Recent progresses in STT/SOT-MRAMs for low power AI/IoT Processors

Tetsuo Endoh
(Tohoku University)

Introduction

STT-MRAM with Double CoFeB/MgO interface perpendicular MTJ has become the de facto standard technology. However, there are still issues for high density memory application and high reliable application. Furthermore, achievement of high speed and high endurance features is still challenging for STT-MRAM, because of its tradeoff relationship among the retention, endurance, and operational speed. Recently, Spin Orbit Torque (SOT) devices have been intensively researched and developed because SOT devices have the potential to achieve high speed and high endurance. To realize the practical use of SOT-MRAM for LSI applications, these issues must be addressed.

This invited speech reviews our recent progresses in STT-MTAM, SOT-MRAM for future AI/IoT processor and its systems that require ultra-low-power and high-performance computing at the same time.

STT-MRAM

We developed novel damage control integration process technologies including new low-damage MgO deposition process, low-damage RIE process, and low temperature cap process. By applying the developed damage control integration process technologies to double interface p-MTJ fabrication, TMR ratio, thermal stability factor, and switching efficiency of Double p-MTJ were successfully improved. Moreover, it was shown that despite the significant increase in thermal stability factor, the endurance of the fabricated Double p-MTJs was over 1010. Finally, with our double-interface p-MTJ technology and novel damage control integration process technologies, fabricated 128Mb STT-MRAMs successfully achieved 14ns/7ns write speed at Vdd of 1.2V/1.8V, respectively.

Next, for further scaling of STT-MRAM, we proposed novel Quad-interface p-MTJ technology which brings forth an increase of thermal stability factor compared with conventional Double-interface p-MTJ technology. We successfully fabricated the quad-interface MTJ using 300mm process based on the damage control integration process. The fabricated Quad p-MTJs achieved an enhancement of switching efficiency in addition to an approximately two times larger thermal stability factor without degradation of TMR ratio. The developed Quad p-MTJ technology will become an essential technology for the scaling of the STT-MRAM beyond 20nm without changing material and process sets from mass-production STT-MRAM. Moreover, the high reliable Quad p-MTJ technology with enough thermal stability factor is suitable for Automobile applications that require high temperature operation such as 150℃.

Field-free SOT-MRAM

To realize practical use of SOT-MRAM for LSI applications, we demonstrated 55 nm-CMOS/SOT-device hybrid MRAM cell with magnetic field free writing for the first time. For field free writing, we developed canted SOT device under 300 mm BEOL process with 400° C thermal tolerance. In addition, we developed its advanced process as follows; PVD process of SOT channel layer for high spin Hall angle under 400°C thermal tolerance, low damage RIE technology for high TMR/thermal stability factor, and ultra-smooth surface metal via process under SOT device to reduce contact resistance. By using the developed technologies, our canted SOT devices achieved fast write speed of 0.35 ns without magnetic field, an enough thermal stability factor of 70 for non-volatile memory (over 10 years retention), and a high TMR ratio of 167%, simultaneously. Moreover, we fabricated a field free canted SOT-MRAM cell with 55 nm CMOS technology and demonstrated its write/read performance. These technologies will open to high speed write non-volatile memory such as 1 level cache application of many kinds of application processors.

Reference

1) T. Endoh, “Embedded Nonvolatile Memory with STT-MRAMs and its Application for Nonvolatile Brain-Inspired VLSIs”, 2017 International Symposia on VLSI Technology, Systems and Applications. (*Invited*)
2) T. Endoh, "Nonvolatile Logic and Smart Nonvolatile Processors with CMOS/MTJ Hybrid Technology for IoT and AI (AIoT) Edge System", 2020 International Solid-State Circuits Conference (ISSCC). (*Invited*)
3) T. Endoh, H. Honjo, K. Nishioka and S. Ikeda, "Recent progresses in STT-MRAM and SOT-MRAM for next generation MRAM", 2020 Symposia on VLSI Technology and Circuits, TMFS.1 (*Invited*)

Long-Range Interlayer Chiral Exchange – Known and Unknown

Chi-Feng Pai
Department of Materials Science and Engineering, National Taiwan University

In recent years, there has been a surge of interest in the practical application of interlayer Dzyaloshinskii-Moriya interaction (DMI) in multilayers and in magnetic random-access memory (MRAM) devices, driven by its ability to facilitate long-range and chirally distinct spin textures. In several of our works [1-4], we experimentally investigated the characteristics and the possible origins of interlayer DMI. The growth condition has been identified as the key to generate robust interlayer DMI. An oscillatory interlayer DMI strength is observed with the cap layer through layer thickness dependence. This characteristic demonstrates the capability of optimizing the robust field-free switching, implementing such long-range interaction into practical spintronic devices. Furthermore, by exploiting the origin mechanism of interlayer DMI, innovative azimuthal symmetry engineering protocol enables tuning it through controlled deposition conditions of individual layers at wafer-scale.

Reference

1) Y.-H. Huang, C.-C. Huang, W.-B. Liao, T.-Y. Chen, and C.-F. Pai, “Growth-dependent Interlayer Chiral Exchange and Field-free Switching,” Physical Review Applied 18, 034046 (2022).
2) Y.-C. Li, Y.-H. Huang, C.-C. Huang, Y.-T. Liu, and C.-F. Pai, “Field-Free Switching in Symmetry-Breaking Multilayers: The Critical Role of Interlayer Chiral Exchange,” Physical Review Applied 20, 024032 (2023).
3) C.-Y. Lin, P.-C. Wang, Y.-H. Huang, W.-B. Liao, M.-Y. Song, X. Bao, and C.-F. Pai, “Field-free Spin-Orbit Torque Switching via Oscillatory Interlayer Dzyaloshinskii–Moriya Interaction for Advanced Memory Applications,” ACS Materials Letters 6, 400 (2024).
4) Y.-H. Huang, J.-H. Han, W.-B. Liao, C.-Y. Hu, Y.-T. Liu, and C.-F. Pai, “Tailoring Interlayer Chiral Exchange by Azimuthal Symmetry Engineering,” Nano Letters 24, 649 (2024).

Advanced MTJ technology for high-density cross-point STT-MRAM

M. Nakayama, R. Takashima, C. Kamata, M. Toko, S. Itai, T. Koike, H. Sugiyama
H. Kanaya, Y. M. Lee, and M. Saitoh
Frontier Technology R&D Institute, KIOXIA Corporation, Yokohama, 221-0022, Japan

Latest innovative technologies such as AI, DX and deep learning increase demands for high-speed non-volatile memory with much higher density and lower cost as compared to conventional DRAM. Spin-Transfer-Torque (STT)-MRAM using perpendicular magnetic anisotropy is a promising candidate for high-density MRAM [1, 2]. Recently, demonstrations of 1-selector/1-MTJ (1S1M) cells for high-density cross-point STT-MRAM have attracted much attention [3, 4]. Toward much higher density, higher speed, and lower cost of cross-point STT-MRAM, sustainable scaling and improvement in performance and reliability of MTJ are strongly required.

In this talk, we present recent progress in key MTJ technologies toward high-density cross-point STT-MRAM. For the scaled MTJ, we introduced a novel MTJ design and demonstrated high retention and high-speed writing simultaneously towards 1Z (15-14) nm STT-MRAM [5]. Key design concept of our MTJ, called as AccelHR-MTJ (Accelerated STT-Switching and High-Retention MTJ), is to assign the functions of high retention and high-speed writing to separate magnetic layers in a storage layer. We demonstrated excellent performance such as high retention of > 10 years at 90 ℃, high-speed writing at 5 ns pulse in our 14 nm AccelHR-MTJs as predicted by its design concept. Furthermore, large H_c of 4 kOe for strong magnetic immunity, large TMR ratio of 100 % with low RA of 1.7 Ωμm^2 for large read margin, and large H_{ex} of 14 kOe in SAF for stable STT switching were successfully achieved.

We also investigated the mechanism of the time-dependent degradation of MgO barrier in scaled MTJ [6]. The stress-time dependent degradation of resistance and TMR ratio was experimentally observed in tail bits under an applied voltage. The degradation can be theoretically explained by the generation of oxygen Frenkel defects at the Fe–MgO interface. A reduction in the initial oxygen vacancy in MgO is an effective method for suppressing degradation.

Our MTJ technologies provide the potential for high-density cross-point STT-MRAM.

Reference

[1] M. Nakayama *et al*., J. Appl. Phys. 103, 07A710 (2008).
[2] S.-W. Chung *et al*., 2016 International Electron Devices Meeting (IEDM), p.659.
[3] S. Seo *et al*., 2022 International Electron Devices Meeting (IEDM), pp. 218-221.
[4] E. Ambrosi *et al*., 2023 International Electron Devices Meeting (IEDM), 21-5.
[5] M. Nakayama *et al*., 2023 International Electron Devices Meeting (IEDM), 31-1.
[6] R. Takashima *et al*., 2024 IEEE International Reliability Physics Symposium (IRPS), P10.EM.

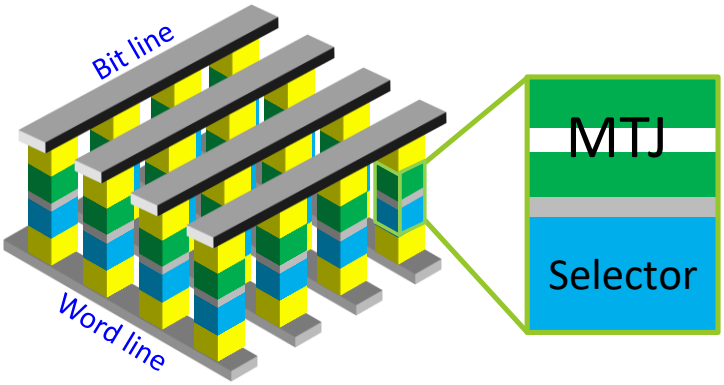

Fig.1: Schematic of cross-point STT-MRAM

Spin-Orbit Torque Based Domain Wall Motion Logic: Spin Torque Majority Gate

Dongryul Kim[1], Suhyeok An[1], Eunchong Baek[1], Jun-Su Kim[1], Soobeom Lee[2], June-Seo Kim[3], Chun-Yeol You[1]

[1]Department of Physics and Chemistry, DGIST, Daegu, South Korea

[2]Center of Basic Science, DGIST, Daegu, South Korea

[3]Division of Nanotechnology, DGIST, Daegu, South Korea

We conducted a comprehensive study on domain wall (DW) motion-based spin torque majority gate (STMG) for advanced logic gate applications. The majority gate, a pivotal functional logic gate, outputs a true signal when half or more of the inputs are true, and false otherwise. For instance, a three-input majority gate can dynamically function as an AND/OR gate based on the third input, offering a streamlined approach to circuit design for logic family devices such as full adders and MUX. Our implementation of a majority gate leveraged spin-orbit torque (SOT)-driven DW motion within Ta/Pt/Co/Ru heterostructures. Through our unique field-free SOT switching technique via SOT[1)], we successfully generated three input signals for eight distinct cases without reliance on an external magnetic field. Subsequently, we orchestrated SOT-induced DW motion seamlessly to execute the majority gate operations. To address critical requirements like cascading and fanout in logic gates, we adopted a sophisticated dynamic logic technique. Capitalizing on the inherent non-volatile properties of spin-based devices, our simulations indicate that STMG-based 32, 64, and 128-bit full adders exhibit superior performance compared to modern CMOS devices. Further in-depth insights into our research findings will be elucidated during the discussion.

Reference

1) S. An, *et al.* Appl. Phys. Lett. **120**, 262402 (2022).

Trends of Embedded MRAM IP Development for MCUs

Tomoya Saito
(Renesas Electronics Corporation, Tokyo, Japan)

IoT and AI technology are driving a paradigm shift toward a smart society. In the process of this shift, microcontroller units (MCUs) play a major role in a wide range of applications with secure and high-performance operation in home automation, robotics, and medical applications as well as intermittent low-energy operation in IoT endpoint applications. MCUs with embedded flash memories (eFlash) have the advantages in terms of security and faster boot load time without initial program code loading from external flash memories. Endpoint MCUs with eFlashs also contribute to low power operation powered by a battery or energy harvesting sources. On the other hand, it has become increasingly difficult to provide eFlash at advanced technology nodes such as 2Xnm and beyond because of its complex process steps, low affinity with advanced CMOS process, and the need for high-voltage transistors to support 10V-class write/erase voltages. Accordingly, embedded emerging memory have been developed and launched into mass production to replace eFlash thanks to fewer additional masks, BEOL process, and lower write voltage.
This presentation will provide an overview of emerging memory and forecast future trends. Circuit technology for high-temperature read and high-speed rewrite of embedded STT-MRAM (eMRAM) will be presented, as well as the evaluation results of test chips. In addition, future prospects for MRAM technology and other emerging memory technologies are also discussed.

Currently, MRAM, ReRAM, and PCM are available as embedded non-volatile memory (eNVM), and Yole's report predicts that the eNVM market will grow at more than two times per year between 2022 and 2028, with eMRAM leading the market growth for MCUs in particular [1]. This is because MRAM IP is available from major foundries for the 2Xnm generation and beyond, which is expected to replace eFlash in a various application.
Next, the evaluation results of test chip will be presented. There are several challenges in accelerating read speed and write throughput to enhances the performance of MCUs with eMRAM, and to realize low energy write for expanding new MCU applications. These are due to intrinsic characteristics of smaller read margin especially at high temperature, and write current variation of MRAM, respectively.
In this presentation, a high-precision boosted cross-coupled sense amplifier [2], global- and local-trimming with parallel-connected resistors, and cascode-clamp MOS scheme to achieve random read access frequency of over 200MHz at high temperature of 125°C will be introduced [3]. In addition, we will also present novel write schemes (a variable parallel bit write scheme and a self-termination write scheme) to enhance the advantage of eMRAM, achieving 10.4MB/s fast rewrite throughput and 65-69% lower write energy [3-4]. These achievements using 2Xnm and/or 1Xnm process technology will enable us to continuously provide advanced MCU products with embedded non-volatile memory to expand new MCU applications.
Finally, a future emerging memory technology will be discussed. Although MRAM leads eNVM market at present, other emerging memories also have the potential to change the situation drastically with the improvements in material development and so on. The next generation eNVM and expansion of its applications will be discussed from the perspective of memory IP design.

Reference
1) Yole Group, Emerging Non Volatile Memory 2023 Market & Technology Report
2) T. Shimoi et al., VLSI2022, pp.134-135.
3) T. Ogawa et al., ISSCC2024, pp.290-292.
4) T. Ito et al., IEDM 2021, pp. 2.2.1-2.2.4.

Spintronic security devices based on magnetic random-access memory

Soogil Lee[1], Jaimin Kang[2], and Byong-Guk Park[2]

[1]Department of Electronic Engineering, Gachon University, Seongnam 13120, Korea

[2]Department of Materials Science and Engineering, KAIST, Daejeon 34141, Korea

Security of information in Internet-of-Things (IoT) era is becoming a critical challenge because present software-based security technology is vulnerable to adversarial attacks. Recently, hardware-based security technology that utilizes natural randomness of materials and devices (e.g., physical unclonable function, PUF) has received much attention as an alternative to overcome this vulnerability [1]. In this study, we present a spintronic PUF utilizing field-free spin-orbit torque switching in ferromagnet (FM)/nonmagnet/FM trilayer structures [2]. This spintronic PUF exhibits ideal uniformity and uniqueness, which are essential PUF metrics. In addition, we also discuss the reconfigurability and reliability of this spintronic PUF. Furthermore, we demonstrate a magnetic random-access memory (MRAM) based PUF that consists 70-nm-diameter magnetic tunnel junctions [3]. We believe that our spintronic PUF offers excellent potential for enhancing the security of IoT applications because it could be compatible with current complementary-metal-oxide-semiconductor and MRAM technology.

Reference

1) Yansong Gao, Said F. Al-Sawari, and Derek Abbott, Nature Electronics 3, 81 (2020).
2) Soogil Lee et al, Advanced Materials 34, 2203558 (2022).
3) Jaimin Kang et al, ACS Nano 18, 12853 (2024).

MRAM growing into 3 terminal device

Tomoyuki Sasaki
Advanced Products Development Center, Technology & Intellectual Property HQ, TDK Corporation, Ichikawa, Chiba 272-8558, Japan

Recently, increased energy consumption with the development of AI is becoming a major social issue. The human perception by AI technology is used in our general life. According to Jones [1], the national energy consumption in ICT, which has a 2000 TWh in 2018, is rapidly increasing. The weight of data center is changing from traditional to hyperscale data center. In particular, the energy consumption in networks and data centers regions will be increased by 5 times in 2030 on a 2018 basis. Human society must suspend this power increase. One of the solutions is edge AI. Next generation AI is demanded low power consumption with keeping computing power at edges. Currently data from sensors were sent from edge to cloud AI, after that, answer is sent from cloud AI to edge. In this case, sending data band should be wide. If we realize edge AI which is mean the recognition and judgments within edge, we can achieve social level power consumption reduction and keeping personal security data. Recently, we can use embedded-MRAM in semiconductor circuits. Mass-produced MRAM is STT-MRAM which replaces NOR-Flash, is not enough for edge AI because of slower write speed and lower integrated density than DRAM and SRAM. To use MRAM for AI, advances in MRAM technology are required.

Present STT-MRAM is trying to achieve L3 cash level memory. The challenge is to improve the write endurance because tunnel barriers break down at a certain rate under a high speed write operation. One of the solutions is SOT-MRAM which is expected to be high endurance with high write speed [2]. SOT-MRAM must be compared and win that it is more suitable than an SRAM in case of cutting-edge node.

We have a different choice which is Spin-memristor in Neuromorphic device [3]. Neuromorphic device mimics human brain. It needs memristors and analog or spike signal computing. Memristor is defined as analog memory whose conductance depends on number of electrons passing through the element. Memristor is integrated in logic circuits. Its important properties are dynamic range, linearity, symmetry and non-volatility [4]. An ideal memristor should have a linear and symmetric response with a constant and small conductance change under each write pulses.

2-terminal STT-MRAM can grow into 3-terminal SOT-MRAM and Spin-memristor. The SRAM-class generation will be tipping point for MRAM.

Reference
1) N. Jones: Nature **561**, 163 (2018).
2) Y. Shiokawa, E. Komura, Y. Ishitani, A. Tsumita, K. Suda, Y. Kakinuma and T. Sasaki: AIP Advances **9**, 035236 (2019)
3) T. Shibata, T. Shinohara, T. Ashida, M. Ohta, K. Ito, S. Yamada, Y. Terasaki and T. Sasaki: Appl. Phys. Express **3**, 043004 (2020)
4) G. W. Burr *et al*., IEEE Trans. Elec. Dev. **62**, 3498 (2015)

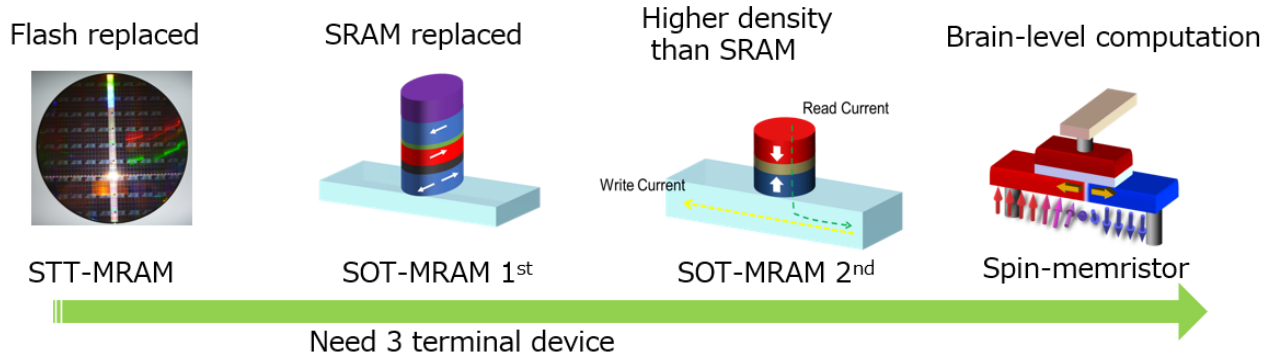

Fig. 1 MRAM Growth Roadmap