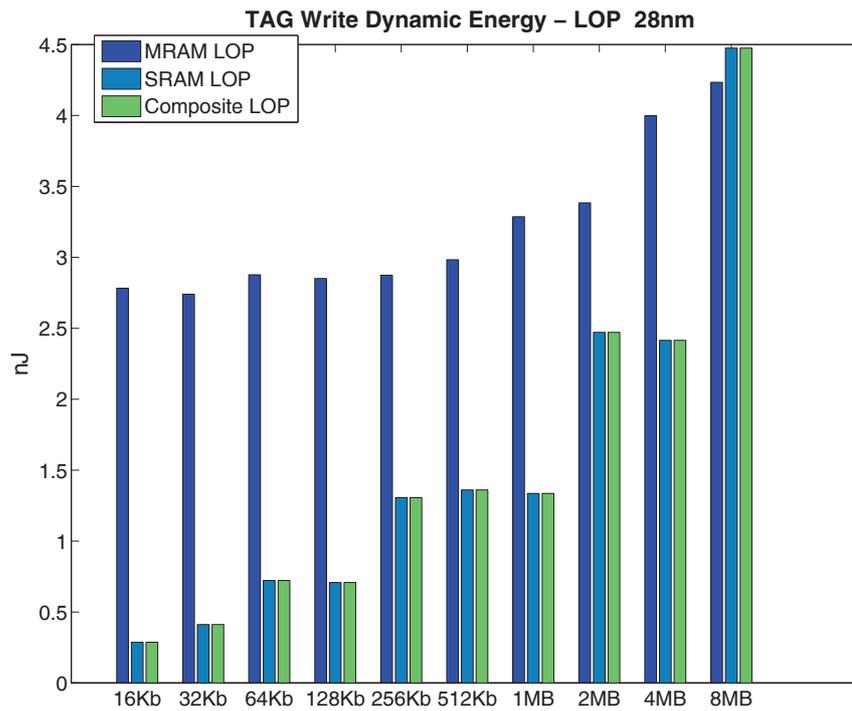


Figure F.55: TAG Write Dynamic Energy

Table F.52: TAG Write Dynamic Energy (nJ).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	2.7823	0.2868	0.2868
32KB	2.7392	0.4114	0.4114
64KB	2.8772	0.7224	0.7224
128KB	2.8507	0.7077	0.7077
256KB	2.8738	1.3061	1.3061
512KB	2.9834	1.3608	1.3608
1MB	3.2861	1.3356	1.3356
2MB	3.3839	2.4711	2.4711
4MB	3.9978	2.4147	2.4147
8MB	4.2337	4.4750	4.4750

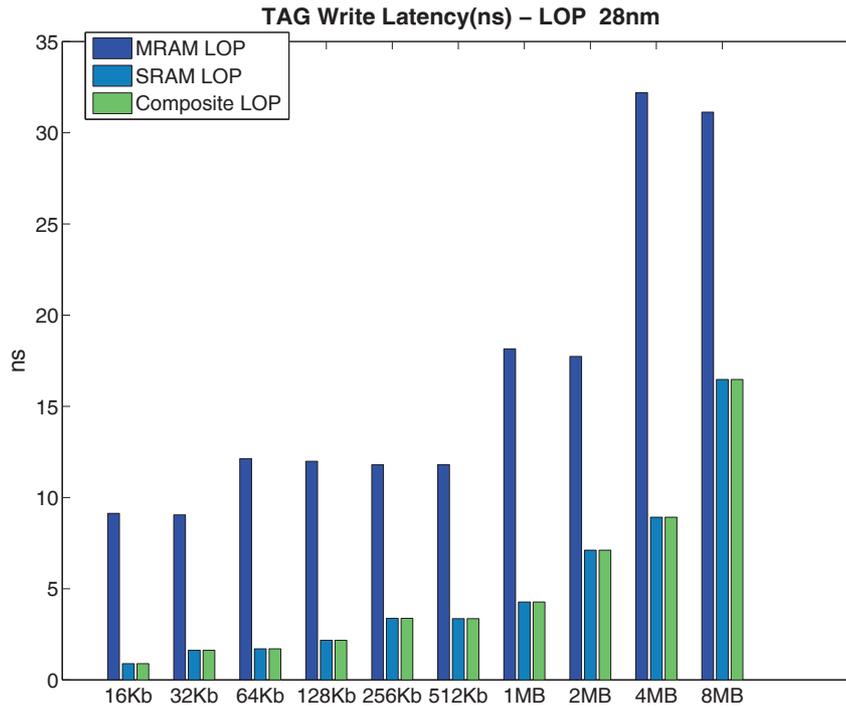


Figure F.56: TAG Write Latency

Table F.53: TAG Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	9.1320	0.8893	0.8893
32KB	9.0482	1.6204	1.6204
64KB	12.1337	1.7068	1.7068
128KB	11.9828	2.1748	2.1748
256KB	11.8007	3.3801	3.3801
512KB	11.8022	3.3582	3.3582
1MB	18.1500	4.2694	4.2694
2MB	17.7366	7.1148	7.1148
4MB	32.1938	8.9197	8.9197
8MB	31.1254	16.4726	16.4726

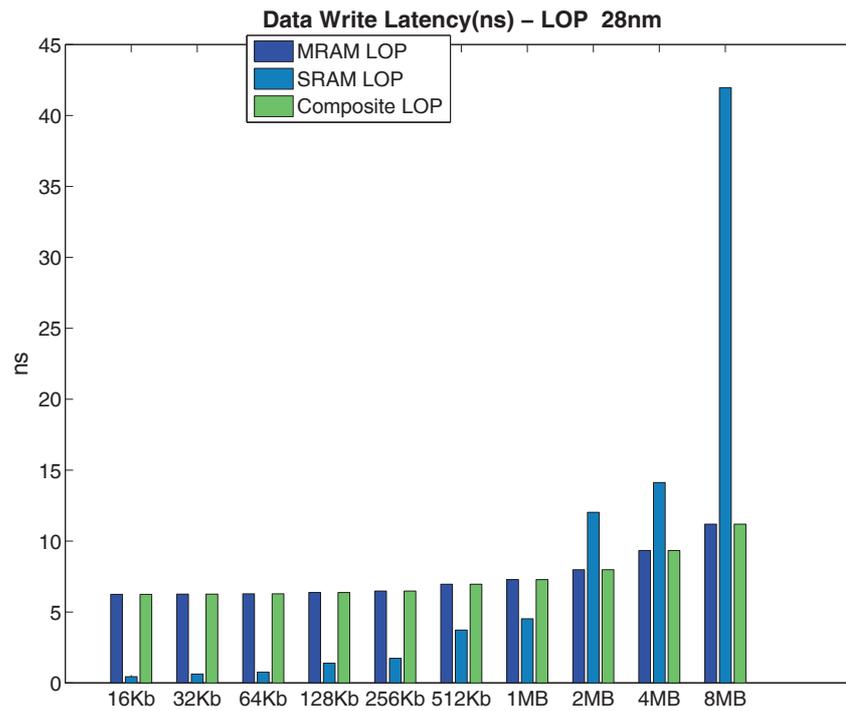


Figure F.57: DATA Write Latency

Table F.54: DATA Write Latency (ns).

	MRAM LOP	SRAM LOP	Composite LOP
16KB	6.2446	0.4389	6.2446
32KB	6.2592	0.6305	6.2592
64KB	6.2846	0.7661	6.2846
128KB	6.3747	1.3957	6.3747
256KB	6.4801	1.7345	6.4801
512KB	6.9570	3.7236	6.9570
1MB	7.2825	4.5230	7.2825
2MB	7.9810	12.0243	7.9810
4MB	9.3365	14.1231	9.3365
8MB	11.1918	41.9588	11.1918

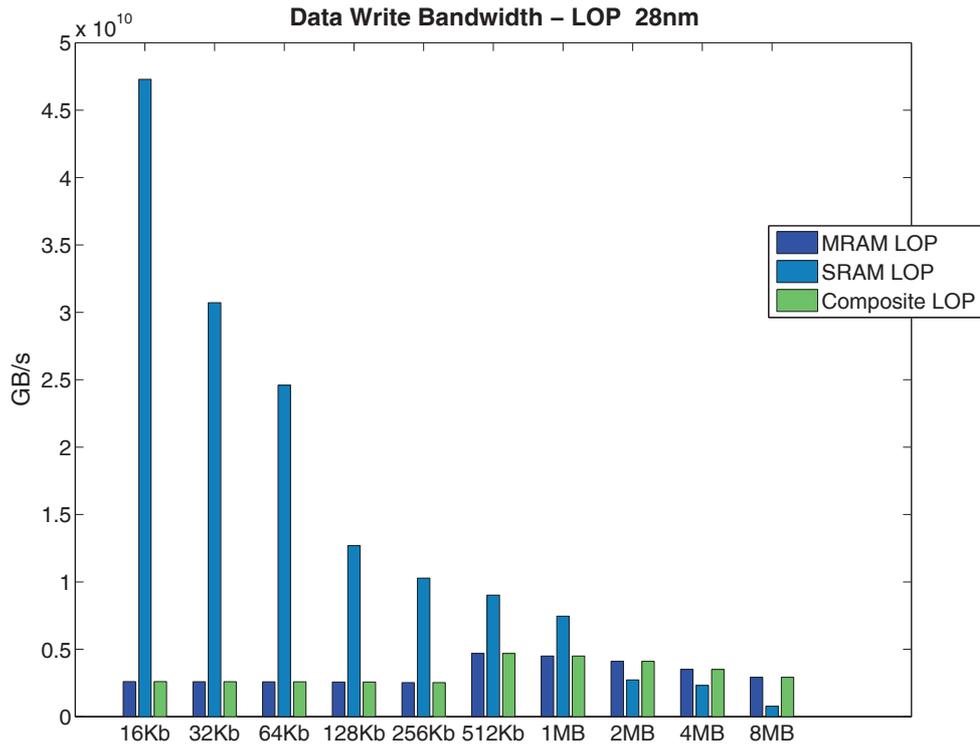


Figure F.58: DATA Write Bandwidth.

Table F.55: DATA Write Bandwidth (B/s)

	MRAM LOP	SRAM LOP	Composite LOP
16KB	2.423e + 00	4.404e + 01	2.423e + 00
32KB	2.419e + 00	2.861e + 01	2.419e + 00
64KB	2.410e + 00	2.290e + 01	2.410e + 00
128KB	2.389e + 00	1.181e + 01	2.389e + 00
256KB	2.352e + 00	9.572e + 00	2.352e + 00
512KB	4.367e + 00	8.399e + 00	4.367e + 00
1MB	4.191e + 00	6.941e + 00	4.191e + 00
2MB	3.823e + 00	2.539e + 00	3.823e + 00
4MB	3.266e + 00	2.161e + 00	3.266e + 00
8MB	2.723e + 00	7.180e - 01	2.723e + 00

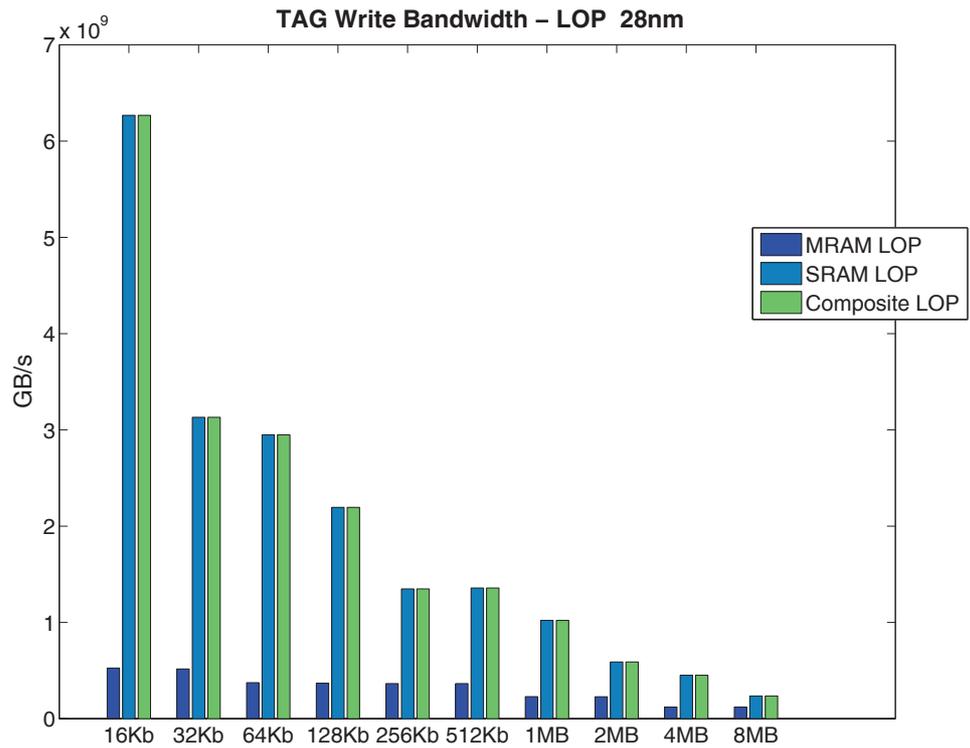


Figure F.59: TAG Write Bandwidth.

Table F.56: TAG Write Bandwidth (GB/s)

	MRAM LOP	SRAM LOP	Composite LOP
16KB	4.895e - 01	5.837e + 00	5.837e + 00
32KB	4.811e - 01	2.916e + 00	2.916e + 00
64KB	3.484e - 01	2.746e + 00	2.746e + 00
128KB	3.440e - 01	2.044e + 00	2.044e + 00
256KB	3.394e - 01	1.255e + 00	1.255e + 00
512KB	3.394e - 01	1.264e + 00	1.264e + 00
1MB	2.135e - 01	9.509e - 01	9.509e - 01
2MB	2.120e - 01	5.479e - 01	5.479e - 01
4MB	1.128e - 01	4.196e - 01	4.196e - 01
8MB	1.129e - 01	2.190e - 01	2.190e - 01

BIBLIOGRAPHY

- Magnetic tunnel junction magnetoresistive read head with sensing layer as flux guide. *google.com*. URL <http://www.google.com/patents/US5898547?dq=ininventor:Slonczewski+1997>.
- Propriétés Magnétiques des ferrites: ferrimagnétisme et astiferrimagnétisme. URL <http://garfield.library.upenn.edu/classics1984/A1984TA22400001.pdf>.
- Status and outlook of emerging nonvolatile memory technologies*, December 2004. doi: 10.1109/IEDM.2004.1419223. URL <http://dx.doi.org/10.1109/IEDM.2004.1419223>.
- Next-generation non-volatile memory*, December 2007. doi: 10.1109/MTDT.2007.4547616. URL <http://dx.doi.org/10.1109/MTDT.2007.4547616>.
- Session 12: Memory technology - resistive memory and magnetic memory*, December 2008. doi: 10.1109/IEDM.2008.4796674. URL <http://dx.doi.org/10.1109/IEDM.2008.4796674>.
- International Technology Roadmap for Semiconductors, 2011. URL <http://www.itrs.net/Links/2011ITRS/Home2011.htm>.
- M H Abu-Rahma, M Anis, and Sei Seung Yoon. Reducing SRAM Power Using Fine-Grained Wordline Pulsewidth Control. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 18(3):356–364, 2010. doi: 10.1109/TVLSI.2009.2012511. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5071192&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- A Aharoni. *Introduction to the Theory of Ferromagnetism*. International Series of Monographs on Physics. Clarendon Press, 2000. ISBN 9780198508090. URL <http://books.google.fr/books?id=Ru-z9b3WcfMC>.
- S J Ahn, Y J Song, C W Jeong, J M Shin, Y Fai, Y N Hwang, S H Lee, K C Ryoo, S Y Lee, J H Park, H Horii, Y H Ha, J H Yi, B J Kuh, G H Koh, G T Jeong, H S Jeong, Kinam Kim, and B I Ryu. Highly manufacturable high density phase change memory of 64Mb and beyond. In *IEDM Technical Digest. IEEE International Electron Devices Meeting, 2004.*, pages 907–910. IEEE. ISBN 0-7803-8684-1. doi: 10.1109/IEDM.2004.1419329. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1419329>.
- Seung-Eon Ahn, Bo Soo Kang, Ki Hwan Kim, Myoung-Jae Lee, Chang Bum Lee, G Stefanovich, Chang Jung Kim, and Youngsoo Park. IEEE Xplore - Stackable All-Oxide-Based Nonvolatile Memory With Al₂O₃ Antifuse and p-CuO/n-InZnO_x Diode. *IEEE Electron Device Letters*, 30(5):550–552, May 2009. doi: 10.1109/LED.2009.2016582. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4808211>.
- H Akinaga and H Shima. Resistive Random Access Memory (ReRAM) Based on Metal Oxides. *Proceedings of the IEEE*, 98(12):2237–2251, November 2010. doi: 10.

- 1109/JPROC.2010.2070830. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5607274.
- J Alvarez, E Barkin, Chai-Chin Chao, B Johnson, M D'Addeo, F Lassandro, G Nicoletta, P Patel, P Reed, D Reid, H Sanchez, J Siegel, M Snyder, S Sullivan, S Taylor, and Minh Vo. Solid-State Circuits Conference 1999. Digest of Technical Papers. ISSCC. 1999 IEEE International 1999 Minh Vo. *Solid-State Circuits Conference, 1999. Digest of Technical Papers. ISSCC. 1999 IEEE International*, pages 96–97, March 2004. doi: 10.1109/ISSCC.1999.759141. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=759141&openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526filter%253DAND%2528NOT%25284283010803%2529%2529%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All%2526queryText%253D%2528%2528sram+cache%2529+AND+%25281+cycle+delay%2529%2529.
- S Amara, R C Sousa, H Bea, C Baraduc, and B Dieny. Barrier Breakdown Mechanisms in MgO-Based Magnetic Tunnel Junctions and Correlation With Low-Frequency Noise. *Magnetics, IEEE Transactions on*, 48(11):4340–4343, 2012. doi: 10.1109/TMAG.2012.2200243. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6332636&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22CoFeB+MgO+CoFeB%22%29.
- Jana Andzane, Nikolay Petkov, Aleksandrs I Livshits, John J Boland, Justin D Holmes, and Donats Erts. Two-terminal nanoelectromechanical devices based on germanium nanowires. *Nano Letters*, 9(5):1824–1829, April 2009. doi: 10.1021/nl8037807. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009NanoL...9.1824A&link_type=EJOURNAL.
- Tomonori Arakawa, Koji Sekiguchi, Shuji Nakamura, Kensaku Chida, Yoshitaka Nishihara, Daichi Chiba, Kensuke Kobayashi, Akio Fukushima, Shinji Yuasa, and Teruo Ono. Sub-Poissonian shot noise in CoFeB/MgO/CoFeB-based magnetic tunneling junctions. *arXiv.org*, August 2011. doi: 10.1063/1.3590921. URL <http://arxiv.org/abs/1108.4831v1>.
- C Argyrides, F Vargas, M Moraes, and D K Pradhan. Embedding Current Monitoring in H-Tree RAM Architecture for Multiple SEU Tolerance and Reliability Improvement. *On-Line Testing Symposium, 2008. IOLTS '08. 14th IEEE International*, pages 155–160, 2008. doi: 10.1109/IOLTS.2008.36. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4567078&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- C Augustine, N Mojumder, Xuanyao Fong, H Choday, Sang Phill Park, and K Roy. STT-MRAMs for future universal memories: Perspective and prospective. In *Microelectronics (MIEL), 2012 28th International Conference on*, pages 349–355, 2012. doi: 10.1109/MIEL.2012.6222872. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6222872&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.

- Todd Austin and Doug Burger. The simplescalar tool set. Technical Report 1342, University of Wisconsin CS Department, Technical Report, 1997.
- K Banerjee and A Mehrotra. A power-optimal repeater insertion methodology for global interconnects in nanometer designs. *Electron Devices, IEEE Transactions on*, 49(11):2001–2007, November 2002. doi: 10.1109/TED.2002.804706. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1097918>.
- Xin Bao, Khusro Sajid, and Elisabeth Moseley. Timing Sign-off using CCS Libraries at QUALCOMM. pages 1–20, February 2006. URL https://www.synopsys.com/news/pubs/sjsnug/sj06/bao_final.pdf.
- D Bergeron. More than Moore. In *Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE*, 2008. doi: 10.1109/CICC.2008.4672003. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4672003>.
- N L Binkert, R G Dreslinski, L R Hsu, K T Lim, A G Saidi, and S K Reinhardt. The M5 Simulator: Modeling Networked Systems. *Micro, IEEE*, 26(4):52–60, 2006. doi: 10.1109/MM.2006.82. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1677503&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22The+M5+Simulator%3A+Modeling+Networked+Systems.%22%29.
- Nathan Binkert, Somayeh Sardashti, Rathijit Sen, Korey Sewell, Muhammad Shoaib, Nilay Vaish, Mark D Hill, David A Wood, Bradford Beckmann, Gabriel Black, Steven K Reinhardt, Ali Saidi, Arkaprava Basu, Joel Hestness, Derek R Hower, and Tushar Krishna. The gem5 simulator. *ACM SIGARCH Computer Architecture News*, 39(2):1–7, August 2011. doi: 10.1145/2024716.2024718. URL <http://portal.acm.org/citation.cfm?id=2024716.2024718&coll=DL&dl=ACM&CFID=219507219&CFTOKEN=22791001>.
- P Blom, R Wolf, J Cillessen, and M Krijn. Ferroelectric Schottky Diode. *Physical Review Letters*, 73(15):2107–2110, October 1994. doi: 10.1103/PhysRevLett.73.2107. URL <http://link.aps.org/doi/10.1103/PhysRevLett.73.2107>.
- B Boschma, D Burns, R Chin, N Fiduccia, C Hu, M Reed, T Rueth, F Schumacher, and V Shen. Solid-State Circuits Conference 1989. Digest of Technical Papers. 36th ISSCC. 1989 IEEE International 1989 Shen. *Solid-State Circuits Conference, 1989. Digest of Technical Papers. 36th ISSCC., 1989 IEEE International*, pages 82–83, April 2004. doi: 10.1109/ISSCC.1989.48191. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=48191&openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526filter%253DAND%2528NOT%25284283010803%2529%2529%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All%2526queryText%253D%2528%2528sram+cache%2529+AND+%25281+cycle+delay%2529%2529.
- El Mostafa Bourim, Sangsoo Park, Xinjun Liu, Kuyyadi P Biju, Hyunsang Hwang, and Alex Ignatiev. Ferroelectric Polarization Effect on Al-Nb Codoped Pb(Zr_{0.52}Ti_{0.48})O₃/Pro.7Ca_{0.3}MnO₃ Heterostructure Resistive Memory. *Electrochemical and Solid-State Letters*, 14(5):H225, 2011. doi: 10.1149/1.3556977. URL <http://esl.ecsd.org/cgi/doi/10.1149/1.3556977>.

- Arne Brataas, Andrew D Kent, and HIDEO OHNO. Nature Materials 2012 Ohno-1. *Nature Publishing Group*, 11(5):372–381, April 2012. doi: 10.1038/nmat3311. URL <http://dx.doi.org/10.1038/nmat3311>.
- Phil Brinkley and Carl Carmichael. SEU Mitigation Design Techniques for the XQR4000XL. page 14, March 2000. URL <http://ieeexplore.ieee.org/iel5/6607/17646/00815306.pdf?arnumber=815306>.
- Nicolas BRUCHON. Evaluation, validation and design of hybrid CMOS - Non Volatile emerging technology cells for dynamically reconfigurable fine grain architecture. pages 1–141, March 2007a.
- Nicolas BRUCHON. Evaluation, validation and design of hybrid CMOS - Non Volatile emerging technology cells for dynamically reconfigurable fine grain architecture. pages 1–141, December 2007b.
- Doug Burger and Todd M Austin. The SimpleScalar tool set, version 2.0. *ACM SIGARCH Computer Architecture News*, 25:13–25, June 1997. doi: <http://doi.acm.org/10.1145/268806.268810>. URL <http://doi.acm.org/10.1145/268806.268810>.
- W Butler, X G Zhang, T Schulthess, and J MacLaren. Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches. *Phys. Rev. B*, 63(5):054416, January 2001. doi: 10.1103/PhysRevB.63.054416. URL <http://link.aps.org/doi/10.1103/PhysRevB.63.054416>.
- R B Buxton. *An Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, 2002. ISBN 9780521581134. URL <http://books.google.fr/books?id=pi3ogRUXxgYC>.
- Yun Cao, Hiroyuki Tomiyama, Takanori Okuma, and Hiroto Yasuura. Data memory design considering effective bitwidth for low-energy embedded systems. In *ISSS '02: Proceedings of the 15th international symposium on System Synthesis*. ACM, October 2002. doi: 10.1145/581199.581245. URL <http://portal.acm.org/citation.cfm?id=581199.581245&coll=DL&dL=ACM&CFID=232708260&CFTOKEN=22541540>.
- Laurent Cario, Cristian Vaju, Benoit Corraze, Vincent Guiot, and Etienne Janod. Electric-Field-Induced Resistive Switching in a Family of Mott Insulators: Towards a New Class of RRAM Memories. *Advanced Materials*, 22(45):5193–5197, 2010. URL <http://onlinelibrary.wiley.com/doi/10.1002/adma.201002521/full>.
- Carl Carmichael. Triple Module Redundancy Design Techniques for Virtex FPGAs. page 37, July 2006. URL http://www.xilinx.com/support/documentation/application_notes/xapp197.pdf.
- Carl Carmichael, Michael Caffrey, and Anthony Salazar. Correcting Single-Event Upsets Through Virtex Partial Configuration. page 12, June 2000. URL http://www.xilinx.com/support/documentation/application_notes/xapp216.pdf.
- Jolanta Celinska, Christopher McWilliams, Carlos Paz de Araujo, and Kan-Hao Xue. Material and process optimization of correlated electron random access memories. *Journal of Applied Physics*, 109(9):091603–091603, 2011. doi: 10.1063/1.3581197. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5770323.

- B G Chae, H T Kim, D H Youn, and K Y Kang. Abrupt metal–insulator transition observed in VO₂ thin films induced by a switching voltage pulse. *Physica B: Condensed Matter*, 2005. URL <http://www.sciencedirect.com/science/article/pii/S0921452605009245>.
- Claude Chappert, Albert Fert, and Frédéric Nguyen Van Dau. The emergence of spin electronics in data storage. *Nature Publishing Group*, 6(11):813–823, November 2007a. doi: 10.1038/nmat2024. URL <http://www.nature.com/doifinder/10.1038/nmat2024>.
- Claude Chappert, Albert Fert, and Frederic Nguyen Van Dau. The emergence of spin eletronics in data storage. *Nature Materials*, 6(11):813 – 823, nov. 2007b. doi: 10.1038/nmat2024.
- S Chellappa, A Dey, and L T Clark. Improved circuits for microchip identification using SRAM mismatch. In *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, pages 1–4, 2011. doi: 10.1109/CICC.2011.6055318. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6055318&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- An Chen. Ionic memories: Status and challenges. *Non-Volatile Memory Technology Symposium, 2008. NVMTS 2008. 9th Annual*, pages 1–5, 2008. doi: 10.1109/NVMT.2008.4731188. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4731188&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22ionic+memories%22%29.
- E Chen, D Apalkov, A Driskill-Smith, A Khvalkovskiy, D Lottis, K Moon, V Nikitin, A Ong, X Tang, S Watts, R Kawakami, M Krounbi, S A Wolf, S J Poon, J W Lu, A W Ghosh, M Stan, W Butler, T Mewes, S Gupta, C K A Mewes, P B Visscher, and R A Lukaszew. Progress and Prospects of Spin Transfer Torque Random Access Memory. *Magnetics, IEEE Transactions on*, 48(11):3025–3030, 2012. doi: 10.1109/TMAG.2012.2198451. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6332660&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- X Chen, S Samavedam, V Narayanan, K Stein, C Hobbs, C Baiocco, W Li, D Jaeger, M Zaleski, H S Yang, N Kim, Y Lee, D Zhang, L Kang, J Chen, H Zhuang, A Sheikh, J Wallner, M Aquilino, J Han, Z Jin, J Li, G Massey, S Kalpat, R Jha, N Moumen, R Mo, S Kirshnan, X Wang, M Chudzik, M Chowdhury, D Nair, C Reddy, Y W Teh, C Kothandaraman, D Coolbaugh, S Pandey, D Tekleab, A Thean, M Sherony, C Lage, J Sudijono, R Lindsay, J H Ku, M Khare, and A Steegen. A cost effective 32nm high-K/metal gate CMOS technology for low power applications with single-metal/gate-first process. In *2008 Symposium on VLSI Technology*, pages 88–89. IEEE, 2008. ISBN 978-1-4244-1802-2. doi: 10.1109/VLSIT.2008.4588573. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4588573>.
- Y S Chen, H Y Lee, P S Chen, P Y Gu, C W Chen, W P Lin, W H Liu, Y Y Hsu, S S Sheu, P C Chiang, W S Chen, F T Chen, C H Lien, and M-J Tsai. Highly scalable hafnium oxide memory with improvements of resistive distribution and

- read disturb immunity. In *2009 IEEE International Electron Devices Meeting (IEDM)*, pages 1–4. IEEE, a. ISBN 978-1-4244-5639-0. doi: 10.1109/IEDM.2009.5424411. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5424411>.
- Yen Huei Chen, G Chan, Shao Yu Chou, Hsien-Yu Pan, Jui-Jen Wu, R Lee, H J Liao, and H Yamauchi. A 0.6 V Dual-Rail Compiler SRAM Design on 45 nm CMOS Technology With Adaptive SRAM Power for Lower VDD_{min} VLSIs. *Solid-State Circuits, IEEE Journal of*, 44(4):1209–1215, 2009. doi: 10.1109/JSSC.2009.2014208. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4804986&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Yi-Chou Chen, C F Chen, C T Chen, J Y Yu, S Wu, S L Lung, R Liu, and Chih-Yuan Lu. An access-transistor-free ($0T/1R$) non-volatile resistance random access memory (RRAM) using a novel threshold switching, self-rectifying chalcogenide device. In *IEEE International Electron Devices Meeting 2003*, pages 37.4.1–37.4.4. IEEE, b. ISBN 0-7803-7872-5. doi: 10.1109/IEDM.2003.1269425. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1269425>.
- Yiran Chen. The Applications of Spintronic Memory in Microprocessors. In *flashmemorysummit.com*. University of Pittsburgh - Swanson School of Engineering Department of Electrical and Computer Engineering, August 2012. URL http://www.flashmemorysummit.com/English/Collaterals/Proceedings/2012/20120823_S301D_Chen.pdf.
- Yiran Chen and Hai Li. Emerging sensing techniques for emerging memories. In *ASPDAC '11: Proceedings of the 16th Asia and South Pacific Design Automation Conference*. IEEE Press, January 2011. URL <http://portal.acm.org/citation.cfm?id=1950815.1950866&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Yong Chen, Douglas A A Ohlberg, Xuema Li, Duncan R Stewart, R Stanley Williams, Jan O Jeppesen, Kent A Nielsen, J F Stoddart, Deirdre L Olynick, and Erik Anderson. Nanoscale molecular-switch devices fabricated by imprint lithography. *Applied Physics Letters*, 82(10):1610–1612, 2003. doi: 10.1063/1.1559439. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4868592.
- Byungjin Cho, Tae-Wook Kim, Sunghoon Song, Yongsung Ji, Minseok Jo, Hyunsang Hwang, Gun-Young Jung, and Takhee Lee. Rewritable switching of one diode–one resistor nonvolatile organic memory devices. *Advanced Materials*, 22(11):1228–1232, 2010. URL <http://onlinelibrary.wiley.com/doi/10.1002/adma.200903203/full>.
- G M Choi, K H Shin, S A Seo, S O Kim, W C Lim, and T D Lee. Substrate Biasing Effect during MgO Deposition in CoFeB/MgO/CoFeB MTJs. *Magnetics, IEEE Transactions on*, 45(6):2371–2373, June 2009. doi: 10.1109/TMAG.2009.2018577. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4957788&queryText%253D%2528%2528mram%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526pageNumber%253D3%2526rowsPerPage%253D50%2526searchField%253DSearch+All.

- Woo Young Choi, Woo Young Choi, T Osabe, and Tsu-Jae King Liu. Nano-Electro-Mechanical Nonvolatile Memory (NEMory) Cell Design and Scaling. *Electron Devices, IEEE Transactions on*, 55(12):3482–3488, November 2008. doi: 10.1109/TED.2008.2006540. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2008ITED...55.3482C&link_type=EJOURNAL.
- Y Choi, K Lee, C H Park, K H Lee, and J W Nam. High current fast switching n-ZnO/p-Si diode - Abstract - Journal of Physics D: Applied Physics - IOPscience. *Journal of Physics D: ...*, 2010. URL <http://iopscience.iop.org/0022-3727/43/34/345101>.
- M Chshiev, I Theodonis, A Kalitsov, N Kioussis, and W H Butler. Voltage Dependence of Spin Transfer Torque In Magnetic Tunnel Junctions. *Magnetics, IEEE Transactions on*, 44(11):2543–2546, November 2008. doi: 10.1109/TMAG.2008.2002605. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4717562>.
- Ki Chul Chun, Hui Zhao, Jonathan D Harms, Tae-Hyoung Kim, Jian-Ping Wang, and Chris H Kim. A Scaling Roadmap and Performance Evaluation of In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory. *Solid-State Circuits, IEEE Journal of*, 48(2):598–610, 2013. doi: 10.1109/JSSC.2012.2224256. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6374706>.
- B D Cullity and C D Graham. *Introduction to Magnetic Materials*. John Wiley & Sons, 2 edition, 2011. ISBN 9781118211496. URL http://books.google.fr/books?id=fh_F0G9KuSgC.
- Cummings, Savchenko, and Ren. Functionalization of flat Si surfaces with inorganic compounds-Towards molecular CMOS hybrid devices. *Coordination Chemistry Reviews*, 255(15-16):16–16, July 2011. doi: 10.1016/j.ccr.2010.12.030. URL <http://www.sciencedirect.com/science/article/pii/S0010854511000154>.
- I B M Dason, V R Kumar, and A A Kirubaraj. Realization of Magnetic RAM using Magnetic Tunneling Junction in atomic level. In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, pages 397–401, 2011. doi: 10.1109/ICECTECH.2011.5941929. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5941929>.
- Vivek De and Shekhar Borkar. *Technology and design challenges for low power and high performance*. ACM, August 1999. ISBN 1-58113-133-X. doi: 10.1145/313817.313908. URL <http://dl.acm.org/citation.cfm?id=313817.313908>.
- G De Sandre, L Bettini, A Pirola, L Marmonier, M Pasotti, M Borghi, P Mattavelli, P Zuliani, L Scotti, Gianfranco Mastracchio, F Bedeschi, R Gastaldi, and R Bez. A 4 Mb LV MOS-Selected Embedded Phase Change Memory in 90 nm Standard CMOS Technology. *Solid-State Circuits, IEEE Journal of*, 46(1):52–63, 2011. doi: 10.1109/JSSC.2010.2084491. URL <http://dx.doi.org/10.1109/JSSC.2010.2084491>.
- A DeHon, S C Goldstein, P J Kuekes, and P Lincoln. Nonphotolithographic nanoscale memory density prospects. *Nanotechnology, IEEE Transactions on*, 4(2):215–228, 2005. doi: 10.1109/TNANO.2004.837849. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1405998.

- Zhitao Diao, Mahendra Pakala, Alex Panchula, Yunfei Ding, Dmytro Apalkov, Lien-Chang Wang, Eugene Chen, and Yiming Huai. Spin-transfer switching in MgO-based magnetic tunnel junctions (invited). *Journal of Applied Physics*, 99(8):08G510, 2006. doi: 10.1063/1.2165169. URL <http://link.aip.org/link/JAPIAU/v99/i8/p08G510/s1&Agg=doi>.
- S Dietrich, M Angerbauer, M Ivanov, D Gogl, H Hoenigschmid, M Kund, C Liaw, M Markert, R Symanczyk, L Altimime, S Bournat, and G Mueller. A Nonvolatile 2-Mbit CBRAM Memory Core Featuring Advanced Read and Program Control. *Solid-State Circuits, IEEE Journal of*, 42(4):839–845, 2007. doi: 10.1109/JSSC.2007.892207. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4140579&queryText%253D%2528%2528Document+Title%253AA+Nonvolatile+2-Mbit+CBRAM+Memory+Core+Featuring+Advanced+Read+and+Program+Control%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Aaron Dingler, Steve Kurtz, Michael Niemier, Xiaobo Sharon Hu, Gyorgy Csaba, Joseph Nahas, Wolfgang Porod, Gary Bernstein, Peng Li, and Vjiay Karthik Sankar. Making non-volatile nanomagnet logic non-volatile. In *DAC '12: Proceedings of the 49th Annual Design Automation Conference*. ACM Request Permissions, June 2012. doi: 10.1145/2228360.2228445. URL <http://portal.acm.org/citation.cfm?id=2228360.2228445&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- M Q Do, M Drazdziulis, P Larsson-Edefors, and L Bengtsson. Parameterizable architecture-level SRAM power model using circuit-simulation backend for leakage calibration. *Quality Electronic Design, 2006. ISQED '06. 7th International Symposium on*, page 7, 2006. doi: 10.1109/ISQED.2006.97. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1613197&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- M Q Do, P Larsson-Edefors, and M Drazdziulis. High-Accuracy Architecture-Level Power Estimation for Partitioned SRAM Arrays in a 65-nm CMOS BPTM Process. In *Digital System Design Architectures, Methods and Tools, 2007. DSD 2007. 10th Euromicro Conference on*, pages 249–256, 2007. doi: 10.1109/DSD.2007.4341476. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4341476&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- DoD-USA. RELIABILITY PREDICTION OF ELECTRONIC EQUIPMENT. pages 1–205, February 1995.
- DoD-USA. Test Method Standard Microcircuits. pages 1–641, December 1996. URL http://www.google.fr/search?client=safari&rls=en-us&q=Test+Method+Standard+Microcircuits&ie=UTF-8&oe=UTF-8&redir_esc=&ei=nH2cTJq2LsnT4waP67TSDQ.
- DoD-USA. ELECTRONIC RELIABILITY DESIGN HANDBOOK. pages 1–1042, April 1999.

- Xiangyu Dong, Xiaoxia Wu, Guangyu Sun, Yuan Xie, H Li, and Yiran Chen. Circuit and microarchitecture evaluation of 3D stacking magnetic RAM (MRAM) as a universal memory replacement. *Design Automation Conference, 2008. DAC 2008. 45th ACM/IEEE*, pages 554–559, 2008. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4555878&queryText%253D%2528%2528Document+Title%253AMRAM%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Xiangyu Dong, Norman P Jouppi, and Yuan Xie. PCRAMsim: System-level performance, energy, and area modeling for Phase-Change RAM. *ieeexplore.ieee.org*, pages 269–275, 2009. ISSN 1092-3152. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5361282&navigation=1>.
- Xiangyu Dong, Cong Xu, Yuan Xie, and N P Jouppi. NVSim: A Circuit-Level Performance, Energy, and Area Model for Emerging Nonvolatile Memory. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 31(7):994–1007, July 2012a. doi: 10.1109/TCAD.2012.2185930. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6218223>.
- Xiangyu Dong, Cong Xu, Yuan Xie, and Norman P Jouppi. Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 31(7):994–1007, 2012b. doi: 10.1109/TCAD.2012.2185930. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6218223.
- E Donkoh, A Lowery, and E Shriver. A hybrid and adaptive model for predicting register file and SRAM power using a reference design. In *Design Automation Conference (DAC), 2012 49th ACM/EDAC/IEEE*, pages 62–67, 2012. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6241491&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- R Dorrance, J G Alzate, S S Cherepov, P Upadhyaya, I N Krivorotov, J A Katine, J Langer, K L Wang, P K Amiri, and D Markovic. Diode-MTJ Crossbar Memory Cell Using Voltage-Induced Unipolar Switching for High-Density MRAM. *Electron Device Letters, IEEE*, 34(6):753–755, 2013. doi: 10.1109/LED.2013.2255096. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6513288&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- A Driskill-Smith, D Apalkov, V Nikitin, X Tang, S Watts, D Lottis, K Moon, A Khvalkovskiy, R Kawakami, X Luo, A Ong, E Chen, and M Krounbi. Latest Advances and Roadmap for In-Plane and Perpendicular STT-RAM. In *Memory Workshop (IMW), 2011 3rd IEEE International*, pages 1–3, 2011. doi: 10.1109/IMW.2011.5873205. URL <http://dx.doi.org/10.1109/IMW.2011.5873205>.
- C Dufour, P Garnier, T Carriere, J Beaucour, R Ecoffet, and M Labrunee. Heavy ion induced single hard errors on submicronic memories [for space application]. *Nuclear Science, IEEE Transactions on*, 39(6):1693–1697, 1992. doi: 10.1109/23.211355. URL <http://dx.doi.org/10.1109/23.211355>.

- M Durlam, D Addie, J Akerman, B Butcher, P Brown, J Chan, M DeHerrera, B N Engel, B Feil, G Grynkewich, J Janesky, M Johnson, K Kyler, J Molla, J Martin, K Nagel, J Nahas, J Ren, N D Rizzo, T Rodriguez, L Savtchenko, J Salter, J M Slaughter, K Smith, J J Sun, M Lien, K Papworth, P Shah, W Qin, R Williams, L Wise, and S Tehrani. A 0.18 μm 4 Mbit toggling MRAM. In *Integrated Circuit Design and Technology, 2004. ICICDT '04. International Conference on*, pages 27–30, 2004. doi: 10.1109/ICICDT.2004.1309899.
- B F Dutton and C E Stroud. Built-In Self-Test of Embedded SEU Detection Cores in Virtex-4 and Virtex-5 FPGAs. *ESA*, pages 149–155, 2009. URL [message: %3CAANLkTimZopqGtjB59rGdTTPXhYo1gDSxxEAXEY2VsX3K@gmail.com%3E](mailto:3CAANLkTimZopqGtjB59rGdTTPXhYo1gDSxxEAXEY2VsX3K@gmail.com).
- Christopher Evans. Conversation: Jay W. Forrester. *IEEE Annals of the History of Computing*, 5(3):297–301. doi: 10.1109/MAHC.1983.10081. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4640715>.
- Everspin Technologies Inc. MRAM Technical Guide. Technical report, Everspin Technologies Inc., December 2010. URL http://www.everspin.com/PDF/TSP-12545_MRAM_Bro_Upd_v3db.pdf.
- Wang Fang, Jia Song, and Ji Lijiu. Power Compleity Analysis of Adiabatic SRAM. In *ASIC, 2005. ASICON 2005. 6th International Conference On*, pages 334–337, 2005. doi: 10.1109/ICASIC.2005.1611318. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1611318&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Albert Fert and Peter Grünberg. Gmr: An attractive resistance. *Europhysics News*, 28: 114–118, September 1997.
- M Fitsilis. DC MetaData for: Scaling of the Ferroelectric Field Effect Transistor and Programming Concepts for Non-volatile Memory Applications. 2005. URL <http://sylvester.bth.rwth-aachen.de/dissertationen/2005/044/>.
- Michael Fitsilis, Yacoub Mustafa, and Rainer Waser. Scaling the Ferroelectric Field Effect Transistor. *Integrated Ferroelectrics*, 70(1):29–44, April 2005. doi: 10.1080/10584580590926657. URL <http://www.informaworld.com/openurl?genre=article&doi=10.1080/10584580590926657&magic=crossref|D404A21C5BB053405B1A640AFFD44AE3>.
- X Fong, Y Kim, S H Choday, and K Roy. Failure Mitigation Techniques for 1T-1MTJ Spin-Transfer Torque MRAM Bit-cells. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, (99):1, 2013. doi: 10.1109/TVLSI.2013.2239671. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6473918&matchBoolean%3Dtrue%26pageNumber%3D3%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM%29.
- Xuanyao Fong, S K Gupta, N N Mojumder, S H Choday, C Augustine, and K Roy. KNACK: A hybrid spin-charge mixed-mode simulator for evaluating different genres of spin-transfer torque MRAM bit-cells. In *Simulation of Semiconductor Processes and Devices (SISPAD), 2011 International Conference on*, pages 51–54, 2011. doi: 10.1109/SISPAD.2011.6035047. URL <http://dx.doi.org/10.1109/SISPAD.2011.6035047>.

- The International Technology Roadmap for Semiconductors. Emerging research devices. Technical report, The International Technology Roadmap for Semiconductors, 2011.
- Jay W Forrester. Digital Information Storage in Three Dimensions Using Magnetic Cores. *Journal of Applied Physics*, 22(1):44–48, January 1951. doi: 10.1063/1.1699817. URL <http://link.aip.org/link/JAPIAU/v22/i1/p44/s1&Agg=doi>.
- Wangyang Fu, Zhi Xu, Xuedong Bai, Changzhi Gu, and Enge Wang. Intrinsic Memory Function of Carbon Nanotube-based Ferroelectric Field-Effect Transistor. *Nano Letters*, 9(3):921–925, March 2009. doi: 10.1021/nl801656w. URL <http://pubs.acs.org/doi/abs/10.1021/nl801656w>.
- A Fukushima, T Seki, K Yakushiji, H Kubota, S Yuasa, and K Ando. Statistical Variance in Switching Probability of Spin-Torque Switching in MgO-MTJ. *Magnetics, IEEE Transactions on*, 48(11):4344–4346, 2012. doi: 10.1109/TMAG.2012.2202100. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6332710&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22CoFeB+MgO+CoFeB%22%29.
- Aeroflex Gaisler. LEON3 Multiprocessing CPU Core. Technical report, February 2010. URL http://www.gaisler.com/doc/leon3_product_sheet.pdf.
- S Gangwal, S Mukhopadhyay, and K Roy. Optimization of Surface Orientation for High-Performance, Low-Power and Robust FinFET SRAM. In *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE*, pages 433–436, 2006. doi: 10.1109/CICC.2006.321009. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4114996&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29.
- M A Garcia-Ramirez, Yoshishige Tsuchiya, and Hiroshi Mizuta. Hybrid circuit analysis of a suspended gate silicon nanodot memory (SGSNM) cell. *MICROELECTRONIC ENGINEERING*, 87(5-8):1284–1286, April 2010. doi: 10.1016/j.mee.2009.10.019. URL <http://www.sciencedirect.com/science/article/pii/S016793170900611X>.
- A Gerber, M Fitsilis, R Waser, Timothy J Reece, E Rije, Stephen Ducharme, and H Kohlstedt. Ferroelectric field effect transistors using very thin ferroelectric polyvinylidene fluoride copolymer films as gate dielectrics. *Journal of Applied Physics*, 107(12):124119–124119, 2010. doi: 10.1063/1.3437638. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5498047.
- Kim GH, Kim KM, Seok JY, Lee HJ, Cho DY, Han JH, and Hwang CS. A theoretical model for Schottky diodes for excluding the sneak current in cross bar array resistive memory. *Nanotechnology*, August 2010. doi: 10.1088/0957-4484/21/38/385202. URL <http://dx.doi.org/10.1088/0957-4484/21/38/385202>.
- Kourosh Gharachorloo, Anoop Gupta, and John Hennessy. Performance evaluation of memory consistency models for shared-memory multiprocessors. *ACM SIGARCH Computer Architecture News*, 19(2):245–257, January 4. doi: 10.1145/106972.106997. URL <http://dl.acm.org/citation.cfm?id=106973.106997>.

- K Gopalakrishnan, R S Shenoy, C T Rettner, K Virwani, D S Bethune, R M Shelby, G W Burr, A Kellock, R S King, K Nguyen, A N Bowers, M Jurich, B Jackson, A M Friz, T Topuria, P M Rice, and B N Kurdi. Highly-scalable novel access device based on Mixed Ionic Electronic conduction (MIEC) materials for high density phase change memory (PCM) arrays. In *VLSI Technology (VLSIT), 2010 Symposium on*, pages 205–206, 2010. ISBN 1424454514. doi: 10.1109/VLSIT.2010.5556229. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5556229.
- Jonathan E Green, Jang Wook Choi, Akram Boukai, Yuri Bunimovich, Ezekiel Johnston-Halperin, Erica DeIonno, Yi Luo, Bonnie A Sheriff, Ke Xu, Young Shik Shin, Hsian-Rong Tseng, J Fraser Stoddart, and James R Heath. A 160-kilobit molecular electronic memory patterned at 10(11) bits per square centimetre. *Nature*, 445(7126):414–417, January 2007. doi: 10.1038/nature05462. URL <http://www.nature.com/nature/journal/v445/n7126/abs/nature05462.html>.
- L M Grupp, A M Caulfield, J Coburn, S Swanson, E Yaakobi, P H Siegel, and J K Wolf. Characterizing flash memory: Anomalies, observations, and applications. *ieeexplore.ieee.org*, pages 24–33, 2009. ISSN 1072-4451. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=5375312&navigation=1>.
- Y Guillemenet, L Torres, G Sassatelli, and N Bruchon. On the Use of Magnetic RAMs in Field-Programmable Gate Arrays. *International Journal of Reconfigurable Computing*, 2008:1–10, August 2008a. doi: 10.1155/2008/723950. URL <http://www.hindawi.com/journals/ijrc/2008/723950/>.
- Y Guillemenet, L Torres, G Sassatelli, N Bruchon, and I Hassoune. A non-volatile run-time FPGA using thermally assisted switching MRAMS. *Field Programmable Logic and Applications, 2008. FPL 2008. International Conference on DOI - 10.1109/FPL.2008.4629974*, pages 421–426, 2008b. doi: 10.1109/FPL.2008.4629974. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4629974>.
- Y Guillemenet, S Z Ahmed, L Torres, A Martheley, J Eydoux, J B Cuelle, L Rouge, and G Sassatelli. MRAM Based eFPGAs: Programming and Silicon Flows, Exploration Environments, MRAM Current State in Industry and Its Unique Potentials for FPGAs. *Reconfigurable Computing and FPGAs, 2009. ReConFig '09. International Conference on DOI - 10.1109/ReConFig.2009.25*, pages 18–23, 2009. URL 10.1109/ReConFig.2009.25.
- Y Guillemenet, L Torres, and G Sassatelli. Non-volatile run-time field-programmable gate arrays structures using thermally assisted switching magnetic random access memories. *Computers & Digital Techniques, IET*, 4(3):211–226, 2010. URL 10.1049/iet-cdt.2009.0019.
- Sumeet Kumar Gupta, Sang Phill Park, Niladri Narayan Mojumder, and Kaushik Roy. Layout-aware optimization of stt mrams. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2012*, pages 1455–1458, 2012. doi: 10.1109/DATE.2012.6176595. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6176595&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.

- A Gutierrez, R Dreslinski, T Wensch, T Mudge, A Saidi, C Emmons, and N Paver. Full-system analysis and characterization of interactive smartphone applications. *Workload Characterization (IISWC), 2011 IEEE International Symposium on*, pages 81–90, 2011. doi: 10.1109/IISWC.2011.6114205. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=6114205&openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526filter%253DAND%2528NOT%25284283010803%2529%2529%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All%2526queryText%253D%2528%2528Full-System+Analysis+and+Characterization+of+Interactive+Smartphone+Applications%2529%2529.
- Guillermo Guzman, Fabien Beteille, Roger Morineau, and Jacques Livage. Electrical switching in VO₂ sol-gel films. *Journal of Materials Chemistry*, 6(3):505, 1996. doi: 10.1039/jm9960600505. URL <http://xlink.rsc.org/?DOI=jm9960600505>.
- Sieu D. Ha, Gulgun H. Aydogdu, and Shriram Ramanathan. Metal-insulator transition and electrically driven memristive characteristics of smnio₃ thin films. *Applied Physics Letters*, 98(1):012105, 2011. doi: 10.1063/1.3536486. URL <http://link.aip.org/link/?APL/98/012105/1>.
- R Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):14, January 1950. URL <http://www.lee.eng.uerj.br/~gil/redesII/hamming.pdf>.
- Jin-Woo Han, Jae-Hyuk Ahn, Min-Wu Kim, Jun-Bo Yoon, and Yang-Kyu Choi. Monolithic integration of NEMS-CMOS with a Fin Flip-flop Actuated Channel Transistor (FinFACT). In *2009 IEEE International Electron Devices Meeting (IEDM)*, pages 1–4. IEEE, 2009. ISBN 978-1-4244-5639-0. doi: 10.1109/IEDM.2009.5424264. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5424264>.
- Jin-Woo Han, Jae-Hyuk Ahn, and Yang-Kyu Choi. FinFACT-Fin Flip-Flop Actuated Channel Transistor. *IEEE Electron Device Letters*, 31(7):764–766, 2010. doi: 10.1109/LED.2010.2048093. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5491086>.
- T.P. Haraszti. *CMOS Memory Circuits*. Springer, 2000. ISBN 9780792379508. URL http://books.google.fr/books?id=_6cY6oNKrOAC.
- Kristjan Haule and Gabriel Kotliar. Strongly correlated superconductivity: A plaquette dynamical mean-field theory study. *Phys. Rev. B*, 76(10):104509, September 2007. doi: 10.1103/PhysRevB.76.104509. URL <http://link.aps.org/doi/10.1103/PhysRevB.76.104509>.
- O G Heinonen, S W Stokes, and J Y Yi. Perpendicular Spin Torque in Magnetic Tunnel Junctions. *Physical Review Letters*, 105(6):066602, August 2010. doi: 10.1103/PhysRevLett.105.066602. URL <http://link.aps.org/doi/10.1103/PhysRevLett.105.066602>.
- K HELD, M ULMKE, and D VOLLHARDT. CORRELATED-ELECTRON THEORY OF METAMAGNETISM IN STRONGLY ANISOTROPIC ANTIFERROMAGNETS. *Modern Physics Letters B*, 10(06):203–210, March 1996. doi: 10.1142/S0217984996000249. URL <http://www.worldscientific.com/doi/abs/10.1142/S0217984996000249>.

- John Hennessy and David Patterson. *Computer Architecture. A quantitative Approach*. Fourth edition, June 2006. URL http://www.scribd.com/doc/7279033/Patterson-Hennessy-Computer-Architecture?secret_password=&autodown=pdf.
- John L Hennessy and David A Patterson. *Computer architecture: a quantitative approach*, volume 1. Elsevier - Morgan Kaufmann - Denise E. M. Penrose, 500 Sansome Street, Suite 400, San Francisco, CA 94111, 4 edition, January 2007a. ISBN 978-0-12-370490-0. URL <http://books.google.com/books?id=57UIPoLt3tkC&printsec=frontcover>.
- John L Hennessy and David A Patterson. *Computer architecture: a quantitative approach*. Morgan Kaufmann, January 2007b. URL <http://books.google.com/books?id=57UIPoLt3tkC&printsec=frontcover>.
- D Hentrich, E Oruklu, and J Saniie. Performance evaluation of SRAM cells in 22nm predictive CMOS technology. In *Electro/Information Technology, 2009. eit '09. IEEE International Conference on*, pages 470–475, 2009. doi: 10.1109/EIT.2009.5189662. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5189662&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- Paul Heremans, Gerwin H Gelinck, Robert Müller, Kang-Jun Baeg, Dong-Yu Kim, and Yong-Young Noh. Polymer and Organic Nonvolatile Memory Devices †. *Chemistry of Materials*, 23(3):341–358, February 2011. doi: 10.1021/cm102006v. URL <http://pubs.acs.org/doi/abs/10.1021/cm102006v>.
- ChiaHua Ho, Cho-Lun Hsu, Chun-Chi Chen, Jan-Tsai Liu, Cheng-San Wu, Chien-Chao Huang, Chenming Hu, and Fu-Liang Yang. 9nm half-pitch functional resistive memory cell with $1A$ programming current using thermally oxidized sub-stoichiometric W₂O₃ film. In *2010 IEEE International Electron Devices Meeting (IEDM)*, pages 19.1.1–19.1.4. IEEE. ISBN 978-1-4424-7418-5. doi: 10.1109/IEDM.2010.5703389. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5703389>.
- R Ho, K W Mai, and M A Horowitz. The future of wires. *Proceedings of the IEEE*, 89(4):490–504, May 2001. doi: 10.1109/5.920580. URL <http://ieeexplore.ieee.org/ielx5/5/19892/00920580.pdf?tp=&arnumber=920580&isnumber=19892>.
- R Ho, Ken Mai, and M Horowitz. Managing wire scaling: a circuit perspective. In IEEE, editor, *Interconnect Technology Conference, 2003. Proceedings of the IEEE 2003 International*, pages 177–179. IEEE, August 2004. doi: 10.1109/IITC.2003.1219747. URL <http://dx.doi.org/10.1109/IITC.2003.1219747>.
- Wolfgang HöFLICH. Using the XC4000 Readback Capability. November 1999. URL <http://www.xilinx.com/xapp/xapp015.pdf>.
- Randolph Hornreich and Richard M. UPPER MAGNETIC FILM. URL <http://www.google.com/patents/US3623038?dq=ininventor:Slonczewski+1997>.
- K. Hoya, D. Takashima, S. Shiratake, R. Ogiwara, T. Miyakawa, H. Shiga, S.M. Doumae, S. Ohtsuki, Y. Kumura, S. Shuto, T. Ozaki, K. Yamakawa, I. Kunishima, A. Nitayama, and S. Fujii. A 64Mb Chain FeRAM with Quad-BL Architecture and 200MB/s Burst

- Mode. In *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International*, pages 459–466, feb. 2006. doi: 10.1109/ISSCC.2006.1696078.
- Zhigang Hu, Philo Juang, Phil Diodato, Stefanos Kaxiras, Kevin Skadron, Margaret Martonosi, and Douglas W Clark. Managing leakage for transient data: decay and quasi-static 4T memory cells. In *ISLPED '02: Proceedings of the 2002 international symposium on Low power electronics and design*. ACM Request Permissions, August 2002. doi: 10.1145/566408.566423. URL <http://portal.acm.org/citation.cfm?id=566408.566423&coll=DL&dl=ACM&CFID=232708260&CFTOKEN=22541540>.
- Y Huai, Z Diao, A Panchula, and E Y Chen. Current-switched spin-transfer magnetic devices with reduced spin-transfer ... - Google Patents. US Patent Office, 2008. URL <http://www.google.com/patents?hl=en&lr=&vid=USPAT7430135&id=cP0tAAAAEBAJ&oi=fnd&dq=Spin+torque+tunnel+current+spin+polarization+and+magnetoresistance+in+MgO+magnetic+tunnel+junctions&printsec=abstract>.
- Yiming Huai. Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects. *AAPPS Bulletin*, 18(6):33–40, 2008. URL <http://www.cospa.ntu.edu.tw/aappsbulletin/data/18-6/33spin.pdf>.
- Yiming Huai, Frank Albert, Paul Nguyen, Mahendra Pakala, and Thierry Valet. Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions. *arXiv.org*, April 2005a. doi: 10.1063/1.1707228. URL <http://arxiv.org/abs/cond-mat/0504486v1>.
- Yiming Huai, M Pakala, Zhitao Diao, and Yunfei Ding. Spin-transfer switching current distribution and reduction in magnetic tunneling junction-based structures. *Magnetics, IEEE Transactions on*, 41(10):2621–2626, October 2005b. doi: 10.1109/TMAG.2005.855346. URL <http://dx.doi.org/10.1109/TMAG.2005.855346>.
- Yiming Huai, Yuchen Zhou, I. Tudosa, R. Malmhall, R. Ranjan, and Jing Zhang. Progress and outlook for stt-mram. In *Computer-Aided Design (ICCAD), 2011 IEEE/ACM International Conference on*, pages 235–235, 2011. doi: 10.1109/ICCAD.2011.6105332.
- J Hubbard. Electron Correlations in Narrow Energy Bands. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 276(1365):238–257, November 1963. doi: 10.1098/rspa.1963.0204. URL <http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1963.0204>.
- N Huby, G Tallarida, M Kutrzeba, and S Ferrari. New selector based on zinc oxide grown by low temperature atomic layer deposition for vertically stacked non-volatile memory devices. *Microelectronic ...*, 2008. URL <http://www.sciencedirect.com/science/article/pii/S0167931708003614>.
- S Huda and A Sheikholeslami. A Novel STT-MRAM Cell With Disturbance-Free Read Operation. *Circuits and Systems I: Regular Papers, IEEE Transactions on*, 60(6):1534–1547, 2013. doi: 10.1109/TCSI.2012.2220458. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6472264&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.

- Hybrid Memory Products Ltd. SRAM Module Mean Time Between Failure Analysis (MTBF). pages 1–11, February 1999.
- Y Iba, K Tsunoda, Y M Lee, C Yoshida, H Noshiro, A Takahashi, Y Yamazaki, M Nakabayashi, A Hatada, M Aoki, and T Sugii. Strain-engineering for high-performance STT-MRAM. *VLSI Technology (VLSIT), 2011 Symposium on*, pages 212–213, 2011. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5984637&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- Daniele Ielmini, Rainer Bruchhaus, and Rainer Waser. Thermochemical resistive switching: materials, mechanisms, and scaling projections. *Phase Transitions*, 84(7):570–602, July 2011. doi: 10.1080/01411594.2011.561478. URL <http://www.tandfonline.com/doi/abs/10.1080/01411594.2011.561478>.
- Hisakazu Iizuka and Fujio Masuoka. Semiconductor memory device and method for manufacturing the same. *google.com*. URL <http://www.google.com/patents/US4531203>.
- S Ikeda, K Miura, H Yamamoto, K Mizunuma, H D Gan, M Endo, S Kanai, J Hayakawa, F Matsukura, and H Ohno. A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction. *Nature Publishing Group*, 9(9):721–724, July 2010. doi: 10.1038/nmat2804. URL <http://www.nature.com/doifinder/10.1038/nmat2804>.
- SHOJI IKEDA, Jun Hayakawa, Young Min Lee, FUMIHIRO MATSUKURA, Yuzo Ohno, Takahiro Hanyu, and HIDEO OHNO. Magnetic Tunnel Junctions for Spintronic Memories and Beyond. *Electron Devices, IEEE Transactions on*, 54(5):991–1002. doi: 10.1109/TED.2007.894617. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4160113>.
- SHOJI IKEDA, HIDEO SATO, MICHIIHIKO YAMANOUCHI, HUADONG GAN, KATSUYA MIURA, KOTARO MIZUNUMA, SHUN KANAI, SHUNSUKE FUKAMI, FUMIHIRO MATSUKURA, NAOKI KASAI, and HIDEO OHNO. RECENT PROGRESS OF PERPENDICULAR ANISOTROPY MAGNETIC TUNNEL JUNCTIONS FOR NONVOLATILE VLSI. *SPIN*, 02(03):1240003, September 2012. doi: 10.1142/S2010324712400036. URL <http://www.worldscientific.com/doi/abs/10.1142/S2010324712400036>.
- Everspin Technologies Inc. Mr2a08am - 512k x 8 mram memory. Datasheet MR2A08AM Rev. 0,2/2011, Everspin Technologies Inc., 2011.
- K. Iniewski. *CMOS Processors and Memories*. Analog circuits and signal processing series. Springer, 2010. ISBN 9789048192168. URL <http://books.google.fr/books?id=2E0r6BRo2Vkc>.
- International Business Machines Corporation. Magnetic tunnel junctions with controlled magnetic response. US Patent Office, March 1996. URL http://www.google.com/patents/about?id=ELQaAAAAEBAJ&dq=ininventor:Slonczewski+1997&as_drrb_ap=q&as_drrb_is=q&num=10&ie=ISO-8859-1.

- S Ishibashi, K Ando, T Seki, T Nozaki, H Kubota, S Yakata, H Maehara, A Fukushima, S Yuasa, and Y Suzuki. High Spin-Torque Diode Sensitivity in CoFeB/MgO/CoFeB Magnetic Tunnel Junctions Under DC Bias Currents. *Magnetics, IEEE Transactions on*, 47(10):3373–3376, 2011. doi: 10.1109/TMAG.2011.2159830. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6028146&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22CoFeB+MgO+CoFeB%22%29.
- K Ishida, T Yasufuku, S Miyamoto, H Nakai, M Takamiya, T Sakurai, and K Takeuchi. A 1.8V 30nJ adaptive program-voltage (20V) generator for 3D-integrated NAND flash SSD. In *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 238–239–239a, 2009. doi: 10.1109/ISSCC.2009.4977396. URL <http://dx.doi.org/10.1109/ISSCC.2009.4977396>.
- Hiroshi Ishiwara. Current status of ferroelectric-gate si transistors and challenge to ferroelectric-gate {CNT} transistors. *Current Applied Physics*, 9(1, Supplement):S2 – S6, 2009a. ISSN 1567-1739. doi: <http://dx.doi.org/10.1016/j.cap.2008.02.013>. URL <http://www.sciencedirect.com/science/article/pii/S1567173908001740>. <ce:title>Nano Korea 2007 Symposium</ce:title>.
- Hiroshi Ishiwara. Current status of ferroelectric-gate Si transistors and challenge to ferroelectric-gate CNT transistors. *Current Applied Physics*, 9(1):S2–S6, January 2009b. doi: 10.1016/j.cap.2008.02.013. URL <http://linkinghub.elsevier.com/retrieve/pii/S1567173908001740>.
- ITRS. <http://public.itrs.net/links/2005itrs/home2005.htm>, a.
- ITRS. <http://public.itrs.net/links/2012itrs/home2012.htm>, b.
- ITRS. Emerging Research Devices. pages 1–83, February 2012a. URL <http://www.itrs.net/Links/2011ITRS/2011Chapters/2011ERD.pdf>.
- ITRS. EmergingResearchMaterials. Technical report, The International Technology Roadmap for Semiconductors, December 2012b.
- Hiroshi Iwai. Technology Scaling and Roadmap. In *iwailab.ep.titech.ac.jp*. Tokyo Institute of Technology, December 2008. URL <http://www.iwailab.ep.titech.ac.jp/pdf/iwaironbun/0812iedm.pdf>.
- Amin Jadidi, Mohammad Arjomand, and Hamid Sarbazi-Azad. High-endurance and performance-efficient design of hybrid cache architectures through adaptive line replacement. In *ISLPED '11: Proceedings of the 17th IEEE/ACM international symposium on Low-power electronics and design*. IEEE Press, August 2011. URL <http://portal.acm.org/citation.cfm?id=2016802.2016836&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- R Jammy. Life beyond Si: More Moore or More than Moore? *Integrated Reliability Workshop Final Report (IRW), 2010 IEEE International*, 2010. doi: 10.1109/IIRW.2010.5706469. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5706469&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.

- C H Jan, M Agostinelli, M Buehler, Z P Chen, S J Choi, G Curello, H Deshpande, S Ganavaram, W Hafez, U Jalan, M Kang, P Kolar, K Komeyli, B Landau, A Lake, N Lazo, S H Lee, T Leo, J Lin, N Lindert, S Ma, L McGill, C Meining, A Paliwal, J Park, K Phoa, I Post, N Pradhan, M Prince, A Rahman, J Rizk, L Rockford, G Sacks, A Schmitz, H Tashiro, C Tsai, P Vandervoorn, J Xu, L Yang, J Y Yeh, J Yip, K Zhang, Y Zhang, and P Bai. A 32nm SoC platform technology with 2nd generation high-k/metal gate transistors optimized for ultra low power, high performance, and high density product applications. In *2009 IEEE International Electron Devices Meeting (IEDM)*, pages 1–4. IEEE, 2009. ISBN 978-1-4244-5639-0. doi: 10.1109/IEDM.2009.5424258. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5424258>.
- H Jarollahi and R F Hobson. Power and area efficient 5T-SRAM with improved performance for low-power SoC in 65nm CMOS. *Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on*, pages 121–124, 2010. doi: 10.1109/MWSCAS.2010.5548577. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5548577&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- JC-42.3. Double Data Rate (DDR) SDRAM Standard. 2008. URL <http://www.jedec.org/standards-documents/docs/jesd-79f>.
- An Quan Jiang, Can Wang, Kui Juan Jin, Xiao Bing Liu, James F Scott, Cheol Seong Hwang, Ting Ao Tang, Hui Bin Lu, and Guo Zhen Yang. A resistive memory in semi-conducting BiFeO thin-film capacitors. *Advanced Materials*, 23(10):1277–1281, January 2011. doi: 10.1002/adma.201004317. URL <http://pubget.com/site/paper/21381130?institution=>.
- Lei Jiang, Bo Zhao, Youtao Zhang, and Jun Yang. Constructing large and fast multi-level cell STT-MRAM based cache for embedded processors. In *DAC '12: Proceedings of the 49th Annual Design Automation Conference*, pages 907–912. ACM Request Permissions, June 2012. doi: 10.1145/2228360.2228521. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6241611&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM%29.
- Lixian Jiang, Hiroshi Naganuma, Mikihiro Oogane, and Yasuo Ando. Large Tunnel Magnetoresistance of 1056Based Double Barrier Magnetic Tunnel Junction. *Applied Physics Express*, 2:083002, July 2009. doi: 10.1143/APEX.2.083002. URL <http://apex.ipap.jp/link?APEX/2/083002/>.
- R Colin Johnson. Memristors ready for prime time, July 2008. URL http://www.eetimes.com/document.asp?doc_id=1168871.
- M Julliere. Tunneling between ferromagnetic films. *Physics Letters A*, 54(3):225–226, September 1975. doi: 10.1016/0375-9601(75)90174-7. URL <http://linkinghub.elsevier.com/retrieve/pii/0375960175901747>.
- M H Jung, S Park, C Y You, and S Yuasa. Phys. Rev. B 81, 134419 (2010): Bias dependences of in-plane and out-of-plane spin-transfer torques in symmetric MgO-

- based magnetic tunnel junctions. *Physical Review B*, 2010. URL <http://prb.aps.org/abstract/PRB/v81/i13/e134419>.
- A B Kahng. Scaling: More than Moore's law. *Design & Test of Computers, IEEE*, 27(3):86–87, 2010. doi: 10.1109/MDT.2010.71. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5465130&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- K Kanda, H Sadaaki, and T Sakurai. 90 Solid-State Circuits, *IEEE Journal of*, 39(6):927–933, 2004. doi: 10.1109/JSSC.2004.827793. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1302269&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Y Kaneko, H Tanaka, M Ueda, Y Kato, and Eiji Fujii. A Dual-Channel Ferroelectric-Gate Field-Effect Transistor Enabling nand-Type Memory Characteristics. *Electron Devices, IEEE Transactions on*, 58(5):1311–1318, 2011. doi: 10.1109/TED.2011.2110653. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5720295.
- DerChang Kau, Stephen Tang, Ilya V Karpov, Rick Dodge, Brett Klehn, Johannes A Kalb, Jonathan Strand, Aleshandre Diaz, Nelson Leung, Jack Wu, Sean Lee, Tim Langtry, Kuo wei Chang, Christina Papagianni, Jinwook Lee, Jeremy Hirst, Swetha Erra, Eddie Flores, Nick Righos, Hernan Castro, and Gianpaolo Spadini. A stackable cross point Phase Change Memory. In *2009 IEEE International Electron Devices Meeting (IEDM)*, pages 1–4. IEEE, 2009. ISBN 978-1-4244-5639-0. doi: 10.1109/IEDM.2009.5424263. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5424263&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22A+Stackable+Cross+Point+Phase+Change+Memory%22%29.
- T Kawahara. Challenges toward gigabit-scale spin-transfer torque random access memory and beyond for normally off, green information technology infrastructure (Invited). *Journal of Applied Physics*, 109(7):07D325–07D325–6, 2011. doi: 10.1063/1.3556681. URL <http://dx.doi.org/10.1063/1.3556681>.
- Changkyu Kim, Doug Burger, and Stephen W Keckler. An adaptive, non-uniform cache structure for wire-delay dominated on-chip caches. In *Proceedings of the 10th international conference on Architectural support for programming languages and operating systems*, pages 211–222, New York, NY, USA, 2002. ACM. ISBN 1-58113-574-2. doi: 10.1145/605397.605420. URL <http://doi.acm.org/10.1145/605397.605420>.
- Joong-Sik Kim, Sung-Woong Chung, Tae-Su Jang, Seung-Hwan Lee, Dong-Hee Son, Seoung-Ju Chung, Sang-Min Hwang, S Banna, S Bhardwaj, M Gupta, Jungtae Kwon, D Kim, G Popov, V Gopinath, M Van Buskirk, Sang-Hoon Cho, Jae-Sung Roh, Sung-Joo Hong, and Sung-Wook Park. Vertical double gate Z-RAM technology with remarkable low voltage operation for DRAM application. In *VLSI Technology (VLSIT), 2010 Symposium on*, pages 163–164, 2010a. doi: 10.1109/VLSIT.2010.5556212. URL <http://dx.doi.org/10.1109/VLSIT.2010.5556212>.

- Keunwoo Kim, Ching-Te Chuang, J B Kuang, H C Ngo, and K J Nowka. Low-Power High-Performance Asymmetrical Double-Gate Circuits Using Back-Gate-Controlled Wide-Tunable-Range Diode Voltage. *Electron Devices, IEEE Transactions on*, 54(9):2263–2268, 2007. doi: 10.1109/TED.2007.902693. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4294215&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Kuk-Hwan Kim, Sung Hyun Jo, Siddharth Gaba, and Wei Lu. Nanoscale resistive memory with intrinsic diode characteristics and long endurance. *Applied Physics Letters*, 96(5):053106–053106–3, 2010b. doi: 10.1063/1.3294625. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5406703.
- Nam Sung Kim, David Blaauw, and Trevor Mudge. Leakage Power Optimization Techniques for Ultra Deep Sub-Micron Multi-Level Caches. In *ICCAD '03: Proceedings of the 2003 IEEE/ACM international conference on Computer-aided design*. IEEE Computer Society, November 2003. URL <http://portal.acm.org/citation.cfm?id=996070.1009957&coll=DL&dl=ACM&CFID=232708260&CFTOKEN=22541540>.
- SangBum Kim, Yuan Zhang, James P McVittie, H Jagannathan, Yoshio Nishi, and H S P Wong. Integrating Phase-Change Memory Cell With Ge Nanowire Diode for Crosspoint Memory—Experimental Demonstration and Analysis. *Electron Devices, IEEE Transactions on*, 55(9):2307–2313, 2008. doi: 10.1109/TED.2008.927631. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4603170.
- Seokjoong Kim and M R Guthaus. SNM-aware power reduction and reliability improvement in 45nm SRAMs. In *VLSI and System-on-Chip (VLSI-SoC), 2011 IEEE/IFIP 19th International Conference on*, pages 204–207, 2011. doi: 10.1109/VLSISoC.2011.6081666. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6081666&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Wanki Kim, Sung Il Park, Zhiping Zhang, Young Yang-Liau, D. Sekar, H.P. Wong, and S.S. Wong. Forming-free nitrogen-doped AIOX RRAM with sub-uA programming current. In *VLSI Technology (VLSIT), 2011 Symposium on*, pages 22–23, June 2011a.
- Woojin Kim, J H Jeong, Y Kim, W C Lim, J H Kim, J H Park, H J Shin, Y S Park, K S Kim, S H Park, Y J Lee, K W Kim, H J Kwon, H L Park, H S Ahn, S C Oh, J E Lee, S O Park, S Choi, H K Kang, and C Chung. Extended scalability of perpendicular STT-MRAM towards sub-20nm MTJ node. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 24.1.1–24.1.4, 2011b. doi: 10.1109/IEDM.2011.6131602. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6131602>.
- E Kitagawa and S Fujita. STT-MRAM cuts power use by 80pages 1–11, April 2013. URL http://www.eetimes.com/document.asp?doc_id=1280753.
- E Kitagawa, S Fujita, K Nomura, H Noguchi, K Abe, K Ikegami, T Daibou, Y Kato, C Kamata, S Kashiwada, N Shimomura, J Ito, and H Yoda. Impact of ultra low power and fast write operation of advanced perpendicular MTJ on power reduction for high-performance mobile CPU. *Electron Devices Meeting (IEDM)*,

- 2012 *IEEE International*, page 29, 2012. doi: 10.1109/IEDM.2012.6479129. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6479129&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- U F Kocks, C N Tomei, and H R Wenk. *Texture and Anisotropy: Preferred Orientations in Polycrystals and Their Effect on Materials Properties*. Cambridge University Press, 2000. ISBN 9780521794206. URL <http://books.google.fr/books?id=vkyU9KZBTioC>.
- H Kohlstedt, N Pertsev, J Rodríguez Contreras, and R Waser. Theoretical current-voltage characteristics of ferroelectric tunnel junctions. *Phys. Rev. B*, 72(12):125341, September 2005. doi: 10.1103/PhysRevB.72.125341. URL <http://link.aps.org/doi/10.1103/PhysRevB.72.125341>.
- P Kolar, E Karl, U Bhattacharya, F Hamzaoglu, H Nho, Yong-Gee Ng, Yih Wang, and K Zhang. A 32 nm High-k Metal Gate SRAM With Adaptive Dynamic Stability Enhancement for Low-Voltage Operation. *Solid-State Circuits, IEEE Journal of*, 46(1):76–84, 2011. doi: 10.1109/JSSC.2010.2084490. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5621844&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- U Krey and A Owen. *Basic Theoretical Physics: A Concise Overview*. SpringerLink: Springer e-Books. Springer-Verlag Berlin Heidelberg, 2007. ISBN 9783540368052. URL http://books.google.fr/books?id=xZ_QelBmkxYC.
- I N Krivorotov, G E Rowlands, T Rahman, J A Katine, J Langer, A Lyle, H Zhao, J G Alzate, A A Kovalev, Y Tserkovnyak, Z M Zeng, H W Jiang, K Galatsis, Y M Huai, P K Amiri, K L Wang, and J Wang. Ultrafast spin torque memory based on magnetic tunnel junctions with combined in-plane and perpendicular polarizers. In *Device Research Conference (DRC), 2012 70th Annual*, pages 211–212, 2012. doi: 10.1109/DRC.2012.6256944. URL <http://dx.doi.org/10.1109/DRC.2012.6256944>.
- J B Kuang, H C Ngo, K J Nowka, J C Law, and R V Joshi. A low-overhead virtual rail technique for SRAM leakage power reduction. In *Computer Design: VLSI in Computers and Processors, 2005. ICCD 2005. Proceedings. 2005 IEEE International Conference on*, pages 574–579, 2005. doi: 10.1109/ICCD.2005.11. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1524209&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- Yongbian Kuang Yongbian Kuang, Ru Huang Ru Huang, Yu Tang Yu Tang, Wei Ding Wei Ding, Lijie Zhang Lijie Zhang, and Yangyuan Wang Yangyuan Wang. Flexible Single-Component-Polymer Resistive Memory for Ultrafast and Highly Compatible Nonvolatile Memory Applications. *Electron Device Letters, IEEE*, 31(7):758–760, June 2010. doi: 10.1109/LED.2010.2048297. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5466137.

- Carsten Kügeler, Roland Rosezin, Eike Linn, Rainer Bruchhaus, and Rainer Waser. Materials, technologies, and circuit concepts for nanocrossbar-based bipolar RRAM. *Applied Physics A*, 102(4):791–809, January 2011. doi: 10.1007/s00339-011-6287-2. URL <http://link.springer.com/article/10.1007/s00339-011-6287-2>.
- H P Labs. CACTI An integrated cache and memory access time, cycle time, area, leakage, and dynamic power model.
- Y Lakys, Wei Sheng Zhao, T Devolder, Yue Zhang, J O Klein, D Ravlosona, and C Chappert. Self-Enabled “Error-Free” Switching Circuit for Spin Transfer Torque MRAM and Logic. *Magnetics, IEEE Transactions on*, 48(9):2403–2406, 2012. doi: 10.1109/TMAG.2012.2194790. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6184315&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- Chun Ning Lau, Duncan R Stewart, R Stanley Williams, and Marc Bockrath. Direct Observation of Nanoscale Switching Centers in Metal/Molecule/Metal Structures. *Nano Letters*, 4(4):569–572, April 2004. doi: 10.1021/nl035117a. URL <http://pubs.acs.org/doi/abs/10.1021/nl035117a>.
- Byoungil Lee and H S P Wong. Fabrication and Characterization of Nanoscale NiO Resistance Change Memory (RRAM) Cells With Confined Conduction Paths. *Electron Devices, IEEE Transactions on*, 58(10):3270–3275, 2011. doi: 10.1109/TED.2011.2161311. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5993527&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28ReRAM+AND+RRAM%29.
- Ching-Ting Lee, Li-Zhen Yu, and Hung-Chun Chen. Memory bistable mechanisms of organic memory devices. *Applied Physics Letters*, 97(4):043301–043301, 2010a. doi: 10.1063/1.3467050. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5529980.
- Chunho Lee, Miodrag Potkonjak, and William H Mangione-Smith. MediaBench: a tool for evaluating and synthesizing multimedia and communications systems. In *Proceedings of the 30th annual ACM/IEEE international symposium on Microarchitecture*, pages 330–335, Washington, DC, USA, 1997. IEEE Computer Society. ISBN 0-8186-7977-8. URL <http://dl.acm.org/citation.cfm?id=266800.266832>.
- Daeseok Lee, Jubong Park, Seungjae Jung, Godeuni Choi, Joonmyoung Lee, Seonghyun Kim, Jiyong Woo, M Siddik, Eujun Cha, and Hyunsang Hwang. Operation Voltage Control in Complementary Resistive Switches Using Heterodevice. *Electron Device Letters, IEEE*, 33(4):600–602, 2012a. doi: 10.1109/LED.2012.2186113. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6165638&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28ReRAM+AND+%28RRAM+AND+%28%22Cross+point%22%29%29%29.
- H Y Lee, Y S Chen, P S Chen, P Y Gu, Y Y Hsu, S M Wang, W H Liu, C H Tsai, S S Sheu, P C Chiang, W P Lin, C H Lin, W S Chen, F T Chen, C H Lien, and M-J Tsai. Evidence and solution of over-RESET problem for HfOX based resistive

- memory with sub-ns switching speed and high endurance. In *2010 IEEE International Electron Devices Meeting (IEDM)*, pages 19.7.1–19.7.4. IEEE, 2010b. ISBN 978-1-4424-7418-5. doi: 10.1109/IEDM.2010.5703395. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5703395>.
- Kwang-Jin Lee, Beak-Hyung Cho, Woo-Yeong Cho, Sangbeom Kang, Byung-Gil Choi, Hyung-Rok Oh, Chang-Soo Lee, Hye-Jin Kim, Joon-Min Park, Qi Wang, Mu-Hui Park, Yu-Hwan Ro, Joon-Yong Choi, Ki-Sung Kim, Young-Ran Kim, In-Cheol Shin, Ki-Won Lim, Ho-Keun Cho, Chang-Han Choi, Won-Ryul Chung, Du-Eung Kim, Yong-Jin Yoon, Kwang-Suk Yu, Gi-Tae Jeong, Hong-Sik Jeong, Choong-Keun Kwak, Chang-Hyun Kim, and Kinam Kim. A 90 nm 1.8 V 512 Mb Diode-Switch PRAM With 266 MB/s Read Throughput. *Solid-State Circuits, IEEE Journal of*, 43(1):150–162, January 2008. doi: 10.1109/JSSC.2007.908001. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4443196&queryText%253D%2528%2528A+90+nm+1.8+V+512+Mb+Diode-Switch+PRAM+With+266+MB%252Fs+Read+Throughput%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Kwangseok Lee and Woo Young Choi. Nanoelectromechanical Memory Cell (T Cell) for Low-Cost Embedded Nonvolatile Memory Applications. *Electron Devices, IEEE Transactions on*, 57(1):1264–1267, March 2011. doi: 10.1109/TED.2010.2104154. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2011ITED...58.1264L&link_type=EJOURNAL.
- M J Lee, Y Park, D S Suh, E H Lee, S Seo, D C Kim, R Jung, B S Kang, S E Ahn, C B Lee, D H Seo, Y K Cha, I K Yoo, J S Kim, and B H Park. Two Series Oxide Resistors Applicable to High Speed and High Density Nonvolatile Memory. *Advanced Materials*, 19(22):3919–3923, 2007a. doi: 10.1002/adma.200700251. URL <http://dx.doi.org/10.1002/adma.200700251>.
- M J Lee, Y Park, D S Suh, E H Lee, S Seo, D C Kim, R Jung, B S Kang, S E Ahn, C B Lee, D H Seo, Y K Cha, I K Yoo, J S Kim, and B H Park. Two Series Oxide Resistors Applicable to High Speed and High Density Nonvolatile Memory. *Advanced Materials*, 19(22):3919–3923, November 2007b. doi: 10.1002/adma.200700251. URL <http://doi.wiley.com/10.1002/adma.200700251>.
- Myoung-Jae MJ Lee, Chang Bum CB Lee, Dongsoo D Lee, Seung Ryul SR Lee, Man M Chang, Ji Hyun JH Hur, Young-Bae YB Kim, Chang-Jung CJ Kim, David H DH Seo, Sunae S Seo, U-In UI Chung, In-Kyeong IK Yoo, and Kinam K Kim. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta₂O_(5-x)/TaO_(2-x) bilayer structures. *Nature Materials*, 10(8):625–630, July 2011a. doi: 10.1038/nmat3070. URL <http://www.nature.com/doifinder/10.1038/nmat3070>.
- Sang Wook Lee, Seung Joo Park, Eleanor E B Campbell, and Yung Woo Park. A fast and low-power microelectromechanical system-based non-volatile memory device. *Nature Communications*, 2:220–220, February 2011b. doi: 10.1038/ncomms1227. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2011NatCo...2E.220L&link_type=ABSTRACT.

- Sungchul Lee, Kwangseok Kim, Keewon Kim, Unghwan Pi, Youngman Jang, U-In Chung, Inkyung Yoo, and Kinam Kim. Highly scalable STT-MRAM with 3-dimensional cell structure using in-plane magnetic anisotropy materials. *VLSI Technology (VLSIT), 2012 Symposium on*, pages 65–66, 2012b. doi: 10.1109/VLSIT.2012.6242463. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6242463&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM%29.
- A Lesea, S Drimer, J J Fabula, C Carmichael, and P Alfke. The rosetta experiment: atmospheric soft error rate testing in differing technology FPGAs. *Device and Materials Reliability, IEEE Transactions on DOI - 10.1109/TDMR.2005.854207*, 5(3):317–328, 2005. URL [10.1109/TDMR.2005.854207](http://dx.doi.org/10.1109/TDMR.2005.854207).
- Austin Lesea. Xilinx WP286 Continuing Experiments of Atmospheric Neutron Effects on Deep Submicron Integrated Circuits. pages 1–9, May 2009.
- Hai Li, Xiaobin Wang, Zhong-Liang Ong, Weng-Fai Wong, Yaojun Zhang, Peiyuan Wang, and Yiran Chen. Performance, Power, and Reliability Tradeoffs of STT-RAM Cell Subject to Architecture-Level Requirement. *Magnetics, IEEE Transactions on*, 47(10):2356–2359, 2011a. doi: 10.1109/TMAG.2011.2159262. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6027811&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22Magnetic+Tunneling+Junction%22+AND+%28MTJ+AND+%28MRAM+AND+%28Spintronics+AND+STT%29%29%29%29.
- YingTao Li, ShiBing Long, Qi Liu, HangBing Lü, Su Liu, and Ming Liu. An overview of resistive random access memory devices. *Chinese Science Bulletin*, 56(28-29):3072–3078, September 2011b. doi: 10.1007/s11434-011-4671-0. URL <http://link.springer.com/10.1007/s11434-011-4671-0>.
- YingTao Li, Hangbing Lv, Qi Liu, ShiBing Long, Ming Wang, Hongwei Xie, Kangwei Zhang, Zongliang Huo, and Ming Liu. Bipolar one diode–one resistor integration for high-density resistive memory applications. *Nanoscale*, 5(11):4785, 2013. doi: 10.1039/c3nr33370a. URL <http://xlink.rsc.org/?DOI=c3nr33370a>.
- Elliott Lieb. The Hubbard Model: Some Rigorous Results and Open Problems. November 1993. URL <http://arxiv.org/abs/cond-mat/9311033>.
- Eike Linn, Roland Rosezin, Carsten Kügeler, and Rainer Waser. Complementary resistive switches for passive nanocrossbar memories. *Nature Materials*, 9(5):403–406, May 2010a. doi: 10.1038/nmat2748. URL <http://www.nature.com/nmat/journal/v9/n5/abs/nmat2748.html#supplementary-information>.
- Eike Linn, Roland Rosezin, Carsten Kügeler, and Rainer Waser. Complementary resistive switches for passive nanocrossbar memories. *Nature Materials*, 9(5):403–406, April 2010b. doi: 10.1038/nmat2748. URL <http://www.nature.com/nmat/journal/v9/n5/abs/nmat2748.html#supplementary-information>.
- Karen Linser. CCS Power Library Characterization Guideline. pages 1–29, January 2010. URL ftp://ftp.synopsys.com/doc/Power_Char_Guide_4.1/ccs_power_char_guide_41.pdf.

- Dannngis Liu. Integrated circuit having memory cell array, and method of manufacturing same. Technical report, 2009.
- S J Liu, Z H Lin, Q Zhao, Y Ma, and H F Shi. FlashMemory Effect for Polyfluorenes with OnChain Iridium (III) Complexes. *Advanced Functional ...*, 2011. URL <http://onlinelibrary.wiley.com/doi/10.1002/adfm.201001884/full>.
- Owen O Loh, Xiaoding X Wei, Changhong C Ke, John J Sullivan, and Horacio D HD Espinosa. Robust carbon-nanotube-based nano-electromechanical devices: understanding and eliminating prevalent failure modes using alternative electrode materials. *Small*, 7(1):79–86, January 2011. doi: 10.1002/smll.201001166. URL <http://onlinelibrary.wiley.com/doi/10.1002/smll.201001166/full>.
- Alvin Loke and Jordan Lai. Session 15 - IC Technology - more Moore and more than Moore. In *Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE*, 2008. doi: 10.1109/CICC.2008.4672099. URL <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4672099>.
- Z Luo, A Steegen, M Eller, and R Mann. IEEE Xplore Download. ... *Meeting*, 2004. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1419254.
- Ken Mackay. TAS, TAS+STT-MRAM and Magnetic Logic Unit. Technical report, Gardanne, Provence-Alpes-Côte d'Azur, France, November 2011.
- R Mandhdapu and J Samson Immanuel. Realization of spin-torque transfer magnetoresistive RAM. In *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*, pages 513–517, 2012. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6216056&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22Magnetic+Tunneling+Junction%22+AND+%28MTJ+AND+%28MRAM+AND+%28STT%29%29%29.
- Mengjie Mao, Hai Helen Li, Alex K Jones, and Yiran Chen. Coordinating prefetching and STT-RAM based last-level cache management for multicore systems. In *GLSVLSI '13: Proceedings of the 23rd ACM international conference on Great lakes symposium on VLSI*. ACM Request Permissions, May 2013. doi: 10.1145/2483028.2483060. URL <http://portal.acm.org/citation.cfm?id=2483028.2483060&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- M Marinella. The future of memory. In *Aerospace Conference, 2013 IEEE*, pages 1–11, 2013. doi: 10.1109/AERO.2013.6507427. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6507427&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM+OR+%28PCM+OR+%28RRAM+OR+OxRAM%29%29.
- Milo M K Martin, Daniel J Sorin, Bradford M Beckmann, Michael R Marty, Min Xu, Alaa R Alameldeen, Kevin E Moore, Mark D Hill, and David A Wood. Multifacet's general execution-driven multiprocessor simulator (GEMS) toolset. *ACM SIGARCH Computer Architecture News*, 33(4):92, November 2005. doi: 10.1145/1105734.1105747. URL <http://portal.acm.org/citation.cfm?doid=1105734.1105747>.

- F Masuoka, M Momodom, Y Iwata, and R Shirota. New ultra high density EPROM and flash EEPROM with NAND structure cell. In *1987 International Electron Devices Meeting*, pages 552–555. IRE. doi: 10.1109/IEDM.1987.191485. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1487443>.
- J Mathon and A Umerski. Theory of tunneling magnetoresistance of an epitaxial Fe/MgO/Fe(001) junction. *Phys. Rev. B*, 63(22):220403, May 2001. doi: 10.1103/PhysRevB.63.220403. URL <http://link.aps.org/doi/10.1103/PhysRevB.63.220403>.
- Shoun Matsunaga, Kimiyuki Hiyama, Atsushi Matsumoto, SHOJI IKEDA, Haruhiro Hasegawa, KATSUYA MIURA, Jun Hayakawa, Tetsuo Endoh, HIDEO OHNO, and Takahiro Hanyu. Standby-Power-Free Compact Ternary Content-Addressable Memory Cell Chip Using Magnetic Tunnel Junction Devices. *Applied Physics Express*, 2:023004, February 2009. doi: 10.1143/APEX.2.023004. URL <http://apex.ipap.jp/Link?APEX/2/023004/>.
- Shoun Matsunaga, Akira Katsumata, Masanori Natsui, Tetsuo Endoh, HIDEO OHNO, and Takahiro Hanyu. Design of a 27ops-access 7-transistor/2-magnetic-tunnel-junction cell circuit for a high-speed-search nonvolatile ternary content-addressable memory. *Journal of Applied Physics*, 111(7):07E336–07E336–3, April 2012. doi: 10.1063/1.3677875. URL http://jap.aip.org/resource/1/japiau/v111/i7/p07E336_s1.
- Christopher R McWilliams, Jolanta Celinska, Carlos A Paz de Araujo, and Kan-Hao Xue. Device characterization of correlated electron random access memories. *Journal of Applied Physics*, 109(9):091608–091608–6, 2011. doi: 10.1063/1.3581206. URL <http://link.aip.org/link/JAPIAU/v109/i9/p091608/s1&Agg=doi>.
- W H Meiklejohn and C P Bean. New Magnetic Anisotropy. *Phys. Rev.*, 105:904–913, February 1957. doi: 10.1103/PhysRev.105.904. URL <http://link.aps.org/doi/10.1103/PhysRev.105.904>.
- A Mejdoubi, B Lacoste, G Prenat, and B Dieny. Macrospin model of precessional spin-transfer-torque switching in planar magnetic tunnel junctions with perpendicular polarizer. *Applied Physics Letters*, 102(15):152413–152413–4, 2013. doi: 10.1063/1.4802720. URL <http://link.aip.org/link/APPLAB/v102/i15/p152413/s1&Agg=doi>.
- Vincent Meunier, Sergei Kalinin, and Bobby Sumpter. Nonvolatile Memory Elements Based on the Intercalation of Organic Molecules Inside Carbon Nanotubes. *Physical Review Letters*, 98(5):056401, February 2007. doi: 10.1103/PhysRevLett.98.056401. URL <http://link.aps.org/doi/10.1103/PhysRevLett.98.056401>.
- Byoung-Chul Min, Il-Jae Shin, Gyung-Min Choi, Chiyui Ahn, J Langer, B Ocker, W Maass, and Kyung-Ho Shin. MgO-based magnetic tunnel junctions for spin-transfer-torque random access memory. In *Nanotechnology (IEEE-NANO), 2010 10th IEEE Conference on*, pages 144–147, 2010. doi: 10.1109/NANO.2010.5697730. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5697730&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22CoFeB+MgO+CoFeB%22%29.
- Asit K Mishra, Xiangyu Dong, Guangyu Sun, Yuan Xie, N Vijaykrishnan, and Chita R Das. Architecting on-chip interconnects for stacked 3D STT-RAM caches in CMPs.

- ISCA '11: *Proceeding of the 38th annual international symposium on Computer architecture*, June 2011. doi: 10.1145/2000064.2000074. URL <http://portal.acm.org/citation.cfm?id=2000064.2000074&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- T Miyazaki and N Tezuka. Giant magnetic tunneling effect in Fe/Al₂O₃/Fe junction. *Journal of Magnetism and Magnetic Materials*, 1995. URL <http://www.sciencedirect.com/science/article/pii/0304885395900012>.
- T Miyazaki, T Yaoi, and S Ishio. Large magnetoresistance effect in 82Ni-Fe/Al-Al₂O₃/Co magnetic tunneling junction. *Journal of Magnetism and Magnetic Materials*, 98(1-2):L7–L9, July 1991. doi: 10.1016/0304-8853(91)90417-9. URL <http://linkinghub.elsevier.com/retrieve/pii/0304885391904179>.
- Baker Mohammad, Percy Dadabhoy, Ken Lin, and Paul Bassett. Comparative study of current mode and voltage mode sense amplifier used for 28nm SRAM. In *Microelectronics (ICM), 2012 24th International Conference on*, pages 1–6, 2012. doi: 10.1109/ICM.2012.6471396. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6471396&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29.
- M Monchiero, R Canal, and A Gonzalez. Power/Performance/Thermal Design-Space Exploration for Multicore Architectures. *Parallel and Distributed Systems, IEEE Transactions on*, 19(5):666–681, 2008. doi: 10.1109/TPDS.2007.70756. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4359440&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22memory+hierarchy%22+AND+%28power+AND+L2%29%29.
- J S Moodera, Lisa R Kinder, Terrilyn M Wong, and R Meservey. Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions. *Physical Review Letters*, 74(16):3273–3276, April 1995. doi: 10.1103/PhysRevLett.74.3273. URL <http://link.aps.org/doi/10.1103/PhysRevLett.74.3273>.
- Gordon E Moore. Competitive potential of semiconductor memories. *Magnetics, IEEE Transactions on*, 6(3):590, 1970. doi: 10.1109/TMAG.1970.1066900. URL <http://dx.doi.org/10.1109/TMAG.1970.1066900>.
- Gordon E Moore. Cramming more components onto integrated circuits. *IEEE Solid-State Circuits Newsletter*, 11(5):33–35, May 2006. doi: 10.1109/N-SSC.2006.4785860. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4785860>.
- Daniel Morris, David Bromberg, Jian-Gang Jimmy Zhu, and Larry Pileggi. mLogic: Ultra-low voltage non-volatile logic circuits using STT-MTJ devices. In *Design Automation Conference (DAC), 2012 49th ACM/EDAC/IEEE*, pages 486–491, 2012. doi: 10.1145/2228360.2228446. URL <http://portal.acm.org/citation.cfm?id=2228360.2228446&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Biswanath Mukherjee, Moumita Mukherjee, Jae-eun Park, and Seungmoon Pyo. High-Performance Molecular Memory Device Using AgTCNQ Crystals Grown by Solution Process. *The Journal of Physical Chemistry C*, 114(1):567–571, January 2010. doi: 10.1021/jp907342r. URL <http://pubs.acs.org/doi/abs/10.1021/jp907342r>.

- R Müller, S De Jonge, K Myny, D J Wouters, J Genoe, and P Heremans. Organic CuTCNQ integrated in complementary metal oxide semiconductor copper back end-of-line for nonvolatile memories. *Applied Physics Letters*, 89(22): 223501–223501, 2006. doi: 10.1063/1.2388883. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4823557&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22Organic+CuTCNQ+integrated+in+complementary+metal+oxide+semiconductor+copper+back+end-of-line+for+nonvolatile+memories+%7C+Browse+-+Applied+Physics+Letters%22%29.
- R Müller, O Rouault, A Katzenmeyer, L Goux, D J Wouters, J Genoe, and P Heremans. Electrodeposition of copper tetracyanoquinodimethane for bipolar resistive switching non-volatile memories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1905):4191–4201, September 2009. doi: 10.1021/nl0619204. URL <http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2008.0300>.
- Naveen Muralimanohar, Rajeev Balasubramonian, and Norman P. Jouppi. Cacti 6.0: A tool to model large caches. Technical Report 85, HP Laboratories, Published in International Symposium on Microarchitecture, Chicago, Dec 2007, April 2009.
- Computer History Museum. <http://www.computerhistory.org/timeline/?category=cmpnt>, a.
- Computer History Museum. <http://www.computerhistory.org/semiconductor/timeline.html>, b.
- Masahiko Nakayama, Tadashi Kai, Naoharu Shimomura, Minoru Amano, Eiji Kitagawa, Toshihiko Nagase, Masatoshi Yoshikawa, Tatsuya Kishi, Sumio Ikegawa, and Hiroaki Yoda. Spin transfer switching in TbCoFeCoFeBMgOCoFeBTbCoFe magnetic tunnel junctions with perpendicular magnetic anisotropy. *Journal of Applied Physics*, 103(7):07A710, 2008. doi: 10.1063/1.2838335. URL <http://link.aip.org/link/JAPIAU/v103/i7/p07A710/s1&Agg=doi>.
- H Nambu, K Kanetani, K Yamasaki, K Higeta, M Usami, Y Fujimura, K Ando, T Kusunoki, K Yamaguchi, and N Homma. A 1.8-ns access, 550-MHz, 4.5-Mb CMOS SRAM. *Solid-State Circuits, IEEE Journal of*, 33(11):1650–1658, 1998. doi: 10.1109/4.726553. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=726553&openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526filter%253DAND%2528NOT%25284283010803%2529%2529%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All%2526queryText%253D%2528sram+cache%2529+AND+%25281+cycle+delay%2529%2529.
- S Natarajan, S Chung, L Paris, and A Keshavarzi. Searching for the dream embedded memory. *Solid-State Circuits Magazine, IEEE*, 1(3):34–44, 2009. doi: 10.1109/MSSC.2009.933521. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=5191434&queryText%253D%2528SRAM%2529+AND+%2528Search+Index+Terms%253AMRAM%2529+AND+%2528Search+Index+Terms%253Acache%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_

- Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- E J Ng, JBW Soon, N Singh, N Shen, VXH Leong, T Myint, V Pott, and J M Tsai. High density vertical silicon NEM switches with CMOS-compatible fabrication. *Electronics Letters*, 47(1):759–760, June 2011. doi: 10.1049/el.2011.1073. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5931008.
- Manish K MK Niranjan, Yong Y Wang, Sitaram S SS Jaswal, and Evgeny Y EY Tsymbal. Prediction of a switchable two-dimensional electron gas at ferroelectric oxide interfaces. *Physical Review Letters*, 103(1):016804–016804, July 2009. doi: 10.1103/PhysRevLett.103.016804. URL <http://link.aps.org/doi/10.1103/PhysRevLett.103.016804>.
- NVIDIA Corporation. NVIDIA Kepler Compute Architecture | High Performance Computing | NVIDIA, a. URL <http://www.nvidia.com/object/nvidia-kepler.html>.
- NVIDIA Corporation. Kepler™ GK110, b. URL <http://www.nvidia.com/content/PDF/kepler/NVIDIA-Kepler-GK110-Architecture-Whitepaper.pdf>.
- J Oh, J Park, Y Lim, H Lim, Y Oh, J Kim, J Shin, Y Song, K Ryoo, D Lim, S Park, J Yu, F Yeung, C Jeong, J Kong, D Kang, G Koh, G Jeong, H Jeong, and Kinam Kim. Full Integration of Highly Manufacturable 512Mb PRAM based on 90nm Technology. *Electron Devices Meeting, 2006. IEDM '06. International*, pages 1–4, 2006. doi: 10.1109/IEDM.2006.346905. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4154324&queryText%253D%2528Document+Title%253AFull+Integration+of+Highly+Manufacturable+512Mb+PRAM+based+on+90nm+Technology%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Se-Chung Oh, Seung Young Park, Aurélien Manchon, Mairbek Chshiev, Jae-Ho Han, Hyun-Woo Lee, Jang-Eun Lee, Kyung-Tae Nam, Younghun Jo, Yo-Chan Kong, Bernard Dieny, and Kyung-Jin Lee. Bias-voltage dependence of perpendicular spin-transfer torque in asymmetric MgO-based magnetic tunnel junctions. *Nature Physics*, 5(12):898–902, October 2009. doi: 10.1038/nphys1427. URL <http://www.nature.com/doifinder/10.1038/nphys1427>.
- H Ohno. Magnetic tunnel junction for magnetoresistive random access memory and beyond. *Silicon Nanoelectronics Workshop (SNW), 2012 IEEE*, pages 1–2, 2012. doi: 10.1109/SNW.2012.6243328. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6243328&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22Magnetic+Tunnel+Junction+for+Magnetoresistive+Random+Access+Memory+and+Beyond%22%29.
- A A Ohtomo and H Y HY Hwang. A high-mobility electron gas at the LaAlO₃/SrTiO₃ heterointerface. *Nature*, 427(6973):423–426, January 2004. doi: 10.1038/nature02308. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2004Natur.427..423O&link_type=ABSTRACT.

- S Okabe and K Abe. A CAM-based Low-power Highly Associative Cache for High-performance Embedded Processors. *Proc*, 2010. URL http://almond.cs.uec.ac.jp/papers/pdf/2010/IWMST2010_okabe.pdf.
- Xin Ou, Pratyush Das Kanungo, Reinhard Kögler, Peter Werner, Ulrich Gösele, Wolfgang Skorupa, and Xi Wang. Carrier profiling of individual si nanowires by scanning spreading resistance microscopy. *Nano Letters*, 10(1):171–175, 2010. doi: 10.1021/nl903228s. URL <http://pubs.acs.org/doi/abs/10.1021/nl903228s>. PMID: 20014820.
- Georgios Panagopoulos, Charles Augustine, and Kaushik Roy. A framework for simulating hybrid MTJ/CMOS circuits: Atoms to system approach. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2012*, pages 1443–1446, 2012. doi: 10.1109/DATE.2012.6176592. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6176592&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- S Panda, N M Kumar, and C K Sarkar. Power, delay and noise optimization of a SRAM cell using a different threshold voltages and high performance output noise reduction circuit. In *Computers and Devices for Communication, 2009. CODEC 2009. 4th International Conference on*, pages 1–4, 2009. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5407194&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- J W JW Park, D F DF Bogorin, C C Cen, D A DA Felker, Y Y Zhang, C T CT Nelson, C W CW Bark, C M CM Folkman, X Q XQ Pan, M S MS Rzchowski, J J Levy, and C B CB Eom. Creation of a two-dimensional electron gas at an oxide interface on silicon. *Nature Communications*, 1:94–94, December 2009. doi: 10.1038/ncomms1096. URL <http://pubget.com/site/paper/20981022?institution=>.
- Jeong-Heon Park, Y Kim, W C Lim, J H Kim, S H Park, J H Kim, W Kim, K W Kim, J H Jeong, K S Kim, H Kim, Y J Lee, S C Oh, J E Lee, S O Park, S Watts, D Apalkov, V Nikitin, M Krounbi, S Jeong, S Choi, H K Kang, and C Chung. Enhancement of data retention and write current scaling for sub-20nm STT-MRAM by utilizing dual interfaces for perpendicular magnetic anisotropy. In *VLSI Technology (VLSIT), 2012 Symposium on*, pages 57–58, 2012a. doi: 10.1109/VLSIT.2012.6242459. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6242459&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22Enhancement+of+data+retention+and+write+current+scaling+for+sub-20nm+STT-MRAMby+utilizing+dual+interfaces+for+perpendicular+magnetic+anisotropy%22%29.
- Samdae S Park, Kyungtae K Kim, Dong Min DM Kim, Wonsang W Kwon, Junman J Choi, and Moonhor M Ree. High temperature polyimide containing anthracene moiety and its structure, interface, and nonvolatile memory behavior. *ACS Applied Materials & Interfaces*, 3(3):765–773, February 2011. doi: 10.1021/am101125d. URL <http://pubs.acs.org/doi/abs/10.1021/am101125d>.

- Sang Phill Park, Sumeet Gupta, Niladri Mojumder, Anand Raghunathan, and Kaushik Roy. Future cache design using STT MRAMs for improved energy efficiency: devices, circuits and architecture. In *DAC '12: Proceedings of the 49th Annual Design Automation Conference*, pages 492–497. ACM Request Permissions, June 2012b. doi: 10.1145/2228360.2228447. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6241551&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- Seong-Geon Park, Min Kyu Yang, Hyunsu Ju, Dong-Jun Seong, Jung Moo Lee, Eunmi Kim, Seungjae Jung, Lijie Zhang, Yoo Cheol Shin, In-Gyu Baek, Jungdal Choi, Ho-Kyu Kang, and Chilhee Chung. A non-linear ReRAM cell with sub-1 μ A ultralow operating current for high density vertical resistive memory (VRRAM). *Electron Devices Meeting (IEDM), 2012 IEEE International*, page 20, 2012c. doi: 10.1109/IEDM.2012.6479084. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6479084&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28ReRAM+AND+RRAM%29.
- W Y Park, G H Kim, J Y Seok, K M Kim, and S J Song. A Pt/TiO₂/Ti Schottky-type selection diode for alleviating the sneak current in resistance switching memory arrays - Abstract - Nanotechnology - IOPscience. . . ., 2010. URL <http://iopscience.iop.org/0957-4484/21/19/195201>.
- Yong-Ha Park, Jeonghoon Kook, and Hoi-Jun Yoo. Embedded DRAM (eDRAM) power-energy estimation for system-on-a-chip (SoC) applications. In *Design Automation Conference, 2002. Proceedings of ASP-DAC 2002. 7th Asia and South Pacific and the 15th International Conference on VLSI Design. Proceedings*, pages 625–630, 2002. doi: 10.1109/ASPDAC.2002.995006. URL <http://dx.doi.org/10.1109/ASPDAC.2002.995006>.
- Stuart S P Parkin, Christian Kaiser, Alex Panchula, Philip M Rice, Brian Hughes, Mahesh Samant, and See-Hun Yang. Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers. *Nature Publishing Group*, 3(12):862–867, October 2004. doi: 10.1038/nmat1256. URL <http://www.nature.com/doi/finder/10.1038/nmat1256>.
- Abhay Pasupathy. Single-Molecule Electronic Devices. *APS Meeting Abstracts*, -1:16002, March 2003. URL <http://adsabs.harvard.edu/abs/2003APS..MARB16002P>.
- Patterson. Design Considerations for Single-Chip Computers of the Future. *Solid-State Circuits, IEEE Journal of*, 15(1):44–52, 1980. doi: 10.1109/JSSC.1980.1051337. URL <http://dx.doi.org/10.1109/JSSC.1980.1051337>.
- David A Patterson and Jhon L Hennessy. *Computer Organization and Design*. The hardware / software interface. fourth edition, April 2012.
- David A Patterson and John L Hennessy. *Computer organization and design: the hardware/software interface 3th*. Morgan Kaufmann, 2005. URL http://books.google.com/books?hl=en&lr=&id=3b63x-0P3_UC&oi=fnd&pg=PP1&dq=%2522Computer+Organization+and+Design:+The+Hardware/Software+Interface%2522+&ots=NyenZZp8M8&sig=A0baxg2Jc5fqRNluVqBg7dSnU6s.

- David A Patterson and John L Hennessy. *Computer organization and design: the hardware/software interface 4th*. UC Berkeley & Stanford, 2012.
- H Pilo, C A Adams, I Arsovski, R M Houle, S M Lamphier, M M Lee, F M Pavlik, S N Sambatur, A Seferagic, R Wu, and M I Younus. A 64Mb SRAM in 22nm SOI technology featuring fine-granularity power gating and low-energy power-supply-partition techniques for 37reduction. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, pages 322–323, 2013. doi: 10.1109/ISSCC.2013.6487753. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6487753&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- D W Plass and Y H Chan. IBM POWER6 SRAM arrays. *IBM Journal of Research and Development*, 51(6):747–756, 2007. doi: 10.1147/rd.516.0747. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5388631&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- M Powell, A Agarwal, T Vijaykumar, B Falsafi, and K Roy. Reducing set-associative cache energy via way-prediction and selective direct-mapping. *Microarchitecture, 2001. MICRO-34. Proceedings. 34th ACM/IEEE International Symposium on*, pages 54–65, 2001. doi: 10.1109/MICRO.2001.991105. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=991105&openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526filter%253DAND%2528NOT%25284283010803%2529%2529%2526matchBoolean%253Dtrue%2526pageNumber%253D4%2526rowsPerPage%253D50%2526searchField%253DSearch+All%2526queryText%253D%2528%2528cache+cycle+delay%2529%2529.
- D D Prime, S S Paul, and P W PW Josephs-Franks. Gold nanoparticle charge trapping and relation to organic polymer memory devices. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1905):4215–4225, October 2009. doi: 10.1098/rsta.2009.0141. URL <http://rsta.royalsocietypublishing.org/content/367/1905/4215.short>.
- T Pro, J Buckley, K Huang, A Calborean, M Gely, G Delapierre, G Ghibaudo, F Duclairoir, J-C Marchon, E Jalaguier, P Maldivi, B De Salvo, and S Deleonibus. Investigation of Hybrid Molecular/Silicon Memories With Redox-Active Molecules Acting as Storage Media. *Nanotechnology, IEEE Transactions on*, 8(2):204–213, February 2009. doi: 10.1109/TNANO.2008.2009875. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4689413.
- Sarath C Puthentheradam, Dieter K Schroder, and Michael N Kozicki. Inherent diode isolation in programmable metallization cell resistive memory elements. *Applied Physics A*, 102(4):817–826, 2011a. URL <http://link.springer.com/article/10.1007/s00339-011-6292-5>.
- Sarath C Puthentheradam, Dieter K Schroder, and Michael N Kozicki. Inherent diode isolation in programmable metallization cell resistive memory elements. In *Applied*

- Physics A*, pages 817–826. Springer, March 2011b. doi: 10.1007/s00339-011-6292-5. URL <http://link.springer.com/article/10.1007%2Fs00339-011-6292-5>.
- Moinuddin K Qureshi, Vijayalakshmi Srinivasan, and Jude A Rivers. Scalable high performance main memory system using phase-change memory technology. In *ISCA '09: Proceedings of the 36th annual international symposium on Computer architecture*, pages 1–10. ACM Request Permissions, June 2009. doi: 10.1145/1555754.1555760. URL <http://portal.acm.org/citation.cfm?id=1555754.1555760&coll=DL&dl=ACM&CFID=232708260&CFTOKEN=22541540>.
- S Rajendra Prasad, B K Madhavi, and K Lal Kishore. Design of 32nm Forced Stack CNTFET SRAM cell for leakage power reduction. In *Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on*, pages 629–633, 2012. doi: 10.1109/ICCEET.2012.6203904. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6203904&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- S Raoux, G W Burr, M J Breitwisch, C T Rettner, Y C Chen, R M Shelby, M Salinga, D Krebs, S H Chen, H L Lung, and C H Lam. Phase-change random access memory: A scalable technology. *IBM Journal of Research and Development*, 52(4.5):465–479, July 2008. doi: 10.1147/rd.524.0465. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5388621>.
- M A Reed, J Chen, A M Rawlett, D W Price, and J M Tour. Molecular random access memory cell. *Applied Physics Letters*, 78(23):3735–3737, 2001. doi: 10.1063/1.1377042. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4896815&queryText%253D%2528%2528Document+Title%253AMolecular+random+access+memory+cell%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Glen Reinman and Norman P Jouppi. CACTI 2.0: An Integrated Cache Timing, Power, and Area Model. Technical report, 250 University Avenue Palo Alto, California 94301 USA, February 2000.
- W Rippard, M Pufall, S Kaka, T Silva, S Russek, and J Katine. Injection Locking and Phase Control of Spin Transfer Nano-oscillators. *Physical Review Letters*, 95(6):067203, August 2005. doi: 10.1103/PhysRevLett.95.067203. URL <http://link.aps.org/doi/10.1103/PhysRevLett.95.067203>.
- R Rosezin, E Linn, L Nielen, C Kugeler, R Bruchhaus, and R Waser. Integrated Complementary Resistive Switches for Passive High-Density Nanocrossbar Arrays. *Electron Device Letters, IEEE*, 32(2):191–193, 2011. doi: 10.1109/LED.2010.2090127. URL <http://dx.doi.org/10.1109/LED.2010.2090127>.
- K Roy, B Jung, D Peroulis, and A Raghunathan. Integrated Systems In The More-Than-Moore Era: Designing Low-Cost Energy-Efficient Systems Using Heterogeneous Components. *Design & Test, IEEE*, (99):1, 2013. doi: 10.1109/MDT.2011.49.

- URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5765911&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- Joshua Rubin, Ravishankar Sundararaman, Moonkyung Kim, and Sandip Tiwari. A Low-Voltage Torsion Nanorelay. *IEEE Electron Device Letters*, 32(3):414–416, February 2011. doi: 10.1109/LED.2010.2099199. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2011IEDL...32..414R&link_type=EJOURNAL.
- Dmitry Ruzmetov, Gokul Gopalakrishnan, Jiangdong Deng, Venkatesh Narayanamurti, and Shriram Ramanathan. Electrical triggering of metal-insulator transition in nanoscale vanadium oxide junctions. *Journal of Applied Physics*, 106(8):3702, October 2009. doi: 10.1063/1.3245338. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2009JAP...106h3702R&link_type=ABSTRACT.
- Lee Sang-Sun, Kim Jungha, Ahn Chang-Yong, and Lee Jong-Hoon. Operation Method Verification for Two-terminal Polymer RAM. *Journal of the Korean Physical Society*, 57(61):1816, December 2010. doi: 10.3938/jkps.57.1816. URL http://www.kps.or.kr/jkps/abstract_view.asp?articleid=37FD5A8B-C8F8-4759-B94F-4D9023C9BFE7.
- Y Sasago, M Kinoshita, T Morikawa, K Kurotsuchi, S Hanzawa, T Mine, A Shima, Y Fujisaki, H Kume, and H Moriya. Cross-point phase change memory with 4F 2 cell size driven by low-contact-resistivity poly-Si diode. pages 24–25, 2009. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5200620.
- L Savtchenko, A A Korokin, B N Engel, and N D Rizzo. Method of writing to scalable magnetoresistance random access memory element. US Patent Office, October 2001. URL http://www.google.com/patents/about?id=mqUMAAAAEBAJ&dq=Savtchenko+switching+toggle+ininventor:savtchenko&as_drrb_ap=q&as_drrb_is=q&num=10&ie=IS0-8859-1.
- T Schloesser, D Manger, R Weis, S Slesazeck, F Lau, S Tegen, M Sesterhenn, K Muemmler, J Nuetzel, D Temmler, B Kowalski, U Scheler, M Stavrev, and D Koehler. Highly scalable sub-50nm vertical double gate trench DRAM cell. In *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International*, pages 57–60, 2004. doi: 10.1109/IEDM.2004.1419064. URL <http://dx.doi.org/10.1109/IEDM.2004.1419064>.
- J C Scott and L D Bozano. Nonvolatile Memory Elements Based on Organic Materials. *Advanced Materials*, 19(11):1452–1463, June 2007. doi: 10.1002/adma.200602564. URL <http://doi.wiley.com/10.1002/adma.200602564>.
- E Seevinck, P J van Beers, and H Ontrop. Current-mode techniques for high-speed VLSI circuits with application to current sense amplifier for CMOS SRAM's. *IEEE Journal of Solid-State Circuits*, 26(4):525–536, April 1991. doi: 10.1109/4.75050. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=75050>.
- International SEMATECH. Semiconductor Device Reliability Failure Models. pages 1–34, May 2000.
- G Servalli. A 45nm generation Phase Change Memory technology. In *Electron Devices Meeting (IEDM), 2009 IEEE International*, pages 1–4, 2009. doi: 10.1109/IEDM.2009.5424409. URL <http://dx.doi.org/10.1109/IEDM.2009.5424409>.

- M A Shea and D F Smart. Tables of Asymptotic Directions and Vertical Cutoff Rigidities for a Five Degree by Fifteen Degree World Grid as Calculated Using the International Geomagnetic Reference Field for Epoch 1975.0. pages 1–166, April 2010.
- Premkishore Shivakumar and Norman P. Jouppi. Cacti 3.0: An integrated cache timing, power, and area model. Technical Report 2, Compaq Western Research Laboratory, 250 University Avenue Palo Alto, California 94301 USA, August 2001.
- B S Simpkins, M A Mastro, Jr C R Eddy, and P E Pehrsson. Surface depletion effects in semiconducting nanowires. *Journal of Applied Physics*, 103(10):104313, 2008. doi: 10.1063/1.2932072. URL <http://link.aip.org/link/?JAP/103/104313/1>.
- A Singhee. PTrace: Derivative-free local tracing of bicriterial design trade-offs. In *Computer-Aided Design (ICCAD), 2011 IEEE/ACM International Conference on*, pages 498–502, 2011. doi: 10.1109/ICCAD.2011.6105375. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6105375&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Radu Sion and Yao Chen. Fighting Mallory the Insider: Strong Write-Once Read-Many Storage Assurances. *Information Forensics and Security, IEEE Transactions on*, 7(2): 755–764, 2012. doi: 10.1109/TIFS.2011.2172207. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6046131>.
- Sklar. *Digital Communications: Fundamentals and Applications*. page 1104, January 2000.
- Clinton W Smullen, Vidyabhushan Mohan, Anurag Nigam, Sudhanva Gurusurthi, and Mircea R Stan. Relaxing non-volatility for fast and energy-efficient STT-RAM caches. In *2011 IEEE 17th International Symposium on High Performance Computer Architecture (HPCA)*, pages 50–61. IEEE. ISBN 978-1-4244-9432-3. doi: 10.1109/HPCA.2011.5749716. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5749716>.
- J I Sohn, S S Choi, S M Morris, J S Bendall, and H J Coles. Novel nonvolatile memory with multibit storage based on a ZnO nanowire transistor. *Nano ...*, 2010. URL <http://pubs.acs.org/doi/abs/10.1021/nl1013713>.
- Dong Ick Son, Jae Ho Shim, Dong Hee Park, Jae Hun Jung, Jung Min Lee, Won Il Park, Tae Whan Kim, and Won Kook Choi. Polymer-ultrathin graphite sheet-polymer composite structured flexible nonvolatile bistable organic memory devices. *Nanotechnology*, 22(29):295203–295203, July 2011. doi: 10.1088/0957-4484/22/29/295203. URL <http://iopscience.iop.org/0957-4484/22/29/295203>.
- Hyunwook Song, Mark A Reed, and Takhee Lee. Single Molecule Electronic Devices. *Advanced Materials*, 23(14):1583–1608, February 2011. doi: 10.1002/adma.201004291. URL <http://doi.wiley.com/10.1002/adma.201004291>.
- Ki-Whan Song, Jin-Young Kim, Jae-Man Yoon, Sua Kim, Huijung Kim, Hyun-Woo Chung, Hyungi Kim, Kanguk Kim, Hwan-Wook Park, Hyun Chul Kang, Nam-kyun Tak, Dukha Park, Woo-Seop Kim, Yeong-Taek Lee, Yong Chul Oh, Gyo-Young Jin,

- Jeihwan Yoo, Donggun Park, Kyungseok Oh, Changhyun Kim, and Young-Hyun Jun. A 31 ns Random Cycle VCAT-Based 4F DRAM With Manufacturability and Enhanced Cell Efficiency. *Solid-State Circuits, IEEE Journal of*, 45(4):880–888, 2010a. doi: 10.1109/JSSC.2010.2040229. URL <http://dx.doi.org/10.1109/JSSC.2010.2040229>.
- Sunghoon S Song, Byungjin B Cho, Tae-Wook TW Kim, Yongsung Y Ji, Minseok M Jo, Gunuk G Wang, Minhyeok M Choe, Yung Ho YH Kahng, Hyunsang H Hwang, and Takhee T Lee. Three-dimensional integration of organic resistive memory devices. *Advanced Materials (FRG)*, 22(44):5048–5052, November 2010b. doi: 10.1002/adma.201002575. URL <http://pubget.com/site/paper/20839254?institution=>.
- Scott Speaks. Reliability and MTBF Overview. pages 1–10, August 2005.
- G Stefanovich, A Pergament, and D Stefanovich. Electrical switching and Mott transition in VO₂. *Journal of Physics Condensed Matter*, 12(41):8837–8845, September 2000. doi: 10.1088/0953-8984/12/41/310. URL <http://stacks.iop.org/0953-8984/12/i=41/a=310?key=crossref.136e4e9f2278c04266016cddb7155440>.
- M Stiles and A Zangwill. Anatomy of spin-transfer torque. *Phys. Rev. B*, 66(1):014407, June 2002. doi: 10.1103/PhysRevB.66.014407. URL <http://link.aps.org/doi/10.1103/PhysRevB.66.014407>.
- Guangyu Sun, Xiangyu Dong, Yuan Xie, Jian Li, and Yiran Chen. High Performance Computer Architecture 2009. HPCA 2009. IEEE 15th International Symposium on 2009 Yiran Chen-2. *High Performance Computer Architecture, 2009. HPCA 2009. IEEE 15th International Symposium on*, pages 239–249, December 2008. doi: 10.1109/HPCA.2009.4798259. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=4798259&openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526filter%253DAND%2528NOT%25284283010803%2529%2529%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All%2526queryText%253D%2528%2528DOI%253A10.1109%252FHPCA.2009.4798259%2529%2529.
- Guangyu Sun, Xiangyu Dong, Yuan Xie, Jian Li, and Yiran Chen. A novel architecture of the 3D stacked MRAM L2 cache for CMPs. *High Performance Computer Architecture, 2009. HPCA 2009. IEEE 15th International Symposium on*, pages 239–249, 2009a. URL [10.1109/HPCA.2009.4798259](http://dx.doi.org/10.1109/HPCA.2009.4798259).
- Guangyu Sun, Xiangyu Dong, Yuan Xie, Jian Li, and Yiran Chen. A novel architecture of the 3d stacked mram l2 cache for cmps. In *High Performance Computer Architecture, 2009. HPCA 2009. IEEE 15th International Symposium on*, pages 239–249, feb. 2009b. doi: 10.1109/HPCA.2009.4798259.
- Guangyu Sun, Huazhong Yang, and Yuan Xie. Performance/Thermal-Aware Design of 3D-Stacked L2 Caches for CMPs. *Transactions on Design Automation of Electronic Systems (TODAES)*, 17(2), April 2012a. doi: 10.1145/2159542.2159545. URL <http://portal.acm.org/citation.cfm?id=2159542.2159545&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Guangyu Sun, Yaojun Zhang, Yu Wang, and Yiran Chen. Improving energy efficiency of write-asymmetric memories by log style write. In *ISLPED '12: Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics*

- and design. ACM Request Permissions, July 2012b. doi: 10.1145/2333660.2333705. URL <http://portal.acm.org/citation.cfm?id=2333660.2333705&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Hongbin Sun, Chuanyin Liu, Nanning Zheng, Tai Min, and Tong Zhang. Design techniques to improve the device write margin for MRAM-based cache memory. In *GLSVLSI '11: Proceedings of the 21st edition of the great lakes symposium on Great lakes symposium on VLSI*. ACM Request Permissions, May 2011a. doi: 10.1145/1973009.1973030. URL <http://portal.acm.org/citation.cfm?id=1973009.1973030&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- J Z Sun, D J Monsma, T S Kuan, M J Rooks, D W Abraham, B Oezylmaz, A D Kent, and R H Koch. Spin-torque transfer in batch-fabricated spin-valve magnetic nanojunctions (invited). *Journal of Applied Physics*, 93(10):6859, 2003. doi: 10.1063/1.1538170. URL <http://link.aip.org/link/JAPIAU/v93/i10/p6859/s1&Agg=doi>.
- Zhenyu Sun, Xiuyuan Bi, Hai Helen Li, Weng-Fai Wong, Zhong-Liang Ong, Xiaochun Zhu, and Wenqing Wu. Multi retention level STT-RAM cache designs with a dynamic refresh scheme. In *MICRO-44 '11: Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 329–338, Porto Alegre, Brazil, October 2011b. ACM Request Permissions. doi: 10.1145/2155620.2155659. URL <http://portal.acm.org/citation.cfm?id=2155620.2155659&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Zhenyu Sun, Xiuyuan Bi, and Hai Li. Process variation aware data management for STT-RAM cache design. In *ISLPED '12: Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design*. ACM Request Permissions, July 2012c. doi: 10.1145/2333660.2333706. URL <http://portal.acm.org/citation.cfm?id=2333660.2333706&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Zhenyu Sun, Xiang Chen, Yaojun Zhang, Hai Li, and Yiran Chen. Nonvolatile Memories as the Data Storage System for Implantable ECG Recorder. *Journal on Emerging Technologies in Computing Systems (JETC)*, 8(2), June 2012d. doi: 10.1145/2180878.2180885. URL <http://portal.acm.org/citation.cfm?id=2180878.2180885&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- J W Swonger, G Horvath, and A Singh. Radiation-Hardened Point-of-Load Buck Regulators for Space Applications. *Radiation Effects Data Workshop (REDW), 2012 IEEE*, pages 1–8, 2012. doi: 10.1109/REDW.2012.6353711. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6353711&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- Synopsys Inc. CCS Power Liberty Syntax. pages 1–22, October 2006. URL ftp://ftp.synopsys.com/doc/Power_Format_3.0/ccs_power_format_3.0.pdf.
- Synopsys Inc. CCS Timing Library Characterization Guidelines. pages 1–45, March 2008. URL ftp://ftp.synopsys.com/doc/Timing_Char_Guide_3.2/ccs_timing_char_guide_32.pdf.

- Synopsys Inc. CCS Timing Liberty Syntax. pages 1–22, May 2013a. URL ftp://ftp.synopsys.com/doc/Timing_Format_1.2/ccs_timing_format_1.2.pdf.
- Synopsys Inc. Liberty User Guide and Reference Manual . pages 1–1566, May 2013b.
- Minoru Takahashi. Magnetization Curve for the Half-Filled Hubbard Model. *Exactly Solvable Models Of Strongly Correlated Electrons. Series: Advanced Series in Mathematical Physics, ISBN: 978-981-02-1534-7. WORLD SCIENTIFIC, Edited by Vladimir E Korepin and Fabian HL Ebetaler, vol. 18, pp. 86-94, 18:86–94, 1994.* URL <http://books.google.com/books?hl=en&lr=&id=mvSerYmnLIkC&oi=fnd&pg=PA86&dq=Magnetization+Curve+for+the+Half+Filled+Hubbard+Model&ots=Y-nB2voxJq&sig=n3MjpcxYKPRC3z2TqdjYLJUHIKU>.
- D Takashima, H Shiga, D Hashimoto, T Miyakawa, S Shiratake, K Hoya, R Ogiwara, R Takizawa, S Doumae, R Fukuda, Y Watanabe, S Fujii, Tohru Ozaki, H Kanaya, S Shuto, K Yamakawa, I Kunishima, T Hamamoto, and A Nitayama. A Scalable Shield-Bitline-Overdrive Technique for Sub-1.5 V Chain FeRAMs. *Solid-State Circuits, IEEE Journal of*, 46(9):2171–2179, 2011. doi: 10.1109/JSSC.2011.2159053. URL <http://dx.doi.org/10.1109/JSSC.2011.2159053>.
- K Takeda, H Ikeda, Y Hagihara, M Nomura, and H Kobatake. Redefinition of Write Margin for Next-Generation SRAM and Write-Margin Monitoring Circuit. In *Solid-State Circuits Conference, 2006. ISSCC 2006. Digest of Technical Papers. IEEE International*, pages 2602–2611, 2006. doi: 10.1109/ISSCC.2006.1696326. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1696326&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- P J Tallerico. Development of a RAMI model for LANSCE and high power APT accelerators. *AIP Conference Proceedings*, 337:395, 1995.
- Simon Tam. Single Error Correction and Double Error Detection. page 12, August 2006. URL http://www.xilinx.com/support/documentation/application_notes/xapp645.pdf.
- Fern Nee Tan, Sze Geat Pang, Chin Leng Ng, Kam Yee Wong, and Lee Kee Yong. SRAM Core Modeling Methodology for Efficient Power Delivery Analysis. In *SoC Design Conference (ISOCC), 2009 International*, pages 224–227, 2009. doi: 10.1109/SOCCDC.2009.5423803. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5423803&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- C T Tanaka and J S Moodera. Spin polarized tunneling in halfmetallic ferromagnets (abstract). *Journal of Applied Physics*, 79(8):6265, 1996. doi: 10.1063/1.362029. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5014839.
- M Tang, Xiaolei Xu, Zhi Ye, Y Sugiyama, and Hiroshi Ishiwara. Impact of Hf-TaO Buffer Layer on Data Retention Characteristics of Ferroelectric-Gate FET for Nonvolatile Memory Applications. *Electron Devices, IEEE Transactions on*, 58(2):370–375, 2011. doi: 10.1109/TED.2010.2090883. URL [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5659901&matchBoolean%](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5659901&matchBoolean%26queryText%3D%28%22SRAM+power%22%29)

- 3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22Impact+of+HfTa0+buffer+layer+on+data+retention+characteristics+of+ferroelectric-gate+FET+for+nonvolatile+memory+application%22%29.
- David Tarjan, Shyamkumar Thoziyoor, and Norman P. Jouppi. Cacti 4.0. Technical Report 86, HP Laboratories Palo Alto, Hewlett-Packard Development Company, L.P., June 2006.
- Tasuku Nagami, Yoshishige Tsuchiya, Ken Uchida, Hiroshi Mizuta, and Shunri Oda. Scaling Analysis of Nanoelectromechanical Memory Devices. *Japanese Journal of Applied Physics*, 49(4):4304, April 2010. doi: 10.1143/JJAP.49.044304. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010JaJAP...49d4304T&link_type=ABSTRACT.
- M Terai, S Kotsuji, H Hada, Noriyuki Iguchi, Toshinari Ichihashi, and S Fujieda. Effect of ReRAM-stack asymmetry on read disturb immunity. *Reliability Physics Symposium, 2009 IEEE International*, pages 134–138, 2009. doi: 10.1109/IRPS.2009.5173238. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5173238&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28ReRAM+AND+RRAM%29.
- Ioannis Theodonis, Nicholas Kioussis, Alan Kalitsov, Mairbek Chshiev, and W Butler. Anomalous Bias Dependence of Spin Torque in Magnetic Tunnel Junctions. *Physical Review Letters*, 97(23):237205, December 2006. doi: 10.1103/PhysRevLett.97.237205. URL <http://link.aps.org/doi/10.1103/PhysRevLett.97.237205>.
- S Thiel, G Hammerl, A Schmehl, C W Schneider, and J Mannhart. Tunable Quasi-Two-Dimensional Electron Gases in Oxide Heterostructures. *Science*, 313(5):1942–1945, September 2006. doi: 10.1126/science.1131091. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2006Sci...313.1942T&link_type=ABSTRACT.
- S Thoziyoor, Jung Ho Ahn, M Monchiero, J B Brockman, and N P Jouppi. A Comprehensive Memory Modeling Tool and Its Application to the Design and Analysis of Future Memory Hierarchies. *Computer Architecture, 2008. ISCA '08. 35th International Symposium on*, pages 51–62, 2008a. doi: 10.1109/ISCA.2008.16. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4556715&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22memory+hierarchy%22+AND+%28power+AND+L2%29%29.
- Shyamkumar Thoziyoor, Naveen Muralimanohar, Jung Ho Ahn, and Norman P. Jouppi. Cacti 5.1. Technical Report 20, HP Laboratories, Palo Alto, Hewlett-Packard Development Company, L.P., April 2008b.
- Haruki Toda. Three-dimensional programmable resistance memory device with a read/write circuit stacked under a memory cell array. Technical report, January 2012.
- Lionel Torres and Weisheng Zhao. Magnetic memory (MRAM), a new area for 2D and 3D SoC/SiP design. In *GLSVLSI '11: Proceedings of the 21st edition of the great lakes symposium on Great lakes symposium on VLSI*. ACM, May 2011. doi: 10.1145/1973009.1973103. URL <http://portal.acm.org/citation.cfm?id=1973009.1973103&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.

- Lionel Torres, Yoann Guillemenet, and Syed Zahid Ahmed. A Dynamic Reconfigurable MRAM based FPGA. *International Conference on Engineering of Reconfigurable Systems and Algorithms (ERSA 2010)*, Las Vegas, USA., pages 31–40, 2010.
- Lionel Torres, Raphael Martins Brum, Luís Vitório Cargnini, and Gilles Sassatelli. Trends on the application of emerging nonvolatile memory to processors and programmable devices. *Circuits and Systems (ISCAS), 2013 IEEE International Symposium on*, pages 101–104, 2013. doi: 10.1109/ISCAS.2013.6571792. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6571792&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Authors%3ACargnini%29.
- James M JM Tour, Long L Cheng, David P DP Nackashi, Yuxing Y Yao, Austen K AK Flatt, Sarah K SK St Angelo, Thomas E TE Mallouk, and Paul D PD Franzon. NanoCell electronic memories. *Journal of the American Chemical Society*, 125(43):13279–13283, October 2003. URL <http://pubget.com/site/paper/14570505?institution=>.
- TSMC. Tsmc expects to enter 45nm production in september, September 2007. URL <http://www.tsmc.com/tsmcdotcom/PRListingNewsAction.do?action=detail&&newsid=2120>.
- TSMC. Tsmc first to deliver 40nm process technology, March 2008. URL <http://www.tsmc.com/tsmcdotcom/PRListingNewsAction.do?action=detail&language=E&newsid=2561&newsdate=...>
- TSMC and Synopsys Inc. Composite Current Source Model Accuracy Study. pages 1–9, July 2007.
- K Tsunoda, H Noshiro, C Yoshida, Y Yamazaki, A Takahashi, Y Iba, A Hatada, M Nakabayashi, T Takenaga, M Aoki, and T Sugii. A novel MTJ for STT-MRAM with a dummy free layer and dual tunnel junctions. *Electron Devices Meeting (IEDM), 2012 IEEE International*, page 29, 2012. doi: 10.1109/IEDM.2012.6479126. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6479126&matchBoolean%3Dtrue%26pageNumber%3D5%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM%29.
- T Tuan and B Lai. Leakage power analysis of a 90nm FPGA. In *CICC Custom Integrated Circuits Conference*, pages 57–60. IEEE, 2003. ISBN 0-7803-7842-3. doi: 10.1109/CICC.2003.1249359. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1249359>.
- A A Tulapurkar, Y Suzuki, A Fukushima, H Kubota, H Maehara, K Tsunekawa, D D Djayaprawira, N Watanabe, and S Yuasa. Spin-torque diode effect in magnetic tunnel junctions. *Nature*, 438(7066):339–342, November 2005. doi: 10.1038/nature04207. URL <http://www.nature.com/doifinder/10.1038/nature04207>.
- L L Vadasz, A S Grove, T A Rowe, and G E Moore. Silicon-gate technology. *Spectrum, IEEE*, 6(10):28–35, 1969. doi: 10.1109/MSPEC.1969.5214116. URL <http://dx.doi.org/10.1109/MSPEC.1969.5214116>.

- A Valaee and A J Al-Khalili. SRAM read-assist scheme for high performance low power applications. In *SoC Design Conference (ISOCC), 2011 International*, pages 179–182, 2011. doi: 10.1109/ISOCC.2011.6138676. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6138676&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- A Valaee and A J Al-Khalili. High-performance low-power sensing scheme for nanoscale SRAMs. *Computers & Digital Techniques, IET*, 6(6):406–413, 2012. doi: 10.1049/iet-cdt.2012.0038. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6403644&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28SRAM+AND+%28%22leakage+power%22+AND+%28%22low+power%22+AND+%22high+performance%22%29%29%29.
- Ilia Valov, Rainer Waser, John R Jameson, and Michael N Kozicki. Electrochemical metallization memories—fundamentals, applications, prospects. *Nanotechnology*, 22(25):254003–254003, June 2011. doi: 10.1088/0957-4484/22/25/254003. URL <http://iopscience.iop.org/0957-4484/22/25/254003>.
- Le Van Hai, Mitsue Takahashi, and Shigeki Sakai. Fabrication and characterization of sub-0.6- μm ferroelectric-gate field-effect transistors. *Semiconductor Science and Technology*, 25(11):115013, October 2010. doi: 10.1088/0268-1242/25/11/115013. URL <http://stacks.iop.org/0268-1242/25/i=11/a=115013?key=crossref.136d64db7e00f3eac529c3742592faa0>.
- A J van Roosmalen. Roadmap challenges - there is more than Moore [integrated system complexity growth]. In *Thermal and Mechanical Simulation and Experiments in Microelectronics and Microsystems, 2004. EuroSimE 2004. Proceedings of the 5th International Conference on, 2004*. doi: 10.1109/ESIME.2004.1304012. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1304012&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- Rangharajan Venkatesan, Vivek Kozhikkottu, Charles Augustine, Arijit Raychowdhury, Kaushik Roy, and Anand Raghunathan. TapeCache: a high density, energy efficient cache based on domain wall memory. In *ISLPED '12: Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design*. ACM Request Permissions, July 2012. doi: 10.1145/2333660.2333707. URL <http://portal.acm.org/citation.cfm?id=2333660.2333707&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- B Vigna. More than Moore: micro-machined products enable new applications and open new markets. *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, page 8, 2005. doi: 10.1109/IEDM.2005.1609252. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1609252&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- William J Vigrass. Calculation of Semiconductor Failure Rates. pages 1–5, June 2010.

- K von Arnim, E Augendre, C Pacha, T Schulz, K T San, F Bauer, A Nackaerts, R Rooyackers, T Vandeweyer, B Degroote, N Collaert, A Dixit, R Singanamalla, W Xiong, A Marshall, C R Cleavelin, K Schrufer, and M Jurczak. A Low-Power Multi-Gate FET CMOS Technology with 13.9ps Inverter Delay, Large-Scale Integrated High Performance Digital Circuits and SRAM. In *2007 IEEE Symposium on VLSI Technology*, pages 106–107. IEEE, 2007. ISBN 978-4-900784-03-1. doi: 10.1109/VLSIT.2007.4339745. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4339745>.
- Ching-Hua Wang, Yi-Hung Tsai, Kai-Chun Lin, Meng-Fan Chang, Ya-Chin King, Chrong-Jung Lin, Shyh-Shyuan Sheu, Yu-Sheng Chen, Heng-Yuan Lee, F T Chen, and Ming-Jinn Tsai. Three-dimensional 4F₂ ReRAM cell with CMOS logic compatible process. In *Electron Devices Meeting (IEDM), 2010 IEEE International*, pages 29.6.1–29.6.4, 2010a. doi: 10.1109/IEDM.2010.5703446. URL <http://dx.doi.org/10.1109/IEDM.2010.5703446>.
- Jue Wang, Xiangyu Dong, and Yuan Xie. OAP: an obstruction-aware cache management policy for STT-RAM last-level caches. In *DATE '13: Proceedings of the Conference on Design, Automation and Test in Europe*. EDA Consortium, March 2013. URL <http://portal.acm.org/citation.cfm?id=2485288.2485493&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- K L Wang and P K Amiri. Spintronics for instant-on nonvolatile electronics. In *Optoelectronic and Microelectronic Materials & Devices (COMMAD), 2012 Conference on*, pages 117–118, 2012. doi: 10.1109/COMMAD.2012.6472388. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6472388&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MTJ+AND+%28MRAM+AND+STT%29%29.
- K L Wang, A Khitun, and K Galatsis. More Than Moore's Law: Nanofabrics and Architectures. *Bipolar/BiCMOS Circuits and Technology Meeting, 2007. BCTM '07. IEEE*, pages 139–143, 2007. doi: 10.1109/BIPOL.2007.4351855. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4351855&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- Y Wang, U Bhattacharya, F Hamzaoglu, P Kolar, Y Ng, L Wei, Y Zhang, K Zhang, and M Bohr. A 4.0 GHz 291Mb voltage-scalable SRAM design in 32nm high-κ metal-gate CMOS with integrated power management. In *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International*, pages 456–457, 2009a. doi: 10.1109/ISSCC.2009.4977505. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4977505.
- Y F Wang, W T Hsieh, Y F Chiu, and S Y Hsu. Patent US7475367 - Memory power models related to access information and methods thereof - Google Patents. US Patent Office, 2009b. URL <http://www.google.com/patents?hl=en&lr=&vid=USPAT7475367&id=h3zVAAAAEBAJ&oi=fnd&dq=SRAM+power+consumption&printsec=abstract>.
- Yih Wang, Uddalak Bhattacharya, Fatih Hamzaoglu, Pramod Kolar, Yong-Gee Ng, Liqiong Wei, Ying Zhang, Kevin Zhang, and Mark Bohr. A 4.0 GHz 291 Mb Voltage-Scalable SRAM Design in a 32 nm High-k + Metal-Gate CMOS Technology With

- Integrated Power Management. *Solid-State Circuits, IEEE Journal of*, 45(1):103–110, 2010b. doi: 10.1109/JSSC.2009.2034082. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=5357552&queryText%253D%2528%2528Document+Title%253AA+4.0+GHz+291+Mb+Voltage+Scalable+SRAM+Design+in+a+32+nm+High-k+++Metal-Gate+CMOS+Technology+With+Integrated+Power+Management%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Yongge Wang and Yuliang Zheng. Fast and Secure Magnetic WORM Storage Systems. In *Security in Storage Workshop, 2003. SISW '03. Proceedings of the Second IEEE International*, pages 11–11, May 2004. doi: 10.1109/SISW.2003.10002. URL <http://dx.doi.org/10.1109/SISW.2003.10002>.
- Yuhao Wang, Chun Zhang, Hao Yu, and Wei Zhang. Design of low power 3D hybrid memory by non-volatile CBRAM-crossbar with block-level data-retention. In *ISLPED '12: Proceedings of the 2012 ACM/IEEE international symposium on Low power electronics and design*. ACM Request Permissions, July 2012. doi: 10.1145/2333660.2333709. URL <http://portal.acm.org/citation.cfm?id=2333660.2333709&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- R Waser. *Nanoelectronics and Information Technology*. Wiley, 2012. ISBN 9783527409273. URL <http://books.google.fr/books?id=1PgYS7zDCM8C>.
- R Waser, R Dittmann, G Staikov, and K Szot. RedoxBased Resistive Switching Memories—Nanoionic Mechanisms, Prospects, and Challenges. *Advanced Materials*, 2009. URL <http://onlinelibrary.wiley.com/doi/10.1002/adma.200900375/full>.
- Wujie Wen, Yaojun Zhang, Yiran Chen, Yu Wang, and Yuan Xie. PS₃-RAM: a fast portable and scalable statistical STT-RAM reliability analysis method. In *DAC '12: Proceedings of the 49th Annual Design Automation Conference*. ACM Request Permissions, June 2012. doi: 10.1145/2228360.2228580. URL <http://portal.acm.org/citation.cfm?id=2228360.2228580&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- M Wilczyński, J Barnas, and R Świrkowicz. Free-electron model of current-induced spin-transfer torque in magnetic tunnel junctions. *Phys. Rev. B*, 77(5):054434, February 2008. doi: 10.1103/PhysRevB.77.054434. URL <http://link.aps.org/doi/10.1103/PhysRevB.77.054434>.
- F C Williams and T Kilburn. A storage system for use with binary-digital computing machines. In *Proceedings of the IEE - Part II: Power Engineering*, pages 183–200, 1949. doi: 10.1049/pi-2.1949.0078. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5239649&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28p_Title%3A%22A+storage+system+for+use+with+binary+digital+computing+machines%22%29.
- Richard Wilson. Electronics Weekly News | Process R&D | TSMC pushes SiON process technology to 28nm for SRAM, June 2009. URL <http://www.electronicweekly.com/news/research/process-rd/tsmc-pushes-sion-process-technology-to-28nm-for-sram-2009-06/>.

- Steven J E Wilton and Norman P Jouppi. An Enhanced Access and Cycle Time Model for On-Chip Caches. Technical report, 250 University Avenue Palo Alto, California 94301 USA, July 1993.
- P Woerlee. Electrical device and method of manufacturing therefore. Technical report, 2005.
- S A Wolf, D D Awschalom, R A Buhrman, J M Daughton, S Von Molnár, M L Roukes, A Y Chtchelkanova, and D M Treger. Spintronics: a spin-based electronics vision for the future. *Science*, 294(5546):1488–1495, 2001. URL <http://www.ncbi.nlm.nih.gov/pubmed/11711666>.
- H S P Wong, Heng-Yuan Lee, Shimeng Yu, Yu-Sheng Chen, Yi Wu, Pang-Shiu Chen, Byoungil Lee, F T Chen, and Ming-Jinn Tsai. Metal–Oxide RRAM. In *Proceedings of the IEEE*, pages 1951–1970, 2012. doi: 10.1109/JPROC.2012.2190369. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6193402&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28ReRAM+AND+RRAM%29.
- H S Philip Wong, Simone Raoux, SangBum Kim, Jiale Liang, John P Reifenberg, Bipin Rajendran, Mehdi Asheghi, and Kenneth E Goodson. Phase Change Memory. *Proceedings of the IEEE*, 98(12):2201–2227, 2010. doi: 10.1109/JPROC.2010.2070050. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5609179>.
- Chen Wu, Li-Jun Zhang, Yong Wang, and Jian-Bin Zheng. SRAM power optimization with a novel circuit and architectural level technique. In *Solid-State and Integrated Circuit Technology (ICSICT), 2010 10th IEEE International Conference on*, pages 687–689, 2010. doi: 10.1109/ICSICT.2010.5667374. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5667374&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22SRAM+power%22%29.
- W Wu, G Y Jung, D L Olynick, J Straznicki, Z Li, X Li, D A A Ohlberg, Y Chen, S Y Wang, J A Liddle, W M Tong, and R Stanley Williams. One-kilobit cross-bar molecular memory circuits at 30-nm half-pitch fabricated by nanoimprint lithography. *Applied Physics A*, 80(6):1173–1178, March 2005. doi: 10.1007/s00339-004-3176-y. URL <http://link.springer.com/10.1007/s00339-004-3176-y>.
- Wenfeng Xiang and Chengkuo Lee. Nanoelectromechanical torsion switch of low operation voltage for nonvolatile memory application. *Applied Physics Letters*, 96(19):193113, 2010. doi: 10.1063/1.3428781. URL <http://link.aip.org/link/APPLAB/v96/i19/p193113/s1&Agg=doi>.
- J Xiao, GEW Bauer, and A Brataas. Phys. Rev. B 77, 224419 (2008): Spin-transfer torque in magnetic tunnel junctions: Scattering theory. *Physical Review B*, 2008. URL <http://prb.aps.org/abstract/PRB/v77/i22/e224419>.
- Xilinx. Virtex FPGA Series Configuration and Readback. page 37, March 2005. URL http://www.xilinx.com/support/documentation/application_notes/xapp138.pdf.
- Xilinx. Xilinx UG116 Device Reliability Report. pages 1–106, May 2010.

- Liu Xing-Hua, Lu Wen-Sheng, Ji Zhuo-Yu, Tu De-Yu, Zhu Xiao-Li, Xie Chang-Qing, and Liu Ming. Fabrication of a 256-bits organic memory by soft x-ray lithography. *Chinese Physics B*, 19(5):057204, May 2010. doi: 10.1088/1674-1056/19/5/057204. URL <http://stacks.iop.org/1674-1056/19/i=5/a=057204?key=crossref.c2f638132cc59cd69936730f1db3d582>.
- Feng Xiong, Albert D Liao, David Estrada, and Eric Pop. Low-Power Switching of Phase-Change Materials with Carbon Nanotube Electrodes. *Science Magazine*, 332(6029):568–570, April 2011. doi: 10.1126/science.1201938. URL <http://dx.doi.org/10.1126/science.1201938>.
- Cong Xu, Xiangyu Dong, N P Jouppi, and Yuan Xie. Design implications of memristor-based RRAM cross-point structures. *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2011*, pages 1–6, 2011. URL http://ieeexplore.ieee.org/search/srchabstract.jsp?tp=&arnumber=5763125&queryText%253D%2528%2528mram%2529%2529%2526openedRefinements%253D*%2526sortType%253Ddesc_Publication+Year%2526matchBoolean%253Dtrue%2526rowsPerPage%253D50%2526searchField%253DSearch+All.
- Chun Jason Xue, Youtao Zhang, Yiran Chen, Guangyu Sun, J Jianhua Yang, and Hai Li. Emerging non-volatile memories: opportunities and challenges. In *CODES+ISSS '11: Proceedings of the seventh IEEE/ACM/IFIP international conference on Hardware/software codesign and system synthesis*. ACM Request Permissions, October 2011a. doi: 10.1145/2039370.2039420. URL <http://portal.acm.org/citation.cfm?id=2039370.2039420&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- Kan-Hao Xue, Carlos A Paz de Araujo, Jolanta Celinska, and Christopher McWilliams. A non-filamentary model for unipolar switching transition metal oxide resistance random access memories. *Journal of Applied Physics*, 109(9):1602, May 2011b. doi: 10.1063/1.3581193. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2011JAP...109i1602X&link_type=ABSTRACT.
- B Yang, K D Buddharaju, S H G Teo, N Singh, G Q Lo, and D L Kwong. Vertical Silicon-Nanowire Formation and Gate-All-Around MOSFET. *Electron Device Letters, IEEE*, 29(7):791–794, 2008a. doi: 10.1109/LED.2008.2000617. URL <http://dx.doi.org/10.1109/LED.2008.2000617>.
- J Joshua Yang, Matthew D Pickett, Xuema Li, Douglas A A Ohlberg, Duncan R Stewart, and R Stanley Williams. Memristive switching mechanism for metal/oxide/metal nanodevices. *Nature Nanotechnology*, 3(7):429–433, June 2008b. doi: 10.1038/nnano.2008.160. URL <http://www.nature.com/doifinder/10.1038/nnano.2008.160>.
- J Joshua Yang, M X Zhang, John Paul Strachan, Feng Miao, Matthew D Pickett, Ronald D Kelley, G Medeiros-Ribeiro, and R Stanley Williams. High switching endurance in TaOx memristive devices. *Applied Physics Letters*, 97(2):2102, December 2010. doi: 10.1063/1.3524521. URL http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2010ApPhL..97w2102Y&link_type=ABSTRACT.
- H Yoda, S Fujita, N Shimomura, E Kitagawa, K Abe, K Nomura, H Noguchi, and J Ito. Progress of STT-MRAM technology and the effect on normally-off computing systems. *Electron Devices Meeting (IEDM), 2012 IEEE International*, page 11, 2012a. doi:

- 10.1109/IEDM.2012.6479023. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6479023.
- H Yoda, S Fujita, N Shimomura, E Kitagawa, K Abe, K Nomura, H Noguchi, and J Ito. Progress of STT-MRAM technology and the effect on normally-off computing systems. pages 11.3. 1–11.3. 4, 2012b. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6479023.
- H. Yoda, S. Fujita, N. Shimomura, E. Kitagawa, K. Abe, K. Nomura, H. Noguchi, and J. Ito. Progress of stt-mram technology and the effect on normally-off computing systems. In *Electron Devices Meeting, 2012. IEDM Technical Digest. IEEE International*, dec. 2012c.
- Hong Sik Yoon, In-Gyu Baek, Jinshi Zhao, Hyunjun Sim, Min Young Park, Hansin Lee, Gyu-Hwan Oh, Jong Chan Shin, In-Seok Yeo, and U-In Chung. Vertical cross-point resistance change memory for ultra-high density non-volatile memory applications. *VLSI Technology, 2009 Symposium on*, pages 26–27, 2009. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=5200621&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28ReRAM+AND+%28RRAM+AND+%28%22Cross+point%22%29%29.
- C Yoshida, T Ochiai, Y Iba, Y Yamazaki, K Tsunoda, A Takahashi, and T Sugii. Demonstration of non-volatile working memory through interface engineering in STT-MRAM. *VLSI Technology (VLSIT), 2012 Symposium on*, pages 59–60, 2012. doi: 10.1109/VLSIT.2012.6242460. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6242460&matchBoolean%3Dtrue%26pageNumber%3D6%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM%29.
- Chun-Yeol You. Reduced spin transfer torque switching current density with non-collinear polarizer layer magnetization in magnetic multilayer systems. *Applied Physics Letters*, 100(25):252413–252413–4, 2012. doi: 10.1063/1.4730376. URL <http://dx.doi.org/10.1063/1.4730376>.
- Hung-Chang Yu, Kai-Chun Lin, Ku-Feng Lin, Chin-Yi Huang, Yu-Der Chih, Tong-Chern Ong, J Chang, S Natarajan, and L C Tran. Cycling endurance optimization scheme for 1Mb STT-MRAM in 40nm technology. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2013 IEEE International*, pages 224–225, 2013. doi: 10.1109/ISSCC.2013.6487710. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6487710&matchBoolean%3Dtrue%26pageNumber%3D2%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28MRAM%29.
- Shinji Yuasa, Taro Nagahama, Akio Fukushima, Yoshishige Suzuki, and Koji Ando. Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. *Nature Publishing Group*, 3(12):868–871, October 2004. doi: 10.1038/nmat1257. URL <http://www.nature.com/doifinder/10.1038/nmat1257>.
- Wen Yueh, Minki Cho, and Saibal Mukhopadhyay. Perceptual quality preserving SRAM architecture for color motion pictures. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2013*, pages 103–108,

2013. doi: 10.7873/DATE.2013.035. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=6513481&matchBoolean%3Dtrue%26pageNumber%3D5%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22Static+Random+Access+Memory%22%29.
- A N M Zainuddin, S Hong, L Siddiqui, S Srinivasan, and S Datta. Voltage-controlled spin precession. *Phys. Rev. B*, 84:165306, October 2011. doi: 10.1103/PhysRevB.84.165306. URL <http://link.aps.org/doi/10.1103/PhysRevB.84.165306>.
- Morelos Zaragoza. URL <http://www.eccpage.com/bch3.c>.
- G Q Zhang. "More than Moore" - The changing international landscape, strategy and solutions of micro/nanoelectronics. In *Electronic Packaging Technology, 2007. ICEPT 2007. 8th International Conference on*, page 1, 2007. doi: 10.1109/ICEPT.2007.4441574. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=4441574&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- G Q Zhang, M Graef, and F van Roosmalen. The rationale and paradigm of "more than Moore". In *Electronic Components and Technology Conference, 2006. Proceedings. 56th*, 2006a. doi: 10.1109/ECTC.2006.1645639. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1645639&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- G Q Zhang, M Graef, and F van Roosmalen. Strategic Research Agenda of "More than Moore". In *Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro-Electronics and Micro-Systems, 2006. EuroSime 2006. 7th International Conference on*, pages 1–6, 2006b. doi: 10.1109/ESIME.2006.1644043. URL http://ieeexplore.ieee.org/xpl/articleDetails.jsp?tp=&arnumber=1644043&matchBoolean%3Dtrue%26rowsPerPage%3D30%26searchField%3DSearch_All%26queryText%3D%28%22More+than+Moore%22%29.
- Jishen Zhao and Yuan Xie. Optimizing bandwidth and power of graphics memory with hybrid memory technologies and adaptive data migration. In *ICCAD '12: Proceedings of the International Conference on Computer-Aided Design*. ACM Request Permissions, November 2012. doi: 10.1145/2429384.2429400. URL <http://portal.acm.org/citation.cfm?id=2429384.2429400&coll=DL&dl=ACM&CFID=333593502&CFTOKEN=20555755>.
- W S Zhao, Y Zhang, Y Lakys, J O Klein, D Etienne, D Revelosona, C Chappert, L Torres, L V Cargnini, R M Brum, Y Guillemenet, and G Sassatelli. Embedded MRAM for high-speed computing. In *VLSI and System-on-Chip (VLSI-SoC), 2011 IEEE/IFIP 19th International Conference on*, pages 37–42, 2011. doi: 10.1109/VLSISoC.2011.6081627. URL <http://dx.doi.org/10.1109/VLSISoC.2011.6081627>.
- Weisheng Zhao, Lionel Torres, LuísVitério Cargnini, RaphaelMartins Brum, Yue Zhang, Yoann Guillemenet, Gilles Sassatelli, Yahya Lakys, Jacques-Olivier Klein, Daniel Etienne, Dafiné Ravelosona, and Claude Chappert. High Performance SoC Design Using Magnetic Logic and Memory. In Salvador Mir, Chi-Ying Tsui, Ricardo

- Reis, and Oliver C S Choy, editors, *VLSI-SoC: Advanced Research for Systems on Chip*, pages 10–33. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-32769-8. doi: 10.1007/978-3-642-32770-4_2. URL http://dx.doi.org/10.1007/978-3-642-32770-4_2.
- Y Zheng, G X Ni, S Bae, C X Cong, and O Kahya. Wafer-scale graphene/ferroelectric hybrid devices for low-voltage electronics. *EPL (Europhysics ...)*, 2011. URL <http://iopscience.iop.org/0295-5075/93/1/17002>.
- Victor V Zhirnov, Ralph K Cavin, Stephan Menzel, Eike Linn, Sebastian Schmelzer, Dennis Brauhaus, Christina Schindler, and Rainer Waser. Memory Devices: Energy–Space–Time Tradeoffs. *Proceedings of the IEEE*, 98(12):2185–2200, 2010. doi: 10.1109/JPROC.2010.2064271. URL <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5597913>.
- Victor V Zhirnov, Roy Meade, Ralph K Cavin, and Gurtej Sandhu. Scaling limits of resistive memories. *Nanotechnology*, 22(25):254027, May 2011. doi: 10.1088/0957-4484/22/25/254027. URL <http://stacks.iop.org/0957-4484/22/i=25/a=254027?key=crossref.65214ab2b7f489616851232a9a41243f>.
- Jian Zhu, J A Katine, Graham E Rowlands, Yu-Jin Chen, Zheng Duan, Juan G Alzate, Pramey Upadhyaya, Juergen Langer, Pedram Khalili Amiri, Kang L Wang, and Ilya N Krivorotov. Voltage-Induced Ferromagnetic Resonance in Magnetic Tunnel Junctions. *arXiv.org*, May 2012. doi: 10.1103/PhysRevLett.108.197203. URL <http://arxiv.org/abs/1205.2835v2>.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*", also the references and schema of the thesis follows the "*CHICAGO: Manual of Style*". `classicthesis` is available for both \LaTeX and \LyX :

<http://code.google.com/p/classicthesis/>

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

<http://postcards.miede.de/>

Final Version as of February 4, 2014 (`classicthesis` version 7.3).

DECLARATION

Put your declaration here.

Montpellier, France, November 12, 2013

Luís Vitório Cargnini