Design & Analysis of Sleepy Instruction Caches for High Speed & Memory Efficient STT-RAM

S. Remya[1] S. Sathya[2] S. Shiny G. Deborah[3] G. Soniya[4] Mr. S. Shanmugaraju[5]
[1,2,3,4]Student [5]Assistant Professor
[1,2,3,4,5]Department of Electronics & Communication Engineering
[1,2,3,4,5]Dr. N. G. P. Institute of Technology, Coimbatore, TamilNadu, India

Abstract— The motto of this proposal is based on efficient memory optimization concept. For high speed STT-RAM, the sleepy instruction caches is used not only for write amount aware, but also for efficient read address enable solutions. To allow the selection of STT-RAM in the implementation of cache memories, new cache hierarchy levels of managing policies are required for overcoming such drawbacks. In this, we evaluated several cache hierarchy management policies in the context of STT-RAM L1 caches and STT-RAM L2 caches. We also found that the non-exclusive policy is superior to non-inclusive and exclusive policies in terms of energy consumption and endurance. We also proposed this sleepy instruction caches with sub block-based management policy because the write energy consumption and endurance are proportional and inversely proportional to the amount of written data respectively.

Key words: Sleepy Instruction Cache, Spin-Transfer Torque Random Access Memory (STT-RAM), Sub-Block, Cache Hierarchy, Write Amount

I. INTRODUCTION

Spin Transfer Torque RAM (STT-RAM) is a data storage device in the form of computer and it can be used to read and write data faster. Every hardware circuit will consume some static power even in the OFF condition. The objective is to increase the speed in peripheral circuits with the help of STT-RAM technology. Instead of fetching instructions from main memory, L1 instruction cache which is called as a tiny buffer loop cache is used to provide instructions to the processor to reduce the power consumption. The effective power down strategy and non- volatile characteristics of STT-RAM integrated with the loop caches, leads to reduced power consumption and high speed It is based on the key observation that only a small set of instructions is accessed inside a program loop. We propose to add a small static RAM cache called loop cache between the processor and the L1 instruction cache and L2 instruction cache made of STT-RAM. Whenever the loop cache has whole loop cached, the L1 instruction cache can be turned off in order to save energy and power during the execution of the loop. Experimental results show that the proposed approach achieves 49% reduction in energy consumption over the STT-RAM baseline.

II. RELATED WORK

Hooman Farkhani, et al.(2017) ,states that the write operation in the 1T-1MTJ STT-RAM bit-cell is asymmetric and stochastic, which leads to the high energy consumption and long latency. Therefore a write assist technique is proposed to terminate the magnetic tunnelling junction (MTJ) which leads to an error free write operation. Thus the proposed write assist technique leads to 81% energy saving compared with a cell without write assistant and adds only 9.6% area overhead to a 16-kbit STT-RAM.[1]

Hamed Farbeh,et al.(2016), states that STT-RAM (spin-transfer torque RAM) is an alternative to SRAM for L2 and L3 cache chip-implementation. Since there is no evaluation to determine the impact of ECCs on lifetime of STT-RAM caches,a Floating-ECC architecture increases the lifetime of the ST T-RAM caches was proposed by him. The major idea is to equally distribute the ECC write activity over all bits of cache lines by periodically redirecting the ECC bits inside the cache lines. Here the lifetime of L2 and L3 caches is increased by more than 318 percent and 254 percent respectively.[2]

Jianlei Yang, et al.(2016), states that the reliability of STT-MRAM is severely impacted by environmental disturbances because radiation strike on the access transistor could introduce potential write and read failures for 1T1MTJ cells. The simulation based on 3-D Metal Oxide Semi-conductor transistor modelling is first performed to capture the radiation-induced transient current pulse. Meanwhile, comprehensive wrote and sense circuits are evaluated for bit error rate analysis under random radiation effects and transistor process variation.[3]

Jaeyoung Park, et al.(2016), states that an energy-reduction strategy that overcomes stochastic switching of the spin-torque-transfer magnetic-RAM (STT-MRAM) write operation can be demonstrated and he proposed a circuit for write completion circuit.This variant tolerant design is used specfically for handling certain variation in process parameters for both FETs and magnetic tunnel junctions and the low value of tunnel magnetoresistance of practical MTJs is developed.[4]

Wujie Wen,et al.(2014), states that a fast and scalable semi-analytical method-PS3-RAM can be used for efficient statistical simulation in STT-RAM. PS3-RAM eliminates costly macro-magnetic and SPICE simulations and achieves more than 100 000* runtime speedup with excellent agreement with the result of conventional simulation method. This PS3-RAM can estimate the STT-RAM write error rate and write energy distribution at both magnetic tunnelling junction switching directions under different temperatures, demonstrating great potential in the analysis of reliable chracteristics of STT-RAM and write energy at the early design stage of memory and micro-architecture.[5]

Djaafar Chabi, et al.(2014), states that CMOS downscaling makes advanced memory and computing systems more vulnerable to radiation. Emerging non-volatile memories such as Magnetic RAM (MRAM) are under development to replace SRAM and Flash Memories. A rad-hard design was thus designed for STT-MRAM sensing circuit with low area overhead and negligible performance degradation. Both transient and Monte-Carlo simulation have

been performed to demonstrate its performance and behaviours.[6]

Yue Zhang, et al.(2013), states that Spin transfer torque(STT) with perpendicular magnetic and anisotropy (PMA) exhibits noticeable performance enhancements compared to that with in phase magnetic anisotropy. Therefore a spice-compact model of PMA-Magnetic Tunnel Junction integrating the stochastic behaviour of STT-RAM which is useful for finding the reliability issues during the design and simulation before process fabrication.[7]

Xiangyu Dong,et al.(2012),states that NVM (non-volatile memory) can be used to deploy the new NVM technologies such as STT-RAM (spin-torque-transfer memory), PCRAM (phase-change random-access memory) and ReRAM (resistive random-access memory) into multiple levels in the memory hierarchy. Therefore an NVSim was developed, which is a circuit-level model for NVM performance, energy and area estimation which supports these technologies. This NVSim was validated successfully against industrial NVM prototypes.[8]

Xuanyao Fong, et al.(2012), states that due to the problem of read and write requirements, there is a need of developing optimisation techniques for designing STT-MRAM bit-cells in order to minimise read and write failures by properly selecting the bit-cell configuration and properly accessing the transistor size. Therefore an optimization technique is designed which is applied to STT-MRAM bit-cells that is designed using 45nm bulk and 45nm silicon-insulator. Finally, predictions are made for optimised STT-MRAM bit-cells designed in 16nm predictive technology.[9]

Wei Xu, Hong bin Sun,et al. (2011),states that STT-RAM is inherently subject to a write latency vs read latency trade-off that is determined by the memory cell size. Here different memory cell-sized that impact the overall computing system performance is studied. Leveraging MTJ device switching characteristics, an STT RAM architecture design method is proposed using CACTI-based memory modelling and computing system performance simulations using Simple Scalar.[10]

Yiran Chen,et al.(2010),states that a magnetic and electric level spin-transfer torque random access memory cell model can be used to stimulate the write operation of an STT-RAM and a Dynamic design flow is proposed. A Dynamic design flow is also proposed. This design minimises unnecessary design margin in an STT-RAM cell by leveraging from the new STT-RAM cell model. Thus the STT-RAM Technology was scaled down to a 22-nm Bulk CMOS technology node.[11]

Author	Implementation detail	Remark
Kim et al. (2017)	Sub-block based management Policy(non-exclusive and non- Inclusive)	Reduction in speed with inefficient memory space
Farkhani et al.(2017)	Write assist technique to improve write characteristics	Limited only upto 81% reduction of energy consumption
Farbeh et al. (2016)	Floating-ECC architecture to increase the lifetime of caches	Limited write endurance
Yanget al. (2016)	Comprehensive framework is pro-posed for radiation-induced error analysis of STT-MRAM	Complicated effect arises on relationship between bias vol-tages and sense performance
Park et al.(2016)	STT-RAM architecture with write-completion monitoring	Reduction of energy consumption ranges from 36%-97% only
Chabi et al.(2014)	Novel rad-hard STT-MTJ sensing circuit	Higher energy consumption in rad- hard compared to conventional
Wen et al.(2014)	Fast and scalable statistical STT- RAM reliability/energy analysis Method-PS3-RAM	Speedup is limited only for conventional stage
Zhang et al.(2013)	Compact model of CoFeB/MgO/ CoFeB PMA MTJ for performance analysis and optimization	Increased error probability long-pulse current
Fong et al.(2012)	Bit-cell optimization technique	Reduced failure probability and better yield only at iso-area
Dong et al.(2012)	NVSim, circuit-level model for NVM performance, energy, area	Limited only for NVM tech-nologies
Xu et al.(2011)	Embedded STT-RAM and dual- write-speed STT-RAM	Energy saving is limited only upto30% for all benchmarks
C h e n et al.(2010)	Magnetic and electrical level STT-RAM cell model and dynamic design flow is proposed	Cell area is limited due to the layout rule
Kim et al.(2009)	Area-efficient error protection architecture is implemented	Increase in memory traffic due to small write backs

Table 1: Comparison of Energy Consumption & Speed

III. CONCLUSION

The simple architecture of Spin Transfer Torque - Performance Enhancement Guaranteed Cache is presented. It reduces the static energy consumption of STT-RAM instruction caches with high speed. The improved loop caches supports all kind of loops, and based on it, controlled sleep mode power-gated L1 instruction cache is constructed.

Finally this architecture achieves low power consumption and high speed with the use of STT-RAM technology

REFERENCES

[1] H.Farkhani, M.Tohidi, A.Peiravi, J.Madsen, and F.Moradi,"STT-RAM energy reduction using self-referenced differential write termination technique," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 25, no. 2, Feb.2017.
[2] H. Farbeh, H. Kim, S.-G. Miremadi, and S. Kim, "Floating-ECC: Dynamic repositioning of error correcting code bits for extending the lifetime of non-volatile caches," IEEE Trans. Comput., doi: 10.1109/TC.2016.2557326, Dec. 2016.
[3] J. Yang, et al., "Radiation-induced soft error analysis of STTMRAM: A device to circuit approach," IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst., vol. 35, no. 3, pp. 380–393, Mar. 2016.
[4] J.Park,T.Zheng, M.Erez, and M.Orshansky, "Variation-Tolerant Write Completion Circuit for Variable-Energy Write STT-RAM Architecture," IEEE Trans. Very Large Scale Integr. (VLSI) Syst.,vol. 24, no. 4, apr.2016
[5] D. Chabi, W. Zhao, J.-O. Klein, "Design and analysis of radiation hardened sensing circuits for spin transfer torque magnetic memory and logic," IEEE Trans. Nucl.Sci., vol. 61, no. 6, pp. 3258–3264, Dec. 2014.
[6] W. Wen, Y. Zhang, Y. Chen, Y. Wang, and Y. Xie, "PS3-RAM: A fast portable and scalable statistical STT-RAM reliability/energy analysis method," IEEE Trans. Comput.-Aided Des. Int. Circuits Syst., vol. 33, no. 11, pp. 1644–1656, Nov. 2014.
[7] Y. Zhang, W. S. Zhao, G. Prenat, J-O. Klein, C. Chappert, B. Dieny, and D. Ravelosona, "Electrical modeling of stochastic spin transfer torque writing in magnetic tunnel junctions for memory and logic applications,"IEEETrans.Magn.,vol.49,pp.4375–4378,2013.
[8] X. Fong, S. H. Choday, and K. Roy, "Bit-cell level optimization for non-volatile memories using magnetic tunnel junctions and spintransfer torque switching," IEEE Trans. Nanotechnol., vol. 11, no. 1, pp. 172–181, Jan. 2012.
[9] X. Dong, C. Xu, Y. Xie, and N. P. Jouppi, "NVSim: A circuit-level performance, energy, and area model for emerging nonvolatile memory," IEEE Trans. Comput.-Aided Design Integr. Circuits Syst., vol. 31, no. 7, pp. 994–1007, Jul. 2012.
[10] W. Xu, H. Sun, Y. Chen, and T. Zhang, "Design of last-level on-chip cache using spin–torque transfer RAM (STT-RAM)," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 19, no. 3, pp. 483–493, Mar. 2011.
[11] Y. Chen et al., "Design margin exploration of spin-transfer torque RAM (STT-RAM) in scaled technologies," IEEE Trans. Very Large Scale Integr. (VLSI) Syst., vol. 18, no. 12, pp. 1724–1734, Dec. 2010.

